George Z. Voyiadjis
Editor

Handbook of Damage Mechanics

Nano to Macro Scale for Materials and Structures

Springer Reference

Handbook of Damage Mechanics

George Z. Voyiadjis
Editor

Handbook of Damage Mechanics

Nano to Macro Scale for Materials and Structures

With 699 Figures and 61 Tables

Springer Reference

Editor
George Z. Voyiadjis
Department of Civil and Environmental Engineering
Louisiana State University
Baton Rouge, LA, USA

ISBN 978-1-4614-5588-2 ISBN 978-1-4614-5589-9 (eBook)
ISBN 978-1-4614-5590-5 (print and electronic bundle)
DOI 10.1007/978-1-4614-5589-9
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2014948193

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

The field of damage mechanics has made considerable headway since its initial conception over 50 years ago as a means of associating changes of material compliance and strain localization with evolving creep damage of materials. Development of basic concepts in this field has evolved in several directions, with one stream focused on continuum thermodynamics based on damage field variables and another appealing to continuum micromechanics, attempting to account for explicit defects and microstructure and linking to cooperative behavior through elastic interactions. Although continuum damage mechanics and analytical/computational micromechanics of materials may appear to have taken distinct and parallel courses of development from the mid-1980s onward, the theoretical connections between these two paths are rich and deep, rooted in thermodynamics and kinetics. This is effectively expressed using internal state variables as a means to reduce the degrees of freedom of the model description while maintaining consistency with damaged material response at the level of a representative volume element.

This handbook thoroughly explores these connections, combining works in foundational topics such as fabric tensors to represent networks of cracks or other modes of damage with explicit modeling of microstructure and associated damage mechanisms. Authored by a series of leading international experts in the field, a unifying theme of the articles in this handbook is the treatment of structural-level degradation of response. Damage mechanics has been fruitfully applied as a practical approach for applications that involve complex aspects of distributed, evolving damage. Examples in this volume are varied:

- Low cycle fatigue via damage mechanics coupled with internal state variable crystal plasticity in polycrystalline metals
- Distributed particle/fiber cracking and interfacial debonding in particle- and fiber-reinforced composites, including both implicit and explicit consideration of damage modes
- Internal state variable modeling of distributed damage in polymers
- Damage evolution and failure of electronic materials and packaging
- Damage evolution and failure of materials subjected to extreme conditions, including dynamic loading conditions, irradiation in nuclear power plant components, and large-deformation metal forming
- Ductile fracture and damage localization in metals

In addition to applications, fundamental aspects of thermodynamics of damaged solids are addressed, including isotropic and anisotropic damage mechanics. Readers will also find substantial treatment of multiscale modeling and various approaches to homogenization of damage evolution processes. Emerging concepts in the evolution of damage, such as fractal theory and discrete damage models, round out theoretical aspects and explore the connectivity of damage mechanics with statistical physics.

Finally, the handbook provides an overview of recent experimental methods to measure evolving distributed damage in materials, with an emphasis on digital image correlation, as well as characterization and inverse modeling. The reader will find linkages between state-of-the-art theory, experiments, and applications in this volume, establishing an invaluable contribution to the literature.

I trust that you will find this *Handbook of Damage Mechanics* an inspiring and indispensable reference.

May 2014

David L. McDowell
Regents' Professor and Carter N. Paden,
Jr. Distinguished Chair in Metals Processing
Georgia Institute of Technology
Atlanta, GA
USA

Preface

Damage characterization and mechanics is a broad and highly interdisciplinary field that has been continuously evolving in the last half century. This handbook is an attempt to cover the wide spectrum of topics of damage mechanics in a single book in order to reach a wide audience of readers ranging from students to active researchers in both academia and industry. This was a monumental challenge for the authors involved to overcome. An enormous group of internationally recognized authors from both academia and industry assembled from three continents to write 47 chapters on this topic and its various branches.

Tremendous developments have taken place within the research topic of continuum damage mechanics in the past 50 years. There are currently one dedicated journal to this topic as well as numerous books and thousands of research papers. In the framework of continuum damage mechanics, the collection of micro-defects (like micro-cracks, micro-voids, etc.) are treated as a continuous region within which the laws of continuum mechanics are assumed to apply. This is in contrast to what is done in fracture mechanics where individual defects are treated separately and discontinuities are allowed.

Our goal was to assimilate the existing damage mechanics knowledge of academic interest into one consistent, self-contained volume accessible to engineers in practice, researchers in this field, and interested people in academia and to motivate nonspecialists with a strong desire to learn damage mechanics. Such a task was beyond the scope of each of the collected research papers, which by nature focus on narrow topics using very specialized terminology. Our intent was to provide a detailed presentation of those areas of damage mechanics which we have found to be of great practical utility in our industrial experience, while maintaining a sufficiently formal approach both to be suitable as a trustworthy reference for those whose primary interest is further research and to provide a solid foundation for students and others first learning the subject.

Each chapter was written to provide a self-contained treatment of one major topic. Collectively, however, the chapters have been designed and carefully integrated to be entirely complementary with respect to definitions, terminology, and notation. Furthermore, there is essentially no duplication of material across chapters.

The *Handbook of Damage Mechanics* comprises 12 distinct sections including 47 chapters covering the basics of damage mechanics as well as recent research.

The topics covered include the fundamentals of continuum damage mechanics, damage in disordered media, damage in crystalline metals and alloys, damage in structures, damage in electronic packaging, damage in metal forming, micromechanics of damage in composite materials, coupled elastoplastic damage and healing mechanics in granular materials, damage under dynamic loading, experimental characterization of damage, micromechanics of damage in laminated composites, nuclear damage characterization, and recent trends in damage and healing mechanics.

One of the major features of the *Handbook of Damage Mechanics* is coverage of the latest research in the new topic of healing mechanics of materials. The handbook includes four chapters on this emerging subject. In addition, it includes three chapters on the experimental characterization of damage in materials. The fundamentals of continuum damage mechanics are presented in four chapters of the very first section.

The handbook integrated knowledge from the theoretical, numerical, and experimental areas of damage mechanics. This book mainly targets graduate students of damage mechanics, researchers in academia and industry who are active or intend to become active in this field, and practicing engineers and scientists who work in this topic and would like to solve problems utilizing the tools offered by damage mechanics. This handbook should serve as an excellent text for a series of graduate courses in mechanical engineering, civil engineering, materials science, engineering mechanics, aerospace engineering, applied mathematics, applied physics, and applied chemistry.

The handbook is basically intended as a textbook for university courses as well as a reference for researchers in this field. It will serve as a timely addition to the literature on damage mechanics and as an invaluable resource to members of the international scientific and industrial communities.

We hope that the reader will find this handbook a useful resource as he/she progresses in their study and research in damage mechanics. We would also like to wish the readers much success and welcome their suggestions for future improvement of the handbook.

Each of the individual sections of this handbook could be considered a compact, self-contained mini-book right under its own title. However, these topics are presented in relation to the basic principles of damage mechanics.

What is finally presented in the handbook is the work contributed by celebrated international experts for their best knowledge and practices on specific and related topics in damage characterization and mechanics.

The editor would like to thank all the contributors who wrote chapters for this handbook. Finally, the editor would like to acknowledge the help and support of his family members and the editors at Springer who made this handbook possible.

March 2014

George Z. Voyiadjis
Baton Rouge, LA
USA

About the Editor

Dr. George Z. Voyiadjis
Boyd Professor Chair and Bingham C. Stewart Distinguished Professor of Engineering
Department of Civil and Environmental Engineering
Louisiana State University
Baton Rouge, LA
USA

George Z. Voyiadjis is the Boyd Professor at the Louisiana State University in the Department of Civil and Environmental Engineering. This is the highest professorial rank awarded by the Louisiana State University System. He joined the faculty of the university in 1980. He is currently the Chair of the Department of Civil and Environmental Engineering. He holds this position since February 2001.

Voyiadjis is a Foreign Member of the Polish Academy of Sciences, Division IV (Technical Sciences). He is the recipient of the 2008 Nathan M. Newmark Medal of the American Society of Civil Engineers and the 2012 Khan International Medal for outstanding lifelong contribution to the field of plasticity. Voyiadjis was honored in April 2012 by the International Symposium on "Modeling Material Behavior at Multiple Scales" sponsored by Hanyang University, Seoul, Korea, chaired by T. Park and X. Chen (with a dedicated special issue in the *Journal of Engineering Materials and Technology* of the ASME). He was honored in January 2013 by an International Mini-symposium on "Multiscale and Mechanism Oriented Models: Computations and Experiments" sponsored by the International Symposium on Plasticity and Its Current Applications, chaired by V. Tomar and X. Chen.

He is a Fellow of the American Society of Civil Engineers, the American Society of Mechanical Engineers, the American Academy of Mechanics, and the Engineering

Mechanics Institute of ASCE and Associate Fellow of the American Institute of Aeronautics and Astronautics. He was on the Board of Governors of the Engineering Mechanics Institute of the American Society of Civil Engineers and Past President of the Board of Directors of the Society of Engineering Science. He is currently a member of the Executive Committee of the Materials Division (MD) of the American Society of Mechanical Engineers. Dr. Voyiadjis is the Chief Editor of the *Journal of Nanomechanics and Micromechanics* of the ASCE and is on the Editorial Board of numerous engineering journals. He was also selected by the Korea Science and Engineering Foundation (KOSEF) as one of the only two World Class University foreign scholars in the area of civil and architectural engineering to work on nanofusion in civil engineering. This is a multimillion research grant.

Voyiadjis' primary research interest is in plasticity and damage mechanics of metals, metal matrix composites, polymers and ceramics with emphasis on theoretical modeling, numerical simulation of material behavior, and experimental correlation. Research activities of particular interest encompass macro-mechanical and micro-mechanical constitutive modeling, experimental procedures for quantification of crack densities, inelastic behavior, thermal effects, interfaces, damage, failure, fracture, impact, and numerical modeling.

Dr. Voyiadjis' research has been performed on developing numerical models that aim at simulating the damage and dynamic failure response of advanced engineering materials and structures under high-speed impact loading conditions. This work will guide the development of design criteria and fabrication processes of high-performance materials and structures under severe loading conditions. Emphasis is placed on the survivability area that aims to develop and field a contingency armor that is thin and lightweight, but with a very high level of an overpressure protection system that provides low penetration depths. The formation of cracks and voids in the adiabatic shear bands, which are the precursors to fracture, is mainly investigated.

He has a patent, over 260 refereed journal articles, and 17 books (11 as editor) to his credit. He gave over 400 presentations as plenary, keynote, and invited speaker as well as other talks. Over 55 graduate students (30 Ph.D.) completed their degrees under his direction. He has also supervised numerous postdoctoral associates. Voyiadjis has been extremely successful in securing more than $20.0 million in research funds as a principal investigator from the National Science Foundation, the Department of Defense, the Air Force Office of Scientific Research, the Department of Transportation, and major companies such as IBM and Martin Marietta.

He has been invited to give plenary presentations and keynote lectures in many countries around the world. He has also been invited as guest editor in numerous volumes of the *Journal of Computer Methods in Applied Mechanics and Engineering*, *International Journal of Plasticity*, *Journal of Engineering Mechanics* of the ASCE, *Journal of Mechanics of Materials*, and others. These special issues focus in the areas of damage mechanics, structures, fracture mechanics, localization, and bridging of length scales.

He has extensive international collaborations with universities in France, Republic of Korea, and Poland.

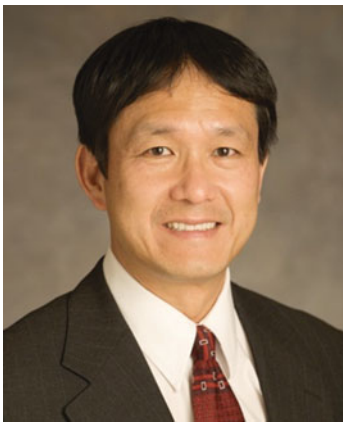
Associate Editors

Mohammed A. Zikry Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC, USA

Cemal Basaran Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA

Antonio Rinaldi Materials Technical Unit, ENEA, C.R. Casaccia, Rome, Italy

Khemais Saanouni Department of Physics, Mechanics, Material and Nanotechnology, ICD/LASMIS, STMR UMR 6281, CNRS, University of Technology of Troyes, Troyes, France



Jiann-Wen Woody Ju Department of Civil and Environmental Engineering, University of California at Los Angeles, Los Angeles, CA, USA

Peter I. Kattan Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA

George A. Kardomateas School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Tomasz Łodygowski Division of Computer Aided Design, Institute of Structural Engineering, Poznan University of Technology, Poznań, Poland

Martin Ostoja-Starzewski Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Taehyo Park World Class University Program, Department of Civil and Environmental Engineering, College of Engineering, Hanyang University, Seoul, Republic of Korea

Contents

Volume 1

Volume 2

Contributors

Boussad Abbès Université de Reims Champagne-Ardenne, GRESPI/MPSE, Faculté des Sciences Exactes et Naturelles, Reims Cedex 2, France

Akrum Abdul-Latif Université Paris 8, Laboratoire d'Ingénierie des Systèmes Mécaniques et des Matériaux (LISMMA), Saint Ouen Cedex, France

F. X. C. Andrade DYNAmore GmbH, Stuttgart, Germany

Houssem Badreddine ICD/LASMIS, STMR UMR CNRS 6281, University of Technology of Troyes, Troyes, France

Cemal Basaran Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA

Zdeněk P. Bažant Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL, USA

A. Amine Benzerga Department of Aerospace Engineering, Texas A&M University, College Station, TX, USA

Department of Materials Science and Engineering, Texas A&M University, College Station, TX, USA

Donald W. Brenner Department of Materials Science and Engineering, North Carolina State University, Raleigh, NC, USA

Vladimir Buljak Faculty of Mechanical Engineering, Department of Strength of Materials, University of Belgrade, Belgrade, Serbia

Esteban P. Busso Materials and Structures Branch, ONERA, Palaiseau, Cedex, France

J. M. A. Cesar de Sa Faculty of Engineering, University of Porto, Porto, Portugal

Giuseppe Cocchetti Department of Civil and Environmental Engineering, Politecnico di Milano, Milano, Italy

Aram Cornaggia Department of Civil and Environmental Engineering, Politecnico di Milano, Milano, Italy

Félix Darve Grenoble INP, UJF, CNRS, Grenoble, France

B. C. Do Department of Mechanical and Aerospace Engineering, University of Miami, Coral Gables, FL, USA

Jérôme Duriez Grenoble INP, UJF, CNRS, Grenoble, France

Rani Elhajjar Department of Civil and Environmental Engineering, University of Wisconsin, Milwaukee, WI, USA

Dan M. Frangopol Department of Civil and Environmental Engineering, Engineering Research Center for Advanced Technology for Large Structural Systems (ATLSS Center), Lehigh University, Bethlehem, PA, USA

Tomasz Garbowski Institute of Structural Engineering, Poznan University of Technology, Poznań, Poland

Juan Gomez Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA

Eray Gunel Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA

Ying-Qiao Guo Université de Reims Champagne-Ardenne, GRESPI/MPSE, Faculté des Sciences Exactes et Naturelles, Reims Cedex 2, France

Rickard Hägglund SCA R&D CENTRE AB, SCA Group, Sundsvall, Sweden

Ali Halouani Université de Reims Champagne-Ardenne, GRESPI/MPSE, Faculté des Sciences Exactes et Naturelles, Reims Cedex 2, France

Mohamed Hamed ICD/LASMIS, STMR UMR CNRS 6281, University of Technology of Troyes, Troyes, France

Neila Hfaiedh ICD/LASMIS, STMR UMR-CNRS 6279, University of Technology of Troyes, Troyes Cedex, France

François Hild LMT-Cachan, ENS Cachan/CNRS/PRES UniverSud Paris, Cachan Cedex, France

Wolfgang Hoffelner RWH consult GmbH, Oberrohrdorf, Switzerland

Nuclear Energy and Safety Research Department, Formerly, Laboratory for Nuclear Materials, Paul Scherrer Institute (PSI), Villigen PSI, Switzerland

Muneo Hori Earthquake Research Institute, The University of Tokyo, Bunkyo, Tokyo, Japan

Mija H. Hubler Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, CO, USA

Per Isaksson Department of Engineering Sciences, Division of Applied Mechanics, Uppsala Universitet, Uppsala, Sweden

Tomasz Jankowiak Division of Computer Aided Design, Institute of Structural Engineering, Poznan University of Technology, Faculty of Civil and Environmental Engineering, Poznań, Poland

Léa Le Joncour ICD/LASMIS, STMR UMR-CNRS 6279, University of Technology of Troyes, Troyes Cedex, France

Jiann-Wen Woody Ju Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, USA

Sohan Kale Department of Mechanical Science & Engineering, also Institute for Condensed Matter Theory and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Peter I. Kattan Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA

Yu-Fu Ko Department of Civil Engineering and Construction Engineering Management, California State University, Long Beach, CA, USA

Valeria La Saponara Department of Mechanical and Aerospace Engineering, University of California, Davis, CA, USA

Carl Labergère ICD/LASMIS, STMR UMR CNRS 6281, University of Technology of Troyes, Troyes, France

Dong Li Department of Materials Science and Engineering, North Carolina State University, Raleigh, NC, USA

Shidong Li Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA

Yuming Li Université de Reims Champagne-Ardenne, GRESPI/MPSE, Faculté des Sciences Exactes et Naturelles, Reims Cedex 2, France

Minghui Lin Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA

H. T. Liu American Bureau of Shipping, Houston, TX, USA

Tomasz Łodygowski Division of Computer Aided Design, Institute of Structural Engineering, Poznan University of Technology, Poznań, Poland

Qiuhai Lu Department of Aeronautical & Astronautical Engineering, School of Aerospace Engineering, Tsinghua University, Beijing, China

Shijing Lu Department of Materials Science and Engineering, North Carolina State University, Raleigh, NC, USA

Giulio Maier Department of Civil and Environmental Engineering, Politecnico di Milano, Milano, Italy

L. Malcher Department of Mechanical Engineering, University of Brasilia, Asa Norte - Brasília, DF, Brazil

Sreten Mastilovic Faculty of Construction Management, Union–Nikola Tesla University, Belgrade, Serbia

Navid Mozaffari Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA

Hakim Naceur Université Lille Nord de France, Laboratoire LAMIH UMR 8201 CNRS, Valenciennes Cedex 9, France

François Nicot IRSTEA, Geomechanics group, ETNA, Grenoble, France

Shihua Nie Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA

Giorgio Novati Department of Civil and Environmental Engineering, Politecnico di Milano, Milano, Italy

Martin Ostoja-Starzewski Department of Mechanical Science & Engineering, also Institute for Condensed Matter Theory and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Benoit Panicaud ICD/LASMIS, STMR UMR-CNRS 6279, University of Technology of Troyes, Troyes Cedex, France

F. M. A. Pires Faculty of Engineering, University of Porto, Porto, Portugal

Jean-Noël Périé LMT-Cachan, ENS Cachan/CNRS/PRES UniverSud Paris, Cachan Cedex, France

Florent Prunier INSA de Lyon, LGCIE, Villeurbanne, France

Jianmin Qu Department of Civil and Environmental Engineering, Department of Mechanical Engineering, Northwestern University, Evanston, IL, USA

Antonio Rinaldi Materials Technical Unit, ENEA, C.R. Casaccia, Rome, Italy

Center for Mathematics and Mechanics of Complex Systems (MEMOCS), University of L'Aquila, Cisterna di Latina (LT), Italy

Stéphane Roux LMT-Cachan, ENS Cachan/CNRS/PRES UniverSud Paris, Cachan Cedex, France

Khemais Saanouni ICD/LASMIS, STMR UMR CNRS 6281, University of Technology of Troyes, Troyes, France

M. R. R. Seabra Faculty of Engineering, University of Porto, Porto, Portugal

Pratheek Shanthraj Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC, USA

Department of Microstructure–Physics and Alloy Design, Max Planck Institut ür Eisenforschung, Düsseldorf, Germany

Amir Shojaei Department of Mechanical and Industrial Engineering, Louisiana State University, Baton Rouge, LA, USA

Mohamed Soliman Department of Civil and Environmental Engineering, Engineering Research Center for Advanced Technology for Large Structural Systems (ATLSS Center), Lehigh University, Bethlehem, PA, USA

Wojciech Sumelka Division of Computer Aided Design, Institute of Structural Engineering, Poznan University of Technology, Poznań, Poland

L. Z. Sun Civil and Environmental Engineering, University of California, Irvine, CA, USA

Hong Tang Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA

Ziad N. Taqieddin Civil Engineering Department, Applied Science University, Amman, Jordan

George Z. Voyiadjis Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA

Richard Wan Department of Civil Engineering, University of Calgary, Calgary, AB, Canada

Michael P. Wnuk Department of Civil Engineering and Mechanics, College of Engineering and Applied Science, University of Wisconsin, Milwaukee, WI, USA

Zhihai Xiang Department of Engineering Mechanics, School of Aerospace Engineering, Tsinghua University, Beijing, China

Chengyong Yan Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA

Q. D. Yang Department of Mechanical and Aerospace Engineering, University of Miami, Coral Gables, FL, USA

Wei Yao Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA

Hua Ye Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA

Mohammed A. Yousef Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA

Qiang Yu Department of Civil and Environmental Engineering, University of Pittsburgh, Swanson School of Engineering, Pittsburgh, PA, USA

K. Y. Yuan Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, USA

Mohammed A. Zikry Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC, USA

Section I

Continuum Damage Mechanics Fundamentals

Some Basic Issues of Isotropic and Anisotropic Continuum Damage Mechanics

1

George Z. Voyiadjis, Peter I. Kattan, and Mohammed A. Yousef

Contents

Abstract

This chapter includes two main topics in isotropic and anisotropic damage mechanics. In the first topic, various new proposed damage variables are introduced, examined, and compared. The scalar case pertaining to isotropic damage is investigated, and several types of new damage variables are proposed. The damage variables introduced in this part can be applied to elastic materials including homogeneous materials like metals and heterogeneous materials like composite laminates. Moreover, higher-order strain energy forms are proposed. These higher-order strain energy forms along with some of the proposed damage variables are used in trying to lay the theoretical groundwork for the design of

G.Z. Voyiadjis (✉) • P.I. Kattan • M.A. Yousef
Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA
e-mail: voyiadjis@eng.lsu.edu; pkattan@orange.jo; myouse2@tigers.lsu.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_1

undamageable materials, i.e., materials that cannot be damaged where the value of the damage variable remains zero throughout the deformation process.

The second topic presents a new concept of anisotropic damage that is examined within the framework of continuum damage mechanics. New proposed damage tensors are studied in order to investigate the damage effect variables in the mechanical behavior of materials. In addition, new hybrid damage tensors are proposed and defined in terms of the damage effect tensor and the new proposed damage tensors. Accordingly, this study demonstrates that most of the new proposed damage tensors are verified within the framework of continuum damage mechanics.

Introduction

Research in the field of continuum damage mechanics has remarkably caused the advancement of the material overall performance and the proliferation in material engineering applications. Continuum damage mechanics addresses among others the study of the growth of micro-cracks and micro-voids (and other defects) and their effect on the mechanical behavior of the material. It also enables engineers and researchers to improve the microstructure of the material. By the skillful manipulation of this development, the correlation between the damage variables and the mechanical properties of the materials is thus obtained. Further consideration in that sense is the direct effect of damage on the mechanical properties of materials which results in a significant influence on the safety aspect of engineering structures. This effect is frequently encountered in safety-critical applications, such as in aeronautical and nuclear industries. An unpredicted failure may cause catastrophic consequences in human life as well as detrimental financial disasters. Accordingly, a great attention has been recently given by many researchers to the field of continuum damage mechanics in order to study the mechanics of micro-cracks and micro-voids and their effects on the mechanical behavior of materials.

Damage causes deterioration in the mechanical properties of materials due to the creation and growth of micro-voids and/or micro-cracks which are discontinuities in a medium considered as continuous at a larger scale (Lemaitre and Desmorat 2005). A representative volume element (RVE) is introduced in the mechanics of continuous media (Lemaitre and Desmorat 2005) on which all properties are represented by homogenized variables. Moreover, the effective stress concept is used in order to obtain the damage variable that represents the average material stiffness degradation within the framework of continuum damage mechanics. The degradation that is included in the material reflects the various types of damage at the microscale level like nucleation and growth of voids, cracks, cavities, microcracks, and other microscopic defects (Krajcinovic 1996; Budiansky and O'Connell 1976; Lubarda and Krajcinovic 1993).

The damage variable is scalar in the case of isotropic damage mechanics, and the evolution equations are quite easy to handle. It has been, however, shown by Cauvin and Testa (1999) that the accurate and consistent description of the special

case of isotropic damage is obtained by using two independent damage variables. New scalar damage variables have been previously proposed by Voyiadjis and Kattan (2009). Lemaitre (1984) argued that isotropic damage gives good predictions of the load carrying capacity, the number of cycles, or the time to local failure in structural components. However, the experimental results have confirmed the development of anisotropic damage (Chow and Wang 1987; Lee et al. 1985) even if the virgin material is isotropic. This has prompted several researchers to investigate the general case of anisotropic damage (Voyiadjis and Kattan 1996, 1999, 2006; Kattan and Voyiadjis 2001a, b).

Kachanov (1958) and Rabotnov (1969) used the case of uniaxial tension to demonstrate the basic idea of the effective stress, and later, Lemaitre (1971) and Chaboche (1981) obtained the three-dimensional states of stress. The second-rank Cauchy stress tensor is applied to the damaged material, and the corresponding effective stress tensor is applied to a fictitious state of the material which is totally undamaged. They used the fictitious undamaged state and the actual damaged state of the material in their formulation. In this regard, one of the following two hypotheses, elastic strain equivalence or elastic energy equivalence, is usually used to derive the corresponding constitutive formulations. Sidoroff (1981), Cordebois and Sidoroff (1979), and Cordebois (1983) developed the theory of anisotropic damage mechanics, and later Lee et al. (1985) and Chow and Wang (1987, 1988) used it to solve simple ductile fracture problems. However, prior to this latest development, Krajcinovic and Foneska (1981), Murakami and Ohno (1981), Murakami (1983), and Krajcinovic (1983) investigated brittle and creep fracture using appropriate anisotropic damage models.

Although these models are based on a sound physical background, they lack vigorous mathematical justification and mechanical consistency. Consequently, more work needs to be done to develop a more involved theory capable of producing results that can be used for practical applications (Krajcinovic and Foneska 1981; Krempl 1981). The damage variable has been shown to be tensorial in nature in order to develop the general case of anisotropic damage (Murakami and Ohno 1981; Leckie and Onat 1981). This damage tensor was shown to be an irreducible even-rank tensor (Onat 1986; Onat and Leckie 1988). Betten (1981, 1986) showed several other basic properties of the damage tensor using the theory of tensor functions. It was pointed out that even for isotropic damage, a damage tensor should be employed (not a scalar damage variable) for characterizing the state of damage in materials (Ju 1990). On the other hand, the damage process is generally anisotropic due to the loading condition or the material nature itself. Although the fourth-rank damage tensor can be directly used as a linear transformation tensor to define the effective stress tensor, it is not easy to computationally use the fourth-rank damage tensor compared to the second-rank damage tensor.

This chapter introduces a new study in the field of continuum damage mechanics and includes two main topics. In the first topic, various new proposed damage variables are introduced, examined, and compared. The scalar case pertaining to isotropic damage is investigated, and several types of new damage variables are proposed. Damage variables are defined in terms of the cross-sectional area and the

elastic modulus or stiffness and the composite damage variables that are defined in terms of two parameters relating to both cross-sectional area and stiffness. The damage variables introduced in this part can be applied to elastic materials including homogeneous materials like metals and heterogeneous materials like composite laminates. Moreover, higher-order strain energy forms are proposed. It is seen that a specific nonlinear stress–strain relationship is associated with each higher-order strain energy form. These higher-order strain energy forms along with some of the proposed damage variables are used in trying to lay the theoretical groundwork for the design of undamageable materials, i.e., materials that cannot be damaged where the value of the damage variable remains zero throughout the deformation process.

On the other hand, the second topic presents a new concept of anisotropic damage that is examined within the framework of continuum damage mechanics. New proposed damage tensors are studied in order to investigate the damage effect variables in the mechanical behavior of materials. All the cases studied here are defined in terms of the elastic stiffness of the material and based on the hypotheses of both the elastic strain equivalence and the elastic energy equivalence. Furthermore, the new concept of anisotropy is significantly applied to define the new proposed damage tensor which is derived and anisotropically expressed in terms of the well-known damage effect tensor. The principal-valued damage effect tensor is utilized in order to obtain the first scalar invariant of that tensor and its inverse which are employed in expressing and verifying the new proposed damage tensors. In addition, new hybrid damage tensors are proposed and defined in terms of the damage effect tensor and the new proposed damage tensors. The new hybrid damage tensors are eventually obtained in terms of the damage effect tensor. Accordingly, this study demonstrates that most of the new proposed damage tensors are verified within the framework of continuum damage mechanics.

Isotropic Damage in Continuum Damage Mechanics

Review of Existing Damage Variables

Two major scalar damage variables used by researchers are elaborately discussed here. The first scalar damage variable is defined in terms of the reduction in cross-sectional area, while the second scalar damage variable is defined in terms of the reduction in the elastic modulus or elastic stiffness. To illustrate the damaged and effective undamaged configurations, consider a body (in the form of a cylinder) in the initial undeformed and undamaged configuration, and also the configuration of the body that is both deformed and damaged after a set of external agencies act on it (see Fig. 1). A fictitious configuration of the body is then obtained from the damaged configuration by removing all the damage that the body has undergone, that is, this is the state of the body after it had only deformed without damage (see Fig. 1). Therefore, in defining a damage variable ϕ, its value must vanish in the fictitious configuration.

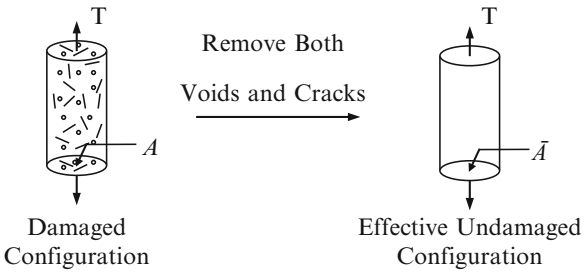


Fig. 1 Damaged and effective undamaged configurations (Reprinted with permission from Voyiadjis and Kattan 2012)

The first damage variable ϕ_1 is usually defined as follows:

$$\phi_1 = \frac{A - \overline{A}}{A} \tag{1}$$

where A is the cross-sectional area in the damaged configuration, while $\overline{A}$ is the cross-sectional area in the fictitious configuration with $A > \overline{A}$. It is clear that when a body is undamaged, i.e., when $A = \overline{A}$ accordingly, then $\phi_1 = 0$.

The stress in the fictitious configuration is called the effective stress and is denoted by $\overline{\sigma}$. The value of the effective stress $\overline{\sigma}$ may be obtained using the relation $\overline{\sigma}\overline{A} = \sigma A$ where σ is the stress in the damaged configuration. Therefore, using this relation along with the definition in Eq. 1, one obtains

$$\overline{\sigma} = \frac{\sigma}{1 - \phi_1} \tag{2}$$

It should be mentioned that the equilibrium condition in the paragraph above reflects a mean-field type of assumption on the stress redistribution (uniform over the resistive section) and therefore appears to be appropriate only in the dilute damage regime. Away from the stress–strain peak, cooperative effects dominate, and damage localization takes place. This holds true, especially for the tensorial case.

Let us denote the ratio $\overline{\sigma}/\sigma$ by the symbol M, and accordingly one obtains the following relation, $M = \frac{1}{1-\phi_1}$. Figure 2 shows a plot of the relation between ϕ_1 and M using the above relation. The graph shown is plotted for the range $0 \leq \phi_1 \leq 1$. Clearly $M \geq 1$ which indicates that $\overline{\sigma} \geq \sigma$. This is the governing condition for the damage variable to be valid. From the graph, it is very clear that $1 \leq M \leq \infty$. These are the limiting conditions for M. The variable M is called the *damage effect variable*.

In order to compute the effective elastic modulus $\overline{E}$ in this case, it is needed to make a certain assumption regarding the energy/strain in the two configurations. Usually, one of the following two hypotheses is assumed:

1. Hypothesis of elastic strain equivalence: in this case, the strain in the damaged state ε is assumed to be equal to the effective strain $\overline{\varepsilon}$ (in the fictitious state).
2. Hypothesis of elastic energy equivalence: in this case the elastic strain energy is assumed to be equal in both configurations.

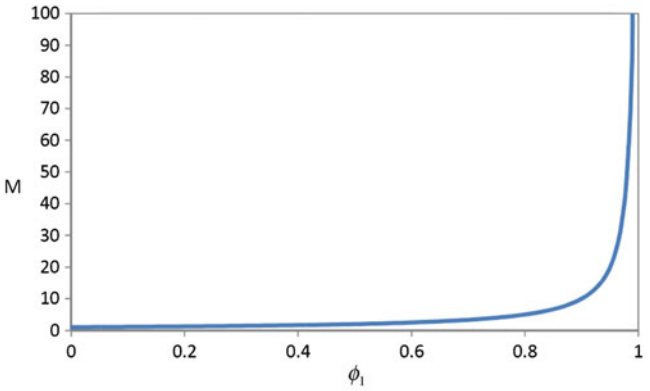


Fig. 2 Relation between the damage variable and M (Reprinted with permission from Voyiadjis and Kattan 2012)

While both types of hypotheses are usually used by researchers in the field of damage mechanics, it is believed that the hypothesis of elastic energy equivalence is more valid than the hypothesis of elastic strain equivalence, primarily because it involves some form of energy formulation. Therefore, it is recommended to use the second hypothesis listed above.

The second scalar damage variable ℓ_1 may be defined in terms of the reduction in the elastic modulus as follows:

$$\ell_1 = \frac{\overline{E} - E}{E} \tag{3}$$

where E is the elastic modulus in the damaged state, while $\overline{E}$ is the effective elastic modulus (in the fictitious state) with $\overline{E} > E$ (see Fig. 3). This damage variable was used recently by Celentano et al. (2004), Nichols and Abell (2003), and Nichols and Totoev (1999). It should also be mentioned that Voyiadjis (1988) and Voyiadjis and Kattan (2009) used a similar relation but in the context of elastoplastic deformation.

The definition of the alternative damage variable of Eq. 3 may be rewritten in the following more appropriate form:

$$\overline{E} = E(1 + \ell_1) \tag{4}$$

It is clear from the definition in Eq. 4 that $\ell_1 = 0$ when the body is undamaged, i.e., when $\overline{E} = E$.

Hypothesis of Elastic Strain Equivalence

Using the hypothesis of elastic strain equivalence, we assume:

$$\overline{\varepsilon} = \varepsilon \tag{5}$$

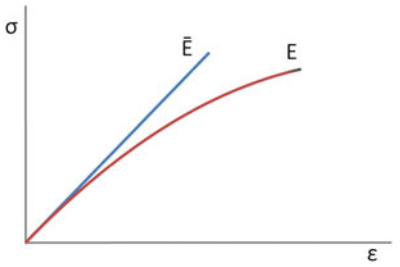


Fig. 3 Damaged and effective moduli of elasticity (Reprinted with permission from Voyiadjis and Kattan 2012)

Using the elastic constitutive relations in both configurations as follows:

$$\sigma = E\varepsilon \tag{6}$$

$$\overline{\sigma} = \overline{E}\overline{\varepsilon} \tag{7}$$

and substituting for $\overline{\sigma}$ and $\overline{\varepsilon}$ using Eqs. 2 and 5, respectively, into Eq. 7, one obtains

$$\overline{E} = \frac{E}{1 - \phi_1} \tag{8}$$

In order to find an appropriate relation between the two damage variables ϕ_1 and ℓ_1, one equates Eqs. 4 and 8 to obtain

$$\ell_1 = \frac{1}{1 - \phi_1} - 1 \tag{9}$$

The above expression defines the exact relation between the two damage variables ℓ_1 and ϕ_1 in the case of the hypothesis of elastic strain equivalence. The relation in Eq. 9 may be rewritten in the following form:

$$\phi_1 = \frac{\ell_1}{1 + \ell_1} \tag{10}$$

The relation of Eq. 9 is plotted between ϕ_1 and ℓ_1 as shown in Fig. 4. In this case, it is clear that $\ell_1 = 0$ when $\phi_1 = 0$. However, when one sets ℓ_1 to its maximum value of 1, one obtains $\phi_1 = 0.5$.

Consequently one obtains the following valid ranges $0 \leq \ell_1 \leq 1$ and $0 \leq \phi_1 \leq 0.5$. This is obviously a limiting maximum value for ϕ_1 based on the definition of stiffness reduction of ℓ_1. Again, one stresses that this holds for the hypothesis of elastic strain equivalence. It should be noted that Eq. 3 does not impose any limitations on the value of ℓ_1. However, it is assumed that the damage variable ℓ_1 has a fractional numerical value between 0 and 1. Thus, the statement that ℓ_1 has a maximum value of 1 arises based on this assumption (see Voyiadjis and Kattan 2009).

Hypothesis of Elastic Energy Equivalence

Using the hypothesis of elastic energy equivalence, one assumes the complementary elastic strain energy $\left(\frac{\sigma^2}{2E}\right)$ to be equal in both configurations, i.e.,

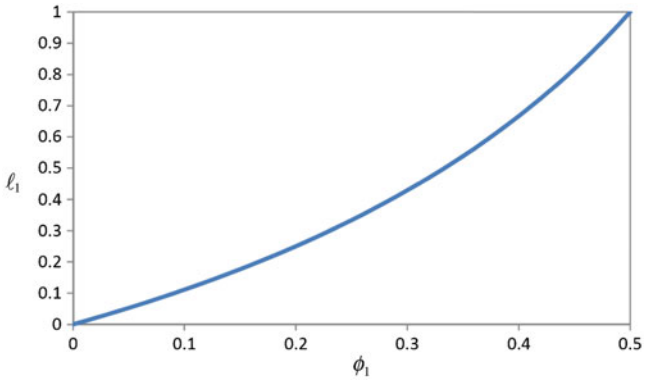


Fig. 4 Relation between the two damage variables in the case of elastic strain equivalence (Reprinted with permission from Voyiadjis and Kattan 2012)

$$\frac{\sigma^2}{2E} = \frac{\overline{\sigma}^2}{2\overline{E}} \tag{11}$$

Substituting for $\overline{\sigma}$ from Eq. 2 into Eq. 11 and simplifying, one obtains

$$\overline{E} = \frac{E}{(1-\phi_1)^2} \tag{12}$$

Next, in order to derive the relation between the two damage variables ϕ_1 and ℓ_1, one equates Eqs. 4 and 12 to obtain

$$\ell_1 = \frac{1}{(1-\phi_1)^2} - 1 \tag{13}$$

The above expression defines the exact relation between the two damage variables ℓ_1 and ϕ_1 in the case of the hypothesis of elastic energy equivalence. The relation in Eq. 13 may be rewritten in the following form:

$$\phi_1 = 1 - \frac{1}{\sqrt{1+\ell_1}} \tag{14}$$

One now plots the relation of Eq. 13 between ϕ_1 and ℓ_1 as shown in Fig. 5. In this case, it is clear that $\ell_1 = 0$ when $\phi_1 = 0$. However, when one sets ℓ_1 to its maximum value of 1, one obtains $\phi = 0.293 \approx 0.3$, and consequently the valid ranges are as follows: $0 \leq \ell_1 \leq 1$ and $0 \leq \phi_1 \leq 0.293$. This is obviously a limiting maximum value for ϕ based on the definition of stiffness reduction of ℓ_1. Again, one should stress that this holds for the hypothesis of elastic energy equivalence. It should be noted that Eq. 3 does not impose any limitations on the value of ℓ_1. However, it is assumed that the damage variable ℓ_1 has a fractional numerical value between 0 and 1. Thus, the statement that ℓ_1 has a maximum value of 1 arises based

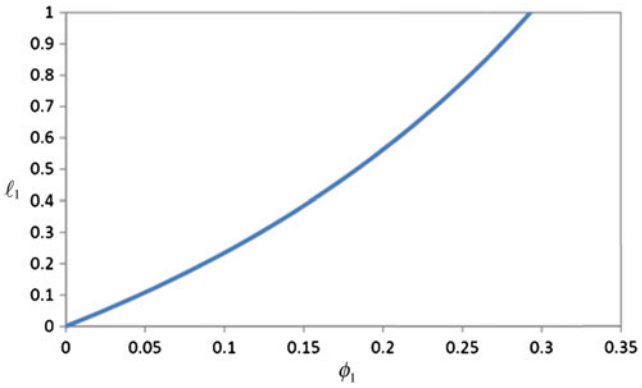


Fig. 5 Relation between the two damage variables for the case of elastic energy equivalence (Reprinted with permission from Voyiadjis and Kattan 2012)

Table 1 Limiting Values of ϕ_1 and ℓ_1 (Reprinted with permission from Voyiadjis and Kattan 2012)

		Hypothesis of elastic strain equivalence	Hypothesis of elastic energy equivalence
ℓ_1	Min. value	0.0	0.0
	Max. value	1.0	1.0
ϕ_1	Min. value	0.0	0.0
	Max. value	0.5	0.293

on this assumption (see Voyiadjis and Kattan 2009). Table 1 summarizes the results for the two hypotheses.

Proposed New Damage Variables

In this section, two alternative proposed new damage variables are presented that are similar to ϕ_1 and ℓ_1. These two new damage variables will be called ϕ_2 and ℓ_2, respectively, defined as follows:

$$\phi_2 = \frac{A - \overline{A}}{\overline{A}} \tag{15}$$

$$\ell_2 = \frac{\overline{E} - E}{\overline{E}} \tag{16}$$

Note that the differences between ϕ_1 and ϕ_2 and between ℓ_1 and ℓ_2 are only in the denominator. These changes are made in order to investigate the relationships between the various damage variables and assess their validity.

Hypothesis of Elastic Strain Equivalence

Using the hypothesis of elastic strain equivalence, the following relation between ϕ_1 and ℓ_2 is obtained:

$$\ell_2 = \phi_1 \tag{17}$$

Thus, that under this hypothesis, the two damage variables ϕ_1 and ℓ_2 are equivalent. Therefore, this relationship is represented by a straight line as shown in Fig. 6. The relation of Eq. 17 between ϕ_1 and ℓ_2 is shown in Fig. 6. In this case, it is clear that $\ell_2 = 0$ when $\phi_1 = 0$. However, when one sets ℓ_2 to its maximum value of 1, then $\phi_1 = 1.0$ with the following valid ranges $0 \leq \ell_2 \leq 1$ and $0 \leq \phi_1 \leq 1$. This is obviously a limiting maximum value for ϕ_1 based on the definition of stiffness reduction of ℓ_2. Again, one should stress that this holds for the hypothesis of elastic strain equivalence. It should be noted that Eq. 16 does not impose any limitations on the value of ℓ_2. However, it is assumed that the damage variable ℓ_2 has a fractional numerical value between 0 and 1. Thus, the statement that ℓ_2 has a maximum value of 1 arises based on this assumption (see Voyiadjis and Kattan 2009).

Next, the relationship between the damage variables ϕ_2 and ℓ_1 is obtained using the hypothesis of elastic strain equivalence. Using this hypothesis, one deduces the following relationship:

$$\ell_1 = \phi_2 \tag{18}$$

Thus, under this hypothesis, the two damage variables ϕ_2 and ℓ_1 are equivalent. Therefore, this relationship is represented by a straight line as shown in Fig. 7. The relation of Eq. 18 between ϕ_2 and ℓ_1 is plotted in Fig. 7. In this case, it is clear that $\ell_1 = 0$ when $\phi_2 = 0$. However, when one sets ℓ_1 to its maximum value of 1, one obtains $\phi_2 = 1.0$ with the following valid ranges $0 \leq \ell_1 \leq 1$ and $0 \leq \phi_2 \leq 1$. This is obviously a limiting maximum value for ϕ_2 based on the definition of stiffness reduction of ℓ_1. This holds for the hypothesis of elastic strain equivalence.

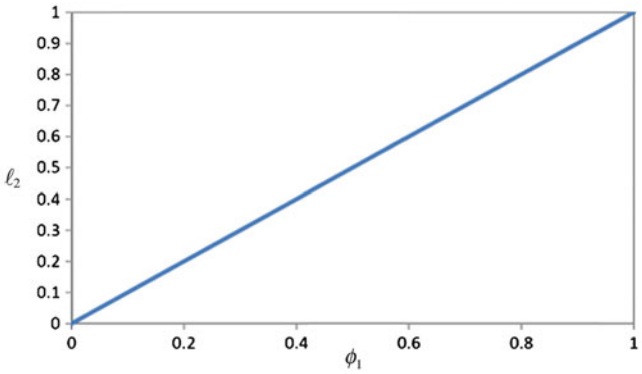


Fig. 6 Relation between the two damage variables in the case of elastic strain equivalence (Reprinted with permission from Voyiadjis and Kattan 2012)

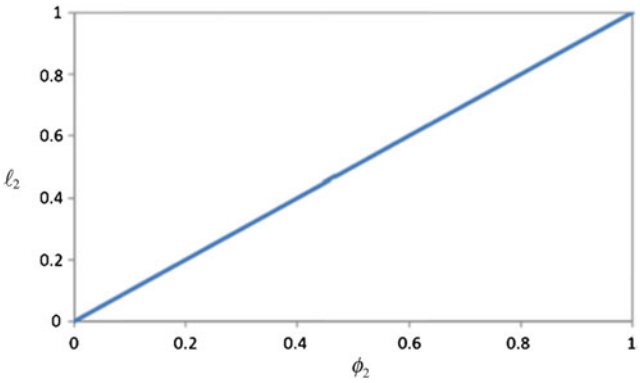


Fig. 7 Relation between the two damage variables in the case of elastic strain equivalence (Reprinted with permission from Voyiadjis and Kattan 2012)

Next, the relationship between the proposed damage variables ϕ_2 and ℓ_2 is studied using the hypothesis of elastic strain equivalence. Using this hypothesis, the following relationship is obtained:

$$\ell_2 = 1 - \frac{1}{1 + \phi_2} \tag{19}$$

The above relationship can also be rewritten as follows:

$$\phi_2 = \frac{\ell_2}{1 - \ell_2} \tag{20}$$

The relation of Eqs. 19 and 20 between ϕ_2 and ℓ_2 is shown in Fig. 8. In this case, it is clear that $\ell_2 = 0$ when $\phi_2 = 0$. However, when one sets ϕ_2 to its maximum value of 1, one obtains $\ell_2 = 0.5$ also, with the valid ranges $0 \leq \ell_2 \leq 0.5$ and $0 \leq \phi_2 \leq 1$. This is obviously a limiting maximum value for ℓ_2 based on the definition of stiffness reduction of ℓ_1. This holds for the hypothesis of elastic strain equivalence.

Hypothesis of Elastic Energy Equivalence

Using the hypothesis of elastic energy equivalence, the following relation is obtained between ϕ_1 and ℓ_2:

$$\ell_2 = 1 - (1 - \phi_1)^2 \tag{21}$$

The above relationship may be rewritten as follows:

$$\phi_1 = 1 - \sqrt{1 - \ell_2} \tag{22}$$

The relation of Eqs. 21 and 22 between ϕ_1 and ℓ_2 is shown in Fig. 9. In this case, it is clear that $\ell_2 = 0$ when $\phi_1 = 0$. However, when by setting ℓ_2 to its maximum

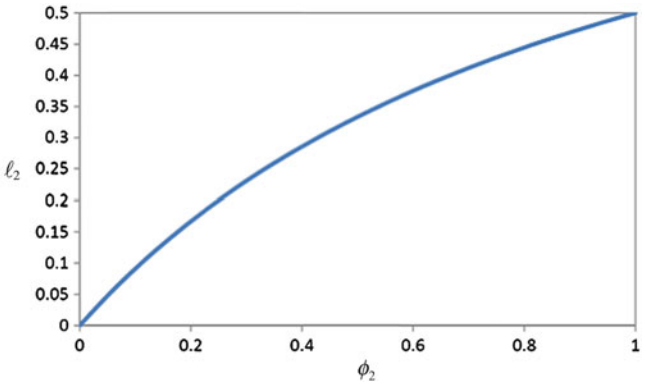


Fig. 8 Relation between the two damage variables in the case of elastic strain equivalence (Reprinted with permission from Voyiadjis and Kattan 2012)

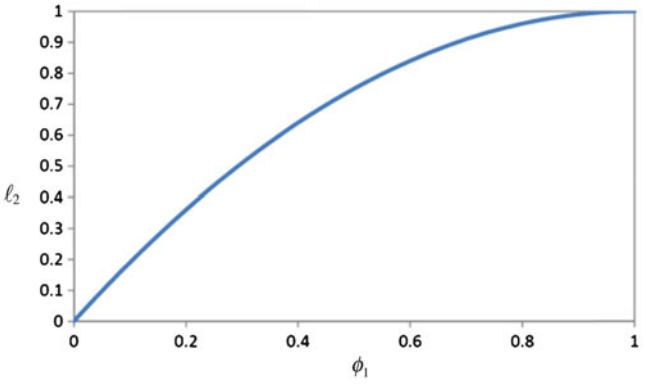


Fig. 9 Relation between the two damage variables in the case of elastic energy equivalence (Reprinted with permission from Voyiadjis and Kattan 2012)

value of 1, one obtains $\phi_1 = 1.0$ for the following valid ranges $0 \leq \ell_2 \leq 1$ and $0 \leq \phi_1 \leq 1$. This is obviously a limiting maximum value for ϕ_1 based on the definition of stiffness reduction of ℓ_2. It should be noted that Eq. 16 does not impose any limitations on the value of ℓ_2. However, it is assumed that the damage variable ℓ_2 has a fractional numerical value between 0 and 1. Thus, the statement that ℓ_2 has a maximum value of 1 arises based on this assumption (see Voyiadjis and Kattan 2009).

The relationship is next studied between the damage variables ϕ_2 and ℓ_1 using the hypothesis of elastic energy equivalence. Using this hypothesis, one deduces the following relationship:

$$\ell_1 = \phi_2 (2 + \phi_2) \tag{23}$$

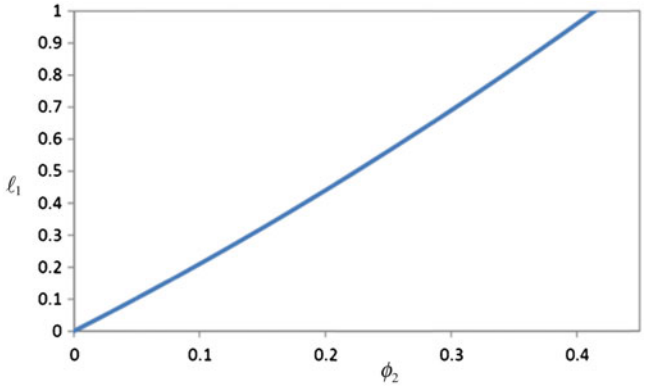


Fig. 10 Relation between the two damage variables in the case of elastic energy equivalence (Reprinted with permission from Voyiadjis and Kattan 2012)

The above relationship can be rewritten in the following form:

$$\phi_2 = \sqrt{1 + \ell_1} - 1 \tag{24}$$

The relation of Eqs. 23 and 24 between ϕ_2 and ℓ_1 is shown in Fig. 10. In this case, it is clear that $\ell_1 = 0$ when $\phi_2 = 0$. However, by setting ℓ_1 to its maximum value of 1, one obtains $\phi_2 = 0.414$ for the valid ranges $0 \leq \ell_1 \leq 1$ and $0 \leq \phi_2 \leq 0.414$. This is obviously a limiting maximum value for ϕ_2 based on the definition of stiffness reduction of ℓ_1.

The relationship between the new proposed damage variables ϕ_2 and ℓ_2 using the hypothesis of elastic energy equivalence is now obtained. Using this hypothesis, the following relation is deduced:

$$\ell_2 = 1 - \frac{1}{(1 + \phi_2)^2} \tag{25}$$

The above relationship can also be rewritten as follows:

$$\phi_2 = \frac{1}{\sqrt{1 - \ell_2}} - 1 \tag{26}$$

The relation of Eqs. 25 and 26 between ϕ_2 and ℓ_2 is shown in Fig. 11. In this case, it is clear that $\ell_2 = 0$ when $\phi_2 = 0$. However, when one sets ℓ_2 to its maximum value of 1, one obtains $\phi_2 \to \infty$ for the valid ranges $0 \leq \ell_2 \leq 1$ and $0 \leq \phi_2 \leq \infty$. Table 2 summarizes the new results for the two hypotheses.

In this section, a new class of damage variables is presented. The new proposed damage variables are investigated and compared with the existing damage variables. It should be noted that the new damage variables fall into two major categories based on the discussion of section "Isotropic Damage in Continuum Damage Mechanics." The first category consists of all the damage variables that are defined in terms of the

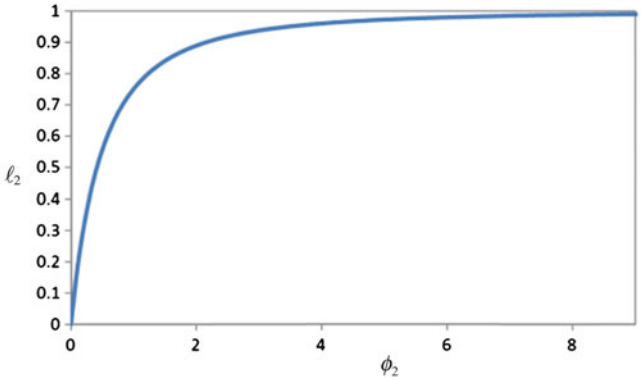


Fig. 11 Relation between the two damage variables in the case of elastic energy equivalence (Reprinted with permission from Voyiadjis and Kattan 2012)

Table 2 Limiting values of ϕ_2 and ℓ_2 (Reprinted with permission from Voyiadjis and Kattan 2012)

		Hypothesis of elastic strain equivalence	Hypothesis of elastic energy equivalence
ℓ_2	Min. value	0.0	0.0
	Max. value	1.0	1.0
ϕ_2	Min. value	0.0	0.0
	Max. value	0.5	0.414

cross-sectional area similar to ϕ_1. The second category consists of all the damage variables that are defined in terms of the elastic stiffness similar to ℓ_1.

Other Proposed New Damage Variables

Damage Variables Defined in Terms of Area

Looking at the damage variable ϕ_1 as defined in Eq. 1, a similar damage variable is proposed ϕ_2 defined as follows (as was shown in Eq. 15):

$$\phi_2 = \frac{A - \overline{A}}{\overline{A}} \tag{27}$$

where the definition of ϕ_2 is different from the definition of ϕ_1 only in the denominator. Based on Eqs. 1 and 27, one can easily show that the relation between these two damage variables is given by

$$\phi_2 = \phi_1 + \phi_1 \phi_2 \tag{28}$$

From the above relation, it is clearly seen that $\phi_2 = \frac{\phi_1}{1-\phi_1}$. The new proposed damage variable ϕ_2 does not have the same limiting values as ϕ_1. It was shown in

the previous section that $0 \leq \phi_1 \leq 1$. However, the same is not true for ϕ_2. From the above equation, it can be deduced that $0 \leq \phi_2 \leq \infty$.

One now obtains the value of the damage effect variable M for this new proposed damage variable. The value of the effective stress $\overline{\sigma}$ may be obtained using the relation $\overline{\sigma}\overline{A} = \sigma A$ where σ is the stress in the damaged configuration. Therefore, using this relation along with the definition in Eq. 27, one obtains

$$\overline{\sigma} = \sigma(1 + \phi_2) \tag{29}$$

Denoting the ratio $\overline{\sigma}/\sigma$ by the letter M, consequently one obtains $M = 1 + \phi_2$. It is now clear that, unlike the classic damage variable, the value of this proposed damage variable can approach infinity. Clearly, the new damage variable satisfies the condition $1 \leq M \leq \infty$. Therefore, this new proposed damage variable is valid.

Next, the following new damage variable is proposed that is slightly more complicated:

$$\psi_1 = \sqrt{\frac{A^2 - \overline{A}^2}{A\overline{A}}} \tag{30}$$

One now investigates whether the above proposed damage variable is valid. The damage effect variable M that is associated with ψ_1 is now obtained. One can easily show that the following relation is obtained based on Eq. 30:

$$\overline{\sigma} = \left(\frac{1}{2}\psi_1^2 \pm \frac{1}{2}\sqrt{4 + \psi_1^4}\right)\sigma \tag{31}$$

Thus, based on the evaluation of the damage effect variable M, the proposed damage variable ψ_1 is valid provided that it satisfies the condition $\left(\frac{1}{2}\psi_1^2 \pm \frac{1}{2}\sqrt{4 + \psi_1^4}\right) \geq 1$, i.e., $M \geq 1$, as discussed in the beginning of this chapter.

One now considers next the following new proposed damage variable which is similar to the one defined in Eq. 30:

$$\psi_2 = \sqrt{\frac{(A - \overline{A})^2}{A\overline{A}}} \tag{32}$$

The validity of the above proposed damage variable is now investigated. One can easily show that the following relation is obtained based on Eq. 32:

$$\overline{\sigma} = \left(1 + \frac{1}{2}\psi_2^2 \pm \frac{1}{2}\psi_2\sqrt{4 + \psi_2^2}\right)\sigma \tag{33}$$

Thus, based on the evaluation of the damage effect variable M, the proposed damage variable ψ_2 is valid provided that it satisfies the condition

$\left(1+\frac{1}{2}\psi_2^2 \pm \frac{1}{2}\psi_2\sqrt{4+\psi_2^2}\right) \geq 1$, i.e., $M \geq 1$, as discussed in the beginning of this chapter.

A third damage variable is now proposed that is defined in terms of the cross-sectional area as follows:

$$\psi_3 = \frac{1}{2}\sqrt{\frac{A^2 - \overline{A}^2}{\overline{A}^2}} \tag{34}$$

One now investigates whether the above proposed damage variable is valid or not. Let us derive the value of the damage effect variable M that is associated with ψ_3. Based on Eq. 34 the following relation is obtained:

$$\overline{\sigma} = \sqrt{1+4\psi_3^2}\,\sigma \tag{35}$$

Thus, based on the evaluation of the damage effect variable M, the proposed damage variable ψ_3 is valid provided that it satisfies the condition $\sqrt{1+4\psi_3^2} \geq 1$, i.e., $M \geq 1$, as discussed in the beginning of this work.

The fourth proposed damage variable in this category is defined as follows:

$$\psi_4 = \frac{1}{2}\sqrt{\frac{A^2 - \overline{A}^2}{A^2}} \tag{36}$$

The value of the damage effect variable M that is associated with ψ_4 is obtained from the relation based on Eq. 36:

$$\overline{\sigma} = \frac{\sigma}{\sqrt{1-4\psi_4^2}} \tag{37}$$

Thus, based on the evaluation of the damage effect variable M, the proposed damage variable ψ_4 is also valid provided that it satisfies the condition $\frac{1}{\sqrt{1-4\psi_4^2}} \geq 1$, i.e., $M \geq 1$, as discussed in the beginning of this work.

Finally, one presents the following two relations between the damage variables ψ_1 and ψ_2 on one hand and ϕ_1 and ϕ_2 on the other hand:

$$\psi_1^2 - \psi_2^2 = 2\phi_1 \tag{38}$$

$$\psi_1^2 + \psi_2^2 = 2\phi_2 \tag{39}$$

Consider now two new damage variables ψ_5 and ψ_6 defined as follows:

$$\psi_5 = \frac{A^2 - \overline{A}^2}{2\overline{A}^2} \tag{40}$$

$$\psi_6 = \frac{\left(A - \overline{A}\right)^2}{2\overline{A}^2} \tag{41}$$

It can be shown that $\psi_5 - \psi_6 = \frac{A}{\overline{\overline{A}}} - 1$ and $\psi_5 + \psi_6 = \left(\frac{A}{\overline{\overline{A}}}\right)^2 - \frac{A}{\overline{\overline{A}}}$. These relations can be used to prove the following relations between the effective stress and the actual stress with regard to the two damage variables defined above:

$$\overline{\sigma} = \frac{1}{2}\left[1 \pm \sqrt{1 + 4(\psi_5 + \psi_6)}\right]\sigma \tag{42}$$

$$\overline{\sigma} = [1 + (\psi_5 - \psi_6)]\sigma \tag{43}$$

Thus, based on the evaluation of the damage effect variable M, the combined damage variable $\psi_5 + \psi_6$ is valid provided that it satisfies the condition $\frac{1}{2}\left[1 \pm \sqrt{1 + 4(\psi_5 + \psi_6)}\right] \geq 1$, while the combined damage variable $\psi_5 - \psi_6$ is valid provided that it satisfies the condition $[1 + (\psi_5 - \psi_6)] \geq 1$.

Consider now two new damage variables ψ_7 and ψ_8 defined as follows:

$$\psi_7 = \frac{A^2 - \overline{A}^2}{2A^2} \tag{44}$$

$$\psi_8 = \frac{\left(A - \overline{A}\right)^2}{2A^2} \tag{45}$$

It can be shown easily that $\psi_7 + \psi_8 = 1 - \frac{1}{\left(\frac{A}{\overline{\overline{A}}}\right)}$ and $\psi_7 - \psi_8 = \frac{1}{\left(\frac{A}{\overline{\overline{A}}}\right)} - \frac{1}{\left(\frac{A}{\overline{\overline{A}}}\right)^2}$. One can use these relations to prove the following relations between the effective stress and the actual stress with regard to the two damage variables defined above:

$$\overline{\sigma} = \frac{\sigma}{1 - (\psi_7 + \psi_8)} \tag{46}$$

$$\overline{\sigma} = \frac{2\sigma}{1 \pm \sqrt{1 - 4(\psi_7 - \psi_8)}} \tag{47}$$

Thus, based on the evaluation of the damage effect variable M, the combined damage variable $\psi_7 + \psi_8$ is valid provided that it satisfies the condition $\frac{1}{1-(\psi_7+\psi_8)} \geq 1$, while the combined damage variable $\psi_7 - \psi_8$ is valid provided that it satisfies the condition $\frac{2}{1\pm\sqrt{1-4(\psi_7-\psi_8)}} \geq 1$.

Damage Variables Defined in Terms of Stiffness

Looking at the damage variable ℓ_1 defined in terms of stiffness in Eq. 3, a similar damage variable ℓ_2 is proposed that is also defined in terms of stiffness as follows (as was shown in Eq. 16):

$$\ell_2 = \frac{\overline{E} - E}{\overline{E}} \tag{48}$$

It can now be easily shown that the following relations between the damage variables ℓ_1 and ℓ_2 are obtained:

$$\ell_1 - \ell_2 = \ell_1 \ell_2 \tag{49}$$

$$\ell_1 + \ell_2 = \frac{\overline{E}^2 - E^2}{\overline{E}E} \tag{50}$$

Using the above results, two new damage variables are defined in terms of stiffness as follows:

$$p_1 = \sqrt{\ell_1 + \ell_2} \tag{51}$$

$$p_2 = \sqrt{\ell_1 - \ell_2}. \tag{52}$$

It can be easily shown that the following two relations are obtained between the effective stress and the actual stress based on the above damage variables using the hypothesis of elastic strain equivalence:

$$\overline{\sigma} = \frac{1}{2}\left(p_1^2 + \sqrt{p_1^4 + 4}\right)\sigma \tag{53}$$

$$\overline{\sigma} = \frac{1}{2}\left(p_2^2 + 2 \pm p_2\sqrt{p_2^2 + 4}\right)\sigma \tag{54}$$

Thus, based on the evaluation of the damage effect variable M, the proposed damage variable p_1 is valid provided that it satisfies the condition $\frac{1}{2}\left(p_1^2 + \sqrt{p_1^4 + 4}\right) \geq 1$, while the combined damage variable p_2 is valid provided that it satisfies the condition $\frac{1}{2}\left(p_2^2 + 2 \pm p_2\sqrt{p_2^2 + 4}\right) \geq 1$.

However, when using the hypothesis of elastic energy equivalence, it can be easily shown that the following two relations are obtained between the effective stress and the actual stress based on the above damage variables:

$$\overline{\sigma} = \sqrt{\frac{1}{2}\left(p_1^2 + \sqrt{p_1^4 + 4}\right)}\sigma \tag{55}$$

$$\overline{\sigma} = \sqrt{\frac{1}{2}\left(p_2^2 + 2 \pm p_2\sqrt{p_2^2+4}\right)}\,\sigma \tag{56}$$

Thus, based on the evaluation of the damage effect variable M, the proposed damage variable p_1 is valid provided that it satisfies the condition $\sqrt{\frac{1}{2}\left(p_1^2 + \sqrt{p_1^4+4}\right)} \geq 1$, while the combined damage variable p_2 is valid provided that it satisfies the condition $\sqrt{\frac{1}{2}\left(p_2^2 + 2 \pm p_2\sqrt{p_2^2+4}\right)} \geq 1$.

The basic damage variable ℓ_1 is now investigated in more detail. For the hypothesis of elastic strain equivalence, $\overline{\sigma} = \left(\frac{\overline{E}}{E}\right)\sigma$. In this case, it is already shown that the damage variable $\ell_1 = \frac{\overline{E}-E}{E}$ will yield the relation $\overline{\sigma} = \sigma(1+\ell_1)$.

Next, using the hypothesis of elastic energy equivalence, one has $\overline{\sigma} = \sqrt{\frac{\overline{E}}{E}}\sigma$. In this case, it can be easily shown that the damage variable $\ell_1 = \frac{\overline{E}-E}{E}$ will yield the relation $\overline{\sigma} = \sigma\sqrt{1+\ell_1}$.

If one postulates a new hypothesis of higher-order energy equivalence in the form $\frac{1}{2}\sigma^2\varepsilon = \frac{1}{2}\overline{\sigma}^2\overline{\varepsilon}$ or $\frac{1}{2}\sigma\varepsilon^2 = \frac{1}{2}\overline{\sigma}\overline{\varepsilon}^2$, the relation $\overline{\sigma} = \sqrt[3]{\frac{\overline{E}}{E}}$ is obtained. In this case, it can be easily shown that using $\ell_1 = \frac{\overline{E}-E}{E}$ will yield the relation $\overline{\sigma} = \sigma\sqrt[3]{1+\ell_1}$.

Postulating another new hypothesis of higher-order energy equivalence in the form $\frac{1}{2}\sigma^2\varepsilon^2 = \frac{1}{2}\overline{\sigma}^2\overline{\varepsilon}^2$, the relation $\overline{\sigma} = \sqrt[4]{\frac{\overline{E}}{E}}$ is obtained. In this case, it can be easily shown that using $\ell_1 = \frac{\overline{E}-E}{E}$ will yield the relation $\overline{\sigma} = \sigma\sqrt[4]{1+\ell_1}$.

Finally, if a new hypothesis is postulated for the generalized equivalence of higher-order energy in terms of the n-powers of σ and ε, then the following relation is obtained $\overline{\sigma} = \sqrt[n]{\frac{\overline{E}}{E}}$. In this case, it can be shown that using $\ell_1 = \frac{\overline{E}-E}{E}$ will yield the relation $\overline{\sigma} = \sigma\sqrt[n]{1+\ell_1}$.

Several curves are plotted on the same graph paper to show the relations between $\frac{\overline{\sigma}}{\sigma}$ (which is M, the ratio of the stresses) and ℓ_1 (see Fig. 12). It is clear that for the limiting case when $n \to \infty$, the curve has a constant value at 1. Note that the lower curves appearing in Fig. 12 are for larger values of n. What does this limiting case signify? Before addressing this question and comment on these results, the above investigation is repeated for the other damage variable ℓ_2.

Next, the basic damage variable ℓ_2 is investigated in more detail. For the hypothesis of elastic strain equivalence, we have $\overline{\sigma} = \left(\overline{E}/E\right)\sigma$. In this case, it has already been shown that the damage variable $\ell_2 = \left(\overline{E}-E\right)/E$ will yield the relation $\overline{\sigma} = \sigma(1-\ell_2)$.

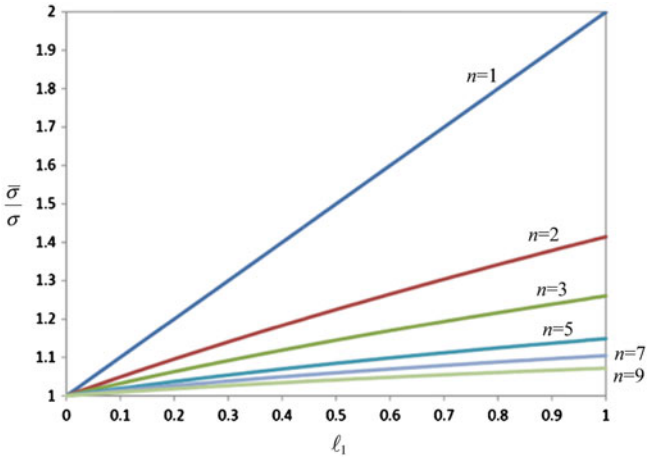


Fig. 12 Relation between ℓ_1 and the ratio of the stresses M (Reprinted with permission from Voyiadjis and Kattan 2012)

For the hypothesis of elastic energy equivalence, one obtains $\overline{\sigma} = \sqrt{\frac{\overline{E}}{E}}\sigma$. In this case, it can be easily shown that the damage variable $\ell_2 = \frac{\overline{E}-E}{\overline{E}}$ will yield the relation $\overline{\sigma} = \frac{\sigma}{\sqrt{1-\ell_2}}$.

If one postulates a new hypothesis of higher-order energy equivalence in the form $\frac{1}{2}\sigma^2\varepsilon = \frac{1}{2}\overline{\sigma}^2\overline{\varepsilon}$ or $\frac{1}{2}\sigma\varepsilon^2 = \frac{1}{2}\overline{\sigma}\overline{\varepsilon}^2$, the following relation is obtained $\overline{\sigma} = \sqrt[3]{\frac{\overline{E}}{E}}\sigma$. In this case, it can be easily shown that using $\ell_2 = \frac{\overline{E}-E}{\overline{E}}$ will yield the relation $\overline{\sigma} = \frac{\sigma}{\sqrt[3]{1-\ell_2}}$.

Postulating another new hypothesis of higher-order energy equivalence in the form $\frac{1}{2}\sigma^2\varepsilon^2 = \frac{1}{2}\overline{\sigma}^2\overline{\varepsilon}^2$, the following relation is obtained $\overline{\sigma} = \sqrt[4]{\frac{\overline{E}}{E}}\sigma$. In this case, it can be easily shown that using $\ell_2 = \frac{\overline{E}-E}{\overline{E}}$ will yield the relation $\overline{\sigma} = \frac{\sigma}{\sqrt[4]{1-\ell_2}}$.

Finally, a generalized equivalence of higher-order energy hypothesis is postulated in terms of the n-powers of σ and ε, such that $\overline{\sigma} = \sqrt[n]{\frac{\overline{E}}{E}}\sigma$. In this case, by using $\ell_2 = \frac{\overline{E}-E}{\overline{E}}$, one obtains the relation $\overline{\sigma} = \frac{\sigma}{\sqrt[n]{1-\ell_2}}$.

Plotting the two curves on the same graph to show the relations between $\frac{\overline{\sigma}}{\sigma}$ (which is M, the ratio of the stresses) and ℓ_2 (see Fig. 13), it is clear that for the limiting case when $n \to \infty$, the curve has a constant value at $n = 1$. Note that the lower curve appearing in Fig. 13 is for the larger value of n. What does this limiting case signify?

The above results are explained using the formulas derived for ℓ_2. A similar treatment may be applied to ℓ_1. One starts with the formula $\overline{\sigma} = \frac{\sigma}{\sqrt[n]{1-\ell_2}}$ which was derived in the previous paragraphs. The case when $n \to \infty$ is now investigated. In this case, the following relation is obtained:

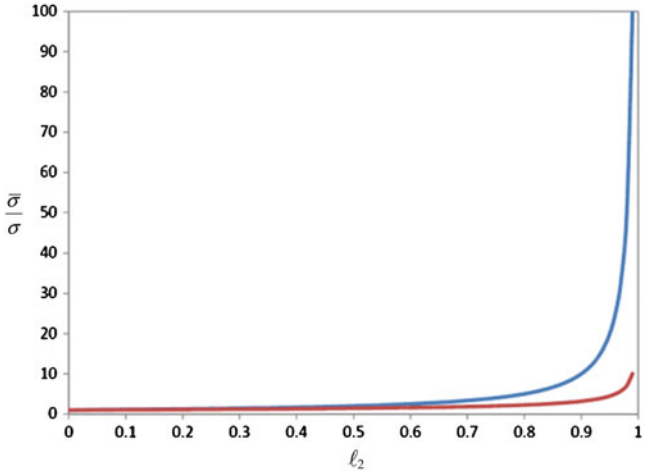


Fig. 13 Relation between ℓ_2 and the ratio of the stresses M (Reprinted with permission from Voyiadjis and Kattan 2012)

$$\overline{\sigma} = \frac{\sigma}{\sqrt[n]{1-\ell_2}} = \frac{\sigma}{(1-\ell_2)^{\frac{1}{n}}} = \frac{\sigma}{(1-\ell_2)^{\frac{1}{\infty}}} = \frac{\sigma}{(1-\ell_2)^{0}} = \frac{\sigma}{1} = \sigma \tag{57}$$

The relation $\overline{\sigma} = \sigma$ is now obtained irrespective of the value of the damage variable ℓ_2. This means that in this limiting case, the material remains totally undamaged irrespective of the value of the damage variable. Of course, this is a hypothetical case as it cannot be reached physically. However, it gives rise to the following new issue. Does a material exist or can be manufactured which remains totally undamaged during the deformation process no matter what the load may be. This is the holy grail of damage mechanics and materials science in general – to design a new type of material that cannot be damaged at all. The above equations will provide some guidelines in this respect. The issue of the proposed higher-order strain energy forms will be discussed in detail in the next section.

Higher-Order Strain Energy Forms

In this section, the nature of the proposed higher-order strain energy forms is investigated that were introduced in section "Damage Variables Defined in Terms of Stiffness." Their exact relationships to the elastic constitutive equations of the material are investigated. It will be seen that each higher-order strain energy form will correspond to an exact nonlinear elastic stress–strain relation. These specific stress–strain relations will be derived in this section. Looking at these proposed higher-order strain energy forms, it is clearly seen that some of them do not have units of energy. Thus, these forms may be called hyper-strain energy forms.

The general formulation is first considered of how to derive a specific stress–strain relation if the corresponding strain energy form is known. For example, in the linear case, the linear stress–strain relation $\sigma = E\,\varepsilon$ corresponds to the

usual strain energy form $U = \frac{1}{2}\sigma\varepsilon$. The types of stress–strain relations are investigated that may be obtained using the proposed higher-order strain energy forms $\frac{1}{2}\sigma\varepsilon^2$, $\frac{1}{2}\sigma^2\varepsilon$, $\frac{1}{2}\sigma\varepsilon^3$, $\frac{1}{2}\sigma^3\varepsilon$, $\frac{1}{2}\sigma^2\varepsilon^2$, and so on with higher powers.

In general, the following general stress–strain relation is assumed:

$\sigma = E\,f(\varepsilon)$, where $f(\varepsilon)$ is to be determined for each specific form of higher-order strain energy. The higher-order strain energy U is defined (some form of hyper-strain energy) as the area under the stress–strain curve. It is given by the following relation:

$$U = \int \sigma\, d\varepsilon \tag{58}$$

Substituting the above general stress–strain relation in the expression for U, one obtains

$$U = E\int f(\varepsilon)\, d\varepsilon \tag{59}$$

Taking the derivative of both sides, the following formula for $f(\varepsilon)$ is obtained:

$$f(\varepsilon) = \frac{dU/d\varepsilon}{E} \tag{60}$$

Thus, the above formula can be used to obtain the specific function of the sought nonlinear stress–strain relation for the material. This is illustrated with one example using the higher-order strain energy form $U = \frac{1}{2}\sigma\varepsilon^2$.

The higher-order strain energy form $U = \frac{1}{2}\sigma\varepsilon^2$ is assumed. Substituting $\sigma = E\,f(\varepsilon)$ in the above formula for U one obtains

$$U = \frac{1}{2}Ef(\varepsilon)\varepsilon^2 \tag{61}$$

Taking the derivative of the above formula the following expression is derived:

$$dU = \frac{1}{2}Ef'(\varepsilon)\varepsilon^2 d\varepsilon + Ef(\varepsilon)\varepsilon\ \ d\varepsilon \tag{62}$$

One now substitutes the above formula into Eq. 60 to obtain

$$f(\varepsilon) = \frac{1}{2}f'(\varepsilon)\varepsilon^2 + f(\varepsilon)\varepsilon \tag{63}$$

or

$$\frac{1}{2}f'(\varepsilon)\varepsilon^2 + f(\varepsilon)\varepsilon - f(\varepsilon) = 0 \tag{64}$$

Next, the above differential equation is solved in order to obtain the desired nonlinear relation between stress and strain as follows. The solution to the above ordinary differential equation is obtained as follows:

$$f(\varepsilon) = e^{C}\,\frac{1}{\varepsilon^2}\,e^{-2/\varepsilon} \tag{65}$$

Substituting back into the original equation $\sigma = E\,f(\varepsilon)$, one obtains

$$\sigma = E\,e^{C}\,\frac{1}{\varepsilon^2}\,e^{-2/\varepsilon} \tag{66}$$

One can assume the constant $C = 0$ to obtain the following nonlinear stress–strain relation:

$$\sigma = E\,\frac{1}{\varepsilon^2}\,e^{-2/\varepsilon} \tag{67}$$

Looking at the above stress–strain relation, it may appear that the stress approaches infinity at the initial condition when the strain is zero, but this is not the case. It is seen that the limit of the above expression for the stress approaches zero as the strain approaches zero. Thus, the initial conditions of zero strain and zero stress are satisfied.

Thus, it is seen that for the higher-order strain energy form $\frac{1}{2}\sigma\varepsilon^2$, the corresponding stress–strain relation is nonlinear and is given by the expression $\sigma = E\,\frac{1}{\varepsilon^2}e^{-2/\varepsilon}$.

The above procedure can be repeated for the other proposed higher-order strain energy forms to obtain their corresponding nonlinear stress–strain relations. These results are summarized in Table 3.

Looking at the results presented in Table 3, one notices that the higher-order strain energy forms are of three types. The first type is the one with powers of ε. This type is well behaved and indeed has units of energy. The other two types are the ones with powers of σ and with mixed powers of both ε and σ. These two types are not well behaved and do not have units of energy. In fact, these powers do not

Table 3 The proposed higher-order strain energy forms and their corresponding stress–strain relations (Reprinted with permission from Voyiadjis and Kattan 2012)

Proposed higher-order strain energy form	Valid or invalid	Corresponding stress–strain relation
$U = \frac{1}{2}\sigma\varepsilon$	Valid	$\sigma = E\,\varepsilon$
$U = \frac{1}{2}\sigma\varepsilon^2$	Valid	$\sigma = E\,\frac{1}{\varepsilon^2}e^{-2/\varepsilon}$
$U = \frac{1}{2}\sigma^2\varepsilon$	Invalid	$\sigma = 2\left[1 - \frac{1}{\left(1-\frac{1}{2}Ef(\varepsilon_o)\right)\sqrt{\varepsilon_o\,\varepsilon}}\right]$
$U = \frac{1}{2}\sigma\varepsilon^3$	Valid	$\sigma = E\,\frac{1}{\varepsilon^3}e^{-1/\varepsilon^2}$
$U = \frac{1}{2}\sigma^3\varepsilon$	Invalid	$\sigma = \frac{\sqrt{2-\left(\varepsilon_o^{2/3}+2-E^2 f_o^2(\varepsilon)\right)\varepsilon^{-2/3}}}{E}$
$U = \frac{1}{2}\sigma^2\varepsilon^2$	Invalid	$\sigma = \frac{1}{\varepsilon}(\ln\varepsilon + CE)$
$U = \frac{1}{2}\sigma\varepsilon^n, n = 1, 2, 3, \ldots.$	Valid	$\sigma = E\,\frac{1}{\varepsilon^n}e^{-2/\left[(n-1)\varepsilon^{(n-1)}\right]}$

satisfy the initial conditions that $\sigma = 0$ when $\varepsilon = 0$. Thus, the results for these two types of higher-order energy forms are discarded.

One now investigates in more detail the higher-order strain energy forms with powers of ε only. The general expression of these types of energy forms is given by $U = \frac{1}{2}\sigma\varepsilon^n$. The corresponding nonlinear stress–strain relation for this general form is given by $\sigma = E\,\frac{1}{\varepsilon^n}e^{-2/\left[(n-1)\varepsilon^{(n-1)}\right]}$ as seen from Table 3. It is seen that this general form satisfies the initial conditions $\sigma = 0$ when $\varepsilon = 0$ because the limit of the expression for the stress approaches zero as the strain approaches zero. Thus, the higher-order strain energy forms with powers of strain only are valid and can be used in this discussion of undamageable materials as illustrated at the end of section "Hypothesis of Elastic Energy Equivalence." Note also that the special valid cases appearing at the top of Table 3 can be deduced directly from the general solution.

Thus, in order to design an undamageable material as explained at the end of section "Hypothesis of Elastic Energy Equivalence," the stress–strain relation of this hypothetical material will have to follow the highly nonlinear stress–strain relation given in the last row of Table 3 when one takes the limit as the value of n goes to infinity. This presentation is given here in its full theoretical framework in the hope that sometime in the future, the manufacturing technology will advance to such a stage that such a material can be realized. The constitutive relations of undamageable materials are similar to those of rubber materials (Arruda and Boyce 1993).

Hybrid Damage Variables

In the remaining part of this work, two new hybrid damage variables are introduced. These variables are more complicated than the previously introduced ones. They are termed "hybrid" because they are defined in terms of both area and stiffness. The hybrid damage variables introduced here may be applied to homogeneous materials like metals as well as for heterogeneous materials like composite laminates.

Using the usual damage variables $\phi = \frac{A-\overline{A}}{A}$ and $\ell = \frac{\overline{E}-E}{E}$, two new damage variables s_1 and s_2 are defined as follows:

$$s_1 = \phi + \ell - \phi\ell \tag{68}$$

$$s_2 = \phi + \ell + \phi\ell \tag{69}$$

These two new damage variables are called hybrid damage variables because each one includes both parameters ϕ and ℓ. It will be shown in this and subsequent derivations that these two damage variables are not independent. Substituting the equations for ϕ and ℓ into the equations of the two new proposed damage variables, one obtains two simultaneous algebraic equations in the variables x and y as follows:

$$s_1 = 1 - 2y + xy \tag{70}$$

$$s_2 = -1 + 2x - xy \tag{71}$$

where $x = \frac{\overline{E}}{E}$ and $y = \frac{\overline{A}}{A}$.

Next, one solves the two simultaneous equations shown above to obtain:

$$x = \frac{1 + \frac{1}{2}s_1 - \frac{1}{2}s_2 + \frac{1}{2}\sqrt{8s_1 - 8s_2 + s_1^2 + 2s_1s_2 + s_2^2}}{1 - \frac{1}{4}s_1 - \frac{1}{4}s_2 + \frac{1}{4}\sqrt{8s_1 - 8s_2 + s_1^2 + 2s_1s_2 + s_2^2}} \tag{72}$$

$$y = 1 - \frac{1}{4}s_1 - \frac{1}{4}s_2 + \frac{1}{4}\sqrt{8s_1 - 8s_2 + s_1^2 + 2s_1s_2 + s_2^2} \tag{73}$$

It can now be easily shown that

$$\overline{\sigma} = \frac{1 + \frac{1}{4}(s_1 + s_2 - c)}{-1 - \frac{1}{2}(s_1 - s_2 + c)}\,\sigma \tag{74}$$

where $c = \sqrt{8(s_2 - s_2) + (s_1 + s_2)^2}$.

Using the result for y in Eq. 73 along with the hypothesis of elastic strain equivalence, one can show that

$$\overline{\sigma} = \left[1 - \frac{1}{4}(s_1 + s_2 - c)\right]\sigma \tag{75}$$

Making use the result for y in Eq. 73 along with the hypothesis of elastic energy equivalence, one obtains:

$$\overline{\sigma} = \sigma\sqrt{1 - \frac{1}{4}(s_1 + s_2 - c)} \tag{76}$$

A Practical Example: Matrix Cracking in Cross-Ply Laminates

In this section, a practical example is provided for the stiffness reduction in carbon fiber-reinforced cross-ply laminated composite materials. This stiffness reduction is linked to the matrix cracking of the material. It has been shown through experimental data analysis of $(90/0)_s$ laminates (Silberschmidt 1997) that the following empirical relation for stiffness reduction is introduced:

$$\frac{E}{\overline{E}} = 1 - \frac{c}{2s} \tag{77}$$

where $2s$ is the average spacing between neighboring cracks in the matrix and c is a constant. The above formula provides for a valid physical and experimental basis for the class of damage variables that are defined in terms of stiffness reduction.

A more general relation that accounts for a wider set of laminate characteristics was also presented by Silberschmidt (1997) which is given by the following formula:

$$\frac{E}{\overline{E}} = \left[1 + \left(\frac{(b+d)\overline{E}}{b\overline{E}_1} - 1\right)\frac{\tanh(\lambda s)}{\lambda s}\right]^{-1} \tag{78}$$

where

$$\lambda^2 = \frac{BG_{23}(b+d)\overline{E}}{bd^2\overline{E}_2\overline{E}_1} \tag{79}$$

b and d are the thickness of the 0° and 90° plies, respectively, G_{23} is a shear modulus, and $\overline{E}_1$ and $\overline{E}_2$ are the uncracked longitudinal and transverse moduli of the ply, respectively. The constant B is equal to 2 under an assumption of linear displacement profile in the cracked 90° plies and equal to 3 for a parabolic profile (Silberschmidt 1997).

The empirical formulae in Eqs. 77 and 78 provide for a practical and experimental validation of the class of proposed damage variables that are defined in terms of stiffness reduction.

Anisotropic Damage in Continuum Damage Mechanics

Review of Existing Damage Variables

The principles of the continuum mechanics theory for the general case of anisotropic damage were recently (Voyiadjis and Park 1997; Voyiadjis and Kattan 1992, 1990, 2012; Kattan and Voyiadjis 1993; Cordebois and Sidoroff 1982) cast in a consistent mathematical and mechanical framework. The generalized form of the transformation equation for the anisotropic case, in indicial notation, is written as follows (Murakami 1988; Chow and Wang 1988):

$$\overline{\sigma}_{ij} = M_{ijkl}\sigma_{kl} \tag{80}$$

where M is a symmetric fourth-rank tensor termed the damage effect tensor, σ is the Cauchy stress tensor, and $\overline{\sigma}$ is the corresponding effective stress tensor (see Fig. 1).

In order to compute the components of the effective elasticity tensor $\overline{E}$ (which is defined in the fictitious configuration), the hypotheses of elastic energy equivalence are used.

The second anisotropic damage tensor L may be defined in terms of the reduction in the elastic stiffness components as follows:

$$L^{(1)}_{ijmn} = \left(\overline{E}_{ijkl} - E_{ijkl}\right)E^{-1}_{klmn} \tag{81a}$$

$$L^{(2)}_{ijmn} = E^{-1}_{ijkl}\left(\overline{E}_{klmn} - E_{klmn}\right) \tag{81b}$$

where E is the elastic stiffness tensor in the damaged state, while $\overline{E}$ is the effective elastic tensor (in the fictitious state). This damage variable was used recently by

Celentano et al. (2004), Nichols and Abell (2003), and Nichols and Totoev (1999) for the scalar case only. It should also be mentioned that Voyiadjis (1988) used a similar relation but in the context of elastoplastic deformation. See also Voyiadjis and Kattan (2009). Two such tensors are proposed in this case as shown in Eqs. 81a and 81b.

The definition of the alternative damage variables of Eqs. 81a and 81b may be rewritten in the following more appropriate form:

$$\overline{E}_{ijkl} = \left(\delta_{im}\delta_{jn} + L^{(1)}_{ijmn}\right)E_{mnkl} \tag{82a}$$

$$\overline{E}_{ijkl} = E_{ijmn}\left(\delta_{mk}\delta_{nl} + L^{(2)}_{mnkl}\right) \tag{82b}$$

Hypothesis of Elastic Strain Equivalence

Making use of the hypothesis of elastic strain equivalence, one assumes

$$\overline{\epsilon}_{ij} = \epsilon_{ij} \tag{83}$$

Using the elastic constitutive relations in both configurations as follows

$$\sigma_{ij} = E_{ijkl}\epsilon_{kl} \tag{84}$$

$$\overline{\sigma}_{ij} = \overline{E}_{ijkl}\overline{\epsilon}_{kl} \tag{85}$$

and substituting for $\overline{\sigma}$ and $\overline{\varepsilon}$ using Eqs. 80 and 83, respectively, into Eq. 85 and comparing the result with Eq. 84, one obtains

$$\overline{E}_{ijmn} = M_{ijkl}E_{klmn} \tag{86}$$

In order to find an appropriate relation between the two damage tensors M and L, one equates Eqs. 82a and 82b with 86 to obtain

$$L^{(1)}_{ijmn} = M_{ijmn} - \delta_{im}\delta_{jn} \tag{87a}$$

$$L^{(2)}_{ijmn} = \delta_{im}\delta_{jn}\left(\frac{1}{9}\alpha - 1\right) \tag{87b}$$

where

$$\alpha = M_{ijij}$$

The above expression defines the exact relation between the two damage tensors L and M in the case of the hypothesis of elastic strain equivalence.

The relations in Eqs. 87a and 87b may be rewritten in the following form:

$$M_{ijmn} = \delta_{im}\delta_{jn} + \frac{1}{9}\delta_{im}\delta_{jn}L^{(2)}_{klkl} \tag{88a}$$

$$M_{ijmn} = L^{(1)}_{ijmn} + \delta_{im}\delta_{jn} \tag{88b}$$

It is clear that the above special relations for the case of anisotropic damage are linear relations in terms of the damage effect tensor and the new proposed damage tensors. It can be verified that the damage tensors $L^{(1)}$ and $L^{(2)}$ are valid for values of $L^{(1)}_{1111}$ and $L^{(2)}_{1111}$ equal to or greater than zero.

Hypothesis of Elastic Energy Equivalence

Using the hypothesis of elastic energy equivalence undamageable material, by assuming the complementary elastic strain energy to be equal in both configurations, one obtains:

$$E^{-1}_{ijkl}\,\sigma_{ij}\,\sigma_{kl} = \overline{E}^{-1}_{mnpq}\,\overline{\sigma}_{mn}\,\overline{\sigma}_{pq} \tag{89}$$

Substituting for $\overline{\sigma}$ from Eq. 80 into Eq. 89 and simplifying, one obtains

$$\overline{E}_{ijkl} = M_{ijmn}\,E_{mnpq}\,M^{T}_{pqkl} \tag{90}$$

Next, in order to derive the relation between the two damage tensors M and L, one equates Eqs. 82a and 82b with 90 to obtain

$$L^{(1)}_{ijmn} = \delta_{im}\delta_{jn} - \frac{1}{9}\,\alpha\,M_{ijmn} \tag{91a}$$

$$L^{(2)}_{ijmn} = \frac{1}{9}\,\alpha\,M_{ijmn} - \delta_{im}\delta_{jn} \tag{91b}$$

The above expression defines the exact relation between the two damage tensors L and M in the case of the hypothesis of elastic energy equivalence.

The relations in Eqs. 91a and 91b may be rewritten in the following form:

$$M_{ijmn} = \frac{9}{\alpha}\left[\delta_{im}\delta_{jn} - L^{(1)}_{ijmn}\right] \tag{92a}$$

$$M_{ijmn} = \frac{9}{\alpha}\left[\delta_{im}\delta_{jn} + L^{(2)}_{ijmn}\right] \tag{92b}$$

It is clear that the above special relations for the case of anisotropic damage are nonlinear relations in terms of the damage effect tensor and the new proposed damage tensors. It can be verified that the damage tensors $L^{(1)}$ and $L^{(2)}$ are valid for values of $L^{(1)}_{1111}$ less than zero and values of $L^{(2)}_{1111}$ greater than zero.

Proposed New Damage Tensors

In this section, two alternative proposed new damage tensors are presented that are similar to $L^{(1)}$ and $L^{(2)}$. These two new damage tensors will be termed $P^{(1)}$ and $P^{(2)}$, respectively, and defined as follows:

The two new damage tensors are given by

$$P^{(1)}_{ijmn} = \left(\overline{E}_{ijkl} - E_{ijkl}\right)\overline{E}_{klmn} \tag{93a}$$

$$P^{(2)}_{ijmn} = \overline{E}_{klmn}\left(\overline{E}_{ijkl} - E_{ijkl}\right) \tag{93b}$$

Note that the difference between $L^{(1)}$ and $P^{(1)}$ and between $L^{(2)}$ and $P^{(2)}$ is that the elasticity tensor in the damaged configuration is replaced by the elasticity tensor in the undamaged configuration. These changes are made in order to investigate the relationships between the various damage tensors and assess their validity. Next, these two new proposed damage tensors are investigated using the two hypotheses outlined above.

Hypothesis of Elastic Strain Equivalence

Using the hypothesis of elastic strain equivalence, the following relations are obtained between M and P:

$$P^{(1)}_{ijmn} = \delta_{im}\delta_{jn} - M^{-1}_{ijmn} \tag{94a}$$

$$P^{(2)}_{ijmn} = \delta_{im}\delta_{jn}\left(1 - \frac{1}{9}\beta\right) \tag{94b}$$

where $\beta = M^{-1}_{ijij}$

The derivation of the above equations is straightforward and follows the same procedure used in section "Hypothesis of Elastic Strain Equivalence." The above relationships can also be rewritten as follows:

$$M_{ijmn} = \left[\delta_{\mathrm{im}}\delta_{\mathrm{jn}} - P^{(1)}_{ijmn}\right]^{-\mathrm{T}} \tag{95a}$$

$$M_{ijmn} = \delta_{im}\delta_{jn}\left[1 - \frac{1}{9}P^{(2)}_{klkl}\right]^{-T} \tag{95b}$$

It is clear that the above special relations for the case of anisotropic damage are nonlinear relations in terms of the damage effect tensor and the new proposed damage tensors. It can be verified that the damage tensors $P^{(1)}$ and $P^{(2)}$ are valid for values of $P^{(1)}_{1111}$ and $P^{(2)}_{1111}$ equal to or greater than zero.

Hypothesis of Elastic Energy Equivalence

Using the hypothesis of elastic energy equivalence, the following relations between M and P are obtained:

$$P^{(1)}_{ijmn} = \delta_{im}\delta_{jn} - \frac{1}{9}\beta\, M^{-1}_{ijmn} \tag{96a}$$

$$P^{(2)}_{ijmn} = \delta_{im}\delta_{jn} - \frac{1}{9}\beta\, M^{-1}_{ijmn} \tag{96b}$$

The above relationships may be rewritten as follows:

$$M_{ijmn} = \frac{\beta}{9}\left[\ \delta_{im}\delta_{jn} - P^{(1)}_{ijmn}\right]^{-T} \tag{97a}$$

$$M_{ijmn} = \frac{\beta}{9}\left[\ \delta_{im}\delta_{jn} - P^{(2)}_{ijmn}\right]^{-T} \tag{97b}$$

It is clear that the above special relations for the case of anisotropic damage are nonlinear relations in terms of the damage effect tensor and the new proposed damage tensors. It can be verified that the damage tensors $P^{(1)}$ and $P^{(2)}$ are valid for values of $P^{(1)}_{1111}$ and $P^{(2)}_{1111}$ equal to or greater than zero.

Other Proposed New Damage Tensors

In this part of research work, new anisotropic damage tensors are defined in terms of the fourth-rank elasticity tensor. The hypotheses of the elastic strain equivalence and the elastic energy equivalence are employed in order to obtain the anisotropic tensors. In this section, the new proposed anisotropic damage tensors are denoted by the letter R.

1. The anisotropic damage tensor R is given by

$$R_{ijkl} = L^{(1)}_{ijkl} + L^{(2)}_{ijkl} \tag{98}$$

Substituting Eqs. 87a and 87b into Eq. 98 and simplifying the terms result in the following equation of the anisotropic damage tensor:

$$R_{ijkl} = M_{ijkl} + \delta_{ik}\delta_{jl}\left(\frac{1}{9}\ \alpha - 2\right) \tag{99a}$$

It is demonstrated that Eq. 99a is derived based on the hypothesis of elastic strain energy equivalence. It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero.

The substitution of Eqs. 91a and 91b into Eq. 98, however, gives the anisotropic damage tensor based on the hypothesis of elastic energy equivalence:

$$R_{ijkl} = 0 \tag{99b}$$

Comparing Eqs. 99a and 99b, one notes that the damage tensor for the second hypothesis is essentially zero. Thus, the damage tensor that is proposed for the hypothesis of elastic energy equivalence in this case is not a valid damage tensor.

2. The anisotropic damage tensor R is given by

$$R_{ijkl} = L^{(1)}_{ijkl} - L^{(2)}_{ijkl} \tag{100}$$

Substituting Eqs. 87a and 87b into Eq. 100 and simplifying the terms result in the following equation of the anisotropic damage tensor based on the hypothesis of elastic strain equivalence:

$$R_{ijkl} = M_{ijkl} - \frac{1}{9}\,\alpha\,\delta_{ik}\delta_{jl} \tag{101}$$

It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero. Substituting Eqs. 91a and 91b into Eq. 100 and simplifying the terms result in the following equation of the anisotropic damage tensor based on the hypothesis of elastic energy equivalence:

$$R_{ijkl} = 2\left(\delta_{ik}\delta_{jl} - \frac{1}{9}\,\alpha\,M_{ijkl}\right) \tag{102}$$

It can be verified that the damage tensor R is valid for values of R_{1111} less than zero.

3. The anisotropic damage tensor R is given by

$$R_{ijkl} = \left(\overline{E}_{ijmn}\overline{E}_{mnpq} - E_{ijmn}E_{mnpq}\right)\overline{E}^{-1}_{pqrs}E^{-1}_{rskl} \tag{103}$$

Equation 103 may be rewritten in the following form:

$$R_{ijkl}E_{klrs}\overline{E}_{rspq} = \overline{E}_{ijmn}\overline{E}_{mnpq} - E_{ijmn}E_{mnpq} \tag{104}$$

Substituting Eq. 86 into Eq. 104 and simplifying the terms, the following equation is derived:

$$R_{ijkl} = M_{ijkl} - \beta\,\delta_{ik}\delta_{jl} \tag{105a}$$

Equation 105a represents the anisotropic damage tensor based on the hypothesis of elastic strain equivalence. It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero.

However, the hypothesis of the elastic energy is applied to obtain the anisotropic damage tensor. Substituting Eq. 90 into Eq. 104 and simplifying give the following equation:

$$R_{ijkl} = \alpha\, M_{ijkl} - \beta^2\, \delta_{ik}\delta_{jl} \tag{105b}$$

Equation 105b represents the anisotropic damage tensor based on the hypothesis of elastic strain equivalence. It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero.

4. The anisotropic damage tensor R is given by

$$R_{ijkl} = \left(\overline{E}_{ijmn}\overline{E}_{mnkl} - E_{ijmn}E_{mnkl}\right)\overline{E}^{-1}_{rspq}E^{-1}_{rspq} \tag{106}$$

Equation 106 may be rewritten in the following form:

$$R_{ijkl}E_{pqrs}\overline{E}_{pqrs} = \overline{E}_{ijmn}\overline{E}_{mnkl} - E_{ijmn}E_{mnkl} \tag{107}$$

Using the hypothesis of elastic strain equivalence and substituting Eq. 86 into Eq. 107 give the following simple expression for the anisotropic damage tensor, Eq. 108a:

$$R_{ijkl} = \frac{1}{9}\left(M_{ijkl} - M^{-1}_{ijkl}\right) \tag{108a}$$

It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero.

Substituting Eq. 90 into Eq. 107, on the other hand, gives the anisotropic damage tensor based on the hypothesis of elastic energy equivalence:

$$R_{ijkl} = M_{ijpq}\, M_{pqkl} - M^{-1}_{ijtu}M^{-1}_{tukl} \tag{108b}$$

It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero.

5. The anisotropic damage tensor R is given by

$$R_{ijkl} = \overline{E}^{-1}_{ijmn}E_{mnpq}\left(\overline{E}_{pqrs}\overline{E}_{rskl} - E_{pqrs}E_{rskl}\right) \tag{109}$$

Equation 109 may be rewritten in the following form:

$$E_{pqmn}\overline{E}_{mnij}R_{ijkl} = \overline{E}_{pqrs}\overline{E}_{rskl} - E_{pqrs}E_{rskl} \tag{110}$$

Substituting Eq. 86 into Eq. 110 and simplifying give the following equation:

$$R_{ijkl} = \frac{1}{9}\delta_{ik}\delta_{jl}[\alpha - 9\,\beta] \tag{111a}$$

It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero.

Equation 111a demonstrates the anisotropic damage tensor based on the hypothesis of elastic strain equivalence. In addition, substitution of Eq. 90 into Eq. 110 and simplification of the equations give the anisotropic damage tensor using the hypothesis of elastic energy equivalence:

$$R_{ijkl} = \frac{1}{9}\alpha^2\delta_{ik}\delta_{jl} - \beta\, M^{-1}_{ijkl} \tag{111b}$$

It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero.

6. The anisotropic damage tensor is given by

$$R_{ijkl} = \overline{E}^{-1}_{rspq}E^{-1}_{rspq}\left(\overline{E}_{ijmn}\overline{E}_{mnkl} - E_{ijmn}E_{mnkl}\right) \tag{112}$$

Equation 112 may be rewritten in the following form:

$$E_{pqrs}\overline{E}_{pqrs}R_{ijkl} = \overline{E}_{ijmn}\overline{E}_{mnkl} - E_{ijmn}E_{mnkl} \tag{113}$$

Substituting Eq. 86 into Eq. 113 and simplifying give the following equation:

$$R_{ijkl} = \frac{1}{9}\left(M_{ijkl} - M^{-1}_{ijkl}\right) \tag{114a}$$

The above expression is exactly the same as the one derived previously in Eq. 108a. It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero.

Equation 114a represents the anisotropic damage tensor based on the hypothesis of elastic strain equivalence. However, the hypothesis of the elastic energy is applied to obtain the anisotropic damage tensor. Substituting Eq. 90 into Eq. 113 and simplifying give the following relation:

$$R_{ijkl} = M_{ijmn}\,M_{mnkl} - M^{-1}_{ijtu}M^{-1}_{tukl} \tag{114b}$$

The above expression is exactly the same as the one derived previously in Eq. 108b. It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero.

7. The anisotropic damage tensor is R given by

$$R_{ijkl} = \left(\overline{E}_{ijmn} - E_{ijmn}\right)\left(\overline{E}_{mnpq} - E_{mnpq}\right)\overline{E}^{-1}_{pqrs}E^{-1}_{rskl} \tag{115}$$

Equation 115 may be rewritten in the following form:

$$R_{ijkl}E_{klrs}\overline{E}_{rspq} = \overline{E}_{mnpq} - \overline{E}_{ijmn}E_{mnpq} - E_{ijmn}\overline{E}_{mnpq} + E_{ijmn}E_{mnpq} \tag{116}$$

Substituting Eq. 86 into Eq. 116 and simplifying give the following equation:

$$R_{ijkl} = \left[\delta_{ik}\delta_{jl} - M_{ijkl}\right][\beta - 9] \tag{117a}$$

It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero.

Equation 117a represents the anisotropic damage tensor based on the hypothesis of elastic strain equivalence. However, the hypothesis of the elastic energy is applied to obtain the anisotropic damage tensor. Substituting Eq. 90 into Eq. 116 and simplifying give the following relation:

$$R_{ijkl} = \delta_{ik}\delta_{jl}[\beta - 9][\beta + 9] \tag{117b}$$

It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero.

8. The anisotropic damage tensor R is given by

$$R_{ijkl} = \overline{E}^{-1}_{ijmn}E^{-1}_{mnpq}\left(\overline{E}_{pqrs} - E_{pqrs}\right)\left(\overline{E}_{rskl} - E_{rskl}\right) \tag{118}$$

Equation 118 may be rewritten in the following form:

$$E_{pqmn}\overline{E}_{mnij}R_{ijkl} = \overline{E}_{pqrs}\overline{E}_{rskl} - \overline{E}_{pqrs}E_{rskl} - E_{pqrs}\overline{E}_{rskl} + E_{pqrs}E_{rskl} \tag{119}$$

Substituting Eq. 86 into Eq. 119 and simplifying give the following equation:

$$R_{ijkl} = \left[M_{ijkl} - \delta_{ik}\delta_{jl}\right][9 - \beta] \tag{120a}$$

It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero.

Equation 120a demonstrates the anisotropic damage tensor based on the hypothesis of elastic strain equivalence. In addition, substitution of Eq. 90 into Eq. 119 and simplification of the equations give the anisotropic damage tensor using the hypothesis of elastic energy equivalence.

$$R_{ijkl} = \frac{1}{9}\delta_{ik}\delta_{jl}[\alpha - 9][\alpha + 9] - M^{-1}_{ijkl}[\alpha - \beta] \tag{120b}$$

It can be verified that the damage tensor R is valid for values of R_{1111} greater than zero. In section "Hybrid Damage Tensors," few new hybrid anisotropic damage tensors are presented.

Hybrid Damage Tensors

New hybrid damage tensors are proposed to be composed of the effective damage tensor and the elasticity tensor, and they are given by

$$N_{ijkl} = M_{ijkl} + L^{(1)}_{ijkl} - M_{ijmn}\, L^{(1)}_{mnkl} \tag{121a}$$

$$N_{ijkl} = M_{ijkl} + L^{(2)}_{ijkl} - M_{ijmn}\, L^{(2)}_{mnkl} \tag{121b}$$

Using Eqs. 87a and 91a, the following relations are obtained:

$$N_{ijkl} = 3M_{ijkl} - M_{ijmn}M_{mnkl} \tag{122a}$$

$$N_{ijkl} = \delta_{ik}\delta_{jl} + \frac{1}{9}\alpha\, M_{ijmn}\,[\, M_{mnkl} - \delta_{mk}\delta_{nl}\,] \tag{122b}$$

Equations 122a and 122b give the hybrid damage tensor expressions using the damage tensor $L^{(1)}$ based on the hypothesis of elastic strain equivalence and the hypothesis of elastic energy equivalence, respectively. It can be verified that the damage tensors N are valid for values of N_{1111} greater than zero.

Using Eqs. 87b and 91b, the following relations are obtained:

$$N_{ijkl} = 2\, M_{ijkl} - \delta_{ik}\delta_{jl} + \frac{1}{9}\,\alpha\,\left(\delta_{ik}\delta_{jl} - M_{ijkl}\right) \tag{123a}$$

$$N_{ijkl} = 2M_{ijkl} - \delta_{ik}\delta_{jl} + \frac{1}{9}\;\alpha\, M_{ijmn}(\delta_{mk}\delta_{nl} - M_{mnkl}) \tag{123b}$$

Equations 117a and 117b give the hybrid damage tensor using the damage tensor $L^{(2)}$ based on the hypothesis of elastic strain equivalence and the hypothesis of elastic energy equivalence, respectively.

Examples

This section demonstrates some examples in order to provide interpretation of the new proposed damage tensors. A two-dimensional (2-D) stress state is employed in these examples, and the damage effect tensor is represented by a 3×3 matrix, as it is a second-order tensor.

Consider the 2-D stress state; the damage effect tensor is given by the following 3×3 matrix.

$$[M] = \frac{1}{\Delta}\begin{bmatrix} 1-\varphi_{22} & 0 & \varphi_{12} \\ 0 & 1-\varphi_{11} & \varphi_{12} \\ \frac{1}{2}\varphi_{12} & \frac{1}{2}\varphi_{12} & \frac{(1-\varphi_{11})+(1-\varphi_{22})}{2} \end{bmatrix} \tag{124}$$

where Δ is given by

$$\Delta = (1-\varphi_{11})(1-\varphi_{22}) - \varphi_{12}^2 \tag{125}$$

The inverse of the damage effect tensor is given by the following 3 × 3 matrix

$$\left[M^{-1}\right] = \Delta \begin{bmatrix} 1-\varphi_{22} & 0 & \varphi_{12} \\ 0 & 1-\varphi_{11} & \varphi_{12} \\ \frac{1}{2}\varphi_{12} & \frac{1}{2}\varphi_{12} & \frac{(1-\varphi_{11})+(1-\varphi_{22})}{2} \end{bmatrix}^{-1} \tag{126}$$

Example 1 Equation 87a can be written in the following form:

$$\left[L^{(1)}\right] = [M] - [I] \tag{127}$$

where I is the identity tensor

For the undamaged material, the damage variables equal to zero, and the damage effect tensor, in this case, equals to the identity tensor.

Substituting Eq. 124 into Eq. 127 and simplifying result in

$$\left[L^{(1)}\right] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{128}$$

Since in this special case, the components of the damage tensor are identically zero, one concludes that this proposed damage tensor is a valid damage tensor.

Example 2 Equation 87b can be written in the following form:

$$\left[L^{(2)}\right] = \left(\frac{1}{3}\alpha - 1\right)[I] \tag{129}$$

Since $\alpha = 3$.

Equation 129 is given by

$$\left[L^{(2)}\right] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{130}$$

Since in this special case, the components of the damage tensor vanish, one concludes that this proposed damage tensor is a valid damage tensor.

Example 3 Equation 91a can be written in the following form:

$$\left[L^{(1)}\right] = [I] - \frac{1}{3}\alpha[M] \tag{131}$$

For the undamaged material, the damage variables equal to zero, and the damage effect tensor, in this case, equals to the identity tensor.

Substituting Eq. 124 into Eq. 131 and simplifying result in

$$\left[L^{(1)}\right] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{132}$$

where $\alpha = 3$.

Since in this special case, the components of the damage tensor vanish, one concludes that this proposed damage tensor is a valid damage tensor.

Example 4 Equation 91b can be written in the following form:

$$\left[L^{(2)}\right] = \frac{1}{3}\alpha[M] - [I] \tag{133}$$

For the undamaged material, the damage variables equal to zero, and the damage effect tensor, in this case, equals to the identity tensor.

Substituting Eq. 124 into Eq. 133 and simplifying give

$$\left[L^{(2)}\right] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{134}$$

where $\alpha = 3$.

Since in this special case, the components of the damage tensor are identically zero, it is concluded that this proposed damage tensor is a valid damage tensor.

Example 5 Equation 108a can be written in the following form:

$$[R] = \frac{1}{3}\left[[M] - \left[M^{-1}\right]\right] \tag{135}$$

For the undamaged material, the damage variables equal to zero, and the damage effect tensor and its inverse, in this case, equal to the identity tensor.

Substituting Eqs. 124 and 126 into Eq. 135 and simplifying result in

$$[R] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{136}$$

Since in this special case, the components of the damage tensor vanish, it is concluded that this proposed damage tensor is a valid damage tensor.

As for the remaining proposed damage tensors, Table 4 shows each tensor along with its validity that is determined in view of its application to the abovementioned two-dimensional damage state.

After studying Table 4, it is seen that only 7 proposed anisotropic damage tensors are invalid out of 27 proposed tensors. Furthermore, looking at the 20 valid ones, it is seen that some of them are identical. For example, tensors 14 and 18 in Table 4 are exactly identical. So eventually it is concluded that there are only seven invalid proposed anisotropic damage tensors. The invalid tensors are numbered 12, 13, 16, 24, 25, 26, and 27 according to Table 4.

Table 4 Proposed anisotropic damage tensors and their validity (Reprinted with permission from Voyiadjis and Kattan 2012)

No.	Proposed damage tensor	Validity
1	$L^{(1)}_{ijmn} = M_{ijmn} - \delta_{im}\delta_{jn}$	Valid
2	$L^{(2)}_{ijmn} = \delta_{im}\delta_{jn}\left(\frac{1}{9}\alpha - 1\right)$	Valid
3	$L^{(1)}_{ijmn} = \delta_{im}\delta_{jn} - \frac{1}{9}\alpha\, M_{ijmn}$	Valid
4	$L^{(2)}_{ijmn} = \frac{1}{9}\alpha\, M_{ijmn} - \delta_{im}\delta_{jn}$	Valid
5	$P^{(1)}_{ijmn} = \delta_{im}\delta_{jn} - M^{-1}_{ijmn}$	Valid
6	$P^{(2)}_{ijmn} = \delta_{im}\delta_{jn}\left(1 - \frac{1}{9}\beta\right)$	Valid
7	$P^{(1)}_{ijmn} = \delta_{im}\delta_{jn} - \frac{1}{9}\beta\, M^{-1}_{ijmn}$	Valid
8	$P^{(2)}_{ijmn} = \delta_{im}\delta_{jn} - \frac{1}{9}\beta\, M^{-1}_{ijmn}$	Valid
9	$R_{ijkl} = M_{ijkl} + \delta_{ik}\delta_{jl}\left(\frac{1}{9}\ \alpha - 2\right)$	Valid
10	$R_{ijkl} = M_{ijkl} - \frac{1}{9}\ \alpha\ \delta_{ik}\delta_{jl}$	Valid
11	$R_{ijkl} = 2\left(\delta_{ik}\delta_{jl} - \frac{1}{9}\ \alpha\, M_{ijkl}\ \right)$	Valid
12	$R_{ijkl} = M_{ijkl} - \beta\ \delta_{ik}\delta_{jl}$	Invalid
13	$R_{ijkl} = \alpha\, M_{ijkl} - \beta^2\ \delta_{ik}\delta_{jl}$	Invalid
14	$R_{ijkl} = \frac{1}{9}\left(M_{ijkl} - M^{-1}_{ijkl}\right)$	Valid
15	$R_{ijkl} = M_{ijmn}M_{mnkl} - M^{-1}_{ijtu}M^{-1}_{tukl}$	Valid
16	$R_{ijkl} = \frac{1}{9}\delta_{ik}\delta_{jl}[\alpha - 9\ \beta]$	Invalid
17	$R_{ijkl} = \frac{1}{9}\alpha^2\delta_{ik}\delta_{jl}\ \ - \beta\, M^{-1}_{ijkl}$	Valid
18	$R_{ijkl} = \frac{1}{9}\left(M_{ijkl} - M^{-1}_{ijkl}\right)$	Valid
19	$R_{ijkl} = M_{ijmn}M_{mnkl} - M^{-1}_{ijtu}M^{-1}_{tukl}$	Valid
20	$R_{ijkl} = [\delta_{ik}\delta_{jl} - M_{ijkl}][\beta - 9]$	Valid
21	$R_{ijkl} = \delta_{ik}\delta_{jl}[\beta - 9][\beta + 9]$	Valid
22	$R_{ijkl} = [M_{ijkl} - \delta_{ik}\delta_{jl}][9 - \beta]$	Valid
23	$R_{ijkl} = \frac{1}{9}\delta_{ik}\delta_{jl}[\alpha - 9][\alpha + 9] - M^{-1}_{ijkl}[\alpha - \beta]$	Valid
24	$N_{ijkl} = 3M_{ijkl} - M_{ijmn}M_{mnkl}$	Invalid
25	$N_{ijkl} = \ \delta_{ik}\delta_{jl} + \frac{1}{9}\alpha\, M_{ijmn}\ [\, M_{mnkl} - \delta_{mk}\delta_{nl}\ ]$	Invalid
26	$N_{ijkl} = 2\, M_{ijkl} - \delta_{ik}\delta_{jl} + \frac{1}{9}\ \alpha\ \left(\delta_{ik}\delta_{jl} - M_{ijkl}\right)$	Invalid
27	$N_{\mathrm{ijkl}} = 2M_{\mathrm{ijkl}} - \delta_{\mathrm{ik}}\delta_{\mathrm{jl}} + \frac{1}{9}\alpha M_{\mathrm{ijmn}}\left(\delta_{\mathrm{mk}}\delta_{\mathrm{nl}} - M_{\mathrm{mnkl}}\right)$	Invalid

Summary and Conclusions

In this chapter the basics of continuum damage mechanics are reviewed first. This is followed by proposing several scalar damage variables. These variables are compared, and several graphs are plotted. Then, the concept of higher-order strain energy forms is presented. In addition, hybrid damage variables are postulated that are defined both in terms of stiffness reduction and cross-sectional area reduction. The mathematical formulation is then extended to the general three-dimensional case using tensors. Finally, several examples are given.

References

E.M. Arruda, M.C. Boyce, A three-dimensional constitutive model for the large stretch behavior of rubber elastic materials. J. Mech. Phys. Solids **41**(2), 389–412 (1993)

J. Betten, Damage tensors in continuum mechanics. J. Mecanique Theorique et Appliquees **2**, 13–32 (1981). Presented at Euromech Colloqium 147 on Damage Mechanics, Paris-VI, Cachan, 22 September

J. Betten, Applications of tensor functions to the formulation of continutive equations involving damage and initial anisotropy. Eng. Fract. Mech. **25**, 573–584 (1986)

B. Budiansky, R.J. O'Connell, Elastic moduli of a cracked solid. Int. J. Solids Struct. **12**, 81–97 (1976)

A. Cauvin, R. Testa, Damage mechanics: basic variables in continuum theories. Int. J. Solids Struct. **36**, 747–761 (1999)

D.J. Celentano, P.E. Tapia, J.-L. Chaboche, experimental and numerical characterization of damage evolution in steels, in *Mecanica Computacional*, ed. by G. Buscaglia, E. Dari, O. Zamonsky, vol. XXIII (Bariloche, 2004)

J.L. Chaboche, Continuous damage mechanics – a tool to describe phenomena before crack initiation. Nucl. Eng. Des. **64**, 233–247 (1981)

C. Chow, J. Wang, An anisotropic theory of elasticity for continuum damage mechanics. Int. J. Fract. **33**, 3–16 (1987)

C.L. Chow, J. Wang, Ductile fracture characterization with an anisotropic continuum damage theory. Eng. Fract. Mech. **30**, 547–563 (1988)

P.J. Cordebois, Criteres d'Instabilite Plastique et Endommagement Ductile en Grandes Deformations, These de Doctorat, Presente a l'Universite Pierre et Marie Curie, 1983

J.P. Cordebois, F. Sidoroff, Damage induced elastic anisotropy, in *Colloque Euromech*, vol. 115 (Villard de Lans, 1979)

J.P. Cordebois, F. Sidoroff, Anisotropic damage in elasticity and plasticity. J. Mech. Theor. Appl. Numerous Special. **1**, 45–60 (1982) (in French)

J.W. Ju, Isotropic and anisotropic damage variables in continuum damage mechanics. J. Eng. Mech. ASCE **116**, 2764–2770 (1990)

L. Kachanov, On the creep fracture time. Izv Akad, Nauk USSR Otd Tech **8**, 26–31 (1958) (in Russian)

P.I. Kattan, G.Z. Voyiadjis, A plasticity-damage theory for large deformation of solids – part II: applications to finite simple shear. Int. J. Eng. Sci. **31**(1), 183–199 (1993)

P.I. Kattan, G.Z. Voyiadjis, Decomposition of damage tensor in continuum damage mechanics. J. Eng. Mech. ASCE **127**(9), 940–944 (2001a)

P.I. Kattan, G.Z. Voyiadjis, *Damage Mechanics with Finite Elements: Practical Applications with Computer Tools* (Springer, Berlin, 2001b)

D. Krajcinovic, Constitutive equations for damaging materials. J. Appl. Mech. **50**, 355–360 (1983)

D. Krajcinovic, *Damage Mechanics* (Elsevier, Amsterdam, 1996)

D. Krajcinovic, G.U. Foneska, The continuum damage theory for brittle materials. J. Appl. Mech. **48**, 809–824 (1981)
E. Krempl, On the identification problem in materials deformation modeling, in *Euromech*, vol. 147, on Damage Mechanics (Cachan, 1981)
F.A. Leckie, E.T. Onat, Tensorial nature of damage measuring internal variables, in *IUTAM Colloquim on Physical Nonlinearities in Structural Analysis* (Springer, Berlin, 1981), pp. 140–155
H. Lee, K. Peng, J. Wang, An anisotropic damage criterion for deformation instability and its application to forming limit analysis of metal plates. Eng. Fract. Mech. **21**, 1031–1054 (1985)
J. Lemaitre, Evaluation of dissipation and damage in metals subjected to dynamic loading, in *Proceedings of I.C.M.*, vol. 1, (Kyoto, 1971)
J. Lemaitre, How to use damage mechanics. Nucl. Eng. Des. **80**, 233–245 (1984)
J. Lemaitre, R. Desmorat, *Engineering Damage Mechanics* (Springer, Berlin/Heidelberg, 2005)
V. Lubarda, D. Krajcinovic, Damage tensors and the crack density distribution. Int. J. Solids Struct. **30**(20), 2859–2877 (1993)
S. Murakami, Notion of continuum damage mechanics and its application to anisotropic creep damage theory. J. Eng. Mater. Technol. Trans. ASME **105**, 99–105 (1983)
S. Murakami, Mechanical modeling of material damage. J. Appl. Mech. **55**, 280–286 (1988)
S. Murakami, N. Ohno, A continuum theory of creep and creep damage, in *Proceedings of 3M IUTAM Symposium on Creep in Structures* (Springer, Berlin, 1981), pp. 422–444
J.M. Nichols, A.B.. Abell, Implementing the degrading effective stiffness of masonry in a finite element model, in *North American Masonry Conference* (Clemson, 2003)
J.M. Nichols, Y.Z. Totoev, Experimental investigation of the damage mechanics of masonry under dynamic in-plane loads, in *North American Masonry Conference* (Austin, 1999)
E.T. Onat, Representation of mechanical behavior in the presence of internal damage. Eng. Fract. Mech. **25**, 605–614 (1986)
E.T. Onat, F.A. Leckie, Representation of mechanical behavior in the presence of changing internal structure. J. Appl. Mech. **55**, 1–10 (1988)
Y. Rabotnov, Creep rupture, in *Proceedings, Twelfth International Congress of Applied Mechanics*, ed. by M. Hetenyi, W.G. Vincenti, Stanford, 1968 (Springer, Berlin, 1969), pp. 342–349
F. Sidoroff, Description of anisotropic damage application to elasticity, in *IUTAM Colloqium on Physical Nonlinearities in Structural Analysis* (Springer, Berlin, 1981), pp. 237–244
V.V. Silberschmidt, Model of matrix cracking in carbon fiber-reinforced cross-ply laminates. Mech. Compos. Mater. Struct. **4**(1), 23–38 (1997)
G.Z. Voyiadjis, Degradation of elastic modulus in elastoplastic coupling with finite strains. Int. J. Plast. **4**, 335–353 (1988)
G.Z. Voyiadjis, P.I. Kattan, A coupled theory of damage mechanics and finite strain elasto-plasticity – part II: damage and finite strain plasticity. Int. J. Eng. Sci. **28**(6), 505–524 (1990)
G.Z. Voyiadjis, P.I. Kattan, A plasticity-damage theory for large deformation of solids – part I: theoretical formulation. Int. J. Eng. Sci. **30**(9), 1089–1108 (1992)
G.Z. Voyiadjis, P.I. Kattan, On the symmetrization of the effective stress tensor in continuum damage mechanics. J. Mech. Behav. Mater. **7**(2), 139–165 (1996)
G.Z. Voyiadjis, P.I. Kattan, *Advances in Damage Mechanics: Metals and Metal Matrix Composites* (Elsevier Science, Amsterdam, 1999)
G.Z. Voyiadjis, P.I. Kattan, *Advances in Damage Mechanics: Metals and Metal Matrix Composites with an Introduction to Fabric Tensors*, 2nd edn. (Elsevier, Amsterdam, 2006)
G.Z. Voyiadjis, P.I. Kattan, A comparative study of damage variables in continuum damage mechanics. Int. J. Damage Mech. **18**(4), 315–340 (2009)
G.Z. Voyiadjis, P. Kattan, A new class of damage variables in continuum damage mechanics. J. Eng. Mater. Technol. Trans. ASME **134**(2), 01210616-1-10 (2012)
G.Z. Voyiadjis, T. Park, Local and interfacial damage analysis of metal matrix composites using the finite element method. Eng. Fract. Mech. **56**(4), 483–511 (1997)

Undamageable Materials and Damage Processes in Series and in Parallel

2

George Z. Voyiadjis, Peter I. Kattan, and Mohammed A. Yousef

Contents

Abstract

This chapter introduces a new study in the field of continuum damage mechanics and includes two main topics. In the first topic, both the concepts of Voyiadjis-Kattan materials and undamageable materials are introduced. The Voyiadjis-Kattan material of order n is defined as a nonlinear elastic material that has a higher-order strain energy form in terms of n. The undamageable material is obtained as the limit of the Voyiadjis-Kattan material of order n as n goes to infinity. The relations of these types of materials to other nonlinear elastic materials from the literature are outlined. Also, comparisons of these types of materials with rubber materials are presented. It is hoped that these proposed

G.Z. Voyiadjis (✉) • P.I. Kattan (✉) • M.A. Yousef
Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA
e-mail: voyiadjis@eng.lsu.edu; pkattan@orange.jo; myouse2@tigers.lsu.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_2

new types of materials will open the way to new areas of research in both damage mechanics and materials science.

The second topic lays special emphasis on the order and sequence of damage processes occurring in materials. These processes can occur in series or in parallel. For example, in a metallic material, the evolution of micro-cracks and the evolution of micro-voids are considered as two separate damage processes. These two different evolutions can occur simultaneously or they can occur following each other. Another example would be matrix cracking and debonding in a composite material. These two different damage processes can occur simultaneously in parallel or sequentially in series. Three-dimensional states of deformation and damage are also presented using the concepts discussed in this work.

Introduction

Considerable interest has been recently given by many researchers to the field of continuum damage mechanics in order to improve the microstructure of the materials. This improvement results in the advancement of the overall material performance and its engineering applications. Moreover, the mechanical properties of the material are directly affected by damage which causes a significant influence on the safety aspect of engineering structures. In many engineering applications, the design considerations include the useful life of components which is a crucial item of information that is carefully considered during the design process. Based on this treatment, the mechanical and structural components that are subjected to severe service conditions, in particular, are therefore developed taking into account the usefulness of the components' life. In certain types of applications, such as aerospace and automotive industries, the prediction of the mechanical failure is mainly obtained by the design engineer. The designer in such cases is responsible for ascertaining the malfunction of the components. However, the damage may cause an unpredicted failure that incurs serious consequences in the mechanical and structural components as well as the economic advancement.

The development of cavities in the microscopic, the mesoscopic, and the macroscopic processes of fracture in materials together with the resulting deterioration in their mechanical properties is termed damage (Murakami 2012). Continuum damage mechanics, in particular, aims at the analysis of the damage development in mesoscopic and macroscopic fracture processes in the framework of continuum mechanics (Murakami 2012). In order to introduce the proposed undamageable material, it is necessary first to review some basic issues of damage mechanics. Kachanov (1958) introduced the concept of effective stress in the context of continuum damage mechanics. This concept was further elaborated on by Rabotnov (1969), Allix et al. (1989), Cauvin and Testa (1999), and Doghri (2000). The effective stress was written in terms of what was called the damage variable. This is a parameter that describes the damage state of the material ranging in values between 0 and 1.

Moreover, the original formulation of Kachanov was limited to the state of isotropic damage. Later, other researchers (Murakami 1988; Lubarda and Krajcinovic 1993; Lee et al. 1985; Voyiadjis and Kattan 1990, 1992, 1996, 1999, 2005, 2006a; Kattan and Voyiadjis 1990, 1993a, b, 2001b) extended Kachanov's work to the general anisotropic damage state. For this purpose, they generalized the concept of effective stress to three-dimensional states of deformation and damage. In this generalization process, they introduced the damage tensor – obviously a generalization of the damage variable. It has been argued (Lemaitre 1984) that the assumption of isotropic damage is sufficient to give good predictions of the load-carrying capacity and the number of cycles or the time to local failure in structural components. However, the development of anisotropic damage has been confirmed experimentally (Chow and Wang 1987; Lee et al. 1985) even if the virgin material is isotropic. For the case of isotropic damage mechanics, the damage variable is scalar and the evolution equations are easy to handle (Lee et al. 1985; Voyiadjis and Kattan 2006a; Kattan and Voyiadjis 2001a, b).

Several expressions of the damage tensor were introduced but mainly consisting of second-rank and fourth-rank tensors (Krajcinovic 1996; Ladeveze et al. 1982; Luccioni and Oller 2003). Silberschmidt (1997) presented an example of matrix cracking in fiber-reinforced composite materials. Recently Voyiadjis and Kattan (2012a, b) introduced the concept of undamageable materials. They also established a link between damage mechanics and the theory of fabric tensors (Voyiadjis and Kattan 2006b, c, 2007a, b). In addition, Budiansky and O'Connell (1976) solved the problem of elastic cracked solids. Hansen and Schreyer (1994) provided a solid thermodynamic basis for damage mechanics. In addition, in continuum damage mechanics, usually a phenomenological approach is adopted. In this approach, the most important concept is that of the representative volume element (RVE). The discontinuous and discrete elements of damage are not considered within the RVE; rather their combined effects are lumped together through the use of a macroscopic internal variable. In this way, the formulation may be derived consistently using sound mechanical and thermodynamic principles (Doghri 2000; Hansen and Schreyer 1994; Luccioni and Oller 2003).

This chapter introduces a new study in the field of continuum damage mechanics and includes two main topics. In the first topic, both the concepts of Voyiadjis-Kattan materials and undamageable materials are introduced. The Voyiadjis-Kattan material of order n is defined as a nonlinear elastic material that has a higher-order strain energy form in terms of n. The undamageable material is obtained as the limit of the Voyiadjis-Kattan material of order n as n goes to infinity. The relations of these types of materials to other nonlinear elastic materials from the literature are outlined. Also, comparisons of these types of materials with rubber materials are presented. Finally, a proof is given to show that the value of the damage variable remains zero in an undamageable material throughout the deformation process. It is hoped that these proposed new types of materials will open the way to new areas of research in both damage mechanics and materials science. Then damage mechanics is finally utilized to show the full details of the proposed undamageable materials.

In the second topic discussed in this chapter, a conceptual framework for damage processes in materials is presented here. In this framework, the mechanics of damage processes in materials is investigated. These processes are categorized into either damage processes described in terms of stiffness degradation or damage processes described in terms of cross-sectional area reduction. Furthermore, the damage processes are visualized to occur as sequences either in series or in parallel. Schematic diagrams are used to illustrate these processes in a similar way to what is done with elastic springs and electric circuits. Different kinds of combinations and interactions of the damage processes are illustrated with various examples. This work is currently limited to linear elastic materials. It is hoped that this work will lay the groundwork to open new areas of research in damage mechanics.

Theory of Elastic Undamageable Materials

Higher-Order Strain Energy Forms

In this section, the nature of the proposed higher-order strain energy forms is investigated and provides their exact relationships to the elastic constitutive equations of the material. It will be seen that each higher-order strain energy form will correspond to an exact nonlinear elastic stress-strain relation. These specific stress-strain relations will be derived in this section. These new proposed types of materials are called here the Voyiadjis-Kattan materials (Voyiadjis and Kattan 2012c).

First, the general formulation is considered of how to derive a specific stress-strain relation if the corresponding strain energy form is known. For example, in the linear case, the linear stress-strain relation $\sigma = E\,\varepsilon$ corresponds to the usual strain energy form $U = \frac{1}{2}\sigma\varepsilon$. The types of stress-strain relations are now investigated that may be derived from the proposed higher-order strain energy forms $\frac{1}{2}\sigma\varepsilon^2, \frac{1}{2}\sigma\varepsilon^3$, and so on with higher powers of ε.

Use is made of the terminology of Voyiadjis-Kattan material of order n to designate any nonlinear elastic material that has a higher-order strain energy of the form $\frac{1}{2}\sigma\varepsilon^n$.

The following general stress-strain relation is assumed:

$\sigma = E\,f(\varepsilon)$, where $f(\varepsilon)$ is to be determined for each specific form of higher-order strain energy. The higher-order strain energy U (some form of hyper-strain energy) is defined as the area under the stress-strain curve. This is given by the following relation:

$$U = \int \sigma\, d\varepsilon \tag{1}$$

The general case using the higher-order strain energy form $U = \frac{1}{2}\sigma\varepsilon^n$ will now be illustrated. Substituting this expression for U into Eq. 1, one obtains

$$\frac{1}{2}\sigma\varepsilon^n = \int \sigma\, d\varepsilon \tag{2}$$

Substituting the general stress-strain relation $\sigma = E\, f(\varepsilon)$ into Eq. 2, the following relation is obtained:

$$\frac{1}{2}Ef(\varepsilon)\ \varepsilon^n = E\int f(\varepsilon)d\varepsilon \tag{3}$$

or

$$f(\varepsilon)\ \varepsilon^n = 2\int f(\varepsilon)d\varepsilon \tag{4}$$

Taking the derivative of both sides of the above equation with respect to ε results in

$$f'(\varepsilon)\,\varepsilon^n + nf(\varepsilon)\,\varepsilon^{n-1} = 2f(\varepsilon) \tag{5}$$

The above expression is the governing differential equation for the Voyiadjis-Kattan material of order n. The solution of the above differential equation is easily obtained as follows using any symbolic algebra system like the MATLAB Symbolic Math Toolbox:

$$f(\varepsilon) = \frac{1}{\varepsilon^n} e^{-2/\left[(n-1)\varepsilon^{(n-1)}\right]} \tag{6}$$

Substituting the above expression into the general constitutive relation $\sigma = E\, f(\varepsilon)$, one obtains

$$\sigma = E\ \frac{1}{\varepsilon^n} e^{-2/\left[(n-1)\varepsilon^{(n-1)}\right]} \tag{7}$$

The above solution is obtained using the initial condition that the stress is zero when the strain is zero. The above equation is a nonlinear stress-strain relationship that governs the behavior of the Voyiadjis-Kattan material of order n.

Looking at the above stress-strain relation, it may appear that the stress approaches infinity at the initial condition when the strain is zero but this is not the case. It is seen that the limit of the above expression for the stress approaches zero as the strain approaches zero. Thus, the initial conditions of zero strain and zero stress are satisfied.

The above procedure can be repeated for the other proposed higher-order strain energy forms to obtain their corresponding nonlinear stress-strain relations. These results are summarized in Table 1.

In Table 2, a comparison is shown between the proposed Voyiadjis-Kattan material of order n and other nonlinear elastic materials from the literature (Bower 2009). For example, it is noted that the Voyiadjis-Kattan material of

Table 1 The proposed higher-order strain energy forms and their corresponding stress-strain relations (constitutive equations for Voyiadjis-Kattan material of order n) (Reprinted with permission from Voyiadjis and Kattan (2012b))

Proposed higher-order strain energy form	Corresponding stress-strain relation	Type of new proposed material
$U = \frac{1}{2}\sigma\varepsilon$	$\sigma = E\ \varepsilon$	Voyiadjis-Kattan material of order 1 (linear elastic)
$U = \frac{1}{2}\sigma\varepsilon^2$	$\sigma = E\ \frac{1}{\varepsilon^2}e^{-2/\varepsilon}$	Voyiadjis-Kattan material of order 2
$U = \frac{1}{2}\sigma\varepsilon^3$	$\sigma = E\ \frac{1}{\varepsilon^3}e^{-1/\varepsilon^2}$	Voyiadjis-Kattan material of order 3
$U = \frac{1}{2}\sigma\varepsilon^n$, $n = 1, 2, 3,....$	$\sigma = E\ \frac{1}{\varepsilon^n}e^{-2/\left[(n-1)\varepsilon^{(n-1)}\right]}$	Voyiadjis-Kattan material of order n

Table 2 Comparison between the Voyiadjis-Kattan material of order n and other nonlinear elastic materials from the literature (Reprinted with permission from Voyiadjis and Kattan (2012b))

Value of n	Proposed material	Comparable material (from the literature)
1	Voyiadjis-Kattan material of order 1	Linear elastic material
2	Voyiadjis-Kattan material of order 2	Mooney-Rivlin material
3	Voyiadjis-Kattan material of order 3	Neo-Hookean material
...	...	...
...	...	...
n (finite)	Voyiadjis-Kattan material of order n	Ogden material
∞	Undamageable material	–

order 2 is comparable to the Mooney-Rivlin material in the sense that the strain energy of both materials includes squared powers of strain. It is noted also that the Voyiadjis-Kattan material of order 3 is comparable to the Neo-Hookean material in the sense that the strain energy of both materials includes cubed powers of strain.

In the last row of Table 2, the concept of undamageable material is introduced. This new proposed material is defined as the limit of the Voyiadjis-Kattan material of order n as n goes to infinity. More details about this material are given in section "The Damage Variable."

As the limiting case when $n \to \infty$ cannot be reached physically, then the concept of an undamageable material is approached by utilizing a very high value for exponent n. The stress-strain curves based on Table 1 are shown in a graph in Fig. 1 for various values of n.

Comparison with Rubber Materials

In this section, a comparison is obtained between the proposed Voyiadjis-Kattan materials, undamageable materials, and rubber materials.

The stress-strain curves appearing in Fig. 1 are similar to those of rubber materials. Arruda and Boyce (1993) conducted extensive investigation of the

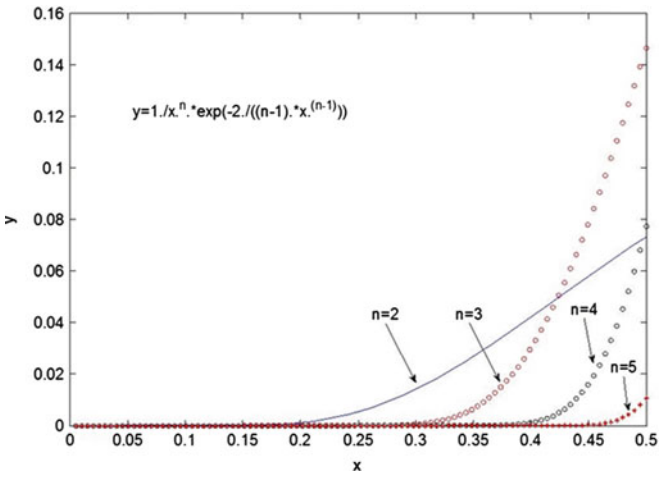


Fig. 1 Valid stress-strain curves for various values of n (Reprinted with permission from Voyiadjis and Kattan (2012b))

constitutive equations of elastic rubber materials. Figures 2, 3, and 4 show the stress-strain curves obtained for different types of rubber materials based on the work of Arruda and Boyce (1993).

It is clear from Fig. 1 that the derived stress-strain relationship for undamageable materials is distinctly different from that for ordinary materials. Clearly, there is no degradation of the elastic modulus. Also, it is seen that the values of the stress remain identically zero until the strain reaches a certain critical value. As the hypothetical undamageable material (for higher values of n in Fig. 1) is approached, it is clearly seen that the stress-strain curve remains almost horizontal indicating absolutely no stress or damage in the material. The value of the stress becomes nonzero only after a considerable amount of strain is accumulated. At these high values of strain, it is also seen that the modulus of elasticity actually strengthens instead of degrading like in ordinary materials. This is the essence of the proposed undamageable materials.

The Damage Variable

In this section, more details are provided for the proposed undamageable materials. In particular, a proof is presented using damage mechanics to show that the value of the damage variable remains zero throughout the deformation process in these types of materials. A scalar damage variable which is defined in terms of reduction in elastic stiffness or modulus is utilized for this purpose.

A linear elastic material with modulus of elasticity E is assumed in this work. Also a fictitious undamaged state of the material is assumed in which the effective elastic modulus is denoted by $\overline{E}$. In order to compute the effective elastic modulus $\overline{E}$

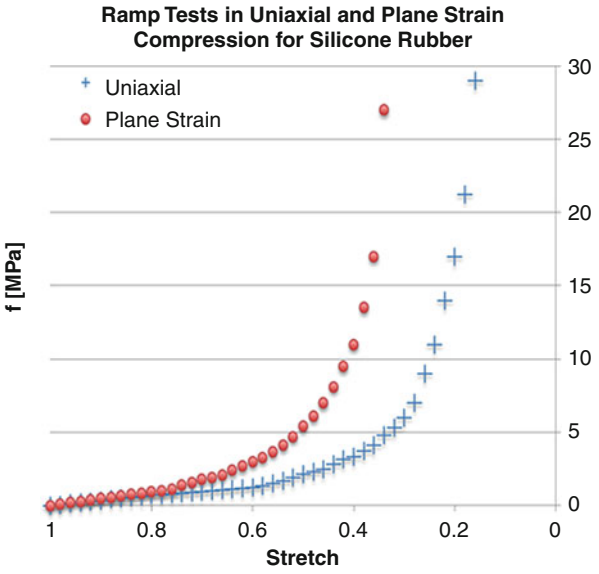


Fig. 2 Stress-strain curve for silicon rubber (Reprinted with permission from Voyiadjis and Kattan (2012b))

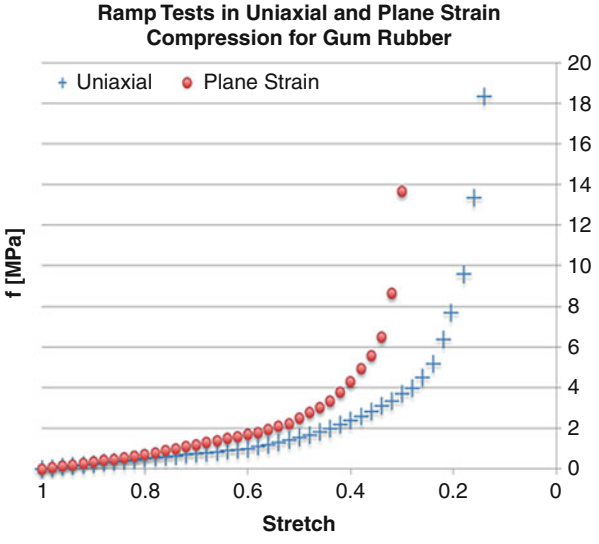


Fig. 3 Stress-strain curve for gum rubber (Reprinted with permission from Voyiadjis and Kattan (2012b))

in this case, use is made of the hypothesis of elastic energy equivalence where the elastic strain energy is assumed to be equal in both configurations (Sidoroff 1981). This will be illustrated below.

The scalar damage variable ℓ is defined in terms of the reduction in the elastic modulus as follows:

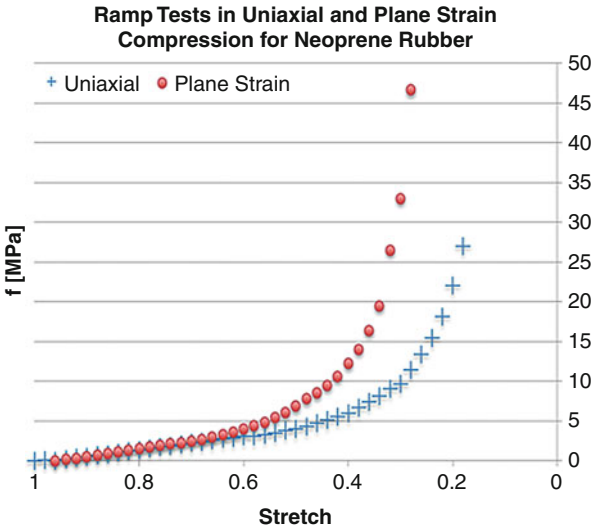


Fig. 4 Stress-strain curve for neoprene rubber (Reprinted with permission from Voyiadjis and Kattan (2012b))

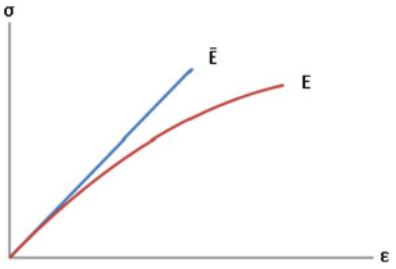


Fig. 5 Damaged and effective moduli of elasticity (Reprinted with permission from Voyiadjis and Kattan (2012b))

$$\ell = \frac{\overline{E} - E}{E} \tag{8}$$

where E is the elastic modulus in the damaged state while $\overline{E}$ is the effective elastic modulus (in the fictitious state) with $\overline{E} > E$ (see Fig. 5). This damage variable was used recently by Celentano et al. (2004), Nichols and Abell (2003), and Nichols and Totoev (1999). Voyiadjis (1988) used a similar relation but in the context of elastoplastic deformation. The reader is also referred to Voyiadjis and Kattan (2009) for more details. The definition of the alternative damage variable of Eq. 8 may be rewritten in the following more appropriate form:

$$\overline{E} = E(1 + \ell) \tag{9}$$

It is clear from the definition in Eq. 8 that $\ell = 0$ when the body is undamaged, i.e., when $\overline{E} = E$.

Using the hypothesis of elastic energy equivalence, the complementary elastic strain energy ($\frac{\sigma^2}{2E}$) is assumed to be equal in both configurations, i.e.,

$$\frac{\sigma^2}{2E} = \frac{\overline{\sigma}^2}{2\overline{E}} \tag{10}$$

Next, the basic damage variable ℓ is explored in more detail and the concept of an undamageable material is introduced using this variable.

Using the hypothesis of elastic energy equivalence and Eq. 10, one obtains $\overline{\sigma} = \sqrt{\frac{\overline{E}}{E}}\sigma$. In this case, it can be easily shown that the damage variable $\ell = \frac{\overline{E}-E}{E}$ will yield $\overline{\sigma} = \sigma\sqrt{1+\ell}$.

A new hypothesis of higher-order energy equivalence is postulated in the form

$$\frac{1}{2}\sigma\varepsilon^2 = \frac{1}{2}\overline{\sigma}\overline{\varepsilon}^2 \tag{11a}$$

to obtain the following relation:

$$\overline{\sigma} = \sqrt[3]{\frac{\overline{E}}{E}} \tag{11b}$$

In this case, it is easily shown that using $\ell = \frac{\overline{E}-E}{E}$ will yield the relation

$$\overline{\sigma} = \sigma\sqrt[3]{1+\ell} \tag{11c}$$

Finally, a new hypothesis is postulated of the generalized equivalence of higher-order energy in terms of the n-powers of σ and ε, in order to obtain the relation

$$\overline{\sigma} = \sqrt[n]{\frac{\overline{E}}{E}} \tag{12a}$$

In this case, it is easily shown that using $\ell = \frac{\overline{E}-E}{E}$ will yield the general relation

$$\overline{\sigma} = \sigma\sqrt[n]{1+\ell} \tag{12b}$$

One now plots several curves on the same graph to show the relations between the ratio of the stresses $\frac{\overline{\sigma}}{\sigma}$ and ℓ using Eqs. 11b and 12b (see Fig. 6). It is clear that for the limiting case when $n \to \infty$, the curve has a constant value at $n = 1$. Note that the lower curves appearing in Fig. 3 are for larger values of n. What does this limiting case signify?

The above results are now explained using the formulae derived for ℓ. The formula $\overline{\sigma} = \sigma\sqrt[n]{1+\ell}$ of Eq. 12b is used here to elaborate on the results. The

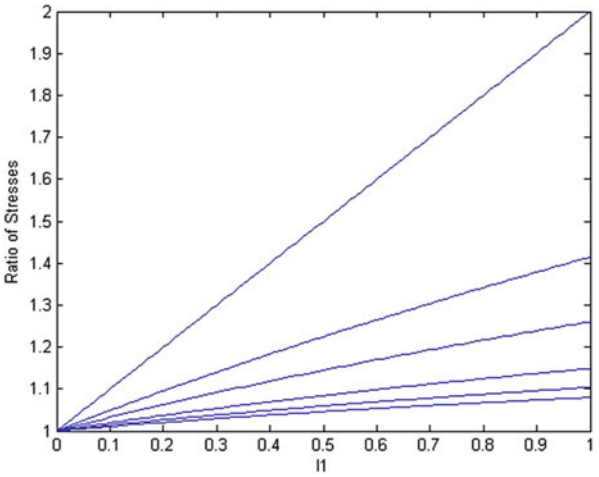


Fig. 6 Relation between ℓ_1 and the ratio of the stresses (Reprinted with permission from Voyiadjis and Kattan (2012b))

case when $n \to \infty$ is now investigated. In this case, the following relation is obtained:

$$\overline{\sigma} = \sigma \sqrt[n]{1+\ell} = \sigma (1+\ell)^{\frac{1}{n}} = \sigma (1+\ell)^{\frac{1}{\infty}} = \sigma (1+\ell)^{0} = \sigma \cdot 1 = \sigma \qquad (13)$$

Therefore, one obtains $\overline{\sigma} = \sigma$ irrespective of the value of the damage variable ℓ. This means that in this limiting case, the material remains totally undamaged irrespective of the value of the damage variable. Of course, this is a hypothetical case as it cannot be reached physically. However, it gives rise to the following new issue. Does a material exists or can be manufactured which remains totally undamaged during the deformation process no matter what the load may be? That will be a seminal achievement of damage mechanics and materials science in general – to design a new type of material that cannot be damaged at all. The above equation will provide some guidelines in this respect.

The following is a summary of the main concepts and results in this work:

1. The Voyiadjis-Kattan material of order n is a nonlinear elastic material which has strain energy of the form $\frac{1}{2}\sigma\varepsilon^n$, where n is greater than 1.
2. The undamageable material is the limit of the Voyiadjis-Kattan material of order n as n goes to infinity.
3. The linear elastic material is a type of Voyiadjis-Kattan material of order 1.
4. In an undamageable material, the value of the stress will remain equal to zero throughout the deformation process. Also, the damage variable will be equal to zero throughout.
5. The undamageable material has zero strain energy.
6. The undamageable material has nonzero strain values. Thus, the undamageable material is a type of deformable body, not a rigid body.

7. The Voyiadjis-Kattan material of order n has nonzero stress values. The range of the nonzero stress values changes depending on the value of n. The higher the value of n, the narrower the range of nonzero stress values.

Thermodynamic Formulation with Internal Variables

In this section, use is made of the thermodynamic theory of Rice (1971). Consider a material sample of size V which is measured in an unloaded reference state and at a reference temperature T_0. Let σ (or ε), T, and ς be the thermodynamic state variables of constrained equilibrium states of the material sample, where T is the temperature and ς is a set of internal state variables that include the damage variable.

Let η be the specific free energy and ψ its Legendre transform where $\eta = \eta(\varepsilon, T, \varsigma)$ and $\psi = \psi(\sigma, T, f) = \varepsilon \frac{\partial \eta}{\partial \varepsilon} - \eta$. Let θ be the specific energy and f be a set of conjugate thermodynamic forces to ς. The following relations are obtained:

$$\sigma = \sigma(\varepsilon, T, \varsigma) = \frac{\partial \eta(\varepsilon, T, \varsigma)}{\partial \varepsilon} \tag{14}$$

$$\varepsilon = \varepsilon(\sigma, T, \varsigma) = \frac{\partial \psi(\sigma, T, \varsigma)}{\partial \sigma} \tag{15}$$

$$\theta = \theta(\sigma, T, \varsigma) = \frac{\partial \psi(\sigma, T, \varsigma)}{\partial T} \tag{16}$$

The thermodynamic conjugate forces f are then given by

$$f = V \frac{\partial \psi}{\partial \varsigma} = -V \frac{\partial \eta}{\partial \varsigma} \tag{17}$$

where $f = f(\sigma, T, \varsigma)$ or $f = f(\varepsilon, T, f)$. From Eq. 15, the following relation for the increment of strains is derived:

$$d\varepsilon = \frac{\partial^2 \psi}{\partial \sigma^2} d\sigma + \frac{\partial^2 \psi}{\partial \sigma \partial T} dT + \frac{\partial^2 \psi}{\partial \sigma \partial \varsigma} d\varsigma \tag{18}$$

Finally, the flow potential $Q = Q(f, T, \varsigma)$ is given by

$$Q = \frac{1}{V} \int_0^f \dot{\varsigma}\, df \tag{19}$$

where $dQ = \frac{1}{V} \dot{\varsigma}\, df$.

Damage Processes in Series and in Parallel

Review of Existing Damage Variables

In this section two major scalar damage variables are discussed that are used by researchers at the present. The first scalar damage variable is defined in terms of cross-sectional area reduction, while the second scalar damage variable is defined in terms of the reduction in the elastic modulus or elastic stiffness (Voyiadjis and Kattan 2009).

Consider a body (in the form of a cylinder) in the initial undeformed and undamaged configuration. Consider also the configuration of the body that is both deformed and damaged after a set of external agencies act on it (see Fig. 7). Next, consider a fictitious configuration of the body obtained from the damaged configuration by removing all the damage that the body has undergone, i.e., this is the state of the body after it had only deformed without damage (see Fig. 7). Therefore, in defining a damage variable ϕ, its value must vanish in the fictitious configuration.

The first damage variable ϕ is usually defined as follows:

$$\phi = \frac{A - \overline{A}}{A} \tag{20}$$

where A is the cross-sectional area in the damaged configuration, while $\overline{A}$ is the cross-sectional area in the fictitious configuration with $A > \overline{A}$. It is clear that when a body is undamaged, i.e., when $A = \overline{A}$, then $\phi = 0$.

The stress in the fictitious configuration is called the effective stress and is denoted by $\overline{\sigma}$. The value of the effective stress $\overline{\sigma}$ may be obtained using the relation $\overline{\sigma}\overline{A} = \sigma A$ where σ is the stress in the damaged configuration. Therefore, using this relation along with the definition in Eq. 20, one obtains

$$\overline{\sigma} = \frac{\sigma}{1 - \phi} \tag{21}$$

It should be mentioned that the equilibrium condition in the paragraph above reflects a mean-field type of assumption on the stress redistribution (uniform over

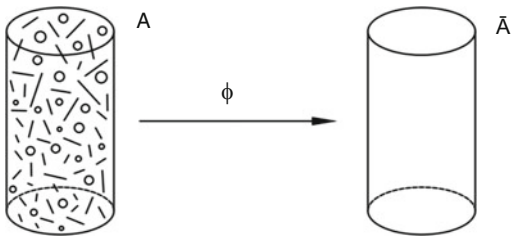


Fig. 7 Damaged and effective undamaged configurations (Reprinted with permission from Voyiadjis et al. (2012))

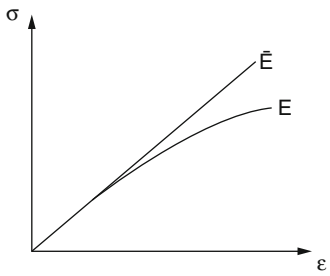


Fig. 8 Stress-strain curves of damage and undamaged material (Reprinted with permission from Voyiadjis and Kattan (2012d))

the resistive section) and therefore appears to be appropriate only in the dilute damage regime, away from the stress-strain peak where cooperate effects dominate and damage localization takes place.

The second scalar damage variable ℓ may be defined in terms of the reduction in the elastic modulus as follows:

$$\ell = \frac{\overline{E} - E}{E} \tag{22}$$

where E is the elastic modulus in the damaged state while $\overline{E}$ is the effective elastic modulus (in the fictitious state) with $\overline{E} > E$ (see Fig. 8). This damage variable was used recently by Celentano et al. (2004), Nichols and Abell (2003), and Nichols and Totoev (1999). It should also be mentioned that Voyiadjis (1988) used a similar relation but in the context of elastoplastic deformation. See also Voyiadjis and Kattan (2009, 2012d).

The definition of the alternative damage variable of Eq. 22 may be rewritten in the following more appropriate form:

$$\overline{E} = E(1 + \ell) \tag{23}$$

It is clear from the definition in Eq. 22 that $\ell = 0$ when the body is undamaged, i.e., when $\overline{E} = E$.

Damage Processes Described in Terms of Stiffness Degradation

The damage processes are studied first that are described in terms of elastic stiffness degradation as outlined in Eqs. 22 and 23. Multiple damage processes are assumed that are operating within the damaged material. All the processes are also assumed to be elastic and undergo elastic damage. The number of these processes is generally assumed to be n.

Consider first a sequence of these damage processes that are described by a sequence of scalar damage variables $\ell_1, \ell_2, \ell_3, \ldots \ell_n$ where damage is characterized by the degradation of the elastic stiffness. Two different cases are considered in which this sequence of n damage processes can occur.

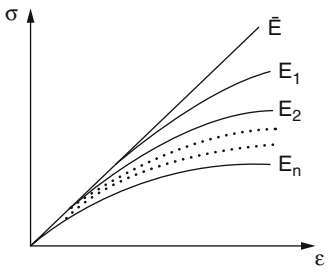


Fig. 9 Damage processes in series that are described in terms of stiffness degradation (Reprinted with permission from Voyiadjis and Kattan (2012d))

The damage processes can occur in a sequence one following the other (which is termed damage processes in series) or they can occur simultaneously, at the same time (which are termed damage processes in parallel). This conceptual description is carried out in this work analogously to what is done in the theory of elastic springs and electric circuits. Combinations of damage processes are studied in series and in parallel later in this section.

Consider first the n damage processes occurring in series. Check Fig. 9 where this process is illustrated using the stress-strain curves. Let ℓ_1 be the first damage variable of the first damage process in which the elastic stiffness degrades from $\overline{E}$ to E_1, and let ℓ_2 be the damage variable of the next damage process in which the elastic stiffness is reduced from E_1 to E_2. It is assumed to continue in this fashion with a sequence of damage processes occurring in series (i.e., one following the other one) until one gets to the final damage variable ℓ_n of the final damage process in which the elastic stiffness is reduced from E_{n-1} to E_n. This conceptual framework is illustrated in Fig. 9.

In this case, the following sequence of equations based on Eq. 22 are obtained:

$$\ell_1 = \frac{\overline{E} - E_1}{E_1} \tag{24a}$$

$$\ell_2 = \frac{E_1 - E_2}{E_2} \tag{24b}$$

$$\ell_3 = \frac{E_2 - E_3}{E_3} \tag{24c}$$

$$\ell_n = \frac{E_{n-1} - E_n}{E_n} \tag{24d}$$

However, if one looks at the total (or cumulative) damage variable ℓ, then one will assume it to represent the reduction of stiffness from the effective undamaged stiffness $\overline{E}$ to the final damaged stiffness E_n. Therefore, the total damage variable is given by the following expression:

$$\ell = \frac{\overline{E} - E_n}{E_n} \tag{25}$$

Next, one considers the following mathematical identity written as follows:

$$\frac{\overline{E}}{E_n} = \frac{\overline{E}}{E_1}\frac{E_1}{E_2}\frac{E_2}{E_3}\cdots\frac{E_{n-1}}{E_n} \tag{26}$$

Substituting Eqs. 24 and 25 into Eq. 26, and simplifying the result, the following relation is obtained between the total damage variable and the individual damage variables for the damage processes in the sequence:

$$1 + \ell = (1 + \ell_1)(1 + \ell_2)(1 + \ell_3)\ldots(1 + \ell_n) \tag{27}$$

In the special case of a sequence of two damage processes ℓ_1 and ℓ_2, Eq. 27 can be reduced to the following simple form:

$$\ell = \ell_1 + \ell_2 + \ell_1\ell_2 \tag{28}$$

Consider next the n damage processes occurring in parallel. Check Fig. 10 where this process is illustrated using the stress-strain curves. Let ℓ_1 be the first damage variable of the first damage process in which the elastic stiffness degrades from $\overline{E}$ to E_1, and let ℓ_2 be the damage variable of the second damage process in which the elastic stiffness is reduced from $\overline{E}$ to E_2. One now continues in this fashion with a sequence of damage processes occurring in parallel (i.e., at the same time) until the final damage variable ℓ_n of the final damage process in which the elastic stiffness is reduced from $\overline{E}$ to E_n. This conceptual framework is illustrated schematically in Fig. 10.

In this case, we write the following sequence of equations based on Eq. 22:

$$\ell_1 = \frac{\overline{E} - E_1}{E_1} \tag{29a}$$

$$\ell_2 = \frac{\overline{E} - E_2}{E_2} \tag{29b}$$

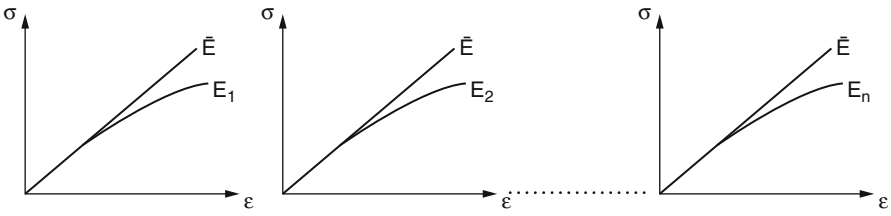


Fig. 10 Stress-strain curves of damage processes in parallel that are described in terms of stiffness degradation (Reprinted with permission from Voyiadjis and Kattan (2012d))

$$\ell_3 = \frac{\overline{E} - E_3}{E_3} \tag{29c}$$

$$\ell_n = \frac{\overline{E} - E_n}{E_n} \tag{29d}$$

However, if one observes the total (or cumulative) damage variable ℓ as shown in Fig. 8, then one will assume it to represent the reduction of stiffness from the effective undamaged stiffness $\overline{E}$ to the final damaged stiffness E. Therefore, the total damage variable is given by the following expression:

$$\ell = \frac{\overline{E} - E}{E} \tag{30}$$

In order to relate the total elastic stiffness to the individual elastic stiffnesses E_1, E_2, ... E_n, some form of elaborate homogenization process should be used. However, for the sake of simplicity and to illustrate the concepts introduced, the following simple relation is chosen:

$$E = c_1 E_1 + c_2 E_2 + \ldots + c_n E_n \tag{31}$$

where $c_1, c_2, \ldots c_n$ are constants to be determined. For example, using $c_1 = c_2 = \ldots = c_n = \frac{1}{n}$, then the total elastic stiffness is taken to be the average value of the individual elastic stiffnesses.

Substituting Eqs. 29 and 30 into Eq. 31, and simplifying the results, the following general relation is obtained between the total damage variable and the individual damage variables for the individual parallel damage processes:

$$\frac{1}{1+\ell} = \frac{c_1}{1+\ell_1} + \frac{c_2}{1+\ell_2} + \ldots + \frac{c_n}{1+\ell_n} \tag{32}$$

Using the special case of two damage processes ℓ_1 and ℓ_2 occurring in parallel, then Eq. 32 is reduced to the following simple form:

$$\ell = \frac{(1+\ell_1)(1+\ell_2)}{c_1(1+\ell_2) + c_2(1+\ell_1)} - 1 \tag{33}$$

It should be noted that Shen et al. (2011) used a sequence of two damage processes that they termed sequential homogenization in a way similar to the framework presented here. However, they only considered the two damage processes to occur in series but have not considered them to occur in parallel.

In order to illustrate various combinations of these damage processes occurring in series and in parallel, one needs to use schematic diagrams. In this case, the elastic spring is used to illustrate an individual damage process that is described in terms of stiffness degradation as shown in Fig. 11.

Fig. 11 Schematic illustration of a damage process in terms of stiffness reduction (Reprinted with permission from Voyiadjis and Kattan (2012d))

Fig. 12 Damage processes in series described in terms of stiffness degradation (Reprinted with permission from Voyiadjis and Kattan (2012d))

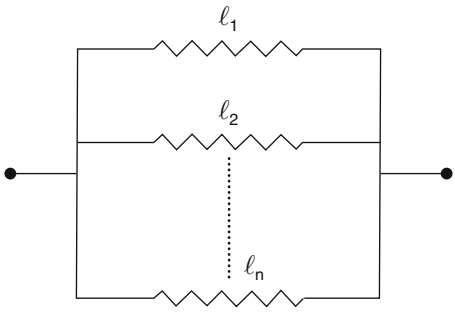


Fig. 13 Damage processes in parallel described in terms of stiffness degradation (Reprinted with permission from Voyiadjis and Kattan (2012d))

Based on the illustration of Fig. 11, these types of damage processes are shown in series in Fig. 12 (corresponding to the stress-strain curves of Fig. 9) and these types of damage processes in parallel in Fig. 13 (corresponding to the stress-strain curves of Fig. 10).

One now considers several combinations of the above damage processes. Consider first a sequence of n damage processes in series (characterized by the damage variables $\ell_1, \ell_2, \ldots \ell_n$) followed by m damage processes in parallel as shown in Fig. 14 (characterized by the damage variables $\ell'_1, \ell'_2, \ldots \ell'_m$).

This case can be investigated by considering that the first sequence of damage processes in series can be characterized by a single total damage variable called ℓ_A, while the second sequence of damage processes in parallel can be characterized by a single total damage variable called ℓ_B as illustrated in Fig. 15.

The damage processes in Fig. 15 are clearly occurring in series. Therefore, the total damage variable ℓ is given by (based on Eq. 27)

$$1 + \ell = (1 + \ell_A)(1 + \ell_B) \tag{34}$$

where ℓ_A and ℓ_B are obtained from Eqs. 27 and 32, respectively. Therefore, expanding Eq. 34 one obtains the following general expression for the total damage variable to describe the damage processes in Fig. 14:

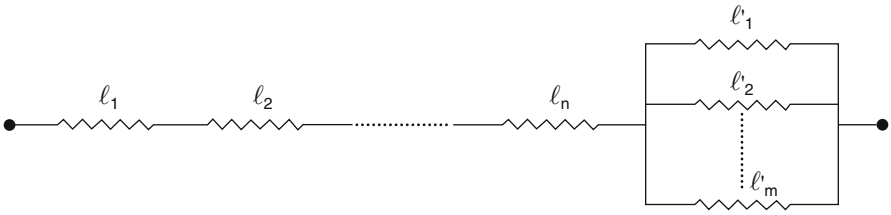


Fig. 14 Damage processes in series followed by damage processes in parallel (Reprinted with permission from Voyiadjis and Kattan (2012d))

Fig. 15 Characterization of the two sequences of damage processes of Fig. 14 (Reprinted with permission from Voyiadjis and Kattan (2012d))

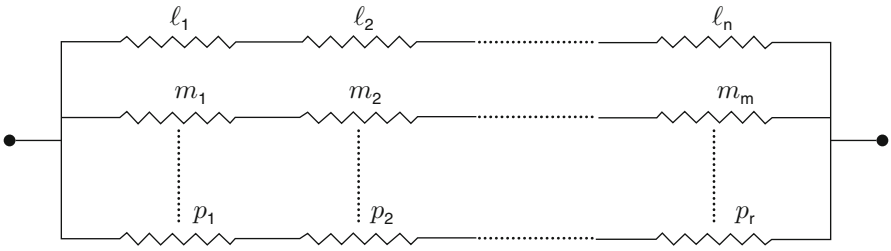


Fig. 16 Several damage processes in series occurring in parallel (Reprinted with permission from Voyiadjis and Kattan (2012d))

$$1+\ell=\frac{(1+\ell_1)(1+\ell_2)\ldots(1+\ell_n)}{\dfrac{c_1}{1+\ell_1'}+\dfrac{c_2}{1+\ell_2'}+\ldots+\dfrac{c_m}{1+\ell_m'}} \tag{35}$$

Next, another combination of damage processes in series and in parallel is considered. The damage processes illustrated schematically in Fig. 16 are considered where several sequences of damage processes in series occur collectively in parallel (characterized by the sequences of damage variables $\ell_1, \ell_2, \ldots \ell_n$ and $m_1, m_2, \ldots m_m, p_1, p_2, \ldots p_r$).

In this case, let each sequence of damage processes in series be characterized by a single total damage variable $\ell_A, \ell_B, \ldots \ell_Z$ as shown in Fig. 17.

In this case, the single total damage variables $\ell_A, \ell_B, \ldots \ell_Z$ can be obtained from Eq. 27 as follows:

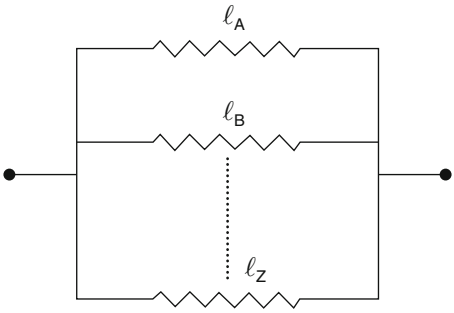


Fig. 17 Characterization of the damage processes of Fig. 16 (Reprinted with permission from Voyiadjis and Kattan (2012d))

$$1 + \ell_A = (1 + \ell_1)(1 + \ell_2) \ldots (1 + \ell_n) \tag{36a}$$

$$1 + \ell_B = (1 + m_1)(1 + m_2) \ldots (1 + m_m) \tag{36b}$$

$$1 + \ell_Z = (1 + p_1)(1 + p_2) \ldots (1 + p_r) \tag{36c}$$

Finally, the expression for the total damage variable can be obtained by using Eqs. 32 and 36 as follows to describe the damage processes in Fig. 16:

$$\frac{1}{1+\ell} = \frac{c_A}{(1+\ell_1)(1+\ell_2)\ldots(1+\ell_n)} + \frac{c_B}{(1+m_1)(1+m_2)\ldots(1+m_m)} + \ldots + \frac{c_Z}{(1+p_1)(1+p_2)\ldots(1+p_r)} \tag{37}$$

Damage Processes Described in Terms of Cross-Sectional Area Reduction

In this section, the damage processes are studied that are described in terms of reduction of the cross-sectional area as outlined in Eq. 20 and Fig. 7. Multiple damage processes are assumed that operate within the damaged material. In addition it is assumed that all the processes are elastic and undergo elastic damage. The number of these processes is generally n processes.

Consider first a sequence of these damage processes that are described by a sequence of scalar damage variables $\phi_1, \phi_2, \ldots \phi_n$ where damage is characterized by the reduction of the cross-sectional area. Two different cases are considered in which this sequence of n damage processes can occur. The damage processes can occur in a sequence one following the other (which is termed damage processes in series) or they can occur simultaneously, at the same time (which is termed damage processes in parallel). This conceptual description is carried out in this work analogously to what is done in the theory of elastic springs and electric circuits. In addition combinations of damage processes are studied that are in series and in parallel.

Consider first the n damage processes occurring in series. Check Fig. 18 where this process is illustrated using cross-sectional area. Let ϕ_1 be the first damage

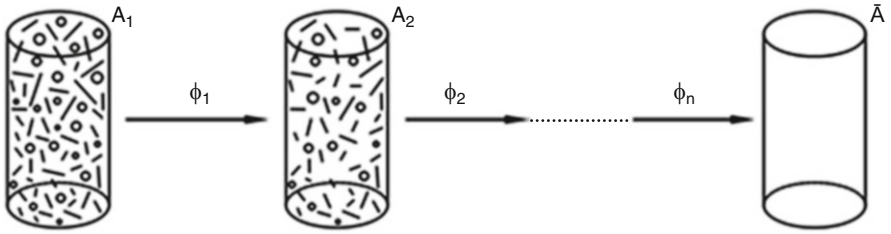


Fig. 18 Damage processes in series that are described in terms of reduction in cross-sectional area (Reprinted with permission from Voyiadjis and Kattan (2012d))

variable of the first damage process in which the cross-sectional area is reduced from A_1 to A_2, and let ϕ_2 be the damage variable of the next damage process in which the cross-sectional area is further reduced from A_2 to A_3. One continues in this fashion with a sequence of damage processes occurring in series (i.e., one following the other one) until the final damage variable ϕ_n of the final damage process is obtained in which the cross-sectional area is reduced from A_n to $\overline{A}$. This conceptual framework is illustrated in Fig. 18.

In this case, the following sequence of equations is obtained based on Eq. 20:

$$\phi_1 = \frac{A_1 - A_2}{A_1} \tag{38a}$$

$$\phi_2 = \frac{A_2 - A_3}{A_2} \tag{38b}$$

$$\phi_3 = \frac{A_3 - A_4}{A_3} \tag{38c}$$

$$\phi_n = \frac{A_n - \overline{A}}{A_n} \tag{38d}$$

The total damage variable ϕ in this case is defined based on the reduction of the initial damaged area A_1 to the effective undamaged area $\overline{A}$ as follows:

$$\phi = \frac{A_1 - \overline{A}}{A_1} \tag{39}$$

The following mathematical identity is defined:

$$\frac{\overline{A}}{A_1} = \frac{\overline{A}}{A_n}\frac{A_n}{A_{n-1}}\cdots\frac{A_3}{A_2}\frac{A_2}{A_1} \tag{40}$$

Substituting Eqs. 38 and 39 into Eq. 40, and simplifying the results, the following general expression for the total damage variable is obtained in terms of the individual damage variables:

$$1 - \phi = (1 - \phi_1)(1 - \phi_2)(1 - \phi_3) \ldots (1 - \phi_n) \tag{41}$$

In the special case of only two such processes ϕ_1 and ϕ_2 occurring in series, then Eq. 41 is reduced to the following simple form:

$$\phi = \phi_1 + \phi_2 - \phi_1 \phi_2 \tag{42}$$

Kattan and Voyiadjis (2001a) derived Eq. 42 considering the presence of voids and cracks in a damaged material within the context of decomposing the damage tensor into two components. The formulation presented here is more general and includes the work of Kattan and Voyiadjis (2001a) as a special case.

Consider next the n damage processes occurring in parallel. Check Fig. 19 where this process is illustrated using the reduction in cross-sectional areas. Let ϕ_1 be the first damage variable of the first damage process in which the cross-sectional area is reduced from A_1 to $\overline{A}$, and let ϕ_2 be the damage variable of the second damage process in which the cross-sectional area is reduced from A_2 to $\overline{A}$. One continues in this fashion with a sequence of damage processes occurring in parallel (i.e., at the same time) until the final damage variable ϕ_n is obtained for the final damage process in which the cross-sectional area is reduced from A_n to $\overline{A}$. This conceptual framework is illustrated schematically in Fig. 19.

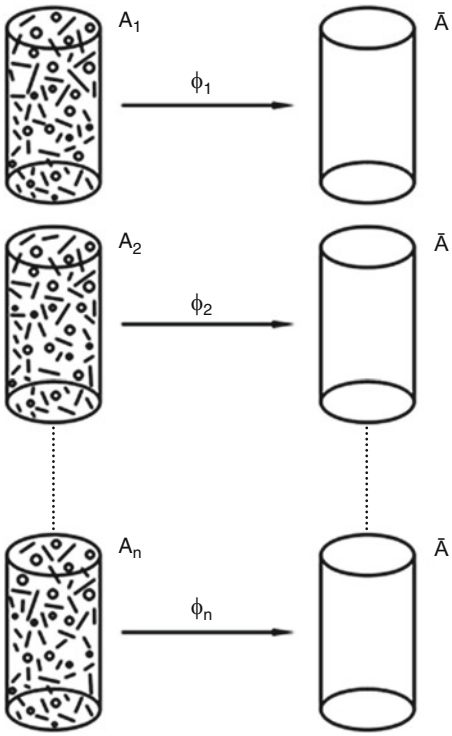


Fig. 19 Damage processes in parallel that are described in terms of reduction in cross-sectional area (Reprinted with permission from Voyiadjis and Kattan (2012d))

In this case the following sequence of equations based on Eq. 20 are obtained:

$$\phi_1 = \frac{A_1 - \overline{A}}{A_1} \tag{43a}$$

$$\phi_2 = \frac{A_2 - \overline{A}}{A_2} \tag{43b}$$

$$\phi_3 = \frac{A_3 - \overline{A}}{A_3} \tag{43c}$$

$$\phi_n = \frac{A_n - \overline{A}}{A_n} \tag{43d}$$

However, if one looks at the total (or cumulative) damage variable ϕ as shown in Fig. 7, then one assumes it to represent the reduction in cross-sectional area from the damaged area A to the effective undamaged area $\overline{A}$. Therefore, the total damage variable is given by the following expression:

$$\phi = \frac{A - \overline{A}}{A} \tag{44}$$

In order to find a way to relate the total damaged area A to the individual damaged areas A_1, A_2, ... A_n, some form of elaborate homogenization process should be used in this case. However, for the sake of simplicity and to illustrate the concepts introduced, the following simple relation is used:

$$A = c_1 A_1 + c_2 A_2 + \ldots + c_n A_n \tag{45}$$

where c_1, c_2, ... c_n are constants to be determined. For example, one may use $c_1 = c_2 = \ldots = c_n = \frac{1}{n}$ in which case the total damaged area is taken to be the average value of the individual damaged areas.

Substituting Eqs. 43 and 44 into Eq. 45, and simplifying the results, the following general relation is obtained between the total damage variable and the individual damage variables for the separate parallel damage processes:

$$\frac{1}{1-\phi} = \frac{c_1}{1-\phi_1} + \frac{c_2}{1-\phi_2} + \ldots + \frac{c_n}{1-\phi_n} \tag{46}$$

Using the special case of two damage processes ϕ_1 and ϕ_2 occurring in parallel, then Eq. 46 is reduced to the following simple form:

$$\phi = 1 - \frac{(1-\phi_1)(1-\phi_2)}{c_1(1-\phi_2) + c_2(1-\phi_1)} \tag{47}$$

In order to illustrate various combinations of these damage processes occurring in series and in parallel, one needs to use schematic diagrams. In this case, an elastic

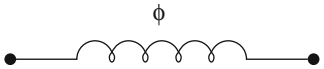


Fig. 20 Schematic illustration of a damage process in terms of reduction in cross-sectional area (Reprinted with permission from Voyiadjis and Kattan (2012d))

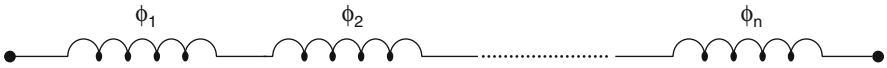


Fig. 21 Damage processes in series described in terms of reduction in cross-sectional area (Reprinted with permission from Voyiadjis and Kattan (2012d))

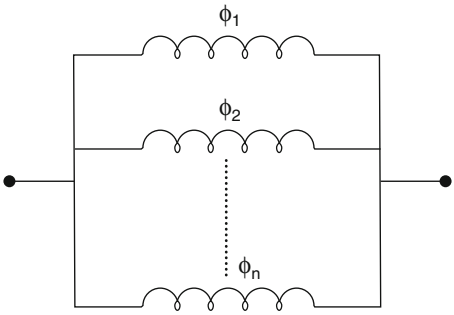


Fig. 22 Damage processes in parallel described in terms of reduction in cross-sectional area (Reprinted with permission from Voyiadjis and Kattan (2012d))

coil is used to illustrate an individual damage process that is described in terms of reduction in cross-sectional area as shown in Fig. 20.

Based on the illustration of Fig. 20, these types of damage processes are shown in series in Fig. 21 (corresponding to the cross-sectional areas of Fig. 18) and these types of damage processes in parallel in Fig. 22 (corresponding to the cross-sectional areas of Fig. 19).

Several combinations are now considered of the above damage processes. Consider first a sequence of n damage processes in series (characterized by the damage variables $\phi_1, \phi_2, \ldots \phi_n$) followed by m damage processes in parallel as shown in Fig. 23 (characterized by the damage variables $\phi'_{11}, \phi'_2, \ldots \phi'_m$).

This case can be investigated by considering that the first sequence of damage processes in series can be characterized by a single total damage variable called ϕ_A, while the second sequence of damage processes in parallel can be characterized by a single total damage variable called ϕ_B as illustrated in Fig. 24.

The damage processes in Fig. 24 are clearly occurring in series. Therefore, the total damage variable ϕ is given by (based on Eq. 41)

$$1 - \phi = (1 - \phi_A)(1 - \phi_B) \tag{48}$$

where ϕ_A and ϕ_B are obtained from Eqs. 41 and 46, respectively. Therefore, expanding Eq. 48, one obtains the following general expression for the total damage variable in order to describe the damage processes in Fig. 23:

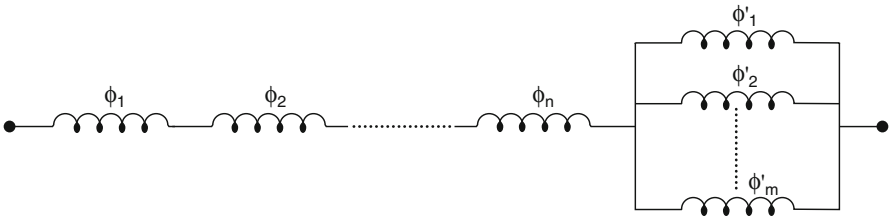


Fig. 23 Damage processes in series followed by damage processes in parallel (Reprinted with permission from Voyiadjis and Kattan (2012d))

Fig. 24 Characterization of the two sequences of damage processes of Fig. 23 (Reprinted with permission from Voyiadjis and Kattan (2012d))

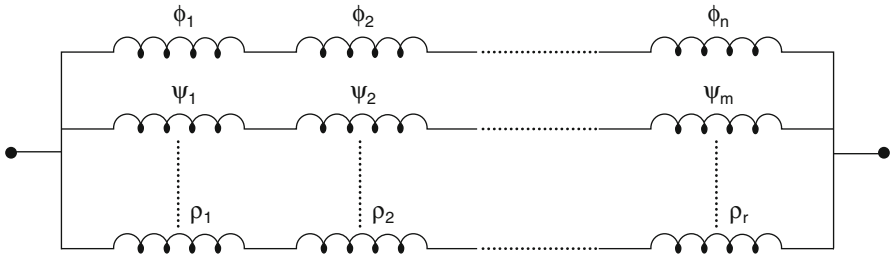


Fig. 25 Several damage processes in series occurring in parallel (Reprinted with permission from Voyiadjis and Kattan (2012d))

$$1-\phi=\frac{(1-\phi_1)(1-\phi_2)\ldots(1-\phi_n)}{\frac{c_1}{1-\phi_1'}+\frac{c_2}{1-\phi_2'}+\ldots+\frac{c_m}{1-\phi_m'}} \tag{49}$$

Another combination is considered of damage processes in series and in parallel. Consider the processes illustrated schematically in Fig. 25 where several sequences of damage processes in series occur collectively in parallel (characterized by the sequences of damage variables $\phi_1, \phi_2, \ldots \phi_n, \psi_1, \psi_2, \ldots \psi_m, \rho_1, \rho_2, \rho_r$).

In this case, let each sequence of damage processes in series be characterized by a single total damage variable ϕ_A, ϕ_B, $\ldots$ ϕ_Z as shown in Fig. 26.

In this case, the single total damage variables $\phi_A, \phi_B, \ldots \phi_Z$ can be obtained from Eq. 41 as follows:

$$1-\phi_A=(1-\phi_1)(1-\phi_2)\ldots(1-\phi_n) \tag{50a}$$

$$1-\phi_B=(1-\psi_1)(1-\psi_2)\ldots(1-\psi_m) \tag{50b}$$

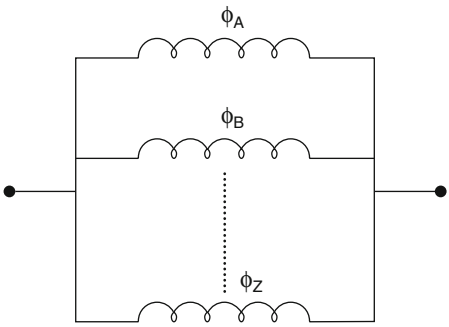


Fig. 26 Characterization of the damage processes of Fig. 25 (Reprinted with permission from Voyiadjis and Kattan (2012d))

$$1 - \phi_Z = (1 - \rho_1)(1 - \rho_2)\ldots(1 - \rho_r) \tag{50c}$$

Finally, the expression for the total damage variable can be obtained by using Eqs. 46 and 50 as follows to describe the damage processes in Fig. 25:

$$\frac{1}{1-\phi} = \frac{c_A}{(1-\phi_1)(1-\phi_2)\ldots(1-\phi_n)} + \frac{c_B}{(1-\psi_1)(1-\psi_2)\ldots(1-\psi_m)} + \ldots + \frac{c_Z}{(1-\rho_1)(1-\rho_2)\ldots(1-\rho_r)} \tag{51}$$

Illustrative Example

In this section, a simple example is illustrated in which four damage processes occur in a damaged system. The individual damage processes are not specified but are indicated with the respective damage variables. The objective of this example is to show how to handle situations in which different types of damage processes are involved. In this case, a mixed combination of two damage processes is described in terms of stiffness degradation (characterized by the damage variables ℓ_o and $2\ell_o$), and two other damage processes are described in terms of cross-sectional area reduction (characterized by ϕ_o and $3\phi_o$). This specific damage system is shown schematically in Fig. 27. The solution of this example will have several interpretations and important implications.

In trying to solve this simple example, one needs first to reduce the two damage processes operating in parallel that are designated with the damage variables $2\ell_o$ and $3\phi_o$. The net results of these two mixed parallel damage processes can be described in two different ways. The net result is described as a resulting damage process in terms of stiffness degradation. In this case, one designates the resulting damage state with the damage variable ℓ'. Alternatively, one can describe the net result as a resulting damage process that is obtained in terms of cross-sectional area reduction. In this case, the resulting damage state is designated with the damage variable ϕ'. These two differing alternative solutions are illustrated in Figs. 28 and 29, respectively.

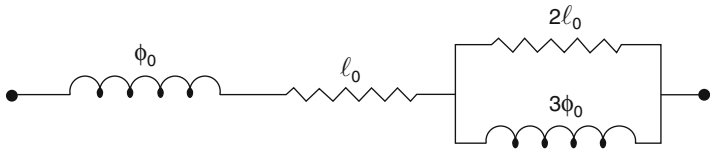


Fig. 27 An illustrative example (Reprinted with permission from Voyiadjis and Kattan (2012d))

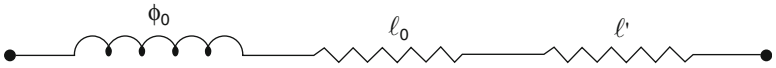


Fig. 28 The first alternative solution of the illustrative example (Reprinted with permission from Voyiadjis and Kattan (2012d))

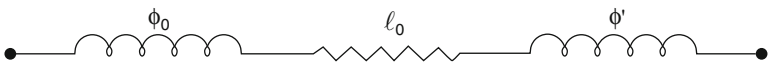


Fig. 29 The second alternative solution of the illustrative example (Reprinted with permission from Voyiadjis and Kattan (2012d))

Based on Eqs. 32 and 46, the following formulae are obtained for the two damage variables ℓ' and ϕ':

$$\frac{1}{1+\ell'} = \frac{c_1}{1+2\ell_o} + \frac{c_2}{1-3\phi_o} \tag{52}$$

$$\frac{1}{1-\phi'} = \frac{d_1}{1+2\ell_o} + \frac{d_2}{1-3\phi_o} \tag{53}$$

where c_1, c_2, d_1, d_2 are constants. In the simplest case, one can assign each one of the constants c_1, c_2, d_1, d_2 the value of $\frac{1}{2}$ in which case this is an averaging procedure. It should be noted that Eqs. 52 and 53 should not be used together; one chooses either Eq. 52 or Eq. 53 depending on the desired description of the resulting damage process.

In the last step, the final damage state of the material system which is schematically shown in Fig. 27 can be characterized in terms of stiffness degradation or cross-sectional area reduction. In the first case of stiffness degradation, the final damage state is designated by the damage variable ℓ, while in the second case in terms of the cross-sectional area reduction. The final damage state is designated by the damage variable ϕ. Utilizing Eqs. 27 and 41 along with Eqs. 52 and 53, one can write the following alternative expressions for the resulting damage variables ℓ and ϕ:

$$1+\ell = \frac{(1-\phi_o)(1+\ell_o)}{\dfrac{c_1}{1+2\ell_o} + \dfrac{c_2}{1-3\phi_o}} \tag{54}$$

$$1+\ell=\frac{(1-\phi_o)(1+\ell_o)}{\dfrac{d_1}{1+2\ell_o}+\dfrac{d_2}{1-3\phi_o}} \tag{55}$$

$$1-\phi=\frac{(1-\phi_o)(1+\ell_o)}{\dfrac{c_1}{1+2\ell_o}+\dfrac{c_2}{1-3\phi_o}} \tag{56}$$

$$1-\phi=\frac{(1-\phi_o)(1+\ell_o)}{\dfrac{d_1}{1+2\ell_o}+\dfrac{d_2}{1-3\phi_o}} \tag{57}$$

It should be noted that Eqs. 54, 55, 56, and 57 are four alternative descriptions for the final resulting damage state of the damage system described by Fig. 27. If one decides to describe the final damage state in terms of stiffness degradation, then one should use either Eq. 54 or Eq. 55 and utilize the final damage variable ℓ. On the other hand, if one decides to describe the final damage state in terms of cross-sectional area reduction, then one should use either Eq. 56 or Eq. 57 and utilize the final damage variable ϕ. It should be kept in mind that Eqs. 54, 55, 56, and 57 are four possible interpretations of the damage state depicted in Fig. 27. These four different interpretations should also be equivalent.

Three-Dimensional States of Deformation and Damage

In this section, a possible generalization of the concepts discussed in the previous sections to three-dimensional states of deformation and damage is proposed. For this purpose, use is made of tensorial damage variables instead of scalar damage variables. In addition, the tensorial damage variables will be represented in the form of their associated matrices.

The damage state described in terms of stiffness degradation can be easily generalized to three dimensions. This is performed directly by using the fourth-rank elasticity tensor. This tensor will be represented here by its associated matrix $[E]$. What remains is the proper form of the damage tensor to describe this case. The other damage state described in terms of cross-sectional area reduction is more difficult to generalize to three dimensions. This is due to the difficulty in generalizing the cross-sectional area along three perpendicular directions. This generalization of the cross-sectional was carried out in the literature using the fourth-rank damage effect tensor $[M]$ (Murakami 1988). However, this case will not be pursued here. The formulation will be limited to the first damage state only.

The generalization of Eq. 22 to three-dimensional states of deformation and damage can be carried out using the following two alternative expressions that were proposed originally by Voyiadjis and Kattan (2009):

$$[L]^{(1)} = \left(\left[\overline{E}\right] - [E]\right)[E]^{-1} \tag{58}$$

$$[L]^{(2)} = [E]^{-1}\left(\left[\overline{E}\right] - [E]\right) \tag{59}$$

where $[E]$ and $\left[\overline{E}\right]$ are the matrix representations of the fourth-rank damaged and effective undamaged elasticity tensors, respectively. In Eqs. 58 and 59, $[L]^{(1)}$ and $[L]^{(2)}$ are two possible matrix generalizations of the scalar damage variable ℓ.

Equations 58 and 59 are two alternative descriptions for the general damage state in terms of the degradation of the elastic modulus. Use is made of the above two equations to study these types of damage states when they occur in series and in parallel.

For damage processes occurring in series and following the same line of derivation presented in Eqs. 24, 25, 26, and 27 for the scalar case, one starts with Eqs. 58 and 59 in order to obtain the following final results:

$$[I] + [L]^{(1)} = \left([I] + [L]_1^{(1)}\right)\left([I] + [L]_2^{(1)}\right)\left([I] + [L]_3^{(1)}\right)\ldots\left([I] + [L]_n^{(1)}\right) \tag{60}$$

$$[I] + [L]^{(2)} = \left([I] + [L]_1^{(2)}\right)\left([I] + [L]_2^{(2)}\right)\left([I] + [L]_3^{(2)}\right)\ldots\left([I] + [L]_n^{(2)}\right) \tag{61}$$

where $[L]_k^{(1)}$ and $[L]_k^{(2)}$ are the proposed damage matrix representations for the individual damage processes in the series sequence $k = 1, 2, 3, \ldots n$ and $[I]$ is the identity matrix representing the identity tensor. It is clear that Eqs. 60 and 61 are the two possible generalizations of Eq. 27.

For damage processes occurring in parallel and following the same line of derivation presented in Eqs. 29, 30, 31, and 32 for the scalar case, one starts with Eqs. 58 and 59 to obtain the following final results:

$$\left([I] + [L]^{(1)}\right)^{-1} = c_1\left([I] + [L]_1^{(1)}\right)^{-1} + c_2\left([I] + [L]_2^{(1)}\right)^{-1} + \ldots + c_n\left([I] + [L]_n^{(1)}\right)^{-1} \tag{62}$$

$$\left([I] + [L]^{(2)}\right)^{-1} = c_1\left([I] + [L]_1^{(2)}\right)^{-1} + c_2\left([I] + [L]_2^{(2)}\right)^{-1} + \ldots + c_n\left([I] + [L]_n^{(2)}\right)^{-1} \tag{63}$$

where $[L]_k^{(1)}$ and $[L]_k^{(2)}$ are the proposed damage matrix representations for the individual damage processes in the parallel sequence $k = 1, 2, 3, \ldots n$. It is clear that Eqs. 62 and 63 are the two possible generalizations of Eq. 32.

For general damage states that are described by the reduction in cross-sectional area, one can use the fourth-rank damage tensor $[M]$ used by Murakami (1988) and Voyiadjis and Kattan (2009) to derive the generalized equations for this case

corresponding to Eqs. 41 and 46 of the scalar case. The generalization of Eq. 21 is well known in the literature and is written in the following form:

$$\{\overline{\sigma}\} = [M]\{\sigma\} \tag{64}$$

However, this type of generalization is not straightforward and will not be discussed here. The reason is that in this case one needs to utilize a hypothesis to derive the required equations. Usually the hypothesis of elastic strain equivalence or the hypothesis of elastic energy equivalence is used. Using these hypotheses in this work will not be carried out in order to maintain the simplicity of the current formulation of the conceptual framework. The interested reader may pursue this line of thought on his/her own after studying these two hypotheses and the associated damage tensors as presented by Voyiadjis and Kattan (2009).

Summary and Conclusions

This chapter consists of two distinct parts. The first part deals with the new topic of undamageable materials. These types of materials are currently hypothetical and are proposed in the hope that future technologies will be able to manufacture them. The theory of undamageable materials provides for a stress-free material that undergoes deformation while maintaining a zero value for the damage variable. The exact stress-strain relationship for these materials is derived. In the second part of this chapter, a conceptual framework is derived for two types of damage processes – those processes occurring in series (consecutively or following each other) and those occurring in parallel (simultaneously or at the same time). The complete mathematics of these types of processes is described. Finally, an example is shown to demonstrate the theory.

References

O.P. Allix, P. Ladeveze, D. Gilleta, R. Ohayon, A damage prediction method for composite structures. Int. J. Numer. Method. Eng. **27**(2), 271–283 (1989)

E.M. Arruda, M.C. Boyce, A three-dimensional constitutive model for the large stretch behavior of rubber elastic materials. J. Mech. Phys. Solids **41**(2), 389–412 (1993)

A. F. Bower, *Advanced Mechanics of Solids* (CRC Press, Boca Ration, FL, USA, 2009)

B. Budiansky, R.J. O'Connell, Elastic moduli of a cracked solid. Int. J. Solids Struct. **12**, 81–97 (1976)

A. Cauvin, R. Testa, Damage mechanics: basic variables in continuum theories. Int. J. Solids Struct. **36**, 747–761 (1999)

D.J. Celentano, P.E. Tapia, J-L.Chaboche, Experimental and numerical characterization of damage evolution in steels, in *Mecanica Computacional*, vol. XXIII, eds. by G. Buscaglia, E. Dari, O. Zamonsky (Bariloche, 2004)

C. Chow, J. Wang, An anisotropic theory of elasticity for continuum damage mechanics. Int. J. Fract. **33**, 3–16 (1987)

I. Doghri, *Mechanics of Deformable Solids: Linear and Nonlinear, Analytical and Computational Aspects* (Springer, Berlin, 2000)
N.R. Hansen, H.L. Schreyer, A thermodynamically consistent framework for theories of elastoplasticity coupled with damage. Int. J. Solids Struct. **31**(3), 359–389 (1994)
L. Kachanov, On the creep fracture time. Izv. Akad. Nauk USSR Otd. Tech. **8**, 26–31 (1958) (in Russian)
P.I. Kattan, G.Z. Voyiadjis, A coupled theory of damage mechanics and finite strain elasto-plasticity – part I: damage and elastic deformations. Int. J. Eng. Sci. **28**(5), 421–435 (1990)
P.I. Kattan, G.Z. Voyiadjis, A plasticity-damage theory for large deformation of solids – part II: applications to finite simple shear. Int. J. Eng. Sci. **31**(1), 183–199 (1993a)
P.I. Kattan, G.Z. Voyiadjis, Overall damage and elastoplastic deformation in fibrous metal matrix composites. Int. J. Plast. **9**, 931–949 (1993b)
P.I. Kattan, G.Z. Voyiadjis, Decomposition of damage tensor in continuum damage mechanics. J. Eng. Mech. ASCE **127**(9), 940–944 (2001a)
P.I. Kattan, G.Z. Voyiadjis, *Damage Mechanics with Finite Elements: Practical Applications with Computer Tools* (Springer, Berlin, 2001b)
D. Krajcinovic, *Damage Mechanics* (Elsevier, Amsterdam, 1996)
P. Ladeveze, M. Poss, L. Proslier, Damage and fracture of tridirectional composites, in *Progress in Science and Engineering of Composites. Proceedings of the Fourth International Conference on Composite Materials*, vol. 1 (Japan Society for Composite Materials, Tokyo, Japan, 1982) pp. 649–658
H. Lee, K. Peng, J. Wang, An anisotropic damage criterion for deformation instability and its application to forming limit analysis of metal plates. Eng. Fract. Mech. **21**, 1031–1054 (1985)
J. Lemaitre, How to use damage mechanics. Nucl. Eng. Des. **80**, 233–245 (1984)
V. Lubarda, D. Krajcinovic, Damage tensors and the crack density distribution. Int. J. Solids Struct. **30**(20), 2859–2877 (1993)
B. Luccioni, S. Oller, A directional damage model. Comput. Method. Appl. Mech. Eng. **192**, 1119–1145 (2003)
S. Murakami, Mechanical modeling of material damage. ASME J. Appl. Mech. **55**, 280–286 (1988)
S. Murakami, Continuum damage mechanics, A Continuum Mechanics Approach to the Analysis of Damage and Fracture (Springer, Dordrecht; New York, USA, 2012)
J.M. Nichols, A.B.. Abell, Implementing the degrading effective stiffness of masonry in a finite element model, in *North American Masonry Conference*, Clemson, 2003
J.M. Nichols, Y.Z. Totoev, Experimental investigation of the damage mechanics of masonry under dynamic in-plane loads, in *North American Masonry Conference*, Austin, 1999
Y. Rabotnov, Creep rupture, in *Proceedings, Twelfth International Congress of Applied Mechanics*, Stanford, 1968, eds. by M. Hetenyi, W.G. Vincenti (Springer, Berlin, 1969) pp 342–349
J.R. Rice, Inelastic constitutive relations for solids: an internal variable theory and its application to metal plasticity. J. Mech. Phys. Solids **19**, 433–455 (1971)
J. Shen, J. Mao, G. Reyes, C.L. Chow, J. Boileau, X. Su, J. Wells, A multiresolution transformation rule of material defects. Int. J. Damage Mech. **18**(11), 739–758 (2011)
F. Sidoroff, Description of anisotropic damage application to elasticity, in *IUTAM Colloquium on Physical Nonlinearities in Structural Analysis* (Springer, Berlin, 1981), pp. 237–244
V.V. Silberschmidt, Model of matrix cracking in carbon fiber-reinforced cross-ply laminates. Mech. Compos. Mater. Struct. **4**(1), 23–38 (1997)
G.Z. Voyiadjis, Degradation of elastic modulus in elastoplastic coupling with finite strains. Int. J. Plast. **4**, 335–353 (1988)
G.Z. Voyiadjis, P.I. Kattan, A coupled theory of damage mechanics and finite strain elasto-plasticity – part II: damage and finite strain plasticity. Int. J. Eng. Sci. **28**(6), 505–524 (1990)
G.Z. Voyiadjis, P.I. Kattan, A plasticity-damage theory for large deformation of solids – part I: theoretical formulation. Int. J. Eng. Sci. **30**(9), 1089–1108 (1992)

G.Z. Voyiadjis, P.I. Kattan, On the symmetrization of the effective stress tensor in continuum damage mechanics. J. Mech. Behav. Mater. **7**(2), 139–165 (1996)
G.Z. Voyiadjis, P.I. Kattan, *Advances in Damage Mechanics: Metals and Metal Matrix Composites* (Elsevier Science, Amsterdam, 1999)
G.Z. Voyiadjis, P.I. Kattan, *Damage Mechanics* (Taylor and Francis/CRC Press, Boca Raton, FL, USA, 2005)
G.Z. Voyiadjis, P.I. Kattan, *Advances in Damage Mechanics: Metals and Metal Matrix Composites with an Introduction to Fabric Tensors*, 2nd edn. (Elsevier, Amsterdam, 2006a)
G.Z. Voyiadjis, P.I. Kattan, A new fabric-based damage tensor. J. Mech. Behav. Mater. **17**(1), 31–56 (2006b)
G.Z. Voyiadjis, P.I. Kattan, Damage mechanics with fabric tensors. Mech. Adv. Mater. Struct. **13**(4), 285–301 (2006c)
G.Z. Voyiadjis, P.I. Kattan, Evolution of fabric tensors in damage mechanics of solids with microcracks: part I – theory and fundamental concepts. Mech. Res. Commun. **34**(2), 145–154 (2007a)
G.Z. Voyiadjis, P.I. Kattan, Evolution of fabric tensors in damage mechanics of solids with microcracks: part II – evolution of length and orientation of micro-cracks with an application to uniaxial tension. Mech. Res. Commun. **34**(2), 155–163 (2007b)
G.Z. Voyiadjis, P.I. Kattan, A comparative study of damage variables in continuum damage mechanics. Int. J. Damage Mech. **18**(4), 315–340 (2009)
G.Z. Voyiadjis, P.I. Kattan, A new class of damage variables in continuum damage mechanics. J. Eng. Mater. Technol. Trans. ASME **134**(2), 021016-1-10 (2012a)
G.Z. Voyiadjis, P.I. Kattan, Introduction to the mechanics and design of undamageable materials. Int. J. Damage Mech. **22**(3), 323–335 (2012b)
G.Z. Voyiadjis, P.I. Kattan, On the theory of elastic undamageable materials. J. Eng. Mater. Technol. Trans. ASME **135**(2), 021002-1-6 (2012c)
G.Z. Voyiadjis, P.I. Kattan, Mechanics of damage processes in series and in parallel: a conceptual framework. Acta Mech. **223**(9), 1863–1878 (2012d)
G.Z. Voyiadjis, M.A. Yousef, P.I. Kattan, New tensors for anisotropic damage in continuum damage mechanics. J. Eng. Mater. Technol. Trans. ASME **134**(2), 021015-1-7 (2012)

Use of Fabric Tensors in Continuum Damage Mechanics of Solids with Micro-cracks

3

George Z. Voyiadjis, Peter I. Kattan, and Ziad N. Taqieddin

Contents

Abstract

In this chapter, a new formulation is presented to link continuum damage mechanics with the concept of fabric tensors within the framework of classical elasticity theory. A fourth-rank damage tensor is used and its exact relationship to the fabric tensors is illustrated. A model of damage mechanics for directional data is formulated using fabric tensors. The applications of the new formulation to micro-crack distributions are well illustrated in two solved examples. In the first example, a micro-crack distribution is considered with its data represented by a circular histogram. The values of the fabric tensors and damage tensor are

G.Z. Voyiadjis (✉) • P.I. Kattan
Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA
e-mail: voyiadjis@eng.lsu.edu; pkattan@orange.jo

Z.N. Taqieddin
Civil Engineering Department, Applied Science University, Amman, Jordan
e-mail: z_taqieddin@asu.edu.jo

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_3

calculated in this case. In the second example, two sets of parallel micro-crack distributions with two different orientations are investigated.

A general hypothesis for damage mechanics is postulated. It is seen that the two available hypotheses of elastic strain equivalence and elastic energy equivalence may be obtained as special cases of the postulated general hypothesis. This general hypothesis is then used to derive the sought relationship between the damage tensor and fabric tensors. Finally, the evolution of the damage tensor is derived in a mathematically consistent manner that is based on sound thermodynamic principles.

Introduction

The major objective of this work is to try to find a physical interpretation of the damage tensor. Since its beginnings, the subject of continuum damage mechanics has been plagued with controversy as the concept of the damage tensor was not based on a sound physical ground. The subject of this work is to try to link the damage tensor with the concept of fabric tensors which have valid and convincing physical interpretation. The fabric tensors have been formulated by Kanatani (1984a) to describe directional data and microstructural anisotropy and further elaborated upon by Lubarda and Krajcinovic (1993) to describe crack distributions.

Satake (1982) applied the concept of fabric tensors to granular materials. The anisotropy due to the fabric (of the distributed data like crack distributions or granular particles) is represented by a tensor in terms of the normals (to the cracks or to the contact surfaces in granular materials). This tensor is usually called the fabric tensor (Satake 1982; Kanatani 1984a; Oda et al. 1982). The fabric tensor is usually related to the probability density function of the distributed data (crack normals or contact normals).

Kanatani (1984a) formulated the concept of fabric tensors based on a rigorous mathematical treatment. He used fabric tensors to describe distributions of directional data like crack distributions in a damaged material element. He applied the least square approximation (a well-known statistical technique) to derive equations for the various fabric tensors he postulated. He defined three types of fabric tensors: fabric tensors of the first kind, denoted by $\mathbf{N}$; fabric tensors of the second kind, denoted by $\mathbf{F}$; and fabric tensors of the third kind, denoted by $\mathbf{D}$. He derived the exact mathematical relations between these three types of fabric tensors. The work of Kanatani (1984a) on fabric tensors is very important and is used extensively here.

Zysset and Curnier (1995, 1996) formulated an alternative model for anisotropic elasticity based on fabric tensors. Actually Cowin (1989) made an attempt to relate the microstructure (through the use of fabric tensors) to the fourth-rank elasticity tensor. He used a normalized second-rank tensor and presented expressions for the elastic constants in terms of the invariant of the fabric tensors. Zysset and Curnier (1995) introduced a general approach for relating the material microstructure to the fourth-rank elasticity tensor based on the Fourier series decomposition. They proposed an approximation based on a scalar and a symmetric traceless

second-rank fabric tensor. Using the representation theorem for anisotropic functions with tensorial arguments, Zysset and Curnier (1995) derived a general expression for the elastic free energy and discussed the resulting material symmetry in terms of the fabric tensors. Finally, they derived a general explicit expression for the fourth-rank elasticity tensor in terms of the fabric tensor. This last result is very important and is used extensively here (Cowin 1989).

Lubarda and Krajcinovic (1993) applied the definitions of fabric tensors (Kanatani 1984a) to crack density distributions. They actually recast Kanatani's general work on directional data (Kanatani 1984a) in terms of crack distributions. Lubarda and Krajcinovic (1993) examined the relationship between a given, experimentally determined, distribution of cracks and the scalar second-rank and fourth-rank fabric tensors. They employed the usual representation of experimentally measured micro-crack densities in planes with different orientations in the form of circular histogram (rose diagram). They then used the data contained in the circular histogram to approximate the distribution function defined on a unit sphere and centered in a material point. They solved several examples with different crack distributions to illustrate this point. They assumed that one of the three types of fabric tensors is identical to the damage tensor of continuum damage mechanics.

The damage variable (or tensor), based on the effective stress concept, represents average material degradation which reflects the various types of damage at the microscale level like nucleation and growth of voids, cracks, cavities, micro-cracks, and other microscopic defects.

For the case of isotropic damage mechanics, the damage variable is scalar and the evolution equations are easy to handle. However, it has been show by Cauvin and Testa (1999) that two independent damage variables must be used in order to describe accurately and consistently the special case of isotropic damage. It has been argued (Lemaitre 1984) that the assumption of isotropic damage is sufficient to give good predictions of the load carrying capacity, the number of cycles, or the time to local failure in structural components. However, the development of anisotropic damage has been confirmed experimentally (Hayhurst 1972; Chow and Wang 1987; Lee et al. 1985) even if the virgin material is isotropic. This has prompted several researchers to investigate the general case of anisotropic damage (Voyiadjis and Kattan 1996, 1999; Kattan and Voyiadjis 2001a, b).

In continuum damage mechanics, usually a phenomenological approach is adopted. In this approach, the most important concept is that of the representative volume element (RVE). The discontinuous and discrete elements of damage are not considered within the RVE; rather their combined effects are lumped together through the use of a macroscopic internal variable. In this way, the formulation may be derived consistently using sound mechanical and thermodynamic principles.

In this chapter, a new formulation is investigated to find a relationship between the damage tensor of continuum damage mechanics and the concept of the fabric tensors within the framework of classical elasticity theory. Actually what is attempted here is the coupling of the three theories of damage mechanics, fabric tensors, and classical elasticity in formulating a new theory that can accurately

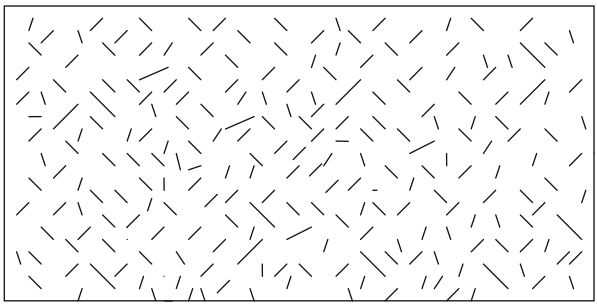


Fig. 1 A cross section showing a typical random micro-crack distribution (Reprinted with permission from G. Voyiadjis, P. Kattan, Mech. Mater. Struct. 13, 4 (2006))

describe real practical engineering problems involving anisotropy and directional data like complicated micro-crack distributions. An explicit expression of the fourth-rank damage tensor is derived in terms of the fabric tensors. The exact relationship between the damage tensor and the fabric tensors is also illustrated in detail.

The formulation is presented within the framework of the usual classical theory of elasticity. Starting with an RVE with a micro-crack distribution that is experimentally determined like that in Fig. 1, one proceeds to calculate the fabric tensors for the micro-crack distribution based on the data contained within the circular histogram. This step was performed by many authors (Kanatani 1984a; Lubarda and Krajcinovic 1993). The next step involves calculating the damage tensor using the fabric tensors determined in the previous step. This step is performed using the new formulas derived in the current formulation. Using the values of the damage tensor components thus calculated, one proceeds to calculate the classical elasticity tensor. The elasticity tensor calculated in this manner represents the elasticity tensor of the damaged material reflecting the characteristics of the microstructure. This final elasticity tensor can then be used to solve boundary-value problems involving the aforementioned micro-crack distribution.

Two applications are presented and solved numerically to illustrate the new formulation. The first application involves a micro-crack distribution within an RVE of a damaged material. A circular histogram is presented for the data in the micro-crack distribution. This histogram is then used to determine the fabric tensors needed. These fabric tensors are then used to determine the damage tensor. Finally, various relations between the damage tensor and fabric tensors are illustrated for this particular example. The second application involves the investigation of two sets of parallel micro-cracks. It is seen that in this case, the interchange of the angle of orientation of the parallel micro-cracks causes a corresponding interchange in the first two diagonal terms in the elastic stiffness matrix.

It should be noted that the theory presented here is for general directional data. When applying this theory for micro-crack distributions, then the theory depends exclusively on the orientations of micro-cracks. This theory does not consider the growth of micro-crack lengths, the opening of the micro-cracks, the closeness of the micro-cracks, and the interaction among the micro-cracks. These effects are beyond the scope of this work. Detailed description of the theory that depends on the

orientation and length of micro-cracks can be found in ▶ Chap. 4, "Evolution of Fabric Tensors in Continuum Damage Mechanics of Solids with Micro-cracks: Studying the Effects of Length and Orientation."

The reader is referred to the book by Nemat-Nasser (2004) regarding damage mechanics and fabric. In particular the reader should refer to ▶ Chap. 7. This chapter covers the issue of fabric rather thoroughly, both theoretically and experimentally. There is a detailed discussion and review of the fundamentals of the characterization of the distribution of the contact normals in granular materials, the relation with stress tensor, and many other issues. In addition the book by Nemat-Nasser and Hori (1999) addresses many examples of micro-crack distributions for several conditions, including opening and closing of micro-cracks, random distributions, effect of friction, load-induced anisotropy, and interactions of micro-cracks. They point out in detail the effect of these issues on the elastic moduli. However, in all of these examples, they assume that the micro-crack distributions are dilute. In the work presented in this chapter, the authors do not have this assumption of dilute micro-crack distributions in the theory or examples. Although the proposed approach does not consider many of the issues mentioned above, the micro-crack distributions are general in the sense that there are no limits on the volume fraction of the micro-cracks. However, the model presented here is limited to considering only the orientation of micro-cracks, without considering opening and closing of micro-cracks, effect of friction, or interaction of micro-cracks. Based on the above remarks, the authors do not see a basis for a consistent comparison of their results with those of Nemat-Nasser and Hori (1999). Effectively, the authors cannot compare the results of dilute distributions with those of non-dilute systems. However, they do acknowledge the work and examples in the book of Nemat-Nasser and Hori (1999).

The reader is also referred to the work of Voyiadjis et al. (2007a) and (2007b) where the applications of fabric tensors in damage mechanics discussed here are further extended to study composite materials.

The tensor notation used here is as follows. All vectors and tensors appear in bold type. The following operations are also defined. For second-rank tensors **A** and **B**, the following notation is used:

$$(\mathbf{A} \pm \mathbf{B})_{ij} = A_{ij} \pm B_{ij}, \quad \mathbf{A} : \mathbf{B} = A_{ij}B_{ij}, \quad (\mathbf{A} \otimes \mathbf{B})_{ijkl} = A_{ij}B_{kl},$$
$$\left(\mathbf{A} \underline{\otimes} \mathbf{B}\right)_{ijkl} = \frac{1}{2}\left(A_{ik}B_{jl} + A_{il}B_{jk}\right)$$

For fourth-rank tensors **C** and **D**, the following notation is used: $(\mathbf{C} \pm \mathbf{D})_{ijkl} = C_{ijkl} \pm D_{ijkl}$, $(\mathbf{C} : \mathbf{D})_{ijkl} = C_{ijmn}D_{mnkl}$, $\mathbf{C} :: \mathbf{D} = C_{ijkl}D_{ijkl}$, $(\mathbf{C} \otimes \mathbf{D})_{ijklmnpq} = C_{ijkl}D_{mnpq}$. For second-rank tensor **A** and fourth-rank tensor **C**, the following notation is used: $(\mathbf{C} : \mathbf{A})_{ij} = C_{ijkl}A_{kl}$. For fourth-rank tensor **C** and eighth-rank tensor **F**, the following notation is used: $(\mathbf{F} :: \mathbf{C})_{ijkl} = F_{ijklmnpq}C_{mnpq}$. For damage tensors, fabric tensors, and identity tensors, a superscript with braces is used to indicate the order of the tensor. For all other tensors, the order of the tensor is clear from the text and equations.

Fabric Tensors

In this section, the use of fabric tensors to describe directional data and microstructural anisotropy are reviewed. Kanatani (1984a, b) introduced the idea of fabric tensors with regard to the distribution of directional data. He used fabric tensors for the stereological determination of structural isotropy. Zysset and Curnier (1995) used fabric tensors to derive an alternative model of anisotropic elasticity. They derived a new formula for the general elasticity tensor of damaged materials in terms of fabric tensors. He and Curnier (1995) formulated a more fundamental approach to damaged elastic stress–strain relations using fabric tensors. Zysset and Curnier (1996) used a damage model based on fabric tensors in the analysis of trabecular bone. Sutcliffe (1992) presented a rigorous mathematical formulation for the spectral decomposition of the general elasticity tensor. This important result is of paramount importance and will be used later in this work.

Consider a distribution of directional data that is radially symmetric, i.e., symmetric with respect to the origin. Let **n** be a unit vector specifying the orientation and consider the orientation distribution function $f(\mathbf{N})$ where **N** is given by

$$N^{(0)} = 1 \tag{1.1}$$

$$N_{ij}^{(2)} = < n_i n_j > = \frac{1}{N} \sum_{\alpha=1}^{N} n_i^{(\alpha)} n_j^{(\alpha)} \tag{1.2}$$

where N is the number of the directional data. In Eqs. 1.1 and 1.2, $N^{(0)}$ and $\mathbf{N}^{(2)}$ are called the zero-rank and second-rank fabric tensors of the first kind, respectively (Kanatani 1984a). Kanatani (1984a) also defined two other fabric tensors **F** and **D** as follows:

$$F^{(0)} = 1 \tag{1.3}$$

$$F_{ij}^{(2)} = \frac{15}{2}\left(N_{ij}^{(2)} - \frac{1}{5}\delta_{ij}\right) \tag{1.4}$$

$$D^{(0)} = 1 \tag{1.5}$$

$$D_{ij}^{(2)} = \frac{15}{2}\left(N_{ij}^{(2)} - \frac{1}{3}\delta_{ij}\right) \tag{1.6}$$

where $F^{(0)}$ and $\mathbf{F}^{(2)}$ are the zero-rank and second-rank fabric tensors of the second kind, respectively; $D^{(0)}$ and $\mathbf{D}^{(2)}$ are the zero-rank and second-rank fabric tensors of the third kind, respectively; and δ_{ij} is the Kronecker delta.

It is assumed here that the distribution function f is to be always positive and square integrable. Then, the function f can be expanded in a convergent Fourier series as follows (Jones 1985; Zysset and Curnier 1995):

$$f(\mathbf{N}) = G^{(0)}.1 + \mathbf{G}^{(2)} : \mathbf{F}^{(2)}(\mathbf{N}) + \mathbf{G}^{(4)} :: \mathbf{F}^{(4)}(\mathbf{N}) + \ldots \tag{2}$$

for each $\mathbf{N}$, where $G^{(0)}$, $\mathbf{G}^{(2)}$, and $\mathbf{G}^{(4)}$ are zero-rank (i.e., scalar), second-rank, and fourth-rank fabric tensors, respectively, while 1, $\mathbf{F}^{(2)}(\mathbf{N})$, and $\mathbf{F}^{(4)}(\mathbf{N})$ are zero-rank (i.e., scalar), second-rank, and fourth-rank basis functions, respectively. It should be noted that $G^{(0)}$ and $\mathbf{G}^{(2)}$ are exactly the same fabric tensors $D^{(0)}$ and $\mathbf{D}^{(2)}$ of the third kind of Kanatani (1984a). Furthermore, note that the basis function $\mathbf{F}^{(2)}(\mathbf{N})$ of Zysset and Curnier (1995) in Eq. 2 is different from the second-rank fabric tensor of the second kind $\mathbf{F}^{(2)}$ of Kanatani (1984a) in Eq. 1.4, although the same symbols are used for both quantities. The basis functions $\mathbf{F}^{(2)}(\mathbf{N})$ and $\mathbf{F}^{(4)}(\mathbf{N})$ are given by (Kanatani 1984a, b; Zysset and Curnier 1995)

$$\mathbf{F}^{(2)}(\mathbf{N}) = \mathbf{N} - \frac{1}{3}\mathbf{I}^{(2)} \tag{3.1}$$

$$\begin{aligned}\mathbf{F}^{(4)}(\mathbf{N}) = \mathbf{N}\otimes\mathbf{N} - \frac{1}{7}\left(\mathbf{I}^{(2)}\otimes\mathbf{N} + \mathbf{N}\otimes\mathbf{I}^{(2)}\right) - \frac{2}{7}\left(\mathbf{I}^{(2)}\underline{\otimes}\mathbf{N} + \mathbf{N}\underline{\otimes}\mathbf{I}^{(2)}\right) \\ + \frac{1}{35}\mathbf{I}^{(2)}\otimes\mathbf{I}^{(2)} + \frac{2}{35}\mathbf{I}^{(2)}\underline{\otimes}\mathbf{I}^{(2)}\end{aligned} \tag{3.2}$$

The three fabric tensors $G^{(0)}$, $\mathbf{G}^{(2)}$, and $\mathbf{G}^{(4)}$ are determined using the following integrals (Zysset and Curnier 1995):

$$G^{(0)} = \frac{1}{4\pi}\int_S f(\mathbf{N})da \tag{4}$$

$$\mathbf{G}^{(2)} = \frac{15}{8\pi}\int_S f(\mathbf{N})\,\mathbf{F}^{(2)}(\mathbf{N})da \tag{5}$$

$$\mathbf{G}^{(4)} = \frac{315}{32\pi}\int_S f(\mathbf{N})\,\mathbf{F}^{(4)}(\mathbf{N})da \tag{6}$$

where S is the surface of the unit sphere and a is the integration parameter.

Kanatani (1984a, b) showed that the first two terms in the expansion given in Eq. 2 are enough and they can describe material anisotropy sufficiently and accurately. Therefore, the third term in the expansion is neglected and only the first two terms are retained as follows:

$$f(\mathbf{N}) \approx G^{(0)}.1 + \mathbf{G}^{(2)} : \mathbf{F}^{(2)}(\mathbf{N}) \tag{7}$$

Thus, it is clear from the above expression that only zero-rank (scalar) and second-rank fabric tensors will be dealt with – there is no need to deal with the fourth-rank fabric tensor. It also should be noted that the function f in the above approximation (Eq. 7) must remain always positive.

The approximation of the distribution function $f(\mathbf{N})$ given in Eq. 7 characterizes anisotropy, i.e., the traceless second-rank tensor $\mathbf{G}^{(2)}$ describes orthotropy with three orthogonal planes of symmetry and all three eigenvalues being distinct. Using only the first term in Eq. 7, i.e., $f(\mathbf{N}) = G^{(0)}$, will characterize the special case of isotropy. The case of transverse isotropy is characterized if the second-rank tensor $\mathbf{G}^{(2)}$ has only two eigenvalues that are distinct (Zysset and Curnier 1995).

Let $\overline{\mathbf{E}}$ be the fourth-rank constant elasticity tensor for the virgin material in the undamaged configuration and is assumed to be here for isotropic materials. Next, the expression of the fourth-rank constant elasticity tensor $\overline{\mathbf{E}}$ is written as follows:

$$\overline{\mathbf{E}} = \lambda \mathbf{I}^{(2)} \otimes \mathbf{I}^{(2)} + 2\mu \mathbf{I}^{(2)} \underline{\bar{\otimes}} \mathbf{I}^{(2)} \tag{8}$$

where λ and μ are Lame's constants. Zysset and Curnier (1995) showed that by replacing the identity tensor $\mathbf{I}^{(2)}$ in the expression of $\overline{\mathbf{E}}$ by the tensor $G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}$, the fourth-rank tensor $\mathbf{E}$ (a fourth-rank variable elasticity tensor for the damaged material) which includes the effects of microstructural anisotropy and directional data, i.e., effects of damage, is obtained. Thus, the following expression for $\mathbf{E}$ is now available (see Eq. 8 in Zysset and Curnier (1995)):

$$\begin{aligned}\mathbf{E} = {} & \lambda\left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \otimes \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \\ & + 2\mu\left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \underline{\bar{\otimes}} \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right)\end{aligned} \tag{9}$$

It is clear that the expression given in Eq. 9 provides a formula for the elasticity tensor $\mathbf{E}$ of the damaged material in terms of the two fabric tensors $G^{(0)}$ and $\mathbf{G}^{(2)}$.

Next, the spectral decomposition of the second-rank fabric tensor $\mathbf{G}^{(2)}$ is considered as follows:

$$\mathbf{G}^{(2)} = \sum_{i=1}^{3} g_i(\mathbf{g}_i \times \mathbf{g}_i) \tag{10}$$

where $g_i (i = 1, 2, 3)$ are the eigenvalues of $\mathbf{G}^{(2)}$ and $\mathbf{g}_i (i = 1, 2, 3)$ are the corresponding eigenvectors. Zysset and Curnier (1995) used the terminology $\mathbf{G}_i$ to denote the dyadic product $\mathbf{g}_i \times \mathbf{g}_i$ as follows:

$$\mathbf{G}_i = \mathbf{g}_i \times \mathbf{g}_i \quad \text{(no sum over i)} \tag{11}$$

where it is clear that $\sum_{i=1}^{3} \mathbf{G}_i = \mathbf{I}^{(2)}$. Using this new terminology, the following alternative expression for $\mathbf{E}$, which was derived in detail by Zysset and Curnier (1995) in the principal coordinate system, can be written as (see Eq. 11 in Zysset and Curnier (1995))

$$\begin{aligned}\mathbf{E} = (\lambda + 2\mu) m_i^{2k} \left(\mathbf{G}_i \otimes \mathbf{G}_i\right) + \lambda m_i^k m_j^k \left(\mathbf{G}_i \otimes \mathbf{G}_j + \mathbf{G}_j \otimes \mathbf{G}_i\right) \\ + 2\boldsymbol{\mu} m_i^k m_j^k \left(\mathbf{G}_i \underline{\otimes} \mathbf{G}_j + \mathbf{G}_j \underline{\otimes} \mathbf{G}_i\right)\end{aligned} \tag{12}$$

where k is a constant scalar parameter and m_i is given by the following:

$$m_i = G^{(0)} + g_i \tag{13}$$

In the above equation, it should be noted that $\sum_{i=1}^{3} m_i = \text{constant}$. It should also be noted that Eq. 12 is valid for damaged materials while Eq. 9 is valid only for granular materials. This is because of the different properties of the microstructure of damaged materials and granular materials. The effects of the fabric tensor on the elasticity tensor for these two kinds of materials are totally in contrary, in that if one principal value of the fabric tensor is larger, the associated Young's modulus will be larger for granular materials and smaller for damaged materials. For the special case of isotropy, $g_i = 0$, $(i = 1, 2, 3)$ and $m_i = G^{(0)}$, $(i = 1, 2, 3)$, so Eq. 12 reduces to the following equation of isotropic elasticity:

$$\begin{aligned}\mathbf{E} = (\lambda + 2\boldsymbol{\mu}) g^{2k} \left(\mathbf{G}_i \otimes \mathbf{G}_i\right) + \lambda g^{2k} \left(\mathbf{G}_i \otimes \mathbf{G}_j + \mathbf{G}_j \otimes \mathbf{G}_i\right) \\ + 2\boldsymbol{\mu} g^{2k} \left(\mathbf{G}_i \otimes \mathbf{G}_j + \mathbf{G}_j \otimes \mathbf{G}_i\right)\end{aligned} \tag{14}$$

Finally, the stress tensor σ_{ij} is related to the strain tensor ε_{ij} through the fourth-rank variable elasticity tensor $\mathbf{E}$ as follows:

$$\sigma_{ij} = E_{ijkl} \varepsilon_{kl} \tag{14.1}$$

General Hypothesis and New Formulation of Damage Mechanics

In this section, derivations are presented for the important concepts of damage mechanics that are relevant to this work, particularly to fabric tensors. This derivation is presented within the general framework of continuum damage mechanics (Cauvin and Testa 1999; Voyiadjis and Kattan 1999) using a general hypothesis that is postulated here. It is shown that general states of anisotropic damage in the material must be described by a fourth-rank damage tensor.

Let $\overline{\mathbf{E}}$ be the fourth-rank constant elasticity tensor of the virgin material and let $\mathbf{E}$ be the elasticity tensor of the damaged material. Then, the two tensors $\overline{\mathbf{E}}$ and $\mathbf{E}$ are related by the following general relation (Cauvin and Testa 1999):

$$\mathbf{E} = \left(\mathbf{I}^{(8)} - \boldsymbol{\varphi}^{(8)}\right) :: \overline{\mathbf{E}} \tag{15}$$

where $\mathbf{I}^{(8)}$ is the eighth-rank identity tensor and $\boldsymbol{\varphi}^{(8)}$ is the general eighth-rank damage tensor.

Next, a new formulation is derived and a general hypothesis is postulated to show that Eq. 15 can be reduced to a similar equation involving a damage tensor of rank four at most. Cauvin and Testa (1999) have shown this result only for the special case of the hypothesis of elastic strain equivalence. Therefore, there will be no need to deal with the eighth-rank general damage tensor $\boldsymbol{\varphi}^{(8)}$ in the constitutive equations.

Kachanov (1958) and Rabotnov (1969) introduced the concept of effective stress for the case of uniaxial tension. This concept was later generalized to three-dimensional states of stress by Lemaitre (1971) and Chaboche (1981). Let $\boldsymbol{\sigma}$ be the second-rank Cauchy stress tensor and $\overline{\boldsymbol{\sigma}}$ be the corresponding effective stress tensor. The effective stress $\overline{\boldsymbol{\sigma}}$ is the stress applied to a fictitious state of the material which is totally undamaged, i.e., all damage in this state has been removed. This fictitious state is assumed to be mechanically equivalent to the actual damaged state of the material. In this regard, one of the two hypotheses (elastic strain equivalence or elastic energy equivalence) is usually used. However, in this work, a general hypothesis of strain transformation is postulated. The elastic strain tensor $\boldsymbol{\varepsilon}^e$ in the actual damaged state is related to the effective elastic strain tensor $\overline{\boldsymbol{\varepsilon}}^e$ in the fictitious state by the following transformation law:

$$\overline{\boldsymbol{\varepsilon}}^e = \mathbf{L}\left(\boldsymbol{\varphi}^{(8)}\right) : \boldsymbol{\varepsilon}^e \tag{16}$$

where $\mathbf{L}(\boldsymbol{\varphi}^{(8)})$ is a fourth-rank tensorial function of the damage tensor $\boldsymbol{\varphi}^{(8)}$. It is noted that both the two hypotheses (elastic strain equivalence and elastic energy equivalence) are obtained as special cases of Eq. 16. By using $\mathbf{L}(\boldsymbol{\varphi}^{(8)}) = \mathbf{I}^{(4)}$, the hypothesis of elastic strain equivalence is obtained, and by using $\mathbf{L}(\boldsymbol{\varphi}^{(8)}) = \mathbf{M}^{-T}$, the hypothesis of elastic energy equivalence is obtained, where the fourth-rank tensor $\mathbf{M}$ is the damage effect tensor as used by Voyiadjis and Kattan (1999).

Equation 15 may be postulated even in the absence of the concept of the effective stress space as a relation that evolves the process of degradation of the elastic stiffness. It may be compared in form to Eqs. 10, 13, and 15. In the absence of the presence of an effective stress space, Eq. 16 is nonexisting and may be interpreted as an identity relation.

The elastic constitutive relation is written in the actual damage state as follows:

$$\boldsymbol{\sigma} = \mathbf{E} : \boldsymbol{\varepsilon}^e \tag{17}$$

A similar elastic constitutive relation in the fictitious state can now be written as follows:

$$\overline{\boldsymbol{\sigma}} = \overline{\mathbf{E}} : \overline{\boldsymbol{\varepsilon}}^e \tag{18}$$

Substituting Eq. 16 into Eq. 18, one obtains

$$\overline{\boldsymbol{\sigma}} = \overline{\mathbf{E}} : \mathbf{L}\left(\boldsymbol{\varphi}^{(8)}\right) : \boldsymbol{\varepsilon}^e \tag{19}$$

Next, Eq. 15 is substituted into Eq. 17 to obtain

$$\boldsymbol{\sigma} = \left(\mathbf{I}^{(8)} - \boldsymbol{\varphi}^{(8)}\right) :: \overline{\mathbf{E}} : \boldsymbol{\varepsilon}^{e} \tag{20}$$

Solving Eq. 19 for ε^e and substituting the result into Eq. 20, one obtains

$$\boldsymbol{\sigma} = \left(\mathbf{I}^{(8)} - \boldsymbol{\varphi}^{(8)}\right) :: \overline{\mathbf{E}} : \left(\mathbf{L}^{-1}\left(\boldsymbol{\varphi}^{(8)}\right) : \overline{\mathbf{E}}^{-1} : \overline{\boldsymbol{\sigma}}\right) \tag{21}$$

Equation 21 above can be rewritten in the following simpler form (note that only fourth-rank tensors are used if adopting the hypothesis of elastic strain equivalence):

$$\sigma = (\mathbf{I}_4 - \varphi_4) : \overline{\mathbf{E}} : \mathbf{L}^{-1}\left(\varphi^{(4)}\right) : \overline{\mathbf{E}}^{-1} : \overline{\sigma} \tag{22}$$

where $\mathbf{I}^{(4)}$ is the fourth-rank identity tensor and $\varphi^{(4)}$ is the fourth-rank damage tensor. In deriving Eq. 22, the following relation is used:

$$\mathbf{I}^{(4)} - \boldsymbol{\varphi}^{(4)} = \left(\left(\mathbf{I}^{(8)} - \boldsymbol{\varphi}^{(8)}\right) :: \overline{\mathbf{E}}\right) : \mathbf{L}^{-1}\left(\boldsymbol{\varphi}^{(8)}\right) : \mathbf{L}\left(\boldsymbol{\varphi}^{(4)}\right) : \overline{\mathbf{E}}^{-1} \tag{23}$$

It can be shown that using Eq. 23, one obtains the following relation between the two damage tensors:

$$\boldsymbol{\varphi}^{(4)} = \left(\boldsymbol{\varphi}^{(8)} :: \overline{\mathbf{E}}\right) : \mathbf{L}^{-1}\left(\boldsymbol{\varphi}^{(8)}\right) : \mathbf{L}\left(\boldsymbol{\varphi}^{(4)}\right) : \overline{\mathbf{E}}^{-1} \tag{24}$$

where the identity tensors $\mathbf{I}^{(4)}$ and $\mathbf{I}^{(8)}$ are given by

$$\mathbf{I}^{(4)}_{ijkl} = \frac{1}{2}\left(\boldsymbol{\delta}_{ik}\boldsymbol{\delta}_{jl} + \boldsymbol{\delta}_{il}\boldsymbol{\delta}_{jk}\right) \tag{25}$$

$$\mathbf{I}^{(8)}_{ijklmnpq} = \frac{1}{4}\left(\boldsymbol{\delta}_{im}\boldsymbol{\delta}_{jn}\boldsymbol{\delta}_{kp}\boldsymbol{\delta}_{lq} + \boldsymbol{\delta}_{im}\boldsymbol{\delta}_{jn}\boldsymbol{\delta}_{kq}\boldsymbol{\delta}_{lp} + \boldsymbol{\delta}_{in}\boldsymbol{\delta}_{jm}\boldsymbol{\delta}_{kp}\boldsymbol{\delta}_{lq} + \boldsymbol{\delta}_{in}\boldsymbol{\delta}_{jm}\boldsymbol{\delta}_{kq}\boldsymbol{\delta}_{lp}\right) \tag{26}$$

Next, Eq. 15 is expanded as follows:

$$\mathbf{E} = \overline{\mathbf{E}} - \boldsymbol{\varphi}^{(8)} :: \overline{\mathbf{E}} \tag{27}$$

Post-multiplying Eq. 24 by $\overline{\mathbf{E}}$, one obtains

$$\boldsymbol{\varphi}^{(4)} : \overline{\mathbf{E}} : \mathbf{L}^{-1}\left(\boldsymbol{\varphi}^{(4)}\right) : \mathbf{L}\left(\boldsymbol{\varphi}^{(8)}\right) = \boldsymbol{\varphi}^{(8)} :: \overline{\mathbf{E}} \tag{28}$$

Substituting Eq. 28 into Eq. 27 and simplifying, one obtains the desired relation as follows:

$$\mathbf{E} = \overline{\mathbf{E}} - \varphi^{(4)} : \overline{\mathbf{E}} : \mathbf{L}^{-1}\left(\varphi^{(4)}\right) : \mathbf{L}\left(\varphi^{(8)}\right) \tag{29}$$

It has now been shown that using the general hypothesis of strain transformation of Eq. 16, Eq. 15 (which involves an eighth-rank damage tensor) was reduced to Eq. 29 (which involves a fourth-rank damage tensor and an eighth-rank damage tensor).

Cauvin and Testa (1999) have shown that for the case of orthotropic damage, the fourth-rank damage tensor $\boldsymbol{\varphi}^{(4)}$ can be represented by the following 6 x 6 matrix:

$$\boldsymbol{\varphi}^{(4)} = \begin{bmatrix} \boldsymbol{\varphi}_{1111} & \boldsymbol{\varphi}_{1122} & \boldsymbol{\varphi}_{1133} & 0 & 0 & 0 \\ \boldsymbol{\varphi}_{2211} & \boldsymbol{\varphi}_{2222} & \boldsymbol{\varphi}_{2233} & 0 & 0 & 0 \\ \boldsymbol{\varphi}_{3311} & \boldsymbol{\varphi}_{3322} & \boldsymbol{\varphi}_{3333} & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\boldsymbol{\varphi}_{2323} & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\boldsymbol{\varphi}_{1313} & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\boldsymbol{\varphi}_{1212} \end{bmatrix} \tag{30}$$

where it is clear that $\boldsymbol{\varphi}^{(4)}$ has twelve independent components. In writing the matrix representation in Eq. 30, it is assumed that the stress and strain tensors can be represented as 6×1 column matrices as follows:

$$\boldsymbol{\sigma} = \begin{bmatrix} \boldsymbol{\sigma}_{11} & \boldsymbol{\sigma}_{22} & \boldsymbol{\sigma}_{33} & \boldsymbol{\sigma}_{23} & \boldsymbol{\sigma}_{13} & \boldsymbol{\sigma}_{12} \end{bmatrix}^T \tag{31.1}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_{11} & \boldsymbol{\varepsilon}_{22} & \boldsymbol{\varepsilon}_{33} & \boldsymbol{\varepsilon}_{23} & \boldsymbol{\varepsilon}_{13} & \boldsymbol{\varepsilon}_{12} \end{bmatrix}^T \tag{31.2}$$

The Damage Tensor and Fabric Tensors

In this section, an explicit expression is derived for the damage tensor in terms of the fabric tensors. The expression to be derived will provide a link between damage mechanics and fabric tensors. It will provide the theory of damage mechanics with a solid physical basis that directly depends on the microstructure.

In the remaining part of this section, one goes back to the general case of anisotropy and Eq. 7. Looking at Eqs. 10 and 30, it can be realized that they both describe the same quantity. Equation 29 describes the elasticity tensor for the damaged material in terms of the damage tensor. On the other hand, Eq. 9 describes the same elasticity tensor in terms of the fabric tensors. Therefore, equating the two equations yields the following:

$$\begin{aligned} \overline{\mathbf{E}} - \varphi^{(4)} : \overline{\mathbf{E}} : \mathbf{L}^{-1}\left(\varphi^{(4)}\right) : \mathbf{L}\left(\varphi^{(8)}\right) = {} & \lambda \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \otimes \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \\ & + 2\mu \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \bar{\otimes} \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \end{aligned} \tag{32}$$

Solving the above equation for $\boldsymbol{\varphi}^{(4)}$, one obtains the following expression:

$$\begin{aligned} \varphi^{(4)} = {} & \left[\overline{\mathbf{E}} - \lambda \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \otimes \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \right. \\ & \left. + 2\mu \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \bar{\otimes} \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right)\right] : \mathbf{L}^{-1}\left(\varphi^{(8)}\right) : \mathbf{L}\left(\varphi^{(4)}\right) : \overline{\mathbf{E}}^{-1} \end{aligned} \tag{33.1}$$

Equation 33.1 represents an explicit expression for the fourth-rank damage tensor $\boldsymbol{\varphi}^{(4)}$ in terms of the zero-rank fabric tensor (scalar) $G^{(0)}$ and the

second-rank fabric tensor $\mathbf{G}^{(2)}$. The other elements appearing in this expression are all constant scalars like λ and μ or constant tensors like $\mathbf{I}^{(2)}$, $\mathbf{I}^{(4)}$, and $\overline{\mathbf{E}}$. The fourth-rank tensorial function $\mathbf{L}$ must be substituted for in terms of other parameters.

For the special case of the hypothesis of elastic strain equivalence, one sets $\mathbf{L}(\boldsymbol{\varphi}^{(8)}) = \mathbf{L}(\boldsymbol{\varphi}^{(4)}) = \mathbf{I}^{(4)}$. In this case, Eq. 33.1 reduces to the following simpler form:

$$\begin{aligned}\boldsymbol{\varphi}^{(4)} = \mathbf{I}^{(4)} - \Big[&\lambda\left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \otimes \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \\ &+ 2\boldsymbol{\mu}\left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \underline{\bar{\otimes}} \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right)\Big] : \overline{\mathbf{E}}^{-1}\end{aligned} \tag{33.2}$$

For the other special case of the hypothesis of elastic energy equivalence, one sets $\mathbf{L}(\boldsymbol{\varphi}^{(8)}) = \mathbf{M}^{-T}(\boldsymbol{\varphi}^{(8)})$ and $\mathbf{L}(\boldsymbol{\varphi}^{(4)}) = \mathbf{M}^{-T}(\boldsymbol{\varphi}^{(4)})$. In this case, Eq. 33.1 reduces to the following form:

$$\begin{aligned}\varphi^{(4)} = \Big[&\overline{\mathbf{E}} - \lambda\left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \otimes \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \\ &+ 2\mu\left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right) \underline{\bar{\otimes}} \left(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\right)\Big] : \mathbf{M}^{T}\left(\varphi^{(8)}\right) : \mathbf{M}^{-T}\left(\varphi^{(4)}\right) : \overline{\mathbf{E}}^{-1}\end{aligned} \tag{33.3}$$

where $\mathbf{M}$ is the fourth-rank damage effect tensor as used by Voyiadjis and Kattan (1999).

For the remaining part of this work, the simpler formula of Eq. 33.2 is adopted and used for the special case of elastic strain equivalence. This simple equation is selected to be used in the derivation of the four cases to be studied in the subsequent sections and in the numerical application involving micro-crack distributions. Therefore, Eq. 33.2 may be rewritten in indicial notation as follows:

$$\begin{aligned}\boldsymbol{\varphi}^{(4)}_{ijkl} = \frac{1}{2}\left(\boldsymbol{\delta}_{ik}\boldsymbol{\delta}_{jl} + \boldsymbol{\delta}_{il}\boldsymbol{\delta}_{jk}\right) - \Big[&\boldsymbol{\lambda}\left(G^{(0)}\boldsymbol{\delta}_{ij} + G^{(2)}_{ij}\right)\left(G^{(0)}\boldsymbol{\delta}_{mn} + G^{(2)}_{mn}\right) \\ &+ \boldsymbol{\mu}\left(G^{(0)}\boldsymbol{\delta}_{im} + G^{(2)}_{im}\right)\left(G^{(0)}\boldsymbol{\delta}_{jn} + G^{(2)}_{jn}\right) + \boldsymbol{\mu}\left(G^{(0)}\boldsymbol{\delta}_{in} + G^{(2)}_{in}\right)\left(G^{(0)}\boldsymbol{\delta}_{jm} + G^{(2)}_{jm}\right)\Big]\overline{E}^{-1}_{mnkl}\end{aligned} \tag{34}$$

Equation 34 can be expanded to obtain the following explicit expression:

$$\begin{aligned}\boldsymbol{\varphi}^{(4)}_{ijkl} = \frac{1}{2}\left(\boldsymbol{\delta}_{ik}\boldsymbol{\delta}_{jl} + \boldsymbol{\delta}_{il}\boldsymbol{\delta}_{jk}\right) - \lambda\left(G^{(0)^2}\boldsymbol{\delta}_{ij}\boldsymbol{\delta}_{mn} + G^{(0)}\boldsymbol{\delta}_{ij}G^{(2)}_{mn} + G^{(0)}\boldsymbol{\delta}_{mn}G^{(2)}_{ij} + G^{(2)}_{ij}G^{(2)}_{mn}\right)\overline{E}^{-1}_{mnkl} \cdot \\ - \boldsymbol{\mu}\left(G^{(0)^2}\boldsymbol{\delta}_{im}\boldsymbol{\delta}_{jn} + G^{(0)}\boldsymbol{\delta}_{im}G^{(2)}_{jn} + G^{(0)}\boldsymbol{\delta}_{jn}G^{(2)}_{im} + G^{(2)}_{im}G^{(2)}_{jn}\right)\overline{E}^{-1}_{mnkl} \\ - \boldsymbol{\mu}\left(G^{(0)^2}\boldsymbol{\delta}_{in}\boldsymbol{\delta}_{jm} + G^{(0)}\boldsymbol{\delta}_{in}G^{(2)}_{jm} + G^{(0)}\boldsymbol{\delta}_{jm}G^{(2)}_{in} + G^{(2)}_{in}G^{(2)}_{jm}\right)\overline{E}^{-1}_{mnkl}\end{aligned} \tag{35}$$

Alternatively, one may use Eq. 12 instead of Eq. 9. Thus, equating Eqs. 13 and 30 and solving for $\boldsymbol{\varphi}^{(4)}$, one obtains

$$\boldsymbol{\varphi}^{(4)} = \mathbf{I}^{(4)} - \Big[(\lambda + 2\boldsymbol{\mu})\, m_i^{2k}\big(\mathbf{G}_i \otimes \mathbf{G}_i\big) + \lambda\, m_i^k m_j^k\big(\mathbf{G}_i \otimes \mathbf{G}_j + \mathbf{G}_j \otimes \mathbf{G}_i\big) + 2\boldsymbol{\mu}\, m_i^k m_j^k\left(\mathbf{G}_i \underline{\otimes} \mathbf{G}_j + \mathbf{G}_j \underline{\otimes} \mathbf{G}_i\right)\Big] : \mathbf{L}^{-1}\big(\boldsymbol{\varphi}^{(8)}\big) : \mathbf{L}\big(\boldsymbol{\varphi}^{(4)}\big) : \overline{\mathbf{E}}^{-1} \tag{36}$$

Equation 36 provides an alternative expression for the fourth-rank damage tensor $\boldsymbol{\varphi}^{(4)}$ in terms of the fabric tensors. It should be noted that in this alternative expression, the fabric tensors do not appear explicitly. However, the variables $m_i (i = 1, 2, 3)$ and $\mathbf{G}_i (i = 1, 2, 3)$ are obtained directly from the fabric tensors through the use of Eqs. 12 and 14 where $\mathbf{G}_i$ is defined as the cross product of the eigenvectors $\mathbf{g}_i$ of the second-rank fabric tensor $\mathbf{G}^{(2)}$, while m_i is defined as the sum of the zero-rank fabric tensor (scalar) $G^{(0)}$ and the eigenvalue g_i of the second-rank fabric tensor $\mathbf{G}^{(2)}$.

Equation 36 can be rewritten in indicial notation for the special case of the hypothesis of elastic strain equivalence as follows:

$$\boldsymbol{\varphi}^{(4)}_{ijkl} = \frac{1}{2}\big(\boldsymbol{\delta}_{ik}\boldsymbol{\delta}_{jl} + \boldsymbol{\delta}_{il}\boldsymbol{\delta}_{jk}\big) - (\lambda + 2\boldsymbol{\mu}) m_i^{2k} G_{i_{ij}} G_{i_{mn}} \overline{E}^{-1}_{mnkl} - \lambda\, m_i^k m_j^k \big(G_{i_{ij}} G_{j_{mn}} + G_{j_{ij}} G_{i_{mn}}\big)\overline{E}^{-1}_{mnkl}. - \boldsymbol{\mu}\, m_i^k m_j^k \Big(G_{i_{im}} G_{j_{jn}} + G_{i_{in}} G_{j_{jm}} + G_{j_{im}} G_{i_{jn}} + G_{j_{in}} G_{i_{jm}}\Big)\overline{E}^{-1}_{mnkl} \tag{37}$$

Next, the 6×6 matrix representations of the various tensors involved are written in detail using the notation adopted in Eqs. 31 and 32. Eventually, one will derive explicit expressions for the damage tensor components $\boldsymbol{\varphi}_{ijkl}$ in terms of the other variables based on the fabric tensors. This is performed in the remaining part of this work for the special case of the hypothesis of elastic strain equivalence.

The general 6×6 matrix representation of the fourth-rank elasticity tensor of the damaged material is given as follows based on Eq. 12 – see Zysset and Curnier (1995):

$$\mathbf{E} = \begin{bmatrix} (\lambda + 2\mu) m_1^{2k} & \lambda m_1^k m_2^k & \lambda m_1^k m_3^k & 0 & 0 & 0 \\ \lambda m_2^k m_1^k & (\lambda + 2\mu) m_2^{2k} & \lambda m_2^k m_3^k & 0 & 0 & 0 \\ \lambda m_3^k m_1^k & \lambda m_3^k m_2^k & (\lambda + 2\mu) m_3^{2k} & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\mu m_2^k m_3^k & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\mu m_3^k m_1^k & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\mu m_1^k m_2^k \end{bmatrix} \tag{38}$$

where λ and μ are Lame's constants, k is a constant scalar parameter with a value less than zero, and $m_i (i = 1, 2, 3)$ are related to the fabric tensors as given by Eq. 13. Thus, Eq. 38 is a matrix representation of the elasticity tensor $\mathbf{E}$ of damaged materials in terms of the fabric tensors. Next, the inverse elasticity tensor $\overline{\mathbf{E}}^{-1}$ (also called the compliance tensor) of the virgin material as a 6×6 matrix is written as follows:

$$\overline{\mathbf{E}}^{-1} = \begin{bmatrix} \frac{1}{E_1} & -\frac{\nu_{12}}{E_1} & -\frac{\nu_{13}}{E_1} & 0 & 0 & 0 \\ -\frac{\nu_{12}}{E_1} & \frac{1}{E_2} & -\frac{\nu_{23}}{E_2} & 0 & 0 & 0 \\ -\frac{\nu_{13}}{E_1} & -\frac{\nu_{23}}{E_2} & \frac{1}{E_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2G_{23}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2G_{31}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2G_{12}} \end{bmatrix} \tag{39}$$

where E_1, E_2, E_3, ν_{12}, ν_{13}, ν_{23}, G_{12}, G_{23}, and G_{31} are the nine independent material constants of orthotropic elasticity. Then, the 6×6 matrix representation of the fourth-rank identity tensor $\mathbf{I}^{(4)}$ is written as follows:

$$\mathbf{I}^{(4)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{40}$$

Solving Eq. 29 for $\boldsymbol{\varphi}^{(4)}$ while using $\mathbf{L} = \mathbf{I}^{(4)}$, one obtains the following expression:

$$\boldsymbol{\varphi}^{(4)} = \mathbf{I}^{(4)} - \mathbf{E} : \overline{\mathbf{E}}^{-1} \tag{41}$$

Substituting the matrix representations of Eqs. 39, 40, and 41 into Eq. 41, one obtains the 6×6 matrix representation of the fourth-rank damage tensor $\boldsymbol{\varphi}^{(4)}$ for the general case of orthotropic damage. Comparing the matrix obtained with the matrix in Eq. 30, one obtains the following explicit expressions for the damage tensor components $\boldsymbol{\varphi}_{ijkl}$ in terms of λ, μ, k, and $m_i (i = 1, 2, 3)$ where $m_i (i = 1, 2, 3)$ represents the fabric tensors:

$$\boldsymbol{\varphi}_{1111} = 1 - \frac{(\lambda + 2\mu) m_1^{2k}}{E_1} + \frac{\lambda m_1^k m_2^k \nu_{21}}{E_2} + \frac{\lambda m_1^k m_3^k \nu_{31}}{E_3} \tag{42.1}$$

$$\boldsymbol{\varphi}_{1122} = \frac{(\lambda + 2\mu) m_1^{2k} \nu_{12}}{E_1} - \frac{\lambda m_1^k m_2^k}{E_2} + \frac{\lambda m_1^k m_3^k \nu_{31}}{E_3} \tag{42.2}$$

$$\boldsymbol{\varphi}_{1133} = \frac{(\lambda + 2\mu) m_1^{2k} \nu_{13}}{E_1} + \frac{\lambda m_1^k m_2^k \nu_{23}}{E_2} - \frac{\lambda m_1^k m_3^k}{E_3} \tag{42.3}$$

$$\boldsymbol{\varphi}_{2211} = -\frac{\lambda m_2^k m_1^k}{E_1} + \frac{(\lambda + 2\mu) m_2^{2k} \nu_{21}}{E_2} + \frac{\lambda m_2^k m_3^k \nu_{31}}{E_3} \tag{42.4}$$

$$\boldsymbol{\varphi}_{2222} = 1 + \frac{\lambda m_2^k m_1^k \nu_{12}}{E_1} - \frac{(\lambda + 2\mu) m_2^{2k}}{E_2} + \frac{\lambda m_2^k m_3^k \nu_{32}}{E_3} \tag{42.5}$$

$$\boldsymbol{\varphi}_{2233} = \frac{\lambda m_2^k m_1^k \nu_{13}}{E_1} + \frac{(\lambda + 2\mu) m_2^{2k} \nu_{23}}{E_2} - \frac{\lambda m_2^k m_3^k}{E_3} \tag{42.6}$$

$$\boldsymbol{\varphi}_{3311} = -\frac{\lambda m_3^k m_1^k}{E_1} + \frac{\lambda m_3^k m_2^k \nu_{21}}{E_2} + \frac{(\lambda + 2\mu) m_3^{2k} \nu_{31}}{E_3} \tag{42.7}$$

$$\boldsymbol{\varphi}_{3322} = \frac{\lambda m_3^k m_1^k \nu_{12}}{E_1} - \frac{\lambda m_3^k m_2^k}{E_2} + \frac{(\lambda + 2\mu) m_3^{2k} \nu_{32}}{E_3} \tag{42.8}$$

$$\boldsymbol{\varphi}_{3333} = 1 + \frac{\lambda m_3^k m_1^k \nu_{13}}{E_1} + \frac{\lambda m_3^k m_2^k \nu_{23}}{E_2} - \frac{(\lambda + 2\mu) m_3^{2k}}{E_3} \tag{42.9}$$

$$\boldsymbol{\varphi}_{2323} = \frac{\mu m_2^k m_3^k}{2G_{23}} \tag{42.10}$$

$$\boldsymbol{\varphi}_{3131} = \frac{\mu m_3^k m_1^k}{2G_{31}} \tag{42.11}$$

$$\boldsymbol{\varphi}_{1212} = \frac{\mu m_1^k m_2^k}{2G_{12}} \tag{42.12}$$

It is clear from the expressions of the damage tensor components $\boldsymbol{\varphi}_{ijkl}$ of Eqs. 42 that the damage tensor $\boldsymbol{\varphi}^{(4)}$ is not symmetric.

In the next section, the special case of plane stress is considered and the damage tensor and fabric tensor equations are illustrated for this case.

Case of Plane Stress

The case of plane stress in the $x_1 - x_2$ plane is considered here. In this case the three stress components $\boldsymbol{\sigma}_{33}$, $\boldsymbol{\sigma}_{13}$, and $\boldsymbol{\sigma}_{23}$ vanish, i.e., $\boldsymbol{\sigma}_{33} = \boldsymbol{\sigma}_{13} = \boldsymbol{\sigma}_{23} = 0$. Therefore, the stress and strain tensors can be represented in this case by the following 3×1 column matrices:

$$\boldsymbol{\sigma} = \left[\, \boldsymbol{\sigma}_{11} \quad \boldsymbol{\sigma}_{22} \quad \boldsymbol{\sigma}_{12} \,\right]^T \tag{43.1}$$

$$\boldsymbol{\varepsilon} = \left[\, \boldsymbol{\varepsilon}_{11} \quad \boldsymbol{\varepsilon}_{22} \quad \boldsymbol{\varepsilon}_{12} \,\right]^T \tag{43.2}$$

It should be noted that in this case, the out-of-plane strain component $\boldsymbol{\varepsilon}_{33}$ does not vanish, i.e., $\boldsymbol{\varepsilon}_{33} \neq 0$. In this case, the damage state may be described by a fourth-rank tensor $\varphi^{(4)}$ which is represented by the following general 3×3 matrix:

$$\varphi^{(4)} = \begin{bmatrix} \varphi_{1111} & \varphi_{1212} & \varphi_{1313} \\ \varphi_{2121} & \varphi_{2222} & \varphi_{2323} \\ \varphi_{3131} & \varphi_{3232} & \varphi_{3333} \end{bmatrix} \tag{44}$$

In this case, Eq. 29 may now be rewritten in matrix form as follows:

$$[\mathbf{E}] = \left(\left[\mathbf{I}^{(4)} \right] - \left[\varphi^{(4)} \right] \right) [\overline{\mathbf{E}}] \tag{45}$$

where $\mathbf{I}^{(4)}$ is the fourth-rank identity tensor represented by the following 3×3 identity matrix:

$$\mathbf{I}^{(4)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{46}$$

The elasticity tensor $\overline{\mathbf{E}}$ of the virgin material may be represented as follows for the case of plane stress:

$$\overline{\mathbf{E}} = \frac{E}{1-\nu^2} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \dfrac{1-\nu}{2} \end{bmatrix} \tag{47}$$

where E and ν are the modulus of elasticity and Poisson's ratio of the virgin material, respectively. The relations between E, ν and Lame's constants λ, μ are given by the following two equations:

$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)} \tag{48.1}$$

$$\mu = \frac{E}{2(1+\nu)} \tag{48.2}$$

In this case, the matrix representation of the elasticity tensor $\mathbf{E}$ of the damaged material may be written as follows (see Eq. 38):

$$\mathbf{E} = \frac{E}{1-\nu^2} \begin{bmatrix} m_1^{2k} & \nu m_1^k m_2^k & 0 \\ \nu m_2^k m_1^k & m_2^{2k} & 0 \\ 0 & 0 & \dfrac{1-\nu}{2} m_1^k m_2^k \end{bmatrix} \tag{49}$$

Next, Eqs. 45, 47, and 48 are substituted into Eq. 45 and the resulting equation is then simplified. Comparing the resulting matrix with the matrix in Eq. 49, one

obtains the following nine linear simultaneous algebraic equations in the damage tensor components φ_{ijkl}:

$$1 - \varphi_{1111} - \nu\varphi_{1212} = m_1^{2k} \tag{50.1}$$

$$\nu - \nu\varphi_{1111} - \varphi_{1212} = \nu m_1^k m_2^k \tag{50.2}$$

$$\varphi_{1313} = 0 \tag{50.3}$$

$$\nu - \nu\varphi_{2222} - \varphi_{2121} = \nu m_2^k m_1^k \tag{50.4}$$

$$-\nu\varphi_{2121} + 1 - \varphi_{2222} = m_2^{2k} \tag{50.5}$$

$$\varphi_{2323} = 0 \tag{50.6}$$

$$\varphi_{3131} + \nu\varphi_{3232} = 0 \tag{50.7}$$

$$\nu\varphi_{3131} + \varphi_{3232} = 0 \tag{50.8}$$

$$1 - \varphi_{3333} = m_1^k m_2^k \tag{50.9}$$

It is possible now to immediately use Eqs. 50.3, 50.6, 50.7, and 50.8 to conclude that $\varphi_{1313} = \varphi_{2323} = \varphi_{3131} = \varphi_{3232} = 0$. Therefore, four of the damage tensor components φ_{ijkl} vanish in the case of plane stress. This leaves the following system of five linear simultaneous algebraic equations:

$$1 - \varphi_{1111} - \nu\varphi_{1212} = m_1^{2k} \tag{51.1}$$

$$\nu - \nu\varphi_{1111} - \varphi_{1212} = \nu m_1^k m_2^k \tag{51.2}$$

$$\nu - \nu\varphi_{2222} - \varphi_{2121} = \nu m_2^k m_1^k \tag{51.3}$$

$$-\nu\varphi_{2121} + 1 - \varphi_{2222} = m_2^{2k} \tag{51.4}$$

$$1 - \varphi_{3333} = m_1^k m_2^k \tag{51.5}$$

Next, one combines the two Eqs. 51.2 and 51.3 in order to obtain the following relation between φ_{1212} and φ_{2121}:

$$\varphi_{2121} = \varphi_{1212} - \nu(\varphi_{2222} - \varphi_{1111}) \tag{52}$$

Equation 52 clearly shows that the damage tensor $\varphi^{(4)}$ is not symmetric.

Equation 51.5 may be solved directly for φ_{3333} in order to obtain the following explicit expression:

$$\varphi_{3333} = 1 - m_1^k m_2^k \tag{53}$$

Equation 53 clearly indicates that the out-of-plane damage tensor component φ_{3333} does not vanish in the case of plane stress. This damage tensor component is

clearly given in Eq. 53 in terms of the fabric tensor parameters m_1 and m_2. This conclusion indicates that the case of plane stress does not imply a case of plane damage also.

The remaining four damage tensor components φ_{1111}, φ_{2222}, φ_{1212}, and φ_{2121} can be obtained by solving the remaining four implicit Eqs. 51.1, 51.2, 51.3, and 51.4 simultaneously to obtain (note that Eqs. 51.1 and 51.2 may be solved simultaneously while the other two Eqs. 51.3 and 51.4 may also be solved simultaneously):

$$\varphi_{1111} = 1 - \frac{m_1^k\left(m_1^k - \nu^2 m_2^k\right)}{1-\nu^2} \tag{54.1}$$

$$\varphi_{1212} = \frac{\nu m_1^k\left(m_1^k - m_2^k\right)}{1-\nu^2} \tag{54.2}$$

$$\varphi_{2222} = 1 - \frac{m_2^k\left(m_2^k - \nu^2 m_1^k\right)}{1-\nu^2} \tag{54.3}$$

$$\varphi_{2121} = \frac{\nu m_2^k\left(m_2^k - m_1^k\right)}{1-\nu^2} \tag{54.4}$$

Equations 54 clearly show explicit expressions for the four damage tensor components φ_{1111}, φ_{2222}, φ_{1212}, and φ_{2121} in terms of the fabric tensor parameters m_1 and m_2 and Poisson's ratio ν. It is clear that the relation between the damage tensor components and the fabric tensor parameters is independent of Young's modulus E of the material. Only the material constant ν plays a role in this relationship.

Application to Micro-crack Distributions

In this section, an application of the damage and fabric tensors to the case of micro-crack distributions is presented. This can be effectively illustrated by solving a practical example. Consider a two-dimensional (planar) micro-crack distribution whose circular histogram (rose diagram) is shown in Fig. 2. It is assumed that the micro-crack distribution is symmetric with respect to the origin. The circular histogram shows the distributions of the normals to the micro-cracks as a rose diagram. The angle is varied in 10° increments from 0° to 360° while the heights of the histogram represent the frequency of the normals to the micro-cracks that are oriented within the specified angle range.

Next, the components of the fabric tensors $G^{(0)}$, $\mathbf{G}^{(2)}$, and $\mathbf{G}^{(4)}$ for this specific example are calculated. It is noted that the fabric tensors $\mathbf{G}^{(2)}$ and $\mathbf{G}^{(4)}$ are taken to correspond to the fabric tensors of the third kind D_{ij} and D_{ijkl} introduced by

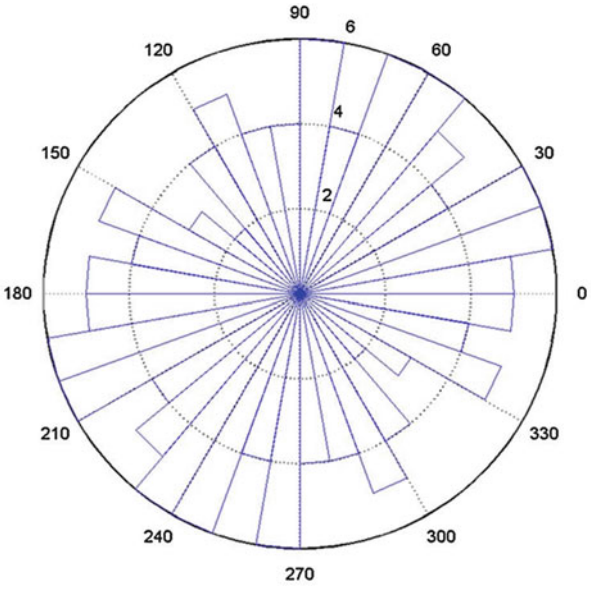


Fig. 2 A circular histogram (rose diagram) for the micro-crack distribution of the application (Reprinted with permission from G. Voyiadjis, P. Kattan, Mech. Mater. Struct. 13, 4 (2006))

Kanatani (1984a). The second-rank fabric tensor of the third kind is calculated as follows:

$$\mathbf{G}^{(2)} = \mathbf{D}^{(2)} = \begin{bmatrix} 1.2305 & 0.4065 & 0 \\ 0.4065 & 1.2695 & 0 \\ 0 & 0 & -2.500 \end{bmatrix} \tag{55}$$

Using the above matrix, the eigenvalues of the second-rank fabric tensor $\mathbf{G}^{(2)}$ are calculated as follows:

$$g_1 = 1.6570 \tag{56.1}$$

$$g_2 = 0.8430 \tag{56.2}$$

where obviously $g_3 = -2.5$. Next, the second-rank fabric tensor of the second kind $\mathbf{F}^{(2)}$ is calculated to obtain

$$\mathbf{F}^{(2)} = \begin{bmatrix} 2.2305 & 0.4065 & 0 \\ 0.4065 & 2.2695 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{57}$$

Using the second-rank fabric tensor of the second kind $\mathbf{F}^{(2)}$, the following approximation $\boldsymbol{\rho}(\mathbf{n})$ of the crack distribution (Kanatani 1984a) can be used:

$$\rho(\mathbf{n}) = F_{ij}^{(2)} n_i n_j \tag{58}$$

Substituting Eq. 57 into Eq. 58 and using the data from the circular histogram of Fig. 3, one obtains the following second-order approximation of the crack distribution:

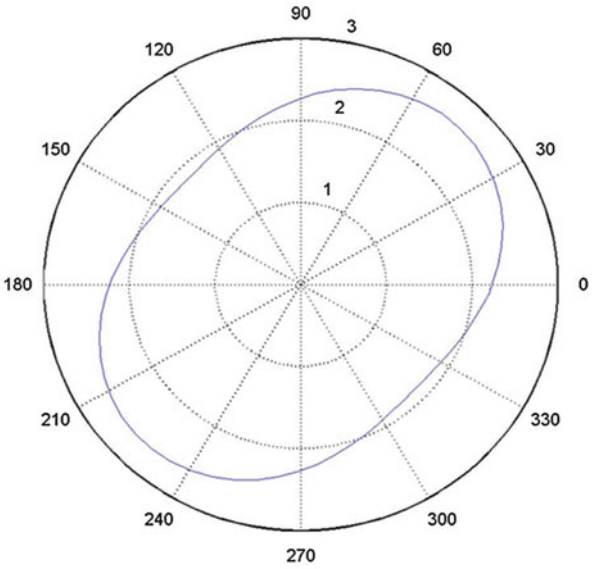


Fig. 3 A polar plot showing the second-order approximation to the micro-crack distribution data. This second-order approximation was obtained using second-rank fabric tensors (Reprinted with permission from G. Voyiadjis, P. Kattan, Mech. Mater. Struct. 13, 4 (2006))

$$\rho(\mathbf{n}) = 2.2305\mathbf{cos}^2\theta + 2.2695\mathbf{sin}^2\theta + 0.8130\mathbf{sin}\,\theta\mathbf{cos}\,\theta \tag{59}$$

where $0 \leq \boldsymbol{\theta} \leq 2\boldsymbol{\pi}$. The approximate distribution of Eq. 59 is now plotted as shown in Fig. 3. A comparison between Figs. 2 and 3 shows the close relationship between the actual distribution and the approximate distribution.

Next, the fourth-rank fabric tensor $\mathbf{N}^{(4)}$ is calculated in order to plot a fourth-order approximation to the micro-crack distribution. The fourth-rank fabric tensor $\mathbf{N}^{(4)}$ is calculated from the following formula (Kanatani 1984a):

$$N^{(4)}_{ijkl} = < n_i n_j n_k n_l > = \frac{1}{N}\sum_{\alpha=1}^{N} n_i^{(\alpha)} n_j^{(\alpha)} n_k^{(\alpha)} n_l^{(\alpha)} \tag{60}$$

The fourth-rank fabric tensor of the second kind $\mathbf{F}^{(4)}$ defined by Kanatani (1984a) can now be introduced as follows:

$$F^{(4)}_{ijkl} = \frac{315}{8}\left(N^{(4)}_{ijkl} - \frac{2}{3}\delta_{ij}N^{(2)}_{kl} + \frac{1}{21}\delta_{ij}\delta_{kl}\right) \tag{61}$$

Using Eq. 61, one obtains the values of the components of the fourth-rank fabric tensor of the second kind $\mathbf{F}^{(4)}$ as follows: $F^{(4)}_{1111} = 3.6698$, $F^{(4)}_{2222} = 3.7369$, $F^{(4)}_{1212} = 4.7368$, $F^{(4)}_{1222} = 1.0316$, and $F^{(4)}_{1112} = -0.3163$.

Using the fourth-rank fabric tensor of the second kind $\mathbf{F}^{(4)}$, one can use the following approximation $\boldsymbol{\rho}(\mathbf{n})$ of the crack distribution (Kanatani 1984a):

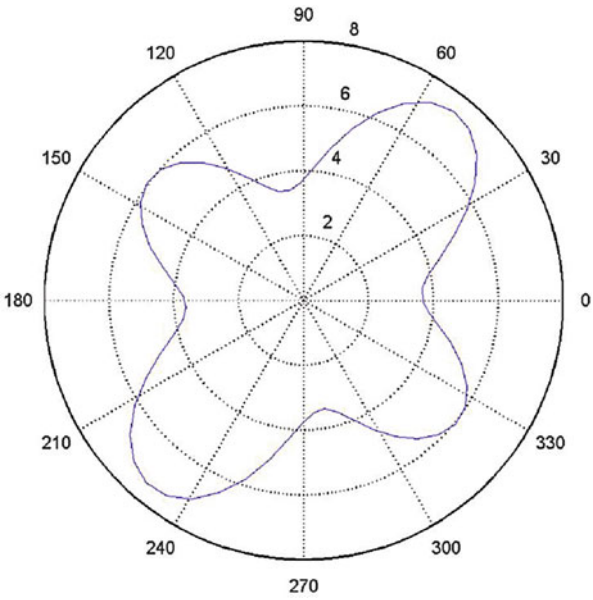


Fig. 4 A polar plot showing the fourth-order approximation to the micro-crack distribution data. This fourth-order approximation was obtained using fourth-rank fabric tensors (Reprinted with permission from G. Voyiadjis, P. Kattan, Mech. Mater. Struct. 13, 4 (2006))

$$\rho(\mathbf{n}) = F^{(4)}_{ijkl} n_i n_j n_k n_l \tag{62}$$

Substituting the values of the components of the fourth-rank fabric tensor of the second kind $\mathbf{F}^{(4)}$ into Eq. 62 and using the data from the circular histogram of Fig. 2, one obtains the following fourth-order approximation of the crack distribution:

$$\begin{aligned}\rho(\mathbf{n}) = {} & 3.6698\,\mathbf{cos}^4\theta + 3.7369\,\mathbf{sin}^4\theta + 18.9474\,\mathbf{sin}^2\theta\,\mathbf{cos}^2\theta \\ & + 4.1264\,\mathbf{cos}\,\theta \sin^3\theta - 1.2652\,\mathbf{cos}^3\theta\,\mathbf{sin}\,\theta\end{aligned} \tag{63}$$

where $0 \leq \boldsymbol{\theta} \leq 2\boldsymbol{\pi}$. The approximate distribution of Eq. 63 is now plotted as shown in Fig. 4. A comparison between Figs. 2, 3, and 4 shows the close relationship between the actual distribution and the approximate distribution.

Next, the precise relationship between the damage tensor components and the fabric tensors' components for this example is illustrated. In this example, one can use the simple expressions for the damage tensor components derived earlier for the case of plane stress, i.e., Eqs. 54 and 55. First the expression of φ_{1111} given in Eq. 54.1 is investigated. It is seen from this equation that the value of φ_{1111} depends on four parameters, namely, the fabric tensor parameters m_1 and m_2, Poisson's ratio ν, and the constant k. The values of the two constants are taken in this example as $\nu = 0.3$ and $k = -0.2$ (several values of k were investigated and the value used here was found to give realistic results). Furthermore, in order to simplify the resulting equation, one may find a relation between m_1 and m_2 for this particular example as follows. Using Eq. 13, the following relation is available:

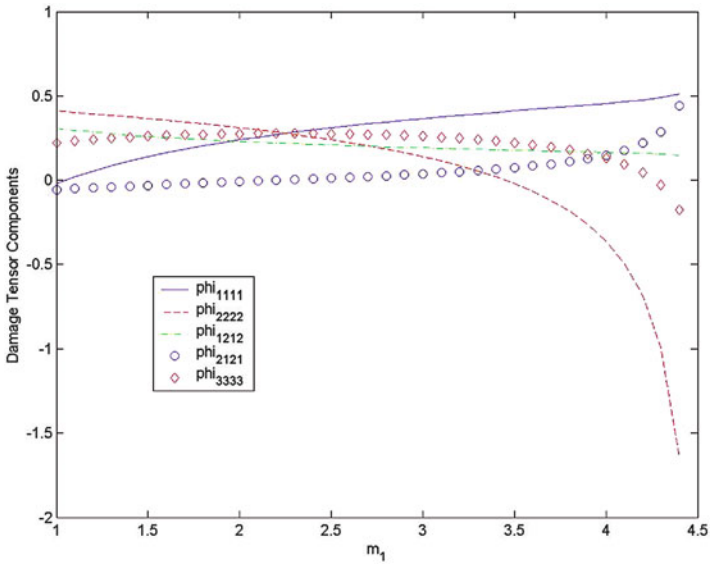


Fig. 5 Variation of damage tensor components versus m_1 (Reprinted with permission from G. Voyiadjis, P. Kattan, Mech. Mater. Struct. 13, 4 (2006))

$$m_1 + m_2 = 2G^{(0)} + g_1 + g_2 \tag{64}$$

However, one has $G^{(0)} = 1$. Substituting the values from Eqs. 56 and 57 into Eq. 64, one concludes that the sum of m_1 and m_2 is equal to 4.5 which is a constant in this example. Therefore, one obtains the following relation:

$$m_2 = 4.5 - m_1 \tag{65}$$

Substituting Eq. 65 and the values of ν and k given above into Eq. 54.1, one obtains the following explicit formula for φ_{1111}:

$$\varphi_{1111} = 1 - \frac{m_1^{-0.2}\left[m_1^{-0.2} - 0.09(4.5 - m_1)^{-0.2}\right]}{1 - 0.09} \tag{66}$$

Equation 66 above clearly shows that φ_{1111} is a function of one variable (m_1) for this example. This function is plotted in Fig. 5 to show clearly that φ_{1111} is a monotonically increasing positive function of m_1. In realistic problems, usually the values of m_1 are within the range $1.5 < m_1 < 3.5$, so one can see that in this region of the plot, the relation is almost linear and the values of φ_{1111} are between 0 and 0.5.

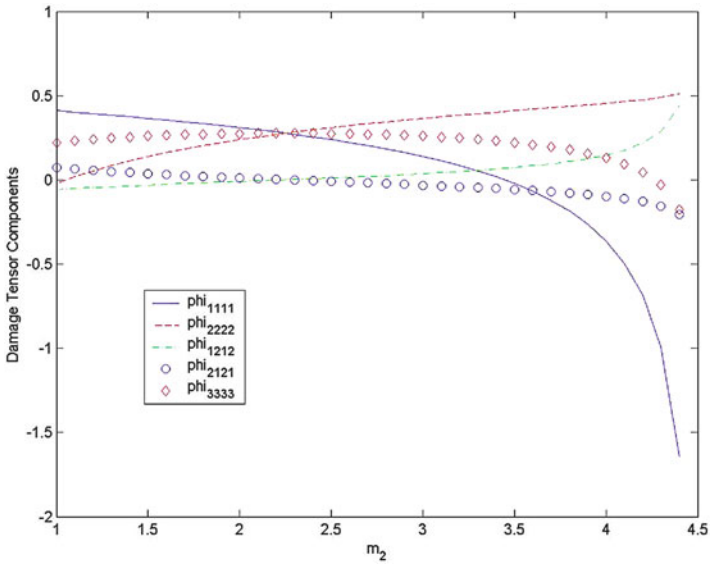


Fig. 6 Variation of damage tensor components versus m_2 (Reprinted with permission from G. Voyiadjis, P. Kattan, Mech. Mater. Struct. 13, 4 (2006))

Rewriting Eq. 66 in terms of m_2 instead of m_1, one obtains the following relation:

$$\varphi_{1111} = 1 - \frac{(4.5 - m_2)^{-0.2}\left[(4.5 - m_2)^{-0.2} - 0.09 m_2^{-0.2}\right]}{1 - 0.09} \tag{67}$$

The plot of Eq. 67 is shown in Fig. 6. It clearly shows that φ_{1111} is a monotonically decreasing function of m_2. Within the specified range $1.5 < m_2 < 3.5$ that is found in practical applications, it is seen that the values of φ_{1111} are positive and range between 0 and 0.5.

Equations 54.2, 54.3, and 54.4 are now rewritten for this example as follows. Each equation is rewritten twice – once in terms of m_1 and then in terms of m_2:

$$\varphi_{1212} = \frac{0.3 m_1^{-0.2}\left[m_1^{-0.2} - (4.5 - m_1)^{-0.2}\right]}{1 - 0.09} \tag{68.1}$$

$$\varphi_{1212} = \frac{0.3(4.5 - m_2)^{-0.2}\left[(4.5 - m_2)^{-0.2} - m_2^{-0.2}\right]}{1 - 0.09} \tag{68.2}$$

$$\varphi_{2222} = 1 - \frac{(4.5 - m_1)^{-0.2}\left[(4.5 - m_1)^{-0.2} - 0.09 m_1^{-0.2}\right]}{1 - 0.09} \tag{68.3}$$

$$\varphi_{2222} = 1 - \frac{m_2^{-0.2}\left[m_2^{-0.2} - 0.09(4.5 - m_2)^{-0.2}\right]}{1 - 0.09} \tag{68.4}$$

$$\varphi_{2121} = \frac{0.3(4.5 - m_1)^{-0.2}\left[(4.5 - m_1)^{-0.2} - m_1^{-0.2}\right]}{1 - 0.09} \tag{68.5}$$

$$\varphi_{2121} = \frac{0.3m_2^{-0.2}\left[m_2^{-0.2} - (4.5 - m_2)^{-0.2}\right]}{1 - 0.09} \tag{68.6}$$

The graphs of Eqs. 68.1, 68.2, 68.3, 68.4, 68.5, and 68.6 are also shown in Figs. 5 and 6. In these two figures, the same trends are observed. In the range $1.5 < m_1 < 3.5$, one obtains positive realistic results for the damage tensor components.

Similarly, Eq. 53 is also rewritten twice as follows:

$$\varphi_{3333} = 1 - m_1^{-0.2}(4.5 - m_1)^{-0.2} \tag{69.1}$$

$$\varphi_{3333} = 1 - (4.5 - m_2)^{-0.2}m_2^{-0.2} \tag{69.2}$$

Figures 5 and 6 show also the graphs of Eqs. 69.1 and 69.2, respectively. The same trends are observed here also in that realistic results are obtained in the range $1.5 < m_1 < 3.5$. In conclusion, it is noted that in this example, actual values for the damage tensor components based on the micro-crack distribution data given in the circular histogram are not possible to obtain.

Application to Parallel Micro-cracks

A second application is presented now to compare the elasticity matrix of two different sets of parallel micro-cracks. Consider the first set A of parallel micro-cracks to be oriented such that their normals are at an angle of $\theta = 0^\circ$ as shown in Fig. 7a. Let the second set B of parallel micro-cracks be oriented such that their normals are at an angle $\theta = 90^\circ$ as shown in Fig. 7b. One can calculate both the fabric tensors and damage tensors for these two sets of micro-cracks and conclude by calculating and comparing the damage elasticity matrices for these two orientations.

It should be pointed out that the number of these parallel micro-cracks is immaterial since an RVE is considered. One will obtain the same fabric tensors irrespective of the number of these parallel micro-cracks.

It should be noted that the number of micro-cracks N cancels out from the equations in order to be able to obtain the above matrices. Next, the second-rank fabric tensors of the third kind $\mathbf{G}^{(2)}$ are calculated for both sets as follows noting that $\mathbf{G}_A^{(0)} = \mathbf{G}_B^{(0)} = 1$:

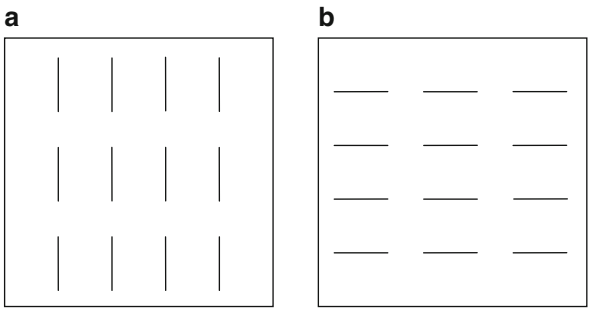


Fig. 7 Two sets of parallel micro-cracks: (**a**) set A with angle of 0, (**b**) set B with angle of 90 (Reprinted with permission from G. Voyiadjis, P. Kattan, Mech. Mater. Struct. 13, 4 (2006))

$$\mathbf{G}_A^{(2)} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & -2.5 & 0 \\ 0 & 0 & -2.5 \end{bmatrix} \tag{70.1}$$

$$\mathbf{G}_B^{(2)} = \begin{bmatrix} -2.5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -2.5 \end{bmatrix} \tag{70.2}$$

Next, the eigenvalues of $\mathbf{G}^{(2)}$ for each set are calculated using Eqs. 70 as follows:

$$g_1 = 5, \quad g_2 = -2.5 \quad \text{for set } A \tag{71.1}$$

$$g_1 = -2.5, \quad g_2 = 5 \quad \text{for set } B \tag{71.2}$$

Substituting Eq. 71 along with $\mathbf{G}_A^{(0)} = \mathbf{G}_B^{(0)} = 1$ into Eq. 13, one obtains the values for the fabric tensor parameters m_1 and m_2 for each set as follows:

$$m_1 = 6, \quad m_2 = -1.5 \quad \text{for set } A \tag{72.1}$$

$$m_1 = -1.5, \quad m_2 = 6 \quad \text{for set } B \tag{72.2}$$

Using the values of material parameters $\nu = 0.3$ and $k = -0.2$, substituting Eqs. 72 into Eqs. (55), one obtains the following principal values for the damage tensor $\varphi^{(4)}$ (special care must be taken when evaluating the negative value -1.5 raised to a negative exponent):

$$\varphi_A^{(4)} = \begin{bmatrix} 0.5181 & 0 & 0 \\ 0 & 0.1383 & 0 \\ 0 & 0 & 0.3556 \end{bmatrix} \tag{73.1}$$

$$\phi_B^{(4)} = \begin{bmatrix} 0.1383 & 0 & 0 \\ 0 & 0.5181 & 0 \\ 0 & 0 & 0.3556 \end{bmatrix} \tag{73.2}$$

It is clear from the above two matrices that the values of the damage variable are interchanged for the two cases. This makes sense since changing the orientation

from $\theta = 0^\circ$ to $\theta = 90^\circ$ clearly causes an interchange in the diagonal terms in the matrix representation of the damage tensor.

Finally, substituting Eqs. 71 along with $\mathbf{G}_A^{(0)} = \mathbf{G}_B^{(0)} = 1$ into Eqs. 28 and 34.3, the components E_{ijkl} of the damage elasticity tensor for each set are calculated and written in matrix form as follows:

$$\mathbf{E}_A = \begin{bmatrix} 36(\lambda + 2\mu) & -9\lambda & 0 \\ -9\lambda & 2.25(\lambda + 2\mu) & 0 \\ 0 & 0 & 2\mu \end{bmatrix} \tag{74.1}$$

$$\mathbf{E}_B = \begin{bmatrix} 2.25(\lambda + 2\mu) & -9\lambda & 0 \\ -9\lambda & 36(\lambda + 2\mu) & 0 \\ 0 & 0 & 2\mu \end{bmatrix} \tag{74.2}$$

Notice that the two elasticity matrices in Eqs. 74 are identical except that the first two diagonal terms are interchanged. This interchange effect results from changing the orientation of the parallel micro-cracks from $\theta = 0^\circ$ to $\theta = 90^\circ$. Since the two elasticity matrices are different for the two sets of distributions, one expects to obtain different stresses and strains in each case. Therefore, it is clear that the model presented here accounts for this clear anisotropy in the microstructure.

Thermodynamics and General Damage Evolution

In this section, thermal elastic damage material behavior is considered. A thermodynamic framework is presented for damage evolution. The dependent constitutive variables are function of the strain tensor $\boldsymbol{\varepsilon}$ (totally elastic), the absolute temperature T, the temperature gradient vector $\nabla_i T$, and a number of phenomenological state variables $N_k (k = 1, 2, 3)$. Hence, within the thermodynamic framework and considering the assumption of infinitesimal displacements/strain relationships, the Helmholtz free energy density function Ψ can be written as follows (Coleman and Gurtin 1967; Lemaitre and Chaboche 1990; Lubliner 1990; Doghri 2000; Voyiadjis and Kattan 1999):

$$\Psi = \widetilde{\Psi}\left(\varepsilon_{ij}, T, \nabla_i T, N_k\right) \tag{75}$$

In order to describe the various micro-damage mechanisms, a finite set of internal state variables N_k representing either a scalar or a tensorial variable are assumed such that

$$N_k = \widetilde{N}_k(\Xi_n) \tag{76}$$

where Ξ_n is a set of damage hardening internal state variables. This set of macro internal state variables Ξ_n is postulated as follows:

$$\Xi_n = \widetilde{\Xi}_n\left(r, \Gamma_{ijkl}, \varphi_{ijkl}\right) \tag{77}$$

where r denotes the accumulative damage, Γ denotes the flux of the residual stress in the damage growth process, and $\boldsymbol{\varphi}$ denotes the fourth-rank damage tensor. These damage hardening variables are introduced in the Helmholtz free energy density in order to provide sufficient details of the deformation defects (micro-cracks and micro-voids) and their interactions in order to properly (i.e., physically) characterize the material microstructural behavior. These variables will provide an adequate characterization of these defects in terms of size, orientation, distribution, spacing, interaction among defects, and so forth.

The determination of the evolution of the assumed internal state variables is the main challenge to modern constitutive modeling. This can be effectively achieved, so far, through the use of the thermodynamic principles for the development of a continuum thermoelastic damage-based model. That is, use is made of the balancing laws, the conservation of mass, linear and angular momentum, and the first and second laws of thermodynamics (Coleman and Gurtin 1967; Lemaitre and Chaboche 1990; Lubliner 1990; Doghri 2000; Voyiadjis and Kattan 1999).

The Clausius-Duhem inequality can be written for our case as follows:

$$\sigma_{ij}\dot{\varepsilon}_{ij} - \rho\left(\dot{\Psi} + \eta\dot{T}\right) - \frac{1}{T}q_i\nabla_i T \geq 0 \tag{78}$$

where ρ, η, and q are the mass density, specific entropy, and the heat flux vector, respectively. Meanwhile, Ψ, T, and η are related by

$$\Psi = e - T\eta \tag{79}$$

where e is the internal energy density. Next, it is assumed that the decomposition of the specific free energy density function Ψ into thermoelastic and thermo-damage parts is as follows:

$$\widetilde{\Psi}\left(\varepsilon_{ij}, T, \nabla_i T, N_k\right) = \widetilde{\Psi}^{te}\left(\varepsilon_{ij}, T, \nabla_i T, \varphi_{ijkl}\right) + \widetilde{\Psi}^{td}\left(T, \nabla_i T, r, \Gamma_{ij}, \varphi_{ijkl}\right) \tag{80}$$

where Ψ^{ts} is the thermoelastic stored energy while Ψ^{td} is the energy stored due to material hardening due to the damage mechanisms.

According to the definition given above for Ψ, the time derivative of Eq. 75 with respect to its internal state variables is given by

$$\dot{\Psi} = \frac{\partial\Psi}{\partial\varepsilon_{ij}}\dot{\varepsilon}_{ij} + \frac{\partial\Psi}{\partial T}\dot{T} + \frac{\partial\Psi}{\partial\nabla_i T}\nabla_i\dot{T} + \frac{\partial\Psi}{\partial N_k}\dot{N}_k \tag{81}$$

where (from Eq. 76)

$$\frac{\partial \Psi}{\partial N_k}\dot{N}_k = \frac{\partial \Psi}{\partial \Xi_n}\dot{\Xi}_n \tag{82}$$

with (from Eq. 77)

$$\frac{\partial \Psi}{\partial \Xi_n}\dot{\Xi}_n = \frac{\partial \Psi}{\partial r}\dot{r} + \frac{\partial \Psi}{\partial \Gamma_{ijkl}}\dot{\Gamma}_{ijkl} + \frac{\partial \Psi}{\partial \varphi_{ijkl}}\dot{\varphi}_{ijkl} \tag{83}$$

Substituting the rate of the Helmholtz free energy density (Eq. 81) into the Clausius-Duhem inequality (Eq. 78), one obtains the following thermodynamic constraint:

$$\left(\sigma_{ij} - \rho\frac{\partial \Psi}{\partial \varepsilon_{ij}}\right)\dot{\varepsilon}_{ij} - \rho\left(\frac{\partial \Psi}{\partial T} + \eta\right)\dot{T} - \rho\frac{\partial \Psi}{\partial \nabla_i T}\nabla_i \dot{T} - \rho\frac{\partial \Psi}{\partial N_k}\dot{N}_k - \frac{q_i}{T}\nabla T \geq 0 \tag{84}$$

Assuming that the axiom of entropy production holds, then the above inequality equation results in the following thermodynamic state laws:

$$\sigma_{ij} = \rho\frac{\partial \Psi}{\partial \varepsilon_{ij}} \tag{85.1}$$

$$\eta = -\frac{\partial \Psi}{\partial T} \tag{85.2}$$

$$\frac{q_i}{\dot{T}} = \rho\frac{\partial \Psi}{\partial \nabla_i T} \tag{85.3}$$

$$\sum\nolimits_k = \rho\frac{\partial \Psi}{\partial N_k} \quad (k = 1, 2, 3) \tag{85.4}$$

where the above equations describe the relations between the state variables and their associated thermodynamic conjugate forces. Note that the three thermodynamic conjugate forces $\sum_k$ denote the following three quantities: K which is associated with r, H_{ijkl} which is associated with Γ_{ijkl}, and Y_{ijkl} which is associated with φ_{ijkl}. The stress $\boldsymbol{\sigma}$ is a measure of the elastic changes in the internal structure, while $\mathbf{Y}$ is a measure of the elastic damage changes in the internal structure resulting from crack closure and void contraction during the unloading process. The conjugate forces K and $\mathbf{H}$ are measures of the damage changes in the internal structure of the material.

Substituting Eqs. 85 into Eq. 84, one reduces the Clausius-Duhem inequality to express the fact that the dissipation energy Π is necessarily positive:

$$\Pi = -\Pi_{\text{int}} - q_i\left(\frac{\frac{\nabla_i T}{T} + \nabla_i \dot{T}}{\dot{T}}\right) \geq 0 \tag{86}$$

where the internal dissipation energy Π_{int} can be written as follows:

$$\Pi_{\text{int}} = \sum_{k=1}^{3} \Sigma_k \dot{N}_k = K\dot{r} + H_{ijkl}\dot{\Gamma}_{ijkl} - Y_{ijkl}\dot{\varphi}_{ijkl} \geq 0 \tag{87}$$

One may rewrite the dissipation energy Π as the summation of dissipations due to damage and thermal effects as follows:

$$\Pi = \Pi^d + \Pi^{th} \tag{88}$$

where

$$\Pi^d = -K\dot{r} - H_{ijkl}\dot{\Gamma}_{ijkl} + Y_{ijkl}\dot{\varphi}_{ijkl} \geq 0 \tag{89}$$

$$\Pi^{th} = -q_i\left(\frac{\nabla_i T}{T} + \frac{\nabla_i \dot{T}}{\dot{T}}\right) \geq 0 \tag{90}$$

Complementary laws can be related to the dissipation processes given by Eqs. 89 and 90, which implies the existence of the dissipation potential expressed as a continuous and convex scalar valued function of the flux variables as shown below:

$$\Theta\left(\dot{N}_k, q_i/T\right) = \Theta^d\left(\dot{N}_k\right) + \Theta^{th}(T, \nabla_i T) \tag{91}$$

The complementary laws are then expressed by the normality property as follows:

$$\sum\nolimits_k = -\frac{\partial \Theta^d}{\partial \dot{N}_k} \tag{92}$$

$$\frac{q_i}{\dot{T}} = -\frac{\partial \Theta^{th}}{\partial(\nabla_i T)} \tag{93}$$

Using the Legendre-Fenchel transformation of the dissipation potential Θ, one can define the corresponding dual potential with respect to the force variables as follows:

$$\begin{aligned} \Theta^*\left(\sum\nolimits_k, \nabla_i T\right) &= \Pi\left(\sum\nolimits_k, q_i; \dot{N}_k, T, \nabla_i T\right) - \Theta\left(\dot{N}_k, T, \nabla_i T\right) \\ &= \Theta^{*d}\left(\sum\nolimits_k\right) + \Theta^{*th}\left(T, \nabla_i T\right) \end{aligned} \tag{94}$$

from which the complementary laws in the form of the evolution laws of the flux variables as a function of the dual variables can then be written as follows:

$$-\dot{N}_k = \frac{\partial \Theta^{*d}}{\partial \sum\nolimits_k} \tag{95}$$

$$-\frac{\nabla_i T}{\dot{T}} = \frac{\partial \Theta^{*th}}{\partial q_i} \tag{96}$$

It is clearly seen that the definitions of Ψ, Θ^{*d}, and consequently $\dot{N}_k (k = 1, 2, 3)$ are essential features of the formulation in order to describe the thermomechanical/microstructural behavior of the material involved in the deformation and damage processes. The associative evolution law of $\boldsymbol{\varphi}$ can be obtained by utilizing the calculus of several variables with the Lagrange multiplier $\dot{\lambda}^{\mathrm{d}}$. The dissipation function Π^d (Eqs. 88 and 89) is subjected to the constraint $g = 0$ (Voyiadjis and Kattan 1990, 1992, 1999; Kattan and Voyiadjis 1990, 1993, 2001b) such that the following objective function is formed:

$$\Omega = \Pi^d - \dot{\lambda}^{\mathrm{d}} g \tag{97}$$

where g is the damage surface (i.e., criterion) to be defined later. One now makes use of the maximum dissipation principle which states that the actual state of the thermodynamic force $\mathbf{Y}$ is that which maximizes the dissipation function over all other possible admissible states. Therefore, the objective function Ω is maximized by using the following necessary condition:

$$\frac{\partial \Omega}{\partial Y_{ijkl}} = 0 \tag{98}$$

Substitution of Eq. 97 into Eq. 98 along with Eq. 89 yields the thermodynamic law corresponding to the evolution of the damage tensor $\dot{\varphi}$ as follows:

$$\dot{\varphi}_{ijkl} = \dot{\lambda}^{\mathrm{d}} \frac{\partial g}{\partial Y_{ijkl}} \tag{99}$$

The above equation represents the evolution equation for the fourth-rank damage tensor $\boldsymbol{\varphi}^{(4)}$. On the other hand, an evolution equation for the fourth-rank damage tensor $\boldsymbol{\varphi}^{(4)}$ can be written in terms of the evolution of the fabric tensors. This is performed by taking the time derivative of Eq. 33.2 as follows:

$$\begin{aligned} \dot{\boldsymbol{\varphi}}^{(4)} = -2\Big[\lambda\Big(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\Big) \otimes \Big(\dot{G}^{(0)}\mathbf{I}^{(2)} + \dot{\mathbf{G}}^{(2)}\Big) \\ + 2\mu\Big(G^{(0)}\mathbf{I}^{(2)} + \mathbf{G}^{(2)}\Big) \underline{\overline{\otimes}} \Big(\dot{G}^{(0)}\mathbf{I}^{(2)} + \dot{\mathbf{G}}^{(2)}\Big)\Big] : \overline{\mathbf{E}}^{-1} \end{aligned} \tag{100}$$

The accumulative damage rate $\dot{r}$ may be defined as follows:

$$\dot{r} = \sqrt{\dot{\varphi}_{ijkl}\dot{\varphi}_{ijkl}} \tag{101}$$

The thermoelastic energy Ψ^{te} can be postulated as follows:

$$\Psi^{te} = \frac{1}{2\rho}\varepsilon_{ij}E_{ijkl}(\varphi)\varepsilon_{kl} - \frac{1}{\rho}\beta_{ij}\varepsilon_{ij}(T - T_r) - \eta_r(T - T_r) - \frac{1}{2}c(T - T_r)^2 - \frac{1}{2\rho}k_{ij}\nabla_i T\nabla_j T \tag{102}$$

On the other hand, the thermo-damage energy Ψ^{td} is assumed as follows:

$$\rho\Psi^{td} = \frac{1}{2}a_1 r^2 V + \frac{1}{2}a_2\Gamma_{ijkl}\Gamma_{ijkl}V \tag{103}$$

where $\mathbf{E}(\boldsymbol{\varphi})$ is the fourth-rank damage elasticity tensor, $\boldsymbol{\beta}$ is the tangent conjugate tensor of thermal dilatation (Lubliner 1990), c is the coefficient of thermal expansion, η_r is the reference entropy, T_r is the reference temperature, a_1 and a_2 are material-dependent constants which are considered independent of temperature, $\mathbf{k} = k\delta$ is the heat conductivity second-rank tensor (k being the conductivity coefficient and δ is the Kronecker delta), and V is the homologous temperature defined as $V = 1 - (T/T_m)^n$, where T_m is the melting temperature and n is the temperature softening component.

The proposed definition of Ψ allows the derivation of the constitutive equations and the internal dissipation described next. The constitutive equations for stress (Eq. 85.1) can be written from the thermodynamic potential of Eq. 102 as follows:

$$\sigma_{ij} = E_{ijkl}\varepsilon_{kl} - \beta_{ij}(T - T_r) \tag{104}$$

where

$$E_{ijkl} = \rho\frac{\partial^2\Psi}{\partial\varepsilon_{ij}\partial\varepsilon_{kl}} \tag{105}$$

$$\beta_{ij} = -\rho\frac{\partial^2\Psi}{\partial\varepsilon_{ij}\partial T} \tag{106}$$

The constitutive equations for the entropy (Eq. 85.2) can be written from the thermodynamic potential of Eqs. 102 and 103, assuming a decoupling between the thermal effects induced through elasticity and damage, such that

$$\eta = \eta^{te} + \eta^{td} \tag{107}$$

where

$$\eta^{te} = \eta_r + c(T - T_r) + \frac{1}{\rho}\beta_{ij}\varepsilon_{ij} \tag{108}$$

$$\eta^{td} = \frac{1}{2\rho}\left(a_1 r^2 + a_2 \Gamma_{ijkl}\Gamma_{ijkl}\right)\frac{\partial V}{\partial T} \tag{109}$$

In the above equation, $\frac{\partial V}{\partial T}$ is given by

$$\frac{\partial V}{\partial T} = \frac{n}{T_m}\left(\frac{T}{T_m}\right)^{n-1} \tag{110}$$

The constitutive equation for the heat flux vector **q** can be obtained from Eq. 85.3 as follows:

$$q_i = -k_{ij}\nabla_j T \tag{111}$$

which is the well-known Fourier heat conduction law. The negative sign indicates that the heat flow is opposite to the direction of temperature increase.

The next important step is the selection of the appropriate form of the damage potential function in order to establish the desired constitutive equations that describe the mechanical behavior of the material. In order to be consistent and to satisfy the generalized normality rule of thermodynamics, a proper analytical form for the damage potential function needs to be postulated to obtain evolution equations of the assumed flux variables, such that

$$G = g + \frac{1}{2}h_1 K^2 + \frac{1}{2}h_2 H_{ijkl}H_{ijkl} \tag{112}$$

where h_1 and h_2 are material constants used to adjust the units of the equation and g is the damage surface (criterion) defined as follows:

$$g = \sqrt{\left(Y_{ijkl} - H_{ijkl}\right)\left(Y_{ijkl} - H_{ijkl}\right)} - l - K \leq 0 \tag{113}$$

where the damage forces **Y** and **H** characterize damage evolution and damage kinematic hardening; l is the initial damage threshold as a function of temperature which has the form $l = l_0 V$, where l_0 is the initial damage threshold at zero absolute temperature; while K is the damage isotropic hardening function.

The model response in the damage domain is then characterized by the Kuhn-Tucker complementary conditions as follows:

$$g \leq 0, \quad \dot{\lambda}^{\mathrm{d}} \geq 0, \quad \dot{\lambda}^{\mathrm{d}} g = 0 \tag{114}$$

In order to derive the hardening evolution equations associated with the damage process, Eq. 112 is substituted into the evolution law of $\dot{r}$ so that one obtains the following relation:

$$\dot{r} = \dot{\lambda}^{d}(1 - h_1 K) \tag{115}$$

The evolution equations for the damage isotropic hardening function K can be obtained by first making use of Eq. 115 and substituting it into the evolution law for $\dot{K}$ so that one can obtain the following relation:

$$\dot{K} = a_1(1 - h_1 K)\dot{\lambda}^{\mathrm{d}} V \tag{116}$$

Furthermore, the evolution equation for the damage kinematic hardening parameter can be obtained by using Eq. 112 and substituting it into the evolution law of $\dot{\Gamma}$ while realizing that $\partial g/\partial \mathbf{H} = -\partial g/\partial \mathbf{Y}$ (clear from Eq. 113) and using Eq. 99 to obtain

$$\dot{\Gamma}_{ijkl} = \dot{\varphi}_{ijkl} - h_2 \dot{\lambda}^{\mathrm{d}} H_{ijkl} \tag{117}$$

Finally, it can be easily seen that by substituting Eq. 117 into the evolution law of $\dot{\mathbf{H}}$, one may obtain the following relation:

$$\dot{H}_{ijkl} = \left(a_2 \dot{\varphi}_{ijkl} - h_2 a_2 \dot{\lambda}^{\mathrm{d}} H_{ijkl}\right) V \tag{118}$$

Equations 115, 116, 117, and 118 represent the evolution laws for the various parameters involved in the damage process and damage hardening. One may continue and derive an explicit expression for the thermodynamic force $\mathbf{Y}$. But this is not performed here as this step may limit the theory by invoking the effective stress space and the special case of using the damage effect tensor.

Summary and Conclusions

A new theory of damage mechanics in terms of fabric tensors is postulated within the framework of classical elasticity assuming small strains. The new theory is called damage mechanics with fabric tensors. First, a review of the concepts of fiber tensors is presented. Then, a general hypothesis is postulated in damage mechanics where it is seen that the two hypotheses of elastic strain equivalence and elastic energy equivalence are obtained as special cases. This is followed by the derivation of an explicit expression of the fourth-rank damage tensor of the zero-rank and second-rank fabric tensors. This is then followed by the investigation of the case of plane stress.

An application of the theory to micro-crack distributions is presented next. The data for a micro-crack distribution is presented first in the form of a circular histogram (rose diagram). This histogram is then used to determine the values of the components of the second-rank and fourth-rank fabric tensors. Then the components of the damage tensor are calculated using the determined values of the fabric tensor. Two approximations of the micro-crack distributions are also given – one is a second-order approximation based on the second-rank fabric tensor, while the other one is a fourth-order approximation based on the fourth-rank fabric tensor.

These two approximations of the distribution are then compared with the real distribution given in the circular histogram. Finally, certain relationships are derived between the damage tensor components and the fabric tensor parameters, and several of these relationships are plotted in a graphical format.

Another application of the theory is presented to compare two sets of parallel micro-crack distributions with two different orientations. In this case, it is seen that interchanging the orientation of the parallel micro-cracks causes an interchange in the first two diagonal terms in the elastic stiffness matrix.

Finally, this work is concluded with an exposition of thermodynamical concepts and derivation of the equations of damage evolution. This is performed in a mathematically consistent manner that is based on sound thermodynamic principles. It should be pointed out that this work will form the basis for future work in this area including visualization and finite element implementation with application to composite materials.

References

A. Cauvin, R. Testa, Damage mechanics: basic variables in continuum theories. Int. J. Solids Struct. **36**, 747–761 (1999)

J.L. Chaboche, Continuous damage mechanics – a tool to describe phenomena before crack initiation. Nucl. Eng. Des. **64**, 233–247 (1981)

C. Chow, J. Wang, An anisotropic theory of elasticity for continuum damage mechanics. Int. J. Fract. **33**, 3–16 (1987)

B. Coleman, M. Gurtin, Thermodynamics with internal state variables. J. Chem. Phys. **47**(2), 597–613 (1967)

S. Cowin, Properties of the anisotropic elasticity tensor. Q. J. Mech. Appl. Math. **42**(Pt. 2), 249–266 (1989)

I. Doghri, *Mechanics of Deformable Solids: Linear and Nonlinear, Analytical and Computational Aspects* (Springer, Berlin, 2000)

D. Hayhurst, Creep rupture under multiaxial states of stress. J. Mech. Phys. Solids **20**, 381–390 (1972)

Q. He, A. Curnier, A more fundamental approach to damaged elastic stress–strain relations. Int. J. Solids Struct. **32**(10), 1433–1457 (1995)

M. Jones, *Spherical Harmonics and Tensors in Classical Field Theory* (Wiley, New York, 1985)

L. Kachanov, On the creep fracture time. Izv. Akad. Nauk. USSR Otd. Tech. **8**, 26–31 (1958) (in Russian)

K. Kanatani, Distribution of directional data and fabric tensors. Int. J. Eng. Sci. **22**(2), 149–164 (1984a)

K. Kanatani, Stereological determination of structural anisotropy. Int. J. Eng. Sci. **22**(5), 531–546 (1984b)

P.I. Kattan, G.Z. Voyiadjis, A coupled theory of damage mechanics and finite strain elasto-plasticity – part I: damage and elastic deformations. Int. J. Eng. Sci. **28**(5), 421–435 (1990)

P.I. Kattan, G.Z. Voyiadjis, A plasticity-damage theory for large deformation of solids – part II: applications to finite simple shear. Int. J. Eng. Sci. **31**(1), 183–199 (1993)

P.I. Kattan, G.Z. Voyiadjis, Decomposition of damage tensor in continuum damage mechanics. J. Eng. Mech. ASCE **127**(9), 940–944 (2001a)

P.I. Kattan, G.Z. Voyiadjis, *Damage Mechanics with Finite Elements: Practical Applications with Computer Tools* (Springer, Berlin, 2001b)

H. Lee, K. Peng, J. Wang, An anisotropic damage criterion for deformation instability and its application to forming limit analysis of metal plates. Eng. Fract. Mech. **21**, 1031–1054 (1985)
J. Lemaitre, Evaluation of dissipation and damage in metals subjected to dynamic loading, in *Proceedings of I.C.M. 1*, Kyoto, 1971
J. Lemaitre, How to use damage mechanics. Nucl. Eng. Des. **80**, 233–245 (1984)
J. Lemaitre, J.L. Chaboche, *Mechanics of Solid Materials* (Cambridge University Press, London, 1990)
V. Lubarda, D. Krajcinovic, Damage tensors and the crack density distribution. Int. J. Solids Struct. **30**(20), 2859–2877 (1993)
J. Lubliner, *Plasticity Theory* (Macmillan, New York, 1990)
S. Nemat-Nasser, *Plasticity, a Treatise on Finite Deformation of Heterogeneous Inelastic Materials* (Cambridge University Press, Cambridge, UK, 2004)
S. Nemat-Nasser, M. Hori, *Microfiche: Overall Properties of Heterogeneous Solids*, 2nd rev edn (Elsevier, Amsterdam, 1999)
M. Oda, S. Nemat-Nasser, M. Mehrabadi, A statistical study of fabric in a random assembly of spherical granules. Int. J. Numer. Anal. Methods Geomech. **6**, 77–94 (1982)
Y. Rabotnov, Creep rupture, in *Proceedings, Twelfth International Congress of Applied Mechanics*, Stanford, 1968, eds. by M. Hetenyi, W.G. Vincenti (Springer, Berlin, 1969), pp. 342–349
M. Satake, Fabric tensors in granular materials, in *IUTAM Conference on Deformation and Failure of Granular Materials*, Delft, 31 Aug– 3 Sept 1982, pp. 63–68
S. Sutcliffe, Spectral decomposition of the elasticity tensor. ASME J. Appl. Mech. **59**, 762–773 (1992)
G.Z. Voyiadjis, P.I. Kattan, A coupled theory of damage mechanics and finite strain elastoplasticity – part II: damage and finite strain plasticity. Int. J. Eng. Sci. **28**(6), 505–524 (1990)
G.Z. Voyiadjis, P.I. Kattan, A plasticity-damage theory for large deformation of solids – part I: theoretical formulation. Int. J. Eng. Sci. **30**(9), 1089–1108 (1992)
G.Z. Voyiadjis, P.I. Kattan, On the symmetrization of the effective stress tensor in continuum damage mechanics. J. Mech. Behav. Mater. **7**(2), 139–165 (1996)
G.Z. Voyiadjis, P.I. Kattan, *Advances in Damage Mechanics: Metals and Metal Matrix Composites* (Elsevier Science, Amsterdam, 1999)
G.Z. Voyiadjis, P.I. Kattan, Damage mechanics with fabric tensors. Mech. Adv. Mater. Struct. **13**, 285–301 (2006)
G.Z. Voyiadjis, P.I. Kattan, Z.N. Taqieddin, Continuum approach to damage mechanics of composite materials with fabric tensors. Int. J. Damage Mech. **16**(7), 301–329 (2007a). http://online.sagepub.com
G.Z. Voyiadjis, Z.N. Taqieddin, P.I. Kattan, Micromechanical approach to damage mechanics of composite materials with fabric tensors. Compos. Part B: Eng. **38**(7–8), 862–877 (2007b). www.sciencedirect.com
P. Zysset, A. Curnier, An alternative model for anisotropic elasticity based on fabric tensors. Mech. Mater. **21**, 243–250 (1995)
P. Zysset, A. Curnier, A 3D damage model for trabecular bone based on fabric tensors. J. Biomech. **29**(12), 1549–1558 (1996)

Evolution of Fabric Tensors in Continuum Damage Mechanics of Solids with Micro-cracks: Studying the Effects of Length and Orientation

4

George Z. Voyiadjis, Peter I. Kattan, and Ziad N. Taqieddin

Contents

Abstract

In this chapter, the evolution of fabric tensors, based on micro-crack distributions, is formulated within the framework of thermodynamics. The exact definition of fabric tensors based on micro-crack distributions is presented. This definition is seen to incorporate both the orientation and length of a micro-crack. In this regard, the micro-crack distribution is assumed to be radially symmetric, i.e., symmetric about a line through the origin. A thermodynamic force that is associated with the fabric tensor is defined and utilized in the derivation of the evolution equations. The application of the theory to the case of uniaxial tension is derived and presented.

Specific uncoupled equations for the evolution of the length and orientation of micro-cracks are also derived. In this regard, some interesting results are obtained. It is concluded that the micro-crack length and orientation cannot evolve simultaneously for the same set of micro-cracks. However, two different

G.Z. Voyiadjis (✉) • P.I. Kattan
Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA
e-mail: voyiadjis@eng.lsu.edu; pkattan@orange.jo

Z.N. Taqieddin
Civil Engineering Department, Applied Science University, Amman, Jordan
e-mail: z_taqieddin@asu.edu.jo

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_4

sets of micro-cracks may be considered in the same representative volume element (RVE) where in one set the micro-crack length evolves, while in the second set the micro-crack orientation evolves.

Introduction

Satake (1982) applied the concept of fabric tensors to granular materials. The anisotropy due to the fabric (of the distributed data like crack distributions or granular particles) is represented by a tensor in terms of the normals (to the cracks or to the contact surfaces in granular materials). This tensor is usually called the fabric tensor (Satake 1982; Kanatani 1984a; Oda 1982). The fabric tensor is usually related to the probability density function of the distributed data (crack normals or contact normals).

Kanatani (1984a) formulated the concept of fabric tensors based on a rigorous mathematical treatment. He used fabric tensors to describe distributions of directional data like crack distributions in a damaged material element. He applied the least square approximation (a well-known statistical technique) to derive equations for the various fabric tensors he postulated. He defined three types of fabric tensors: fabric tensors of the first kind, denoted by **N**; fabric tensors of the second kind, denoted by **F**; and fabric tensors of the third kind, denoted by **D**. He derived the exact mathematical relations between these three types of fabric tensors. The work of Kanatani (1984a) on fabric tensors is very important and is used extensively here.

Zysset and Curnier (1995, 1996) formulated an alternative model for anisotropic elasticity based on fabric tensors. Zysset and Curnier (1995) introduced a general approach for relating the material microstructure to the fourth-rank elasticity tensor based on the Fourier series decomposition. They proposed an approximation based on a scalar and a symmetric, traceless second-rank fabric tensor. Using the representation theorem for anisotropic functions with tensorial arguments, Zysset and Curnier (1995) derived a general expression for the elastic free energy and discussed the resulting material symmetry in terms of the fabric tensors. Finally, they derived a general explicit expression for the fourth-rank elasticity tensor in terms of the fabric tensor. This last result is very important and is used extensively here.

Lubarda and Krajcinovic (1993) applied the definitions of fabric tensors (Kanatani 1984a) to crack density distributions. They actually recast Kanatani's general work on directional data (Kanatani 1984a) in terms of crack distributions. Lubarda and Krajcinovic (1993) examined the relationship between a given, experimentally determined, distribution of cracks and the scalar, second-rank and fourth-rank fabric tensors. They employed the usual representation of experimentally measured micro-crack densities in planes with different orientations in the form of circular histogram (rose diagram). They then used the data contained in the circular histogram to approximate the distribution function defined on a unit sphere and centered in a material point. They solved several examples with different crack

distributions to illustrate this point. They assumed that one of the three types of fabric tensors is identical to the damage tensor of continuum damage mechanics.

The damage variable (or tensor), based on the effective stress concept, represents average material degradation which reflects the various types of damage at the microscale level like nucleation and growth of voids, cracks, cavities, micro-cracks, and other microscopic defects (Chaboche 1981; Kattan and Voyiadjis 1990, 1993; Lemaitre 1971; Lemaitre and Chaboche 1990; Rabotnov 1969; Voyiadjis and Kattan 1990, 1992).

For the case of isotropic damage mechanics, the damage variable is scalar and the evolution equations are easy to handle. However, it has been shown by Cauvin and Testa (1999) that two independent damage variables must be used in order to describe accurately and consistently the special case of isotropic damage. It has been argued (Lemaitre 1984) that the assumption of isotropic damage is sufficient to give good predictions of the load-carrying capacity, the number of cycles, or the time to local failure in structural components. However, the development of anisotropic damage has been confirmed experimentally (Hayhurst 1972; Chow and Wang 1987; Lee et al. 1985) even if the virgin material is isotropic. This has prompted several researchers to investigate the general case of anisotropic damage (Voyiadjis and Kattan 1996, 1999; Kattan and Voyiadjis 2001a, b).

In continuum damage mechanics, usually a phenomenological approach is adopted. In this approach, the most important concept is that of the representative volume element (RVE). The discontinuous and discrete elements of damage are not considered within the RVE; rather their combined effects are lumped together through the use of a macroscopic internal variable. In this way, the formulation may be derived consistently using sound mechanical and thermodynamic principles.

In this work, the evolution equations of fabric tensors are formulated based on sound thermodynamic principles. For this purpose, a general thermodynamic force that is associated with the fabric tensor is defined. Then, both the fabric tensor and its associated thermodynamic force are used in the derivation of the evolution equations. In this chapter, the exact definition of fabric tensors based on micro-crack distributions is presented. This definition is seen to incorporate both the orientation and length of a micro-crack. In this regard, the micro-crack distribution is assumed to be radially symmetric, i.e., symmetric about a line through the origin.

The equations of thermodynamics are employed in order to derive the exact evolution equations of the fabric tensors defined in Part I. In this regard, a thermodynamic force that is associated with the fabric tensor is defined and utilized in the derivation of the evolution equations. The application of the theory to the case of uniaxial tension is derived in this chapter.

It should be noted that the theory presented here is for general directional data. When applying this theory for micro-crack distributions, then the theory depends exclusively on the orientations and the lengths of micro-cracks. This theory does not consider the opening and closing of the micro-cracks, the closeness of the micro-cracks, and the interaction among the micro-cracks. These effects are beyond

the scope of this work. Detailed description of the theory that depends exclusively on the orientation of micro-cracks can be found in this chapter.

Specific uncoupled equations for the evolution of the length and orientation of micro-cracks are also derived. In this regard, some interesting results are obtained. It is concluded that the micro-crack length and orientation cannot evolve simultaneously for the same set of micro-cracks. However, two different sets of micro-cracks may be considered in the same RVE where in one set the micro-crack length evolves, while in the second set the micro-crack orientation evolves.

The tensor notation used here is as follows. All vectors and tensors appear in bold type. The following operations are also defined. For second-rank tensors **A** and **B**, the following notation is used: $(\mathbf{A} \pm \mathbf{B})_{ij} = A_{ij} \pm B_{ij}$, $\mathbf{A} : \mathbf{B} = A_{ij}B_{ij}$, $(\mathbf{A} \otimes \mathbf{B})_{ijkl} = A_{ij}B_{kl}$. For fourth-rank tensors **C** and **D**, the following notation is used: $(\mathbf{C} \pm \mathbf{D})_{ijkl} = C_{ijkl} \pm D_{ijkl}$, $(\mathbf{C} : \mathbf{D})_{ijkl} = C_{ijmn}D_{mnkl}$. For second-rank tensor **A** and fourth-rank tensor **C**, the following notation is used: $(\mathbf{C} : \mathbf{A})_{ij} = C_{ijkl}A_{kl}$. For damage tensors, fabric tensors, and identity tensors, a superscript with braces is used to indicate the order of the tensor. For all other tensors, the order of the tensor is clear from the text and equations.

Review of Fabric Tensors and the Mesoscopic Theory

In this section, a solid with micro-cracks is investigated and a representative volume element (RVE) is considered as shown in Fig. 1a. In addition, the assumptions of the mesoscopic theory of micro-cracks of Papenfuss et al. (2003) are adopted and summarized as follows:

1. The micro-cracks are assumed to have a linear cross section as shown in Fig. 1b. This assumption arises due to the fact that the opening of the micro-crack does not have an effect on the elastic stiffness in this theory. This also means that the opening and closing of micro-cracks are not considered in this theory.
2. The diameter of each micro-crack is much smaller than the linear dimension of the RVE. Therefore, the micro-cracks may be considered as part of the micro-structure of the RVE. It is assumed that the micro-cracks are small enough such that there is a whole distribution of micro-crack sizes and orientations in the RVE.

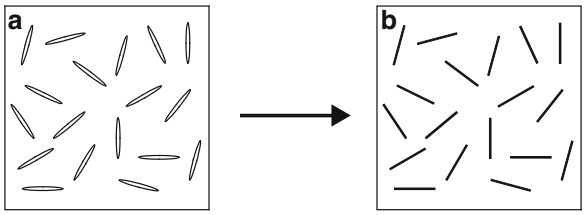


Fig. 1 A typical micro-crack distribution (Reprinted with permission from G. Voyiadjis, P. Kattan, Mech. Res. Commun. 34, 2 (2007))

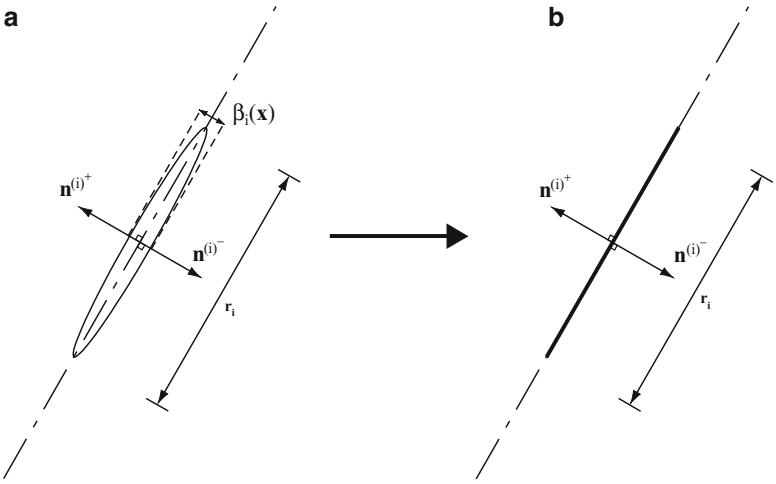


Fig. 2 Geometry of a micro-crack (Reprinted with permission from G. Voyiadjis, P. Kattan, Mech. Res. Commun. 34, 2 (2007))

3. The micro-cracks are fixed to the material. This means that the motion of micro-cracks is directly linked to the motion of the RVE.
4. All micro-cracks within the RVE move and rotate with the same velocity. The micro-cracks cannot rotate independently of the material.
5. The behavior of the material of the RVE is linear elastic. Furthermore, small elastic strains are assumed throughout the RVE.

Consider a typical penny-shaped micro-crack "*i*" as shown in Fig. 2a. Utilizing assumption 1 above, one can reduce the geometry of the micro-crack as shown in Fig. 2b. Let $\mathbf{n}^{(i)}$ be a unit vector normal to the surface of the micro-crack. Then, the vector $\mathbf{n}^{(i)}$ represents the orientation of micro-crack "*i*." Let the length of the micro-crack be denoted by r_i which is considered to be a scalar quantity. Note that in Fig. 2, the relation $\mathbf{n}_i^{(+)} = -\mathbf{n}_i^{(-)}$ holds.

The above definition is a typical definition for a micro-crack. However, Oda (1982) considered r_i to be defined as follows:

1. Consider a flat micro-crack "*i*" with an occupied area A_i as shown in Fig. 3a.
2. The micro-crack consists of two micro-crack surfaces, each of which has a unit normal vector $\mathbf{n}^{(i)}$ or $-\mathbf{n}^{(i)}$.
3. Replace the micro-crack with an equivalent circle with the same occupied area A_i as shown in Fig. 3b.
4. The radius of the equivalent circle r_i is calculated by the relation $r_i = \sqrt{\mathrm{A}_i/\pi}$.

Consider again the RVE of the micro-crack distribution. The micro-crack density distribution function $f(\mathbf{n})$ varies in all direction and is an orientation distribution function (Qiang et al. 2001). Let θ and ϕ be the spherical coordinates

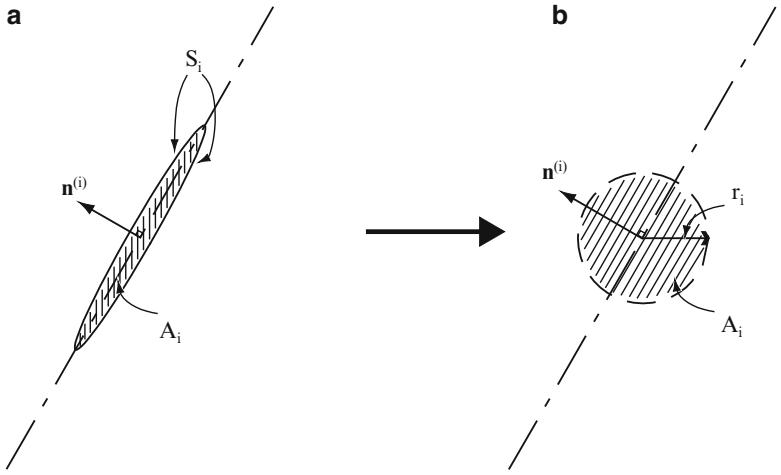


Fig. 3 Concept of a micro-crack when compared to a spherical void (Reprinted with permission from G. Voyiadjis, P. Kattan, Mech. Res. Commun. 34, 2 (2007))

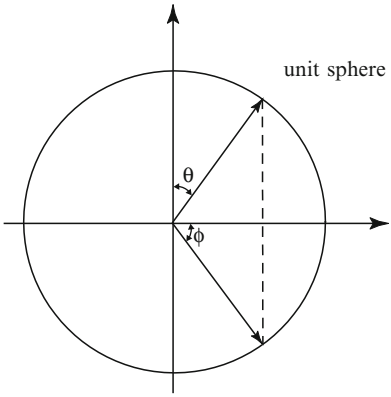


Fig. 4 Spherical coordinates on the unit sphere (Reprinted with permission from G. Voyiadjis, P. Kattan, Mech. Res. Commun. 34, 2 (2007))

of the unit sphere as shown in Fig. 4. Then, any possible direction **n** may be represented by one point on the surface of the unit sphere where

$$n_1 = \sin\theta\cos\phi \tag{1.1}$$

$$n_2 = \sin\theta\sin\phi \tag{1.2}$$

$$n_3 = \cos\theta \tag{1.3}$$

Consider an infinitesimal area on the unit sphere denoted by $d\mathbf{n}$ where $d\mathbf{n} = \sin\theta d\theta d\phi$. This small area represents a bundle of directions about the unit vector **n**. The number of micro-cracks in this infinitesimal area is given by

$f(\mathbf{n})d\mathbf{n}/(4\pi)$. Therefore, the micro-crack volumetric density f_o in the RVE is given by

$$f_o = \frac{1}{4\pi}\oint f(\mathbf{n})dn = \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} f(\mathbf{n})\sin\theta d\theta d\phi \tag{2}$$

If the micro-crack distribution is uniform, then $f(\mathbf{n}) = f_o$. Next, the definition of the fabric tensor for this micro-crack distribution is considered. In the definition of the fabric tensor, both the orientation $\mathbf{n}_i$ and length r_i of each micro-crack are included. Therefore, the zero-rank, second-rank, and fourth-rank fabric tensors are defined by (Qiang et al. 2001; Yang et al. 2004)

$$G^{(0)} = \frac{1}{V}\oint_i r_i^3 d\mathbf{n} \tag{3.1}$$

$$\mathbf{G}^{(2)} = \frac{1}{V}\oint_i r_i^3 \mathbf{n}^{(i)}\mathbf{n}^{(i)} d\mathbf{n} \tag{3.2}$$

$$\mathbf{G}^{(4)} = \frac{1}{V}\oint_i r_i^3 \mathbf{n}^{(i)}\mathbf{n}^{(i)}\mathbf{n}^{(i)}\mathbf{n}^{(i)} d\mathbf{n} \tag{3.3}$$

(no sum over i)
where V is the volume of the RVE.

It is clear from Eq. 3.1 that $G^{(0)}$ is a scalar. Using indicial notation, one may write the components of the second-rank and fourth-rank fabric tensors $\mathbf{G}^{(2)}$ and $\mathbf{G}^{(4)}$, respectively, as follows:

$$\mathbf{G}^{(2)}_{mn} = \frac{1}{V}\oint_i r_i^3 \mathbf{n}^{(i)}_m \mathbf{n}^{(i)}_n d\mathbf{n} \tag{4.1}$$

$$\mathbf{G}^{(4)}_{mnkl} = \frac{1}{V}\oint_i r_i^3 \mathbf{n}^{(i)}_m \mathbf{n}^{(i)}_n \mathbf{n}^{(i)}_k \mathbf{n}^{(i)}_l d\mathbf{n} \tag{4.2}$$

It is noted that the fabric tensors $\mathbf{G}^{(\alpha)}$ ($\alpha = 0, 2, 4$) are exactly the fabric tensors of the first kind of Kanatani (1984a, b). However, Kanatani considered only the orientation vectors without the length parameter. It should be noted that the fabric tensors are not independent. For example, a second-rank fabric tensor may be fully determined by a fourth-rank fabric tensor (Qiang et al. 2001; Yang et al. 2004). In using fabric tensors in constitutive equations, it is essential that the state variables be independent with each other in the constitutive equations (He and Curnier 1995; Voyiadjis and Kattan, 2006).

Next, the evolution equations of fabric tensors using thermodynamic principles of Coleman and Gurtin (1967) and Doghri (2000) derived. Consider a general even-rank fabric tensor $\mathbf{G}^{(\alpha)}$ where $\alpha = 0, 2, 4, \ldots$ as defined in equations (3). Let ψ be the Helmholtz free energy density function. Define a generalized even-rank tensorial thermodynamic force $\mathbf{H}^{(\alpha)}$, ($\alpha = 0, 2, 4, \ldots$) that is associated with the fabric tensor $\mathbf{G}^{(\alpha)}$ as follows:

$$\mathbf{H}^{(\alpha)} = \frac{\partial \psi}{\partial \mathbf{G}^{(\alpha)}} \tag{5}$$

where $\psi \equiv \psi(\mathbf{G}^{(\alpha)})$. Using indicial notation, Eq. 5 may be written explicitly as follows for $\alpha = 0, 2, 4$:

$$H^{(0)} = \frac{\partial \psi}{\partial G^{(0)}} \tag{6.1}$$

$$H_{ij}^{(2)} = \frac{\partial \psi}{\partial G_{ij}^{(2)}} \tag{6.2}$$

$$H_{ijkl}^{(4)} = \frac{\partial \psi}{\partial G_{ijkl}^{(4)}} \tag{6.3}$$

Next, one assumes the existence of m damage surfaces where each surface is denoted by the scalar-valued function $g^{(\alpha)}$, where $\alpha = 0, 2, 4, \ldots, m$. The damage criterion $g^{(\alpha)}$ is defined as follows:

$$g^{(\alpha)} = \sqrt{\frac{1}{2}\mathbf{H}^{(\alpha)} : \mathbf{J}^{(\alpha)} : \mathbf{H}^{(\alpha)}} - l_o^{(\alpha)} - L^{(\alpha)}\left(l^{(\alpha)}\right) = 0 \tag{7}$$

where $\mathbf{J}^{(\alpha)}$ is a constant tensor represented by the following 6×6 matrix:

$$[\mathbf{J}] \equiv \begin{bmatrix} 1 & \mu & \mu & 0 & 0 & 0 \\ \mu & 1 & \mu & 0 & 0 & 0 \\ \mu & \mu & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2(1-\mu) & 0 & 0 \\ 0 & 0 & 0 & 0 & 2(1-\mu) & 0 \\ 0 & 0 & 0 & 0 & 0 & 2(1-\mu) \end{bmatrix} \tag{8}$$

where μ is a material constant satisfying $-\frac{1}{2} \leq \mu \leq 1$. In Eq. 7, the scalar function $L^{(\alpha)}$ is a damage strengthening criterion that is a function of the overall damage parameter $l^{(\alpha)}$. For the case $\alpha = 0$, Eq. 7 is rewritten as follows:

$$g^{(0)} = \sqrt{\frac{1}{2}\mathbf{H}^{(0)} : \mathbf{J}^{(0)} : \mathbf{H}^{(0)}} - l_o^{(0)} - L^{(0)}\left(l^{(0)}\right) = 0 \tag{9}$$

where $J^{(0)}$ is a scalar parameter given as the first term in Eq. 8, i.e., $J^{(0)} = 1$. Therefore, Eq. 8 reduces to

$$g^{(0)} = \frac{\mathbf{H}^{(0)}}{\sqrt{2}} - l_o^{(0)} - L^{(0)}\left(l^{(0)}\right) = 0 \tag{10}$$

To simplify the generated expressions for $g^{(\alpha)}$ given in Eq. 7, one may assume that $L^{(\alpha)}(l^{(\alpha)}) = L(l)$ where $l^{(0)} = l^{(2)} = l^{(4)} = \ldots = l^{(\alpha)} = l$.

Next, consider the power of dissipation function Π given as follows:

$$\Pi = -\sum_{\alpha=0,2,4,\ldots} \mathbf{H}^{(\alpha)} : d\mathbf{G}^{(\alpha)} - \sum_{\alpha=0,2,4,\ldots} L^{(\alpha)} dl^{(\alpha)} \tag{11}$$

Define the objective functions Ψ such that

$$\Psi = \Pi - \sum_{\alpha=0,2,4,\ldots} d\lambda^{(\alpha)} \cdot g^{(\alpha)} \tag{12}$$

where $d\lambda^{(\alpha)}$, $(\alpha = 0, 2, 4, \ldots)$ are scalar Lagrange multipliers. In order to extremize the objective function Ψ, the following conditions are applied:

$$\frac{\partial \Psi}{\partial \mathbf{H}^{(\alpha)}} = 0 \tag{13.1}$$

$$\frac{\partial \Psi}{\partial \mathbf{L}^{(\alpha)}} = 0 \tag{13.2}$$

Substituting Eqs. 11 and 12 into Eq. 13, one obtains

$$\sum_{\alpha=0,2,4,\ldots} d\mathbf{G}^{(\alpha)} = -\sum_{\alpha=0,2,4,\ldots} d\lambda^{(\alpha)} \frac{\partial g^{(\alpha)}}{\partial \mathbf{H}^{(\alpha)}} \tag{14}$$

Equation 14 represents a sum of the evolutions of all fabric tensors of rank α, $(\alpha = 0, 2, 4, \ldots)$. At this stage, one needs to make the assumption of decoupled evolution of these fabric tensors. Based on this assumption, the summation signs in Eq. 14 may be dropped to obtain

$$d\mathbf{G}^{(\alpha)} = -d\lambda^{(\alpha)} \frac{\partial g^{(\alpha)}}{\partial \mathbf{H}^{(\alpha)}}, \quad \alpha = 0, 2, 4, \ldots \tag{15}$$

Equation 15 represents an elementary form of the evolution of the fabric tensor of rank α. Next, the scalar Lagrange multiplier $d\lambda^{(\alpha)}$ is evaluated. Substituting Eqs. 11 and 12 into Eq. 13.2 and simplifying, one obtains

$$\sum_{\alpha=0,2,4,\ldots} dl^{(\alpha)} = -\sum_{\alpha=0,2,4,\ldots} d\lambda^{(\alpha)} \frac{\partial g^{(\alpha)}}{\partial \mathbf{L}^{(\alpha)}} \tag{16}$$

Again, the assumption of decoupling of all the $dl^{(\alpha)}$ is made, $(\alpha = 0, 2, 4, \ldots)$. Therefore, one may drop the summation signs in Eq. 16 to obtain

$$dl^{(\alpha)} = -d\lambda^{(\alpha)} \frac{\partial g^{(\alpha)}}{\partial \mathbf{L}^{(\alpha)}}, \quad \alpha = 0, 2, 4, \ldots \tag{17}$$

Using Eq. 7, it is noted that $\partial g^{(\alpha)}/\partial \mathbf{L}^{(\alpha)} = -1$. Substituting this result into Eq. 17, one obtains

$$dl^{(\alpha)} = d\lambda^{(\alpha)}, \quad \alpha = 0, 2, 4, \ldots \tag{18}$$

Actually, the general form of Eq. 18 is

$$\sum_{\alpha=0,\,2,\,4,\,\ldots} dl^{(\alpha)} = \sum_{\alpha=0,\,2,\,4,\,\ldots} d\lambda^{(\alpha)} \tag{19}$$

However, by employing the assumption of the independence of Lagrange multipliers, one may arrive at Eq. 18. Next, the Lagrange multiplier $d\lambda^{(\alpha)}$ is evaluated.

Using the consistency condition $dg^{(\alpha)} = 0$ where $g^{(\alpha)} \equiv g^{(\alpha)}(\mathbf{H}^{(\alpha)}, L^{(\alpha)})$ as given in Eq. 7, one obtains

$$\frac{\partial g^{(\alpha)}}{\partial \mathbf{H}^{(\alpha)}} : d\mathbf{H}^{(\alpha)} + \frac{\partial g^{(\alpha)}}{\partial L^{(\alpha)}} dL^{(\alpha)} = 0 \tag{20}$$

Using the chain rule, one may write

$$dL^{(\alpha)} = \frac{dL^{(\alpha)}}{dl^{(\alpha)}} dl^{(\alpha)} \tag{21}$$

Substituting the result that $\partial g^{(\alpha)}/\partial L^{(\alpha)} = -1$ along with Eq. 21 into Eq. 20, one obtains

$$\frac{\partial g^{(\alpha)}}{\partial \mathbf{H}^{(\alpha)}} : d\mathbf{H}^{(\alpha)} - \frac{dL^{(\alpha)}}{dl^{(\alpha)}} dl^{(\alpha)} = 0 \tag{22}$$

Using Eq. 7, one may write the following derivative:

$$\frac{\partial g^{(\alpha)}}{\partial \mathbf{H}^{(\alpha)}} = \frac{\mathbf{J}^{(\alpha)} : \mathbf{H}^{(\alpha)}}{\sqrt{2\mathbf{H}^{(\alpha)} : \mathbf{J}^{(\alpha)} : \mathbf{H}^{(\alpha)}}} \tag{23}$$

Substituting Eq. 23 into Eq. 22, simplifying, and using Eq. 18, one obtains

$$d\lambda^{(\alpha)} = dl^{(\alpha)} = \frac{\mathbf{J}^{(\alpha)} : \mathbf{H}^{(\alpha)} : d\mathbf{H}^{(\alpha)}}{\dfrac{dL^{(\alpha)}}{dl^{(\alpha)}} \sqrt{2\mathbf{H}^{(\alpha)} : \mathbf{J}^{(\alpha)} : \mathbf{H}^{(\alpha)}}} \tag{24}$$

Substituting the expression of the Lagrange multiplier of Eq. 24 into Eq. 14, while also using Eq. 23, one obtains

$$d\mathbf{G}^{(\alpha)} = -\frac{2\mathbf{J}^{(\alpha)} : \mathbf{H}^{(\alpha)} : d\mathbf{H}^{(\alpha)} : \mathbf{J}^{(\alpha)} : \mathbf{H}^{(\alpha)}}{\dfrac{dL^{(\alpha)}}{dl^{(\alpha)}}\mathbf{H}^{(\alpha)} : \mathbf{J}^{(\alpha)} : \mathbf{H}^{(\alpha)}}, \quad \alpha = 0, 2, 4, \ldots \tag{25}$$

where $dL^{(\alpha)}/dl^{(\alpha)}$ is considered to be the material damage parameter. For more details on the evaluation of this parameter and its physical significance, the reader is referred to the book by Voyiadjis and Kattan (1999). Equation 25 represents the evolution of the fabric tensor $d\mathbf{G}^{(\alpha)}$ in terms of the evolution of its associated thermodynamic tensorial function $d\mathbf{H}^{(\alpha)}$. In order to proceed further, the evaluation of the tensor $\mathbf{H}^{(\alpha)}$ and the derivation of an appropriate expression for its evolution $d\mathbf{H}^{(\alpha)}$ need more elaboration. The generalized thermodynamic tensorial force function $\mathbf{H}^{(\alpha)}$ that is associated with the fabric tensor $\mathbf{G}^{(\alpha)}$ is assumed here to be a function of the Cauchy stress tensor $\boldsymbol{\sigma}$ in the deformed and damaged configuration of the material, i.e., $\mathbf{H}^{(\alpha)} \equiv \mathbf{H}^{(\alpha)}(\mathbf{G}^{(\alpha)}, \boldsymbol{\sigma})$. Therefore, one obtains the following by evaluating the differential $d\mathbf{H}^{(\alpha)}$:

$$d\mathbf{H}^{(\alpha)} = \frac{\partial H^{(\alpha)}}{\partial \mathbf{G}^{(\alpha)}} : d\mathbf{G}^{(\alpha)} + \frac{\partial H^{(\alpha)}}{\partial \boldsymbol{\sigma}} : d\boldsymbol{\sigma} \tag{26}$$

Substituting Eq. 25 for $d\mathbf{G}^{(\alpha)}$ into Eq. 26, simplifying, and solving for $d\mathbf{H}^{(\alpha)}$, one obtains

$$d\mathbf{H}^{(\alpha)} = \left(\mathbf{I}^{(2\alpha)} - \frac{\partial \mathbf{H}^{(\alpha)}}{\partial \mathbf{G}^{(\alpha)}} : \mathbf{P}^{(2\alpha)}\right)^{-1} : \frac{\partial \mathbf{H}^{(\alpha)}}{\partial \boldsymbol{\sigma}} : d\boldsymbol{\sigma} \tag{27}$$

where $\mathbf{P}^{(2\alpha)}$ is an even-rank tensor of rank 2α given by

$$\mathbf{P}^{(2\alpha)} = -\frac{2\mathbf{J}^{(\alpha)} : \mathbf{H}^{(\alpha)} : \mathbf{J}^{(\alpha)} : \mathbf{H}^{(\alpha)}}{\dfrac{dL^{(\alpha)}}{dl^{(\alpha)}}\mathbf{H}^{(\alpha)} : \mathbf{J}^{(\alpha)} : \mathbf{H}^{(\alpha)}} \tag{28}$$

and $\mathbf{I}^{(2\alpha)}$ is the identity tensor of rank 2α.

It should be noted that the tensor $\mathbf{P}^{(2\alpha)}$ may be directly derived from Eq. 25 such that

$$d\mathbf{G}^{(\alpha)} = \mathbf{P}^{(2\alpha)} : d\mathbf{H}^{(\alpha)} \tag{29}$$

Finally, substituting Eq. 27 into Eq. 29, one obtains

$$d\mathbf{G}^{(\alpha)} = \mathbf{L}^{(\alpha+2)} : d\boldsymbol{\sigma} \tag{30}$$

where the tensor $\mathbf{L}^{(\alpha+2)}$ is of rank $(\alpha + 2)$ and is given by

$$\mathbf{L}^{(\alpha+2)} = \mathbf{P}^{(2\alpha)} : \left(\mathbf{I}^{(2\alpha)} - \frac{\partial \mathbf{H}^{(\alpha)}}{\partial \mathbf{G}^{(a)}} : \mathbf{P}^{(2\alpha)} \right)^{-1} : \frac{\partial H^{(\alpha)}}{\partial \boldsymbol{\sigma}} \tag{31}$$

Equation 30 represents the general evaluation law for the fabric tensor $\mathbf{G}^{(\alpha)}$ of rank α. It is clear from this equation that once the loading evolution $d\boldsymbol{\sigma}$ is determined, one may evaluate the evolution of the fabric tensor using Eqs. 30 and 31. However, in this case, the tensor $\mathbf{L}^{(\alpha+2)}$ needs to be evaluated from Eq. 31, which is not an easy task. Finally, it should be noted that the evolution of fabric tensors of micro-cracks based on thermodynamic principles has been addressed by Yang et al. (1999, 2005) and Swoboda and Yang (1999). However, they employed a different approach than the one used here.

Evolution of Fabric Tensors for the Case of Uniaxial Tension

In this section, the evolution equation derived in section "Review of Fabric Tensors and the Mesoscopic Theory" for the special case of uniaxial tension is evaluated. In this case, the stress tensor and the stress increment tensor are presented as 3×1 vectors as follows:

$$\boldsymbol{\sigma} \equiv \begin{bmatrix} \sigma_{11} & 0 & 0 \end{bmatrix}^T \tag{32.1}$$

$$d\boldsymbol{\sigma} \equiv \begin{bmatrix} d\sigma_{11} & 0 & 0 \end{bmatrix}^T \tag{32.2}$$

where σ_{11} is the uniaxial stress and is the only nonzero component of the stress tensor $\boldsymbol{\sigma}$.

The evolution of the zero-rank fabric tensor $G^{(0)}$ is evaluated first. For this case, one sets $\alpha = 0$ in Eq. 30 to obtain

$$dG^{(0)} = \mathbf{L}^{(2)} : d\boldsymbol{\sigma} \tag{33}$$

where $\mathbf{L}^{(2)}$ is a second-rank tensor. Substituting Eq. 1.2 into Eq. 33, one obtains

$$dG^{(0)} = L_{11}^{(2)} d\boldsymbol{\sigma}_{11} \tag{34}$$

Next, the component $L_{11}^{(2)}$ is evaluated by using Eq. 31 to obtain

$$L_{11}^{(2)} = \frac{P^{(0)}}{1 - P^{(0)} \dfrac{\partial H^{(0)}}{\partial G^{(0)}}} \frac{\partial H^{(0)}}{\partial \sigma_{11}} \tag{35}$$

where $P^{(0)}$ is a scalar variable evaluated using Eq. 28 as follows:

$$P^{(0)} = -\frac{2}{\left(\dfrac{dL^{(0)}}{dl^{(0)}}\right)} \tag{36}$$

Substituting Eq. 36 into Eq. 35 and then substituting the result into Eq. 34, one finally obtains

$$dG^{(0)} = \left(\frac{-2}{\dfrac{dL^{(0)}}{dl^{(0)}} + 2\dfrac{\partial H^{(0)}}{\partial G^{(0)}}}\right)\frac{\partial H^{(0)}}{\partial \sigma_{11}} d\sigma_{11} \tag{37}$$

Equation 37 represents the general evolution law for the zero-rank fabric tensor $G^{(0)}$ for the case of uniaxial tension. It is clear that for the special case when $G^{(0)}$ is a constant, one obtains $dG^{(0)} = 0$ from Eq. 37.

Next, the evolution of the second-rank fabric tensor $\mathbf{G}^{(2)}$ for the case of uniaxial tension is evaluated. In this case, one sets $\alpha = 2$ and substitutes Eq. 1.2 into Eq. 30 to obtain

$$dG_{ij}^{(2)} = L_{ij11}^{(4)}\, d\sigma_{11} \tag{38}$$

where the component $L_{ij11}^{(4)}$ is obtained from Eq. 31 as follows:

$$L_{ij11}^{(4)} = P_{ijmn}^{(4)}\left(I_{mnpq}^{(4)} - \frac{\partial H_{mn}^{(2)}}{\partial H_{rs}^{(2)}} P_{rspq}^{(4)}\right)^{-1} \frac{\partial H_{pq}^{(2)}}{\partial \sigma_{11}} \tag{39}$$

where $I_{mnpq}^{(4)}$ is given by

$$I_{mnpq}^{(4)} = \frac{1}{2}\left(\delta_{mp}\delta_{nq} + \delta_{mq}\delta_{np}\right) \tag{40}$$

and δ_{ij} is the Kronecker delta.

Substituting for $\mathbf{P}^{(4)}$ from Eq. 28 into Eq. 39, and simplifying, and substituting the result into Eq. 38, one obtains

$$dG_{ij}^{(2)} = P_{ij\gamma\alpha}^{(4)}\left[\frac{\partial H_{\gamma\alpha}^{(2)}}{\partial G_{mn}^{(2)}}\left(P_{mnab}^{(4)} - \frac{\partial H_{ab}^{(2)}}{\partial G_{ef}^{(2)}} P_{efcd}^{(4)}\right)^{-1} \frac{\partial H_{cd}^{(2)}}{\partial \sigma_{11}} + \frac{\partial H_{\gamma\alpha}^{(2)}}{\partial \sigma_{11}}\right] d\sigma_{11} \tag{41}$$

where the component $P_{ijkl}^{(4)}$ may be readily obtained from Eq. 28. Equation 41 represents the general law of evolution of the second-rank fabric tensor $\mathbf{G}^{(2)}$ for the special case of uniaxial tension.

Next, it is reasonable to discuss how to evaluate the expression $\partial H_{ij}^{(2)}/\partial \sigma_{11}$ appearing in Eq. 41. In order to evaluate the partial derivative $\partial H_{ij}^{(2)}/\partial \sigma_{11}$, one needs an explicit expression for the tensor $\mathbf{H}^{(2)}$ in terms of the stress. This may be

done using the concept of effective stress in damage mechanics. For the case of uniaxial tension, the effective stress $\overline{\sigma}_{11}$ is given as follows:

$$\overline{\sigma}_{11} = \frac{\sigma_{11}}{1-\phi} \tag{42}$$

where ϕ is the scalar damage variable in one dimension. Let $M = 1/(1-\phi)$ where M is a scalar variable. Therefore, one will write Eq. 42 as follows:

$$\overline{\sigma}_{11} = M\sigma_{11} \tag{43}$$

It has been shown by Voyiadjis and Kattan (2006) that there is a relationship between the damage tensor (a scalar variable in this case) and fabric tensors. Specifically, the damage variable ϕ is a function of the second-rank fabric tensor $\mathbf{G}^{(2)}$, i.e., $\phi \equiv \phi(G_{ij}^{(2)})$ – see Eq. 53.

Using the hypothesis of elastic energy equivalence, one can write the following relation between the elastic strain components ε_{11} and its effective counterpart $\overline{\varepsilon}_{11}$:

$$\overline{\varepsilon}_{11} = N\varepsilon_{11} \tag{44}$$

where N is a scalar variable given as $N = 1 - \phi$ (see Voyiadjis and Kattan (1999) for details of this derivation). Based on the above discussion, it should be noted that both M and N are functions of the second-rank fabric tensor $\mathbf{G}^{(2)}$, i.e., $M \equiv M(G_{ij}^{(2)})$ and $N \equiv N(G_{ij}^{(2)})$.

Let U be the elastic strain energy in the deformed and damaged configuration and let $\overline{U}$ be its effective counterpart. This leads to $U = \frac{1}{2}\sigma_{ij}\varepsilon_{ij}$ and $\overline{U} = \frac{1}{2}\overline{\sigma}_{ij}\overline{\varepsilon}_{ij}$. For the case of uniaxial tension, these relations reduce to the following expressions:

$$U = \frac{1}{2}\sigma_{11}\varepsilon_{11} \tag{45.1}$$

$$\overline{U} = \frac{1}{2}\overline{\sigma}_{11}\overline{\varepsilon}_{11} \tag{45.2}$$

Substituting Eqs. 43 and 44 into Eq. 45.2, one obtains

$$\overline{U} = \frac{1}{2}MN\sigma_{11}\varepsilon_{11} \tag{46}$$

where the product $MN = 1$ in this case. This confirms the assumption of the hypothesis of elastic energy equivalence that $\overline{U} = U$.

Next, use is made of the elastic constitutive relation (Hooke's law) written as follows in both configurations:

$$\sigma_{11} = E\varepsilon_{11} \tag{47.1}$$

$$\overline{\sigma}_{11} = \overline{E}\overline{\varepsilon}_{11} \tag{47.2}$$

where E is Young's modulus and $\overline{E}$ is its effective counterpart. Using the hypothesis of elastic energy equivalence, i.e., $\overline{U} = U$, substituting equations (163) along with Eqs. 43 and 44 and Eq. 16, and solving for E, one obtains

$$E = \frac{N}{M}\overline{E} \tag{48}$$

where $\overline{E}$ is the constant Young's modulus of the undamaged material. It is clear from Eq. 48 that $E \equiv E(G_{ij}^{(2)})$.

Next, Eq. 47.1 is substituted into Eq. 45.1 and written in terms of σ_{11} to obtain

$$U = \frac{1}{2E}\sigma_{11}^2 \tag{49}$$

Substituting for E using Eq. 48 into Eq. 49, one obtains

$$U = \frac{M}{2NE}\sigma_{11}^2 \tag{50}$$

Since $M \equiv M(G_{ij}^{(2)})$ and $N \equiv N(G_{ij}^{(2)})$, it is concluded from Eq. 50 that $U \equiv U(G_{ij}^{(2)}, \sigma_{11})$.

Using Eq. 6.2 and substituting Eq. 50, one obtains

$$H_{ij}^{(2)} = \alpha\frac{\sigma_{11}^2}{2N\overline{E}}\frac{\partial M}{\partial G_{ij}^{(2)}} - \alpha\frac{M\sigma_{11}^2}{2\overline{E}N^2}\frac{\partial N}{\partial G_{ij}^{(2)}} \tag{51}$$

where α is a scalar variable that is a function of the density ρ of the material. Taking the derivative of Eq. 51, one finally obtains

$$\frac{\partial H_{ij}^{(2)}}{\partial \sigma_{11}} = \alpha\frac{\sigma_{11}}{N\overline{E}}\left(\frac{\partial M}{\partial G_{ij}^{(2)}} - \frac{M}{N}\frac{\partial N}{\partial G_{ij}^{(2)}}\right) \tag{52}$$

where it is clear that both $\partial M/\partial G_{ij}^{(2)}$ and $\partial N/\partial G_{ij}^{(2)}$ are functions of $G_{ij}^{(2)}$.

Utilizing the relation between the damage tensor and fabric tensors that is formulated by Voyiadjis and Kattan (2006), it is possible to derive the following expression for the scalar damage variable ϕ:

$$\phi = -G_{11}^{(2)} - \frac{\left(1+G_{11}^{(2)}\right)\left[(1-\nu)G_{11}^{(2)} - \nu^2 G_{22}^{(2)} - \nu^2 G_{33}^{(2)}\right]}{(1+\nu)(1-2\nu)} + \frac{\nu}{1+\nu}\left(\left(G_{12}^{(2)}\right)^2 + \left(G_{13}^{(2)}\right)^2\right) \tag{53}$$

where ν is Poisson's ratio. Equation 53 represents an explicit relation between the scalar damage variable ϕ and the components of the second-rank fabric tensor $G_{ij}^{(2)}$ for the case of uniaxial tension.

Next, use is made of the expression for M and N in terms of ϕ, i.e., $M = 1/(1-\phi)$ and $N = (1-\phi)$, and the partial derivatives $\partial M/\partial G_{ij}^{(2)}$ and $\partial N/\partial G_{ij}^{(2)}$ are evaluated using Eq. 53. Once these derivatives are evaluated, one substitutes the results into Eq. 52. Finally, the resulting expression for the partial derivative $\partial H_{ij}^{(2)}/\partial \sigma_{11}$ is substituted into Eq. 41 in order to obtain an explicit expression for the evolution of the second-rank fabric tensor $\mathbf{G}^{(2)}$. At this stage, use should be made of a computer algebra package like MAPLE or MATHEMATICA to obtain the final result.

Evolution of Length and Orientation of Micro-cracks

In this section, explicit expressions are developed for the evolution of the length and orientation of micro-cracks in terms of the evolution of the fabric tensors. In this regard, one may obtain some interesting results as shown below.

Consider first the zero-rank (scalar) fabric tensor $G^{(0)}$ given in Eq. 3.1. This relation may be rewritten in the following equivalent form:

$$G^{(0)} = \gamma \int_V r^3 dV \tag{54}$$

where γ is function of the volume V of the RVE. Taking the time derivative of both sides of Eq. 3.1, one obtains

$$\dot{G}^{(0)} = 3\gamma \int_V r^2 \dot{r} dV \tag{55}$$

where a superimposed dot indicates the time derivative. Differentiating Eq. 55 with respect to V and solving for $\dot{r}$, one obtains

$$\dot{r} = \frac{1}{3\gamma r^2} \frac{d\dot{G}^{(0)}}{dV} \tag{56}$$

Equation 56 represents the evolution law of the length of micro-cracks in terms of $d\dot{G}^{(0)}$, which may be obtained from Eq. 30. There is no evolution of the

orientation of micro-cracks in this case as this case deals with an isotropic distribution of micro-cracks with no anisotropy.

Next, the second-rank fabric tensor $\mathbf{G}^{(2)}$ as defined in Eq. 3.2 is considered. This expression may be rewritten in the following equivalent form:

$$G_{ij}^{(2)} = \gamma \int_V r^3 n_i n_j dV \tag{57}$$

Taking the time derivative of both sides of Eq. 57, one obtains

$$\dot{G}_{ij}^{(2)} = \gamma \int_V \left(3r^2 \dot{r} n_i n_j + r^3 \dot{n}_i n_j + r^3 n_i \dot{n}_j\right) dV \tag{58}$$

Differentiating Eq. 58 with respect to V and simplifying, one obtains

$$3r^2 \dot{r} n_i n_j + r^3 \dot{n}_i n_j + r^3 n_i \dot{n}_j = \frac{1}{\gamma} \frac{d\dot{G}_{ij}^{(2)}}{dV} \tag{59}$$

Multiply both sides of Eq. 59 by $\dot{n}_j$ and use the relations $n_j \dot{n}_j = 0$ (obtained by taking the time derivative of $n_j n_j = 1$) such that

$$r^3 n_i \dot{n}_j \dot{n}_j - \frac{1}{\gamma} \dot{n}_j \frac{d\dot{G}_{ij}^{(2)}}{dV} = 0 \tag{60}$$

Equation 60 may be rewritten as follows:

$$\left(r^3 n_i \dot{n}_j - \frac{1}{\gamma} \frac{d\dot{G}_{ij}^{(2)}}{dV} \right) \dot{n}_j = 0 \tag{61}$$

It is clear from Eq. 61 that two different solutions are possible. For the first solution of Eq. 61, one would have

$$r^3 n_i \dot{n}_j = \frac{1}{\gamma} \frac{d\dot{G}_{ij}^{(2)}}{dV} \tag{62}$$

Multiply both sides of Eq. 62 by n_i and solve for $\dot{n}_j$ to obtain

$$\dot{n}_j = \frac{1}{\gamma r^3} n_i \frac{d\dot{G}_{ij}^{(2)}}{dV} \tag{63}$$

Equation 63 represents the evolution law for the orientation of micro-cracks in this case. Substituting Eq. 63 into Eq. 59 while using the relations $n_i n_i = n_j n_j = 1$ and simplifying, one obtains

$$3r^2\dot{r}\,n_i\,n_j = -\frac{1}{\gamma}\frac{d\dot{G}_{ij}^{(2)}}{dV} \tag{64}$$

Multiply both sides of Eq. 64 by $n_i n_j$ and solve for $\dot{r}$ to obtain

$$\dot{r} = -\frac{1}{3\gamma r^2}\,n_i\,n_j\frac{d\dot{G}_{ij}^{(2)}}{dV} \tag{65}$$

Equation 65 represents the evolution law for the length of micro-cracks in this case. It can be shown that Eq. 65 reduces identically to zero. Substitute for the expression $\dot{n}_j$ of Eq. 63 into Eq. 65, and use the relation $n_j\dot{n}_j = 0$ to obtain $\dot{r} = 0$. Thus, the micro-crack lengths do not evolve in this case.

Next, one obtains the second solution of Eq. 61. In this case,

$$\dot{n}_j = 0 \tag{66}$$

Thus, the orientation of micro-cracks does not change in this solution. Substituting Eq. 66 into Eq. 59, we obtain

$$3r^2\dot{r}\,n_i\,n_j = \frac{1}{\gamma}\frac{d\dot{G}_{ij}^{(2)}}{dV} \tag{67}$$

Multiply both sides of Eq. 67 by $n_i n_j$ and solve for $\dot{r}$ to obtain

$$\dot{r} = \frac{1}{3\gamma r^2}\,n_i\,n_j\frac{d\dot{G}_{ij}^{(2)}}{dV} \tag{68}$$

We summarize the two solutions obtained as follows:

First Solution

$$\dot{n}_j = \frac{1}{\gamma r^3}n_i\frac{d\dot{G}_{ij}^{(2)}}{dV}$$

$$\dot{r} = 0$$

Second Solution

$$\dot{n}_j = 0$$

$$\dot{r} = \frac{1}{3\gamma r^2}\,n_i\,n_j\frac{d\dot{G}_{ij}^{(2)}}{dV}$$

Thus, it is clear from the above results that we obtain two completely different solutions. The distribution of micro-cracks may evolve according to one of these solutions. Alternatively, both solutions may apply to the same distribution but affect two different sets of micro-cracks. It is also clear from the first solution

that if the change in micro-crack orientation is nonzero, then the change in micro-crack length will be zero. This is a very interesting result. Also, it is noticed that if the change in micro-crack orientation is zero, then the change in micro-crack length will be nonzero. These results apply for a system of micro-cracks embedded within an RVE as a whole, not for a single micro-crack.

Next, we investigate the case of the fourth-rank fabric tensor $\mathbf{G}^{(4)}$ given in Eq. 3.3. This equation may be rewritten in the following equivalent form:

$$G^{(4)}_{ijkl} = \gamma \int_V r^3 n_i n_j n_k n_l dV \tag{69}$$

Taking the time derivative of both sides of Eq. 69, one obtains

$$\dot{G}^{(4)}_{ijkl} = \gamma \int_V \left(3r^2 \dot{r}\, n_i n_j n_k n_l + r^3 \dot{n}_i n_j n_k n_l + r^3 n_i \dot{n}_j n_k n_l + r^3 n_i n_j \dot{n}_k n_l + r^3 n_i n_j n_k \dot{n}_l\right) dV \tag{70}$$

Differentiating Eq. 70 with respect to V and rearranging the terms, one obtains

$$3r^2 \dot{r}\, n_i n_j n_k n_l + r^3 \dot{n}_i n_j n_k n_l + r^3 n_i \dot{n}_j n_k n_l + r^3 n_i n_j \dot{n}_k n_l + r^3 n_i n_j n_k \dot{n}_l = \frac{1}{\gamma} \frac{d\dot{G}^{(4)}_{ijkl}}{dV} \tag{71}$$

Multiply both sides of Eq. 71 by $\dot{n}_j$ and use $n_j \dot{n}_j = 0$ to obtain

$$r^3 n_i \dot{n}_j \dot{n}_j n_k n_l = \frac{1}{\gamma} \frac{d\dot{G}^{(4)}_{ijkl}}{dV} \tag{72}$$

Equation 72 may be rewritten as follows:

$$\left(r^3 n_i \dot{n}_j n_k n_l - \frac{1}{\gamma} \frac{d\dot{G}^{(4)}_{ijkl}}{dV}\right) \dot{n}_j = 0 \tag{73}$$

It is clear that Eq. 73 has two solutions. For the first solution, one obtains

$$r^3 n_i \dot{n}_j n_k n_l = \frac{1}{\gamma} \frac{d\dot{G}^{(4)}_{ijkl}}{dV} \tag{74}$$

Multiply both sides of Eq. 74 by $n_i n_k n_l$ and solve for $\dot{n}_j$ to obtain

$$\dot{n}_j = \frac{1}{\gamma r^3} n_i n_k n_l \frac{d\dot{G}^{(4)}_{ijkl}}{dV} \tag{75}$$

Equation 75 represents the evolution of the micro-crack orientation according to the first solution. Substituting Eq. 75 into Eq. 71 and using the relations $n_i n_i = n_j n_j = n_k n_k = n_l n_l = 1$, one obtains

$$3r^2 \dot{r}\, n_i n_j n_k n_l = -\frac{3}{\gamma} \frac{d\dot{G}^{(4)}_{ijkl}}{dV} \tag{76}$$

Multiply both sides of Eq. 76 by $n_i n_j n_k n_l$ and solve for $\dot{r}$ to obtain

$$\dot{r} = -\frac{1}{\gamma r^2} n_i n_j n_k n_l \frac{d\dot{G}^{(4)}_{ijkl}}{dV} \tag{77}$$

Equation 77 represents the evolution of micro-crack length according to the first solution. In deriving Eq. 77, the following symmetry of the fourth-rank fabric tensor is assumed: $G^{(4)}_{ijkl} = G^{(4)}_{jikl} = G^{(4)}_{ikjl} = G^{(4)}_{iljk}$. If the symmetry assumption does not hold, then Eq. 77 should be replaced by the following general expression for $\dot{r}$:

$$\dot{r} = -\frac{1}{3\gamma r^2} n_i n_j n_k n_l \left(\frac{d\dot{G}^{(4)}_{jikl}}{dV} + \frac{d\dot{G}^{(4)}_{ikjl}}{dV} + \frac{d\dot{G}^{(4)}_{iljk}}{dV} \right) \tag{78}$$

It can be shown that Eqs. 77 and 78 reduce identically to zero. Substitute for the expression $\dot{n}_j$ of Eq. 75 into Eqs. 77 and 78, and use the relation $n_j \dot{n}_j = 0$ to obtain $\dot{r} = 0$. Thus, the micro-crack lengths do not evolve in this case.

For the second solution of Eq. 73, one obtains

$$\dot{n}_j = 0 \tag{79}$$

Substituting Eq. 79 into Eq. 71, simplifying, and solving for $\dot{r}$, one obtains

$$\dot{r} = \frac{1}{3\gamma r^2} n_i n_j n_k n_l \frac{d\dot{G}^{(4)}_{ijkl}}{dV} \tag{80}$$

Finally, a summary of the results of the two solutions for the fourth-rank fabric tensor is presented as follows:

First Solution

$$\dot{n}_j = \frac{1}{\gamma r^3} n_i n_k n_l \frac{d\dot{G}^{(4)}_{ijkl}}{dV}$$

$$\dot{r} = 0$$

Second Solution

$$\dot{n}_j = 0$$

$$\dot{r} = \frac{1}{3\gamma r^2} n_i n_j n_k n_l \frac{d\dot{G}^{(4)}_{ijkl}}{dV}$$

Thus, it is clear from the above results that one obtains two completely different solutions. Alternatively, both solutions may apply to the same distribution but affect two different sets of micro-cracks. It is also clear from the first solution that if the change in micro-crack orientation is nonzero, then the change in micro-crack length will be zero. This is a very interesting result. Furthermore, it is noted that if the change in micro-crack orientation is zero, then the change in micro-crack length will be nonzero. These results apply for a system of micro-cracks embedded within an RVE as a whole, not for a single micro-crack.

Based on the above solutions, one makes the following observations:

1. The evolution of micro-crack orientation $\dot{\mathbf{n}}$ is proportional to $1/r^3$. Thus, longer micro-cracks undergo smaller rotations, while shorter micro-cracks undergo larger rotations.
2. The evolution of micro-crack length $\dot{r}$ is proportional to $1/r^2$. Thus, longer micro-cracks undergo smaller increases in length, while shorter micro-cracks undergo larger increases in length.
3. In a system of micro-cracks where micro-crack rotations vanish, the following conditions must be satisfied:

$$\dot{G}^{(2)}_{ij} n_i = 0, \quad \text{all } j \tag{81.1}$$

$$\dot{G}^{(4)}_{ijkl} n_i n_k n_l = 0, \quad \text{all } j \tag{81.2}$$

4. In a system of micro-cracks where the changes in micro-crack lengths vanish, the following condition must be satisfied:

$$\dot{G}^{(2)}_{ij} n_i n_j = 0 \tag{82.1}$$

$$\dot{G}^{(4)}_{ijkl} n_i n_j n_k n_l = 0 \tag{82.2}$$

5. In a system of micro-cracks, one may have a set of micro-cracks with $\dot{r} = 0$, $\dot{\mathbf{n}} \neq \mathbf{0}$, while another set of micro-cracks has $\dot{r} \neq 0$, $\dot{\mathbf{n}} = \mathbf{0}$. These two sets of micro-cracks should not be overlapping as shown in Fig. 5.

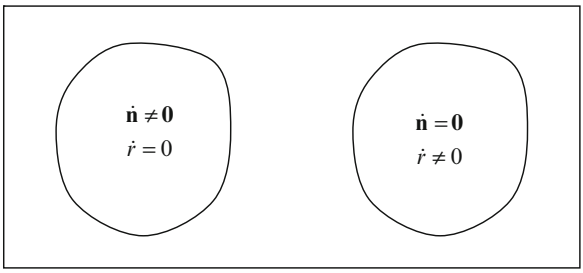


Fig. 5 RVE with two sets of micro-cracks (Reprinted with permission from G. Voyiadjis, P. Kattan, Mech. Res. Commun. 34, 2 (2007))

Summary and Conclusions

This work presents an exposition of thermodynamic concepts and derivation of the equations of damage evolution. This is performed in a mathematically consistent manner that is based on sound thermodynamic principles. This work is very important in forming the basis for future work in this area including visualization and finite element implementation with application to composite materials.

The fabric tensors used represent a system of micro-cracks which is embedded into an RVE. Both the orientation and length of micro-cracks are included in the definition of the fabric tensors. Explicit evolution equations are derived for both the micro-crack orientation and micro-crack length. It is concluded that micro-crack orientations and lengths cannot evolve simultaneously. However, two sets of micro-cracks may be considered where each parameter (length or orientation) evolves separately, while the other parameter's evolution vanishes.

References

A. Cauvin, R. Testa, Damage mechanics: basic variables in continuum theories. Int. J. Solids Struct. **36**, 747–761 (1999)

J.L. Chaboche, Continuous damage mechanics – a tool to describe phenomena before crack initiation. Nucl. Eng. Des. **64**, 233–247 (1981)

C. Chow, J. Wang, An anisotropic theory of elasticity for continuum damage mechanics. Int. J. Fract. **33**, 3–16 (1987)

B. Coleman, M. Gurtin, Thermodynamics with internal state variables. J. Chem. Phys. **47**(2), 597–613 (1967)

I. Doghri, *Mechanics of Deformable Solids: Linear and Nonlinear, Analytical and Computational Aspects* (Springer, Berlin, 2000)

D. Hayhurst, Creep rupture under multiaxial states of stress. J. Mech. Phys. Solids **20**, 381–390 (1972)

Q. He, A. Curnier, A more fundamental approach to damaged elastic stress–strain relations. Int. J. Solids Struct. **32**(10), 1433–1457 (1995)

K. Kanatani, Distribution of directional data and fabric tensors. Int. J. Eng. Sci. **22**(2), 149–164 (1984a)

K. Kanatani, Stereological determination of structural anisotropy. Int. J. Eng. Sci. **22**(5), 531–546 (1984b)

P.I. Kattan, G.Z. Voyiadjis, A coupled theory of damage mechanics and finite strain elasto-plasticity – part I: damage and elastic deformations. Int. J. Eng. Sci. **28**(5), 421–435 (1990)

P.I. Kattan, G.Z. Voyiadjis, A plasticity-damage theory for large deformation of solids – part II: applications to finite simple shear. Int. J. Eng. Sci. **31**(1), 183–199 (1993)
P.I. Kattan, G.Z. Voyiadjis, Decomposition of damage tensor in continuum damage mechanics. J. Eng. Mech., ASCE **127**(9), 940–944 (2001a)
P.I. Kattan, G.Z. Voyiadjis, *Damage Mechanics with Finite Elements: Practical Applications with Computer Tools* (Springer, Berlin, 2001b)
H. Lee, K. Peng, J. Wang, An anisotropic damage criterion for deformation instability and its application to forming limit analysis of metal plates. Eng. Fract. Mech. **21**, 1031–1054 (1985)
J. Lemaitre, Evaluation of dissipation and damage in metals subjected to dynamic loading, in *Proceedings of I.C.M. 1*, Kyoto, Japan, 1971
J. Lemaitre, How to use damage mechanics. Nucl. Eng. Des. **80**, 233–245 (1984)
J. Lemaitre, J.L. Chaboche, *Mechanics of Solid Materials* (Cambridge University Press, London, 1990)
V. Lubarda, D. Krajcinovic, Damage tensors and the crack density distribution. Int. J. Solids Struct. **30**(20), 2859–2877 (1993)
M. Oda, Geometry of discontinuity and its relation to mechanical properties of discontinuous materials, in *IUTAM Conference on Deformation and Failure of Granular Materials*, Delft, 31 Aug–3 Sep, 1982
C. Papenfuss, P. Van, W. Muschik, Mesoscopic theory of microcracks. Arch. Mech. **55**(5/6), 481–500 (2003)
Y. Qiang, L. Zhongkui, L.G. Tham, An explicit expression of second-order fabric tensor dependent elastic compliance tensor. Mech. Res. Commun. **28**(3), 225–260 (2001)
Y. Rabotnov, in *Creep Rupture*, ed. by M. Hetenyi, W.G. Vincenti. Proceedings, Twelfth International Congress of Applied Mechanics. Stanford, 1968 (Springer, Berlin, 1969), pp. 342–349
M. Satake, Fabric tensors in granular materials, in *IUTAM Conference on Deformation and Failure of Granular Materials*, Delft, Aug 31–Sept 3, 1982, pp. 63–68
G. Swoboda, Q. Yang, An energy-based damage model of geomaterials – II: deduction of damage evolution laws. Int. J. Solids Struct. **36**, 1735–1755 (1999)
G.Z. Voyiadjis, P.I. Kattan, A coupled theory of damage mechanics and finite strain elasto-plasticity – part II: damage and finite strain plasticity. Int. J. Eng. Sci. **28**(6), 505–524 (1990)
G.Z. Voyiadjis, P.I. Kattan, A plasticity-damage theory for large deformation of solids – part I: theoretical formulation. Int. J. Eng. Sci. **30**(9), 1089–1108 (1992)
G.Z. Voyiadjis, P.I. Kattan, On the symmetrization of the effective stress tensor in continuum damage mechanics. J. Mech. Behav. Mater. **7**(2), 139–165 (1996)
G.Z. Voyiadjis, P.I. Kattan, *Advances in Damage Mechanics: Metals and Metal Matrix Composites* (Elsevier Science, Amsterdam, 1999)
G.Z. Voyiadjis, P.I. Kattan, Damage mechanics with fabric tensors. Mech. Adv. Mater. Struct. **13**, 285–301 (2006)
G.Z. Voyiadjis, P.I. Kattan, Evolution of fabric tensors in damage mechanics of solids with micro-cracks: part I – theory and fundamental concepts. Mech. Res. Commun. **34**, 145–154 (2007a)
G.Z. Voyiadjis, P.I. Kattan, Evolution of fabric tensors in damage mechanics of solids with micro-cracks: part II – evolution of length and orientation of micro-cracks with an application to uniaxial case. Mech. Res. Commun. **34**, 155–163 (2007b)
Q. Yang, W.Y. Zhou, G. Swoboda, Micromechanical identification of anisotropic damage evolution laws. Int. J. Fract. **98**, 55–76 (1999)
Q. Yang, X. Chen, L.G. Tham, Relationship of crack fabric tensors of different orders. Mech. Res. Commun. **31**, 661–666 (2004)
Q. Yang, X. Chen, W.Y. Zhou, On microscopic thermodynamic mechanisms of damage evolution laws. Int. J. Damage Mech. **14**, 261–293 (2005)
P. Zysset, A. Curnier, An alternative model for anisotropic elasticity based on fabric tensors. Mech. Mater. **21**, 243–250 (1995)
P. Zysset, A. Curnier, A 3D damage model for trabecular bone based on fabric tensors. J. Biomech. **29**(12), 1549–1558 (1996)

Section II

Damage for Disordered Materials

Failure Mechanics of Geomaterials

5

Florent Prunier, François Nicot, Richard Wan, Jérôme Duriez, and Félix Darve

Contents

F. Prunier (✉)
INSA de Lyon, LGCIE, Villeurbanne, France
e-mail: florent.prunier@insa-lyon.fr

F. Nicot
IRSTEA, Geomechanics group, ETNA, Grenoble, France
e-mail: francois.nicot@irstea.fr

R. Wan
Department of Civil Engineering, University of Calgary, Calgary, AB, Canada
e-mail: wan@ucalgary.ca

J. Duriez • F. Darve
Grenoble INP, UJF, CNRS, Grenoble, France
e-mail: jerome.duriez@3sr-grenoble.fr; felix.darve@grenoble-inp.fr; felix.darve@3sr-grenoble.fr

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_21

Abstract

Geomaterials represent an important class of dissipative materials whose mechanical behavior is pressure, density, and fabric dependent. This constitutive characteristic together with the discrete particulate nature of the material leads to the manifestation of a rich variety of failure modes whose precise understanding is elusive within standard failure theories. The present chapter attempts to clarify this issue by invoking plasticity/damage phenomena in geomaterials and exploits their non-associated character in relation to rate-independent irreversible strains. The second-order work criterion provides a basic framework within which failure can be systematically treated as a divergence instability that leads to various forms, including localized and diffuse modes. This new interpretation considers the existence of a bifurcation domain and so-called instability cones whose generators denote the range of loading directions in stress space along which the material response is potentially unstable. As additional important characteristics, macroscopic failure is found to occur with an outburst of kinetic energy with the proper load control parameter in place, as demonstrated in discrete element computations. Finally, the failure analysis of in situ boundary value problems as in a rock and a soil slope is presented using the second-order work.

Introduction: Main Features of Failure in Geomaterials

Failure in geomaterials (soil, rock, and concrete) is a very intricate question in comparison to other solid materials, probably because for geomaterials the plastic limit condition, yield surface, and plastic potential are all mean pressure dependent. All these surfaces have indeed a conical shape in the stress space. In relation with this mean pressure dependency, the plastic strains have a non-associated character. Thus, the yield surface (characterized roughly speaking by a cohesion and a friction angle) does not coincide with the plastic potential (characterized by a dilatancy angle). For an associated material, both friction and dilatancy angles must be equal, while these differ usually by 25–30° for granular media.

As a first approximation, geomaterial behavior can be considered as largely rate independent, which means that this behavior is described by an elastoplastic or elasto-damageable constitutive relation represented by a fourth-order constitutive tensor linking incremental strains (or strain rates) and incremental stresses (or stress rates). Then, by considering the six-dimensional associated stress and strain spaces,

a 6-by-6 constitutive matrix is used to relate incremental strains and stresses. A direct consequence of non-associativity of plasticity is the nonsymmetry of the above constitutive matrix. As such, the possible singularities of this matrix (linked to failure as it will be seen later on) can be of very different kinds in comparison with the case of a symmetric matrix corresponding to the associated plasticity case. A typical aspect of failure in geomaterials appears here as a direct consequence of this nonsymmetry: there is indeed a rich variety of failure modes each differing in character involving either localized or diffuse deformation fields, among others.

An important but fundamental question has to be asked: what is failure? The usual definition is that a failure state is reached when a limit stress state is reached by the material. If some mixed stress–strain loading paths are considered, this definition has to be extended to include other types of limit states such as those constituted by a limiting volume expansion for dilatant materials in the so-called triaxial loading paths with a constant deviatoric stress in geomechanics (Darve et al. 2007; Daouadji et al. 2010, 2011; Laouafa et al. 2011). More precisely, when a failure state is reached, the failure can be effective or not depending on the current control parameter. As a classical example, when the control parameter is a strain rate, the limit stress state can be overcome without any specific change for the body, while if the control parameter is a stress rate, the body suddenly enters into a dynamic regime with a burst of kinetic energy and large uncontrollable strains.

From a theoretical point of view, it is possible to say that an effective failure corresponds to a bifurcation point, because of the sudden change in the strain regime from quasi-static deformations to dynamic ones. It is also characterized by a loss of uniqueness, since the deformations are no more defined in a unique way and depend on some imperfections. Eventually effective failure is also an unstable state, since a small perturbation of the body in a failure state induces large material responses. Of course, the converse of the above argument is not generally true: bifurcations, losses of uniqueness, or instabilities are not always related to an effective failure in solid mechanics. However, an efficient way to investigate the failure states and the failure modes of a given geomaterial is to first consider its unstable states, and then check if these instabilities can lead to effective failures and, if yes, for which conditions. This methodology to analyze failure will be the central theme of this chapter.

Indeed, a general criterion for all types of divergence instabilities has been proposed by Hill (1958), the so-called second-order work criterion, which corresponds to the loss of positive definitiveness of the constitutive matrix and is a necessary condition for instability. This can be reworded as to there is at least one stress direction where the scalar product of the incremental stress by the incremental strain is nil or negative. For a nonsymmetric matrix, the loss of positive definitiveness (linked to the vanishing value of the determinant of the symmetric part of the constitutive matrix) appears before the singularity of the matrix itself (linked to the vanishing value of the determinant of the matrix itself). Thus, there is a whole failure domain in the stress space (Darve et al. 2004) and not only a single plastic limit surface as for associated plastic materials. More recently, two new developments have allowed clarifying two important physical aspects of the

second-order work. First, the link between bursts of kinetic energy and negative values of second-order work has been established (Nicot and Darve 2007, 2011; Nicot et al. 2009, 2012a). Second, for elastic nonconservative systems, the second-order work appears as the lower envelope of all possible instability curves (Lerbet et al. 2012; Nicot et al. 2011), corresponding to all types of divergence instabilities.

This chapter is organized as follows. First the link between second-order work criterion and burst of kinetic energy is recalled. At the grain level, the second-order work takes a discrete form linking increment of force and increment of displacement at the contact point. The relation between the sum of all discrete second-order works calculated at every intergranular contact and the macroscopic second-order work is established and discussed. Then, the features of second-order work as a quadratic form associated to the current incremental stress or the current incremental strain are illustrated by phenomenological rate-independent constitutive relations for the general 3D multiaxial-loading case. The bifurcation domain is plotted as well as the "instability cones" for two elastoplastic relations (one is incrementally piecewise linear and the other fully incrementally nonlinear). In the third part, a careful check of these various aspects (link between second-order work and kinetic energy, bifurcation domain, instability cones, features of the effective failure, influence of the control mode, influence of a perturbation) is performed by considering direct numerical simulations using a discrete element method. Then this methodology is applied in the fourth section to the question of the stability of rock joints. A rate-independent relation for rock joints is introduced and a specific expression of second-order work is considered to analyze failure. Eventually, the finite element method is employed to simulate triaxial tests, exhibiting localized and diffuse failure modes. An example of modeling of an in situ landslide in Italy (Petacciato landslide) is finally presented to illustrate the ability of the second-order work criterion to describe a failure mode of the diffuse type using the finite element method.

A General Criterion for Failure by Divergence Instabilities

Kinetic Energy and Second-Order Work

Let us consider a material system, subjected to external forces. The mechanical (stress–strain) state of this system, after a given loading history, is reputed (mechanically) unstable if the kinetic energy of the system may increase under the effects of infinitesimal loading (disturbance). In particular, if the system is initially at rest, in equilibrium under external forces, an increase in kinetic energy means that there is a transition from a quasi-static regime towards a dynamical regime. This transition is basically a bifurcation and is clearly associated to a failure that can be either localized or diffuse, according to the pattern of the kinematic field developing during this transition. The failure is localized when a localization pattern develops or is diffuse if the kinematic field remains chaotic, without specific localization (Nicot and Darve 2011).

In what follows, the conditions and the growth mode of the kinetic energy are studied. For a material system made up of a volume V_0, initially in a configuration C_0 of boundary (Γ_0), in equilibrium at time t under a prescribed external loading, it was established that the second-order time differentiation of the kinetic energy (at time t) written in Lagrangian formulation (Nicot et al. 2007; Nicot and Darve 2007; Nicot et al. 2012a, b) is

$$\begin{aligned} \delta^2 E_c(t) &= W_2^{\text{ext}} - W_2^{\text{int}} \\ W_2^{\text{ext}} &= \int_{\Gamma_o} \delta f_j \ \delta u_i \ dS_o; \ W_2^{\text{int}} = \int_{V_o} \delta \Pi_{ij} \ \frac{\partial(\delta u_i)}{\partial X_j} \ dV_o \end{aligned} \tag{1}$$

where W_2^{ext} and W_2^{int} denote the external second-order work and internal second-order work, respectively. In these expressions, $\overline{\overline{\Pi}}$ denotes the Piola–Kirchhoff stress tensor of the first kind and $\vec{f}$ the current forces applied to the initial (reference) configuration. $\delta\vec{u}$ denotes the current incremental displacement of the material points initially located at the position $\vec{X}$. In addition, the second-order Taylor's expansion of kinetic energy reads

$$E_c(t+\Delta t) = E_c(t) + \Delta t \ \dot{E}_c(t) + \frac{(\Delta t)^2}{2}\ddot{E}_c(t) + 0(\Delta t)^3 \tag{2}$$

Noting that $E_c(t) = \int_{V_o} \rho_o \ \left\|\dot{\vec{u}}\right\|^2 \ dV_o$ and $\dot{E}_c(t) = \int_{V_o} \rho_o \ \dot{\vec{u}} \cdot \ddot{\vec{u}} \, dV_o$, where ρ_o is the density of the material in the initial configuration at point $M\left(\vec{X}\right)$, and since the system is in an equilibrium state at time t, then $\dot{E}_c(t) = 0$. Thus, ignoring third-order terms, Eq. 2 reads

$$\delta^2 E_c(t) = 2(E_c(t+\delta t) - E_c(t)) \tag{3}$$

Combining Eqs. 1 and 3 gives

$$E_c(t+\delta t) - E_c(t) = \frac{1}{2} \ \left(W_2^{\text{ext}} - W_2^{\text{int}}\right) \tag{4}$$

Consequently, the increase in kinetic energy for a material system initially in equilibrium (at rest) equals to the difference between the external second-order work (that involves the displacements and the forces acting on the boundary) and the so-called internal second-order work (built from the internal stress and strain acting at each point of the system). Finally, the external second-order work is related to the external loading and can be controlled by an external user. In contrast, the internal second-order work is intimately related to the constitutive behavior of the material and is imposed by the material independently of the external user

(Nicot et al. 2012a). As a result, any increase in kinetic energy of the system stems from a conflict between the internal forces (resulting from the internal stress) applying to the internal side of the boundary and the external forces (imposed by the operator or any other external action) applying to the external side of the boundary (Nicot et al. 2012a).

According to Eq. 4, both external and internal second-order works are equal in quasi-static conditions, since $E_c(t + \delta t) - E_c(t) = 0$ (there is no increase in kinetic energy). Particularly of interest is the situation where the loading applied gives rise to a negative value of the internal second-order work. It can be shown that such a situation corresponds to the existence of a limit state (possibly in a generalized strain–stress space; Nicot et al. 2009, 2011, 2012a). Some components of the internal loading pass through a peak and then follow a descending branch, which prevent these components from taking higher values. On the contrary, higher values of the external loads applied to the system can be imposed. The resulting conflict between both internal and external loads induces an increase in kinetic energy, indicating a transition from a quasi-static to a dynamical regime. It is worth noting that this situation is closely related to the existence of a generalized limit state, detected by the vanishing of the internal second-order work. After the peak, along the descending branch, the internal second-order work takes negative values.

As a conclusion, as far as divergence instabilities are concerned, detecting whether the internal second-order work can vanish makes sense. The existence of nil or negative values of the second-order works stands as one of the basic conditions giving rise to the occurrence of a divergence instability.

Micromechanically Based Formulation

The problem is now specialized to granular materials, with an attempt to relate the macroscopic formulation of the internal second-order work to microstructural variables. For this purpose, a homogeneous volume of granular material is comprised of N grains. "p" will denote indiscriminately the grain (as a body) or enumerate a particular grain within the assembly such that $1 \leq p \leq N$. The shape of each grain "p" is arbitrary. The total number of contacts at time t within the assembly is denoted N_c. The system is assumed to be in equilibrium at a given time t under a prescribed external loading. Depending on the type of loading control, each grain "p" belonging to the boundary ∂V of the considered volume is subjected to either a displacement (kinematic control) or an external force $\vec{f}^{\,\mathrm{ext},p}$ (static control), possibly zero.

The second-order work, in its basic formulation, involves both incremental stress and strain. The stress within a granular assembly expresses the transmission of forces in granular materials. The transmission of forces in granular materials operates at contacts of adjoining grains, thereby resulting in a macroscopic average stress at the grain assembly level. The stress tensor in such a body of volume V in equilibrium under external forces $\vec{f}^{\,\mathrm{ext},p}$ applied to the boundary particles "p" of

position $\vec{x}^p$ can be defined by the classical Love–Weber formula (Love 1927; Mehrabadi et al. 1982), i.e.,

$$\sigma_{ij} = \frac{1}{V} \sum_{p \in \partial V} f_i^{\text{ext},p} \; x_j^p \tag{5}$$

The above expression can be transformed as follows, by accounting for the interparticle contact forces $\vec{f}^c$ (Nicot et al. 2012c):

$$\sigma_{ij} = \frac{1}{V} \sum_{c=1}^{N_c} f_i^c \; l_j^c + \frac{1}{V} \sum_{p \in V} f_i^p \; x_j^p \tag{6}$$

where $\vec{l}^c$ is the branch vector relating the centers of contacting particles and $\vec{f}^p$ denotes the resultant force applied to the particle "p." In the absence of inertial effects or when all particles are in static equilibrium, the second term in Eq. 6 vanishes. However, this term may subsist in the presence of internal dynamical effects that arise from local force imbalances, even if the whole granular body may be in equilibrium macroscopically.

In the Eulerian formulation represented in Eq. 6, the contact forces, the branch vectors, the location of each particle, and the volume of the specimen are bound to evolve over a given loading history from an initial configuration C_0 ($\vec{f}_0^c$, $\vec{l}_0^c$, $\vec{x}_0^c$, V_0). Thus, referring to the initial configuration, the analogous form of the stress tensor in Lagrangian description is

$$\Pi_{ij} = \frac{1}{V_0} \sum_{p,q} f_i^c \; l_{0,j}^c + \frac{1}{V_0} \sum_{p \in V_0} f_i^p \; x_{0,j}^p \tag{7}$$

The Lagrangian formulation given in Eq. 7 can be readily differentiated, then providing the following expression of the internal second-order work (see Nicot et al. 2012c, for more details):

$$W_2 = \sum_{p,q} \delta f_i^c \; \delta l_i^c + \sum_{p \in V} \delta f_i^p \; \delta x_i^p \tag{8}$$

As specified in Nicot et al. (2012c), the creation or the deletion of contacts is accounted for in this approach. The symbol $\sum_{p,q}$ denotes the summation over p and q varying over $[1, N]$ with $q \leq p$, and c refers to the contacting pair (p, q).

It is worth noting that in the absence of incremental unbalanced force and in quasi-static regime, Eq. 8 simplifies into

$$W_2 = \sum_{p,q} \delta f_i^c \; \delta l_i^c \tag{9}$$

Equation 9 expresses the internal second-order work from micromechanical variables, namely, the contact forces existing between contacting granules and

the branch vectors joining these granules. The attempt of such a formulation is to go down to the microscopic scale and to try to elucidate what are the basic microstructural origins giving rise to the vanishing of the internal second-order work and therefore what are the microstructural contexts prone to instabilities.

The Second-Order Work Criterion, Features, and Illustrative 3D Examples (Multiaxial Loading)

As mentioned in the previous section, determining the sign of the second-order work allows the detection of material instability by divergence. Neglecting geometrical effects and assuming small strains, the internal second-order work takes the following expression:

$$W_2 = \int_{V_o} \delta\sigma_{ij}\delta\varepsilon_{ij} dV_o \tag{10}$$

where $\delta\sigma_{ij}$ is the incremental Cauchy stress tensor and $\delta\varepsilon_{ij}$ the incremental small strain tensor. In a local form the stability criterion is expressed as

$$\forall\left(\delta\sigma_{ij}, \delta\varepsilon_{ij}\right) \quad \delta W_2 = \delta\sigma_{ij}\delta\varepsilon_{ij} > 0. \tag{11}$$

In other words, when δW_2 is positive, the specimen is stable, else potentially unstable. It will be seen further that instability occurrence (or the kinetic energy growth) depends on the control loading parameter when δW_2 is negative or zero.

General Equation of Local Second-Order Work Criterion

In the following of this section, only rate-independent materials are considered. In this case the constitutive relation takes the general following form:

$$F^h_{ijkl}\left(\delta\sigma_{ij}, \ \delta\varepsilon_{kl}\right) = 0 \tag{12}$$

where F^h is a nonlinear function and h a memory parameter like a hardening parameter. Along the actual loading direction function F^h can be linearized ("directional linearization"), and Eq. 12 can take the simple form:

$$\delta\varepsilon_{ij} = N_{ijkl}\delta\sigma_{kl} \tag{13}$$

Using dual representations of tensors where a second-order tensor can be a six-dimensional vector and a fourth-order tensor, a matrix of rank six, the matrix N denoting the tangent elastoplastic operator along the current loading direction is introduced. As such, when substituting Eq. 13 into Eq. 11, the boundary between positive and negative values of δW_2 is given by

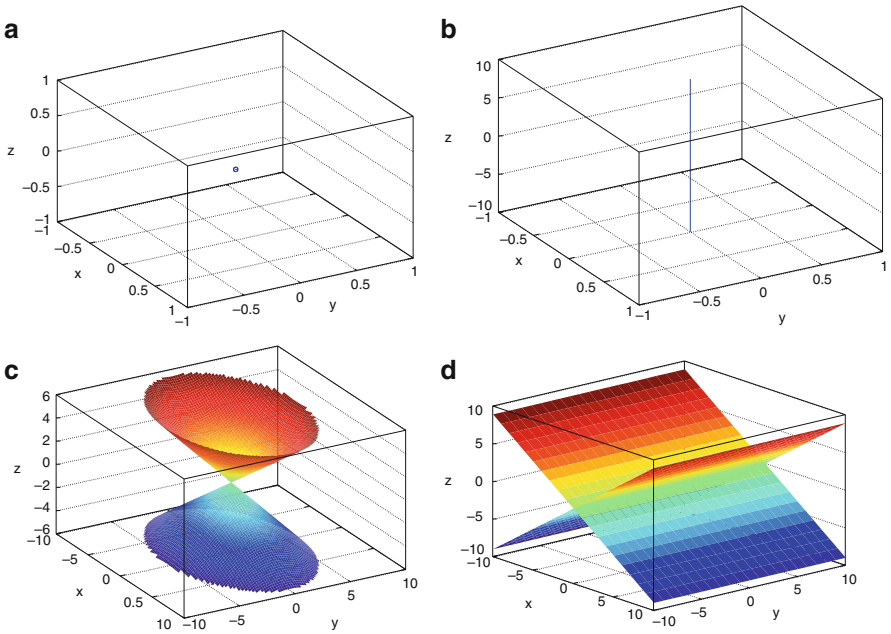


Fig. 1 Graphical representation of solutions of Eq. 15: empty solution, straight line, elliptical cone, and two planes intersection

$$\delta\sigma_i N_{ij}\delta\sigma_j = \delta\sigma_i N^s_{ij}\delta\sigma_j = 0 \tag{14}$$

where N^s is the symmetrical part of N. Developing Eq. 14 in the case of orthotropic incrementally piecewise linear relation leads to the general equation of an elliptical cone in the three-dimensional space of principal incremental stress.

$$\frac{\delta\sigma_1^2}{E_1}+\frac{\delta\sigma_2^2}{E_2}+\frac{\delta\sigma_3^2}{E_3}-\left(\frac{\nu_1^2}{E_1}+\frac{\nu_2^1}{E_2}\right)\delta\sigma_1\delta\sigma_2-\left(\frac{\nu_3^2}{E_3}+\frac{\nu_2^3}{E_2}\right)\delta\sigma_2\delta\sigma_3-\left(\frac{\nu_1^3}{E_1}+\frac{\nu_3^1}{E_3}\right)\delta\sigma_3\delta\sigma_1=0 \tag{15}$$

Geometrical representations of the solutions of Eq. 15 in terms of loading directions are displayed in Fig. 1.

If the solution of Eq. 15 is empty, no instability can develop. In the other cases, instability can develop if the loading path has a direction included inside directions given by the solution of Eq. 15. It is to be noted that Eq. 15 can be used practically only with an incrementally piecewise linear model, and solutions have to be truncated in parts of the space where the constitutive relation is linear. These particular parts of the space are also denoted as tensorial zones (Darve and Labanieh 1982). Nevertheless, most of rate-independent models are piecewise linear. In the case of fully incrementally nonlinear models, nonempty solutions of Eq. 15 reduce necessarily to a nonelliptical cone gathering straight lines. An

illustration of this proposal is given in the next subsection using the octo-linear model (8L model) and incrementally nonlinear model of Darve (INL2 model) (Darve and Labanieh 1982).

Illustration of Instability Cones Using Darve Model

Constitutive relations developed by Darve do not rely on classical concepts of elastoplasticity and therefore assumptions of (a) strain decomposition into an elastic and a plastic part, (b) existence of an elastic limit, and (c) existence of a flow rule are not needed.

In order to describe the nonlinear behavior of geomaterials, an incrementally nonlinear relation of second order is used and written in principal axes as follows:

$$\begin{Bmatrix} \delta\varepsilon_1 \\ \delta\varepsilon_2 \\ \delta\varepsilon_3 \end{Bmatrix} = \frac{1}{2}\left[\underline{\underline{N}}^{+} + \underline{\underline{N}}^{-}\right]\begin{Bmatrix} \delta\sigma_1 \\ \delta\sigma_2 \\ \delta\sigma_3 \end{Bmatrix} + \frac{1}{2\|\underline{\delta\sigma}\|}\left[\underline{\underline{N}}^{+} + \underline{\underline{N}}^{-}\right]\begin{Bmatrix} \delta\sigma_1^2 \\ \delta\sigma_2^2 \\ \delta\sigma_3^2 \end{Bmatrix} \tag{16}$$

with

$$\underline{\underline{N}}^{\pm} = \begin{bmatrix} \frac{1}{E_1^{\pm}} & -\frac{\nu_{21}^{\pm}}{E_2^{\pm}} & -\frac{\nu_{31}^{\pm}}{E_3^{\pm}} \\ -\frac{\nu_{12}^{\pm}}{E_1^{\pm}} & \frac{1}{E_2^{\pm}} & -\frac{\nu_{32}^{\pm}}{E_3^{\pm}} \\ -\frac{\nu_{13}^{\pm}}{E_1^{\pm}} & -\frac{\nu_{23}^{\pm}}{E_2^{\pm}} & \frac{1}{E_3^{\pm}} \end{bmatrix} \tag{17}$$

Coefficients E_i^{+} and ν_{ij}^{+} are defined on generalized triaxial loading paths when $\delta\sigma_i > 0$ and, respectively, E_i^{-} and ν_{ij}^{-} when $\delta\sigma_i < 0$. For $\delta\sigma_i = 0$ it can be verified that the relation is continuous (Gudehus 1979). More detailed information about this constitutive model and, in particular, how the tangent moduli and Poisson's ratios evolve with stress–strain history can be found in Darve et al. (1995). In one dimension, this relationship is piecewise linear. By extension, the octo-linear model (eight tensorial zones) is defined in the following with the previous notations:

$$\begin{Bmatrix} \delta\varepsilon_1 \\ \delta\varepsilon_2 \\ \delta\varepsilon_3 \end{Bmatrix} = \frac{1}{2}\left[\underline{\underline{N}}^{+} + \underline{\underline{N}}^{-}\right]\begin{Bmatrix} \delta\sigma_1 \\ \delta\sigma_2 \\ \delta\sigma_3 \end{Bmatrix} + \frac{1}{2}\left[\underline{\underline{N}}^{+} + \underline{\underline{N}}^{-}\right]\begin{Bmatrix} |\delta\sigma_1| \\ |\delta\sigma_2| \\ |\delta\sigma_3| \end{Bmatrix} \tag{18}$$

It can be seen that the eight tensorial zones are bounded by the intersection of the three planes of respective expression: ($\delta\sigma_1 = 0$, $\delta\sigma_2 = 0$, $\delta\sigma_3 = 0$). Figure 2 shows instability cones obtained along a drained triaxial path with both models (octo-linear and incrementally nonlinear) of Darve on a dense Hostun sand. According to the incrementally nonlinear model, a numerical procedure has been used (Prunier

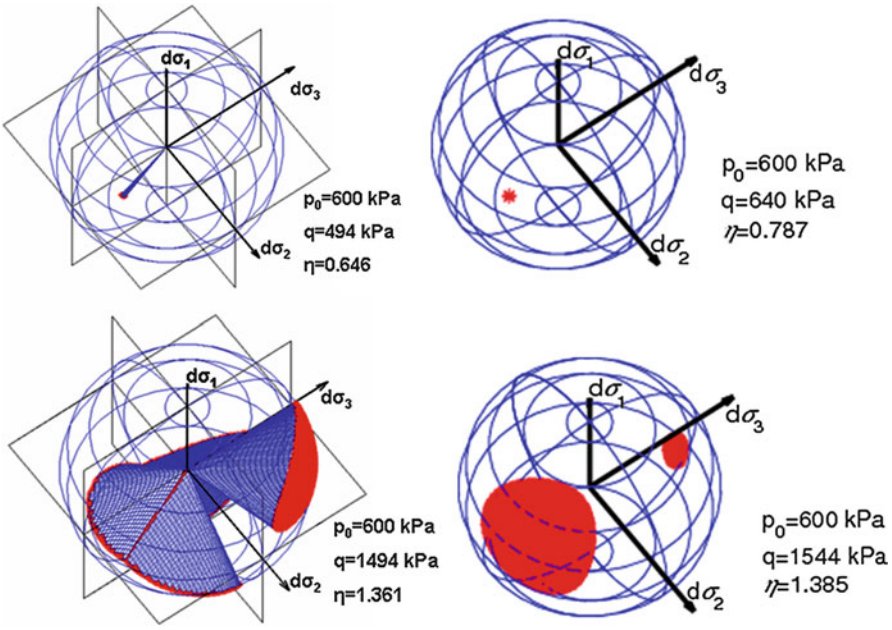


Fig. 2 3D instability cones given by octo-linear model of Darve (*left* side) and by the nonlinear model of Darve (*right* side). p_0 is the initial confinement pressure, q the actual stress deviator, and $\eta = q/p$

et al. 2009a; Darve et al. 2004). For the octo-linear model, solutions are displayed using analytical form of Eq. 15 as well as the numerical procedure.

In the stress space, the set of stress states for which instability cones exist constitutes the bifurcation domain. When the cones are reduced to a single direction, the latter gives the limit of the bifurcation domain. For non-associated materials this bifurcation limit is located strictly inside the plasticity limit, whereas for associated materials, the limits coincide. The limit of bifurcation domain has been plotted for Darve models and compared to Mohr–Coulomb limit (Prunier et al. 2009a, b, c); see Fig. 3.

Conditions of Effective Failure

Previously, conditions that lead to negative values of second-order work have been seen, i.e., the stress–strain state has to be included inside the bifurcation domain, and the loading path should follow a direction inside an instability cone. When second-order work is negative or zero, effective instability which leads to a burst of kinetic energy is conditioned by the control parameter chosen by the experimenter (or natural boundary loading occurring according to in situ conditions). From a basic point of view, when pulling on steel sample if the axial strain is controlled, the

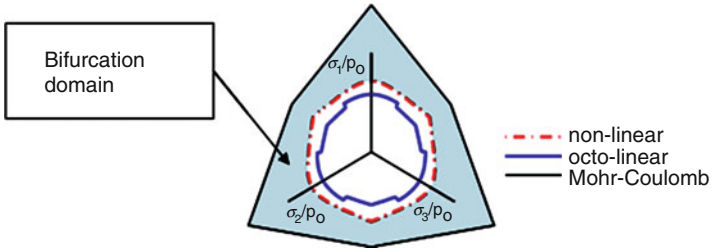


Fig. 3 Limit of bifurcation domain plotted according Darve models calibrated on a dense sand of Hostun plotted in the deviatoric plane

test can be continued after reaching the limit state given by the maximum axial stress, while if the axial stress is controlled, a sudden failure occurs. In fact the vanishing of the second-order work corresponds to a generalized limit state achieved strictly inside the plasticity limit. Let us consider proportional stress path as follows:

$$\begin{cases} \delta\sigma_1 = cst \quad cst \in \mathrm{R} \\ \delta\sigma_1 + R\delta\sigma_3 = 0 \quad R \in \boldsymbol{R}^* \\ \delta\sigma_2 - R'\delta\sigma_3 = 0 \quad R' \in \boldsymbol{R} \end{cases} \tag{19}$$

where $\boldsymbol{R}$ denotes the set of real numbers and $\boldsymbol{R}^*$ the set of real numbers excluding zero. For such loading paths, it is possible to write a generalized constitutive relation:

$$\begin{bmatrix} \delta\varepsilon_1 - \dfrac{\delta\varepsilon_3}{R} - \dfrac{R'}{R}\delta\varepsilon_2 \\ \delta\sigma_1 + R\delta\sigma_3 \\ \delta\sigma_2 - R'\delta\sigma_3 \end{bmatrix} = \underline{\underline{S}} \begin{bmatrix} \delta\sigma_1 \\ \dfrac{\delta\varepsilon_3}{R} + \dfrac{R'}{R}\delta\varepsilon_2 \\ \delta\varepsilon_2 \end{bmatrix} \tag{20}$$

where R and R' are variables that allow the description of all possible loading directions. Particular values of R and R' which describe directions included in an instability cone can be found, such that $\varepsilon_1 - \varepsilon_3/R - R'/R.\varepsilon_2$ reaches an extremum. Hence, because of the statical constraint in Eq. 20, at the $\varepsilon_1 - \varepsilon_3/R - R'/R.\varepsilon_2$ peak, $\det(S)$ vanishes for nontrivial solutions of Eq. 20 and a generalized failure rule can be defined. Furthermore, knowing that the second-order work can be rewritten as in Eq. 21,

$$\begin{aligned} \delta W_2 = &\left(\delta\varepsilon_1 - \frac{\delta\varepsilon_3}{R} - \frac{R'}{R}\delta\varepsilon_2\right) d\sigma_1 + (\delta\sigma_1 + R\delta\sigma_3)\left(\frac{\delta\varepsilon_3}{R} + \frac{R'}{R}\delta\varepsilon_2\right) \\ &+ (\delta\sigma_2 - R'\delta\sigma_3)\delta\varepsilon_2, \end{aligned} \tag{21}$$

the second-order work vanishes as well at this extremum. Figure 4 illustrates previous discussions on a particular stress path. First, a drained triaxial path is followed until reaching a stress deviator just before the bifurcation domain limit.

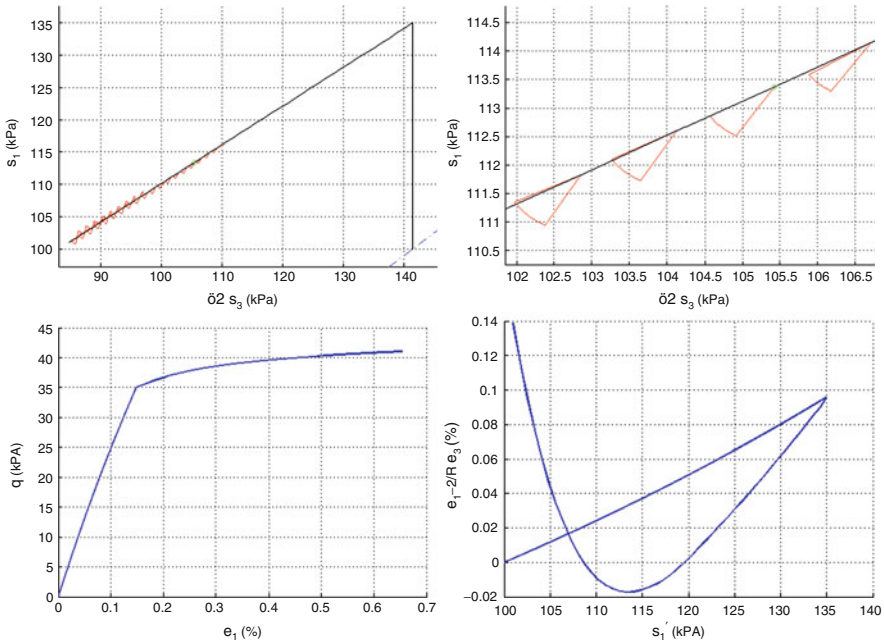


Fig. 4 Simulation of proportional stress path using INL model. First a drained triaxial path is applied until $q = 35$ kPa, and then a proportional stress path has been followed with $R = 0.85$ and $R' = 1$. At the $\varepsilon_1 - \varepsilon_3/R - R'/R.\varepsilon_2$ extremum, second-order work vanishes and a limit state is reached

Then, a proportional stress path is imposed with values of $R = 0.85$ and $R' = 1$ (Laouafa et al. 2011; Daouadji et al. 2011).

Based on Fig. 4, the following observations can be made. The first cone (reduced to a single direction) gives the bifurcation domain limit, whereas the limit state described by the extremum is reached whenever the loading path is tangent to a cone and $W_2 = 0$. As such, the loading path is strictly inside a cone when $W_2 < 0$. As concluding remarks, in the present case, the limit state is defined by a strain state. Similar studies can be made with proportional strain paths in which case limit states are defined by stress states instead.

Failure Analysis in Granular Materials by the Discrete Element Method

The Discrete Element Method (DEM) is a direct numerical simulation of granular media (Cundall and Strack 1979). In the granular assembly each grain is described as an elastic body interacting with a Coulombian friction with the neighboring grains in contact. Thus, pre-failure, failure state, and post-failure regime can be simulated in a very natural and realistic way without addressing the numerical difficulties usually encountered in continuum mechanics with the finite element

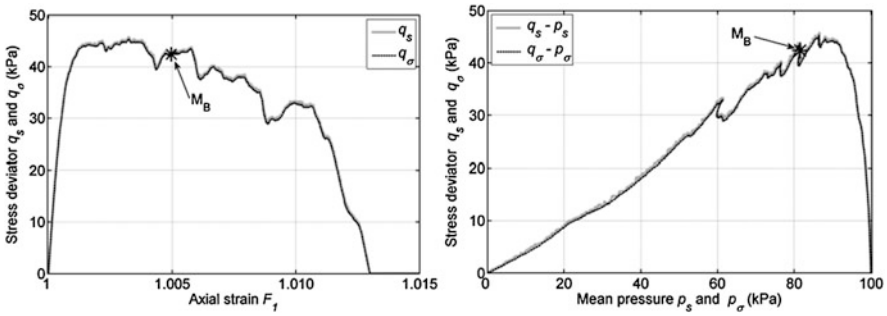


Fig. 5 Undrained (isochoric) triaxial compression of a loose particle assembly. q_s and p_s are computed from the stress state at the boundary of the sample, and q_σ and p_σ are computed from the internal stress state (Nicot et al. 2012a)

method. This is the main reason why the DEM approach is chosen here to validate numerically the features of diffuse failure as exhibited previously in sections "A General Criterion for Failure by Divergence Instabilities" and "The Second-Order Work Criterion, Features, and Illustrative 3D Examples (Multiaxial Loading)" by the second-order work criterion. A numerical cubical specimen of about 10,000 spheres of various radii is prepared, then some loading paths can be applied at the boundaries of the cube. For a more detailed description of the numerical method, the reader can refer to Nicot et al. (2012a) and Sibille et al. (2008).

The first point to be investigated now is the link between second-order work and kinetic energy as established theoretically in section "Introduction: Main Features of Failure in Geomaterials." To this end, an axisymmetric undrained triaxial path is simulated by applying numerically to the specimen, in a loose density state, an axisymmetric isochoric loading path. As observed experimentally (Daouadji et al. 2010), the deviatoric stress q (q is equal to the axial stress minus the lateral stress) passes through a maximum value. In Fig. 5, the peak of q can be observed as well as the ultimate vanishing of the intergranular stresses when the effective stress path reaches the origin of the q-p plane (p is the intergranular mean effective pressure). In addition to the above plot, a comparison is presented between the externally applied stresses at the boundary of the sample (noted by q_s and p_s) and the internal stresses q_σ and p_σ obtained from the internal intergranular force field by applying Love–Weber relation (Love 1927). The observed excellent agreement is an indicator of the consistency of the numerical computations. At stress state M_B, just after the peak in q, a small additional axial deviatoric stress is applied, i.e., $\Delta q_s = 1.6$ kPa. According to section "A General Criterion for Failure by Divergence Instabilities," the sample cannot sustain this additional force. This is illustrated in Fig. 6, where the applied external load q_s is no more equal to the internal stress q_σ. The failure induced by this additional load produces a burst of kinetic energy, which is equal for a small time increment (here, on Fig. 7, about 0.08 s) to half the difference between the external applied second-order work (noted by $[V.\Delta S_i.\Delta F_i]_{\Delta t}$ on Fig. 7) and the internal second-order work (noted by $[W_2]_{\Delta t}$), as

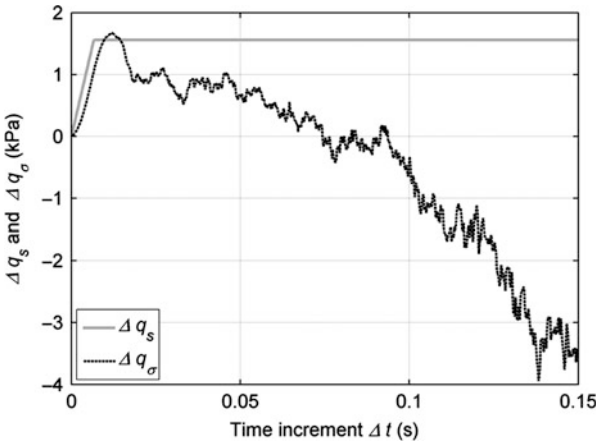


Fig. 6 Stress increments resulting from the application of $\Delta q_s = 1.6$ kPa from the state M_B defined in Fig. 5 (Nicot et al. 2012a)

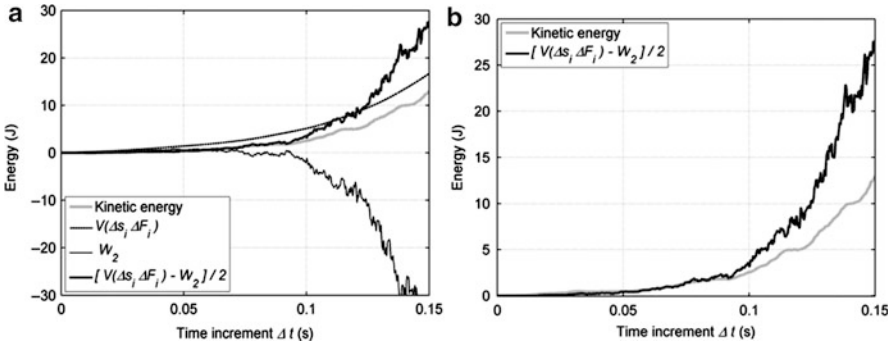


Fig. 7 Time series of the kinetic energy and terms $[W_2]_{\Delta t}$, $[V.\Delta S_i.\Delta F_i]_{\Delta t}$, and $[V(\Delta S_i.\Delta F_i - W_2)/2]_{\Delta t}$ during and after the application of $\Delta q_s = 1.6$ kPa from the state M_B (**a**). Focus on shorter time increments (**b**) (Nicot et al. 2012a)

established in section "A General Criterion for Failure by Divergence Instabilities." The kinetic energy is obtained by adding both the translational and rotational kinetic energies of each grain, whereas the external second-order is the product of the small additional axial surface force and the axial strain (multiplied by the sample volume V). On the other hand, the internal second-order work is simply calculated as the inner product of the Love macroscopic stress tensor and the macroscopic strain tensor (multiplied by V).

A second aspect (shown in section "The Second-Order Work Criterion, Features, and Illustrative 3D Examples (Multiaxial Loading)") is the fact that the second-order work is a quadratic form of incremental stresses or strains, which means – according to linear algebra – that there are some instability cones capturing all stress directions for which the second-order work is negative. Figure 8 shows such a cone for different values of the deviatoric stress $\eta = q/p$ by plotting the polar

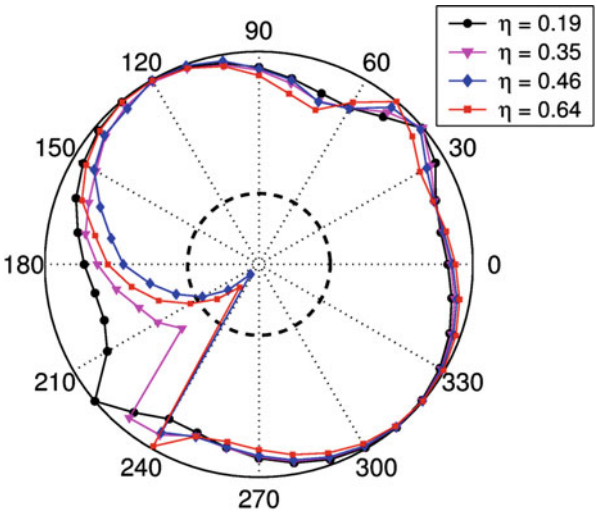


Fig. 8 Polar diagram of the normalized second-order work for a loose granular assembly at confinement pressure of 100 kPa. The dashed circles represent zero value of the second-order work; $\eta = q/p$ represents the mobilized deviatoric stress (Sibille et al. 2008)

variations of the normalized second-order work with respect to stress directions in axisymmetric conditions. The normalized second-order work is obtained by dividing the second-order work by the norms of the incremental stress and the incremental strain. As such, the normalized second-order work represents the cosine of the angle between both the stress and strain vectors in the axisymmetric plane and varies between −1 and +1. The second-order work takes negative values inside the dashed circle of Fig. 8. The fact that inside the instability cone an effective failure is possible (with a proper control variable) while outside no failure can occur is checked in Fig. 5. For both stress directions outside of the cone (alpha angle = 200° and 240°), stresses and strains remain constant under a small additional load. For both directions inside the cone (220° and 215.3°), the failure is characterized by suddenly and strongly growing strains with decreasing stresses. Eventually for both directions at the boundary of the cone (210° and 230°), some erratic responses from the sample are obtained (Fig. 9).

The question of the control variable is investigated in Fig. 10 for a q constant axisymmetric drained loading path (which is indeed an isotropic unloading from a deviatoric stress–strain state, corresponding to a stress direction of 215.3°). Experimentally, it has been observed (Daouadji et al. 2010) that the volume changes go through a maximum dilation along this path. Thus, if this loading path is fully stress controlled, no effective failure is expected according to the theory described in section "The Second-Order Work Criterion, Features, and Illustrative 3D Examples (Multiaxial Loading)." On the other hand, if this loading is partially volume variations controlled (mixed loading), an effective failure must be observed as given by the theory. These phenomena are well illustrated in Fig. 10, where both kinds of control are compared.

Finally, Fig. 11 establishes the important observation that an effective failure can be indeed induced by a perturbation close to the failure state. Because failure

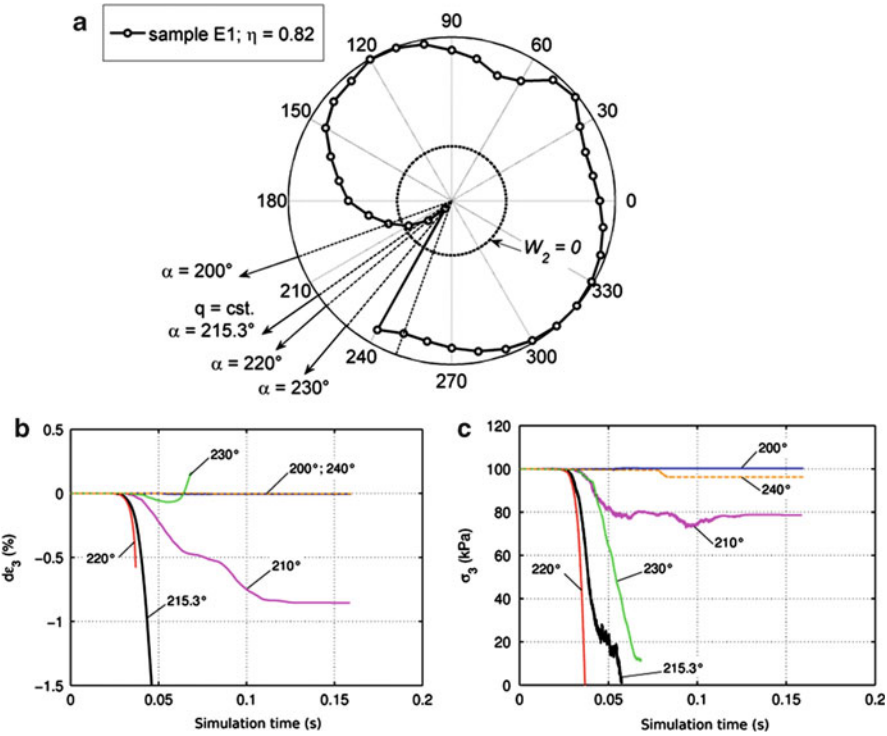


Fig. 9 (**a**) Polar plot of the normalized second-order work for a loose granular assembly at a confinement pressure of 100 kPa, with some particular stress directions highlighted. (**b**) Simulated responses in terms of strains (**b**) and stresses (**c**) to loading paths defined by $\delta\sigma_1 - \delta\sigma_3/R = 0$ and $\delta\varepsilon_1 + 2R\,\delta\varepsilon_3 < 0$, for a loose granular assembly at a confinement pressure of 100 kPa and for $\eta = q/p = 0.46$

states are essentially unstable states, they are naturally sensitive to perturbations. Experimentally, it is well known that failure can be triggered by a proper perturbation. Here, the numerical perturbation is constituted by a small "injection" of kinetic energy to the sample. Since all the computations are performed without gravity forces, at each computation step few grains are floating inside the specimen. To perturb the sample's state, a small additional velocity is numerically imparted to these grains. An effective failure is clearly observed in Fig. 11 just after the perturbation of the q constant axisymmetric path through a burst of kinetic energy with strongly growing strains and decreasing stresses.

Rock Joint Failure Modeling

In the following rocks as another example of geomaterials are considered. In order to assess rock slopes stability, a focus is made on rock joints whose failure can trigger rock falls. Among the possible different types of defects that can occur in

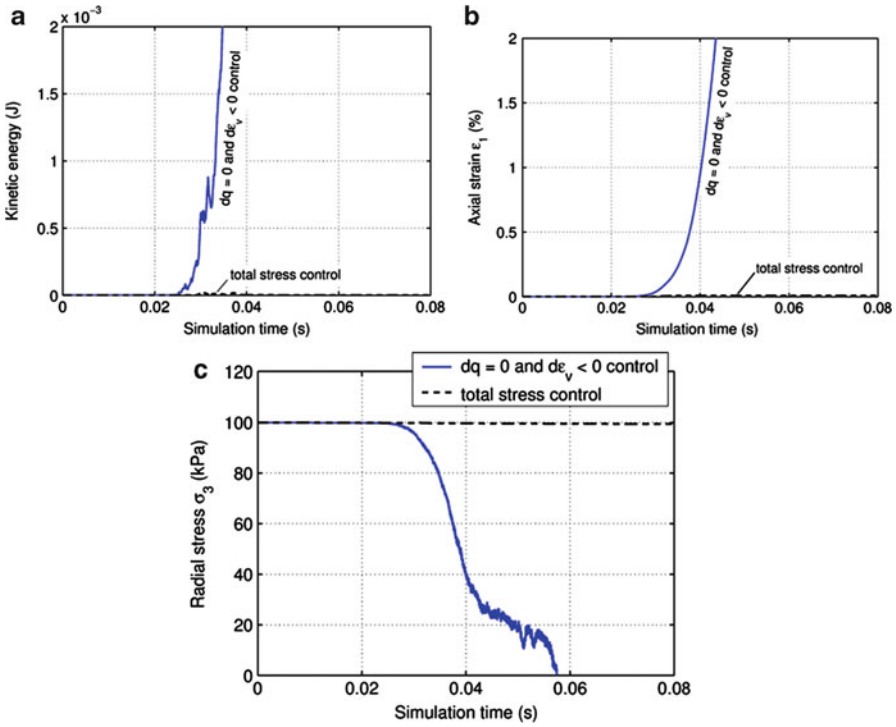


Fig. 10 Comparison of a loose granular assembly responses along a q-constant path for a full stress control ($\delta\sigma_1 = \delta\sigma_3 < 0$) and a control with $\delta q = 0$ and $\delta\varepsilon_V < 0$; kinetic energy (**a**), axial strain (**b**), and radial stress (**c**) versus time of simulation (Sibille et al. 2008)

rock slopes at different scales, the discontinuities in mechanical properties that exist at a macroscopic scale are called "rock joints." In a fractured rock slope, it can be assumed that the stability of the slope is mainly governed by the mechanical behavior of existing rock joints. Compared to the geomaterials covered in the previous sections, these rock joints constitute an interface medium (material). Thus, their mechanical behavior can be described using only four scalar variables as obtained from two vectors. The first vector relates to stress with σ as its component normal to the joint (positive in compression) and τ its tangential one. The second vector refers to relative displacement mobilized between the two rock blocks in contact along the joint, with u being the normal component (positive in compression) and γ the tangential component.

An internal second-order work is in this case defined as

$$\delta W_2 = \delta\vec{\sigma}\,\delta\vec{l} = \delta\sigma\,\delta u + \delta\tau\,\delta\gamma \tag{22}$$

As explained in section "A General Criterion for Failure by Divergence Instabilities," a negative second-order work can lead to failure by divergence

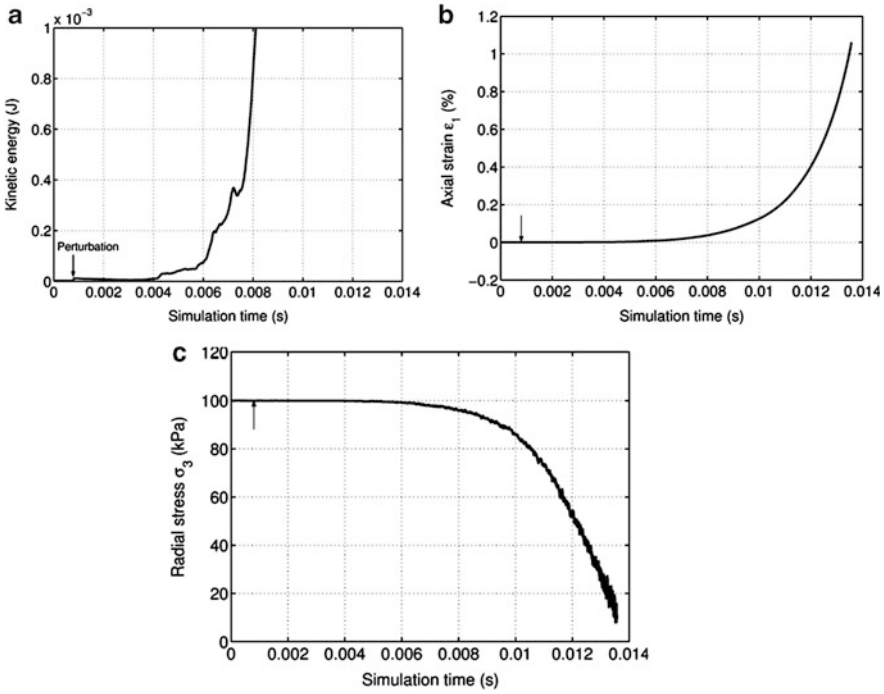


Fig. 11 Simulation of loss of sustainability for loose granular assembly at a mechanical state governed by the control parameters $\delta q = 0$ and $\delta\varepsilon_V = 0$; the arrows in the diagrams represent the time where perturbation was applied (Sibille et al. 2008)

instabilities. In this context, relative displacements along the joint will increase towards a dynamic regime, ultimately triggering a rockfall. Since the second-order work is governed by the material properties of the joint, two given rock joint are studied with a focus on their constitutive relations.

Two Rock Joint Constitutive Relations

In the following, the rock joint relations are derived from the soils relations presented in section "Illustration of Instability Cones Using Darve Model." An INL2 relation, with an infinity of tensorial zones, and a quadri-linear relation with four tensorial zones for rock joints were recently proposed in Duriez et al. (2011a, b). The development of these models relies on calibration paths along which the behavior of the joint is known, i.e.,

- A constant normal displacement (CND) path, with ($\delta u = 0$; $\delta\gamma = cst$). Four moduli are thus defined from this path:

$$G_{\gamma}^{+} = \frac{\partial \tau}{\partial \gamma_{u\,cst,\,d\gamma>0}} \qquad G_{\gamma}^{-} = \frac{\partial \tau}{\partial \gamma_{u\,cst,\,d\gamma<0}}$$

$$N_{\gamma}^{+} = \frac{\partial \sigma}{\partial \gamma_{u\,cst,\,d\gamma>0}} \qquad N_{\gamma}^{-} = \frac{\partial \sigma}{\partial \gamma_{u\,cst,\,d\gamma<0}}$$

Moduli $G_{\gamma}^{+/-}$ correspond to tangential stiffnesses, while moduli $N_{\gamma}^{+/-}$ take into account the dilatant (or contractant) feature of the joint.

- A constant tangential displacement (CTD) path, with ($\delta u = cst$; $\delta\gamma = 0$). This corresponds to an oedometric compression. Four other moduli are herein defined:

$$G_{u}^{+} = \frac{\partial \tau}{\partial u_{\gamma\,cst,\,du>0}} \qquad G_{u}^{-} = \frac{\partial \tau}{\partial u_{\gamma\,cst,\,du<0}}$$

$$N_{u}^{+} = \frac{\partial \sigma}{\partial u_{\gamma\,cst,\,du>0}} \qquad N_{u}^{-} = \frac{\partial \sigma}{\partial u_{\gamma\,cst,\,du<0}}$$

Moduli $N_{u}^{+/-}$ are normal stiffnesses. Changes in τ during compression can be described through moduli $G_{u}^{+/-}$. Indeed, numerical and experimental data revealed such variations, depending on the state of shearing in the joint (Duriez et al. 2011a; see Fig. 12).

Along any other loading path, the behavior of the joint is computed as an interpolation between the behavior of these two calibration paths. For a piecewise linear interpolation, a quadri-linear relation, with four tensorial zones is derived. For constant signs of du and $d\gamma$, a linear relation with a constant matrix links in this case $\delta\vec{\sigma}$ to $\delta\vec{l}$. For example, if $\delta u > 0$ and $\delta\gamma < 0$, it comes

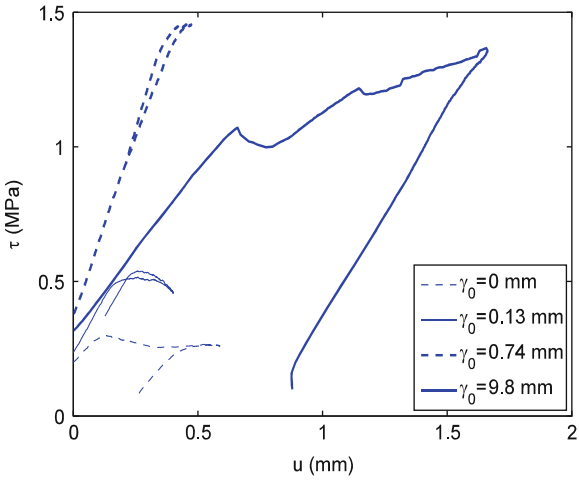


Fig. 12 Changes in τ during different oedometrical compressions on a rock joint in gneiss, since different initial states in shearing (Duriez et al. 2011a). The slopes of the curves correspond to moduli $G_{u}^{+/-}$

$$\delta\vec{\sigma} = \begin{pmatrix} \delta\tau \\ \delta\sigma \end{pmatrix} = \begin{pmatrix} G_\gamma^- & G_u^+ \\ N_\gamma^- & N_u^+ \end{pmatrix} \vec{\delta l} = \begin{pmatrix} G_\gamma^- & G_u^+ \\ N_\gamma^- & N_u^+ \end{pmatrix} \begin{pmatrix} \delta\gamma \\ \delta u \end{pmatrix} \tag{23}$$

Using a quadratic interpolation leads to the INL2 relation. The expression (not presented here) involves matrices depending on the eight moduli and the direction of $\vec{\delta l}$, leading to an infinite number of tensorial zones, meaning that for each direction, there is a different constitutive matrix. Predictions of the INL2 relation are generally more efficient than the quadri-linear relation, but both are generally close. Furthermore, the quadri-linear relation allows analytical derivations, as it will be seen in section "The Use of the Second-Order Work Criterion."

Appropriate expressions of the moduli were proposed in (Duriez et al. 2011b). They depend on the state of shearing in the joint, expressed by the ratio τ/σ. Due to the nature of the expressions used in the model, a plastic limit criterion is intrinsically defined in the relations. The stress states they predict do not exceed a Mohr–Coulomb-like criterion. Both models describe a non-associated behavior (Duriez et al. 2011b), which justifies the use of the second-order work criterion.

The Use of the Second-Order Work Criterion

For any rock joint constitutive relation in one tensorial zone, e.g., for the quadri-linear relation with constant signs of δu and $\delta\gamma$, it is easy to show (Duriez et al. 2012) that the sign of the second-order work is given by a second-order polynomial P, with respect to the loading direction $\delta u/\delta\gamma$: $P(\delta u/\delta\gamma)$. Under the assumption of positive normal stiffnesses $N_u^{+/-}$ – this could be untrue in high-porosity rocks showing compaction bands (Mollema and Antonellini 1996) – directions of instability exist under the following condition:

$$\left(G_u^{+/-} + N_\gamma^{+/-}\right)^2 \geq 4N_u^{+/-}G_\gamma^{+/-} \tag{24}$$

The above equation involves a comparison between "diagonal" stiffnesses, $N_u^{+/-}$ and $G_\gamma^{+/-}$, and coupled stiffnesses, $G_u^{+/-}$ and $N_\gamma^{+/-}$. Applying Eq. 24 to the four tensorial zones of the quadri-linear relation allows determining directly its bifurcation domain in the stress space (Fig. 13). Directions of instability exist under sheared states of the joint, but before the Mohr–Coulomb criterion. Shearing reduces indeed diagonal (e.g., normal) stiffnesses (Bandis et al. 1983) and increases coupled ones, allowing Eq. 23 to be verified. For the INL2 relation, the bifurcation domain is obtained using a numerical procedure. It is reduced in the INL2 case, compared to the quadri-linear case (see Fig. 13) (as for soils, see previous Fig. 3).

Directions of instability are determined analytically for the quadri-linear relation, by solving the inequality $P(\delta u/\delta\gamma) \leq 0$. A cone of directions is found; see Fig. 14. The two branches of the cone are the roots of polynomial P, and all directions $\delta u/\delta\gamma$ between the two roots trigger $P(\delta u/\delta\gamma) < 0 \Leftrightarrow \delta W_2 < 0$. The cone is more and more open as the rock joint is sheared, i.e., as the ratio

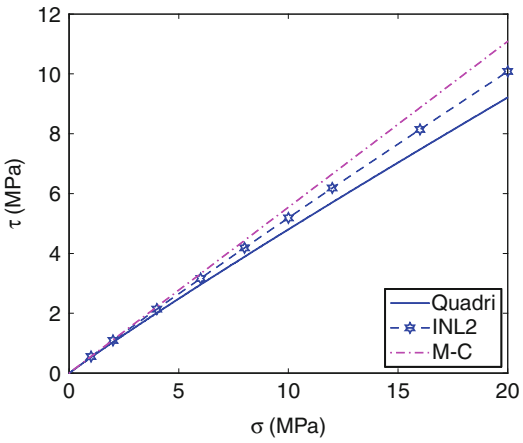


Fig. 13 Bifurcation domains of quadri and INL2 rock joint constitutive relations (Duriez et al. 2012). A Mohr–Coulomb (M–C) criterion gathers the admissible stress states

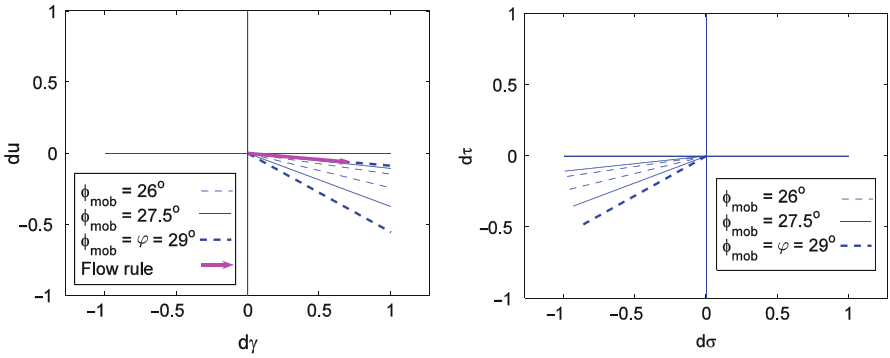


Fig. 14 Cones of directions of instability for the quadri-linear relation, in the displacement space (*left*) or the stress space (*right*) (Duriez et al. 2012). The cones get more and more open while the shearing increases

$\tau/\sigma = \tan(\Phi_{\text{mob}})$ increases. Once on the plastic limit criterion, for $\tau/\sigma = \tan(\varphi)$, one branch of the cone in the displacement space corresponds to the flow rule of the joint. Indeed, for a loading $\delta\vec{l}$ corresponding to the flow rule, $\delta\vec{\sigma} = f\left(\delta\vec{l}\right) = \vec{0}$ (the unstable direction in stress space is then not defined), and obviously $\delta W_2 = 0$.

For the INL2 relation, similar results are obtained with a numerical procedure (Duriez et al. 2012). For both relations, in stress space, both normal and shear stresses decrease for unstable loadings, as for soils in axisymmetric conditions, considering mean and deviatoric stresses. It was shown in Duriez et al. (2012) that the orientation in the stress space of the directions of instability is given by a comparison between the coupled stiffnesses of the joint: the moduli Gu and $N\gamma$. These coupled stiffnesses govern the instability of a rock joint, since they appear in the equation of the bifurcation domain.

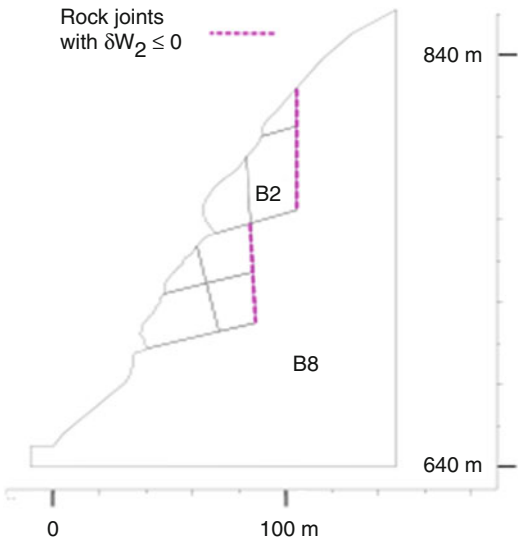


Fig. 15 Numerical model of the slope with the rock joints where $\delta W_2 \leq 0$ during the simulated loading (Merrien-Soukatchoff et al. 2011)

Application

An existing rock slope, in south of France, is studied using this approach (Merrien-Soukatchoff et al. 2011). A numerical model of the cliff was proposed, in two dimensions, with eight blocks of intact matrix and several rock joints inbetween. The INL2 constitutive relation is used to describe rock joint behavior. A continuous increase in gravity is applied to the cliff, as a loading. Another loading could be chosen by excavating the top of the cliff, in order to better reproduce the geological history of the slope.

During the simulated loading, negative values of the second-order work are obtained in some joints (Fig. 15). It is noteworthy that this is the case for one joint which has not yet reached its plastic limit criterion, between blocks B2 and B8. No rockfall is obtained here; this would probably be the case if the concerned joints surround a set of blocks. A global second-order work, taking into account the local values along all joints, is required to deal with such cases.

Failure Modeling by Finite Element Method: Homogeneous Cases and Boundary Value Problems

Material Instabilities in the Triaxial Test

The preceding sections delved into the description of failure in geomaterials as a bifurcation problem with the second-order work criterion acting as a general framework within which various forms of instabilities can be captured such as diffuse and localized modes. Examples of diffuse and localized failure are manifest

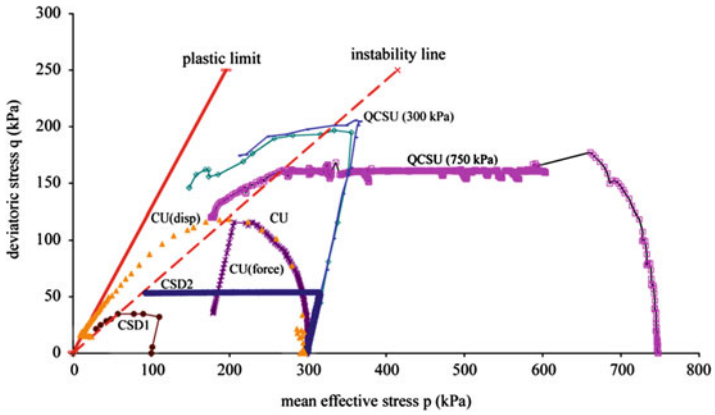


Fig. 16 Experimental evidence of material instabilities in various triaxial tests on Hostun S28 sand (Darve et al. 2007)

in conventional triaxial testing of soils where both stress and strain are axisymmetric. The most well-known manifestation of diffuse failure is in the consolidated undrained (CU) shearing of a loose sand when the effective stress path passes through a peak where the second-order work is violated. With the proper operative control parameters, herein load control mode, a diffuse failure emerges in a CU test on loose sand with spontaneous release of kinetic energy. While this peak point lies within the bifurcation domain, it is most importantly associated with a loading direction that is inside the bifurcation cone. The above two conditions together with the proper control parameters are required for an effective failure to occur. Figure 16 shows how diffuse material instability occurs in several types of triaxial tests, all having a common feature of violating the second-order work well below the plastic limit.

In the so-called CSD test where the specimen is sheared at constant deviatoric stress with decreasing mean effective stress in drained conditions, controllability is lost at some point during the test. The ensuing diffuse failure is governed by the same phenomenon as in the CU test, provided that all of the proper conditions discussed previously are in place. The QCSU test, a variant of the CSD test where drainage is prevented while the deviatoric stress fluctuates about a constant value, also shows the same instability phenomenon. All tests exhibited unstable material response by loss of controllability, and hence collapse, roughly at the same stress ratio defined by a reference bifurcation line with slope of $\eta = q/p = 0.6$ in the p-q space, which is reminiscent of the well-known instability line introduced by (Lade 1992). However, in the framework described in this chapter, a wider scope using bifurcation theory provides theoretical and physical insights in explaining the collapse behavior of sands.

Finite Element Analysis of Diffuse Failure

In the next diffuse failure with respect to control parameters as evidenced in the triaxial tests described in the previous subsection is numerically captured.

Two initial densities are examined, namely, void ratios $e_0 = 0.7$ and 0.8, referring to medium–loose to loose sand, respectively. An elastoplastic model having all the essential ingredients that are prerequisites for capturing unstable material behavior is employed (see Wan and Guo 2004). These include density, stress, and anisotropy dependencies through a micromechanical enrichment of Rowe's stress dilatancy equations. As additional important properties, they account for plastic-strain softening and non-associativity of plastic yield and flow that provide sources of material instability, as discussed in the beginning of this chapter.

Loading Direction, Control Parameters, and Bifurcation Domain

The numerical triaxial tests are herein modeled as an initial boundary value problem involving coupled solid–fluid interaction to mimic the actual triaxial test with drainage conditions. As such, a three-dimensional finite element mesh with 5,220 linear elements representing a cylindrical sand sample is used in the simulations.

Figure 17 summarizes the numerical results of CSD, CU, and QCSU tests performed on a loose sand specimen, $e_0 = 0.8$ with a focus on second-order work evolution. The undrained (CU) test was run in both displacement control and stress control in order to highlight the role of the load control parameters on the resulting failure mode. Figure 17a shows that the points at which the second-order work first vanishes for all tests fall almost on a straight line with a slope $\eta = 0.56$. This line actually lies above the lower limit of the bifurcation domain where the second-order

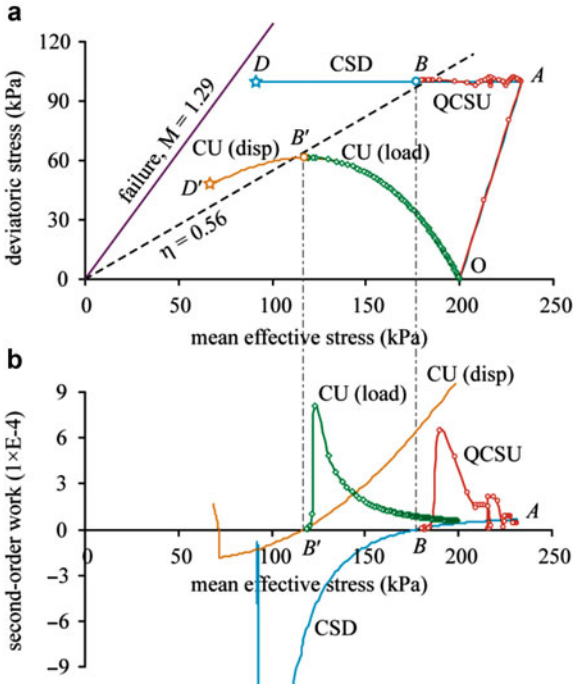


Fig. 17 Simulation of various triaxial tests together with identification of bifurcation and ultimate failure points

work first vanishes with all but only one possible loading direction for material instability to occur. When compared with the Mohr–Coulomb failure line at a slope $M = 1.29$ in the p-q space, it is evident that the second-order work first vanishes at a stress state well below the classic plastic limit.

The evolution of the non-normalized second-order work with mean effective stress for these tests is presented in Fig. 17b. It is interesting to observe that the second-order work becomes zero approximately at the same point B for the two q-constant tests, but with different evolution due to the loading history. It is noteworthy to point out that contrary to the QCSU test, the CSD test carried past the bifurcation point B but finally failed at point D near the plastic limit surface. For the QCSU test, failure as a loss of controllability occurred much earlier at the bifurcation point B, due to the proper load control parameters.

Also of particular interest are the numerical results obtained for the consolidated undrained compression test case where it is found that the second-order work vanishes at the same stress point B′ irrespective of the load control parameters. However, in the force-controlled case, as the second-order work vanishes, the test cannot be controlled further and the computations break down, meaning collapse of the specimen. The undrained test performed under strain-controlled mode carries on past the peak B′ until it fails at point D′ for lack of numerical convergence. Since, by contrast, the same test stops at the very same peak point B′ under stress controlled mode, this observation conveys two important results. First, to obtain diffuse failure with collapse, it is not solely sufficient to have the second-order work criterion violated, but it is also necessary to have the proper control variables in place, here a force-controlled loading program. Second, the direction of the loading matters, and here it appears to be a horizontal direction which is necessarily contained within the instability cone in the p-q plot. The instability cone essentially defines the range of loading directions in the p-q space for which the second-order work is violated.

Second-Order Work, Relevance to Loss of Uniqueness, and Localization

It is recalled that the vanishing of the second-order work locally and globally signals the possibility of instability and the loss of uniqueness of solutions of the boundary value problem. This numerical aspect is checked in the subsection whereby two or more solutions can emerge starting from the same base state. As an illustration, the undrained test simulations for medium–loose sand ($e_0 = 0.7$) are considered.

Figure 18a shows the branching of effective stress paths after the peak was reached where second-order work vanished. For clarity of illustration, only four strategic gauss points where choosen as indicated in Fig. 18b showing the final nonhomogeneous deformed configuration when localization of deviatoric plastic strain occurred. As expected, inside this shearing zone (#078), the deviatoric plastic strains are higher than in other parts (#004, #068, #059) of the specimen. In Fig. 18, these are reflected into distinctly different paths followed by the four

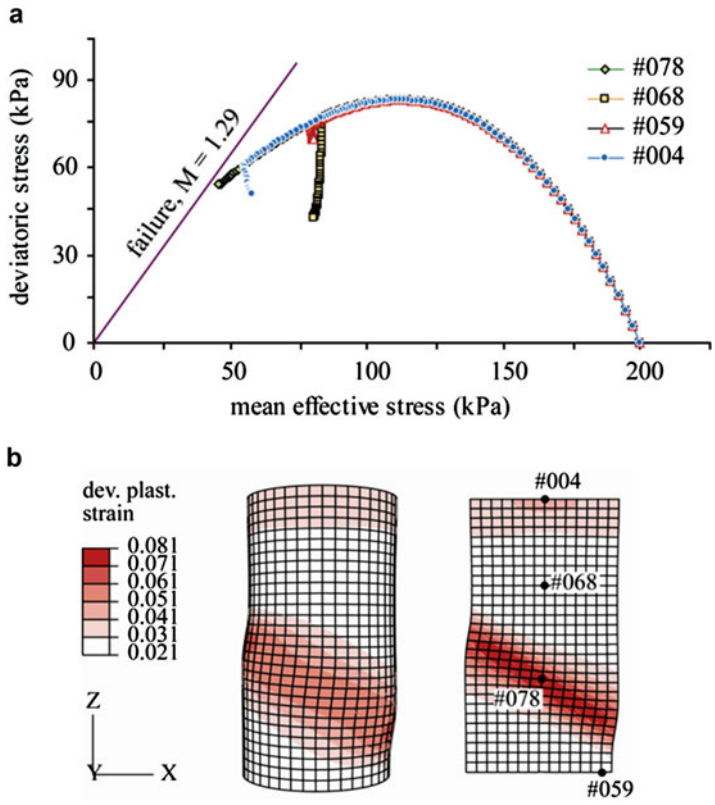


Fig. 18 Loss of homogeneity after diffuse failure with ensuing localization mode when reaching plastic limit

strategic nodes, with unloading occurring at nodes #068 and #59 found outside the shear band.

The loss in homogeneity arising from bifurcation is reflected in all field variables, for example, excess pore water pressure, void ratio, and deviatoric plastic-strain fields as illustrated in Fig. 19a–c for the same four nodes mentioned above. As seen in Fig. 19a, there is a steady and uniform increase in pore pressures until when bifurcation is first met. Thereafter, a distinct region of localized deformations eventually appears. Material points inside this region experience escalating pore water pressures while points outside this shear zone undergo a decrease in pore pressures. In line with the pore pressure field response, Fig. 19b indicates that the localized zones compact, leading to a decrease in void ratio and an increase in pore pressures. Such local compactive response is expected since the sample is medium–loose even though the sample is globally undrained.

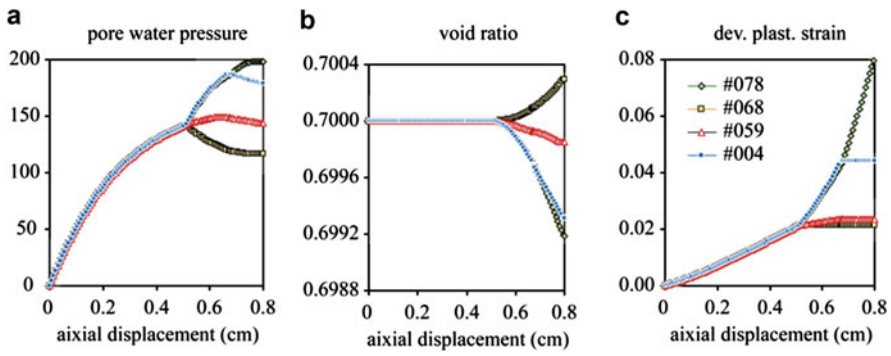


Fig. 19 Evolution of (**a**) excess pore water pressure, (**b**) void ratio, and (**c**) deviatoric plastic strain at four selected nodes in the FE mesh

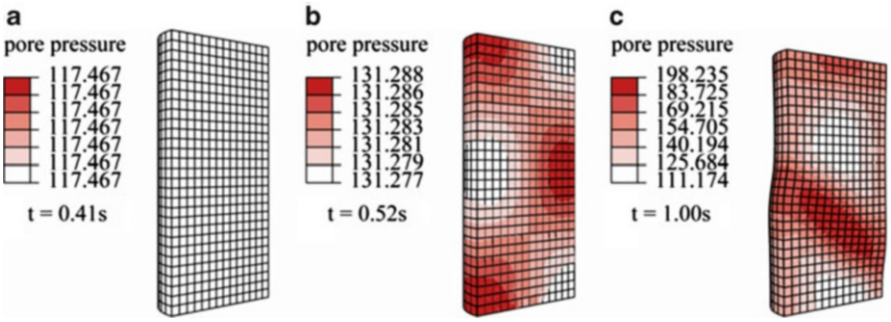


Fig. 20 Snapshots of the excess of pore water pressure field at (**a**) $t = 0.41$ s, (**b**) $t = 0.52$ s, and (**c**) $t = 1.00$s ($1\times$ magnification)

Figure 20 shows a transverse section of the sample at various stages during the undrained compression test. The pore pressure field is still uniform at the peak ($t = 0.41$ s), where the second-order work first becomes zero (Fig. 19a). A departure from homogeneity is only observed at $t = 0.52$ s when a bifurcated response eventually emerges (Fig. 19b). The pore pressure increases throughout the specimen in a somewhat heterogeneous manner. Moving further along loading history, the pore pressure field develops into a more organized state as shown in Fig. 19c where a well-defined band is revealed with concentrated high values of pore pressures.

The numerical results demonstrate that the vanishing of the second-order work refers to a failure mode which is diffuse in the absence of any localization or organized displacement pattern. Localized deformations distinctly appear crossing through the sand specimen diagonally only when reaching close to the plastic limit surface in the stress space and well after diffuse failure has occurred at the peak, confirming that diffuse failure precedes here localization during loading history. The hierarchy of failure modes and other issues were discussed in Wan et al. (2012).

FEM Modeling of the Petacciato Landslide

The area of Petacciato is located along the Adriatic coast of Italy. The landslide occurred in 1996, after an intense rainy period. The landslide covers an area of 2 km wide, 7 km long, and over 200 m of difference in elevation. Geological and geotechnical surveys have been carried out by the "Studio Geotecnico Italiano" (SGI) (SGI 2004). The failure surface was found to be between 70 and 90 m in depth. The mean slope of the site was only 6°. The soil underlying the first 5–6 m of the slope is assumed to be homogeneous and composed of blue clayey silt.

To conduct the numerical modeling of this landslide, the finite element code LAGAMINE developed at Technical University of Liège (Belgium) is used. A constitutive model describing the behavior of unsaturated soils has been used. It is based on a frictional non-associative elastoplastic relation for the solid phase, and a water retention description for the fluid phase. The elastoplastic model uses the Van-Eekelen (Van-Eekelen 1980) surface as elastic and plastic limit. The hardening parameter chosen here is the equivalent plastic shear strain that makes the elastic limit evolve isotropically until the plastic limit is ultimately reached. The non-associative flow rule is described using a plastic potential surface with the same form as the yield surface, except that the friction angle is replaced with a dilation angle. The evolution of the dilation angle follows the same rule as for the friction angle during hardening. As for describing the water retention behavior of the soil, the empirical relation of Van-Genuchten (1980) is used. Hydromechanical coupling is established using Bishop's effective stress principle (Bishop 1959) in which the parameter χ is assumed to be the current degree of saturation. Finally, fluid flow is described by the generalized Darcy's law, the so-called Richards equation (Richards 1931). In this expression, the permeability is proportional to the degree of saturation degree with the proportionality coefficient being the permeability at the current saturation state.

The finite element mesh with associated boundary conditions that are used to model the Petacciato slope are shown in Fig. 21.

According to the mechanical boundary conditions, vertical displacements are fixed at the bottom edge, while horizontal displacements are fixed on left and right edges. Gravity loads are then applied to the whole body. For hydraulic boundary conditions, the bottom edge is impervious, whereas the upper edge is permeable with water pressure equal to zero. Water pressures are imposed on both left and

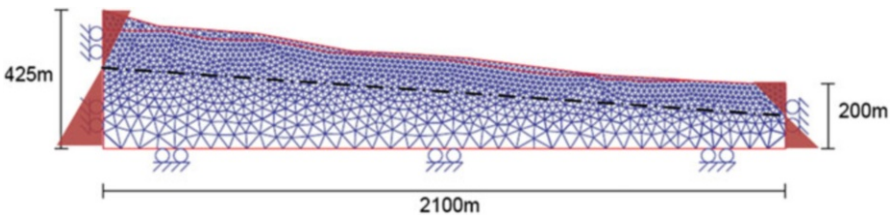


Fig. 21 Finite element mesh and boundary conditions of the Petacciato slope

Table 1 Mechanical and retention parameters of the Petacciato soils

Soil parameters	Symbols	Unit	Blue-gray clay
Grain-specific weight	ρ_s	kN/m^3	27.40
Young modulus	E	MPa	95.0
Poisson's ratio	ν	–	0.21
Porosity	n	–	0.3
Intrinsic permeability	k_w	m^2	10^{-17}
Friction angle	φ	°	19.0
Cohesion	C	kPa	171
Dilatancy angle	ψ	°	0
Maximal saturation degree	S_w	–	1
Residual saturation degree	S_{rw}	–	0.1
First retention parameter (Van-Genuchten)	α	P_a^{-1}	1.10^{-5}
Second retention parameter (Van-Genuchten)	β	–	1.35

right boundaries so that they can evolve at each computation step in order to impose a chosen water table level in the steady-state regime. Fully saturated conditions are considered below the water table, while the above soil is partially saturated. The numerical simulation involves computing the degree of saturation as well as suction by integrating Richard's equation.

The soil parameters have been determined by calibrating SGI's laboratory tests (SGI 2004), mainly oedometer and triaxial tests. In particular, calibration has been based on the more representative CD and CU tests. Results of this calibration exercise are summarized in Table 1.

The numerical simulation of the Petacciato slope is as follows (Lignon et al. 2009). Initially, the whole soil medium is dry with the water table located at the bottom boundary. In subsequent steps, hydrostatic water pressures are imposed in view of mimicking a rise in the water table. This was achieved in a total number of 100 loading steps, whereby hydrostatic water pressures were imposed at elevations between 0 and 425 m on the left boundary and between 0 and 200 m on the right boundary; see Fig. 21. As such, the water table is roughly parallel to the slope. To evaluate any instability at every loading step, the second-order work is computed at each integration point throughout the whole body. In order to enhance the contrast of isovalues of the second-order work maps and to have a proper interpretation of the stability evolution of a given material point during the loading process, the second-order work has been normalized as follows:

$$\delta W_2^n = \frac{\delta W_2}{\|\delta\sigma_i\| \, \|\delta\varepsilon_i\|} \tag{25}$$

This quantity, as seen previously, corresponds to the cosine of the angle between vectors $\delta\sigma_i$ and $\delta\varepsilon_i$ and varies between -1 and 1. Values less or equal to zero signal unstable states, while positive values refer to stable states. Furthermore, states with

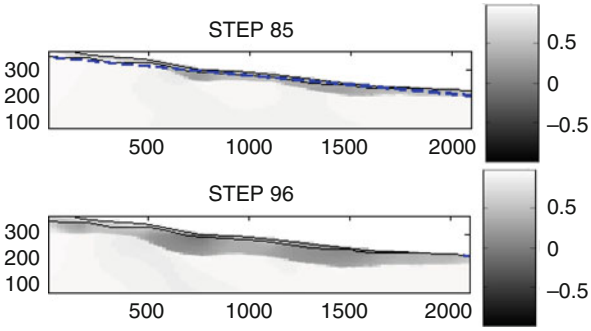


Fig. 22 Unstable domain of the Petacciato slope during saturation process

values near one can be considered as being further away from instability than those near zero. Hence, δW_2^n can be used in the same manner as the local safety factor centered around the value of zero instead of one. Figure 22 illustrates the evolution of the normalized second-order work with loading history.

When the water table approaches the free surface, the results of loading step 85 plotted in Fig. 22, show that the area of low δW_2^n values develops extensively along the slope surface. As soon as the slope reaches 95 % saturation, the region of low or negative δW_2^n extends further deep into the slope as shown for loading step 96 in Fig. 22. At the same time, a slip surface located at a depth between 80 and 90 m is clearly visible. These numerical results are in close agreement with borehole and inclinometer measurements made in the field. It is worth noting that the natural inclination of the slope is around 6° and consequently the failure, which is in the diffuse mode, cannot be explained using either classical empirical methods or limit analyses.

Conclusion

In this chapter, an attempt to demonstrate why and how plasticity and damage phenomena in geomaterials are fundamentally different from classical interpretation was made. The main reason is the non-associated character of the rate-independent irreversible strains in these materials. The second-order work criterion has been introduced as a general and necessary condition for all types of divergence instabilities to occur and to give rise to localized or diffuse failure modes. This criterion has been analyzed in relation to granular microstructure and the characteristic expression of macroscopic failure through the phenomenon of bursts of kinetic energy. Its main features have been investigated (i.e., existence of a bifurcation domain and of instability cones) and they have been carefully checked using a discrete element model. Eventually some examples were presented using the finite element method to show the applicability of this criterion to in situ boundary value problems implicating both soil and rock slopes.

References

S.C. Bandis, A.C. Lumsden, N.R. Barton, Fundamentals of rock joint deformation. Int. J. Rock Mech. Min. Sci. Geomech. Abstr. **20**(6), 249–268 (1983)

A.W. Bishop, Principle of effective stress. Teknisk. Ukeblad. **106**(39), 859–863 (1959)

P.A. Cundall, O.D.L. Strack, A discrete numerical model for granular assemblies. Geotechnique **29**(1), 47–65 (1979)

A. Daouadji, H. Al Gali, F. Darve, A. Zeghloul, Instability in granular materials, an experimental evidence of diffuse mode of failure for loose sands. J. Eng. Mech. **136**(5), 575–588 (2010)

A. Daouadji, F. Darve, H. Al-Gali, P.Y. Hicher, F. Laouafa, S. Lignon, F. Nicot, R. Nova, M. Pinheiro, F. Prunier, L. Sibille, R. Wan, Diffuse failure in geomaterials: experiments, theory and modeling. Int. J. Numer. Anal. Methods Geomech. **35**(16), 1731–1773 (2011)

F. Darve, S. Labanieh, Incremental constitutive law for sands and clays: simulations of monotonic and cyclic tests. Int. J. Numer. Anal. Methods Geomech. **6**(2), 243–275 (1982)

F. Darve, E. Flavigny, M. Meghachou, Yield surfaces and principle of superposition revisited by incrementally non-linear constitutive relations. Int. J. Plast. **11**(8), 927–948 (1995)

F. Darve, G. Servant, F. Laouafa, H.D.V. Khoa, Failure in geomaterials: continuous and discrete analyses. Comp. Methods Appl. Mech. Eng. **193**, 3057–3085 (2004)

F. Darve, L. Sibille, A. Daouadji, F. Nicot, Bifurcations in granular media: macro- and micro-mechanics approaches. Comptes Rendus Acad. Sci. Mec. **335**, 496–515 (2007)

J. Duriez, F. Darve, F.V. Donzé, A discrete modeling-based constitutive relation for infilled rock joints. Int. J. Rock Mech. Min. Sci. **48**(3), 458–468 (2011a)

J. Duriez, F. Darve, F.V. Donzé, Incrementally non-linear plasticity applied to rock joint modeling. Int. J. Numer. Anal. Methods Geomech. (2011b). doi:10.1002/nag.1105

J. Duriez, F. Darve, F.V. Donzé, F. Nicot, Material stability analysis of rock joints. Int. J. Numer. Anal. Methods Geomech. (2012). doi:10.1002/nag.2149

G. Gudehus, A comparison of some constitutive laws for soils under radially loading symmetric loading unloading, *in Third International Conference on Numerical Methods in Geomechanics*, 1979, edn Balkema, pp. 1309–1323

R. Hill, A general theory of uniqueness and stability in elasto-plastic solids. J. Mech. Phys. Solids **6**, 236–249 (1958)

P.V. Lade, Static instability and liquefaction of loose fine sandy slopes. J. Geotech. Eng. Div. Am. Soc. Civ. Eng. **118**(1), 51–71 (1992)

F. Laouafa, F. Prunier, A. Daouadji, H. Al-Gali, F. Darve, Stability in geomechanics, experimental and numerical analyses. Int. J. Numer. Anal. Methods Geomech. **35**(2), 112–139 (2011)

J. Lerbet, M. Aldowadji, N. Challamel, F. Nicot, F. Prunier, F. Darve, P-positive definite matrices and stability of nonconservative systems. J. Appl. Math. Mech. ZAMM **92**(5), 409–422 (2012)

S. Lignon, F. Laouafa, F. Prunier, F. Darve, H.D.V. Khoa, Hydro-mechanical modelling of landslides with a material instability criterion. Geotechnique **59**(6), 513–524 (2009)

A.E.H. Love, *A Treatise of Mathematical Theory of Elasticity* (Cambridge University Press, Cambridge, 1927)

M.M. Mehrabadi, M. Oda, S. Nemat-Nasser, On statistical description of stress and fabric in granular materials. Int. J. Numer. Anal. Methods Geomech. **6**, 95–108 (1982)

V. Merrien-Soukatchoff, J. Duriez, M. Gasc, F. Darve, F.V. Donzé, Mechanical Stability Analyses of Fractured Rock Slopes, in *Rockfall Engineering*, ed. by S. Lambert, F. Nicot (Wiley/ISTE, New York/London, 2011)

P. Mollema, M. Antonellini, Compaction bands: a structural analog for anti-mode I cracks in aeolian sandstone. Tectonophysics **267**, 209–228 (1996)

F. Nicot, F. Darve, A micro-mechanical investigation of bifurcation in granular materials. Int. J. Solids Struct. **44**, 6630–6652 (2007)

F. Nicot, F. Darve, Diffuse and localized failure modes: two competing mechanisms. Int. J. Numer. Anal. Methods Geomech. **35**(5), 586–601 (2011)

F. Nicot, F. Darve, H.D.V. Khoa, Bifurcation and second-order work in geomaterials. Int. J. Numer. Anal. Methods Geomech. **31**, 1007–1032 (2007)

F. Nicot, L. Sibille, F. Darve, Bifurcation in granular materials: an attempt at a unified framework. Int. J. Solids Struct. **46**, 3938–3947 (2009)

F. Nicot, N. Challamel, J. Lerbet, F. Prunier, F. Darve, Bifurcation and generalized mixed loading conditions in geomaterials. Int. J. Numer. Anal. Methods Geomech. **35**(13), 1409–1431 (2011)

F. Nicot, L. Sibille, F. Darve, Failure in rate-independent granular materials as a bifurcation toward a dynamic regime. Int. J. Plast. **29**, 136–154 (2012a)

F. Nicot, N. Challamel, J. Lerbet, F. Prunier, F. Darve, Some insights into structure instability and the second-order work criterion. Int. J. Solids Struct. **49**(1), 132–142 (2012b)

F. Nicot, N. Hadda, F. Bourrier, L. Sibille, R. Wan, F. Darve, Inertia effects as a possible missing link between micro and macro second-order work in granular media. Int. J. Solids Struct. **49** (10), 1252–1258 (2012c)

F. Prunier, F. Nicot, F. Darve, F. Laouafa, S. Lignon, 3D multi scale bifurcation analysis of granular media. J. Eng. Mech. ASCE **135**(6), 493–509 (2009a)

F. Prunier, F. Laouafa, S. Lignon, F. Darve, Bifurcation modeling in geomaterials: from the second-order work criterion to spectral analyses. Int. J. Numer. Anal. Methods Geomech. **33**, 1169–1202 (2009b)

F. Prunier, F. Laouafa, F. Darve, 3D bifurcation analysis in geomaterials, investigation of the second order work criterion. Eur. J. Env. Civ. Eng. **13**(2), 135–147 (2009c)

L.A. Richards, Capillary conduction of liquids through porous mediums. Physics **1**(5), 318–333 (1931)

SGI (Studio Geotechnico Italiano), The Petacciato landslide: geological and geotechnical data. LESSLOSS report, 2004

L. Sibille, F.V. Donzé, F. Nicot, B. Chareyre, F. Darve, From bifurcation to failure in a granular material, a DEM analysis. Acta Geotechnica **3**(1), 15–24 (2008)

H.A.M. Van-Eekelen, Isotropic yield surfaces in three dimensions for use in soil mechanics. Int. J. Numer. Anal. Methods Geomech. **4**, 89–101 (1980)

M.T. Van-Genuchten, A closed form for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. **4**, 892–898 (1980)

R.G. Wan, P.J. Guo, Stress dilatancy and fabric dependencies on sand behavior. J. Eng. Mech. ASCE **130**(6), 635–645 (2004)

R.G. Wan, M. Pinheiro, A. Daouadji, M. Jrad, F. Darve, Diffuse instabilities with transition to localization in loose granular materials. Int. J. Numer. Anal. Methods Geomech. (2012). doi:10.1002/nag.2085

Fractals and Mechanics of Fracture

6

Michael P. Wnuk

Contents

Abstract

Classical mechanics including mechanics of fracture is often unsatisfactory when a solution predicts a singularity and the need arises to interpret the underlying physical meaning or lack thereof. A customary practice to deal with the singularity problem is to exclude a small region near the singular point, for which a different constitutive law – usually nonelastic – is postulated. This approach is adequate provided that the stress field outside the singular region is dominated by the elastic behavior. An alternative approach that successfully resolves problems involving singularities is the averaging process, also known as the quantization procedure – or – equivalently, discretization of the condition of the minimum of the potential energy of the system. In addition to the constitutive law, a certain "rule of decohesion" must be incorporated into the theory of fracture. An example of such a rule is the δCOD or the so-called "final stretch" criterion employed to describe the onset and the stable growth of a crack

M.P. Wnuk (✉)
Department of Civil Engineering and Mechanics, College of Engineering and Applied Science, University of Wisconsin, Milwaukee, WI, USA
e-mail: mpw@csd.uwm.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_18

contained in a ductile solid. This criterion generalizes the well-known criteria of Griffith, Irwin–Orowan, Rice, and Wells.

Success of the novel approaches is particularly remarkable in the nanoscale domain, where the fractal geometry of cracks and the quantization rules need to be combined in order to describe adequately fracture processes at the lattice and/or atomistic level. Discrete cohesive crack representation with the fractal geometry incorporated into the mathematical model appears to produce most straightforward and useful results. Application of the Wnuk–Yavari correspondence principle relating the fractal and smooth blunt cracks demonstrates that even a minute amount of roughness of the crack surface is sufficient to cause a drop in the maximum stress measured at the tip of the crack from infinity to a well-defined finite value.

Early stages of fracture and the pre-fracture deformation states associated with a stable propagation of a subcritical crack in viscoelastic and/or ductile solids are described in some detail. The initial stable growth of crack manifests itself as a sequence of the local instability points, while the onset of catastrophic fracture corresponds to attainment of the global instability. The locus of these critical states supplants the Griffith result. Only in the limit of ideally brittle material behavior that both results, the present one and the classic one, coincide.

In the present review, introductory concepts of fractal and quantized fracture mechanics followed by the studies of delayed fracture in viscoelastic solids and the instabilities occurring in the process of ductile fracture are discussed.

Introduction

In the seminal paper (Mandelbrot et al. 1984) entitled *Fractal character of fracture surfaces in metals* and published in 1984 in the prestigious British journal *Nature*, Mandelbrot, Passoja, and Paullay wrote:

> When a piece of metal is fractured either by tensile or impact loading the fracture surface that is formed is rough and irregular. Its shape is affected by the metal's microstructure (such as grains, inclusions and precipitates where characteristic length is large relative to the atomic scale), as well as by 'macrostructural' influences (such as the size, the shape of the specimen, and the notch from which the fracture begins). However, repeated observations at various magnifications also reveal a variety of additional structures that fall between 'micro' and 'macro' and have not yet been described satisfactorily in a systematic manner. The experiments reported here reveal the existence of broad and clearly distinct zone of intermediate scales in which the fracture is modeled very well by a fractal surface.

The purpose of this review is to show how the concept of fractal geometry can be incorporated into the mathematical analysis underlying the existing mechanics of fracture. Some progress has been achieved and certain novel ideas pertinent to this goal have been developed by Carpinteri (1994), Carpinteri et al. (2002), Carpinteri and Spagnoli (2004), Spagnoli (2005), Balankin (1997), and Cherepanov et al. (1995). From the text of Mandelbrot et al. (1984) quoted above, it follows

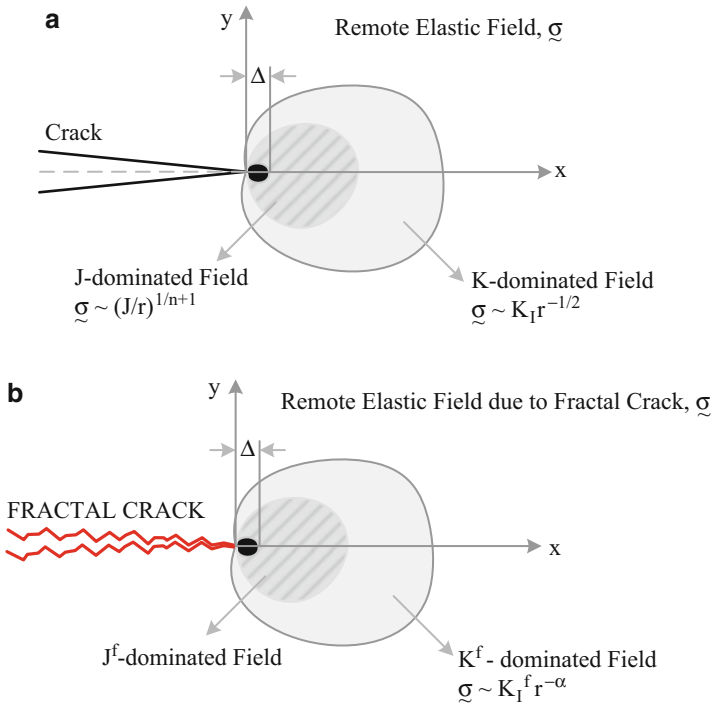


Fig. 1 (**a**) Singular asymptotic fields for a smooth crack: J-dominated HRR field and the K-field. Embedded inside the J-controlled field is the process zone, Δ. (**b**) Singular asymptotic fields for a fractal crack: J^f -dominated HRR field and K^f -field. Embedded inside the J^f -controlled field is the process zone, Δ

that fracture is not only of fractal character, but it is also of multi-scale nature. To fill the gap between the "micro" and "macro" scale levels, the group of Panin (Pugno and Ruoff 2004) in Siberia, Russia, has developed an entirely new branch of mechanics, named "Mesomechanics." In the second part of this review, certain theoretical considerations are presented, which are based on the experimental observations of the Siberian group.

For several past decades, the linear elastic fracture mechanics (LEFM) has been subject to multiple attempts to eliminate or alter the singularities present in the stress and strain fields generated by a crack contained in a brittle solid. As the singular point is approached and the distance measured from the crack tip shrinks to zero, certain new fields derived on the basis of nonelastic constitutive laws are embedded within the known asymptotic K-field. An example is described in Fig. 1, where a nonlinear stress and strain fields proposed by Hutchinson–Rice–Rosengren (the HRR fields) are based on the constitutive law of Ramberg and Osgood and – strictly speaking – valid for a nonlinearly elastic power strain-hardening material. In these studies, the hardening exponent n present in the Ramberg–Osgood

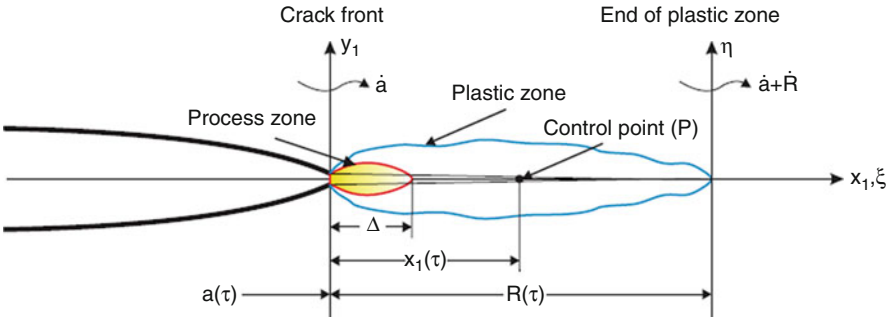


Fig. 2 Discrete cohesive zone crack model of Wnuk (1974). Note that of the two length parameters Δ and R in addition, the time-dependent crack length "a" is needed to describe quasi-static crack motion. While both "a" and R are time dependent, the process zone size Δ is the material property and it remains constant during the crack growth. Ratio R/Δ serves as a measure of material ductility; for $R/\Delta >> 1$, material is ductile, while for $R/\Delta \rightarrow 1$, material is brittle. An increasing R due to the subcritical crack extension in the early stages of fracture demonstrates the toughening phenomenon in ductile fracture, often described in terms of the material R-curve

constitutive law $\sigma \sim \varepsilon^n$ enters the expressions for the HRR stress and strain near-tip asymptotic fields, as follows: $\underset{\sim}{\sigma} \sim (J/r)^{\frac{1}{n+1}}$ and $\underset{\sim}{\varepsilon} \sim (J/r)^{\frac{n}{n+1}}$. Subsequently, such a representation has been shown to be valid for elastic–plastic fields, too; and thus, the concept of Rice's path-independent integral has been extended to elastic–plastic problems including large-scale yielding cases. Note that the product of stress and strain is inversely proportional to r, thus making all the energy integrals finite – analogous to the integrals involved in the problem solved by Griffith. Since the theory underlying the HRR field is the deformation theory of plasticity, the results are not valid whenever unloading occurs, meaning that the J-controlled field described here is valid only for stationary cracks. As soon as the crack begins to propagate, the HRR fields are no longer pertinent. New solutions valid for a moving crack remain to be found.

An interesting alternative is offered in the form of the cohesive forces crack model, for which a certain cohesion mechanism of separation is considered (either Wells' CTOD criterion for stationary cracks or Wnuk's δCOD criterion for moving cracks); see Fig. 2. This model has produced very useful results, especially when equipped with an additional element – the Neuber particle, known also as "unit step growth," "fracture quantum," or "process zone." Incorporation of the Neuber particle into the mathematical analysis leads to the quantization procedure; cf. Pugno and Ruoff (2004) and also in Wnuk and Yavari (2008, 2009). This type of cohesive forces model, which includes the Neuber-type zone, is known as "discrete cohesive crack model" and it works well both for the stationary and quasi-statically moving cracks.

In addition to the basic equations of mechanics of solids, such as the equilibrium equations and the constitutive law, certain additional relations are needed in order

to predict the onset of fracture, which is, of course, considered as propagation of the discontinuity in displacements. A certain postulate regarding the final act of decohesion is required. Many physical concepts have been added to the list of this kind of postulates, beginning with the Griffith energy criterion and then the "driving force" criterion of Irwin–Orowan and the by-product of it in the form of the critical stress intensity factor (SIF) criterion. As shown by Rice (1968), all of these criteria are equivalent – even though they have different physical interpretations. They can also be shown to be equivalent to Rice's J-integral criterion, which uses the rate of energy release as the controlling factor in the process of cracking (decohesion). Based on the notions associated with the cohesive crack model, the "CTOD criterion" or δ-criterion was proposed by Wells and then several other conceptually related conditions followed, and they now form the well-established foundation of the contemporary mechanics of fracture. One such criterion will be discussed here in more detail, and it is the δCOD criterion of Wnuk (1974). This "incremental displacement" criterion implies quantization of the relevant calculus, and it is used when instabilities of ductile fracture preceded by a slow crack growth are analyzed. In its essence it is a critical strain criterion for the onset and stable growth of a subcritical crack.

Success of the existing continuum mechanics solutions is only partial, as the singularity still persists. The problem is resolved if the cohesive forces model of the crack with the Neuber particle built into it is employed. Surprisingly, the essential results derived by the use of the structured cohesive crack model are very similar to those obtained by the approximate technique of Irwin's "plasticity correction."

In recent years some new approaches such as quantized fracture mechanics and fractal fracture mechanics have been developed. These techniques necessitate implementation of certain mathematical tools, which are novel to mechanics of fracture such as the root-mean-square averaging and a finite difference formula replacing Newton's expression for the derivative. These techniques are briefly reviewed in the present work.

Basic Concepts in Fractal Fracture Mechanics

In hope of removing the singularities in stress and strain fields generated in the vicinity of crack tip, it has been postulated that fractal geometry can be used to better represent a stationary or a propagating crack than the smooth Euclidean crack considered in the classic theory of Griffith–Irwin–Orowan; cf. Mandelbrot et al. (1984). Unfortunately, the singularity does not disappear; it acquires a new order; instead of $r^{-1/2}$ in the LEFM domain valid for the smooth crack representation, it becomes $r^{-\alpha}$ where the fractal exponent α is related to other measures of roughness of the crack surface, namely, the fractal dimension D or the Hurst fractal measure H. One has

$$\alpha = \frac{2 - D}{2}, \quad 1 \leq D \leq 2 \tag{1}$$

for a self-similar crack, and

$$\alpha = \begin{cases} \dfrac{2H-1}{2H}, & \dfrac{1}{2} \le H \le 1 \\ 0 & 0 \le H \le \dfrac{1}{2} \end{cases} \tag{2}$$

for a self-affine crack. Variation of the fractal dimension D within the range (Pugno and Ruoff 2004; Wnuk and Yavari 2008) corresponds to α changing from ½ when $D = 1$ (a smooth crack case) to zero when $D = 2$. Only for this particular case of $\alpha = 0$ the singularity in the near-tip stress field associated with a fractal crack $\sigma \sim r^{-\alpha}$ disappears. This is an interesting limit when the fractal crack "fills the plane" as it becomes a two-dimensional object. However, when the fractal geometry is taken into account, even the basic concepts of calculus such as line or area integrals are not defined (Harrison and Norton 1991; Harrison 1994). The "embedded crack model" of Wnuk and Yavari bypasses this problem, since it assumes a smooth crack contained within the stress field generated by a fractal crack. Thus, the basic mathematical manipulations are enabled. The price paid when this model is employed is a rather substantial limitation imposed on the range of crack surface roughness, which is not supposed to differ significantly from the case of the smooth crack. Somewhat unexpectedly Yavari and Khezrzadeh have recently shown (Khezrzadeh et al. 2011) that reaching a high level of crack surface roughness is physically impossible. Using a branching argument, these authors have shown that there exists a limiting roughness estimated at about $\alpha = 0.25$.

Indeed, it is expected that the model of an embedded crack can provide reliable results only for the limited range of crack roughness. In what follows influence of the fractal geometry for only moderately rough cracks has been studied, say for the exponent α not dropping below 0.4, which is not significantly different from the case of smooth crack, for which $\alpha = 0.5$. The model of embedded crack assumes that a smooth crack is placed within a field generated by a fractal crack, so it is sort of "putting cart in front of a horse." In designing this model, one first considers a specific stress field associated with a fractal crack, and then a smooth crack is embedded in this field. Of course, a better model would be desirable. Despite of these limitations, the study does provide a valuable insight into behavior of the fractal cracks.

Wnuk and Yavari (2003) were the first to use the model of embedded crack in order to estimate the stress intensity factor. Their result was

$$K_{WY}^{f} = \frac{C(a,\alpha)}{(\pi a)^{\alpha}} \int_0^a p(x) \frac{(a-x)^{2\alpha} + (a+x)^{2\alpha}}{(a^2-x^2)^{\alpha}} dx \tag{3}$$

With $p(x)$ identified with the applied stress σ and "a" denoting the crack length and with certain dimensional considerations, the constant C was chosen as $C = (a/\sqrt{\pi})^{2\alpha-1}$ in order to extend validity of the formula in Eq. 3 to a "true fractal

crack" range, which implies surpassing the limits of the embedded crack model. For the fractal dimension D contained in the interval (Pugno and Ruoff 2004; Wnuk and Yavari 2008), the final result is

$$\begin{aligned} K_{WY}^{f} &= \frac{a^{\alpha-1}\sigma}{\pi^{2\alpha-1/2}}\int_0^a \frac{(a-x)^{2\alpha}+(a+x)^{2\alpha}}{(a^2-x^2)^{\alpha}}dx = \chi_0(\alpha)\sigma\sqrt{\pi a^{2\alpha}} \\ \chi_0(\alpha) &= \frac{1}{\pi^{2\alpha}}\int_0^1 \frac{(1-z)^{2\alpha}+(1+z)^{2\alpha}}{(1-z^2)^{\alpha}}dz \\ \chi_0(\alpha) &\simeq -3.28\alpha^3+6.475\alpha^2-4.42\alpha+2 \end{aligned} \tag{4}$$

This result was later corrected in Khezrzadeh et al. (2011), and the correct K-factor for a fractal crack represented by the embedded crack model was expressed as

$$K_I^f = \left(\frac{a}{\pi}\right)^{1-\alpha}\int_0^a \frac{2\sigma}{(a^2-x^2)^{1-\alpha}}dx \tag{5}$$

or

$$\begin{aligned} K_I^f &= \chi(\alpha)\sigma\sqrt{\pi a^{2\alpha}} \\ \chi(\alpha) &= \frac{2}{\pi^{1-\alpha}}\int_0^1 \frac{dz}{(1-z^2)^{1-\alpha}} \end{aligned} \tag{6}$$

where the function $\chi(\alpha)$ can be expressed in a closed form by the use of the Euler gamma function Γ, namely,

$$\begin{aligned} \chi(\alpha) &= \frac{\pi^{\alpha-1}\Gamma(\alpha)}{\Gamma(\alpha+1/2)} \\ \chi(\alpha) &\simeq -10.37\alpha^3+15.679\alpha^2-8.234\alpha+2.493 \quad \text{for } 0.25 \le \alpha \le 0.5 \end{aligned} \tag{7}$$

For the admissible range of crack roughness, both results in Eqs. 4 and 6 are identical, and for smaller α, when the fractal dimension D approaches 2, they diverge (see Fig. 3) suggesting that the range of surface roughness beyond the limits of validity of the embedded crack model remains yet to be studied. Since the basic manipulations of the calculus are not yet defined along the fractal, a better mathematical model is needed to deal with fractals and their somewhat whimsical mathematical nature.

Once the stress intensity factor for a fractal crack is known, one may write the equations for the asymptotic near-tip stress field; cf. Wnuk and Yavari (2003) and Khezrzadeh et al. (2011). The opening stress for a fractal crack reads

$$\sigma_{yy}(r,\theta,\alpha) = \frac{K_I^f}{(2\pi r)^{\alpha}}\{\sin[\alpha(\pi-\theta)] - \alpha\sin(\theta)\cos[(\alpha+1)(\pi-\theta)]\} \tag{8}$$

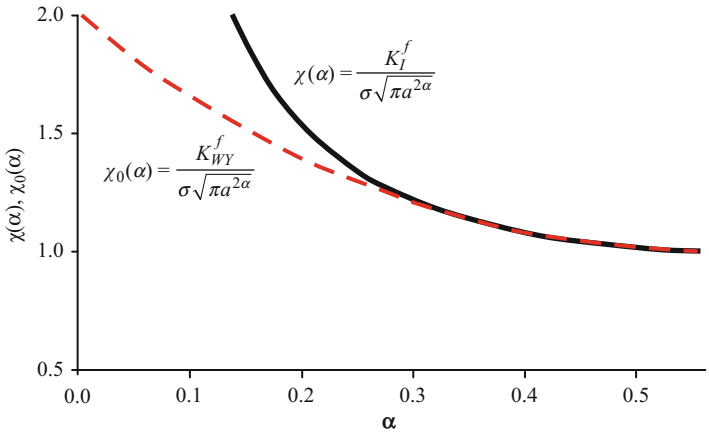


Fig. 3 Comparison between the estimate of the fractal stress intensity factor proposed by Wnuk and Yavari, K_{WY}^f (cf. Wnuk and Yavari 2003), and the solution resulting from the embedded crack model, K_I^f (cf. Khezrzadeh et al. 2011). It appears that within the interval of limited crack surface roughnesses $0.3 \leq \alpha \leq 0.4$, both solutions are identical

where the polar coordinates (r, θ) are anchored at the crack tip and the other two stresses are

$$\begin{aligned}\sigma_{xx}(r,\theta,\alpha) &= \frac{K_I^f}{(2\pi r)^\alpha}\{\sin[\alpha(\pi-\theta)] + \alpha\sin(\theta)\cos[(\alpha+1)(\pi-\theta)]\}\\ \sigma_{xy}(r,\theta,\alpha) &= \frac{K_I^f}{(2\pi r)^\alpha}\alpha\sin(\theta)\sin[(\alpha+1)(\pi-\theta)]\end{aligned} \tag{9}$$

It is understood that the absolute value of θ in Eqs. 8 and 9 should be taken. These formulae resulted from the Westergaard stress function; cf. Khezrzadeh et al. (2011):

$$\begin{aligned}Z(z,\alpha) &= \frac{K_I^f}{e^{i\mathrm{Sign}(\theta)(1/2-\alpha)\pi}(2\pi z)^\alpha}\\ z &= x+iy\end{aligned} \tag{10}$$

To obtain Eqs. 8 and 9, the well-known equations of Westergaard (1939) were used. It is noted that the Westergaard equations represent a special case of a more general formulation of the complex potential functions of Kolosov–Muskhelishvili (1933), namely,

$$\begin{aligned}\sigma_{xx} &= \mathrm{Re}Z(z,\alpha) - y\mathrm{Im}Z'(z,\alpha)\\ \sigma_{yy} &= \mathrm{Re}Z(z,\alpha) + y\mathrm{Im}Z'(z,\alpha)\\ \sigma_{xy} &= -y\mathrm{Re}Z'(z,\alpha)\end{aligned} \tag{11}$$

The origin of the Cartesian system (x,y) is placed at the center of the crack. An alternative mathematical approach was offered by Williams (1957),

who decomposed the Airy stress potential into the radial and angular functions and then obtained an exact solution for the angular part while seeking the solution for the radial part in the form of a generalized Laurent power series. He found the dominant term in this series to be proportional to $r^{-1/2}$ and thus fully confirmed the K-controlled near-tip stress field solutions found by the other researchers such as Kolosov–Muskhelishvili (1933), Westergaard (1939), and Irwin (1956).

In 2005 Wnuk and Yavari (2005) described a transformation, which made possible converting any fractal crack to an equivalent blunt crack or a notch with a finite crack tip radius. This approach not only predicts a finite stress at the root of the notch, but it also shows that the three mathematical representations of discontinuities in the displacement field, a notch, a classic Griffith crack, and a fractal crack, are mathematically related and can be treated interchangeably. The proposed rule was named "a correspondence principle," according to which any fractal crack (not excluding the classic Griffith crack when the fractal dimension reduces to $D = 1$) can be treated as an equivalent blunt crack of a given root radius, say ρ_α, which is a function of the fractal exponent α. The value of this root radius is determined from a condition that the stress certain distance away from the tip of a fractal crack r^* equals the maximum stress $\sigma_{\max}$ generated at the circumference of the blunt crack. Therefore, one needs to compare the opening stresses ahead of the blunt crack

$$\sigma_{yy}^{bc} = \frac{K_I}{\sqrt{2\pi r}} \frac{\rho}{2r} \cos\left(\frac{3\theta}{2}\right) + \frac{K_I}{\sqrt{2\pi r}} \frac{\rho}{2r} \cos\left(\frac{\theta}{2}\right) \left[1 + \sin\frac{\theta}{2} \sin\frac{3\theta}{2}\right] + \ldots \tag{12}$$

with the one valid for a fractal crack

$$\sigma_{yy}^{f} = \frac{K_I}{(2\pi r)^\alpha} \left\{ \cos\left(\alpha\theta + \alpha \sin\theta \sin\left[(\alpha+1)\theta\right]\right\} \right. \tag{13}$$

When the plane of crack is considered, $\theta = 0$, the equations above reduce:

$$\begin{aligned} \sigma_{yy}^{bc} &= \frac{K_I}{\sqrt{2\pi r}} \left(1 + \frac{\rho}{2r}\right) \\ \sigma_{yy}^{f} &= \frac{K_I^f}{(2\pi r)^\alpha} \end{aligned} \tag{14}$$

The stress expressed by the first expression in the above equations refers to a blunted crack (hence the superscript "bc"), while the second formula gives the stress in front of a fractal crack. Now, one requires that the first of these quantities is evaluated at the circumference of the blunt crack, while the second is evaluated at a certain distance from the crack tip $r = r^*$. From the first formula in Eq. 14, the stress at the circumference of the blunt crack is evaluated when the radius r is set

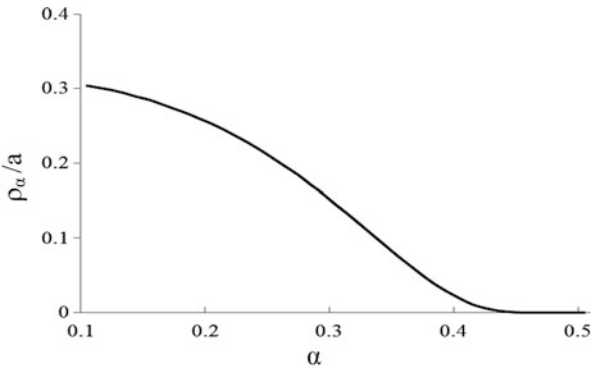


Fig. 4 Root radius of a blunt crack equivalent to a fractal crack obtained from the Wnuk–Yavari correspondence principle and defined by the fractal exponent α. Note that for α approaching ½, the radius ρ_α shrinks to zero rendering the Griffith crack case

equal $\rho/2$. Next, according to the correspondence principle, one needs to put $r = r^*$ in the second expression in Eq. 14, yielding

$$\begin{aligned}\left[\sigma_{yy}^{bc}\right]_{\max} &= \frac{2K_I}{\sqrt{\pi\rho}} = 2\sqrt{\frac{a}{\rho_\alpha}}\sigma \\ \left[\sigma_{yy}^{f}\right]_{\max} &= \frac{K_I^f}{(2\pi r^*)^\alpha} = \frac{\chi_0(\alpha)\sigma\sqrt{\pi a^{2\alpha}}}{(2\pi r^*)^\alpha}\end{aligned} \tag{15}$$

According to the correspondence principle, these two stresses should equal if the distance r^* is chosen to scale down with ρ_α, say $r^* = \varepsilon\,\rho_\alpha$, and then the requirement of equivalency of the two objects considered is expressed by the following equation:

$$\frac{2K_I}{\sqrt{\pi\rho_\alpha}} = \frac{K_I^f}{(2\pi\varepsilon\rho_\alpha)^\alpha} \tag{16}$$

Hence, one can evaluate the required root radius of the equivalent blunt crack:

$$\rho_\alpha = \frac{a}{\pi}\left[\frac{\chi_0(\alpha)}{2^{\alpha+1}\varepsilon^\alpha}\right]^{\frac{2}{2\alpha-1}} \tag{17}$$

With $\varepsilon = 0.05$, chosen after a careful parametric study (cf. Wnuk and Yavari 2005), this equation provides the function shown in Fig. 4. When the stress Eq. 15 is evaluated, the predicted maximum stress of the equivalent blunt crack becomes the following function of the fractal exponent:

$$\sigma_{\max} = 2\sqrt{\pi}\sigma\left[\frac{\chi_0(\alpha)}{2^{\alpha+1}\varepsilon^\alpha}\right]^{\frac{1}{1-2\alpha}} \tag{18}$$

The function is shown in Fig. 5. It is seen that with the exception of the classic Griffith crack ($\alpha = 0.5$), the stress is finite. It is noteworthy that the

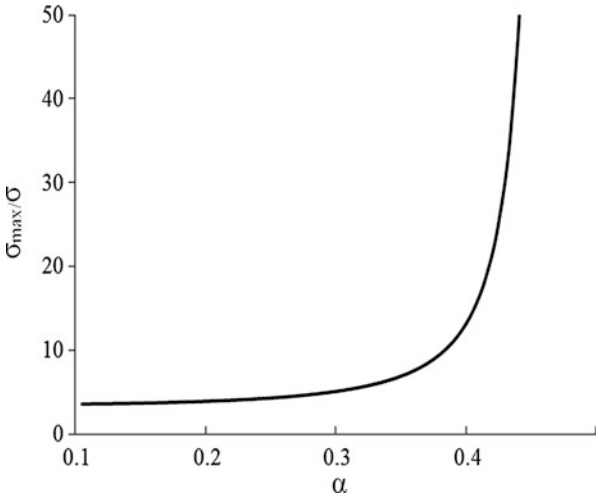


Fig. 5 Maximum stress at the root of a fractal crack equivalent to a certain smooth blunt crack, according to the correspondence principle of Wnuk and Yavari (2005), is shown to be finite for all values of the fractal exponent α with exception of the classic Griffith crack, when $\alpha = ½$. The graph illustrates that even a minute amount of roughness of the crack surface is sufficient to cause a drop in the maximum stress from infinity to a well-defined finite value

maximum stress given in Eqs. 15 and 18 also satisfies the Inglis (Inglis 1913) formula for a notch when $a >> \rho_\alpha$, namely,

$$\sigma_{\max}^{\text{Inglis}} = 2\sigma\sqrt{\frac{a}{\rho_\alpha}} \tag{19}$$

This feature of the basic concept underlying the correspondence formula for fractal cracks appears to close the train of thought, which from Inglis (1913) to Griffith (1921a) led to the present-day foundations of the fracture mechanics. When molecular theory of fracture is pondered, then Eq. 19 corresponds to the molecular strength, which is believed to be one order of magnitude smaller than the Young modulus.

Similarly, the Irwin plastic correction to the LEFM model, say r_y^f for a fractal crack (cf. Wnuk and Yavari 2005), has been evaluated. Using the statically equivalent elastic–plastic stress field, the entity r_y^f has been estimated as follows:

$$\int_0^{r_y^f} \frac{K_I^f}{(2\pi r)^\alpha} dr - r_y^f \sigma_Y = \left(r_p^f - r_y^f\right)\sigma_Y \tag{20}$$

Following Irwin the stress generated by a fractal crack is required to equal the yield stress σ_Y at the distance r_y^f from the crack tip. It has been found that

$$r_y^f = \frac{1}{2\pi}\left(\frac{K_I^f}{\sigma_Y}\right)^{\frac{1}{\alpha}} \tag{21}$$

The size of the plastic zone (in the form of a disc adjacent to the crack front and inserted into the elastic stress field) is calculated from Eq. 20 as follows:

$$r_p^f = \frac{1}{2\pi(1-\alpha)}\left(\frac{K_I^f}{\sigma_Y}\right)^{\frac{1}{\alpha}} \tag{22}$$

The opening stress within the plastic zone $0 \le r \le r_p{}^f$ is constant and equals the yield stress σ_Y, while for distances larger than r_p^f, the stress obeys the elastic fractal law and decays as $K_I^f/(2\pi r)^\alpha$.

In quantized mechanics of fracture, one assumes that a crack propagates in a continuum but in discrete steps. Therefore, a modification of the mathematical analysis is needed. Such quantization procedures applied to the structured cohesive stress crack model have been described in Wnuk and Yavari (2008, 2009). Employing the concept of the configurational force, these authors proposed that the energy release per unit of fractal measure of the crack surface is defined in the following way:

$$G^f = \frac{\left(K_I^f\right)^2}{E'} = \frac{\chi_0(\alpha)^2\sigma^2\pi a^{2\alpha}}{E'} \tag{23}$$

With the help of an auxiliary smooth crack (cf. "embedded crack model"), integration of this expression with respect to the crack length yields

$$-\Pi^f = \int_0^a G^f(\sigma, a)d(2a) = \frac{2\chi_0(\alpha)^2\sigma^2\pi a^{2\alpha+1}}{(2\alpha+1)E'} \tag{24}$$

Discontinuous nature of crack extension calls for replacement of the notion of the derivative expressed in terms of infinitesimals $\mathrm{d}\Pi^f$ and da (or $d\ell$) by the finite quantities $\Delta\Pi^f$ and Δa (or $\Delta\ell$). The derivative $\mathrm{d}\Pi^f/\mathrm{d}\ell$, which defines the crack driving force (or a configurational force), is thus replaced by the following entity involving finite differences $\Delta_\Delta\Pi^f$ and $\Delta\ell$, namely,

$$G^f = -\frac{\Delta_\Delta\Pi^f}{\Delta\ell} = -\frac{\Pi^f(a+\Delta) - \Pi^f(a)}{2\Delta} \tag{25}$$

Symbol Δ denotes the fracture quantum and for a two-ended crack, $\Delta\ell = 2\Delta$, while E' equals the Young modulus E for a plane stress case and $E/(1-\nu^2)$ for the plane strain; ν is the Poisson ratio. Substituting Eq. 24 into Eq. 25 generates

$$G^f = \frac{\chi_0(\alpha)^2\sigma^2\pi}{\Delta(2\alpha+1)E'}\left[(a+\Delta)^{2\alpha+1} - a^{2\alpha+1}\right] \tag{26}$$

At the onset of fracture, it has been postulated $G^f = G_c^f$, where the critical energy release rate $G_c^f = (K_c^f)^2/E'$ has the dimensions of stress x length$^{2\alpha}$;

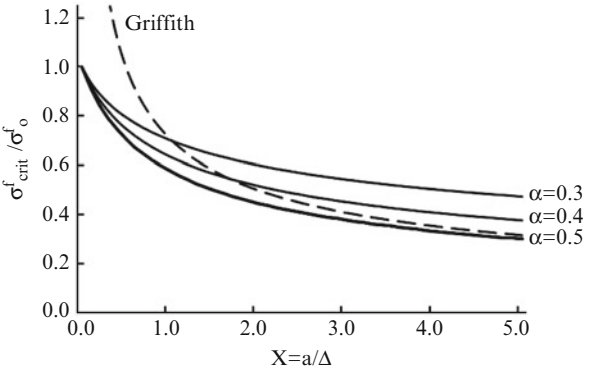


Fig. 6 Critical stress predicted by the fractal fracture mechanics for three levels of the crack surface roughness compared to that obtained from the Griffith theory. All stresses in this graph have been normalized by the inherent strength of an undamaged material, for which crack length is assumed zero

hence the critical stress for a fractal quantized fracture obtained from the equality $G^f = G_c^f$

$$\sigma^f_{\text{crit}} = \sqrt{\frac{(2\alpha+1)E'G^f_c}{\pi}}\frac{\sqrt{\Delta}}{\chi_0(\alpha)\sqrt{(a+\Delta)^{2\alpha+1}-a^{2\alpha+1}}} \tag{27}$$

does indeed have the dimension of stress. For the zero crack length "a," this formula predicts a finite critical stress identified as an inherent material strength, namely,

$$\sigma^f_0 = \sqrt{\frac{(2\alpha+1)E'G^f_c}{\pi}}\left(\frac{1}{\chi_0(\alpha)\Delta^\alpha}\right) = \frac{K^f_c}{K_c}\sigma_0\Delta^{\frac{1-2\alpha}{2}}\frac{\sqrt{\alpha+1/2}}{\chi_0(\alpha)} \tag{28}$$

As can be readily verified for α approaching ½, this equation reduces to the inherent strength σ_0 predicted earlier for a solid containing a smooth crack.

When the expression Eq. 28 is used as a normalizing constant for the critical stress, Eq. 27 assumes the form

$$s^f_{\text{crit}} = \frac{\sigma^f_{\text{crit}}}{\sigma^f_0} = \frac{1}{\sqrt{\left[\left(1+\frac{a}{\Delta}\right)^{2\alpha+1}-\left(\frac{a}{\Delta}\right)^{2\alpha+1}\right]}} \tag{29}$$

This result is illustrated in Fig. 6. For α approaching ½, this formula reduces to that given by Pugno and Ruoff in their treatise (Pugno and Ruoff 2004) of the quantum fracture mechanics (QFM), namely,

$$s^{QFM}_{\text{crit}} = \frac{1}{\sqrt{1+2\left(\frac{a}{\Delta}\right)}} \tag{30}$$

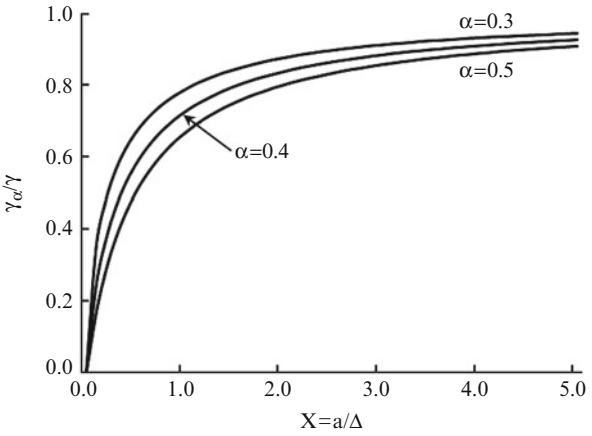


Fig. 7 Surface energy for small fractal cracks appears to be the function of the crack length – at least for the nanoscale range, where the crack lengths are on the order of magnitude of the size of Neuber particle. The surface energy is also affected by the fractal exponent α, showing higher specific surface energy for rougher cracks. This effect is referred to as "toughening phenomenon" ascribed to the influence of the fractal geometry

Using those equations, Wnuk and Yavari (2008) predict the relationship of the surface energy on the fractal exponent and the crack length

$$\frac{\gamma_\alpha}{\gamma} = \frac{(2\alpha+1)(a/\Delta)^{2\alpha}}{\left(1+\frac{a}{\Delta}\right)^{2\alpha+1} - \left(\frac{a}{\Delta}\right)^{2\alpha+1}} \tag{31}$$

The surface energy for a fractal crack γ_s is normalized by the LEFM surface energy γ.

The curve resulting from Eq. 31 shows a rapid increase of γ from zero at $a = 0$ to one over a distance on the order of magnitude of fracture quantum Δ; this trend is in good agreement with the experimental results of Ippolito et al. (2006); see Fig. 7. Even though for cracks of size orders of magnitude greater than the fracture quantum Δ $(X >> 1)$, the surface energy is a material constant, it is not necessarily so in the realm of nanostructured brittle materials, for which the crack length is comparable to the atomistic characteristic length or to the size of the Neuber particle. In this domain the quantized fracture mechanics is superior to the classic mechanics as it provides a correct representation of the basic lattice features at the atomistic level. In this way the domain of the continuum mechanics has been enlarged to include the atomistic structure of a solid. Comparison of the curves shown in Fig. 7 proves that at nanoscale the fractal geometry of cracks leads to an enhancement of the fracture resistance γ. Therefore, one may speak about a "toughening effect" due to an increased material resistance to crack propagation; this statement is in complete agreement with the experimental findings of Ippolito et al. (2006) in their studies of onset and spread of fracture in a silicon carbide matrix.

Similar "toughening effect" is observed if the characteristic lengths R_{coh} and $R_{\text{coh}}^{\text{discrete}}$ are compared. As it follows from Wnuk and Yavari (2009) for the case of $R << a$, the relation between the two measures of inherent material resistance to crack propagation is

$$R_{\text{coh}}^{\text{discrete}} = \frac{\pi K_I^2}{8\sigma_Y^2}\left(1 + \frac{1}{2a/\Delta}\right) \tag{32}$$

The expression in front of the parentheses stands for the R_{coh}, while the correction factor results from the quantization process, and it becomes significant only for very small cracks, when the crack length and the fracture quantum are comparable in size.

It is noteworthy that for α approaching ½ (a smooth crack case), the formula in Eq. 28 predicts the inherent strength of an undamaged solid as

$$\sigma_0 = \sqrt{\frac{2}{\pi\Delta}} K_c \tag{33}$$

while Eq. 29 reduces to

$$\sigma_{\text{crit}}^{QFM} = \frac{\sigma_0}{\sqrt{1 + 2\dfrac{a}{\Delta}}} \tag{34}$$

One of the curves depicted in Fig. 4 represents this equation; see the case of $\alpha = ½$. As expected this result coincides exactly with the curve obtained from the quantized fracture mechanics.

The same can be said about the last two results in Eqs. 33 and 34, which are in complete agreement with the equations derived by Pugno and Ruoff (2004) for the case of quantized fracture. It is noted that these authors have used the method of averages over the length Δ, while here the finite difference technique is applied in conjunction with the potential Π^f of a body containing a fractal crack. The results are the same. An example of a root-mean-square averaging of the stress intensity factor is given as follows:

$$\langle K \rangle = \left(\frac{1}{\Delta}\int_a^{a+\Delta} K_I^2(a)da\right)^{\frac{1}{2}} = \sigma\sqrt{\pi}\left(\frac{1}{\Delta}\int_a^{a+\Delta} ada\right)^{\frac{1}{2}} = \sigma\sqrt{\pi}\sqrt{a + \frac{\Delta}{2}} \tag{35}$$

When this is set equal to K_c, the quantized fracture mechanics results in Eqs. 33 and 34 are recovered. This shows again that our finite difference approach and the root-mean-square averaging technique proposed by Pugno et al. (2004) for K-factors are mathematically equivalent. It is noteworthy that these recent theoretical developments and the ensuing progress in the contemporary mechanics of fracture are continuations of the early intuitive concepts suggested by Neuber (1958) and Novozhilov (1969).

In Wnuk and Yavari (2005) have using the averaging technique to quantize the classic cohesive crack model extended the quantized mechanics analysis onto the fractal cracks as explained in (Wnuk and Yavari 2008, 2009). Valuable contributions to fractal fracture mechanics are due to the Russian scientists Borodich (1992, 1997, 1999), Mosolov (1991), and Goldstein and Mosolov (1992). Two recent works of Wnuk et al. (2012, 2013) illustrate the applications of Wnuk's final stretch criterion (or δCOD criterion) to the problems of time-dependent fracture in viscoelastic and ductile solids. Noteworthy results were also obtained by Balankin (1997) and reviewed by Cherepanov et al. (1995).

The fracture criterion of Wells (1961) has been shown to be equivalent to the criterion based on the Rice path-independent J-integral; cf. Shih (1981). Shih (1981) and also Wnuk et al. (1997, 1998) have studied the relation between the two criteria, related by a simple expression $J = d_n(CTOD/\sigma_Y)$. Numerical solutions for the coefficient d_n, based on the analysis of the HHR J-controlled asymptotic stress field, were given in the form of tables (Shih 1981) and certain approximate closed form formulae in Wnuk and Omidvar (1997) and Wnuk et al. (1998).

An essential conclusion of these considerations is the fact that the fractal geometry presents a new and useful tool in describing fracture of brittle and quasi-brittle solids. When combined with the quantized mechanics of fracture, it allows one to reach into the realm of atomistic modeling of the fracture process. This feature is especially valuable when the objective of one's research is the study of the lattice effects on deformation and fracture. In this way the range of validity of the classical mechanics of fracture has been extended into the nanoscale domain.

Delayed Fracture in Viscoelastic Solids for Euclidian and Fractal Geometries: Motion of a Smooth Crack in a Viscoelastic Medium

Effects of two parameters on the time-dependent fracture manifested by a slow stable crack propagation that precedes catastrophic failure in viscoelastic and in ductile materials have been studied. One of these parameters is related to the material ductility (ρ), and the other describes the geometry (roughness) of crack surface and is measured by the degree of fractality represented by the fractal exponent α or – equivalently – by the Hausdorff fractal dimension D for a self-similar crack.

Within a certain range of the applied load, there are two distinct stages of deformation process and the ensuing fracture. These are the incubation stage, when the crack remains stationary, followed by the onset of slow crack growth that eventually ends up with a transition to the catastrophic fracture. Comparison of the results pertaining to smooth and rough cracks demonstrates that the roughness of the crack surfaces slows down the growth of such "creeping cracks" observed in polymers. This phenomenon enhances the lifetime of the specimen and increases

the crack length, at which the critical point is reached. Interestingly, the incubation time was found to be independent of the fractal geometry.

These studies of early stages of the delayed fracture in polymeric materials, sometimes referred to as "creep rupture," are then compared to the slow stable crack growth in the ductile solids. Despite different physical mechanisms involved in the preliminary stable crack extension and despite different mathematical representations, a remarkable similarity of the end results pertaining to the two phenomena of slow crack growth (SCG) that occur either in viscoelastic or in ductile media has been demonstrated.

In the late 1960s and the early 1970s of the past century, a number of physical models and mathematical theories have been developed to provide a better insight and a quantitative description of the early stages of fracture in polymeric materials. In particular two phases of fracture initiation and subsequent growth have been considered: (1) the incubation phase during which the displacements of the crack surfaces are subject to creep process but the crack remains dormant and (2) slow propagation of a crack embedded in a viscoelastic medium. According to the linear theory of viscoelastic solids, the material response to the deformation process obeys the following constitutive relations:

$$\begin{aligned} s_{ij}(t,x) &= \int_0^t G_1(t-\tau)\frac{\partial e_{ij}(\tau,x)}{\partial \tau}d\tau \\ s(t,x) &= \int_0^t G_2(t-\tau)\frac{\partial e(\tau,x)}{\partial \tau}d\tau \end{aligned} \tag{36}$$

Here s_{ij} is the deviatoric part of the stress tensor, s denotes the spherical stress tensor, while $G_1(t)$ and $G_2(t)$ are time-dependent relaxation moduli for shear and dilatation, respectively. The inverse relations read

$$\begin{aligned} e_{ij}(t,x) &= \int_0^t J_1(t-\tau)\frac{\partial s_{ij}(\tau,x)}{\partial \tau}d\tau \\ e(t,x) &= \int_0^t J_2(t-\tau)\frac{\partial s(\tau,x)}{\partial \tau}d\tau \end{aligned} \tag{37}$$

Symbols e_{ij} and e are used to denote the deviatoric and spherical strain tensors and $J_1(t)$ and $J_2(t)$ are the two creep compliance functions. For a uniaxial state of stress, these last two equations reduce to a simple form

$$\varepsilon(t) = \int_0^t J(t-\tau)\frac{\partial \sigma(\tau)}{\partial \tau}d\tau \tag{38}$$

The relaxation moduli $G_1(t)$ and $G_2(t)$ and the creep compliance functions $J_1(t)$ and $J_2(t)$ satisfy the following integral equations:

$$\begin{aligned}\int_0^t G_1(t-\tau)J_1(\tau)d\tau = t\\ \int_0^t G_2(t-\tau)J_2(\tau)d\tau = t\end{aligned} \tag{39}$$

For a uniaxial state of stress, these equations reduce to a single relation between the relaxation modulus $E_{\text{rel}}(t)$ and the creep compliance function $J(t)$

$$\int_0^t E_{\text{rel}}(t-\tau)J(\tau)d\tau = t \tag{40}$$

Atomistic model of delayed fracture was considered by Zhurkov (1965), but this molecular theory had no great impact on the further development of the theories based in the continuum mechanics approaches. Inspired by Max Williams, W. G. Knauss of Caltech in his doctoral thesis considered time-dependent fracture of viscoelastic materials (Knauss 1965). Similar research was done by Willis (1967) followed by simultaneous researches of Williams (1967, 1968, 1969b), Wnuk and Knauss (1970), Field (1971), Wnuk (1968a, 1969, 1971, 1972), and also by Knauss and Dietmann (1970), Mueller and Knauss (1971a, b), Graham (1968), Kostrov and Nikitin (1970), Mueller (1971), Knauss (1973) and Schapery (1973).

What follows in this section is an attempt to present a brief summary of the essential results, which have had a permanent impact on the development of the mechanics of time-dependent fracture. After this review is completed, an interesting analogy of delayed fracture in polymers will be indicated (a material property intricately related to the ability to creep) with the "slow crack growth" (SCG) occurring in ductile solids due to the redistribution of strains within the yielded zone preceding the front of a propagating crack.

Two stages of delayed fracture in viscoelastic media, incubation and propagation, are described, respectively, by two governing equations: (1) Wnuk–Knauss equation and (2) Mueller–Knauss–Schapery equation. The duration of the incubation stage can be predicted from the Wnuk–Knauss equation:

$$\Psi(t_1) = \frac{J(t_1)}{J(0)} = \left(\frac{K_G}{K_0}\right)^2_{a=a_0=\text{const}} \tag{41}$$

Mueller–Knauss–Schapery equation relates the rate of crack growth $\overset{o}{a}$ to the applied constant load σ_0 and the material properties such as the unit

step growth Δ, usually identified with the process zone size, and the Griffith stress $\sigma_G = \sqrt{{}^{2E\gamma}/_{\pi a_0}}$, namely,

$$\Psi\left(\frac{\Delta}{\overset{\circ}{a}}\right) = \frac{J\left(\Delta/\overset{\circ}{a}\right)}{J(0)} = \left(\frac{K_G}{K_0}\right)^2 \tag{42}$$

For a constant crack length equal to the length of the initial crack a_0, the right-hand side in Eq. 41 reduces to the square of the ratio of the Griffith stress to the applied stress

$$n = \left(\frac{\sigma_G}{\sigma_0}\right)^2 \tag{43}$$

This quantity is sometimes referred to as "crack length quotient" – it determines how many times the actual crack is smaller than the critical Griffith crack. Therefore, the larger is the number "n," the further away is the initial defect from the critical point of unstable propagation predicted for a Griffith crack embedded in a brittle solid. For large "n" the crack is too short to initiate the delayed fracture process; see expression in Eq. 50 for the definition of the $n_{\max}$. Beyond $n_{\max}$ growth of the crack cannot take place. For $n > n_{\max}$ one can assume that these are stable cracks, which – according to the theory presented here – will never propagate. These are so-called dormant cracks that belong to a "no-growth" domain; see Appendix.

When crack length "a" is not constant, but it can vary with time $a = a(t)$, then the right side in Eq. 42 reads

$$\left(\frac{\sigma_G}{\sigma_0}\right)^2 \frac{a_0}{a} = \frac{n}{\varsigma} \tag{44}$$

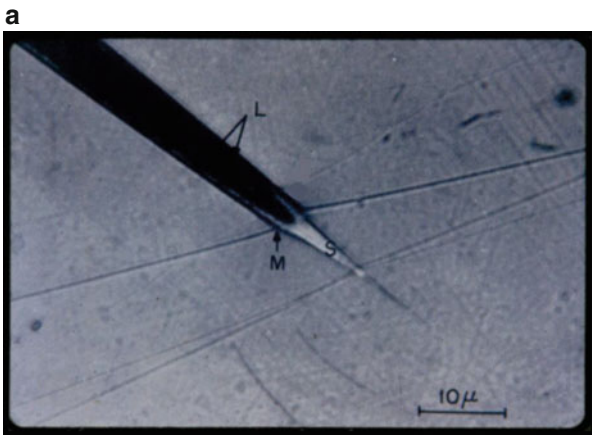
Here ζ denotes the nondimensional crack length, $\varsigma = a/a_0$. It is noteworthy that the physical meaning of the argument $\Delta/\overset{o}{a}$ appearing in Eq. 42 is the time interval needed for the tip of a moving crack to traverse the process zone adjacent to the crack tip, say

$$\delta t = \Delta/\overset{o}{a} \tag{45}$$

The location of the process zone with respect to the cohesive zone which precedes a propagating crack is shown in Fig. 8.

To illustrate applications of Eqs. 41 and 43, the constitutive equations valid for the standard linear solid (see Fig. 9) will be used. With β_1 denoting the ratio of the moduli E_1/E_2, the creep compliance function for this solid is given as

$$J(t) = \frac{1}{E_1}\{1 + \beta_1[1 - \exp(-t/\tau_2)]\} \tag{46}$$



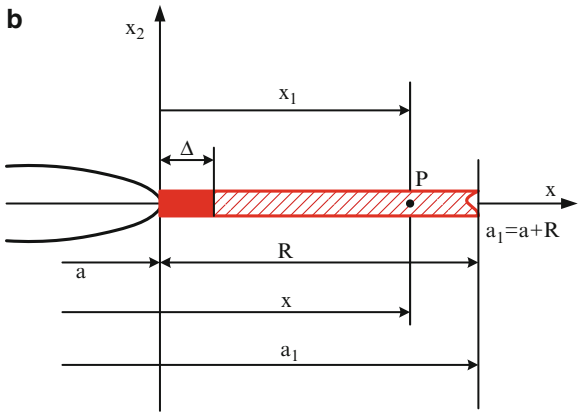


Fig. 8 (**a**) Microcrack preceded by a craze in Solithane 50/50 mimicking mechanical response of the solid rocket fuel and (**b**) details of the discrete cohesive crack model. When employed to glassy polymers, the cohesive zone coincides with the craze preceding the crack, and the Neuber particle Δ is incorporated into the equation of crack motion pertinent to the delayed fracture phenomenon in polymers as proposed by Knauss–Mueller–Dietmann (Borodich 1997, 1999; Mosolov 1991) and also by Schapery (Zhurkov 1965)

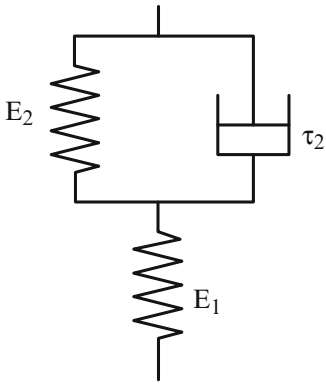


Fig. 9 Schematic diagram of the standard linear model of a viscoelastic solid. Despite its simplicity, the model is able to explain the instantaneous elastic response, creep, and stress relaxation

Therefore, the nondimensional creep compliance function $\Psi(t) = J(t)/J(0)$ reads

$$\Psi(t) = 1 + \beta_1[1 - \exp(-t/\tau_2)] \tag{47}$$

Substituting this expression into Eq. 41, one obtains

$$1 + \beta_1[1 - \exp(-t_1/\tau_2)] = n \tag{48}$$

Solving for t_1 one obtains the following prediction for the incubation time valid for a material represented by standard linear solid:

$$t_1 = \tau_2 \ln\left(\frac{\beta_1}{1 + \beta_1 - n}\right) \tag{49}$$

Inspection of Eq. 49 reveals that the quotient "n" should not exceed a certain limiting level

$$\begin{aligned} n_{\max} &= 1 + \beta_1 \\ s_{\min} &= \frac{1}{\sqrt{1 + \beta_1}} \end{aligned} \tag{50}$$

Physical interpretation of this relation can be stated as follows: for short cracks, when $n > n_{\max}$, there is no danger of initiating the delayed fracture process. These subcritical cracks are permanently dormant and they do not propagate. Perhaps, this is easier to understand when the variable $s = \sigma_0/\sigma_G$ is used instead of n. For $s < s_{\min}$ the preexisting cracks contained in a viscoelastic medium do not propagate.

Figure 10a illustrates the relationship between the incubation time and the loading parameter given either as n or s $(= 1/\sqrt{n} = \sigma_0/\sigma_G)$. Figure 10b shows an analogous relation between the time used in the process of crack propagation and the loading parameter s. Note that the incubation time is expressed in units of the relaxation time τ_2, while the time measured during the crack propagation phase of the delayed fracture is expressed in units of (τ_2/δ), wherein the constant δ contains the initial crack length a_0 and the characteristic material length Δ; cf. Eq. 53. When the variable s is used on the vertical axis and the pertinent function is plotted against the logarithm of time, then it is seen that a substantial portion of the curve appears as a straight line. This confirms the experimental results of Knauss and Dietmann (1970) used also by Schapery (1973) and Mohanty (1972).

To describe motion of a crack embedded in viscoelastic solid represented by the standard linear model, one needs to insert Eq. 46 into the governing Eq. 42. The equation of motion reads then

$$1 + \beta_1[1 - \exp(-\delta t/\tau_2)] = \frac{n}{\varsigma} \tag{51}$$

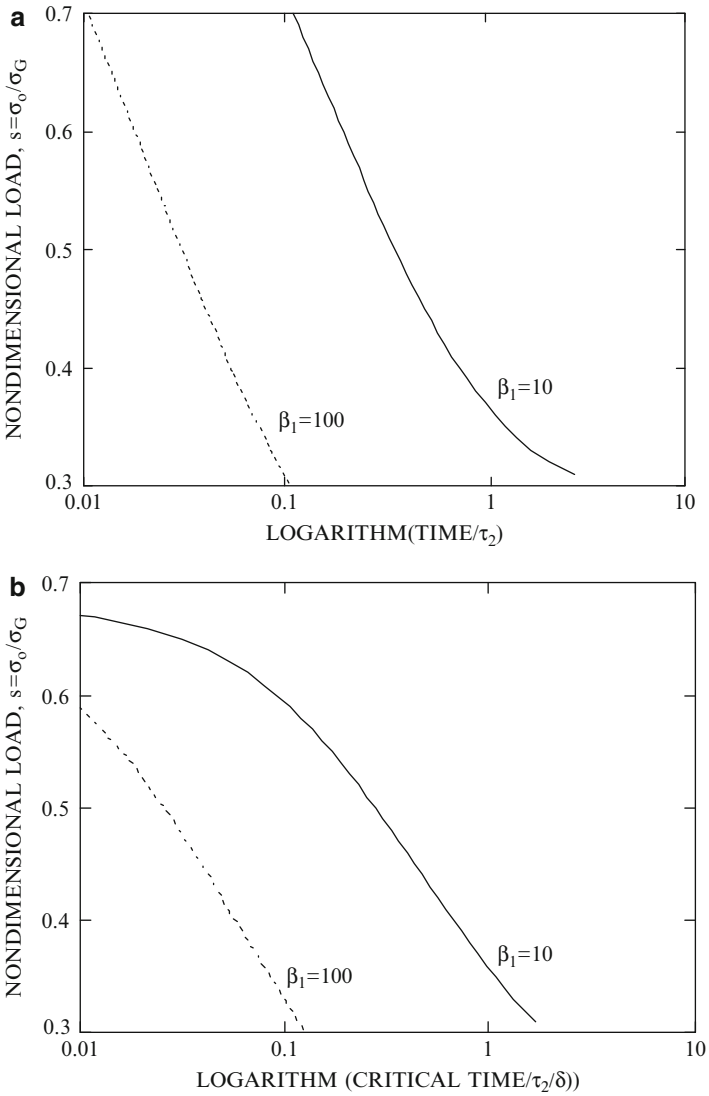


Fig. 10 (**a**) Logarithm of the incubation time in units of τ_2 shown as a function of the loading parameter s for two different values of the material constant $\beta_1 = E_1/E_2$. (**b**) Logarithm of the time to failure used during the crack propagation phase, in units of τ_2/δ, shown as a function of the loading parameter $s = \sigma_0/\sigma_G$ for two different values of the material constant $\beta_1 = E_1/E_2$

Solving it for the time interval $\delta t/\tau_2$ ($= \Delta / \overset{\circ}{a}\, \tau_2$) yields

$$\frac{\Delta}{\tau_2 \overset{o}{a}} = \ln\left(\frac{\beta_1}{1+\beta_1-\dfrac{n}{\varsigma}}\right) \tag{52}$$

It is seen from Eq. 52 that for the motion to exist, the quotient n should not exceed the maximum value defined by Eq. 50. For $n > n_{\max}$ the cracks are too small to propagate.

If nondimensional notation for the length and time variables is introduced

$$\begin{aligned} \delta &= \Delta/a_0 \\ \theta &= t/\tau_2 \end{aligned} \tag{53}$$

the left-hand side of Eq. 52 can be reduced as follows:

$$\frac{\delta t}{\tau_2} = \frac{\Delta}{\tau_2\, a} = \frac{\Delta}{\dfrac{d(\varsigma a_0)}{d(\theta\tau_2)}\tau_2} = \frac{\Delta \Big/ a_0}{d\varsigma \Big/ d\theta} \tag{54}$$

When this is inserted into Eq. 52 and with $\delta = \Delta/a_0$, the following differential equation results:

$$\frac{d\varsigma}{d\theta} = \delta\left[\ln\left(\frac{\beta_1}{1+\beta_1-\frac{n}{\varsigma}}\right)\right]^{-1} \tag{55}$$

or, after separation of variables,

$$(\delta)d\theta = \ln\left(\frac{\beta_1}{1+\beta_1-\dfrac{n}{\varsigma}}\right)d\varsigma \tag{56}$$

Motion begins at the first critical time t_1, which designates the end of the incubation period. Therefore, the lower limit for the integral applied to the left-hand side of Eq. 56 should be $\theta_1 = t_1/\tau_2$, while the upper limit is the current nondimensional time $\theta = t/\tau_2$. The corresponding upper limit to the integral on the right-hand side of Eq. 56 is the current crack length $\zeta = a/a_0$, while the lower limit is one. Upon integration one obtains

$$\int_{t_1/\tau_2}^{t/\tau_2} d\theta = \left(\frac{1}{\delta}\right)\int_1^{\varsigma} \ln\left(\frac{\beta_1}{1+\beta_1-\dfrac{n}{z}}\right)dz \tag{57}$$

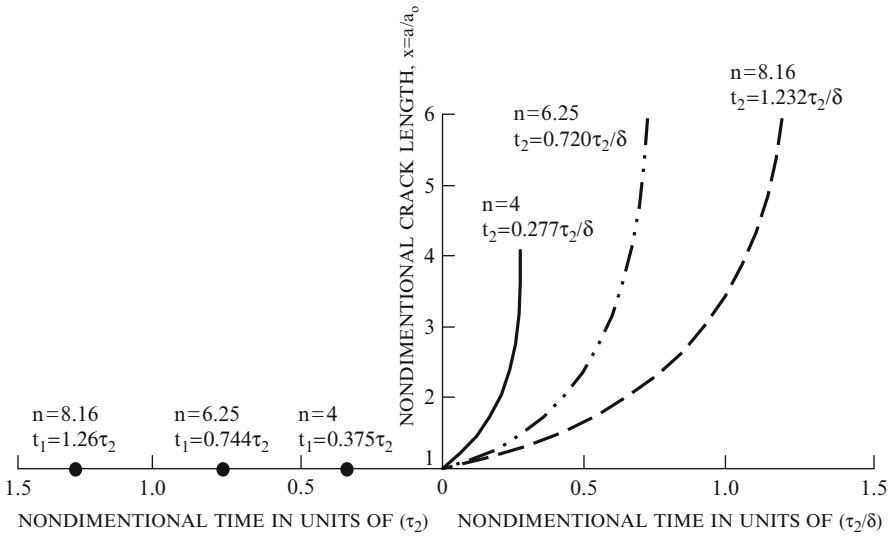


Fig. 11 Slow crack propagation occurring in a linear viscoelastic solid represented by the standard linear model depicted in Fig. 2 at $\beta_1 = 10$. Crack length is shown as a function of time; points marked on the negative time axis designate the incubation times corresponding to the given level of the applied constant load n and expressed in units of τ_2. The time interval between the specific point t_1 and the origin of the coordinates provides the duration of the incubation period. Crack propagation begins at $t = 0$. Symbol t_2 denotes time to failure, which is the time used during the quasi-static phase of crack extension, and it is expressed in units of (τ_2/δ). Constant δ is related to the characteristic material length, the so-called unit growth step Δ

The resulting expression relates the crack length x to time t, namely,

$$t - t_1 = \left(\frac{\tau_2}{\delta}\right)\int_1^{\varsigma} \ln\left(\frac{\beta_1}{1+\beta_1-\frac{n}{z}}\right) dz \tag{58}$$

If the closed form solution for the integral in Eq. 58 is used, then this formula can be cast in the following final form:

$$t = t_1 + \left(\frac{\tau_2}{\delta}\right)\left\{\varsigma \ln\left[\frac{\varsigma\beta_1}{(1+\beta_1)\varsigma - n}\right] + \frac{n}{1+\beta_1}\ln\left[\frac{(1+\beta_1)\varsigma - n}{1+\beta_1-n}\right] + \ln\left(\frac{1+\beta_1-n}{\beta_1}\right)\right\} \tag{59}$$

This equation has been used in constructing the graphs shown in Fig. 11. At $\beta_1 = 10$ three values of n have been used (4.00, 6.25, and 8.16, which corresponds to the following values of s: 0.5, 0.4, and 0.35). It can be observed that at x approaching n, the phase of the slow crack propagation is transformed into

unrestrained crack extension tantamount to the catastrophic fracture. The point in time, at which this transition occurs, can be easily seen on the horizontal axis of Fig. 11. This point of transition into unstable propagation can also be predicted from Eq. 59; substituting n for ζ, one obtains the time to fracture

$$t_2 = \left(\frac{\tau_2}{\delta}\right)\left\{\frac{n}{1+\beta_1}\ln\left[\frac{\beta_1 n}{1+\beta_1-n}\right] + \ln\left[\frac{1+\beta_1-n}{\beta_1}\right]\right\} \tag{60}$$

If the incubation time t_1 given by Eq. 49 is now added to Eq. 60, one obtains the total lifetime of the component, namely,

$$T_{\mathrm{cr}} = t_1 + t_2 = \tau_2 \ln\left(\frac{\beta_1}{1+\beta_1-n}\right) + \left(\frac{\tau_2}{\delta}\right)\left\{\frac{n}{1+\beta_1}\ln\left[\frac{\beta_1 n}{1+\beta_1-n}\right] + \ln\left[\frac{1+\beta_1-n}{\beta_1}\right]\right\} \tag{61}$$

Summarizing the results of this section, one can state that the delayed fracture in a viscoelastic solid can be mathematically represented by four expressions:

- Time of incubation t_1 given by Eq. 49 for standard linear model.
- Equation of motion given by Eq. 59 for the same material model and defining ζ as a function of time, $\zeta = \zeta\,(t)$.
- Time to fracture t_2 due to crack propagation given by Eq. 60.
- Lifetime T_{cr} equal to the sum $t_1 + t_2$, as given by Eq. 61. It is noted that while the first term in the expression Eq. 61 involves the relaxation time, material constant β_1, and the quotient n, the second term in Eq. 61 contains also the internal structural constant δ. It is also noted that for the quotient n approaching one, both terms in Eq. 61 are zero, while for n exceeding $n_{\max}$, the expression looses the physical sense (since in that case there is no propagation). With the constant δ being on the order of magnitude varying within the range 10^{-3}–10^{-6}, the second term in Eq. 61 is substantially greater than the first term which represents the incubation time; see also Appendix.

For $\beta_1 = 10$ and three different levels of n, the resulting functional relationships between the crack length x and time t are shown in Fig. 11 along with the values of the incubation times, expressed in units of (τ_2), and the times-to-failure expressed in units of (τ_2/δ). A numerical example is given in the Appendix.

Example described here, involving the standard linear solid, serves as an illustration of the mathematical procedures necessary in predicting the delayed fracture in polymeric materials. Knauss and Dietmann (1969) and Schapery (1973) have shown how the real viscoelastic materials, for which the relaxation modulus $G(t)$ and the creep compliance function $J(t)$ are measured (or calculated from Eq. 4) and then used in the governing equations of motion discussed above, can provide a good approximation of the experimental data.

Growth of Fractal Cracks in Viscoelastic Media

To extend the theory of creeping cracks into the domain of fractal geometry, one needs to redefine entities of the critical stress intensity factor and the applied stress intensity stress factor K_0, denoted, respectively, by K_G and K_0 in the governing Eq. 42. The new definitions accounting for the fractal geometry of a crack are as follows:

$$\begin{aligned} K_G &\to K^f_{\text{crit}} = \chi_0(\alpha)\sigma_{\text{crit}}\sqrt{\pi a^{2\alpha}} \\ K_0 &\to K^f_0 = \chi_0(\alpha)\sigma_0\sqrt{\pi a_0{}^{2\alpha}} \end{aligned} \tag{62}$$

For glassy polymers, the critical stress σ_{crit} is very closely approximated by the Griffith stress σ_G. Thus, the ratio $(K_G/K_0)^2$ will reduce to an expression

$$\left(\frac{K^f_{\text{crit}}}{K^f_0}\right)^2 = \left(\frac{\sigma_G}{\sigma_0}\right)^2\left(\frac{a_0}{a}\right)^{2\alpha} = \frac{n}{\zeta^{2\alpha}} \tag{63}$$

When substituted into Eqs. 55 or 56, this leads to the following ordinary differential equation describing the slow motion of a fractal crack:

$$d\theta = \delta^{-1}\ln\left(\frac{\beta_1}{1+\beta_1-\dfrac{n}{\zeta^{2\alpha}}}\right)d\zeta \tag{64}$$

The solution of this equation is readily obtained as

$$t = t_1 + \left(\frac{\tau_2}{\delta}\right)\int_1^{\zeta}\ln\left[\frac{\beta_1}{1+\beta_1-\dfrac{n}{z^{2\alpha}}}\right]dz \tag{65}$$

In here t_1 denotes the incubation time, which is given by Eq. 49 – the same formula which was found valid for a smooth crack. However, the motion of a rough crack differs from that of a smooth crack in a rather distinct way; see Fig. 12. Fractal crack moves slower than the smooth crack and it arrives at the critical length significantly greater than the critical length attained by a smooth crack of same initial size and subjected to the same loading configuration and the same load.

To substantiate this statement, let us compare the delayed fracture of a rough crack defined by the fractal exponent $\alpha = 0.3$ (or the fractal dimension $D = 1.4$) to that of a smooth crack, for which $\alpha = 0.5$ and $D = 1$. For the input data

$$\beta_1 = 10, \quad n = 4, \quad \delta = 10^{-3} \tag{66}$$

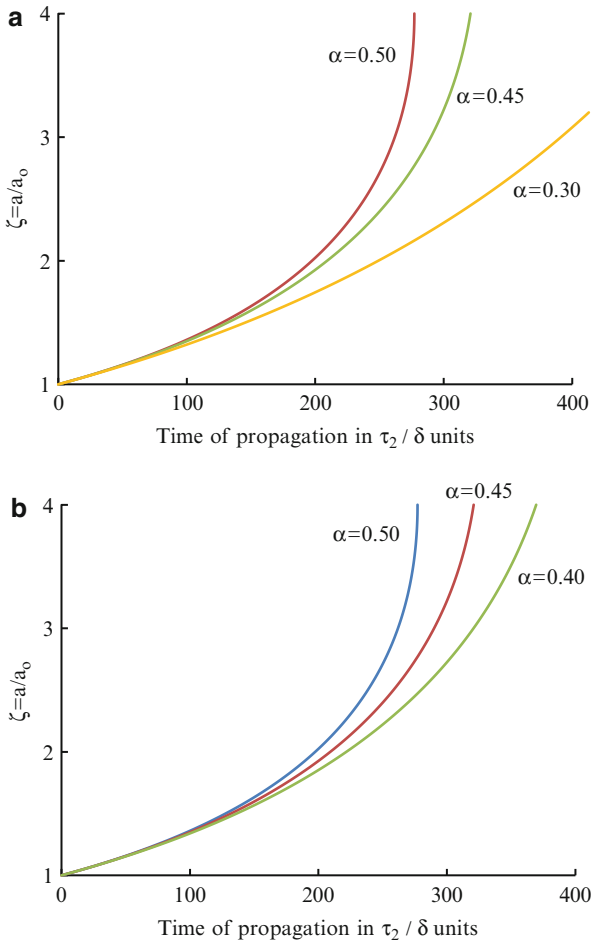


Fig. 12 The graphs in (**a**) and (**b**) show the effects of the roughness of the crack surfaces on the quasi-static propagation of crack contained in the viscoelastic medium. An increased degree of crack fractality is tantamount to an enhanced roughness of the crack surfaces and to a more pronounced phenomenon of time-dependent fracture, which precedes the catastrophic failure

and using Eq. 65 (see also Fig. 12a), one obtains the values of the terminal crack lengths and the times of propagation up to the point of catastrophic fracture for the fractal and smooth crack as follows:

$$a^f_{\text{crit}} = 10a_0 \text{ and } a^{\text{smooth}}_{\text{crit}} = 4a_0, \; t^f_2 = 659.34\tau_2 \text{ and } t^{\text{smooth}}_2 = 485.04\tau_2 \quad (67)$$

This clearly shows that roughness of the crack surfaces has a significant influence on the process of delayed fracture. For the data specified in Eq. 66, one observes 2.5-fold increase in the critical crack length attained at the terminal instability point (150 % enhancement over the smooth crack case) and 1.36-fold increase in the time of propagation, which is 36 % above the result valid for a smooth crack. These numbers illustrate the strong influence of fractal geometry on the outcome of the delayed fracture tests.

Some Fundamental Concepts

Delayed fracture occurring in a linearly viscoelastic solid such as the one discussed in section "Basic Concepts in Fractal Fracture Mechanics" consists of two distinct stages: (1) incubation phase, during which the opening displacement associated with the crack increases in time, but the crack remains stationary, and (2) propagation phase, when the crack advances up to the critical length (Griffith length), at which transition to unstable crack extension takes place. Stage I (incubation) is described by the Wnuk–Knauss equation in Eq. 41, and for the standard linear solid (see Fig. 9), the predicted duration of the incubation phase t_1 is given as

$$t_1 = \tau_2 \ln\left(\frac{\beta_1}{1+\beta_1 - n}\right) \tag{68}$$

The phase II (crack propagation) is governed by the Mueller–Knauss–Schapery equation (42). For the nondimensional creep compliance function $\Psi(t)$ defined by Eq. 47, the resulting equation of motion, which relates crack length ζ to time t, is given by Eq. 59, while the duration of the propagation phase is predicted as follows:

$$t_2 = \left(\frac{\tau_2}{\delta}\right)\left\{\frac{n}{1+\beta_1}\ln\left[\frac{\beta_1 n}{1+\beta_1 - n}\right] + \ln\left[\frac{1+\beta_1 - n}{\beta_1}\right]\right\} \tag{69}$$

The total lifetime T_{cr} of the component manufactured of a polymeric material that obeys the constitutive equations described in section "Basic Concepts in Fractal Fracture Mechanics" is obtained as the sum of Eqs. 68 and 69, namely,

$$T_{\mathrm{cr}} = t_1 + t_2 = \tau_2 \ln\left(\frac{\beta_1}{1+\beta_1 - n}\right) + \left(\frac{\tau_2}{\delta}\right)\left\{\frac{n}{1+\beta_1}\ln\left[\frac{\beta_1 n}{1+\beta_1 - n}\right] + \ln\left[\frac{1+\beta_1 - n}{\beta_1}\right]\right\} \tag{70}$$

For Solithane 50/50, a polymer which is used to model mechanical properties of the solid rocket fuel, the times t_1, t_2, and T_{cr} were evaluated by Knauss (1969) and Mohanty (1972). The moduli E_1 and E_2 and the viscosity η_2 involved in the standard linear solid that was applied in these studies are as follows:

$$\begin{aligned} E_1 &= 6.65^{*}10^{3}\ \ \mathrm{lb/in^2} \\ E_2 &= 3.69^{*}10^{3}\ \ \mathrm{lb/in^2} \\ \eta_2 &= 1.36^{*}10^{3}\ \ s\mathrm{lb/in^2} \end{aligned} \tag{71}$$

This leads to $\beta_1 = 1.8$, the relaxation time $\tau_2 = \eta_2/E_2 = 0.368$ s, and the maximum crack length quotient $n_{\max} = 1 + \beta_1 = 2.8$. The structural length Δ was estimated as 4.5×10^{-4} in., while the precut cracks used in the experiments were on the order of 0.225 in. This yielded the inner structural constant $\delta = 2 \times 10^{-3}$.

From Eq. 71 the "glassy" and the "rubbery" values of the creep compliance function can be readily calculated, namely,

$$\begin{aligned} J_{\text{glassy}} &= J(0) = 1.50^{*}10^{-4} \ \text{in}^2/\text{lb} \\ J_{\text{rubbery}} &= J(\infty) = 4.22^{*}10^{-4} \ \text{in}^2/\text{lb} \end{aligned} \tag{72}$$

For detailed calculations, the reader is referred to Knauss (1969) and Mohanty (1972).

The glassy (instantaneous) and rubbery (upon complete relaxation) compliance function values, as given in Eq. 72, allow one to establish the domains of the delayed fracture, such as "no-growth," incubation, or propagation domains. It should be noted that the creep compliance functions involved in these experimental investigations were obtained by use of the method.

In general, the propagation of a crack embedded in the viscoelastic medium will occur within a certain range of applied load. The two limiting values are (1) the Griffith stress evaluated for the initial crack size a_0, which is

$$\sigma_G = \begin{cases} \sqrt{\dfrac{2E\gamma}{\pi a_0}} \\ \dfrac{K_{IC}}{\sqrt{\pi a_0}} \end{cases} \tag{73}$$

and (2) the propagation threshold stress

$$\sigma_{\text{threshold}} = \sqrt{\frac{J(0)}{J(\infty)}}\sigma_G = \sqrt{\frac{J_{\text{glassy}}}{J_{\text{rubbery}}}}\sigma_G \tag{74}$$

For the standard linear solid expression, Eq. 74 reads

$$\sigma_{\text{threshold}} = \frac{1}{\sqrt{1+\beta_1}}\sigma_G \tag{75}$$

Using these relations one can predict the range of the applied loads for a successful delayed fracture test performed on Solithane 50/50 as being between 6/10 of the Griffith stress and the Griffith stress itself.

Summarizing, for the loads below the threshold stress given in Eqs. 74 and 75, one enters the "no-growth" domain, where propagation does not take place and the cracks in this region remain dormant. The other extreme is attained when the applied constant stress σ_0 reaches the Griffith level σ_G. When σ_0 approaches the Griffith stress, one observes an instantaneous fracture as in a brittle medium with no delay effects. Therefore, one may conclude that the delayed fracture occurs only in the range

$$\begin{aligned} \sigma_{\text{threshold}} &\le \sigma_0 \le \sigma_G \\ \frac{\sigma_G}{\sqrt{1+\beta_1}} &\le \sigma_0 \le \sigma_G \end{aligned} \tag{76}$$

The second expression in Eq. 76 pertains to the standard linear model.

Let us now consider a numerical example for a polymer characterized by the following properties $\beta_1 = 10$, $\tau_2 = 1$ s, and $\delta = 10^{-4}$. Pertinent calculations are performed for three levels of the applied load, measured either by the crack length quotient $n\ (=\sigma_G{}^2/\sigma_0{}^2)$ or by the load ratio $s = \sigma_0/\sigma_G$, namely, $n = 8.16$ ($s = 0.35$), $n = 6.25$ ($s = 0.40$), and $n = 4$ ($s = 0.50$). Applying Eqs. 68 and 69, one obtains the following incubation (t_1) and time-to-failure (t_2) values:

$$\begin{aligned} &n = 8.16; s = 0.35 \\ &t_1 = 1.26\ \ s; t_2 = \left(1/10^{-4}\right)\left(0.277\right)\ \ s = 46.2\ \ \text{min} \\ &n = 6.25; s = 0.40 \\ &t_1 = 0.744\ \ s;\ \ t_2 = \left(1/10^{-4}\right)\left(0.720\right)\ \ s = 120\ \ \text{min} \\ &n = 4; s = .50 \\ &t_1 = 0.375\ \ s; t_2 = \left(1/10^{-4}\right)\left(1.232\right)s = 205\ \ \text{min} \end{aligned} \tag{77}$$

It is noted that for this material, the range of the applied stress for the delayed fracture to occur is contained within the interval [0.3 σ_G, σ_G]. For applied stress less than the threshold stress of $0.3\sigma_G$, the phenomenon of delayed fracture vanishes, and the crack remains stationary.

An interesting study of the interaction between the damage zone, governed by the modified Kachanov's law, and the dominant crack has been described in Wnuk and Kriz (1985). There it has been shown that the existence of the damage zone adjacent to the crack leading edge accelerates motion of the subcritical crack.

Conclusions

From the considerations presented above, it follows that fracture is not only of fractal character, but it is also of multi-scale nature. To fill the gap between the "micro" and "macro" scale levels, the group of Panin (Pugno and Ruoff 2004) in Siberia, Russia, has developed an entirely new branch of mechanics, named "Mesomechanics." The primary objective of this and the following chapter is to construct a theory that explains and supports the findings based on the experimental observations of the Siberian group.

References

A.S. Balankin, Physics of fracture and mechanics of self-affine cracks. Eng. Fract. Mech. **57**(2), 135–203 (1997)

F.M. Borodich, Fracture energy in a fractal crack propagating in concrete or rock. Doklady Russian Acad. Sci. **325**, 1138–1141 (1992)

F.M. Borodich, Some fractal models of fracture. J. Mech. Phys. Solids **45**, 239–259 (1997)

F.M. Borodich, Fractals and fractal scaling in fracture mechanics. Int. J. Fract. **95**, 239–259 (1999)

A. Carpinteri, Scaling laws and renormalization groups for strength and toughness of disordered materials. Int. J. Solids Struct. **31**, 291–302 (1994)

A. Carpinteri, A. Spagnoli, A fractal analysis of the size effect on fatigue crack growth. Int. J. Fatigue **26**, 125–133 (2004)

A. Carpinteri, B. Chiaia, P. Cornetti, A scale invariant cohesive crack model for quasi-brittle materials. Eng. Fract. Mech. **69**, 207–217 (2002)
G.P. Cherepanov, A.S. Balankin, V.S. Ivanova, Fractal fracture mechanics – a review. Eng. Fract. Mech. **51**(6), 997–1033 (1995)
F.A. Field, A simple crack extension criterion for time-dependent spallation. J. Mech. Phys. Solids **19**, 61 (1971); also in AMR, vol. 25 (1972), Rev. 2781
R.V. Goldstein, A.B. Mosolov, Fractal cracks. J. Appl. Math. Mech. **56**, 563–571 (1992)
G.A.C. Graham, The correspondence principle of linear viscoelasticity theory for mixed boundary value problems involving time dependent boundary regions. Q. Appl. Math. **26**, 167 (1968); also in AMR, vol. 22, Rev. 4036
A.A. Griffith, The phenomenon of rupture and flow in solids. Phil. Trans. Roy. Soc. Lond. **A221**, 163–398 (1921a)
J. Harrison, Numerical integration of vector fields over curves with zero area. Proc. Am. Math. Soc. **121**, 715–723 (1994)
J. Harrison, A. Norton, Geometric integration on fractal curves in the plane, research report. Indiana Univ. Math. J. **40**, 567–594 (1991)
C.E. Inglis, Stresses in a plate due to the presence of cracks and sharp corners. Trans. R. Inst. Naval Architects **60**, 219 (1913)
M. Ippolito, A. Mattoni, L. Colombo, Role of lattice discreteness on brittle fracture: Atomistic simulations versus analytical models. Phys. Rev. B **73**, 104111 (2006). 6 pages
G.R. Irwin, *Handbuch der Physik*, vol. 6 (Springer, Berlin, 1956), pp. 551–590
H. Khezrzadeh, M.P. Wnuk, A. Yavari, Influence of material ductility and crack surface roughness on fracture instability. J. Phys. D Appl. Phys. **44**, 395302 (2011) (22 pages)
W.G. Knauss, Stable and unstable crack growth in viscoelastic media. Trans. Soc. Rheol. **13**, 291 (1969)
W.G. Knauss, Delayed failure. The Griffith problem for linearly viscoelastic materials. Int. J. Fract. **6**, 7 (1970); also in AMR, vol. 24, Rev. 5923
W.G. Knauss, The mechanics of polymer fracture. Appl. Mech. Rev. **26**, 1–17 (1973)
W.G. Knauss, H. Dietmann, Crack propagation under variable load histories in linearly viscoelastic solids. Int. J. Eng. Sci. **8**, 643 (1970); also in AMR, vol. 24, Rev. 1097
W.G. Knauss, The time dependent fracture of viscoelastic materials, in *Proceedings of the First International Conference on Fracture*, vol. 2, ed. by M.L. Williams. p. 1139; also see the Ph.D. Thesis, California Institute of Technology 1963 (1965)
B.V. Kostrov, L.V. Nikitin, Some general problems of mechanics of brittle fracture. Archiwum Mechaniki Stosowanej. (English version) **22**, 749; also in AMR, vol. 25 (1972), Rev. 1987 (1970)
B.B. Mandelbrot, D.E. Passoja, A.J. Paullay, Fractal character of fracture surfaces in metals. Nature **308**, 721–722 (1984)
D. Mohanty, *Experimental Study of Viscoelastic Properties and Fracture Characteristics in Polymers*, M.S. Thesis at Department of Mechanical Engineering, South Dakota State University, Brookings, 1972
A.B. Mosolov, Cracks with fractal surfaces. Doklady Akad. Nauk SSSR **319**, 840–844 (1991)
H.K. Mueller, Stress-intensity factor and crack opening for a linearly viscoelastic strip with a slowly propagating central crack. Int. J. Fract. **7**, 129 (1971)
H.K. Mueller, W.G. Knauss, Crack propagation in a linearly viscoelastic strip. J. Appl. Mech. **38** (Series E), 483 (1971a)
H.K. Mueller, W.G. Knauss, The fracture energy and some mechanical properties of a polyurethane elastomer. Trans. Soc. Rheol. **15**, 217 (1971b)
N.I. Muskhelishvili, *Some Basic Problems of the Mathematical Theory of Elasticity* (English translation) (Noordhoff, 1953)
H. Neuber, *Theory of Notch Stresses* (Springer, Berlin, 1958)
V.V. Novozhilov, On a necessary and sufficient criterion for brittle strength. J. Appl. Mech. USSR **33**, 212–222 (1969)

N. Pugno, R.S. Ruoff, Quantized fracture mechanics. Philos. Mag. **84**(27), 2829–2845 (2004)
J.R. Rice, Mathematical analysis in the mechanics of fracture, in *Fracture. An Advanced Treatise*, ed. by H. Liebowitz, vol. II (Academic, New York, 1968)
R.A. Schapery, A theory of crack growth in viscoelastic media. Int. J. Fract. **11**, 141–159 (1973)
C.F. Shih, Relationship between the J-integral and crack opening displacement for stationary and growing cracks. J. Mech. Phys. Solids **29**, 305–326 (1981)
A. Spagnoli, Self-similarity and fractals in the Paris range of fatigue crack growth. Mech. Mater. **37**, 519–529 (2005)
A.A. Wells, Application of fracture mechanics at and beyond general yielding. Br. J. Weld. **11**, 563–570 (1961)
H.M. Westergaard, Bearing pressure and cracks. J. Appl. Mech. **61**(1939), A49–A53 (1939)
M.L. Williams, On stress distribution at the base of a stationary crack. J. Appl. Mech. **24**, 109–114 (1957)
M.L. Williams, The continuum interpretation for fracture and adhesion. J. Appl. Polym.Sci. **13**, 29 (1969a)
M.L. Williams, The kinetic energy contribution to fracture propagation in a linearly viscoelastic material. Int. J. Fract. **4**, 69 (1969b); also in AMR, vol. 22 (1969), Rev. 8521
J.R. Willis, Crack propagation in viscoelastic media. J. Mech. Phys. Solids **15**, 229 (1967); also in AMR, vol.22 (1969), Rev. 8625
M.P. Wnuk, *Energy Criterion for Initiation and Spread of Fracture in Viscoelastic Solids* (Technical Report of the Engineer Experimental Station at SDSU, No.7, Brookings, 1968a)
M.P. Wnuk, Nature of fracture in relation to the total potential energy. Brit. J. Appl. Phys. **1** (Serious 2), 217 (1968b)
M.P. Wnuk, Effects of time and plasticity on fracture. British J. Appl. Phys., Ser. 2 **2**, 1245 (1969)
M.P. Wnuk, Prior-to-failure extension of flaws under monotonic and pulsating loadings, SDSU Technical Report No. 3, Engineering Experimental Station Bulletin at SDSU, Brookings (1971)
M.P. Wnuk, Accelerating crack in a viscoelastic solid subject to subcritical stress intensity, in *Proceedings of the International Conference on Dynamic Crack Propagation, Lehigh University*, ed. by G.C. Sih (Noordhoff, Leyden, 1972), pp. 273–280
M.P. Wnuk, Quasi-static extension of a tensile crack contained in a viscoelastic-plastic solid. J. Appl. Mech. **41**, 234–242 (1974)
M.P. Wnuk, R.D. Kriz, CDM model of damage accumulation in laminated composites. Int. J. Fract. **28**, 121–138 (1985)
M.P. Wnuk, B. Omidvar, Effects of strain hardening on quasi-static fracture in elasto-plastic solid represented by modified yield strip model. Int. J. Fract. **84**, 383–403 (1997)
M.P. Wnuk, A. Yavari, On estimating stress intensity factors and modulus of cohesion for fractal cracks. Eng. Fract. Mech **70**, 1659–1674 (2003)
M.P. Wnuk, A. Yavari, A correspondence principle for fractal and classical cracks. Eng. Fract. Mech. **72**, 2744–2757 (2005)
M.P. Wnuk, A. Yavari, Discrete fractal fracture mechanics. Eng. Fract. Mech. **75**, 1127–1142 (2008)
M.P. Wnuk, A. Yavari, A discrete cohesive model for fractal cracks. Eng. Fract. Mech. **76**, 548–559 (2009)
M.P. Wnuk, B. Omidvar, M. Choroszynski, Relationship between the CTOD and the J-integral for stationary and growing cracks. Closed form solutions. Int. J. Fract. **87**(1998), 331–343 (1998)
M.P. Wnuk, M. Alavi, A. Rouzbehani, Comparison of time dependent fracture in viscoelastic and ductile solids. Phys. Mesomech. **15**(1–2), 13–25 (2012)
M.P. Wnuk, M. Alavi, A. Rouzbehani, A mathematical model of Panin's pre-fracture zones and stability of subcritical cracks, in *Physical Mesomechanics* (Russian Academy of Sciences, Tomsk, 2013 in print)
S.N. Zhurkov, Kinetic concept of the strength of solids. Int. J. Fract. **1**, 311 (1965); also in Appl. Mech. Rev., vol. 20, 1967, Rev. 4080

Lattice and Particle Modeling of Damage Phenomena

7

Sohan Kale and Martin Ostoja-Starzewski

Contents

Abstract

Lattice (spring network) models offer a powerful way of simulating mechanics of materials as a coarse scale cousin to molecular dynamics and, hence, an alternative to finite element models. In general, lattice nodes are endowed with masses, thus resulting in a quasiparticle model. These models, having their origins in spatial trusses and frameworks, work best when the material may

S. Kale (✉) • M. Ostoja-Starzewski
Department of Mechanical Science & Engineering, also Institute for Condensed Matter Theory and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL, USA
e-mail: skale2@illinois.edu; martinos@illinois.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_20

naturally be represented by a system of discrete units interacting via springs or, more generally, rheological elements. This chapter begins with basic concepts and applications of spring networks, in particular the anti-plane elasticity, planar classical elasticity, and planar nonclassical elasticity. One can easily map a specific morphology of a composite material onto a particle lattice and conduct a range of parametric studies; these result in the so-called damage maps. Considered next is a generalization from statics to dynamics, with nodes truly acting as quasiparticles, application being the comminution of minerals. The chapter closes with a discussion of scaling and stochastic evolution in damage phenomena as stepping-stone to stochastic continuum damage mechanics.

Introduction

The need to simultaneously model elastic, plastic, and fracture responses in heterogeneous materials necessitates the introduction of techniques outside the realm of conventional continuum solid mechanics and finite element analysis. One technique that can meet the challenge, especially, when complicated microstructures need to be considered goes under the heading "lattice models." This chapter outlines the basics of lattice (or spring network) models and quasiparticle models in studies of damage phenomena. These models have their origin in spatial trusses and frameworks (from the engineering mechanics side) as well as in crystal structures (from the physics side). They offer a powerful way of simulating mechanics of materials with either periodic or random microstructures and, hence, offer an attractive alternative to finite element models. When dealing with a dynamic problem, nodes of the lattice may be endowed with masses, thus resulting in a quasiparticle model, which is then a coarse scale cousin of molecular dynamics. Lattice models work best when the material may naturally be represented by a system of discrete units interacting via springs or, more generally, rheological elements – examples are fibrous or granular systems. In the latter case, the lattice model turns into the discrete element model.

In this chapter, first, an introduction to basic concepts and applications of spring networks is given and, in particular, to the anti-plane elasticity, planar classical elasticity, and planar nonclassical elasticity. It is shown that one can easily map a specific morphology of a composite material onto a particle lattice and conduct a range of parametric studies; such studies result in the so-called damage maps. Here we discuss the elastic–plastic, elastic–brittle, and elastic–plastic–brittle materials. Considered next is a generalization from statics to dynamics (i.e., to a quasiparticle model), with nodes truly acting as quasiparticles, one application being the comminution of minerals. The chapter closes with a discussion of scaling and stochastic evolution in damage phenomena as stepping-stone to stochastic continuum damage mechanics.

Basic Idea of a Spring Network Representation

The basic idea in setting up lattice spring network (or lattice) models in d (=1, 2, or 3) dimensions is based on the equivalence of potential energies (U) stored in the unit cell of a given network. In the case of static problems, with which this chapter begins, for a cell of volume V (Fig. 1), there holds

$$U_{\text{cell}} = U_{\text{continuum}}. \tag{1}$$

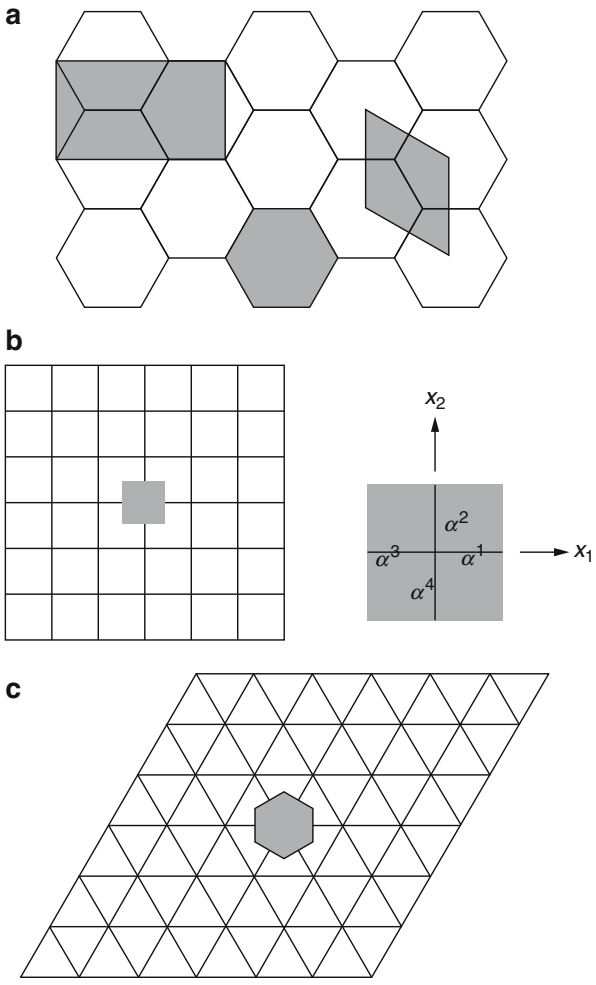


Fig. 1 Three periodic planar lattices: honeycomb, square, and triangular. In each case, a possible periodic unit cell is shown

The unit cell is a periodically repeating part of the network, and it is important to note that:

(i) The choice of the unit cell may be nonunique (see Fig. 1a).
(ii) The microstructure of the unit cell is not necessarily "nicely" ordered – it may be of a disordered microgeometry, with an understanding that it repeats itself in space such as the periodic Poisson–Delaunay network, e.g., Fig. 4.7 in Ostoja-Starzewski (2002a, 2008).

In Eq. 1 the energies of the cell (U_{cell}) and its equivalent continuum ($U_{\text{continuum}}$), respectively, are

$$U_{\text{cell}} = \sum_b E_b = \sum_b^{N_b} \int_0^{\mathbf{u}} \mathbf{F}(\mathbf{u}')d\mathbf{u}' \qquad U_{\text{continuum}} = \int_V \boldsymbol{\sigma}(\varepsilon)dV. \tag{2}$$

The superscript b in Eq. 2_1 stands for the bth spring (bond) and N_b for the total number of bonds. The discussion is set in the $d = 2$ setting so that the volume actually means the area of unit thickness. Here, and in the sequel, the discussion is restricted to linear elastic springs and spatially linear displacement fields $\mathbf{u}$ (i.e., uniform strain fields ε), implying that Eq. 2 becomes

$$U_{\text{cell}} = \frac{1}{2}\sum_b^{|b|} (k\,\mathbf{u}\cdot\mathbf{u})^{(b)} \qquad U_{\text{continuum}} = \frac{V}{2}\varepsilon : \mathbf{C} : \varepsilon. \tag{3}$$

In Eq. 3 $\mathbf{u}$ is a generalized spring displacement and k its corresponding spring constant. The next step, depending on the particular geometry of the unit element and on the particular model of interactions, will involve making a connection between $\mathbf{u}$ and ε and then deriving $\mathbf{C}$ from Eq. 1. The corresponding procedures and resulting formulas are given below for several elasticity problems set in the square and triangular network geometries.

Anti-plane Elasticity on Square Lattice

Of all the elasticity problems, the anti-plane is the simplest one on which to illustrate the spring network idea. In the continuum setting, the constitutive law is

$$\sigma_i = C_{ij}\varepsilon_j \quad i,j = 1,2, \tag{4}$$

where $\boldsymbol{\sigma} = (\sigma_1, \sigma_2) \equiv (\sigma_{31}, \sigma_{32})$, $\varepsilon = (\varepsilon_1, \varepsilon_2) \equiv (\varepsilon_{31}, \varepsilon_{32})$ and $C_{ij} \equiv C_{3i3j}$. Upon the substitution of Eq. 4 into the equilibrium equation

$$\sigma_{i,i} = 0, \tag{5}$$

these result to

$$\left(C_{ij}u_{,j}\right)_{,i} = 0. \tag{6}$$

Henceforth, with interest focused on approximations of locally homogeneous media, the governing equation Eq. 6 becomes

$$C_{ij}u,_{ij} = 0. \tag{7}$$

In the special case of an isotropic medium, Eq. 7 simplifies to the Laplace equation

$$Cu,_{ii} = 0. \tag{8}$$

Next, discretize the material with a square lattice network, Fig. 1b, whereby each node has one degree of freedom (anti-plane displacement u), and the nearest-neighbor nodes are connected by springs of constant k. It follows that the strain energy of the unit cell of such a lattice is

$$U = \frac{1}{2}k\sum_{b=1}^{4} l_i^{(b)} l_j^{(b)} \varepsilon_i \varepsilon_j. \tag{9}$$

Here the uniform strain $\varepsilon = (\varepsilon_1, \varepsilon_2)$ is employed, while $\mathbf{l}^{(b)} = (l_1^{(b)}, l_2^{(b)})$ is the vector of half-length of bond b. In view of Eq. 1, the stiffness tensor is obtained as

$$C_{ij} = \frac{k}{V}\sum_{b=1}^{4} l_i^{(b)} l_j^{(b)} \quad i,j = 1,2, \tag{10}$$

where $V = 4$ if all the bonds are of unit length ($|\mathbf{l}^{(b)}| = 1$). This leads to a relation between the bond spring constant k and the C_{ij} tensor

$$C_{11} = C_{22} = \frac{k}{2} \qquad C_{12} = C_{21} = 0. \tag{11}$$

In order to model an orthotropic medium, different bonds are applied in the x_1 and x_2 directions: $k^{(1)}$ and $k^{(2)}$. The strain energy of the unit cell is now

$$U = \frac{1}{2}\sum_{b=1}^{4} k^{(b)} l_i^{(b)} l_j^{(b)} \varepsilon_i \varepsilon_j, \tag{12}$$

so that the stiffness tensor is

$$C_{ij} = \frac{1}{V}\sum_{b=1}^{4} k^{(b)} l_i^{(b)} l_j^{(b)}, \tag{13}$$

which leads to relations

$$C_{11} = \frac{k^{(1)}}{2} \quad C_{22} = \frac{k^{(2)}}{2} \quad C_{12} = C_{21} = 0. \tag{14}$$

If one wants to model an anisotropic medium (i.e., with $C_{12} \neq 0$), one may either choose to rotate its principal axes to coincide with those of the square lattice and use

the network model just described or introduce diagonal bonds. In the latter case, the unit cell energy is given by the formula (Eq. 12) with $N_b = 8$. The expressions for C_{ij}s are

$$C_{11} = \frac{k^{(1)}}{2} + k^{(5)} \quad C_{22} = \frac{k^{(2)}}{2} + k^{(6)} \quad C_{12} = C_{21} = k^{(5)} - k^{(6)}. \tag{15}$$

It will become clear in the next section how this model can be modified to a triangular spring network geometry.

In-Plane Elasticity: Triangular Lattice with Central Interactions

In the planar continuum setting, assuming linear elastic behavior, Hooke's law

$$\sigma_{ij} = C_{ijkm}\varepsilon_{km} \quad i,j,k,m = 1,2, \tag{16}$$

upon substitution into the balance law

$$\sigma_{ij,j} = 0, \tag{17}$$

results in a planar Navier equation for the displacement u_i:

$$\mu u_{i,jj} + \kappa u_{j,ji} = 0. \tag{18}$$

In Eq. 18 μ is defined by $\sigma_{12} = \mu\varepsilon_{12}$, which makes it the same as the classical 3D shear modulus. On the other hand, κ is the (planar) 2D bulk modulus that is defined by $\sigma_{ii} = \kappa\varepsilon_{ii}$.

As in the foregoing section, approximations of locally homogeneous media are of interest. Consider the regular triangular network of Fig. 1c with central-force interactions, which are described, for each bond b, by

$$F_i = \Phi_{ij}^{(b)} u_j \quad \text{where} \quad \Phi_{ij}^{(b)} u_j = \alpha^{(b)} n_i^{(b)} n_j^{(b)}. \tag{19}$$

Similar to the case of anti-plane elasticity, $\alpha^{(b)}$ is the spring constant of half-lengths of such central (normal) interactions – i.e., of those parts of the springs that fall within the given unit cell (Fig. 2a). The unit vectors $\mathbf{n}^{(b)}$ at respective angles of the first three α-springs are

$$\begin{array}{lll} \theta^{(1)} = 0^0 & n_1^{(1)} = 1 & n_2^{(1)} = 0 \\ \theta^{(2)} = 60^0 & n_1^{(2)} = \frac{1}{2} & n_2^{(2)} = \frac{\sqrt{3}}{2} \\ \theta^{(3)} = 120^0 & n_1^{(3)} = -\frac{1}{2} & n_2^{(3)} = \frac{\sqrt{3}}{2}. \end{array} \tag{20}$$

The other three springs ($b = 4, 5, 6$) must, by the requirement of symmetry with respect to the center of the unit cell, have the same properties as $b = 1, 2, 3,$

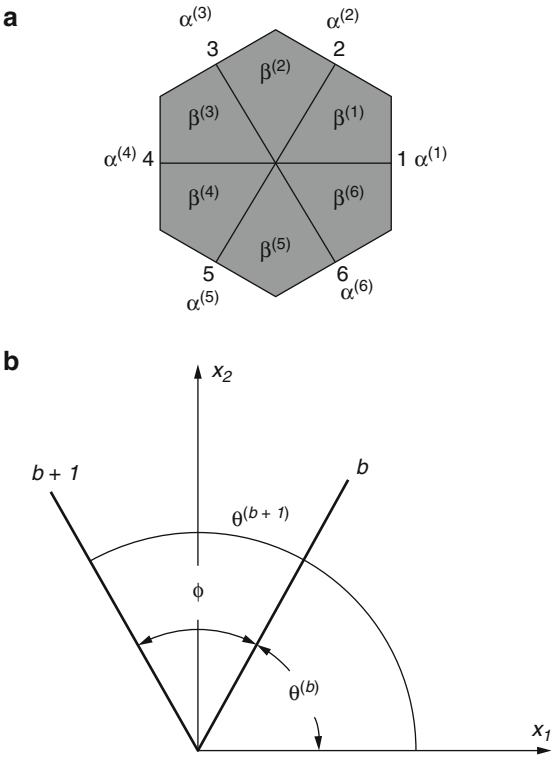


Fig. 2 Unit cell of a triangular lattice model; $\alpha^{(1)}, \ldots, \alpha^{(6)}$ are the normal spring constants; $\beta^{(1)}, \ldots, \beta^{(6)}$ are the angular spring constants; in the isotropic Kirkwood model $\alpha^{(b)} = \alpha^{(b+3)}$ and $\beta^{(b)} = \beta^{(b+3)}$, $b = 1, 2, 3$; (**b**) details of the angular spring model

respectively. All the α-springs are of length l, that is, the spacing of the triangular mesh is $s = 2l$. The cell area is $V = 2\sqrt{3}l^2$.

Every node has two degrees of freedom, and it follows that the strain energy of a unit hexagonal cell of this lattice, under conditions of uniform strain $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{12})$, is

$$U = \frac{l^2}{2} \sum_{b=1}^{6} \alpha^{(b)} n_i^{(b)} n_j^{(b)} n_k^{(b)} n_m^{(b)} \varepsilon_{ij} \varepsilon_{km}, \tag{21}$$

so that, again by Eq. 1, the stiffness tensor becomes

$$C_{ijkm} = \frac{l^2}{V} \sum_{b=1}^{6} \alpha^{(b)} n_i^{(b)} n_j^{(b)} n_k^{(b)} n_m^{(b)}. \tag{22}$$

In particular, taking all $\alpha^{(b)}$ the same, one finds

$$C_{1111} = C_{2222} = \frac{9}{8\sqrt{3}}\alpha \; C_{1122} = C_{2211} = \frac{3}{8\sqrt{3}}\alpha \; C_{1212} = \frac{3}{8\sqrt{3}}\alpha, \tag{23}$$

so that there is only one independent elastic modulus, and the modeled continuum is isotropic.

It is important to note here that the isotropy follows from the triangular lattice having an axis of symmetry of the sixth order. This, combined with the fact that Eq. 22 satisfies the conditions of *Cauchy symmetry* (Love 1934) with respect to the permutations of all the four indices (which is the last equality)

$$C_{ijkm} = C_{ijmk} = C_{jikm} = C_{kmij} = C_{ikjm}, \tag{24}$$

implies that C_{ijkm} is of the form

$$C_{ijkm} = \lambda\left(\delta_{ij}\delta_{km} + \delta_{ik}\delta_{jm} + \delta_{im}\delta_{jk}\right). \tag{25}$$

In view of Eq. 23, there result the classical Lamé constants

$$\lambda = \mu = \frac{3}{4\sqrt{3}}\alpha. \tag{26}$$

The above is a paradigm from the crystal lattice theory that the Cauchy symmetry occurs when:

(i) The interaction forces between the atoms (or molecules) of a crystal are of a central-force type.
(ii) Each atom (or molecule) is a center of symmetry.
(iii) The interaction potential in a crystal can be approximated by a harmonic one.

Note: The Cauchy symmetry reduces the number of independent constants in general 3D anisotropy from 21 to 15. The first case has been called the *multi-constant theory*, while the second one the *rari-constant theory* (Trovalusci et al. 2009; Capecchi et al. 2010). Basically, there is a decomposition of the stiffness tensor into two irreducible parts with 15 and 6 independent components, respectively; see Hehl and Itin (2002) for a group-theoretical study of these issues.

Note: One might try to model anisotropy by considering three different αs in Eqs. 21 and 22, but such an approach would be limited given the fact that only three of those can be varied: one needs to have six parameters in order to freely adjust any planar anisotropy which involves six independent C_{ijkm}s. This can be achieved by introducing the additional angular springs as discussed below. In fact, angular springs are also the device to vary the Poisson ratio.

In-Plane Elasticity: Triangular Lattice with Central and Angular Interactions

The triangular network is now enriched by the addition of angular springs acting between the contiguous bonds incident onto the same node. These are assigned

spring constants $\beta^{(b)}$, and, again by the argument of symmetry with respect to the center of the unit cell, only three of those can be independent. This leads to six spring constants: $\{\alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}, \beta^{(1)}, \beta^{(2)}, \beta^{(3)}\}$. With reference to Fig. 2b, let $\Delta\theta^{(b)}$ be the (infinitesimal) angle change of the bth spring orientation from the undeformed position. On account of $\mathbf{n} \times \mathbf{n} = \mathbf{l}\Delta\theta$,

$$\Delta\theta^{(b)} = e_{kij}\varepsilon_{jp}n_i n_p, \tag{27}$$

where e_{kij} is the Levi-Civita permutation tensor. The angle change between two contiguous α-springs (b and $b+1$) is measured by $\Delta\phi = \Delta\theta^{(b+1)} - \Delta\theta^{(b)}$, so that the energy stored in the spring $\beta^{(b)}$ is

$$E^{(b)} = \frac{1}{2}\beta^{(b)}|\Delta\phi|^2 = \frac{1}{2}\beta^{(b)}\left\{\varepsilon_{kij}\varepsilon_{jp}\left(n_i^{(b+1)}n_p^{(b+1)} - n_i^{(b)}n_p^{(b)}\right)\right\}^2. \tag{28}$$

By superposing the energies of all the angular bonds with the energy (Eq. 21), the elastic moduli are derived after Kirkwood (1939) as

$$\begin{aligned} C_{ijkm} = \frac{l^2}{V}\sum\nolimits_{b=1}^{6}\alpha^{(b)}n_i^{(b)}n_j^{(b)}n_k^{(b)}n_m^{(b)} + \frac{1}{V}\sum\nolimits_{b=1}^{6}\Big\{&\left[\beta^{(b)}+\beta^{(b-1)}\right]\delta_{ik}n_p^{(b)}n_j^{(b)}n_p^{(b)}n_m^{(b)} \\ &- \left[\beta^{(b)}+\beta^{(b-1)}\right]n_i^{(b)}n_j^{(b)}n_k^{(b)}n_m^{(b)} - \beta^{(b)}\delta_{ik}n_p^{(b)}n_j^{(b+1)}n_p^{(b+1)}n_m^{(b)} \\ &+ \beta^{(b)}n_i^{(b)}n_j^{(b+1)}n_k^{(b+1)}n_m^{(b)} - \beta^{(b)}\delta_{ik}n_p^{(b)}n_j^{(b)}n_p^{(b+1)}n_m^{(b+1)} \\ &+ \beta^{(b)}n_i^{(b+1)}n_j^{(b)}n_k^{(b)}n_m^{(b+1)}\Big\}, \end{aligned} \tag{29}$$

where $b = 0$ is the same as $b = 6$.

This provides the basis for a spring network representation of an anisotropic material; it also forms a generalization of the so-called Kirkwood model (Keating 1966) of an isotropic material. The latter is obtained by assigning the same α to all the normal and the same β to all the angular springs

$$\begin{aligned} C_{ijkm} = \frac{\alpha}{2\sqrt{3}}\sum_{b=1}^{6}n_i^{(b)}n_j^{(b)}n_k^{(b)}n_m^{(b)} + \frac{\beta}{2\sqrt{3}l^2}\sum_{b=1}^{6}\Big\{&2\delta_{ik}n_j^{(b)}n_m^{(b)} - 2n_i^{(b)}n_j^{(b)}n_k^{(b)}n_m^{(b)} \\ &- \delta_{ik}n_p^{(b)}n_j^{(b+1)}n_p^{(b+1)}n_m^{(b)} + n_i^{(b)}n_j^{(b+1)}n_k^{(b+1)}n_m^{(b)} \\ &- \delta_{ik}n_p^{(b)}n_j^{(b)}n_p^{(b+1)}n_m^{(b+1)} + n_i^{(b+1)}n_j^{(b)}n_k^{(b)}n_m^{(b+1)}\Big\} \end{aligned} \tag{30}$$

In accordance with the above,

$$\begin{aligned} C_{1111} &= C_{2222} = \frac{1}{2\sqrt{3}}\left(\frac{9}{4}\alpha + \frac{1}{l^2}\beta\right) \\ C_{1122} &= C_{2211} = \frac{1}{2\sqrt{3}}\left(\frac{3}{4}\alpha + \frac{1}{l^2}\frac{9}{4}\beta\right) \\ C_{1212} &= \frac{1}{2\sqrt{3}}\left(\frac{3}{4}\alpha + \frac{1}{l^2}\frac{9}{4}\beta\right). \end{aligned} \tag{31}$$

Condition $C_{1212} = (C_{1111} - C_{1122})/2$ is satisfied, so that there are only two independent elastic moduli.

From Eq. 31, the α and β constants are related to the planar bulk and shear moduli by

$$\kappa = \frac{1}{2\sqrt{3}}\left(\frac{3}{2}\alpha\right) \qquad \mu = \frac{1}{2\sqrt{3}}\left(\frac{3}{4}\alpha + \frac{1}{l^2}\frac{9}{4}\beta\right). \tag{32}$$

It is noted here that the angular springs have no effect on κ, i.e., the presence of angular springs does not affect the dilatational response. The formula for planar Poisson's ratio (Ostoja-Starzewski 2008) gives

$$\nu = \frac{\kappa - \mu}{\kappa + \mu} = \frac{C_{1111} - 2C_{1212}}{C_{1111}} = \frac{1 - 3\beta/l^2\alpha}{3 + 3\beta/l^2\alpha}. \tag{33}$$

From Eq. 33, there follows the full range of Poisson's ratio which can be covered with this model. It has two limiting cases:

$$\begin{aligned} \nu &= \frac{1}{3} \quad \text{if} \quad \frac{\beta}{\alpha} \to 0, \quad \alpha - \text{model} \\ \nu &= -1 \quad \text{if} \quad \frac{\beta}{\alpha} \to \infty, \quad \beta - \text{model}. \end{aligned} \tag{34}$$

For the subrange of Poisson's ratio between $-1/3$ and $1/3$, one may also use a Keating model (Keating 1966) which employs a different calculation of the energy stored in angular bonds.

Triple Honeycomb Lattice

Since 1/3 is the highest Poisson's ratio of central-force triangular lattices with one spring constant, an interesting model permitting higher values, from 1/3 up to 1, was introduced (Garboczi et al. 1991; Buxton et al. 2001). The model sets up three honeycomb lattices, having spring constants α, β, and γ, respectively, overlapping in such a way that they form a single triangular lattice (Fig. 3). The planar bulk and shear moduli of a single phase are

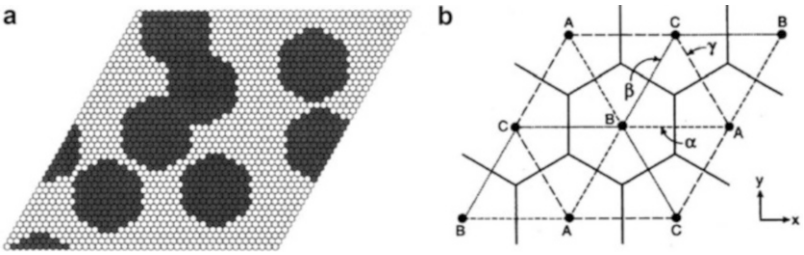


Fig. 3 (**a**) A triple honeycomb lattice made of three different spring types α, β, and γ belonging, respectively, to three sublattices A, B, and C; (**b**) a 42 × 42 unit cell of a triangular lattice of hexagonal pixels, with 11 pixel diameter circular inclusions centered on pixels and randomly placed with periodic boundary conditions; from (Snyder, K.A., Garboczi E.J. & Day, A.R. (1992)

$$\kappa = \frac{1}{\sqrt{12}}(\alpha + \beta + \gamma) \quad \mu = \sqrt{\frac{27}{16}}\left(\frac{1}{\alpha} + \frac{1}{\beta} + \frac{1}{\gamma}\right)^{-1}. \tag{35}$$

In the case of two (or more) phases, a spring that crosses a boundary between any two phases (**1** and **2**) is assigned a spring constant according to a series rule $\alpha = [(2\alpha_1)^{-1} + (2\alpha_2)^{-1}]$, where α_i, $i = 1, 2$, (i. e. α, β, or γ), is a spring constant of the respective phase.

Note: While this chapter is focused on planar lattice models of elastic solids, there also exist extensions of the lattice approach to 3D and inelastic materials (e.g., Buxton et al. 2001).

Spring Network Models

Representation by a Fine Mesh

With reference to section "Basic Idea of a Spring Network Representation," one may employ the square mesh of Fig. 2a in the (x_1, x_2)-plane for discretization of an anti-plane elasticity problem. Indeed, this approach may be applied to model multiphase composites treated as planar, piecewise-constant continua, providing a lattice or mesh (very) much finer than a single inclusion is involved (Fig. 4b). How much finer should actually be assessed on a reference problem according to a preset error criterion? The governing equations for the displacement field $u \equiv u_3$ are

$$\begin{aligned} &u(i,j)\,[k_r + k_l + k_u + k_d] - u(i+1,j)\,k_r - u\big(i-1,j\big)k_l - u\big(i,j+1\big)k_u \\ &\quad - u\big(i,j-1\big)k_d = f\big(i,j\big), \end{aligned} \tag{36}$$

where $f(i, j)$ is the body force (or source) at node (i, j), while i and j are the coordinates of mesh points, and k_r (right), k_l (left), k_u (up), and k_d (down) are defined from the series spring model

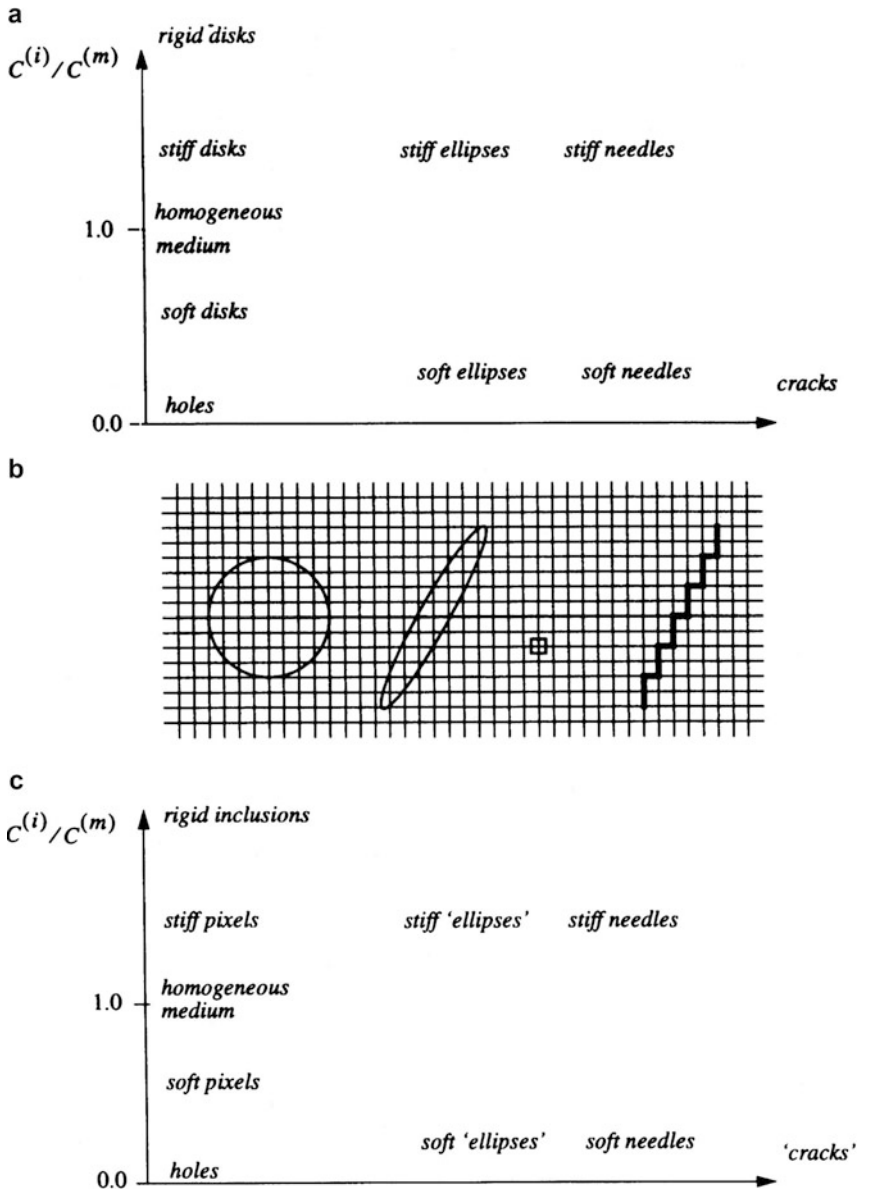


Fig. 4 (**a**) Parameter plane: aspect ratio of inclusions and the contrast; (**b**) spring network as a basis for resolution of round disks, ellipses, pixels, and needles in the parameter plane; (**c**) another interpretation of the parameter plane: from pixels to needles

$$
\begin{aligned}
k_r &= \left[\frac{1}{C(i,j)} + \frac{1}{C(i+1,j)}\right]^{-1} \\
k_l &= \left[\frac{1}{C(i,j)} + \frac{1}{C(i-1,j)}\right]^{-1} \\
k_u &= \left[\frac{1}{C(i,j)} + \frac{1}{C(i,j+1)}\right]^{-1} \\
k_d &= \left[\frac{1}{C(i,j)} + \frac{1}{C(i,j-1)}\right]^{-1}.
\end{aligned}
\tag{37}
$$

In Eq. 37, $C(i, j)$ is the material property at node (i, j).

This type of a discretization is equivalent to a finite difference method that would be derived by considering the expansions

$$
\begin{aligned}
u(i \pm 1, j) &= u(i,j) \pm s \left.\frac{\partial u(i,j)}{\partial x_1}\right|_{i,j} + \frac{s^2}{2!} \left.\frac{\partial^2 u(i,j)}{\partial x_1^2}\right|_{i,j} \\
u(i, j \pm 1) &= u(i,j) \pm s \left.\frac{\partial u(i,j)}{\partial x_2}\right|_{i,j} + \frac{s^2}{2!} \left.\frac{\partial^2 u(i,j)}{\partial x_2^2}\right|_{i,j}
\end{aligned}
\tag{38}
$$

in the governing equation (recall Eq. 8)

$$
C\left(\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2}\right) = 0. \tag{39}
$$

However, in the case of in-plane elasticity problems, the spring network approach is not identical to the finite difference method, because the node–node connections of a spring network do really have a meaning of springs, whereas the finite difference connections do not.

In the case of a composite made of two locally isotropic phases, matrix (m) and inclusions (i), anti-plane Hooke's law is

$$
\sigma_i = C_{i,j}\varepsilon_j \quad i,j = 1,2 \quad C_{ij} = C^{(m)}\delta_{ij} \quad \text{or} \quad C^{(i)}\delta_{ij}. \tag{40}
$$

The above leads to a so-called contrast (or mismatch) $C^{(i)}/C^{(m)}$. It is clear that, with very high contrast, materials with rigid inclusions can approximately be modeled. Similarly, by decreasing the contrast, systems with very soft inclusions (nearly holes) can be simulated.

While the disk is the most basic inclusion shape when dealing with composites, a departure from this is of interest. Thus, another basic parameter specifying the composite is the *aspect ratio* of ellipses a/b, where a (b) is the ellipse's major (minor) semiaxis. By varying the aspect ratio from 1 up through higher values, systems having disk-type, ellipse-type, through needle-type inclusions are simulated. This leads to the concept of a parameter plane shown in Fig. 4a.

Resolution of several different types of inclusions by the spring network is shown in Fig. 4b. Admittedly, this type of modeling is approximate so that a somewhat different interpretation of a parameter plane is given in Fig. 4c. It is seen that disks may most simply be modeled as single pixels or more accurately as finite regions; in the latter case, arbitrary anisotropies can be modeled. The former

case allows one to deal with very large-scale systems, while the latter allows a much better resolution of local stress–strain fields within and around inclusions. By decreasing the spring network mesh size, an increasingly better accuracy can be achieved. Depending on the shape functions employed in finite element models, somewhat more accurate results may be obtained, but this comes at a higher price of costly and cumbersome remeshing for each and every new configuration $B(\omega)$ from the ensemble $\mathcal{B}$, which is required in statistical (Monte Carlo) studies.

It is noteworthy that, in contradistinction to the finite element method, no need for remeshing and constructing of a stiffness matrix exists in the spring network method: spring constants are very easily assigned throughout the mesh, and the conjugate gradient method finds the solution of the equilibrium displacement field $u(i, j)$. In that manner, a system having 10^6 million degrees of freedom (1,000 $\times$ 1,000 nodes) can readily be handled on a computer with 90 MB of random access memory. For 2,000 $\times$ 2,000 nodes, one requires some 360 MB, and so on, because of the linear scaling of memory requirements with the number of degrees of freedom.

The quality of approximation of ellipses and needle-type cracks/inclusions can be varied according to the number of nodes chosen to represent such objects. Local fields cannot be perfectly resolved, but the solution by the spring network is sufficient to rapidly establish the elastic moduli of a number of different B (ω) realizations from the random medium $\mathcal{B}$, and the corresponding statistics with a sufficient accuracy. As indicated below, spring networks are used to study scaling laws of various planar composites.

Note: Interestingly, the computational method for determining effective moduli of composite materials with circular inclusions due to Bird and Steele (1992) would be very well suited for analysis of this type of stationarity and isotropy.

Damage in Macro-Homogeneous Materials

Spring Network for Inelastic Materials

The spring network model can also be used in studies assessing the effect of small disorder on the formation and evolution of damage in elastic–inelastic macro-homogeneous materials under quasi-static assumption. Such damage models are characterized by local constitutive law influenced by an appropriate probability distribution to account for the spatial material disorder. Representation of micro-cracks by removal of the springs from the lattice and accounting for elastic interactions of the micro-cracks are the key advantages. Disorder-induced statistical effects such as crack surface roughness, acoustic emission avalanches, damage localization, and strength-size scaling are well addressed using such damage models (Alava et al. 2006; see also Krajcinovic 1996; Rinaldi et al. 2008).

The random fuse model (RFM) (De Arcangelis et al. (1985)) is the simplest form of a quasi-static lattice damage model. In RFM, monotonically increasing voltage is applied across a network of resistors with randomly assigned maximum current thresholds, exceeding which the resistor burns. The failure of fuse network modeled

using RFM can be mapped onto an anti-plane elastic–brittle transition problem on a spring lattice with randomly assigned spring failure thresholds.

In order to admit plasticity followed by brittle failure, the equations of spring network model need to be modified from what has been presented in section "Introduction." This is done as follows:

$$\begin{array}{rcc} \text{elastic}: & F = ku & u < u_Y \\ \text{plastic}: & F = k^P(u - u_Y) + ku_Y & u_Y \le u < u_F \\ \text{brittle}: & F = 0 & u \ge u_F \\ \text{elastic unloading}: & F = ku - \left(k - k^P\right)\left(u_{\text{unload}} + u_Y\right), & \end{array} \tag{41}$$

where u_Y, u_{unload}, and u_F are the magnitudes of the change in length of a given spring at yield, at unloading, and at failure, respectively, whereas u is the magnitude of change in length of a given spring at the current loading step. The spring stiffnesses before and after yielding are denoted as k and k^P, respectively, and correspond to the elastic and plastic tangent moduli of the given material. The disorder is introduced in the model by constraining the yield and failure thresholds of the springs to follow desired probability distribution.

The simulation progresses by incrementing the boundary conditions in very small steps. It is assumed that the stress redistribution within the system followed by the yielding or failure event is much faster than the load incrementing rate. After a spring is yielded or failed, stiffness matrix of the system is modified and the system of equations is solved again to account for local stress redistribution. This process is repeated until the lattice falls apart in the case of an elastic–brittle transition or until the fully plastic state is reached for elastic–plastic transition.

Hill–Mandel Macrohomogeneity Condition

Any disordered body $B_\delta(\omega)$ with a given (deterministic) microstructure is loaded by either one of two different types of boundary conditions:

Uniform displacement (also called kinematic, essential, or Dirichlet) boundary condition (d)

$$\mathbf{u}(\mathbf{x}) = \varepsilon^0 \cdot \mathbf{x} \quad \forall \mathbf{x} \in \partial B_\delta \tag{42}$$

Uniform traction (also called static, natural, or Neumann) boundary condition (t)

$$\mathbf{t}(\mathbf{x}) = \boldsymbol{\sigma}^0 \cdot \mathbf{n} \quad \forall \mathbf{x} \in \partial B_\delta \tag{43}$$

Here ε^0 and $\boldsymbol{\sigma}^0$ are employed to denote constant tensors, prescribed a priori, whereby the average strain and stress theorems imply $\varepsilon^0 = \overline{\varepsilon}$ and $\boldsymbol{\sigma}^0 = \overline{\boldsymbol{\sigma}}$. Each of these loadings is consistent with the Hill–Mandel macrohomogeneity condition

$$\overline{\boldsymbol{\sigma} : \varepsilon} = \overline{\boldsymbol{\sigma}} : \overline{\varepsilon} \Leftrightarrow \int_{\partial B_\delta} (\mathbf{t} - \overline{\boldsymbol{\sigma}} \cdot \mathbf{n}) \cdot (\mathbf{u} - \overline{\varepsilon} \cdot \mathbf{x}) dS = 0, \tag{44}$$

which means that the volume-averaged scalar product of stress and strain fields should equal the product of their volume averages (Hill 1963; Mandel 1963; Huet 1982, 1990; Sab 1991, 1992). While the Hill–Mandel condition is written above for elastic materials, it also holds for plastic material behavior in the incremental setting.

Each of these boundary conditions results in a different mesoscale (or apparent) stiffness, or compliance tensor, and generally different from the macroscale (or *effective*, *global*, *overall*, etc.) properties that are typically denoted by ^{eff}. So as to distinguish from the effective, Huet introduced the term *apparent*.

For a given realization $B_\delta(\omega)$ of the random medium $\mathcal{B}_\delta$, taken as a linear elastic body ($\boldsymbol{\sigma} = \mathbf{C}(\omega, \mathbf{x}) : \varepsilon$), on some mesoscale δ, condition Eq. 42 yields an apparent random stiffness tensor $\mathbf{C}^d_\delta(\omega)$ – sometimes denoted $\mathbf{C}^e_\delta(\omega)$ – with the constitutive law

$$\overline{\boldsymbol{\sigma}} = \mathbf{C}^d_\delta(\omega) : \varepsilon^0, \tag{45}$$

while the boundary condition (Eq. 43) results in an apparent random compliance tensor $\mathbf{S}^t_\delta(\omega)$ (sometimes denoted $\mathbf{S}^n_\delta(\omega)$) with the constitutive law being stated as

$$\overline{\varepsilon} = \mathbf{S}^t_\delta(\omega) : \boldsymbol{\sigma}^0. \tag{46}$$

For anti-plane loading, the boundary conditions (Eqs. 42 and 43) are implemented through

$$\text{uniform displacement}: \quad \varepsilon^0_{31} = \varepsilon, \quad \varepsilon^0_{32} = 0 \tag{47}$$

$$\text{uniform traction}: \quad \sigma^0_{31} = \sigma, \quad \sigma^0_{32} = 0. \tag{48}$$

Modeling Elastic–Brittle Materials

The elastic–brittle transition is modeled using a constitutive law with linear elastic behavior up to a failure threshold.

Simulation Setup. A spring lattice network based on discussion in section "Basic Idea of a Spring Network Representation" is considered to represent a homogeneous anti-plane elastic medium. The bond strength (t) is defined as the maximum strain the spring can sustain (for RFM, t is equivalent to the maximum current a fuse can take before burning out). Bond strength is assigned to all the bonds in the lattice from a distribution $p(t)$ between $[0, t_{\max}]$. Monotonically increasing displacement boundary conditions are applied on the vertical edges of the lattice, while periodic boundary conditions are applied on the horizontal edges to avoid any boundary effects. At each loading step, a spring having the maximum value of $\epsilon_{\text{spring}}/t$ is removed from the lattice and the modified system of linear equations is solved again to allow stress redistribution. As stress redistribution may cause other springs in the lattice to fail, the removal process at a given loading step is continued until all possible springs are failed or the macroscopic failure with sudden drop in load-carrying capacity is observed (Fig. 5).

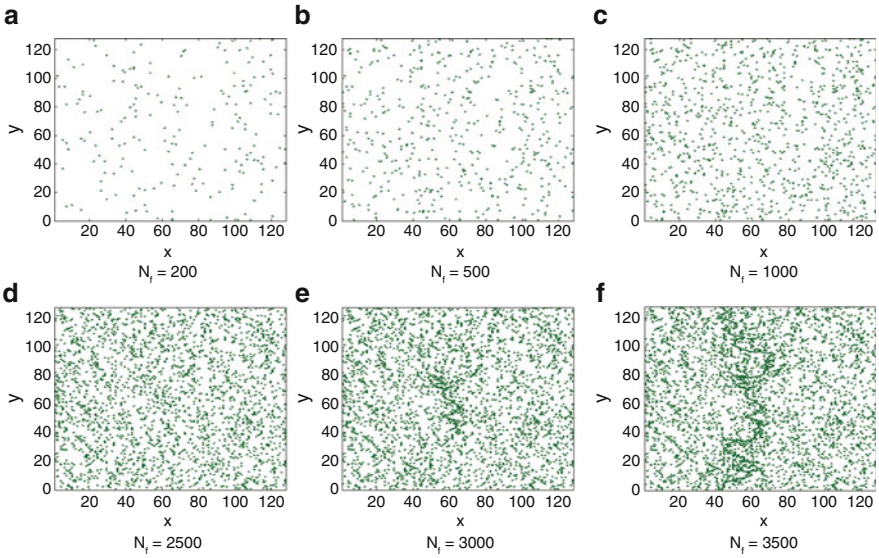


Fig. 5 Evolution of strain localization in an elastic–brittle material

Diffusive Fracture and Brittle Fracture. The evolution of damage is governed by the competition between disorder and stress concentrations at the micro-crack tips. The effect of stress concentrations trying to localize the damage is opposed by the disorder trying to delocalize the damage. For strong disorder case, disorder dominates the stress concentrations leading to distribution of spatially uncorrelated cracks in the initial stages of loading. As the loading increases, stress concentrations are high enough to overcome the barrier due to disorder, whereas for weak disorder, stress concentrations dominate the disorder in the initial stages itself leading to the formation of a crack originating from the weakest zone.

Kahng et al. (1988) have demonstrated using the electrical breakdown process in RFM that the elastic–brittle transition depends on two factors: the amount of disorder and the length scale of the medium under consideration. The schematic phase diagram of the elastic–brittle transition suggested in Kahng et al. (1988) is shown in Fig. 10. The strength of disorder is controlled by the factor w such that $p(t)$ is a uniform distribution on the support $[1 - w/2,\ 1 + w/2]t_{\max}$. Thus, $w = 2$ corresponds to maximum disorder, i.e., a uniform distribution over $[0,\ 2t_{\max}]$.

Two fundamentally different crack formation trends observed are termed as diffusive fracture [ductile fracture in Kahng et al. (1988)] and brittle fracture. The diffusive fracture is characterized by the appearance of spatially uncorrelated micro-cracks prior to the final catastrophic failure, which is observed generally for the systems with strong disorder. For systems with weak disorder, brittle fracture is observed. In brittle fracture, the weakest element within the domain dominates the final crack formation process. Crack nucleates from the weakest element and proceeds in the spanwise direction due to high stress concentration near the crack tip, leading to final failure.

The main idea proposed by the authors in the phase diagram (Fig. 10) is that for a given length scale (L), there exists a critical level of disorder for which transition from brittle to diffusive fracture behavior is observed for increasing disorder strength, whereas, for a fixed disorder strength (w), the brittle fracturing is observed for increasing macroscopic length scale (L), except for $w = 2$ which according to the authors (Kahng et al. 1988) approaches the percolation limit as $L \to \infty$.

Fractal Nature of Fracture Surfaces. That fracture surfaces of metals are self-affine fractals was first pointed out in the seminal paper by Mandelbrot and Paullay (1984). Their work has initiated a new research area dealing with the roughness coefficient (ζ) of the fracture. It is now confirmed through experiments that $\zeta \simeq 0.8$ (out of plane or 3D) is observed (Lapasset and Planes 1990), in a material dependent scaling domain (~down to μm length scales and high crack propagation speeds) for many ductile as well as brittle materials (Bouchad 1997). At smaller length scales (~down to *nm*) and quasi-static conditions (low crack propagation speeds), another ζ in the range of 0.4–0.6 is reported which is often associated with the fracture process zone (FPZ). In the case of 2D fracture surface, ζ is obtained to be in the range of 0.6–0.7 by experiments (mainly on paper samples) (Bonamy and Bouchad 2011).

The universality of ζ (at larger length scales at least) suggests that the fracture surface roughening process is governed by a typical physical phenomenon independent of the material properties, much like the existence of Kolmogorov scaling in the inertial subrange of isotropic turbulence (Hansen et al. 1991). Thus, the topic of crack surface roughness has attracted significant attention over the last 20 years. While the formation of fractal cracks may be observed in Figs. 5–9, the topic of fracture mechanics with fractal cracks is treated by Wnuk (2014a, b) in this handbook (Fig. 10).

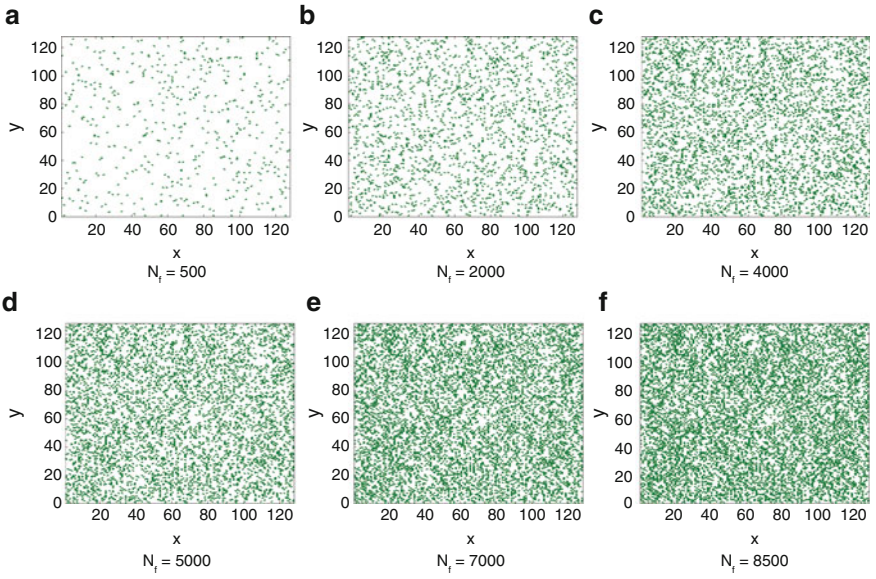


Fig. 6 Evolution of strain localization in an elastic–plastic material

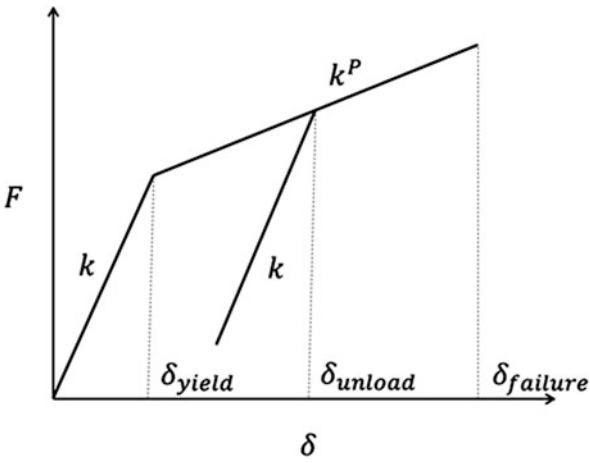


Fig. 7 Schematic of the elastic–plastic–brittle model of an anti-plane spring lattice

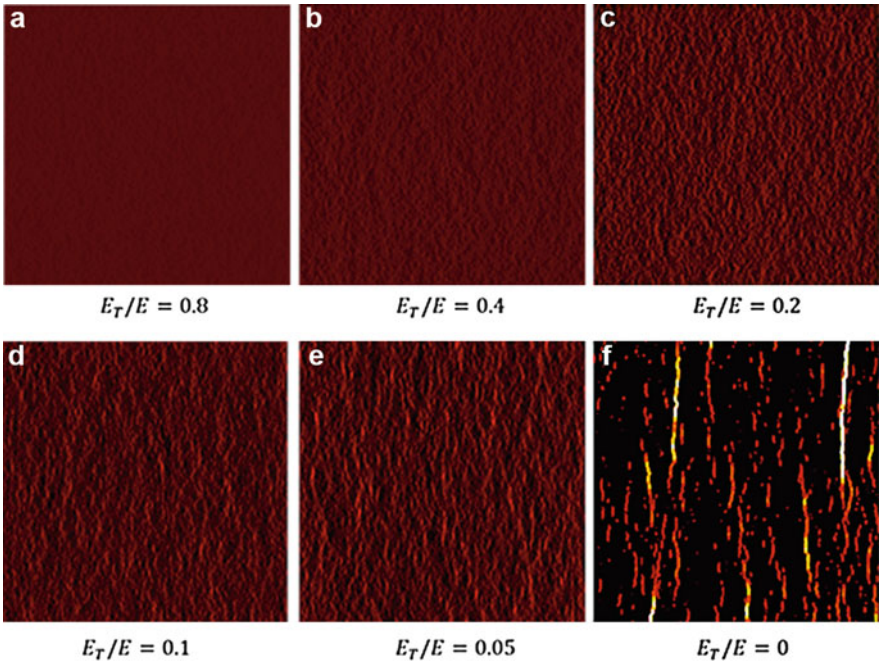


Fig. 8 Normalized strain contour plots for E_T/E values: (**a**) 0.8, (**b**) 0.4, (**c**) 0.2, (**d**) 0.1, (**e**) 0.05, and (**f**) (perfectly plastic). Shear bands due to strain localization of increasing strength are observed for decreasing E_T/E

Modeling Elastic–Plastic Materials

Elastic–plastic response is obtained allowing a hardening slope in the spring constitutive behavior. In the anti-plane elastic setting, the elastic–plastic response (flow rule) follows these equations at the grain level:

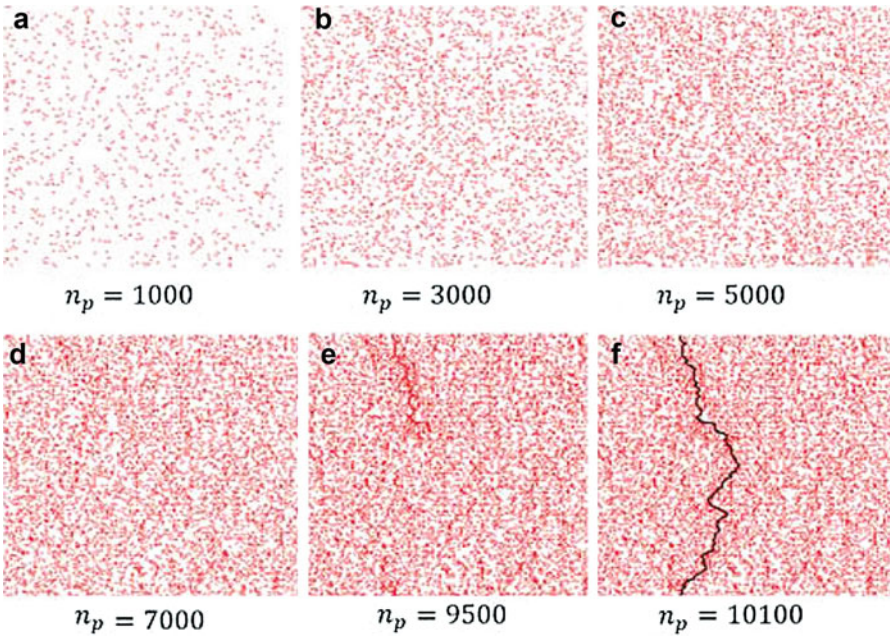


Fig. 9 Damage accumulation and localization leading to formation of the final macroscopic crack in (**f**) with increasing number of failed springs n_p.

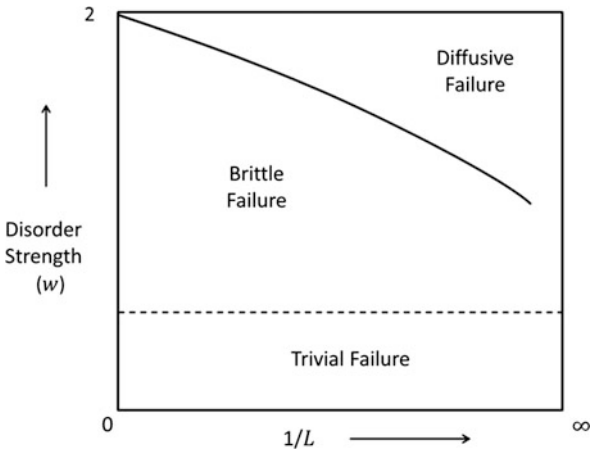


Fig. 10 Schematic phase diagram of elastic–brittle transition (Kahng et al. 1988)

$$\begin{aligned} d\sigma_{3i} &= C_{3i3j}\, d\varepsilon_{3i} \quad \text{when} \quad f_P < 0 \quad \text{or} \quad f_P = 0 \quad \text{and} \quad df_P < 0 \\ d\sigma_{3i} &= C^P_{3i3j}\, d\varepsilon^P_{3i} \quad \text{when} \quad f_P = 0 \quad \text{and} \quad df_P = 0. \end{aligned} \tag{49}$$

Here $d\varepsilon^P_{3i}$ is the plastic strain increment and C^P_{3i3j} represents the hardening modulus of the material. The Tresca criterion is used to define the yield function (f_P):

$$f^P = \max(\sigma_{31}, \sigma_{32}) - \sigma_s, \tag{50}$$

where σ_s is the yield stress in shear of the given grain. In the simulations, disorder is introduced by assigning yield thresholds to the springs following desired probability distribution. The simulation is performed similar to the elastic–brittle one, with loading increasing monotonically in small steps. After each loading step, all the springs exceeding the yield criteria are modified to follow the hardening slope and the system of equations is solved after each modification so as to account for stress redistributions. The simulation progresses until a fully plastic state is reached.

The elastic–plastic transition is essentially different than elastic–brittle one as the stress concentrations near the yielded zones are not as strong. Thus, no damage localization is observed. But, strain localization zones are observed for low-hardening materials due to weak post-yielding material response (Fig. 6).

Modeling Elastic–Plastic–Brittle Materials

The entire process of failure of plastic hardening disordered materials can be captured by implementing the complete bilinear response as shown in Fig. 7. The elastic unloading behavior is an essential part of the model and is explicitly accounted for as the formation of micro-cracks may result into local unloading of a yielded zone. In the simulations, at each loading step after a failure or yielding event, every yielded spring that is unloading is modified to follow the elastic unloading response.

The effective model response is now controlled by three parameters: hardening ratio, strength of disorder in yield, and failure limits. Based on observations from elastic–brittle and elastic–plastic responses, some results can be intuitively expected from the elastic–plastic–brittle model. For high hardening ratio, significant post-yield load-carrying capacity of the material avoids the formation of strain localization zones and hence the system closely follows elastic–brittle behavior. Otherwise, for low hardening ratio, the formation of micro-cracks within the strain-localized zones accelerates the failure process leading to lower effective strength of the lattice. The strength of disorder plays the important role of mitigating the effect of crack-tip stress concentrations by delocalizing the micro-cracks. Thus, a wide spectrum of material responses can be modeled by modifying the disorder distribution and the constitutive spring responses.

A similar idea is studied in the realm of fiber bundle model (FBM) by Rinaldi (2011) using a bilinear fiber response and independent yield and failure threshold probability distributions.

Damage Patterns and Maps of Disordered Elastic–Brittle Composites

As mentioned in section "Introduction," the lattice method can also be used to simulate damage of heterogeneous materials. This works particularly well in the case of elastic–brittle failure of composites, where one uses a mesh (much) finer

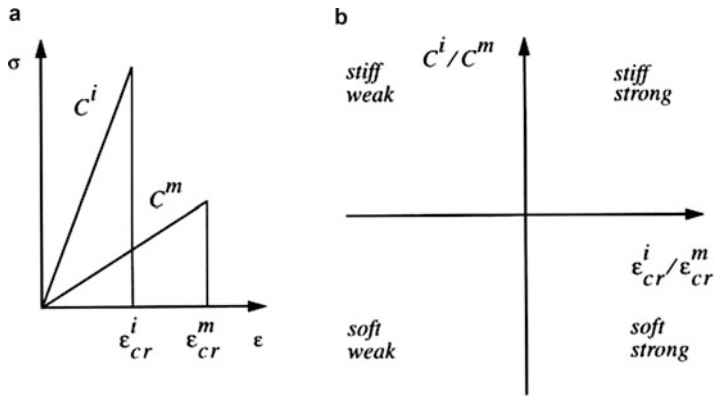


Fig. 11 (**a**) Elastic–brittle stress–strain curves for matrix and inclusion phases; (**b**) sketch of the damage plane

than the typical size of the microstructure. In principle, one needs to determine which lattice spacing ensures mesh independence or nearly so. Such a study has been conducted for a thin aluminum polycrystalline sheet discussed (Ostoja-Starzewski 2008).

With reference to Alzebdeh et al. (1998), the attention is focused on two-phase composites in anti-plane shear, under periodic boundary conditions and (necessarily) periodic geometries. Now, since both phases (inclusion i and matrix m) are isotropic and elastic–brittle (Fig. 11a), the composite can be characterized by two dimensionless parameters:

$$\varepsilon^i_{cr}/\varepsilon^m_{cr} \quad C^i/C^m, \tag{51}$$

where ε^i_{cr} (ε^m_{cr}) is the strain-to-failure of the inclusion (resp. matrix) phase and C^i (C^m) are the corresponding stiffnesses. This leads to the concept of a *damage plane* (Fig. 11b) showing various combinations of strengths and stiffnesses. While the response in the first and third quarters of damage plane is quite intuitive, this is not so for the second and fourth quarters. In those two quarters, there is a competition of either high stiffness with low strength of the inclusions with the reverse properties of the matrix or the opposite of that. The damage plane is useful for displaying effective damage patterns of any particular geometric realization of the random composite while varying its physical properties (Fig. 12) as well as other characteristics, say, statistics of response in the ensemble sense (Fig. 13).

A number of other issues are studied in the referenced papers:

- Stress and strain concentrations
- Finite size scaling of response

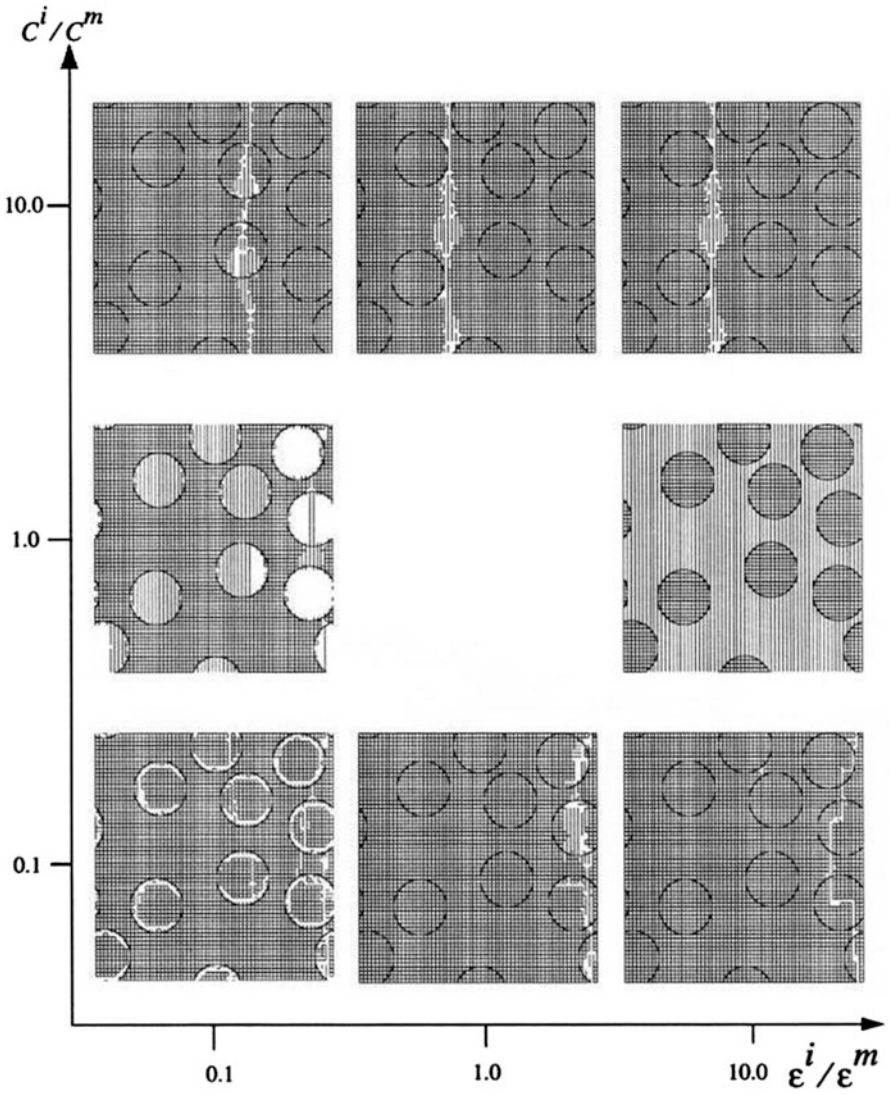


Fig. 12 Crack patterns in the damage plane on a scale 4.5 times larger than that of the inclusion diameter. The center figure of a homogeneous body is not shown as it corresponds to all the bonds failing simultaneously

– Function fitting of statistics (where it turns out that the beta probability distribution offers a more universal fit than either Weibull or Gumbel)
– Effects of disorder versus periodicity

Also, see Ostoja-Starzewski and Lee (1996) for a similar study under in-plane loading; computer movies of evolving damage can be obtained from the author.

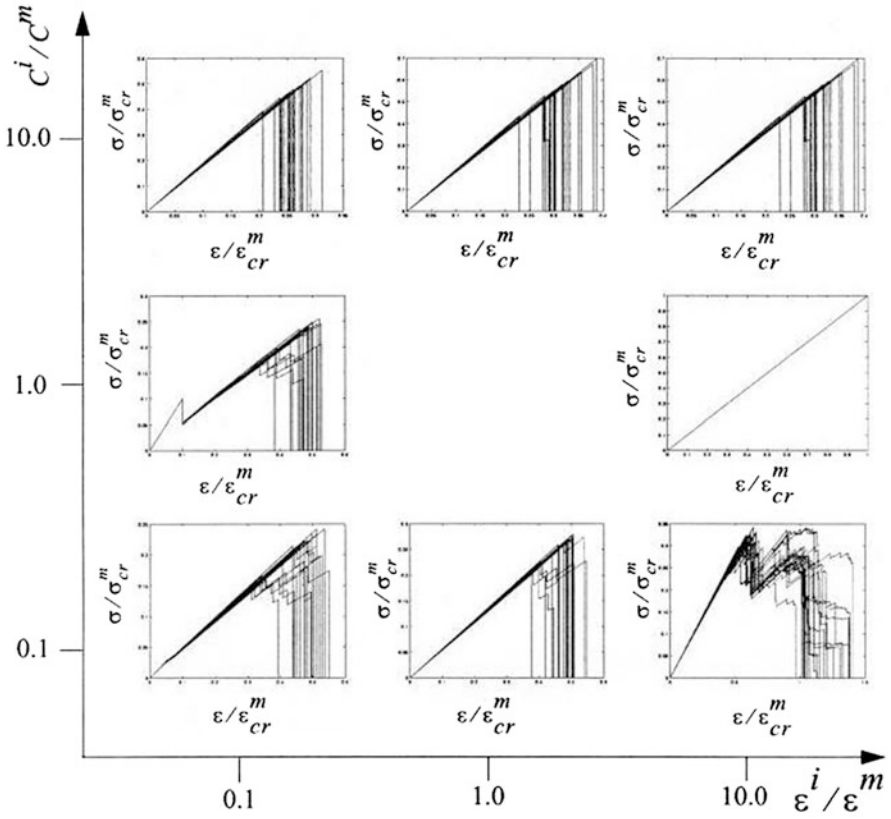


Fig. 13 Damage maps of statistics of constitutive responses for twenty realizations of the random composite, such as that in Fig. 12

Particle Models

Governing Equations

Basic Concepts

Particle models are a generalization of lattice models to include dynamic effects and can also be viewed as an offshoot of *molecular dynamics* (MD). The latter field has developed over the past few decades in parallel with the growth of computers and computational techniques. Its objective has been to simulate many interacting atoms or molecules in order to derive macroscopic properties of liquid or solid materials (Greenspan 1997, 2002; Hockney and Eastwood 1999). The governing Hamiltonian differential equations of motion need to be integrated over long time intervals so as to extract the relevant statistical information about the system from the computed trajectories.

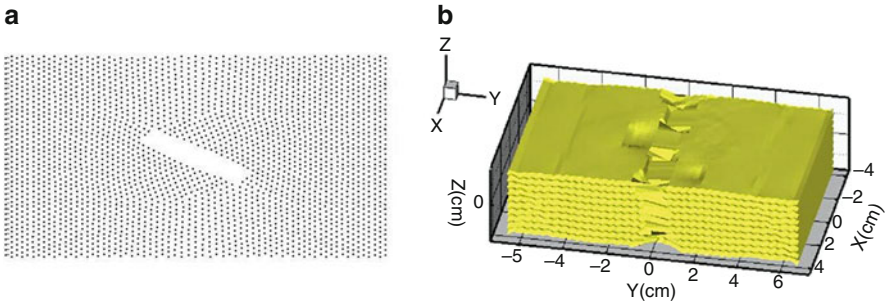


Fig. 14 Particle models and intermediate stages of fracture in (**a**) 2D and (**b**) 3D; after Wang et al. (2006)

Techniques of that type have been adapted over the past two decades to simulate materials at larger length scales, whereby the role of a particle is played by a larger-than-molecular piece of material, a so-called particle or quasiparticle. The need to reduce the number of degrees of freedom in complex systems has also driven models of galaxies as systems of quasiparticles, each representing lumps of large numbers of stars. In all these so-called particle models (PMs), the material is discretized into particles arranged in a periodic lattice, just like in spring network models studied in earlier sections, yet interacting through nonlinear potentials, and accounting for inertial effects, i.e., full dynamics. With reference to Fig. 14, the lattice may be in 2D or in 3D.

Note that, by comparison with finite elements (FE) which indeed also involve a quite artificial spatial partitioning, PMs are naturally suited to involve interparticle potentials of the same functional form as the interatomic potentials, providing one uses the same type of lattice. The PM can therefore take advantage of the same numerical techniques as those of MD and rather easily deal with various highly complex motions. Thus, the key issue is how to pass from a given molecular potential in MD to an interparticle potential in PM. In the case when the molecular interactions are not well known, the PM may still turn out to be superior relative to the FE. Among others, this indeed is the case with comminution of minerals where scales up to meters are involved (Wang and Ostoja-Starzewski 2005; Wang et al. 2006).

In MD, the motion of a system of atoms or molecules is governed by classical molecular potentials and Newtonian mechanics. As an example, let us consider copper. Following Greenspan (1997), its 6–12 Lennard–Jones potential is

$$\phi(r) = -\frac{1.398068}{r^6}10^{-10} + \frac{1.55104}{r^{12}}10^{-8}\text{erg}. \tag{52}$$

Here r is measured in Å. It follows that the interaction force between two copper atoms is

$$F(r) = -\frac{d\phi(r)}{dr} = -\frac{8.388408}{r^7}10^{-2} + \frac{18.61248}{r^{13}}\text{dyn}. \tag{53}$$

In Eq. 52 $F(r) = 0$ occurs at $r_0 = 2.46$Å, and ϕ then attains the minimum: $\phi(r_0) = -3.15045 \cdot 10^{-13}$ erg.

Using a simple methodology from basic materials science (Ashby and Jones 1980), Young's modulus E of the material can be found from $\phi(r)$ as follows:

$$E = \frac{S_0}{r_0} \quad \text{where} \quad S_0 = \frac{d^2\phi(r)}{dr^2}|_{r_0}. \tag{54}$$

With this method, Young's modulus of copper 152.942 GPa is found, a number that closely matches the physical property of copper and copper alloys valued at 120~150 GPa. Then, the continuum-type tensile stress

$$\sigma(r) = NF(r), \tag{55}$$

where N is the number of bonds/unit area, equals to $1/r_0^2$. Tensile strength, σ_{TS}, results when $dF(r)/dr = 0$, that is, at $r_d = 2.73$Å (*bond damage spacing*), and yields

$$\sigma_{TS} = NF(r_d) = 462.84\,\mathrm{MN/m^2}. \tag{56}$$

This value is quite consistent with data for the actual copper and copper-based alloys reported at 250–1,000 Mpa.

In the PM, the interaction force is also considered only between nearest-neighbor (quasi)particles and assumed to be of the same form as in MD:

$$\phi(r) = -\frac{G}{r^p} + \frac{H}{r^q}. \tag{57}$$

Here G, H, p, and q, all positive constants, are yet to be determined, and this will be done below. Inequality $q > p$ must hold so as to obtain the repulsive effect that is necessarily (much) stronger than the attractive one. Three examples of interaction force for three pairs of p and q are displayed in Fig. 15a. The dependence of Young's modulus for a wide range of p and q is shown in Fig. 15b.

The conventional approach in PM, just as in MD, is to take the equation of motion for each particle P_i of the system as

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \alpha \sum_j \left(-\frac{G_i}{r_{ij}^p} + \frac{H_i}{r_{ij}^q} \right) \frac{\mathbf{r}_{ji}}{r_{ij}} \quad i \neq j, \tag{58}$$

where m_i is the mass of P_i and $\mathbf{r}_{ji}$ is the vector from P_j to P_i; summation is taken over all the neighbors of P_i. Also, α is a normalizing constant obtained by requiring that the force between two particles must be small in the presence of gravity:

$$\alpha \left| -\frac{G_i}{D^p} + \frac{H_i}{D^q} \right| < 0.001 \cdot 980 \mathrm{m}_i. \tag{59}$$

Here D is the distance of local interaction ($1.7r_0$ cm in this particular example), where r_0 is the equilibrium spacing of the particle structure. The reason for

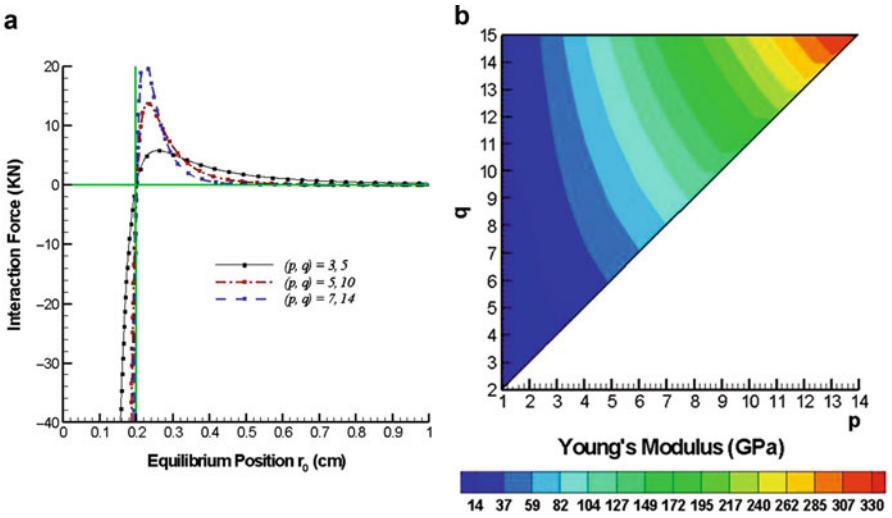


Fig. 15 (**a**) The interaction force for pairs of (p,q) exponents, at $r_0 = 0.2$ cm. (**b**) The variability of Young's modulus in the (p,q) − plane

introducing the parameter α by Greenspan (1997) was to define the interaction force between two particles as local in the presence of gravity. However, since setting α according to Eq. 49 would result in a "pseudo-dynamic" solution, $\alpha = 1$ is set.

According to Eq. 47, different (p, q) pairs result in different continuum-type material properties, such as Young's modulus E. Clearly, changing r_0 and volume of the simulated material V $(= A \times B \times C)$ will additionally influence Young's modulus. Therefore, in general, there is some functional dependence:

$$E = E(p, q, r_0, V). \tag{60}$$

One can formulate four conditions to determine continuum-level Young's modulus and tensile strength while maintaining the conservation of mass and energy of the particle system and satisfying the interaction laws between all the particles in the PM model for a given MD model (Wang and Ostoja-Starzewski 2005).

Leapfrog Method

Just like in MD, there are two commonly used numerical schemes in particle modeling: *completely conservative method* and *leapfrog method*. The first scheme is exact in that it perfectly conserves energy and linear and angular momentum, but requires a very costly solution of a large algebraic problem. The second scheme is approximate. Since in most problems one needs large numbers of particles to adequately represent a simulated body, the completely conservative method is unwieldy and, therefore, usually abandoned in favor of the leapfrog method (Ostoja-Starzewski and Wang 2006). That method is derived by considering Taylor expansions of positions $\mathbf{r}_{i,k+1}$ and $\mathbf{r}_{i,k}$ of the particle $P_i (i = 1, 2, \ldots, N)$ at

times $t_k = k\Delta t$ and $t_{k+1} = (k+1)\Delta t$, respectively, about time $t_{k+1/2} = (k+1/2)\Delta t$ (with Δt being the time step):

$$\begin{aligned}\mathbf{r}_{i,k} &= \mathbf{r}_{i,k+1/2} - \frac{\Delta t}{2}\mathbf{v}_{i,k+1/2} + \frac{\Delta t^2}{4}\mathbf{a}_{i,k+1/2} - \frac{\Delta t^3}{48}\dot{\mathbf{a}}_{i,k+1/2} + O\left(\Delta t^4\right) \\ \mathbf{r}_{i,k+1} &= \mathbf{r}_{i,k+1/2} + \frac{\Delta t}{2}\mathbf{v}_{i,k+1/2} + \frac{\Delta t^2}{4}\mathbf{a}_{i,k+1/2} + \frac{\Delta t^3}{48}\dot{\mathbf{a}}_{i,k+1/2} + O\left(\Delta t^4\right).\end{aligned} \tag{61}$$

Here $\mathbf{v}_i$ and $\mathbf{a}_i$ denote velocity and acceleration. Upon addition and subtraction of these, the new position and velocity are found:

$$\begin{aligned}\mathbf{r}_{i,k+1} &= 2\mathbf{r}_{i,k+1/2} - \mathbf{r}_{i,k} + \frac{\Delta t^2}{4}\mathbf{a}_{i,k+1/2} + O\left(\Delta t^4\right) \\ \mathbf{v}_{i,k+1/2} &= \left(\mathbf{r}_{i,k+1} + \mathbf{r}_{i,k}\right)/\Delta t + O\left(\Delta t^2\right),\end{aligned} \tag{62}$$

showing that the position calculation is two orders of magnitude more accurate than the velocity calculation. However, the error in computation of velocity accumulates only as fast as that in position because it is really being calculated from positions. It is easy to see that the leapfrog method is more accurate than the conventional Euler integration based on $\mathbf{v}_{i,k+1} = \mathbf{v}_{i,k} + (\Delta t)\mathbf{a}_{i,k}$ and $\mathbf{r}_{i,k+1} = \mathbf{r}_{i,k} + (\Delta t)\mathbf{v}_{i,k}$.

Oftentimes, the leapfrog formulas relating position r_i, velocity v_i, and acceleration a_i for all the particles $P_i (i = 1, 2, \ldots, N)$ are written as

$$\begin{aligned}\mathbf{v}_{i,1/2} &= \mathbf{v}_{i,0} + \tfrac{\Delta t}{2}\mathbf{a}_{i,0} \quad \text{(starter formula)} \\ \mathbf{v}_{i,k+1/2} &= \mathbf{v}_{i,k-1/2} + (\Delta t)\mathbf{a}_{i,k} \quad k = 0, 1, 2, \ldots \\ \mathbf{r}_{i,k+1} &= \mathbf{r}_{i,k} + (\Delta t)\mathbf{v}_{i,k+1/2} \quad k = 0, 1, 2, \ldots\end{aligned} \tag{63}$$

Clearly, the name of the method comes from taking velocities at intermediate time steps relative to positions and accelerations; it is also known as a *Verlet algorithm.*

It can be shown that the global (cumulative) error in position going from $\mathbf{r}_{i,k}$ to $\mathbf{r}_{i,k+n}$ (i.e., over $T = n\Delta t$) of P_i is

$$\text{error}\left(\mathbf{r}_{i,k+n} - \mathbf{r}_{i,k}\right) = O\left(\Delta t^2\right), \tag{64}$$

which is also the global error in velocity.

Stability is concerned with the propagation of errors. Even if the truncation and roundoff errors are very small, a scheme would be of little value if the effects of small errors were to grow rapidly with time. Thus, instability arises from the nonphysical solution of the discretized equations. If the discrete equations have solutions which grow much more rapidly than the correct solution of the differential equations, then even a very small roundoff error is certain to eventually seed that solution and render the numerical results meaningless. By the root locus method for an atomistic unit of time, the safe time step used in the leapfrog method meeting this requirement is

$$\Omega\Delta t \ll 2 \quad \Omega = \left(\frac{1}{m}\left|\frac{dF}{dr}\right|_{\max}\right)^{1/2}. \tag{65}$$

Thus, as $r \to 0$, $dF/dr \to \infty$, which results in $\Delta t \to 0$. Since this may well cause problems in computation, it is advisable to introduce the smallest distance between two particles according to these conditions:

(i) For a stretching problem of a plate/beam, take $(dF/dr)_{\max} \simeq dF/dr|_{r=r_0}$, which with () dictates $\Delta t \simeq 10^{-7} - 10^{-6} s$.
(ii) For an impact problem, one often needs to set up a minimum distance limiting the spacing between two nearest particles, e.g., $r_{\min} = 0.1 r_0$. It is easy to see from Fig. 15a that, in this case, this suitable time increment is greatly reduced because of a rapid increase in Ω. This leads to $\Delta t \simeq 10^{-8} s$.

Following the MD methodology (Napier-Munn et al. 1999), one can also set up a criterion for convergence: $\Delta t < 2\sqrt{m/k}$, where m is the smallest mass to be considered and k is the same stiffness as S_0 in Eq. 44_2. An examination of these two criteria shows there is not much quantitative difference between them in the case of elastic or elastic–brittle, but not plastic, materials.

Examples

The maximum entropy formalism is much better suited to deal with quasi-static rather than dynamic fracture. The dynamic character of fracture in these experiments, combined with the presence of multiple incipient spots, was also a big challenge in several computational mechanics models reviewed in Al-Ostaz and Jasiuk (1997) employing commercial finite element programs, as well as in an independent study using a meshless element program (Belytschko et al. 1995). Upon trying various failure criteria and subjective choices (such as being forced to initialize the cracking process in the meshless model), the simulations have run into uncertainty as to which modeling aspect is more critical and whether there is a way to clarify it. A more recent study (Ostoja-Starzewski and Wang 2006) was motivated by that challenge and offered a way to test the PM vis-à-vis experiments.

The same experiments on crack patterns in epoxy plates perforated by holes were also treated by two analyses – one based on the minimum potential energy formulation and another based on the maximum entropy method; they both relied on the assumption of quasi-static response. Strictly speaking, while the loading was static, the fragmentation event was dynamic. Clearly, the preparation of mineral specimens – involving measurement of highly heterogeneous and multiphase microstructures – for a direct comparison with the model prediction is very hard. Thus, the model is applied to the experimentally tested plate with 31 holes according to this strategy:

- Decrease the lattice spacing until mesh-independent crack patterns are attained.
- Find whether the lattice of (i) will also result in the most dominant crack pattern of Fig. 16. Indeed, the crack patterns "stabilize" as the mesh is refined.
- Assuming the answer to (ii) is positive, introduce weak perturbations in the material properties – either stiffness or strength – to determine which one of these has a stronger effect on the deviation away from the dominant crack pattern, i.e., on the scatter in Fig. 16.

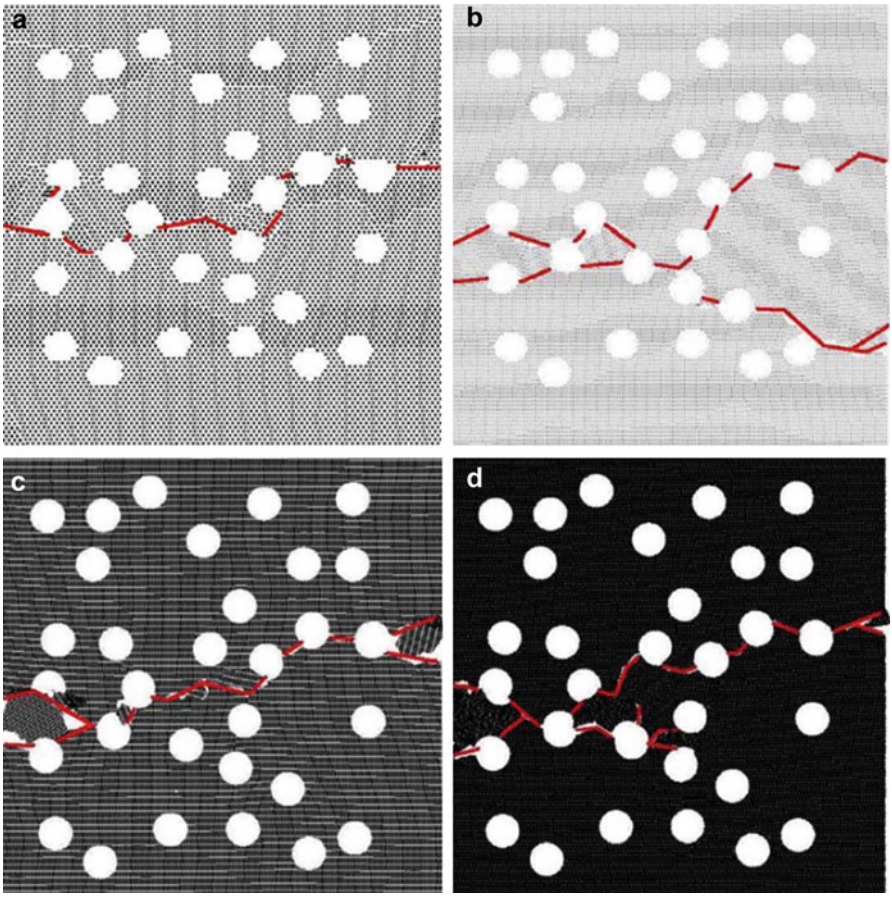


Fig. 16 From (Ostoja-Starzewski and Wang 2006) final crack patterns for four mesh configurations at ever finer lattice spacings: (**a**) $r_0 = 0.1$ cm, (**b**) $r_0 = 0.05$ cm, (**c**) $r_0 = 0.02$ cm, and (**d**) $r_0 = 0.01$ cm

Other Models

The PM is but one of the variations on the theme of MD. Here are some other possibilities:

- Molecular statics (MS) – by disregarding the inertia forces, it involves a static solution of the system of atoms (Vinogradov 2006, 2009, 2010). While the MD allows simulations of large systems with a constraint to very short time scales (transient phenomena of the order of nanoseconds), the MS allows large (macroscopic type) time scales albeit with a limitation by the size of a (nonlinear) algebraic system one is able to solve and a restriction to 0°K.

- Derivation of a continuum model from a microscopic model based on the assumption that the displacements on the macroscopic level are the same as those on the molecular level (Blanc et al. 2002).
- Introduction of a finite extension and spin for continuum-type particles (Yserentant 1997).
- Direct incorporation of interatomic potentials into a continuum analysis on the atomic scale (Zhang et al. 2002).

Scaling and Stochastic Evolution in Damage Phenomena

Consider a material whose elasticity is coupled to damage state, as described by the constitutive equation (Lemaitre and Chaboche 1994)

$$\sigma_{ij} = (1 - D)C_{ijkl}\varepsilon_{kl}. \tag{66}$$

Here C_{ijki} is isotropic, and which must be coupled with a law of isotropic damage, that is,

$$\dot{D} = \frac{\partial \Phi^*}{\partial Y}, \tag{67}$$

with $Y = -\partial\Psi/\partial\varepsilon$, being the Helmholtz free energy. This formulation is set within the TIV (thermomechanics with internal variables) framework. In particular, the scalar D evolves with the elastic strain $\varepsilon = \varepsilon_{ii}$, which is taken as a time-like parameter, according to

$$\frac{\partial D}{\partial \varepsilon} = \begin{cases} (\varepsilon/\varepsilon_0)^{s^*} & \text{when} \quad \varepsilon = \varepsilon_D \quad \text{and} \quad d\varepsilon = d\varepsilon_D > 0, \\ 0 & \text{when} \quad \varepsilon < \varepsilon_D \quad \text{and} \quad d\varepsilon < 0. \end{cases} \tag{68}$$

Integration from the initial conditions $D = \varepsilon_D = 0$ up to the total damage, $D = 1$, gives

$$D = (\varepsilon/\varepsilon_0)^{s^*+1} \quad \varepsilon_R = \left[(1 + s^*)\varepsilon_D^{s^*}\right]^{s^*+1} \quad \sigma = \left[1 - (\varepsilon/\varepsilon_R)^{s^*+1}\right] E\varepsilon, \tag{69}$$

where $\sigma = \sigma_{ii}$.

This formulation is understood as the effective law for the RVE, that is,

$$C_{ijkl}^{eff} = C_{ijkl|\delta\to\infty} \qquad D^{\text{eff}} = D_{|\delta\to\infty} \quad \Psi^{\text{eff}} = \Psi_{|\delta\to\infty} \quad \Phi^{\text{eff}} = \Phi_{|\delta\to\infty}, \tag{70}$$

as well as a guidance for adopting the form of apparent responses on mesoscales. Thus, assuming that the same types of formulas hold for any mesoscale δ, the apparent response for any specimen $B_\delta(\omega)$ is

$$\overline{\boldsymbol{\sigma}} = \left(1 - D_\delta^d\right)\mathbf{C}_\delta^d(\omega) : \varepsilon^0 \tag{71}$$

under uniform displacement boundary condition. The notation D_δ^d expresses the fact that the material damage is dependent on the mesoscale δ and the type of boundary conditions applied (^{d}). In fact, while one could formally write another apparent

response $\overline{\varepsilon} = \left(1D_\delta^d\right)^{-1}\mathbf{S}_\delta^t(\omega) : \boldsymbol{\sigma}^0$, this is not done because the damage process under the traction boundary condition (^{t}) would be unstable.

It is now possible to obtain scale-dependent bounds on D_δ^d through a procedure analogous to that for linear elastic materials, providing one assumes a WSS and ergodic microstructure. One then obtains a hierarchy of bounds on $\langle D_\infty^d \rangle \equiv D^{\text{eff}}$ from above (Ostoja-Starzewski 2002b):

$$\left\langle D_{\delta'}^d \right\rangle \leq \left\langle D_\delta^d \right\rangle \leq \ldots \leq \left\langle D_\infty^d \right\rangle \quad \forall \delta' = \delta/2. \tag{72}$$

These inequalities are consistent with the much more phenomenological Weibull model of scaling of brittle solids saying that the larger is the specimen, the more likely it is to fail.

Next of interest is the formulation of a stochastic model of evolution of D_δ^d with ε to replace Eq. 73_1. Said differently, a stochastic process $D_\delta^d = \{D_\delta^d(\omega, \varepsilon); \omega \in \Omega, \varepsilon \in [0, \varepsilon_R]\}$ is needed. Assuming, for simplicity of discussion, just as in Lemaitre and Caboche (1994) that $s^* = 2$, this setup may be considered:

$$dD_\delta^d(\omega, \varepsilon) = D_\delta^d(\omega, \varepsilon) + 3\varepsilon^2[1 + r_\delta(\omega)]\, dt, \tag{73}$$

where $r_\delta(\omega)$ is a zero-mean random variable taking values from $[-a_\delta, a_\delta]$, $1/\delta = a_\delta < 1$. This process has the following properties:

(i) Its sample realizations display scatter ω-by-ω for $\delta < \infty$, i.e., for finite body sizes.
(ii) It becomes deterministic as the body size goes to infinity in the RVE limit ($\delta \to \infty$)
(iii) Its sample realizations are weakly monotonically increasing functions of ε
(iv) Its sample realizations are continuous.
(v) The scale effect inequality (Eq. 73) is satisfied, providing ε_R is taken as a function of δ with a property

$$\varepsilon_R(\delta) < \varepsilon_R(\delta') \quad \forall \delta' = \delta/2. \tag{74}$$

Let us observe, however, that, given the presence of a random microstructure, mesoscale damage should be considered as a sequence of microscopic events – shown as impulses in Fig. 17a – thus rendering the apparent damage process D_δ^d one with discontinuous sample paths, having increments dD_δ^d occurring at discrete time instants (Fig. 17c). To satisfy this requirement, one should, in place of the above, take a *Markov jump process* whose range is a subset [0,1] of real line (i.e., where D_δ^d takes values). This process would be specified by an evolution propagator or, more precisely, by a *next-jump probability density function* defined as follows:

$p(\varepsilon', D_\delta^{d\prime}|\varepsilon, D_\delta^d)d\varepsilon' dD_\delta^{d\prime}$ = probability that, given the process is in state D_δ^d at time ε, its next jump will occur between times $\varepsilon + \varepsilon'$ and $\varepsilon + \varepsilon' + d\varepsilon'$ and will carry the process to some state between $D_\delta^d + D_\delta^{d\prime}$ and $D_\delta^d + D_\delta^{d\prime} + dD_\delta^{d\prime}$.

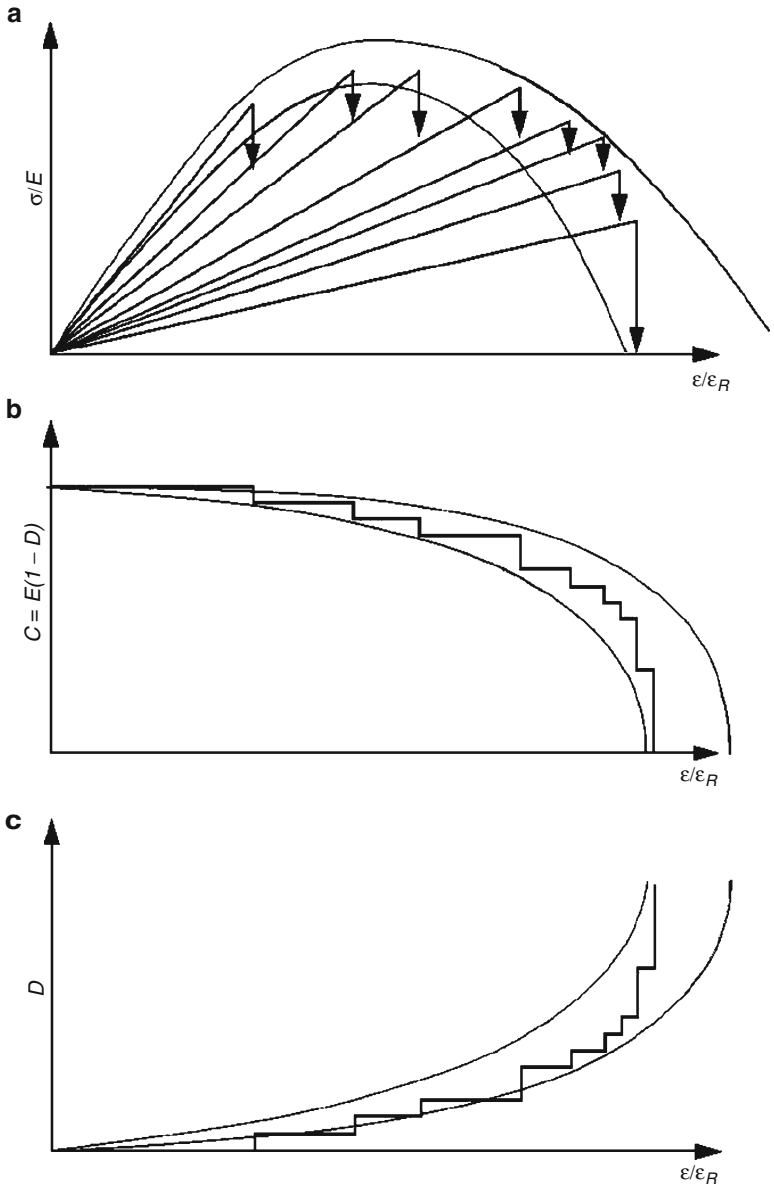


Fig. 17 Constitutive behavior of a material with elasticity coupled to damage where $\varepsilon/\varepsilon_R$ plays the role of a controllable, time-like parameter of the stochastic process. (**a**) Stress–strain response of a single specimen $\mathcal{B}_\delta$ from $\mathcal{B}$ having a zigzag realization, (**b**) deterioration of stiffness, and (**c**) evolution of the damage variable. Curves shown in (**a**–**c**) indicate the scatter in stress, stiffness, and damage at finite scale δ. Assuming spatial ergodicity, this scatter would vanish in the limit $\delta \to \infty$, whereby unique response curves of continuum damage mechanics would be recovered

Figure 17b shows one realization $C_\delta^d(\omega, \varepsilon)$; $\omega \in \Omega, \varepsilon \in [0, \varepsilon_R]$, of the apparent, mesoscale stiffness, corresponding to the realization $D_\delta^d(\omega, \varepsilon)$; $\omega \in \Omega, \varepsilon \in [0, \varepsilon_R]$, of Fig. 17c. The resulting constitutive response $\sigma_\delta(\omega, \varepsilon)$; $\omega \in \Omega, \varepsilon \in [0, \varepsilon_R]$ is depicted in Fig. 17a.

Calibration of this model (just as the simpler one above) – that is, the specification of $p(\varepsilon', D_\delta^{d\prime}|\varepsilon, D_\delta^d)d\varepsilon' dD_\delta^{d\prime}$ – may be conducted by either laboratory or computer experiments such as those discussed earlier in this chapter. Note that in the macroscopic picture ($\delta \to \infty$), the zigzag character and randomness of an effective stress–strain response vanish. However, many studies in mechanics/physics of fracture of random media (e.g., Herrmann and Roux 1990) indicate that the homogenization with $\delta \to \infty$ is generally very slow and hence that the assumption of WSS and ergodic random fields may be too strong for many applications; see Rinaldi (2013) for related work.

Extension of the above model from isotropic to (much more realistic) anisotropic damage will require tensor rather than scalar random fields and Markov processes. This will lead to a greater mathematical complexity which may be balanced by choosing the first model of this subsection rather than the latter. These issues, while technically challenging and offering rich harvest for theoreticians, are quite secondary relative to the underlying goal to outline a stochastic continuum damage mechanics that (i) is based on, and consistent with, micromechanics of random media as well as the classical thermomechanics formalism and (ii) reduces to the classical continuum damage mechanics in the infinite volume limit.

Concluding Remarks

The motivation for lattice models in damage mechanics has been the need to simultaneously model elastic, plastic, and fracture responses in heterogeneous materials, something that cannot easily be delivered by conventional continuum solid mechanics and finite element analysis. Following a review of basic concepts of lattice models in anti-plane, planar classical, and planar nonclassical elasticity settings, various applications to damage mechanics (including fractal characteristics of fracture) have been given. The discussion has then been expanded from statics to dynamics (i.e., to a quasiparticle model). The chapter closes with a discussion of scaling and stochastic evolution in damage phenomena as stepping-stone to stochastic continuum damage mechanics.

References

M.J. Alava, P.K.V.V. Nukala, S. Zapperi, Statistical models of fracture. Adv. Phys. **55**(3–4), 349–476 (2006)

A. Al-Ostaz, I. Jasiuk, Crack initiation and propagation in materials with randomly distributed holes. Eng. Fract. Mech. **58**, 395–420 (1997)

K. Alzebdeh, A. Al-Ostaz, I. Jasiuk, M. Ostoja-Starzewski, Fracture of random matrix-inclusion composites: scale effects and statistics. Int. J. Solids Struct. **35**(19), 2537–2566 (1998)

M.F. Ashby, D.R.H. Jones, *Engineering Materials 1: An Introduction to their Properties and Applications* (Pergamon Press, Oxford, 1980)
T. Belytschko, Y.Y. Lu, L. Gu, Crack propagation by element-free Galerkin method. Eng. Fract. Mech. **51**, 295–313 (1995)
M.D. Bird, C.R. Steele, A solution procedure for Laplace's equation on multiply connected circular domains. J. Appl. Mech. **59**(2), 398–404 (1992)
X. Blanc, C. LeBris, P.-L. Lions, From molecular models to continuum mechanics. Arch. Ration. Mech. Anal. **164**, 341–381 (2002)
A. Bonamy, E. Bouchad, Failure of heterogeneous materials: a dynamic phase transition. Phys. Rep. **498**, 1–44 (2011)
E. Bouchad, Scaling properties of cracks. J. Phys. Conden. Matter **9**, 4319–4343 (1997)
G.A. Buxton, C.M. Care, D.J. Cleaver, A lattice spring model of heterogeneous materials with plasticity. Model. Simul. Mater. Sci. Eng. **9**, 485–497 (2001)
D. Capecchi, G. Giuseppe, P. Trovalusci, From classical to Voigt's molecular models in elasticity. Arch. Hist. Exact Sci. **64**, 525–559 (2010)
L. De Arcangelis, S. Redner, H.J. Hermann, A random fuse model for breaking processes. J. Phys. Lett. **46**, 585–590 (1985)
E.J. Garboczi, M.F. Thorpe, M.S. DeVries, A.R. Day, Universal conductance curve for a plane containing random holes. Phys. Rev. A **43**, 6473–6480 (1991)
D. Greenspan, *Particle Modeling* (Birkhauser Publishing, Basel, 1997)
D. Greenspan, New approaches and new applications for computer simulation of N-body problems. Acta Appl. Math. **71**, 279–313 (2002)
A. Hansen, E.L. Hinrichsen, S. Roux, Scale invariant disorder in fracture and related breakdown phenomena. Phys. Rev. B **43**(1), 665–678 (1991)
F.W. Hehl, Y. Itin, The Cauchy relations in linear elasticity. J. Elast. **66**, 185–192 (2002)
R. Hill, Elastic properties of reinforced solids: Some theoretical principles, J. Mech. Phys. Solids. **11**, 357–372 (1963)
R.W. Hockney, J.W. Eastwood, *Computer Simulation Using Particles* (Institute of Physics Publishing, Bristol, 1999)
C. Huet, Universal conditions for assimilation of a heterogeneous material to an effective medium, Mech. Res. Comm. **9**(3), 165–170 (1982)
C. Huet, Application of variational concepts to size effects in elastic heterogeneous bodies, J. Mech. Phys. Solids 38, 813–841 (1990)
B. Kahng, G. Batrouni, S. Redner, Electrical breakdown in a fuse network with random, continuously distributed breaking strengths. Phys. Rev. B **37**(13), 7625–7637 (1988)
P.N. Keating, Effect of invariance requirements on the elastic strain energy of crystals with application to the diamond structure. Phys. Rev. **145**, 637–645 (1966)
J.G. Kirkwood, The skeletal modes of vibration of long chain molecules. J. Chem. Phys. **7**, 506–509 (1939)
D. Krajcinovic, *Damage Mechanics* (North-Holland, Amsterdam, 1996)
G. Lapasset, J. Planes, Fractal dimension of fractured surfaces: a universal value? EuroPhys. Lett. **13**(1), 73–79 (1990)
J. Lemaitre, J.-L. Chaboche, *Mechanics of Solid Materials* (Cambridge University Press, Cambridge, 1994)
A.E.H. Love, *The Mathematical Theory of Elasticity* (Cambridge University Press, New York, 1934)
M.B. Mandelbrot, A.J. Paullay, Fractal nature of fracture surfaces of metals. Nature **308**(19), 721–722 (1984)
J. Mandel, P. Dantu, Contribution à l'étude théorique et expérimentale du coefficient délasticité d'un milieu hétérogénes mais statisquement homog ène, Annales des Ponts et Chaussées Paris **6**, 115–145 (1963)
T.J. Napier-Munn, S. Morrell, R.D. Morrison, T. Kojovic, *Mineral Comminution Circuits – Their Operation and Optimisation* (Julius Kruttschnitt Mineral Research, The University of Queensland, Indooroopilly, 1999)

M. Ostoja-Starzewski, Lattice models in micromechanics. Appl. Mech. Rev. **55**(1), 35–60 (2002a)
M. Ostoja-Starzewski, Microstructural randomness versus representative volume element in thermomechanics. ASME J. Appl. Mech. **69**, 25–35 (2002b)
M. Ostoja-Starzewski, *Microstructural Randomness and Scaling in Mechanics of Materials* (Chapman & Hall/CRC Modern Mechanics and Mathematics Series, Boca Raton, 2008)
M. Ostoja-Starzewski, J.D. Lee, Damage maps of disordered composites: a spring network approach. Int. J. Fract. **75**, R51–R57 (1996)
M. Ostoja-Starzewski, G. Wang, Particle modeling of random crack patterns in epoxy plates. Probab. Eng. Mech. **21**(3), 267–275 (2006)
A. Rinaldi, Statistical model with two order parameters for ductile and soft fiber bundles on nanoscience and biomaterials. Phys. Rev. E **83**, 046126-1-10 (2011)
A. Rinaldi, Bottom-up modeling of damage in heterogeneous quasi-brittle solids. Contin. Mech. Thermodyn. **25**(2–4), 359–373 (2013)
A. Rinaldi, D. Krajcinovic, P. Peralta, Y.C. Lai, Lattice models of polycrystalline microstructures: a quantitative approach. Mech. Mater. **40**, 17–36 (2008)
K.A. Snyder, E.J. Garboczi, A.R. Day, The elastic moduli of simple two-dimensional composites: Computer simulation and eective medium theory. J. Appl. Phys.**72**, 5948–5955 (1992)
K. Sab, Principe de Hill et homogénéisation des matériaux aléatoires, C.R. Acad. Sci. Paris II. **312**, 1–5 (1991)
K. Sab, On the homogenization and the simulation of random materials. Europ. J. Mech., A Solids **11**, 585–607 (1992)
P. Trovalusci, D. Capecchi, G. Ruta, Genesis of the multiscale approach for materials with microstructure. Arch. Appl. Mech. **79**(11), 981–997 (2009)
O. Vinogradov, A static analog of molecular dynamics method for crystals. Int. J. Comput. Methods **3**(2), 153–161 (2006)
O. Vinogradov, Vacancy diffusion and irreversibility of deformations in the Lennard–Jones crystal. Comput. Mater. Sci. **45**, 849–854 (2009)
O. Vinogradov, On reliability of molecular statics simulations of plasticity in crystals. Comput. Mater. Sci. **50**, 771–775 (2010)
G. Wang, M. Ostoja-Starzewski, Particle modeling of dynamic fragmentation – I: theoretical considerations. Comput. Mater. Sci. **33**(4), 429–442 (2005)
G. Wang, M. Ostoja-Starzewski, P.M. Radziszewski, M. Ourriban, Particle modeling of dynamic fragmentation - II: Fracture in single- and multi-phase materials. Comp. Mat. Sci. **35**(2), 116–133 (2006)
M.P. Wnuk, *Introducing Fractals to Mechanics of Fracture. Basic Concepts in Fractal Fracture Mechanics*. Handbook of Damage Mechanics, Springer, New York (2014a)
M.P. Wnuk, *Introducing Fractals to Mechanics of Fracture. Toughening and Instability Phenomena in Fracture. Smooth and Rough Cracks*. Handbook of Damage Mechanics, Springer, New York (2014b)
H. Yserentant, A new class of particle methods. Numer. Math. **76**, 87–109 (1997)
P. Zhang, Y. Huang, H. Gao, K.C. Hwang, Fracture nucleation in single-wall carbon nanotubes under tension: a continuum analysis incorporating interatomic potentials. ASME J. Appl. Mech. **69**, 454–458 (2002)

Toughening and Instability Phenomena in Quantized Fracture Process: Euclidean and Fractal Cracks

8

Michael P. Wnuk

Contents

Abstract

Basic concept underlying Griffith's theory of fracture of solids was that, similar to liquids, solids possess surface energy and, in order to propagate a crack by increasing its surface area, the corresponding surface energy must be compensated through the externally added or internally released energy. This assumption works well for brittle solids, but is not sufficient for quasi-brittle and ductile solids.

Here some new forms of energy components must be incorporated into the energy balance equation, from which the input of energy needed to propagate the crack and subsequently the stress at the onset of fracture can be determined. The additional energy that significantly dominates over the surface energy is the irreversible energy dissipated by the way of the plastic strains that precede the leading edge of a moving crack. For stationary cracks, the additional terms within the energy balance equation were introduced by Irwin and Orowan.

M.P. Wnuk (✉)
Department of Civil Engineering and Mechanics, College of Engineering and Applied Science, University of Wisconsin, Milwaukee, WI, USA
e-mail: mpw@csd.uwm.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_19

An extension of these concepts is found in the experimental work of Panin, who showed that the irreversible deformation is primarily confined to the pre-fracture zones associated with a stationary or a slowly growing crack.

The present study is based on the structured cohesive crack model equipped with the "unit step growth" or "fracture quantum." This model is capable to encompass all the essential issues such as stability of subcritical cracks, quantization of the fracture process, and fractal geometry of crack surfaces and incorporate them into one consistent theoretical representation.

Introduction

Inspiration for writing this paper was provided by the experimental work of Panin and his group (Panin 1995) relevant to the better understanding of the phenomenon of pre-fracture strain accumulation, concentration, and redistribution, which occurs within the small pre-fracture zone adjacent to the leading edge of crack and being of paramount importance in determining the early stages of fracture, point of fracture onset – followed at first by the stable crack growth – and then by a terminal instability, which when the positive stress intensity factor K-gradient is maintained leads to a catastrophic propagation.

In order to be able to construct a mathematical model of these nonlinear deformation and fracture processes, it is necessary to introduce the "quantized model" of fracture or QFM for "quantized fracture mechanics." One shall be working here with a structured cohesive model of crack equipped with the "unit growth step" or, equivalently, the Neuber's particle (Neuber 1958) or Novozhilov's "fracture quantum" (Novozhilov 1969). Such an approach represents a substantial departure from the classic theory of Griffith who predicts no subcritical cracks. Griffith crack is either stationary or catastrophic under the positive K-gradient loading configuration. What visibly is missing in the classic theory is the transition period from a stationary state to a moving crack, which is accomplished by the insertion of the period of slow stable crack growth (SCG) and made possible by accounting for the highly nonlinear deformation processes preceding fracture.

To this end similar works have been done in recent past by Khezrzadeh et al. (2011) and by Wnuk et al. (2012), but none of these investigations have succeeded in presenting a mathematically complete theory departing from the continuum-based approximations and consistent with the latest trends in the computational fracture mechanics; cf. Prawoto and Tamin (2013). It is noteworthy that the mathematical model of "structured cohesive crack" has been successfully applied to the studies of the effects of specimen geometry and loading configuration on occurrence of instabilities in ductile fracture (cf. Wnuk and Rouzbehani 2005), in modeling the fatigue phenomenon at nanoscale levels; see Wnuk and Rouzbehani (2008).

To follow this line of approach, the quantization of the fracture process is needed, and it was implemented via the δCOD criterion of Wnuk. The notions of Neuber's particle and Novozhilov's fracture quantum are invoked in order to

accomplish the quantization procedure. A prior knowledge of the strain distribution within the Panin zones is required. The governing differential equations of slow stable crack growth based on the Panin's study and on the theoretical model proposed by Wnuk et al. (2012) have been refined. It has been demonstrated that the nature provides certain mechanisms of enhancing or reducing the material resistance to fracture. The first one is related to material ductility and energy dissipation that precedes the final act of decohesion, while the other factor is purely geometrical as it derives from the roughness of the crack surfaces (not accounted for by the Euclidean geometry of a smooth classic crack). The conclusion is that while ductility significantly improves the fracture toughness, the increased roughness of crack surface suppresses the subcritical crack growth and it tends to induce a more brittle-like fracture. This feature is described in our model by the fractal fracture mechanics. The theory developed here is based on certain key equations involving the fractal representation of stationary and growing cracks due to the fundamental research of Wnuk and Yavari (2003), (2009) and Khezrzadeh et al. (2011).

Displacements and Strains Associated with a Discrete Cohesive Crack Model

For many years, one of the primary subjects of V. Panin's research (Panin 1995) were the experimental studies and recording of the strains at micro- and meso-levels as well as observation of their subsequent build up and redistribution occurring within a certain small process zone adjacent to the leading edge of the crack. Here the primary purpose is to construct a simple mathematical model describing such a phenomenon of pre-fracture strain accumulation and concentration within the regions close to the crack front – in what follows referred to as "pre-fracture" or "Panin zones." The intention is to study both the stationary and slowly moving cracks. For this purpose, one shall employ the structured cohesive crack model equipped with a Neuber's particle or fracture quantum Δ. An assumption regarding existence of such particle embedded within the cohesive zone is necessary if the quantization of the fracture process is anticipated and necessary to provide a complete mathematical representation of the fracture process for a ductile (or quasi-brittle) solid. It is noted that the terms "pre-fracture zone," "Panin zone," or "cohesive zone" are to be understood as synonyms.

Material ductility will be one of the parameters of primary concern. For the structured cohesive crack model (see Figs. 1 and 2), the following quantity will be used as a measure of the ductility:

$$\rho = \frac{R_{ini}}{\Delta} = \frac{\varepsilon^f}{\varepsilon_Y} = \frac{\varepsilon_Y + \varepsilon_{pl}^f}{\varepsilon_Y} \tag{1}$$

Here Δ denotes the fracture quantum, while the length of the cohesive zone measured at the onset of fracture (usually occurring in form of slow stable crack

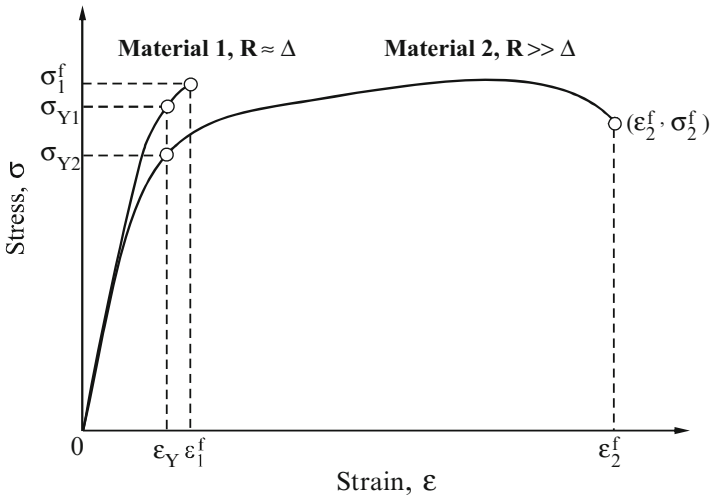


Fig. 1 Examples of quasi-brittle and ductile material behavior. Two materials with identical yield strain ε_Y and similar yield points σ_Y but with widely different ductility are compared: **Material 1** (quasi-brittle) shows the ratio $\varepsilon^f/\varepsilon_Y$ close to one, while for the **Material 2** (ductile) the ratio $\varepsilon^f_2/\varepsilon_Y \gg 1$. In the discrete cohesive crack model, the ductility index R_{ini}/Δ is identified with $\varepsilon^f/\varepsilon_Y$. The length of the cohesive zone at the onset of fracture R_{ini} equals $(\pi/8)(K_c/\sigma_Y)^2$, while Δ is the size of the Neuber's particle

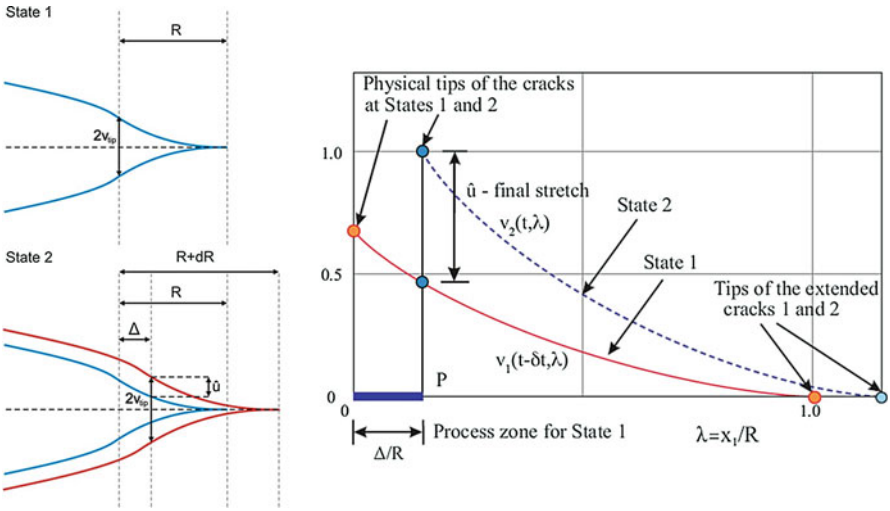


Fig. 2 Distribution of the COD within the cohesive zone corresponding to two subsequent states represented by instants "t- δt" and "t" in the course of quasi-static crack extension as required in Wnuk's criterion of delta COD; $[v_2(t) - v_1(t-\delta t)]_P$ = final stretch

growth that in ductile media precedes the catastrophic propagation) is denoted by R_{ini} and this quantity is related to the yield stress σ_Y, Young modulus E, and the fracture toughness measured by K_c or Rice's integral J_c in a familiar fashion

$$R_{ini} = \frac{\pi}{8}\left(\frac{K_c}{\sigma_Y}\right)^2 = \frac{\pi}{8}\left(\frac{EJ_c}{\sigma_Y}\right)^2 \tag{2}$$

This quantity is often identified with the material characteristic length, say L_{ch}; cf. Taylor (2008). In order to estimate the size of the other important length parameter, the fracture quantum Δ, it suffices at this point to say that in brittle and quasi-brittle materials, Δ and L_{ch} are of the same order of magnitude, while for the ductile materials, Δ is much smaller than the characteristic length given by Eq. 2.

Roughness of the crack surfaces represented via fractal geometry will be treated as a secondary variable that influences the early stages of fracture, i.e., the stable crack extension and the onset of the unstable propagation. For comparison, both the Euclidean and the fractal geometries of a crack will be considered. An approximate model proposed by Wnuk and Yavari (2003) known as the "embedded fractal crack," whose fractal dimension D may differ from one, will be employed; see Fig. 2. Prior to the addition of the cohesive zones, an embedded fractal crack exhibits a singular near tip stress field proportional to $r^{-\alpha}$, where r is the distance measured from the crack tip and the so-called fractal exponent α is related to the dimension D and the roughness measure H as follows:

$$\begin{aligned} \alpha &= \frac{2-D}{2}, 1 \le D \le 2 \\ \alpha &= \frac{2H-1}{2H}, 0 \le H \le 1/2 \end{aligned} \tag{3}$$

Using the fractal crack model, Wnuk and Yavari (2003) and also Khezrzadeh et al. (2011) have estimated the stress intensity factor as

$$K_I^f = \chi(\alpha)\sqrt{\pi a^{2\alpha}}\sigma \tag{4}$$

where σ is the applied stress, a denotes the crack length, and the function $\chi(\alpha)$ is defined by the integral

$$\chi(\alpha) = \frac{2}{\pi^{1-\alpha}}\int_0^1 \frac{dz}{(1-z^2)^{1-\alpha}} = \frac{\pi^{\alpha-1}\Gamma(\alpha)}{\Gamma\left(\alpha+\frac{1}{2}\right)} \tag{5}$$

Here Γ is the Euler gamma function. One notices that for $1 \le D \le 2$ the fractal exponent α varies within the range [0.5, 0]. According to the principle of correspondence, all quantities describing a fractal crack reduce to the classic expressions valid for a smooth crack when $\alpha \to$ ½. As shown by Khezrzadeh et al. (2011), the

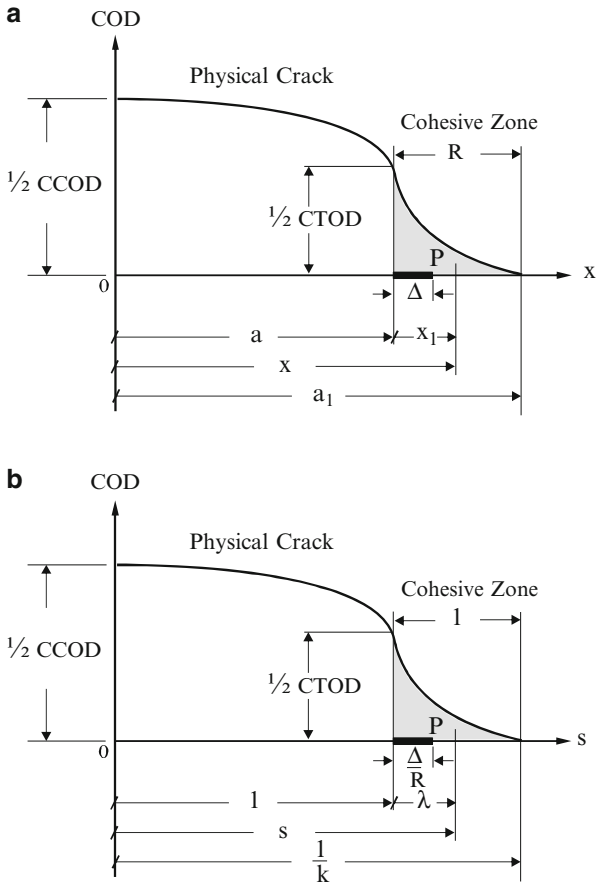


Fig. 3 (**a**) Dimensional coordinates associated with an extended structured cohesive crack. Note the location of the "fracture quantum" Δ, which is adjacent to the crack leading edge and is embedded within the cohesive zone. According to our model, the brittle behavior is observed when Δ and R are of approximately same size, while for the ductile behavior, Δ is deeply embedded within the cohesive zone, and therefore, it is much smaller than R. P denotes the control point for measuring the increment of the COD for a slowly moving crack during the early stages of fracture. Symbols CCOD and CTOD designate the "crack center opening displacement" and the "crack tip opening displacement," respectively. (**b**) Nondimensional coordinates $s = x/a$ and $\lambda = x_1/R$. Note that when these coordinates are used, the tip of the extended crack falls at s = $1/k$, where k is related to the nondimensional loading parameter $Q = (\pi/2)(\sigma/\sigma_Y)$ by this formula: $k = \cos Q$

Wnuk–Yavari model of a fractal crack holds only for cracks with relatively small roughness, and thus in the considerations that follow the range of α will be limited to [0.5, 0.4].

First the case of a smooth Euclidean crack will be represented by a structured cohesive crack model as shown in Fig. 3. Two sets of coordinates are used: the dimensional coordinates are shown in Fig. 3a, while the nondimensional are

explained in Fig. 3b. The distance measured from the origin of coordinates is denoted by x (or $s = x/a$), while the distance measured from the tip of the physical crack is denoted by x_1 (or $\lambda = x_1/R$) and the ratio $a/a_1 = k$. The crack length is "a," the length of the extended crack is $a_1 = a + R$, and the profile of the entire crack is described (cf. Anderson 2004; Khezrzadeh et al. 2011) by the following expression involving the inverse hyperbolic functions

$$u_y = \frac{4\sigma_Y a}{\pi E} \mathrm{Re}\left\{ \coth^{-1}\sqrt{\frac{1-k^2s^2}{1-k^2}} - s\coth^{-1}\left[\frac{1}{s}\sqrt{\frac{1-k^2s^2}{1-k^2}}\right]\right\} \tag{6}$$

At $s = 0$ one obtains the expression for the crack center opening displacement (CCOD), namely,

$$u_y^{\text{center}} = \frac{4\sigma_Y a}{\pi E}\coth^{-1}\left(\frac{1}{1-k^2}\right)^{1/2} = \frac{4\sigma_Y a}{\pi E}\coth^{-1}\left(\frac{1}{\sin Q}\right) \tag{7}$$

The nondimensional loading parameter $Q = \pi\sigma/2\sigma_Y$ enters the last equation due to the known Dugdale formula valid for our model:

$$k = \cos Q \tag{8}$$

At $s = 1$ one obtains the crack tip opening displacement (CTOD), namely,

$$u_y^{tip} = \frac{4\sigma_Y a}{\pi E}\ln\left(\frac{1}{k}\right) = \frac{4\sigma_Y a}{\pi E}\ln\left(\frac{1}{\cos Q}\right) \tag{9}$$

Figure 4a shows the dependence of the tip displacement and the center displacement on the applied load Q, while Fig. 4b illustrates the fact that the ratio of these two quantities CTOD/CCOD remains constant almost throughout the entire range of loading. The constant is 0.504, which provides a good rule of thumb: the tip displacement of the cohesive crack is roughly one half of the mouth displacement measured at the crack center. This observation provides helpful information for an experimentalist, who utilizing various clip gages can access the center of the crack much easier than the tip of the physical crack. Thus, once the CCOD is measured, the tip displacement, the CTOD, can be estimated with a good accuracy.

The profile of the entire extended crack, normalized by the constant $\mathrm{C} = \frac{4\sigma_Y a}{\pi E}$, is shown in Fig. 5a, while Fig. 5b shows the same profile normalized by the CTOD. In constructing these figures, the following equations were used:

$$u = \frac{u_y}{C} = \mathrm{Re}\left\{ \coth^{-1}\sqrt{\frac{1-k^2s^2}{1-k^2}} - s\coth^{-1}\left[\frac{1}{s}\sqrt{\frac{1-k^2s^2}{1-k^2}}\right]\right\} \tag{10}$$

and

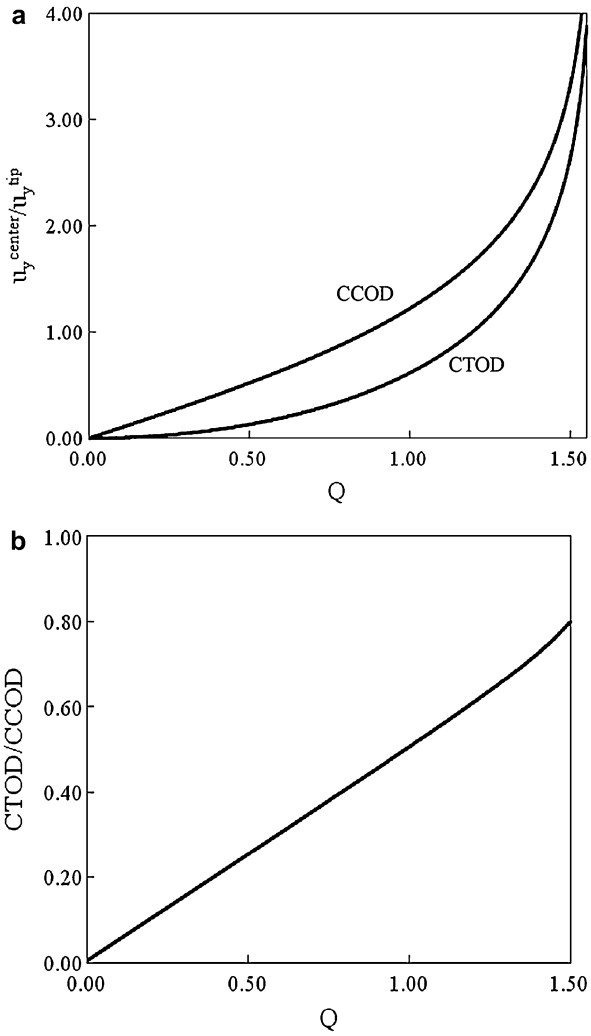


Fig. 4 (**a**) Shows the opening displacements at the center of a cohesive crack and at the tip of the physical crack as functions of the applied load Q. (**b**) Shows the ratio of the CTOD (crack tip opening displacement) to the CCOD (crack center opening displacement). Despite the nonlinear nature of the problem, these results show that the CTOD is roughly one half of the CCOD through the entire range of the loading parameter Q. The ratio depicted in the figure can be very closely approximated by the simple equation CTOD = 0.504CCOD

$$v = \frac{u_y}{u_y^{tip}} = \frac{1}{\ln\left(\frac{1}{k}\right)} \mathrm{Re}\left\{ \coth^{-1}\sqrt{\frac{1-k^2s^2}{1-k^2}} - s\coth^{-1}\left[\frac{1}{s}\sqrt{\frac{1-k^2s^2}{1-k^2}}\right]\right\} \tag{11}$$

Due to this normalization procedure, all v-profiles pass through one at the tip of the physical crack, $x = a$ or $s = 1$. Figure 6a shows the graphs resulting from the expression in Eq. 11 drawn for three values of load Q and plotted within the range of x that corresponds to the "cusp" of the cohesive crack, i.e., for $a \le x \le a_1$ or $1 \le s \le 1/k$.

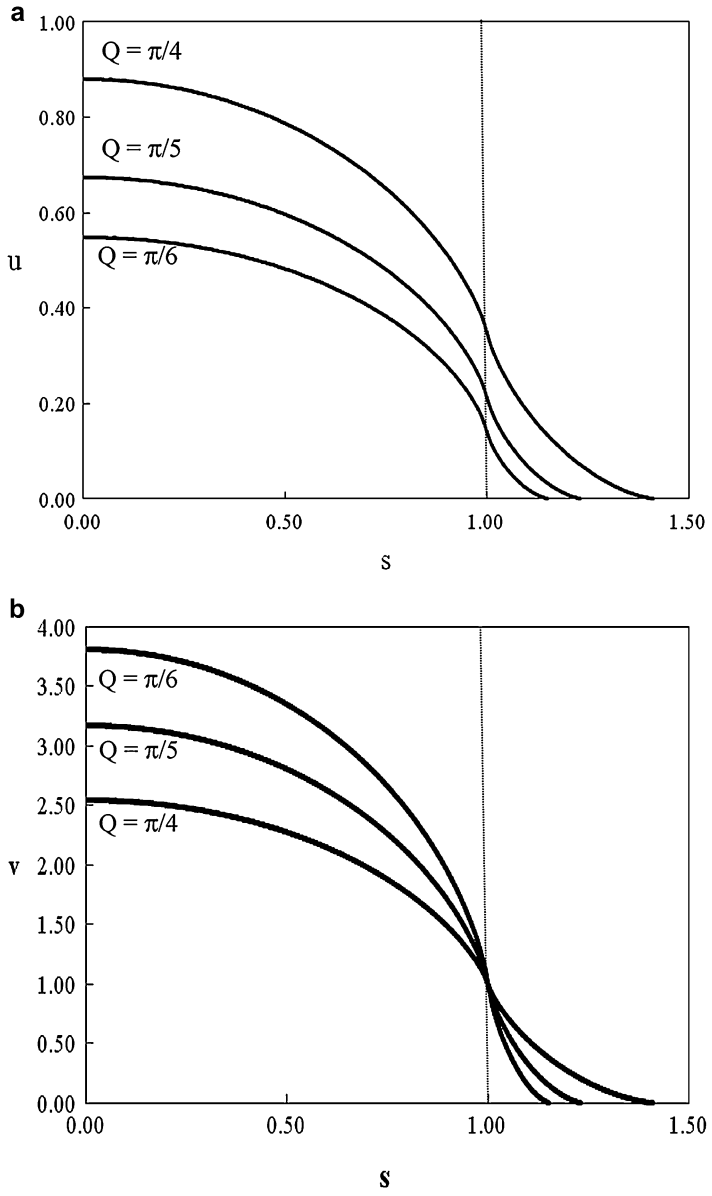


Fig. 5 Profiles of the cusp region of the cohesive crack: (**a**) when the crack opening displacements are normalized by the constant $C = 4\sigma_Y/\pi E$, see Eq. 10, and (**b**) when half of the CTOD is used as the normalization constant, see Eq. 11. Nondimensional loading parameter, proportional to σ/σ_Y, is denoted by Q

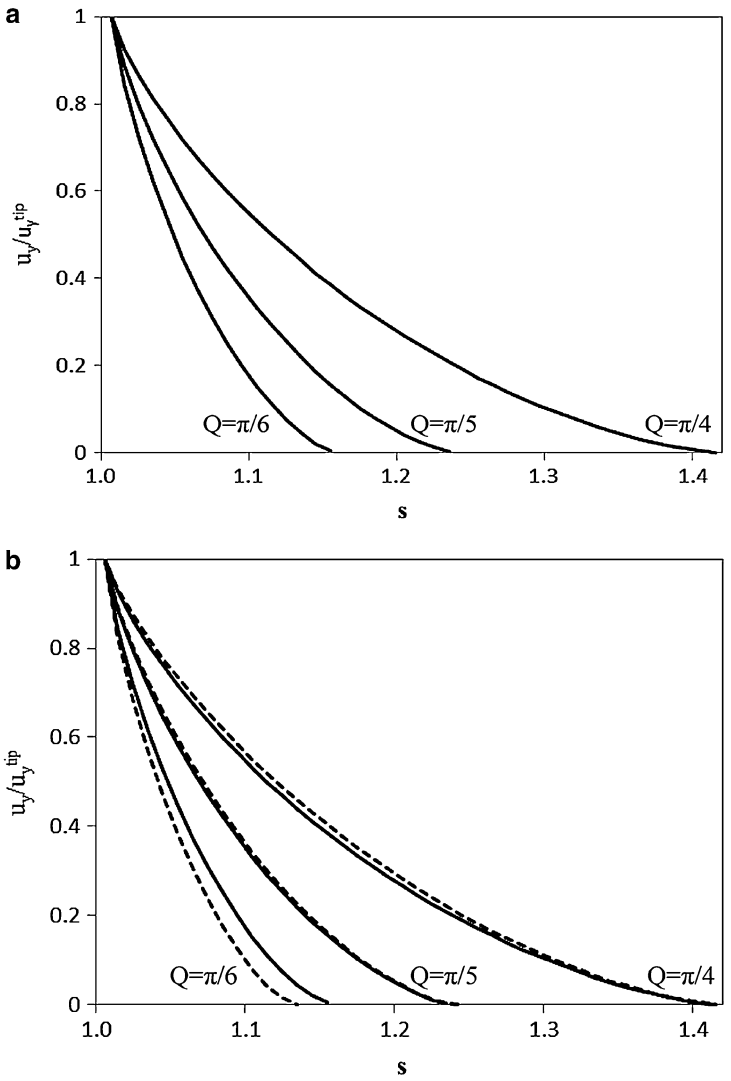


Fig. 6 (**a**) Profiles of the cusp of the cohesive crack plotted for three values of the loading parameter Q according to Eq. 11. (**b**) Comparison of the profiles shown in (**a**) with those that result from simplified formula (*dashed lines*) valid under Barenblatt's restriction of R being much smaller than the crack length; cf. Eq. 14

Finally, in Fig. 6b, these v-graphs representing the cusp are compared to the curves that result from a known (cf. Rice 1968 or Wnuk 1974) approximate formula valid under the Barenblatt's restriction of R being much smaller than the crack length

$$v_{\text{coh}}^{\text{cusp}} = \frac{4\sigma_Y R}{\pi E}\left[\sqrt{1-\frac{x_1}{R}} - \frac{x_1}{2R}\ln\frac{1+\sqrt{1-\frac{x_1}{R}}}{1-\sqrt{1-\frac{x_1}{R}}}\right] \tag{12}$$

It is not difficult to show that for $R \ll a$, the expression for the tip displacement in Eq. 9 reduces to the constant shown in front of the square bracket in Eq. 12, namely,

$$u_y^{\text{tip}} = \frac{4\sigma_Y a}{\pi E}\ln\left(\frac{1}{k}\right) = \frac{4\sigma_Y a}{\pi E}\ln\left(\frac{a_1}{a}\right) = \left[\frac{4\sigma_Y a}{\pi E}\ln\left(\frac{a+R}{a}\right)\right]_{R\ll a} \simeq \frac{4\sigma_Y}{\pi E}R \tag{13}$$

When this constant is used to normalize the displacement in Eq. 12, one obtains

$$v_{\text{coh}} = \frac{v_{\text{coh}}^{\text{cusp}}}{u_y^{\text{tip}}} = \sqrt{1-\frac{x_1}{R}} - \frac{x_1}{R}\ln\frac{1+\sqrt{1-\frac{x_1}{R}}}{1-\sqrt{1-\frac{x_1}{R}}} = \sqrt{1-\lambda} - \frac{\lambda}{2}\ln\frac{1+\sqrt{1-\lambda}}{1-\sqrt{1-\lambda}} \tag{14}$$

Figure 6b shows that the agreement between the exact and the approximate formulae for the cohesive crack opening displacements within the cusp region is indeed good for all values of the loading parameter Q. Therefore, to simplify all further calculations, the formula in Eq. 14 will be employed. Of particular interest will be the strains within the Panin zone, which are defined by the derivative

$$\varepsilon_y^{\text{coh}} = \frac{dv_{\text{coh}}^{\text{cusp}}}{dx_1} = \frac{1}{R}\frac{dv_{\text{coh}}^{\text{cusp}}}{d\lambda} = \frac{1}{R}\frac{4\sigma_Y R}{\pi E}\frac{dv_{\text{coh}}}{d\lambda} = \frac{4\sigma_Y}{\pi E}\frac{dv_{\text{coh}}}{d\lambda} \tag{15}$$

Applying Eq. 14 and carrying out the derivative yields the closed form expression for the strains within the pre-fracture zone. For convenience the strains are expressed in terms of the variable s, related to the variable λ as follows:

$$\begin{aligned} \lambda &= \frac{s-1}{m-1} \\ m &= 1/k \\ k &= \cos Q \end{aligned} \tag{16}$$

Thus, the expression for strains within the pre-fracture zone associated with the structured cohesive crack model reads

$$\begin{aligned} \varepsilon_y^{\text{coh}} &= \frac{4\sigma_Y}{\pi E}\left\{\frac{1}{2}\ln\frac{1-\sqrt{1-\lambda(s,m)}}{1+\sqrt{1-\lambda(s,m)}}\right\} = \frac{4\sigma_Y}{\pi E}\left\{\frac{1}{2}\ln\left[\frac{\sqrt{1-k}-\sqrt{1-ks}}{\sqrt{1-k}+\sqrt{1-ks}}\right]\right\} \\ \lambda(s,m) &= \frac{s-1}{m-1} = \frac{(s-1)\cos Q}{1-\cos Q} \end{aligned} \tag{17}$$

The graphs showing the strains as functions of the loading parameter Q and the coordinate s are shown in Fig. 7.

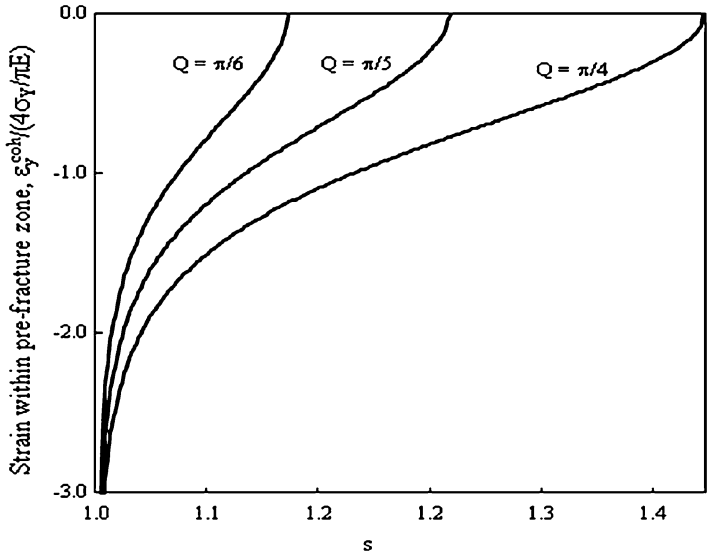


Fig. 7 Strains within the pre-fracture zone obtained as gradient of the crack opening displacement; see Eq. 17

Quantization of the Panin Strain and the Criterion for Subcritical Crack Growth

Closer examination of the expression in Eq. 17 reveals that the strains within the pre-fracture zone are infinite at the tip of the physical crack. Therefore, any use of this entity for the purpose of predicting the onset of fracture propagation will fail, unless it is preceded by the quantization procedure, which in essence is tantamount to evaluation of the strain averaged over the Neuber length Δ, namely,

$$\langle \varepsilon \rangle_{a,\,a+\Delta} = \frac{1}{\Delta}\int_0^{\Delta} \varepsilon_y^{\mathrm{coh}} dx_1 = \frac{1}{\Delta}\int_{\mathrm{state1}}^{\mathrm{state\,2}} \frac{dv_{\mathrm{coh}}^{\mathrm{cusp}}}{dx_1} dx_1 = \frac{1}{\Delta}\left[v_{\mathrm{coh}}^{\mathrm{cusp}}(\mathrm{state\,2}) - v_{\mathrm{coh}}^{\mathrm{cusp}}(\mathrm{state\,1})\right] \tag{18}$$

When this quantity is set equal the average critical strain $\overline{\varepsilon}_{\mathrm{crit}} = \hat{u}/\Delta$, one obtains the following criterion defining the onset of fracture propagation:

$$\begin{aligned} &\langle \varepsilon \rangle_{a,\,a+\Delta} = \overline{\varepsilon}_{\mathrm{crit}} = \hat{u}/\Delta \\ &v_{\mathrm{coh}}^{\mathrm{cusp}}(\mathrm{state\,2}) - v_{\mathrm{coh}}^{\mathrm{cusp}}(\mathrm{state1}) = \hat{u} \end{aligned} \tag{19}$$

Let us define the two neighboring states using the time "*t*" and the time-like variable $x_1(t) = x - a(t)$ for a slowly progressing crack; these states are defined as follows:

$$\begin{array}{l}\text{State 1}, (t-\delta t, x_1=\Delta)\\ \text{State 2}, (t, x_1=0)\end{array} \tag{20}$$

This means that at instant "$t-\delta t$" defining state 1, the front of the advancing crack is a distance Δ away from the control point P (see Fig. 3), while at the instant "t" describing state 2, the tip of the physical crack has reached the control point P. This is indicative that the crack has advanced the "unit growth step" or "fracture quantum" Δ between the two states considered. The constancy of the increment of the crack opening displacement $\hat{u}$ (the so-called final stretch) measured at the control point P constitutes the necessary condition for the stable crack to propagate. In essence this requirement is tantamount to stating Wnuk's criterion of the final stretch or the δ(COD) criterion for subcritical crack extension; cf. Wnuk (1974). It is noteworthy that the physical foundation for the criterion is the same as the one postulated by McClintock (1965), which is the critical strain. As Eqs. 18 and 19 demonstrate, the quantization technique and the attributes of the cohesive crack model allow one to bypass the long expression for pre-fracture strains and to reduce all the essential considerations to the displacements only, namely, the function $v_{\text{coh}}^{\text{cusp}}(\lambda)$ given by Eq. 14. Similar techniques of the "quantized fracture mechanics" (QFM) were employed by Pugno and Ruoff (2004), Taylor et al. (2005), and Wnuk and Yavari (2009).

Using Eq. 12, one may express the opening displacements for both considered states as follows:

$$v_{\text{coh}}^{\text{cusp}}(\text{state 1}) = \frac{4\sigma_Y R(\Delta)}{\pi E}\left[\sqrt{1-\frac{\Delta}{R(\Delta)}} - \frac{\Delta}{2R(\Delta)}\ln\frac{1+\sqrt{1-\dfrac{\Delta}{R(\Delta)}}}{1-\sqrt{1-\dfrac{\Delta}{R(\Delta)}}}\right] \tag{21}$$

and

$$v_{\text{coh}}^{\text{cusp}}(\text{state 2}) = \frac{4\sigma_Y R(0)}{\pi E} = \frac{4\sigma_Y}{\pi E}\left(R(\Delta)+\Delta\frac{dR}{da}\right) \tag{22}$$

Subtracting Eq. 21 from Eq. 22 yields the left hand of the second equation in Eq. 19, which now reads

$$R(\Delta)+\Delta\frac{dR}{da} - R(\Delta)\left[\sqrt{1-\frac{\Delta}{R(\Delta)}} - \frac{\Delta}{2R(\Delta)}\ln\frac{1+\sqrt{1-\dfrac{\Delta}{R(\Delta)}}}{1-\sqrt{1-\dfrac{\Delta}{R(\Delta)}}}\right] = \frac{\hat{u}}{\dfrac{4\sigma_Y}{\pi E}} \tag{23}$$

For $\Delta \ll R$ this expression readily reduces to

$$\frac{dR}{da} = \frac{\pi E}{4\sigma_Y}\left(\frac{\hat{u}}{\Delta}\right) - \frac{1}{2} - \frac{1}{2}\ln\left(\frac{4(R_{ini}/\Delta)R}{R_{ini}}\right) \tag{24}$$

With the notation

$$\begin{aligned} M &= \frac{\pi E}{4\sigma_Y}\left(\frac{\hat{u}}{\Delta}\right) \\ \rho &= R_{ini}/\Delta \end{aligned} \tag{25}$$

this becomes the ordinary differential equation which governs the motion of a stable crack in the early stages of fracture

$$\frac{dR}{da} = M - \frac{1}{2} - \frac{1}{2}\ln\left(\frac{4\rho R}{R_{ini}}\right) \tag{26}$$

Two constants which enter the equation above are (1) the tearing modulus M and (2) the material ductility ρ. For a smooth crack, this is the result of Wnuk (1974) and Rice et al. (1980). In the next section, certain modifications of this equation extending the range of its validity into the fractal geometry domain will be investigated. The rate dR/da reflects the rate of material energy demand; and since R and the integral J differ by just a constant, and $J = -\frac{d\Pi}{2da}$, thus the left-hand side of Eq. 26 also represents the second derivative $-\frac{d^2\Pi}{2da^2}$ where Π denotes the potential energy of the loaded body containing a crack. The "R vs. a" curve defined by the differential in Eq. 26 is often referred to as the material resistance curve. On the other hand, the rate of the energy supply due to external applied stress field is measured by the quantity R hidden in Eq. 8. For the case of $R \ll a$, one may expand both sides of this equation in the corresponding power series:

$$\begin{aligned} \frac{a+R}{a} &= 1 + \frac{R}{a} + \ldots \\ \frac{1}{\cos Q} &= 1 + \frac{Q^2}{2} + \ldots \end{aligned} \tag{27}$$

Setting both expression equal to each other, one obtains for $R \ll a$

$$\begin{aligned} R &= \frac{aQ^2}{2} \\ Q &= \sqrt{\frac{2R}{a}} \end{aligned} \tag{28}$$

The "R" in Eq. 28 represents the rate of energy supplied by the external effort. For the terminal instability to occur, the second derivatives of the energy terms or the rates dR_{MAT}/da and $\partial R_{\mathrm{APPL}}/\partial a$ must equal. The rate dR_{MAT}/da is given by Eq. 26, while differentiation of the first equation in Eq. 28 leads to

$$\left[\frac{\partial R}{\partial a}\right]_{Q=\text{const}} = \frac{Q^2}{2} = \frac{R}{a} \tag{29}$$

This quantity represents the external effort, and thus the conditions for the occurrence of the terminal instability are met when

$$\frac{dR_{\text{MAT}}}{da} = \left[\frac{\partial R_{\text{APPL}}}{\partial a}\right]_{Q=\text{const}} \tag{30}$$

or when

$$M - \frac{1}{2} - \frac{1}{2}\ln\left(\frac{4\rho R}{R_{ini}}\right) = \frac{R}{a} \tag{31}$$

It is noted that R which appears on the left-hand side of Eq. 31 represents the material resistance to an extending crack, so really it should be read as R_{MAT}, while the R shown on the right-hand side of Eq. 31 symbolizes the driving force applied to the crack and truly it should be denoted by R_{APPL}. Since at all points of the stable crack growth including the point of terminal instability described by Eq. 31 both these quantities remain in equilibrium, $R_{\text{MAT}} = R_{\text{APPL}}$; when the subscripts are skipped, the equation defining the critical state can be written in form of Eq. 31. An alternative way to write Eq. 31 is to define the difference between the energy demand and energy supply. A suitable name for such a difference is "stability index" S, namely,

$$S = \frac{dR_{\text{MAT}}}{da} - \left[\frac{\partial R_{\text{APPL}}}{\partial a}\right]_{Q=\text{const}} = M - \frac{1}{2} - \frac{1}{2}\ln\left(\frac{4\rho R}{R_{ini}}\right) - \frac{R}{a} \tag{32}$$

To solve for the parameters characterizing the critical state, i.e., the parameter $R_{c,}$ the critical load Q_c, and the critical crack length a_c, one needs to integrate Eq. 26 and then inspect the results and eventually solve Eq. 31 and/or Eq. 32. This is best done in two steps: first one separates the variables in Eq. 26 obtaining the solution for $R = R(a)$, or $X = X(Y)$, in this implicit form

$$\begin{aligned} a(R) &= a_0 + \int_1^R \frac{dz}{M - \frac{1}{2} - \frac{1}{2}\ln(4\rho z)} \\ X(Y) &= X_0 + \int_1^Y \frac{dz}{M - \frac{1}{2} - \frac{1}{2}\ln(4\rho z)} \end{aligned} \tag{33}$$

The value of the tearing modulus M must be chosen to be somewhat above the value of the minimum modulus $M_{\min}$, below which no stable crack growth may

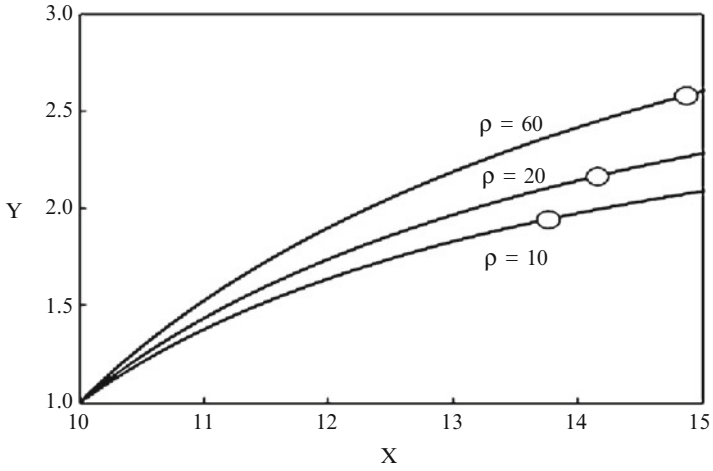


Fig. 8 Apparent material resistance to cracking $Y = R/R_{ini}$ at various levels of material ductility shown as functions of the current crack length during the stable growth process up to the points of terminal instability marked by little circles. All R-curves shown here were obtained from the governing differential in Eq. 26 subject to the initial condition $Y = 1$ at $X_0 = 10$

occur. In this case M is chosen to be 20 % above the minimum value, so the modulus M is determined by this expression

$$M = 1.2\left(\frac{1}{2} + \frac{1}{2}\ln(4\rho)\right) \tag{34}$$

Next one needs to calculate the loading parameter $Q = \sqrt{2Y/X}$ and plot it against the nondimensional crack length $X = a/R_{ini}$. The symbol Y denotes the nondimensional length of the pre-fracture zone R, namely, $Y = R/R_{ini}$, and $X = a/R_{ini}$ denotes the nondimensional length of the crack, while the initial length is $X_0 = a_0/R_{ini}$. Figure 8 shows the R-curves plotted vs. X for $\rho = 10$, 20, and 60 and obtained for the initial crack size of $10R_{ini}$. Figure 9 shows the Q-curves obtained for the same input data. It is noted that the attainment of the maximum on the Q-curve is equivalent to reaching the terminal instability. The point at which the derivative dQ/da approaches zero is best located when the rates of energy demand and energy supply are compared as it is done in Fig. 10a. The intersection of these curves determines the critical state (Q_c, X_c). In addition to these critical parameters, the apparent material fracture toughness encountered at the critical point $Y_c = R_c/R_{ini}$ can readily be evaluated. A convenient way to determine these intersection points numerically is to inspect the stability indices graphs shown in Fig. 10b. The critical states for $\rho = 10$, 20, and 60 and $X_0 = 10$ were established as follows:

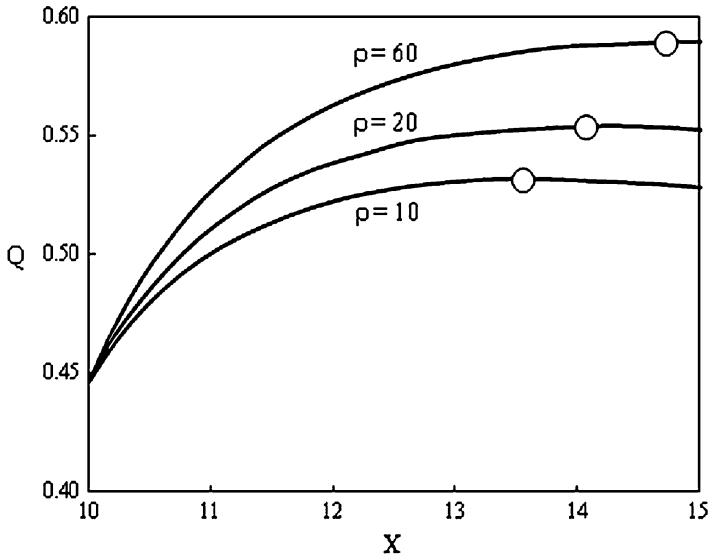


Fig. 9 Loading parameter Q shown during the stable crack growth phase for various material ductility indices. The functions shown pass through the maxima denoted by little *circles*. These points define the critical states (X_c, Q_c)

	Critical values		
Critical parameter	$\rho = 10$	$\rho = 20$	$\rho = 60$
Fracture toughness, Y_c	1.925	2.159	2.581
Crack length, X_c	13.605	14.086	14.842
Load, Q_c	0.532	0.554	0.590

It is readily seen that when the Q_c values are compared with the load prevailing at the onset of fracture, $Q_{ini} = \sqrt{2/X_0}$, one comes to a conclusion that for each case represented in the Table shown above, the loading parameter is enhanced during the process of slow crack growth and the percentage increases of the load are as follows: 19 % for $\rho = 10$, 24 % for $\rho = 20$, and 32 % for $\rho = 60$. These are significant numbers.

Stability of Fractal Cracks

In this section, attention will be focused on the cusp region of the cohesive crack. Following Khezrzadeh et al. (2011), the crack opening displacements at the center of the crack and at the crack tip will be redefined to accommodate the fractal geometry. First let us define four auxiliary functions:

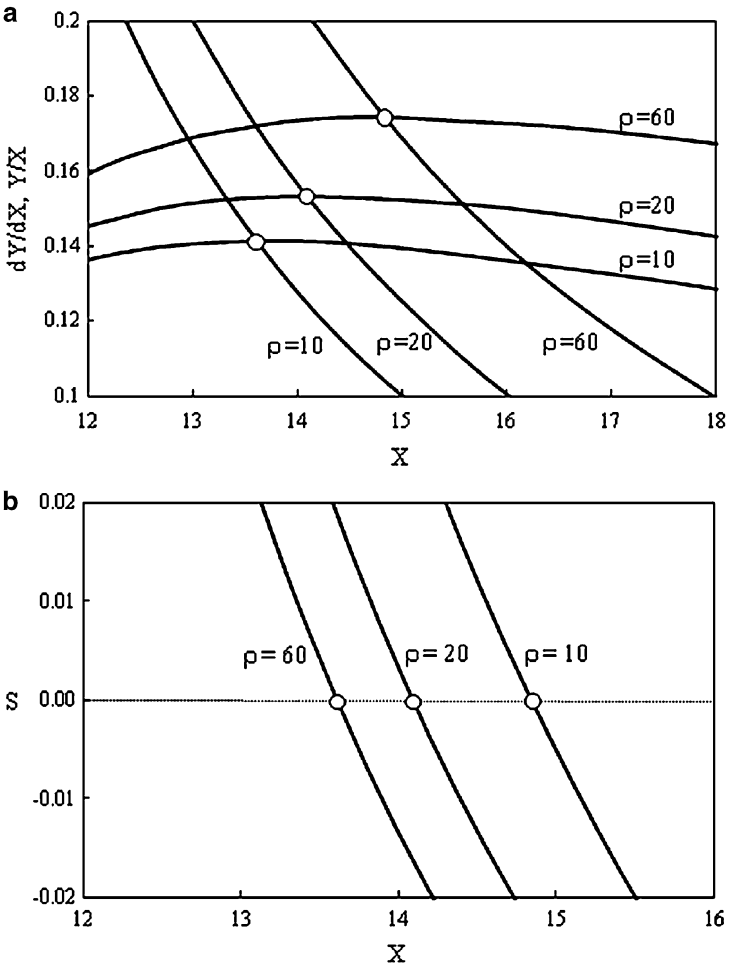


Fig. 10 (**a**) Curves representing the rates of the energy demand (nearly *straight lines*) and the energy supply intersect each other at the points defining the terminal instability (critical states). (**b**) Stability indices shown as functions of the crack length. Zeros of these functions determine the critical states attained at the end of the stable crack growth. Both (**a**) and (**b**) were obtained for the initial crack of length $X_0 = 10$

$$p(\alpha) = 4\pi^{\left(\frac{1}{2\alpha}-2\right)}\left[\frac{\alpha\Gamma(\alpha)}{\Gamma\left(\frac{1}{2}+\alpha\right)}\right]^{\frac{1}{\alpha}} \tag{35}$$

and

$$\kappa(\alpha) = \frac{1+(\alpha-1)\sin(\pi\alpha)}{2\alpha(1-\alpha)} \tag{36}$$

and

$$N(\alpha, X) = p(\alpha)\left[\frac{2}{\pi}\sqrt{\frac{2R}{a}}\right]^{\frac{1-2\alpha}{\alpha}} \tag{37}$$

$$Y_f = R_f/R_{ini} = \frac{R}{R_{ini}}N(\alpha, X) = p(\alpha)Y\left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}} \tag{38}$$

The subscript "f" designates the entities pertinent to the fractal geometry of the crack. The present considerations will be limited to the $R \ll a$ range, and rough cracks described by the fractality parameters such as fractal dimension D, fractal exponent α, and the roughness measure H will be considered. These roughness parameters are related as defined by Eq. 3.

Since the limitations of the Wnuk–Yavari "embedded crack" representation of a fractal crack need to be accounted for, only the limited range of the fractal exponent will be considered, namely, α will be contained within the interval [0.5, 0.4] and it will not fall below 0.4. When this notation is applied, one can cast the results of Khezrzadeh et al. (2011) in the following form:

$$\begin{aligned} u_f^{\text{tip}} &= \kappa(\alpha)\frac{4\sigma_Y}{\pi E}R = \kappa(\alpha)u_y^{\text{tip}} \\ R_f &= N(\alpha, X)R = 4\pi^{\left(\frac{1}{2\alpha}-2\right)}\left[\frac{\alpha\Gamma(\alpha)}{\Gamma\left(\frac{1}{2}+\alpha\right)}\right]^{\frac{1}{\alpha}}\left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}}R \end{aligned} \tag{39}$$

Upon inspection of the latter expression, it is seen that before the length R_f can be determined (and before the profiles of a fractal crack can be sketched), a prior knowledge of the resistance curve $Y_f(X)$ is necessary. Therefore, one must at first establish the differential equation that defines Y_f as a function of the nondimensional crack length X. Let us return to Eq. 23, which in view of the first expression in Eq. 39 has to be rewritten as follows:

$$\kappa(\alpha)\frac{4\sigma_Y}{\pi E}\left\{R_f(\Delta) + \Delta\frac{dR_f}{da} - R_f(\Delta)\left[\sqrt{1-\frac{\Delta}{R_f(\Delta)}} - \frac{\Delta}{2R_f(\Delta)}\ln\frac{1+\sqrt{1-\frac{\Delta}{R_f(\Delta)}}}{1-\sqrt{1-\frac{\Delta}{R_f(\Delta)}}}\right]\right\} = \hat{u} \tag{40}$$

This expression reduces to the ordinary differential equation of the kind similar to Eq. 24. When the ductile behavior of the material ($\Delta \ll R$) is considered, Eq. 40 reads

$$\begin{aligned}\frac{dR_f}{da} &= \frac{1}{\kappa(\alpha)}\frac{\pi E}{4\sigma_Y}\left(\frac{\hat{u}}{\Delta}\right) - \frac{1}{2} - \frac{1}{2}\ln\left(\frac{4(R_{ini}/\Delta)R_f}{R_{ini}}\right)\\ \frac{dY_f}{dX} &= M_f - \frac{1}{2} - \frac{1}{2}\ln\left(4\rho Y_f\right)\\ M_f &= \frac{1}{\kappa(\alpha)}\frac{\pi E}{4\sigma_Y}\left(\frac{\hat{u}}{\Delta}\right) = \frac{M}{\kappa(\alpha)}\end{aligned} \tag{41}$$

The second expression in Eq. 37 defines the function R_f – or its nondimensional equivalent Y_f – namely,

$$\begin{aligned}R_f &= N(\alpha, X)R = p(\alpha)\left[\frac{2}{\pi}\sqrt{\frac{2R}{a}}\right]^{\frac{1-2\alpha}{\alpha}} R\\ Y_f &= N(\alpha, X)Y = p(\alpha)\left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}} Y\end{aligned} \tag{42}$$

Substituting this into Eq. 40 yields

$$\frac{d}{dX}\left[p(\alpha)\left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}} Y\right] = M_f - \frac{1}{2} - \frac{1}{2}\ln\left(4\rho p(\alpha)\left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}} Y\right) \tag{43}$$

Carrying out the differentiation in the left-hand side of this equation gives

$$\frac{d}{dX}\left[p(\alpha)\left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}} Y\right] = p(\alpha)\left\{Y\frac{d}{dX}\left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}} + \left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}}\frac{dY}{dX}\right\} \tag{44}$$

When this expression is substituted back into Eq. 43 and after some simple algebraic manipulations (see the Appendix), one obtains the desired governing differential equation

$$\frac{dY}{dX} = \frac{2\alpha\left[M_f - \frac{1}{2} - \frac{1}{2}\ln\left(4\rho p(\alpha)\left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}} Y\right)\right]}{p(\alpha)\left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}}} + (1-2\alpha)\frac{Y}{X} \tag{45}$$

The fractal tearing modulus M_f will be assumed to be somewhat higher (say by 20 %) than the minimum value of the modulus, at which the stable growth is still possible. The minimum value of the fractal tearing modulus $M^f_{\min}$ is readily established by setting the rate dY/dX equal zero at the point of fracture onset, $X = X_0$ and $Y = 1$, and then evaluating the corresponding modulus. The result is

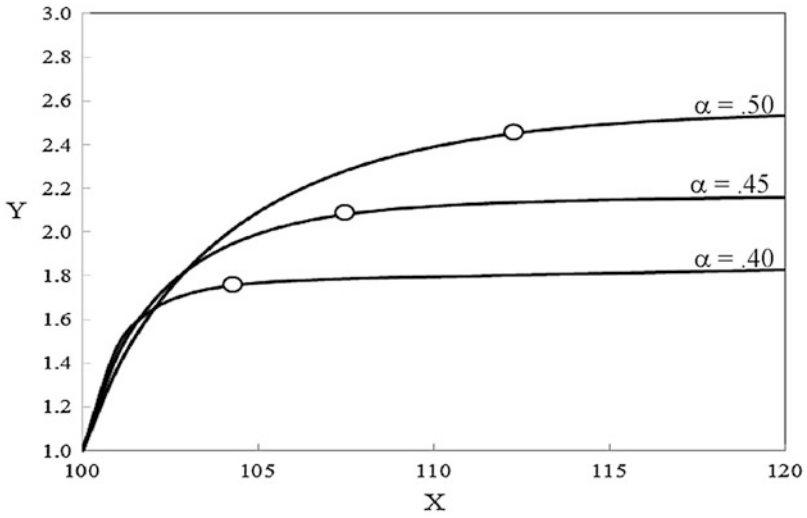


Fig. 11 The resistance curves for stable cracks with the fractal geometry accounted for. The *top curve* corresponds to the case of smooth crack, while the lower curves pertain to the rough cracks with the fractal exponent α designated in the figure. Fractal exponent little circles denote the critical states (Y_c, X_c). Input data $\rho = 10$ and $X_0 = 100$

$$M^f_{\min} = \frac{1}{2} + \frac{1}{2}\ln(4\rho N_0(\alpha, X_0)) - \frac{1-2\alpha}{2\alpha}\frac{N_0(\alpha, X_0)}{X_0}$$
$$N_0(\alpha, X_0) = p(\alpha)\left[\frac{2}{\pi}\sqrt{\frac{2}{X_0}}\right]^{\frac{1-2\alpha}{\alpha}} \tag{46}$$

With the modulus M_f assumed to be $1.2M^f_{\min}$, solutions of Eq. 45 are generated in form $Y = Y(X, \alpha)$, and they are shown in Fig. 11. Three curves shown were drawn for alphas equal to 0.5 (smooth crack) and 0.45 and 0.40, which correspond to the rough cracks of increasing degree of surface roughness. Inspection of Fig. 11 leads to a conclusion that an increased roughness of the crack surface reduces the apparent material fracture toughness attained during the subcritical crack growth. Little circles on the Y-curves in Fig. 11 show the terminal instability points. The location of these points was evaluated by seeking maxima on the Q-curves shown in Fig. 12 or evaluating zeros in the graph representing the stability index Eq. 32 – this has been demonstrated in Fig. 13b. Figure 13a also shows an alternative way of determining the terminal instability points by comparing the rate of energy demand with the rate of energy supplied to the system.

Once the function $Y(X, \alpha)$ has been determined from Eq. 45, one can proceed to evaluate the profiles of the fractal crack within the pre-fracture zone. The equations used for these evaluations read.

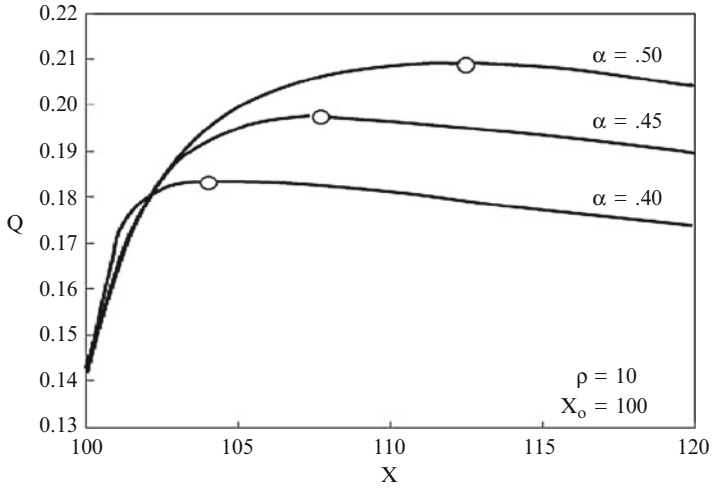


Fig. 12 Loading parameter Q plotted for the rough cracks (two *lower curves*) and for a smooth crack (*top curve*) as a function of the current crack length. It is noted that an increasing roughness of the crack surface reduces the effects of stable crack extension

$$
\begin{aligned}
v_f^{\text{cusp}} &= \frac{4\sigma_Y R}{\pi E} N\kappa(\alpha)\left\{\sqrt{1-\frac{\lambda}{N}} - \frac{\lambda}{2N}\ln\left[\frac{1+\sqrt{1-\frac{\lambda}{N}}}{1-\sqrt{1-\frac{\lambda}{N}}}\right]\right\} \\
N &= N(\alpha, X) = \left[\frac{\alpha\Gamma(\alpha)}{\Gamma\left(\frac{1}{2}+\alpha\right)}\right]^{\frac{1}{\alpha}}\left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}} \\
v_f^{\text{tip}} &= \frac{4\sigma_Y R}{\pi E} N(\alpha, X)\kappa\left(\alpha\right)
\end{aligned}
\tag{47}
$$

Figure 14 shows the profiles of the cusp Eq. 47 normalized by the tip displacement v_f^{tip}. It is now seen that for the enhanced roughness of crack surfaces (diminishing fractal exponent α), the pre-fracture zones diminish and the entire pre-fracture zones shrink. This phenomenon reflects on the earlier attainment of the critical state at the end of stable crack growth phase. To document this fact, all three parameters – apparent fracture toughness Y_c established from the available R-curves, critical nondimensional crack length X_c, and the loading parameter at the terminal instability Q_c – characterizing the critical state (terminal instability) have been grouped in sets (Y_c, X_c, Q_c) and collected in Table 1. The numbers shown in Table 1 were obtained for different initial inputs of the pertinent parameters characterizing a cracked body such as material ductility index ρ, the initial crack length X_0, and the fractal exponent α.

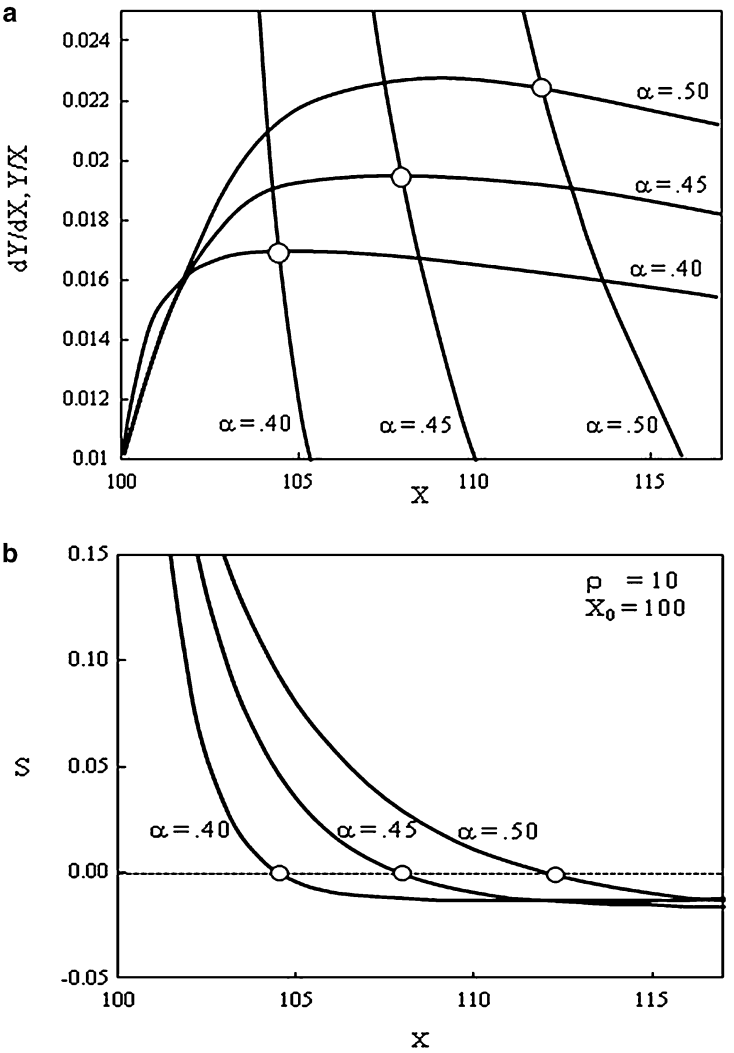


Fig. 13 (**a**) Rates of energy demand (dY/dX) and energy supply (Y/X) for rough cracks, $\alpha = 0.45$ and $\alpha = 0.40$. The curve drawn for $\alpha = 0.5$ corresponds to a smooth crack. Initial crack length is set as $X_0 = 100$, and the material ductility index is $\rho = 10$. (**b**) Stability index S sketched for the same input data as in (**a**). Little *circles* on both graphs indicate the terminal instability points

Conclusions

It has been established that the unstable (catastrophic) fracture propagation in the Griffith sense is almost always preceded by slow stable crack extension that is associated with accumulation and redistribution of strains within the pre-fracture

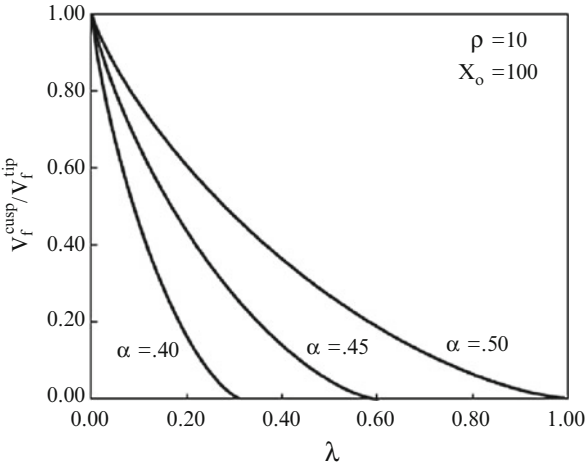


Fig. 14 Profiles of the cusp of the cohesive fractal crack drawn according to Eq. 47. All curves have been normalized by the tip displacement $v_f^{\text{tip}} = N(\alpha, X)\kappa(\alpha)(4\sigma_Y/\pi E)$. An increased crack surface roughness (smaller values of the fractal exponent α) causes the cohesive zone to shrink, which is indicative of a more brittle material behavior

zone adjacent to the front of the propagating crack; cf. Panin (1995). Solutions for advancing cracks significantly differ from those for stationary cracks. Exact solutions address only few loading configurations such as anti-plane mode of loading considered by Hult and McClintock (1956), McClintock (1958), and McClintock and Irwin (1965). By analogy with anti-plane case, the crack advance under a tensile loading has been researched by Krafft et al. (1961), who reformulated the problem and restated it in terms of a universal resistance curve. This view is supported by the studies at the microstructural level of ductile fracture occurring in metals and metallic alloys, where it was found that certain mechanisms exist that facilitate slow crack growth by a sequence of debonding of hard inclusions followed by the formation of voids and their growth and coalescence (Rice 1968). It is noteworthy that due to high strain levels and the redistribution associated with crack motion, the deformation theory of plasticity is not sufficient as a mathematical tool. Perhaps the path-dependent relations between stresses and strains, as those described by the incremental theory of plasticity of Prandtl and Reuss, would be more appropriate to construct a theoretical model based on continuum mechanics.

With exception of Prandtl's slip lines field suggested by Rice et al. (1980), no theory has been proposed that would provide exact mathematical treatment of the problem at hand. Therefore, in this research, one has employed an approximation based on the cohesive crack model equipped with the "unit step growth" or "fracture quantum" combined with an "embedded crack" model of Wnuk and Yavari (2003, 2009), which accounts for a non-Euclidean geometry of a crack represented by a certain fractal. The fractal dimension D for such a crack can vary between 1 (straight line) and 2 (a two-dimensional object). It has been shown that the present "structured cohesive crack model" yields the same essential result as that of Rice et al. (1980), namely, the governing equation which defines the universal R-curve. This statement is true for a smooth crack only. For fractal

geometry, there have been two papers published that address the slow crack advancement, namely, Khezrzadeh et al. (2011) and Wnuk et al. (2012). When the stability of the fractal cracks is reconsidered in the context of the present model, the pertinent results somewhat diverge from the previous findings. Specifically, the relation between the extent of the slow cracking and the fractal exponent α is opposite to what was suggested earlier.

Table 1 Characteristic parameters of the critical states resulting for various input data (ρ, α, X_0)

ρ	α		$X_0 = 3$	$X_0 = 10$	$X_0 = 20$	$X_0 = 60$	$X_0 = 100$	$X_0 = 200$
2	0.50	**Yc**	***Unstable***	1.462	1.619	1.757	1.791	1.820
		Xc		12.407	24.153	67.358	108.998	211.334
		Qc		0.486	0.366	0.228	0.181	0.131
		ΔQ		8.569	15.785	25.109	28.200	31.228
	0.45	**Yc**		1.410	1.515	1.569	1.570	1.559
		Xc		12.096	23.328	65.092	105.801	206.626
		Qc		0.483	0.360	0.220	0.172	0.123
		ΔQ		7.958	13.972	20.255	21.797	22.851
	0.40	**Yc**		1.348	1.399	1.383	1.360	1.326
		Xc		11.734	22.453	63.090	103.207	203.228
		Qc		0.479	0.353	0.209	0.162	0.114
		ΔQ		7.912	11.637	14.676	14.805	14.223
4	0.50	**Yc**	1.088	1.648	1.835	2.006	2.049	2.085
		Xc	3.249	12.945	24.851	68.420	110.269	212.919
		Qc	0.818	0.505	0.384	0.242	0.193	0.140
		ΔQ	0.205	12.848	21.513	32.616	36.315	39.957
	0.45	**Yc**	1.059	1.580	1.704	1.773	1.775	1.765
		Xc	3.170	12.575	23.911	65.880	106.684	207.629
		Qc	0.817	0.501	0.378	0.232	0.182	0.130
		ΔQ	0.051	12.102	19.401	27.066	29.006	30.401
	0.40	**Yc**	1.037	1.501	1.560	1.545	1.52	1.481
		Xc	3.110	12.143	22.912	63.619	103.754	203.782
		Qc	0.817	0.497	0.369	0.220	0.171	0.121
		ΔQ	0.0	11.184	16.705	20.697	21.034	20.579
8	0.50	**Yc**	1.123	1.854	2.076	2.287	2.342	2.389
		Xc	3.591	13.447	25.530	69.503	111.587	214.590
		Qc	0.828	0.525	0.403	0.257	0.205	0.149
		ΔQ	1.403	17.422	27.531	40.516	44.885	49.218
	0.45	**Yc**	1.197	1.768	1.915	2.002	2.008	1.998
		Xc	3.514	13.020	24.478	66.682	107.597	208.681
		Qc	0.825	0.521	0.396	0.245	0.193	0.138
		ΔQ	1.095	16.517	25.095	34.227	36.607	38.391
	0.40	**Yc**	1.172	1.668	1.739	1.725	1.698	1.655
		Xc	3.454	12.521	23.357	64.153	104.312	204.352
		Qc	0.824	0.516	0.386	0.232	0.180	0.127
		ΔQ	0.903	15.419	22.012	27.020	27.587	27.280

(*continued*)

Table 1 (continued)

ρ	α		$X_0 = 3$	$X_0 = 10$	$X_0 = 20$	$X_0 = 60$	$X_0 = 100$	$X_0 = 200$
10	0.50	**Yc**	1.278	1.925	2.160	2.386	2.445	2.496
		Xc	3.692	13.604	25.748	69.859	112.023	215.150
		Qc	0.832	0.532	0.410	0.261	0.209	0.152
		ΔQ	1.908	18.946	29.524	43.142	47.741	52.313
	0.45	**Yc**	1.243	1.832	1.988	2.082	2.089	2.080
		Xc	3.615	13.159	24.660	66.946	107.900	209.036
		Qc	0.829	0.528	0.402	0.249	0.197	0.141
		ΔQ	1.549	17.984	26.978	36.606	39.139	41.059
	0.40	**Yc**	1.216	1.725	1.800	1.787	1.760	1.715
		Xc	3.554	12.640	23.499	64.328	104.497	204.542
		Qc	0.827	0.522	0.391	0.236	0.184	0.130
		ΔQ	1.325	16.823	23.766	29.119	29.766	29.513
100	0.50	**Yc**	1.827	2.800	3.218	3.666	3.794	3.908
		Xc	4.590	15.194	28.060	73.871	117.052	221.752
		Qc	0.892	0.607	0.479	0.315	0.255	0.188
		ΔQ	9.929	35.755	51.460	72.553	80.040	87.743
	0.45	**Yc**	1.762	2.620	2.901	3.107	3.135	3.135
		Xc	4.495	14.575	26.610	69.955	111.430	213.241
		Qc	0.886	0.600	0.467	0.298	0.237	0.171
		ΔQ	8.452	34.085	47.671	63.243	67.736	71.487
	0.40	**Yc**	1.708	2.419	2.561	2.575	2.540	2.479
		Xc	4.402	13.852	25.044	66.349	106.665	206.819
		Qc	0.881	0.591	0.452	0.279	0.218	0.155
		ΔQ	7.887	32.162	43.016	52.600	54.326	54.841
200	0.50	**Yc**	2.017	3.122	3.617	4.163	4.324	4.468
		Xc	4.832	15.678	28.796	75.224	118.784	224.082
		Qc	0.914	0.631	0.501	0.333	0.270	0.200
		ΔQ	11.921	41.123	58.501	82.223	90.788	99.705
	0.45	**Yc**	1.940	2.909	3.243	3.500	3.540	3.546
		Xc	4.728	15.010	27.240	70.984	112.665	214.748
		Qc	0.906	0.623	0.488	0.314	0.251	0.182
		ΔQ	10.964	39.204	54.304	72.008	77.253	81.739
	0.40	**Yc**	1.874	2.671	2.843	2.873	2.837	2.770
		Xc	4.623	14.229	25.551	67.051	107.435	207.646
		Qc	0.900	0.613	0.472	0.293	0.230	0.163
		ΔQ	10.274	37.019	42.186	60.329	62.488	63.334

Perhaps the best way to explain the essential conclusions of this paper is to take a look at Figs. 15 and 16, and also to examine the summary of the results collected in Table 1. From all the pertinent parameters used in the theoretical considerations, one needs to choose just one: the increase in the nondimensional loading parameter ΔQ due to the slow cracking process that precedes the critical point of terminal

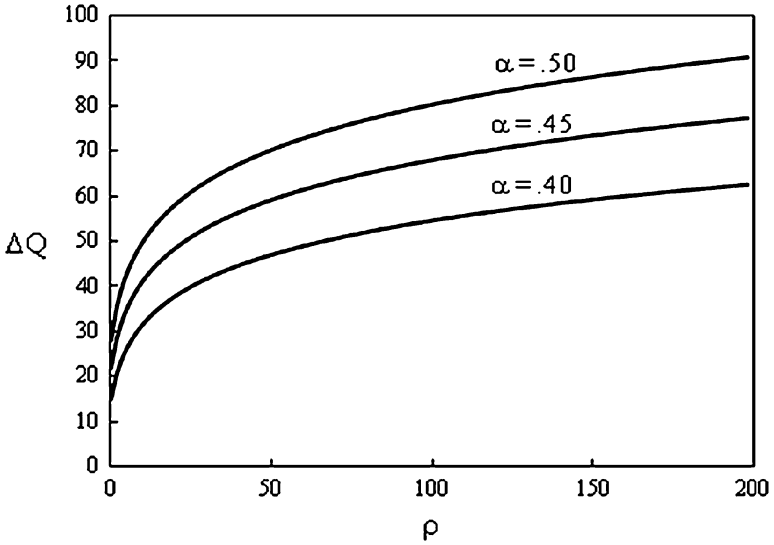


Fig. 15 Enhancement of the load measured at the terminal instability (see Eq. 48) shown as a function of material ductility. Case denoted by $\alpha = 0.5$ corresponds to a smooth crack, while the other curves describe the rough cracks represented by the fractal geometry. The initial crack length for all three curves is $X_0 = 100$

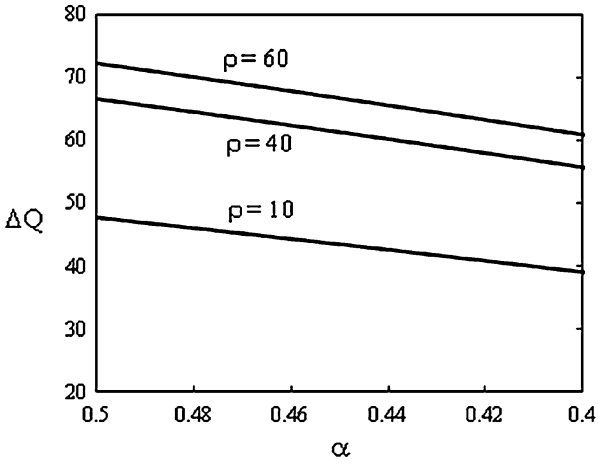


Fig. 16 Load increase attained during the subcritical crack extension at $X_0 = 100$ and various levels of the ductility index shown as a function of the increasing crack surface roughness. It is seen that an increase in roughness causes a reduction of ΔQ

instability. This quantity is represented on the vertical axes in Figs. 15 and 16. It is seen that material ductility (ρ) significantly enhances the slow stable crack growth, leading to an increase in the applied load measured as a difference between the load at the point of catastrophic fracture (Q_c) and the load at the onset of stable crack growth (Q_{ini}):

$$\Delta Q = \frac{Q_c - Q_{ini}}{Q_{ini}}$$
$$Q_{ini} = \sqrt{\frac{2}{X_0}} \tag{48}$$

This observation is entirely in agreement with the previous researches (Khezrzadeh et al. 2011; Wnuk et al. 2012). However, accounting for the fractal geometry leads to an opposite conclusion: higher roughness of crack surface (α less than 0.5) reduces the slow crack growth and results in a decrease of the observed load increase ΔQ; see Fig. 16. In other words, the roughness of the crack surface is conducive to a more brittle material response. Closer examination of the data gathered in Table 1 reveals an interesting phenomenon. It indicates that for a very small crack, $X_0 = 3$, and for low material ductility, $\rho = 2$, the stable growth does not exist at all. It is noted that in this particular case, the initial crack is of the size on the order of magnitude of the characteristic length R_{ini}. For such a small crack, a new effect of "overstressing" comes to light, similar to the phenomenon known in Physics of Fluids as "supercooled" liquid. The effect can be explained as follows: despite the sufficient energy accumulated within the immediate surroundings of the small crack, the crack does not begin to propagate until a certain hypercritical load level is reached. What happens then is a sudden transition from a stationary to a dynamically propagating crack; compare Mott (1948) and Cotterell (1968).

Whether or not the phenomenon of stable crack extension may exist strongly depends on (1) material ductility and (2) tearing modulus of Paris, proportional to the initial slope of the R-curve, and to a much lesser degree on the level of crack surface roughness. If the tearing modulus in the governing Eq. 26 for a smooth crack and Eq. 45 for a rough crack does not meet the condition of being greater than the minimum modulus calculated in sections "Quantization of the Panin Strain and the Criterion for Subcritical Crack Growth" and "Stability of Fractal Cracks," i.e.,

$$\frac{\pi E}{4\sigma_Y}\left(\frac{\hat{u}}{\Delta}\right) \geq M_{\min} = \frac{1}{2} + \frac{1}{2}\ln(4\rho) \tag{49}$$

for a smooth crack, and

$$\frac{1}{\kappa(\alpha)}\frac{\pi E}{4\sigma_Y}\left(\frac{\hat{u}}{\Delta}\right) \geq M^f_{\min} = \frac{1}{2} + \frac{1}{2}\ln(4\rho N_0(\alpha, X_0)) - \frac{1-2\alpha}{2\alpha}\frac{N_0(\alpha, X_0)}{X_0}$$
$$N_0(\alpha, X_0) = p(\alpha)\left[\frac{2}{\pi}\sqrt{\frac{2}{X_0}}\right]^{\frac{1-2\alpha}{\alpha}} \tag{50}$$

for a rough crack, the stable crack growth will vanish.

It is seen that while the tearing modulus for a smooth crack in Eq. 49 depends only on the material property such as the ductility index ρ, the tearing modulus for a fractal crack, as given by Eq. 50, depends also on the purely geometrical parameters such as the measure of the crack surface roughness α and the initial crack length X_0. It is noteworthy that in the case when the conditions stated by the inequalities in

Eqs. 49 and 50 are not met, slow stable crack growth does not exist. Indeed, for a certain combination of the input parameters such as the material ductility, initial crack length, and the fractal exponent, it can be shown that the transition period of slow stable crack growth is missing and one returns to the rules valid for ideally brittle fracture. This may be compared with the experimental data of Alves et al. (2010). Further research of this type is needed to fully understand the Physics behind the present model.

For readers interested in a more advanced (or more detailed) studies of the subject of linear and nonlinear fracture mechanics, the following textbooks are recommended:

1. H. Liebowitz (ed.), 1968, "Fracture. An advanced Treatise", editor H. Liebowitz, Vol. 2: Mathematical fundamentals, Academic Press 1968.
2. D. Broek, 1986, "Elementary engineering fracture mechanics", 4th revised edition, Martinus Nijhoff Publishers, Dordrecht 1986.
3. M. F. Kanninen and C. H. Popelar, 1985, "Advanced fracture mechanics", Oxford University Press, New York and Clarendon Press, Oxford, UK.
4. M. P. Wnuk (ed.), 1990, "Nonlinear fracture mechanics", International Centre for Mechanical Sciences, CISM Course No. 314, Udine 1990, Springer Verlag 1990.
5. T. L. Anderson, 1991, "Fracture mechanics. Fundamentals and Applications", CRC Press 1991.
6. C. T. Sun and Z. H. Jin, 2012, "Fracture mechanics", Elsevier 2012.

Appendix A

Let us recall Eq. 43

$$\frac{d}{dX}\left[p(\alpha)\left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}}Y\right] = M_f - \frac{1}{2} - \frac{1}{2}\ln\left(4\rho p(\alpha)\left[\frac{2}{\pi}\sqrt{\frac{2Y}{X}}\right]^{\frac{1-2\alpha}{\alpha}}Y\right) \tag{51}$$

When the product of $\left(\frac{2}{\pi}\right)^{\frac{1-2\alpha}{\alpha}}$ and $p(\alpha)$ is denoted by $f(\alpha)$, then the left-hand side (LHS) of Eq. 51 can be written as $f(\alpha)\frac{d}{dX}\left[\left(\frac{2Y}{X}\right)^{\frac{1-2\alpha}{2\alpha}}Y\right]$. Now one proceeds with the differentiation

$$LHS = f(\alpha)\left[\left(\frac{2Y}{X}\right)^{\frac{1-2\alpha}{2\alpha}}\frac{dY}{dX} + \left(\frac{1}{2\alpha} - 1\right)Y\left(\frac{2Y}{X}\right)^{\frac{1}{2\alpha}-2} 2\frac{X\frac{dY}{dX} - Y}{X^2}\right] \tag{52}$$

hence

$$LHS = f(\alpha)\left[\left(\frac{2Y}{X}\right)^{\frac{1}{2\alpha}-1}\left(1+\frac{1}{2\alpha}-1\right)\frac{dY}{dX}\right] - f(\alpha)\left(\frac{1}{\alpha}-2\right)\left(\frac{Y}{X}\right)^{2}\left(\frac{2Y}{X}\right)^{\frac{1}{2\alpha}-2} \quad (53)$$

With G denoting the RHS of Eq. 51, one has

$$f(\alpha)\left(\frac{2Y}{X}\right)^{\frac{1}{2\alpha}-1}\left[\frac{1}{2\alpha}\frac{dY}{dX}\right] - f(\alpha)\left(\frac{1}{\alpha}-2\right)\left(\frac{Y}{X}\right)^{2}\left(\frac{2Y}{X}\right)^{\frac{1}{2\alpha}-2} = G(X,Y,\alpha) \quad (54)$$

This reduces to

$$\frac{dY}{dX} = \frac{2\alpha G}{f(\alpha)\left(\frac{2Y}{X}\right)^{\frac{1}{2\alpha}-1}} + 2\alpha\left(\frac{1}{\alpha}-2\right)\left(\frac{Y}{X}\right)^{2}\frac{X}{2Y} \quad (55)$$

With $\left(\frac{2}{\pi}\right)^{\frac{1-2\alpha}{\alpha}} p(\alpha) = f(\alpha)n,$ this equation becomes identical with Eq. 45.

Appendix B

Fracture in an ideally brittle solid (and for the fractal exponent $\alpha = ½$) occurs when the ductility index $\rho = R/\Delta \to 1$. It would be worthwhile to prove that in this case the differential equation governing motion of the subcritical crack in Eq. 23 predicts no stable crack growth and that the δCOD criterion reduces then to the classic case of Griffith. In order to prove this point, let us write the governing equation derived from the δCOD criterion, Eq. 23, in this form:

$$\begin{aligned} \frac{dR}{da} &= M - \frac{R}{\Delta} + \frac{R}{\Delta}F(\Delta/R) \\ M &= \frac{\pi E}{4\sigma_Y}\left(\frac{\hat{u}}{\Delta}\right) \end{aligned} \quad (56)$$

where M is the tearing modulus and the function F is defined as follows:

$$F(\Delta/R) = \sqrt{1-\frac{\Delta}{R}} - \frac{\Delta}{2R}\ln\frac{1+\sqrt{1-\frac{\Delta}{R}}}{1-\sqrt{1-\frac{\Delta}{R}}} \quad (57)$$

For ductile solids, Δ is much smaller than R, and thus $\rho \gg 1$. Under this condition, the function F reduces as follows:

$$F\left(\frac{\Delta}{R}\right)_{\rho\gg 1} = 1 - \frac{\Delta}{2R} + \frac{\Delta}{2R}\ln\left(\frac{\Delta}{4R}\right) \quad (58)$$

This form leads to the differential in Eq. 24 considered in the last section. To obtain the ideally brittle limit, one needs to expand the function F into a power series for ρ approaching one. The results is

$$F\left(\frac{\Delta}{R}\right)_{\rho\to 1} = -\frac{2}{3}\left(1-\frac{\Delta}{R}\right)^{3/2} \tag{59}$$

When this is substituted into Eq. 56, one obtains the differential equation governing an R-curve for quasi-brittle solids, namely,

$$\frac{dR}{da} = M - \frac{R}{\Delta} - \frac{2}{3}\frac{R}{\Delta}\left(1-\frac{\Delta}{R}\right)^{3/2} \tag{60}$$

For the ideally brittle solid, two things happen, first, one has $R = \Delta$ and, second, the slope of the R-curve defined by Eq. 60 equals zero (the R-curve reduces now to a horizontal line drawn at the level $R = R_{ini}$). Therefore, Eq. 60 reduces to

$$\frac{dR}{da} = 0 \quad \text{or,} \quad M = 1 \tag{61}$$

It is also known that for an ideally brittle solid, the size of the Neuber particle Δ can be identified with the length of the cohesive zone

$$\Delta = R = \frac{\pi E}{8\sigma_Y}\text{CTOD} \tag{62}$$

Here symbol CTOD stands for the crack tip opening displacement. The final stretch $\hat{u}$ is now equal half of the CTOD, namely,

$$\hat{u} = \frac{1}{2}\text{CTOD} \tag{63}$$

When Eqs. 62 and 63 are substituted into the definition of the tearing modulus shown in Eq. 56, one gets

$$M = \frac{\pi E}{4\sigma_Y}\left(\frac{\hat{u}}{\Delta}\right) = \frac{\pi E}{4\sigma_Y}\left(\frac{(1/2)\text{CTOD}}{(\pi E/8\sigma_Y)\text{CTOD}}\right) = 1 \tag{64}$$

In this way it is confirmed that the requirement of zero slope of the R-curve in the limiting case of an ideally brittle solid, expressed by Eq. 61, is satisfied when $\hat{u} \equiv (1/2)\text{CTOD}$ and $\Delta \equiv R$. In other words, the δCOD criterion for the onset of fracture reduces to the CTOD criterion of Wells or – equivalently – to the J-integral criterion of Rice. The latter is in full accord with the Irwin driving force criterion $G = G_c$, and this yields the result identical to the ubiquitous Griffith expression for the critical stress

$$\sigma_G = \sqrt{\frac{2E\gamma}{\pi a}} = \sqrt{\frac{G_c E}{\pi a}} = \frac{K_c}{\sqrt{\pi a}} \tag{65}$$

Similar conclusion may be obtained directly from the fact that the R-curve is given as a horizontal line (of zero slope) drawn at the level of R_{ini}. Setting the

equilibrium length of the cohesive zone R equal to its critical value R_c leads to Eq. 65, as expected. To complete this consideration, one is reminded that the quantities R and K_I are related as follows:

$$R = \frac{\pi}{8}\left(\frac{K_I}{\sigma_Y}\right)^2 = R_c = \frac{\pi}{8}\left(\frac{K_c}{\sigma_Y}\right)^2 \quad \text{or,} \quad K_I = K_c \quad \text{or,} \quad \sigma_{\text{crit}} = \sigma_G \tag{66}$$

Therefore, it has been demonstrated that the nonlinear theory described in the preceding sections encompasses the classic theory of fracture, which becomes now a special case of a more general mathematical representation.

References

L.M. Alves, R.V. Da Silva, L.A. Lacerda, Fractal model of the J-R curve and the influence of the rugged crack growth on the stable elastic–plastic fracture mechanics. Eng. Fract. Mech. **77**, 2451–2466 (2010)

T.L. Anderson, *Fracture Mechanics: Fundamentals and Applications*, 2nd edn. (CRC Press, Boca Raton, 2004)

B. Cotterell, Fracture propagation in organics glasses. Int. J. Fract. Mech. **4**(3), 209–217 (1968)

J.A. Hult, F.A. McClintock, in *Proceedings of the 9th International Congress of Applied Mechanics*, vol. 8, (Brussels, 1956), pp. 51–58

H. Khezrzadeh, M.P. Wnuk, A. Yavari, Influence of material ductility and crack surface roughness on fracture instability. J. Phys. D Appl. Phys. **44**, 395302 (2011) (22 pp)

J.M. Krafft, A.M. Sullivan, R.W. Boyle, Effect of dimensions on fast fracture instability of notched sheets. in *Proceedings of the Crack Propagation Symposium* (Cranfield College of Aeronautics, Cranfield, 1961)

F.A. McClintock, J. Appl. Mech. **58**, 582 (1958)

F.A. McClintock, Effect of root radius, stress, crack growth, and rate on fracture instability. Proc. R. Soc. Lond. Ser. A **285**, 58–72 (1965)

F.A. McClintock, G.R. Irwin, in *Fracture Toughness Testing and Its Applications*, ASTM STP 381, (ASTM, Philadelphia, 1965), pp. 84–113

N.F. Mott, Brittle fracture in mild steel plates, Part II. Engineer **165**, 16–18 (1948)

H. Neuber, *Theory of Notch Stresses* (Springer, Berlin, 1958)

V.V. Novozhilov, On a necessary and sufficient criterion for brittle strength. J. Appl. Mech. USSR **33**, 212–222 (1969)

V.E. Panin, *Physical Mesomechanics and Computer-Aided Design of Materials*, vols. 1 and 2, (in Russian) (Nauka, Novosibirsk, 1995)

Y. Prawoto, M.N. Tamin, A new direction in computational fracture mechanics in materials science: will the combination of probabilistic and fractal fracture mechanics become mainstream? Comput. Mater. Sci. **69**, 197–203 (2013)

N. Pugno, R.S. Ruoff, Quantized fracture mechanics. Phil. Mag. **84**(27), 2829–2845 (2004)

J.R. Rice, Mathematical Analysis in the Mechanics of Fracture, in *Fracture, An Advanced Treatise*, ed. by H. Liebowitz, vol. 2 (Academic Press, New York, 1968)

J.R. Rice, W.J. Drugan, T.L. Sham, Elastic–plastic analysis of growing cracks. in *Fracture Mechanics, 12th Conference*, ASTM STP 700 (ASTM, Philadelphia, 1980)

D. Taylor, The theory of critical distances. Eng. Fract. Mech. **75**, 1696–1705 (2008)

D. Taylor, P. Cornetti, N. Pugno, The fracture mechanics of finite crack extension. Eng. Fract. Mech. **72**, 1021–1038 (2005)

M.P. Wnuk, Quasi-static extension of a tensile crack contained in a viscoelastic-plastic solid. J. Appl. Mech. **41**, 234–242 (1974)

M.P. Wnuk, A. Rouzbehani, Instabilities in early stages of ductile fracture. Phys. Mesomech. **8**(5–6), 81–92 (2005)

M.P. Wnuk, A. Rouzbehani, A mesomechanics model of fatigue crack growth for nanoengineering applications. Phys. Mesomech. **11**(5–6), 272–284 (2008)

M.P. Wnuk, A. Yavari, On estimating stress intensity factors and modulus of cohesion for fractal cracks. Eng. Fract. Mech. **70**, 1659–1674 (2003)

M.P. Wnuk, A. Yavari, A discrete cohesive model for fractal cracks. Eng. Fract. Mech. **76**, 545–559 (2009)

M.P. Wnuk, M. Alavi, A. Rouzbehani, Comparison of time dependent fracture in viscoelastic and ductile solids. Phys. Mesomech. **15**(1–2), 13–25 (2012)

Two-Dimensional Discrete Damage Models: Discrete Element Methods, Particle Models, and Fractal Theories

9

Sreten Mastilovic and Antonio Rinaldi

Contents

Abstract

Discrete element methods (DEM) reviewed in this essay are limited to discontinuous models comprised of two-dimensional (2D) basic constitutive units such as circles, ellipses, or polygons given geometrical, structural, and contact

S. Mastilovic (✉)
Faculty of Construction Management, Union–Nikola Tesla University, Belgrade, Serbia
e-mail: smastilovic@fgm.edu.rs

A. Rinaldi
Materials Technical Unit, ENEA, C.R. Casaccia, Rome, Italy

Center for Mathematics and Mechanics of Complex Systems (MEMOCS), University of L'Aquila, Cisterna di Latina (LT), Italy
e-mail: antonio.rinaldi@enea.it

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_23

properties that allow their assemblies to approximate phenomenological response of abstracted materials. The contacts are endowed with energy dissipation mechanisms and cohesive strength, which enables representation of damage-evolution phenomena such as crack nucleation and growth. By the way of dynamic interactions among particles, DEM are capable of coping with the complexity of fracture events in a simple and natural manner. On the other hand, the particle models are, in principle, offshoots of molecular dynamics (MD) adapted to simulation of materials at coarser scales. The role of atom is taken over by a continuum particle or quasi-particle, a basic constitutive unit that can represent, for example, a grain of ceramics, a concrete aggregate, and a particle of a composite. The computation domain is discretized into regular or random network of such particles generally interacting through nonlinear potentials within the realm of Newtonian dynamics. The model parameters should be identifiable with the macroscopic elastic, inelastic, and fracture properties of the material they aspire to represent, and the model should be structured in accordance with its morphology. Finally, a succinct survey of the percolation theory and fractal scaling laws of damage in discrete models is offered.

Introduction

DEM refer to a Lagrangian simulation technique in which the computational domain consists of discrete rigid or deformable elements interacting by the way of contact algorithms. The material is viewed as a Voronoi assembly of discrete particles, representative of the material heterogeneity, with macroscopic continuum behavior determined by dynamic interparticle interactions. The contact algorithms are at the core of a set of numerical techniques designed to solve the applied mechanics problems characterized by strong discontinuities in material or geometric behavior. Consequently, DEM models are most naturally applied to media that have the same topology as the underlying particle network. Recent DEM surveys are presented by Bicanic (2004), Donzé et al. (2008), and Munjiza et al. (2011).

DEM was introduced by Cundall (1971) for the analysis of the discontinuous progressive failure of rock slopes and then implemented to soils by Cundall and Strack (1979). Cundall and Hart (1992) described it as a method that allows finite displacements and rotations of discrete solid macroscopic bodies, which can overlap and detach and recognize new contacts automatically as the simulation progresses. Thus, the mechanical behavior of the abstracted medium results from collective effect of movements of the individual particles within conglomerate, which is governed by the specified interparticle force–displacement law or contact forces. DEM provides detailed time evolution of particle systems by solving Newton's equations of motion of individual particles, which includes complex damage mechanisms that naturally emanate during simulations.

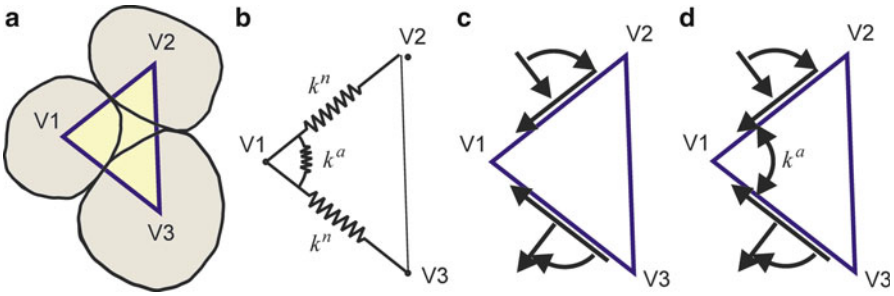


Fig. 1 A cluster of three grains with mutual interaction lines (**a**) and corresponding load transfer mechanisms based on (**b**) central (k^n) and angular (k^a) interactions; (**c**) central, shear, and bending interactions (a typical DEM that can be termed as "locally inhomogeneous micropolar continuum" (Ostoja-Starzewski 2007) with inhomogeneity varying on the grain scale); (**d**) central, shear, bending, and angular interactions (Redrawn from Ostoja-Starzewski 2007)

It is convenient to classify DEM with respect to the load transfer mechanism into the models with:

- Central interactions (generalization of the α models briefly overviewed in the preceding essay of this handbook; Fig. 1b with $k^a \equiv 0$)
- Central and angular interactions (generalization of the α–β models; Fig. 1b)
- Central, shear, and bending interactions (generalization of the beam interaction models; Fig. 1c)
- Central, shear, bending, and angular interactions (Fig. 1d)

The DEM computer-implementation techniques utilize an explicit finite difference scheme and alternate in each calculation cycle between application of Newton's second law of motion and conditions of moment equilibrium and a force–displacement contact law at each interparticle contact. New contacts can be made in the course of deformation and – in the case of cohesive materials – some contacts between particles can break. Because of this, the global stiffness matrix of the complete particle assembly has to be updated constantly. The process of developing a continuum description from the microscopic state parameters of individual elements comprising an assembly is called homogenization. The initial point of that approach is the introduction of representative volume elements that serve as averaging volumes for the macroscopic quantities.

DEM Implementation for Non-cohesive Materials

The difference between cohesive and non-cohesive particular materials is based on existence or nonexistence of normal-direction tension carrying capability. The entire behavior of non-cohesive materials can be described as a multi-body contact

problem, which makes them naturally suitable for DEM. The numerical techniques for modeling the flow of non-cohesive materials are based on the pioneering work of Cundall and Strack (1979). Since granular materials are large conglomerations of individual macroscopic particles that can translate and rotate, this model was based on the basic elements of these materials – granules themselves – and their mutual interactions. Three main aspects of DEM are as follows: (i) particle shape and particle size distribution (physical parameters), (ii) interparticle contact constitutive behavior (mechanical parameters, e.g., contact friction coefficient, contact normal stiffness, contact tensile strength), and (iii) numerical technique for solving equations of motion. A slip feature of the DEM model accounts for a limited shear resistance that the contact, bounded by a friction law, can offer before sliding.

As granular systems evolve, collision, sliding, and rolling contacts produce forces and torques that DEM aspires to evaluate. The DEM model building blocks are randomly generated circles (Cundall and Strack 1979), ellipses (Ting 1992), convex polygons (Cundall 1988), or clusters (Jensen et al. 1999) whose size distributions (the lognormal is a frequent choice) reflect inherent heterogeneity of the system. The circular particles are the simplest since a single parameter defines the particle geometry and there is only one possible type of contact, easy to detect. They are often chosen due to their simplicity; however, they underestimate rolling resistance and cannot capture particle interlock. Consequently, circular granules fail earlier than granules of more complex shape.

The particles are frequently assumed to be rigid, but overlap is permitted in order for relative displacement to occur (soft or smooth contact). (The contact dynamics methods based on "non-smooth" formulation, which exclude possibility of overlaps among particles, are not reviewed herein; the interested reader is referred to Donze et al. 2008.) The assumption of particle rigidity is reasonable when movements along interfaces account for most of the deformation in a particle assembly, which is typically the case for a non-cohesive assembly such as sand.

The following abbreviated overview of the DEM slip model with rolling friction is largely after Xiang et al. (2009).

DEM Slip Model with Rolling Friction

Dry granular media are characterized by dominance of short-range, non-cohesive intergranular interactions: elastic or inelastic contact forces and friction at contact between particles. Thus, each interparticle contact can be rheologically represented by an axial (normal-direction) spring–dashpot element (Kelvin element) and a spring–dashpot–slider element in the tangential direction (Fig. 2d). The contact behavior in normal direction is elastic with no tension limit (tensile strength). A generalized interparticle contact behavior of the simple central and angular interaction type accounts for the normal interaction, shear interaction, and slip.

In general, forces acting on particles i include the following: the gravitational force ($m_i \boldsymbol{g}$) and normal and tangential elastic contact forces between particles i and j ($\boldsymbol{f}_{ij}^{n}$ and $\boldsymbol{f}_{ij}^{t}$), respectively. Regarding notation, the bold symbols refer to vector or

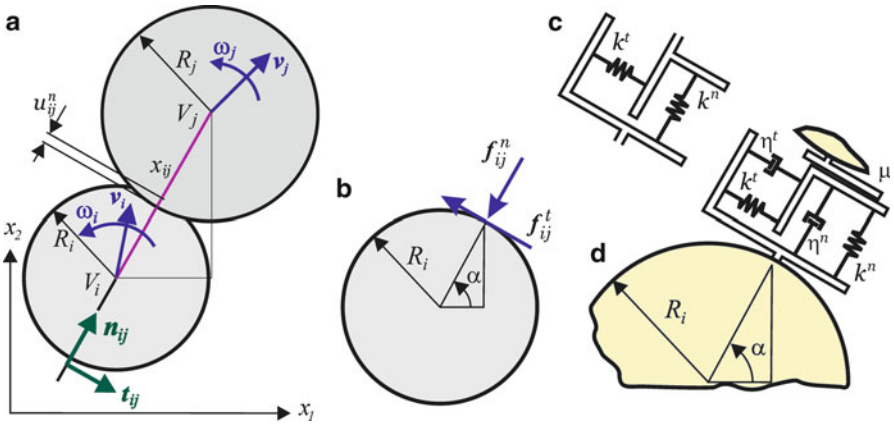


Fig. 2 Two circular particles in contact. (**a**) Definition of microparameters; (**b**) decomposition of the contact force; (**c**) basic model of the contact interface – bonded contact without damping; and (**d**) frictional contact with viscous damping

tensor quantities. According to Newton's second law, the translational and rotational motions of particle are defined by

$$\frac{d\boldsymbol{p}_i}{dt} = m_i\boldsymbol{g} + \sum_{j=1}^{N_i}\left(\boldsymbol{f}_{\boldsymbol{ij}}^{\boldsymbol{n}} + \boldsymbol{f}_{\boldsymbol{ij}}^{\boldsymbol{t}}\right), \quad \frac{d\boldsymbol{L}_i}{dt} = \sum_{j=1}^{N_i}\left(\boldsymbol{T}_{\boldsymbol{ij}}^{\boldsymbol{t}} + \boldsymbol{T}_{\boldsymbol{ij}}^{\boldsymbol{r}}\right) \tag{1}$$

where $\boldsymbol{p}_i = m_i\boldsymbol{v}_i$ and $\boldsymbol{L}_i = I_i\boldsymbol{\omega}_i$ are linear and angular momenta of particle i, defined in a usual way in terms of the mass, moment of inertia (I), translational ($\boldsymbol{v}$), and rotational ($\boldsymbol{\omega}$) velocities, while $\boldsymbol{T}_{ij}^{t}$ and $\boldsymbol{T}_{ij}^{r}$ are torques due to the tangential contact force and the rolling friction, respectively (Xiang et al. 2009). For multiple interactions, the interparticle forces and torques are summed for N_i particles interacting with particle i. In calculating contact forces, the contact of particles is modeled by a pair of linear spring–dashpot–slider contact model (Cundall and Strack 1979) in both the normal and tangential directions. The contact force vector that represents the action of particle j on particle i may be decomposed into a normal and a shear vector $\boldsymbol{f}_{\boldsymbol{ij}} = \boldsymbol{f}_{ij}^{n} + \boldsymbol{f}_{ij}^{t}$ where

$$\boldsymbol{f}_{ij}^{n} = -\left[k^{\boldsymbol{n}}\boldsymbol{u}_{ij}^{n} + \eta^{n}\left(\boldsymbol{v}_{ij}\cdot\boldsymbol{n}_{ij}\right)\boldsymbol{n}_{ij}\right], \quad \boldsymbol{f}_{ij}^{t} = \min\left\{-k^{\boldsymbol{t}}\boldsymbol{u}_{ij}^{t} - \eta^{t}\boldsymbol{v}_{\boldsymbol{ij}}^{\boldsymbol{t}}, \mu_f\left|\boldsymbol{f}_{ij}^{n}\right|\boldsymbol{t}_{ij}\right\} \tag{2}$$

and both force components include dissipative terms. The relative velocities of particle i with respect to particle j in Eq. 2 are defined as

$$\begin{aligned}\boldsymbol{v}_{ij} &= \boldsymbol{v}_i - \boldsymbol{v}_j + \left(\boldsymbol{\omega}_i\times\boldsymbol{R}_i - \boldsymbol{\omega}_j\times\boldsymbol{R}_j\right)\\ \boldsymbol{v}_{ij}^{n} &= \left(\boldsymbol{v}_{ij}\cdot\boldsymbol{n}_{ij}\right)\boldsymbol{n}_{ij}, \quad \boldsymbol{v}_{ij}^{t} = \boldsymbol{v}_{ij} - \boldsymbol{v}_{ij}^{n}\end{aligned} \tag{3}$$

In Eqs. 2 and 3, $\boldsymbol{u}_{ij}^{n}$ and $\boldsymbol{u}_{ij}^{t}$ are normal and tangential displacement vectors between particles i and j; $\boldsymbol{v}_{ij}$ is the relative velocity vector of the contact point; $\boldsymbol{n}_{ij}$

is the center-to-center unit vector; $\boldsymbol{t}_{ij}$ is the unit vector perpendicular to $\boldsymbol{n}_{ij}$, $\boldsymbol{v}_{ij}^{n}$ and $\boldsymbol{v}_{ij}^{t}$ are the relative velocities of contact point in normal and tangential directions, respectively; μ_f is coefficient of sliding friction; and k^n and k^t are the normal and tangential spring constants, respectively. For cohesive materials, k^n is constant, but for non-cohesive materials, it is not because it depends on the normal displacement, $k^n \propto \sqrt{u_{ij}^{n}}$ (Van Baars 1996). The dissipative damping contact coefficients η^n and η^t refer to the normal and tangential directions, respectively. The local viscous damping is often included into the DEM model to dissipate kinetic energy together with frictional sliding in order to reach equilibrium configuration more efficiently. It is important to note that the shear force in Eq. 2_2 is computed in an incremental fashion: when contact is formed, f_{ij}^{t} is set to zero and each subsequent relative shear–displacement (and, generally, velocity) increment $\boldsymbol{u}_{ij}^{t}$ results in a corresponding shear force increment. Thus, k^t is a tangent stiffness since it relates incremental displacement and force, while k^n is a secant stiffness since it relates total displacement and force.

The torque due to tangential contact force is $\boldsymbol{T}_{ij}^{t} = \boldsymbol{f}_{ij}^{t} \times \boldsymbol{R}_i$, while the rolling friction torque is calculated using a particular friction model.

For the frictional contact, the normal-direction spring acting in compression results in a shear force limited by a Coulomb friction law. Consequently, when the computed shear force reaches the Coulomb limit, the contact undergoes sliding and slip occurs, Eq. 2_2.

Alonso-Marroquin and Herrmann (2005) present a similar approach with particles represented by convex polygons.

Expressions for stress and strain tensors in terms of the microscopic contact parameters are given by Kruyt and Rothenburg (1996). The same authors also developed statistical theories for the elastic moduli of 2D particle assemblies (1998).

DEM Application for Non-cohesive Granular Materials

The deformation of non-cohesive granular materials, such as sand, is described well by the particle rigidity assumption since the deformation results primarily from the sliding and rotation of the particles as rigid bodies and the opening and interlocking at interfaces while the individual particle deformation is comparably small. Additionally, an interparticle friction coefficient is introduced to describe the ratio between shear and normal force when the interparticle overlap vanishes.

(The mechanical structure of cohesive soils is much more complex compared to the dry soils (e.g., Yang and Hsiau 2001) due to the presence of water, which results in cohesion between soil particles. In addition to inelastic contact forces and friction at interparticle contacts, the mechanical response of cohesive soil is influenced also by the capillary and the dynamic viscous forces produced by the discrete liquid bridges between particles (Zhang and Li 2006). Prunier et al. examine the DEM applications in geomechanics in this handbook.)

Jensen et al. (1999, 2001) enhanced DEM for modeling non-cohesive granular materials by introducing particles of complex shape obtained by combining several

smaller circular particles of various sizes into a cluster that acts as a single granule. Numerical simulations of the ring shear test, with varying normal loads, void ratio, and surface roughness, were performed to compare effects of complex particle shape with the circular particles. As expected, the computer simulations indicated the shear strength increase of the complex shape particle assemblies due to the reduction of particle rotations compared to the circular particle counterparts. The opportunity to model explicitly grain damage was another important feature of this clustering approach. The grain damage was included by allowing cluster particles to break apart according to a failure criterion based upon sliding work. The particle would become separated from the cluster as soon as the cumulative work done on an individual cluster particle reached a certain threshold. The computer simulations that included this feature revealed very distinct shear zones without significant reduction of the maximum shear strength of the particle assembly. It was demonstrated that the damage level was related to the angularity of clusters.

DEM modeling with the crushable clusters was successfully compared to the salient experimental trends reported in literature. Simulations revealed in detail response mechanisms on the particle level. Cheng et al. (2003) used a similar approach for 3D simulations.

DEM Implementation for Cohesive Materials

The defining characteristic of cohesive materials is the ability to transfer tensile normal force between bonded particles. Thus, Cundall's approach is extended to account for the interfacial tensile strength (Zubelewicz and Mroz 1983; Plesha and Aifantis 1983). In order to do so, DEM models are usually adapted to cohesive materials by adding a bond at the contact between two particles mimicking the presence of a matrix sticking to the particles, which "endows cemented granular materials with cohesion" Topin et al. (2007). This approach is used to model wide class of cemented granulates forming heterogeneous macroscopic materials such as sedimentary rocks, concrete, ceramics, grouted soils, solid propellants and high explosives, and some biomaterials. These materials can be also represented, in principle, by a simple model outlined in section "DEM Implementation for Non-cohesive Materials" (Fig. 2) with an important proviso that for a bonded contact, the normal-direction spring offers resistance in both compression and *tension*. In the course of deformation process, if the bonded contact between two particles fails in accordance with some prescribed failure criterion, the contact becomes frictional if two particles are still pressed against each other. Damage modes emanate naturally through a process of progressive debonding of particles when the strength, the critical strain, or the fracture energy of a bond between particles is exceeded due to external action.

Two different approaches to account for cohesive interactions between edges of two contiguous particles are the following: (i) the beam-enhanced DEM (beam element between particle centers; Kun and Herrmann (1996), D'Addetta

et al. (2001) and (ii) the interface-enhanced DEM (interface elements defined at particle edges; Kun et al. 1999).

DEM Application for Rocks

DEM originated from rock mechanics. Although rocks may not look like granular materials, they can be viewed as assemblies of distinct units bonded together by different models of cohesive forces or cementation effects. Thus, the overall mechanical behavior evolves in a simple and natural way, in the course of loading, through collective contribution of these distinct units where their debonding mimics elementary micro-damage events – basic building blocks of complex damage-evolution phenomena. A comprehensive review article by Jing (2003) presented the techniques, advances, problems, and likely future work in numerical modeling of rocks.

In 2D DEM rock simulations, particles can be randomly generated circles or convex polygons cemented together by insertion of a parallel bond in the contact region. Shapes and packings of particles have a profound effect on the distribution and intensity of interaction forces. The bond strengths are permitted to vary from contact to contact, representing another source of heterogeneity in the simulated material. The most representative explicit DEM method in rock mechanics is Cundall's (1980, 1988) distinct element method with polygonal/polyhedral blocks developed in the commercial computer codes UDEC and 3DEC (ITASCA^{TM}). A simpler DEM – the bonded particle method – based on the circular/spherical rigid particles is presented herein following closely Potyondy and Cundall (2004). This modeling approach is used in the commercial programs PFC2D and PFC3D (ITASCA^{TM}).

The bonded particle model (also referred to as the parallel-bond model) simulates rock as a densely packed assembly of nonuniformly sized circular particles glued together at contact points by insertion of parallel bonds representing the cohesive effects of cement. The model is fully dynamic and hence capable of describing complex phenomena of damage evolution such as nucleation, growth, and coalescence of microcracks resulting in damage-induced anisotropy, hysteresis, dilation, and softening.

The particle diameters are sampled from a uniform distribution bounded by D_{min} and D_{max}, and dense packing is obtained by following appropriate material-genesis procedure. The rigid particles can independently translate and rotate and interact at the soft contacts defined by normal and shear stiffness (Fig. 3). The particle overlaps are assumed to be small compared to their size to ensure that contacts occur "at a point." The set of microproperties consist of stiffness and strength parameters of the particles and the bonds. The modulus–stiffness scaling relations of grain and cement, which include particle size, ensure that the macroscopic elastic constants are independent of the particle size. The force–displacement laws at each contact relate relative particle movements to force and moment exerted on each particle.

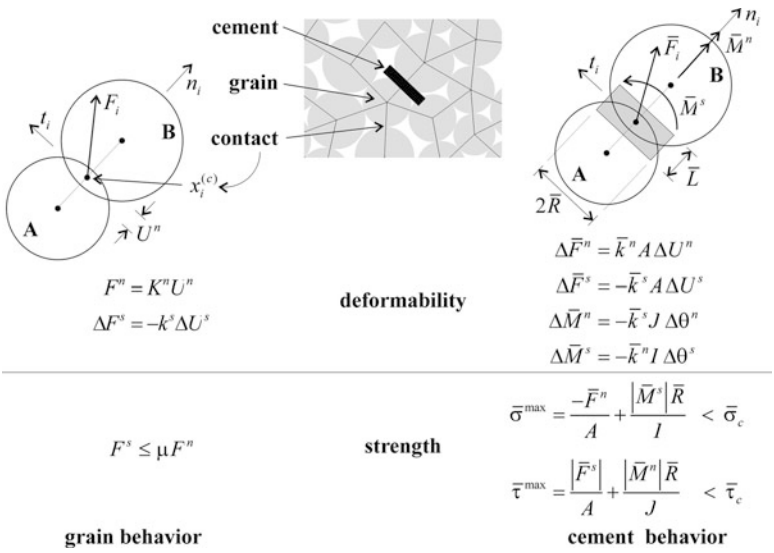


Fig. 3 Force–displacement behavior of bonded particle model of the grain–cement system (Reprinted from Potyondy and Cundall (2004) with permission from Elsevier)

Cundall and Hart (1992) described an explicit finite difference algorithm used to numerically evaluate dynamics of the simulated material. The DEM simulation technique is based on the assumption that time step is so small that, within one calculation step, disturbances cannot propagate from any particle further than its nearest neighbors. Advantages of explicit numerical scheme are discussed by Potyondy and Cundall (2004).

The bonded particle model mimics the mechanical behavior of an assembly of grains (particles) linked by cement (parallel bond) as illustrated in Fig. 3. The total force and moment acting on each contact consists of contact force $\boldsymbol{f}_{ij}$ resulting from the particle overlap and representing grain behavior (Eq. 2 with or without damping) and a force and moment, $\overline{\boldsymbol{f}}_{ij}$ and $\overline{\boldsymbol{M}}_{ij}$, carried by the parallel bond. These quantities contribute to the resultant force and moment acting on both particles that are input into Newton's second law numerically integrated to obtain particle trajectories.

The force–displacement behavior of the grain–cement system is outlined in Fig. 3. The grain behavior is the same as non-cohesive frictional interaction (section "DEM Slip Model with Rolling Friction" without contact dumping) described by the normal and shear stiffness, k^n and k^s, and friction coefficient, μ, per each particle. This contact is formed as soon as two particles overlap and the contact stiffness is determined by serial connections of the individual particle stiffnesses. The overlap, though physically inadmissible, mimics to some extent the local deformation of grains (especially when surfaces are not smooth and have asperities). The contact force vector can be decomposed into a normal and a shear vector as previously in Eq. 2.

The grain (contact) behavior is already discussed in section "DEM Slip Model with Rolling Friction": if a gap exists, both normal and shear forces are zero; otherwise, slip is accommodated in the usual manner by using Coulomb's law. It should be noted that the existence of parallel bond does not prevent slip since the bond elastic interaction acts in *parallel* with grain contact portion of force–displacement interaction.

The cement behavior is represented by the total force and moment, $\overline{\boldsymbol{f}}_{ij}$ and $\overline{\boldsymbol{M}}_{ij}$, carried by the parallel bond (action on particle B, i.e., j). The force and moment can be projected on the normal and tangential directions

$$\overline{\boldsymbol{f}}_{ij} = \overline{f}^{n}_{ij}\boldsymbol{n}_{ij} + \overline{f}^{t}_{ij}\boldsymbol{t}_{ij}, \quad \overline{\boldsymbol{M}}_{ij} = \overline{M}^{n}_{ij}\boldsymbol{n}_{ij} + \overline{M}^{t}_{ij}\boldsymbol{t}_{ij} \tag{4}$$

When a parallel bond is initialized, $\overline{\boldsymbol{f}}_{ij}$ and $\overline{\boldsymbol{M}}_{ij}$ are set to zero. Each subsequent relative translation and rotation increment, Δu^{n}_{ij}, Δu^{t}_{ij} and $\Delta\theta_{ij} = (\omega_j - \omega_i)\Delta t$, produces an increment in elastic force and moment (given in Fig. 3) that is added to the current values in manner described by Potyondy and Cundall (2004). The parameters in Fig. 3, A, I, and J, are the area, the centroidal moment of inertia, and the polar moment of inertia of the parallel-bond cross section, respectively.

The maximum normal and shear stresses acting on the parallel bond are calculated from beam theory and also presented in Fig. 3. If the maximum normal stress exceeds the tensile strength ($\overline{\sigma}^{\max} \geq \overline{\sigma}_c$) or the maximum shear stress exceeds the shear strength ($\overline{\tau}^{\max} \geq \overline{\tau}_c$), the parallel bond ruptures and is removed from the network, which reduces the contact to the basic non-cohesive frictional interaction.

Potyondy and Cundall (2004) used this DEM model for simulation of biaxial and Brazilian tests of granite and demonstrated its capability to reproduce numerous features of rock behavior such as fracturing, damage-induced anisotropy, dilation, softening, and confinement-driven strengthening. Damage evolution was obtained explicitly as a process of progressive accumulation of broken bonds; "no empirical relations are needed to define damage or to quantify its effect on material behavior." The damage patterns, such as those reprinted in Fig. 4, agreed well with experimental observations with respect to dynamics of the damage accumulation and revealed effects of the lateral confinement. They suggested that the effect of confinement on damage accumulation process was more pronounced in the softening region than in the hardening region. The authors argued that lateral confinement reduced the tensile forces that developed in a direction perpendicular to the sample axis and thereby favored formation of shear microcracks.

Potyondy and Cundall (2004) discussed in detail the effect of particle size on macroscopic properties. Specifically, they explored the role of D_{min} (a model parameter that controls mesh resolution and the length scale of the material) as an intrinsic part of material characterization. The elastic constants were independent of the particle size due to the scaling of the parallel-bond stiffnesses as a function of particle size. The unconfined compressive strength appeared to be particle size independent as well, but the simulation results were inconclusive regarding the

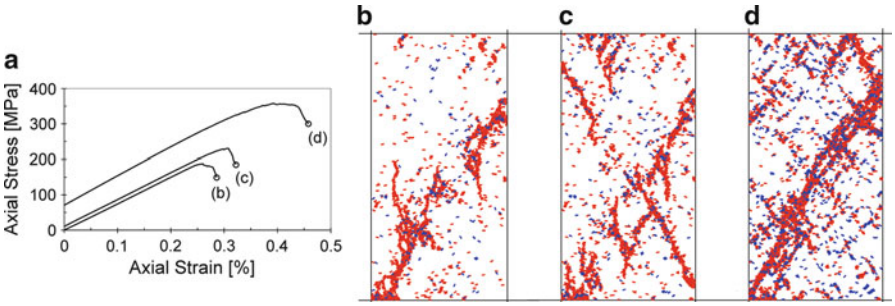


Fig. 4 Biaxial test simulation of granite: (**a**) axial stress versus axial strain and damage patterns in the softening region for three levels of lateral confinement: (**b**) 0.1 MPa, (**c**) 10 MPa, and (**d**) 70 MPa (Reprinted from Potyondy and Cundall (2004) with permission from Elsevier)

particle size effect on the friction angle and cohesion. The Brazilian strength exhibited clear particle size dependence as well as the mode-I fracture toughness

$$K_{\mathrm{Ic}} = \sigma_t' \sqrt{\pi R}, \quad \sigma_t' = \frac{s_n}{2Rt} \tag{5}$$

where σ_t' is the tensile strength of the model, s_n the bond tensile strength, R the particle size, and $t \equiv 1$ for 2D models. Consequently, it appeared that the particle size, controlling the model resolution, could not be chosen arbitrarily since it was related to the material fracture toughness as well. Instead, when modeling damage processes, the particle size and model properties should be selected to match both the material fracture toughness and the unconfined compressive strength.

The most pronounced shortcoming of the bonded particle model observed in these simulations was that material strength matched the granite strength only for stresses near the uniaxial state. Otherwise, the tensile strength was too high, and the slope of the strength envelope as a function of confining stress was too low. The authors attributed this limitation to oversimplified particle shape.

D'Addetta et al. (2001, 2002) developed a 2D model of heterogeneous cohesive frictional solids with material structure represented by an assembly of discrete convex polygons joined together by simple beams accounting for cohesive effects. Depending on the selection of model parameters, it can represent a range of different materials from non-cohesive dry soils to a variety of cohesive materials. This DEM model is an extension of the models used to simulate mechanical response of randomly shaped granular materials (Kun and Herrmann 1996).

The process of model development enveloped three major steps. First, computational domain was discretized into a froth of convex polygons (mimicking grains of the material) by a random Voronoi tessellation. Second, the global mechanical response of the sample was defined by proper interactions among contiguous rigid polygons. The failure criterion necessary to define the element rupture on the microscale was the final model ingredient. The definition of microparameters, contact forces, and torques reduced to a particle center was conceptually similar

to that discussed previously for the circular particles (Fig. 2). Suffice it to mention that in the absence of analytical solution for the real deformational behavior of contacting polygons of arbitrary shape, an approximate technique must be devised instead.

The beam rupture criterion took into account the stretching and bending breaking modes

$$p_{ij}^{(b)} = \left(\frac{\varepsilon_{ij}^{(b)}}{\varepsilon_{\max}^{(b)}}\right)^2 + \frac{\max\left(|\varphi_i|, |\varphi_j|\right)}{\varphi_{\max}} = 1, \quad \varepsilon_{ij}^{(b)} \geq 0 \tag{6}$$

where $\varepsilon_{ij}^{(b)}$ was the longitudinal beam strain, φ_i and φ_j were the rotation angles at the beam ends, and $\varepsilon_{\max}^{(b)}$ and $\varphi_{\max}$ were the threshold values for the two breaking modes. Varying the two threshold values could control the relative importance of the two breaking modes. The time evolution of the system was obtained by solving numerically Newton's equations of motion Eq. 1 for each individual polygon in the assembly.

The model parameters were identifiable with properties of cohesive granular materials up to some extent, but it remained uncertain whether the quality of this identification depended on the rupture criterion or if it was influenced "by the special combination of beam and particle elements within the model" (D'Addetta et al. 2002).

The simulation results of the uniaxial compression test (Fig. 5) and the simple shear test demonstrated the model's ability to capture salient effects of brittle deformation and damage evolution. Polar plots of damage fabric (Fig. 5d) provided a handy description of damage distribution and revealed the tendency of anisotropic damage accumulation in the softening phase.

The authors also used this model to study dynamic fragmentation of heterogeneous brittle solids in various loading configurations: the explosion of a disk-shaped solid, the impact of a projectile into a solid block, and the collision of macroscopic bodies. The simulation results were found to be in reasonable qualitative agreement with experimental observations bearing in mind the limitations of the 2D simulation setup.

DEM Application for Concrete

Kim and Buttler (2009) developed a DEM cohesive fracture model of a disk-shaped compact tension specimen to investigate various aspects of mode-I fracture of asphalt concrete. Their objective was to use DEM to integrate experimental, analytical, and numerical approaches based on the image-based microstructure (high-resolution image analysis mapped into a clustered mesh as illustrated in Fig. 6). The material properties were identified based on experimental bulk and fracture tests and comprehensive inverse analyses and assigned to all DEM model contacts (the phase contact rule schematics is indicated in Fig. 6).

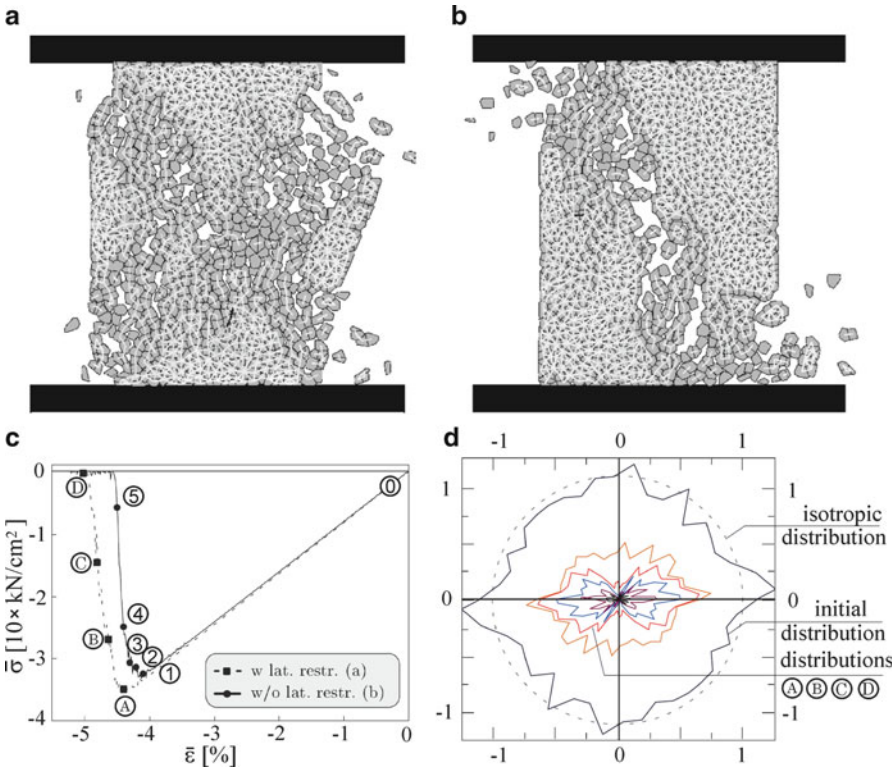


Fig. 5 Simulation of the compression test: fractured configurations with (**a**) and without (**b**) lateral confinement; (**c**) stress–strain curves for the two compression cases, and (**d**) damage distribution of evolving compressive failure for the setup with lateral confinement (D'Addetta et al. 2002)

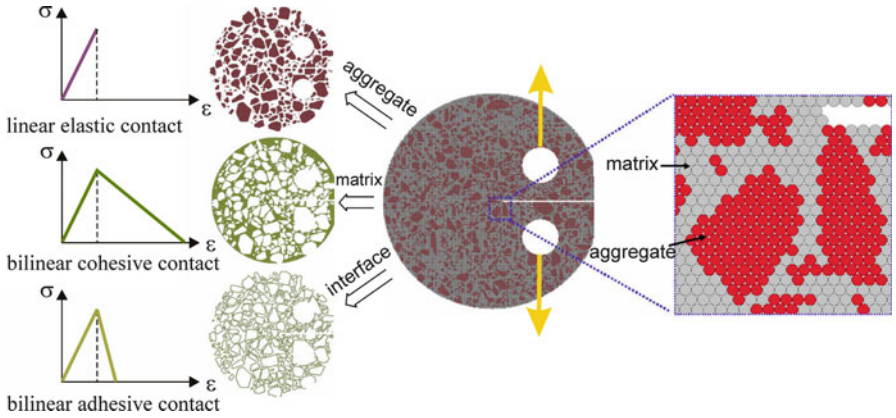


Fig. 6 Digitized images of three-phase geometry of asphalt concrete with indicated phase contact models and DEM mesh of compact tension specimen (Reprinted from Kim and Buttlar (2009) with permission from Elsevier)

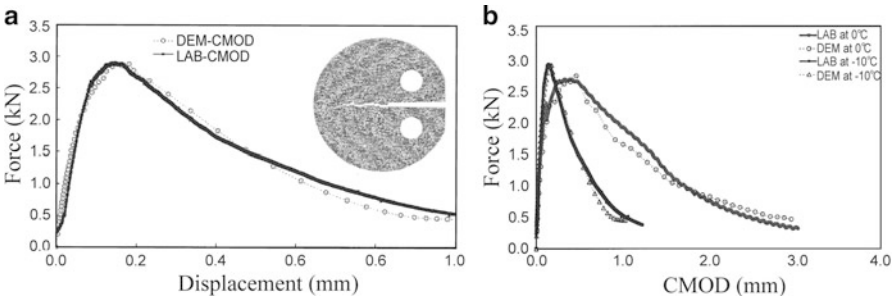


Fig. 7 A selection of comparisons of experimental and DEM simulation results: (**a**) two different nominal maximum aggregate sizes (9.5 mm and 19 mm) at −10 °C, (**b**) nominal maximum aggregate size 19 mm at two temperatures (0 and −10 °C) (Reprinted from Kim and Buttlar (2009) with permission from Elsevier)

The DEM simulations of mode-I loading of the specimen were performed with different nominal maximum aggregate sizes, temperatures, and aggregate types to gain understanding of various fracture mechanisms. The simulation results showed a very good agreement with experimental results at different temperatures and with different mixture types (e.g., Fig. 7). It was demonstrated that, in addition to global fracture response, the heterogeneous DEM model could potentially capture the stress and damage distributions as well. The authors concluded that the DEM modeling approach "appears to have significant potential to aid in the understanding of fracture behavior in asphalt concrete."

Particle Models

Particle models are a generalization of spring networks that includes dynamic effects and can be also viewed as an engineering MD offspring on a coarser scale. As such, they take advantage of the time-honored MD techniques to cope in a straightforward manner with various highly complex motions and physics of extremes. MD is a method frequently used in various branches of computational physics to analyze motion of large ensembles of atoms or molecules. Hence, the MD computer simulation techniques have been treated extensively in literature. Traditional references for particle modeling include Allen and Tildesley (1987) and Greenspan (1997), while Wang et al. (2010) and Munjiza et al. (2011) may be consulted for more recent developments. Overviews of the basic concepts of particle modeling of damage have been presented by Ostoja-Starzewski (2007) and Kale and Ostoja-Starzewski in an abridged form in this handbook.

A particle model consists of N particles of known masses m_i and positions $\boldsymbol{r}_i$ arranged in a lattice. (Hereinafter, the lower case alphabetic indices (i, j) designate particles, while the lower case Greek letter indices $(\alpha, \beta, \gamma, \delta)$ are reserved for tensor components.) The continuum particles often interact according to the central-force law that completely defines the strain energy density function in terms of particle

positions. Ignoring many-body interactions, the system energy can be approximated by the sum of isolated empirical pair potentials φ (pairwise additivity assumption). The system of Newton's equations of motion

$$\frac{d\boldsymbol{p}_i}{dt} = \boldsymbol{F}_i = -\sum_{j \neq i} \frac{d\varphi}{dr_{ij}} \frac{\boldsymbol{r}_{ij}}{r_{ij}} \tag{7}$$

is approximated by corresponding finite difference equations and subsequently solved using one of several mature MD techniques at our disposal for that purpose. Assuming that interparticle forces are conservative, the intensity of central force exerted on the particle i by the particle j and the total force exerted on the particle i by all nearest neighbors are

$$f_{ij} = \left|\boldsymbol{f}_{ij}\right| = -\frac{d\varphi\left(r_{ij}\right)}{dr_{ij}}, \quad \boldsymbol{F}_i = -\sum_j f_{ij} \frac{\boldsymbol{r}_{ij}}{r_{ij}} \tag{8}$$

where $r_{ij} = \left|\boldsymbol{r}_j - \boldsymbol{r}_i\right|$ is the interparticle distance equal to r_{0ij} at equilibrium.

An example of pairwise interparticle potential frequently used in particle models for mesoscale simulations of brittle materials is a combination of the Born–Mayer (Eq. 9_1) and Hookean (Eq. 9_2) interatomic potentials

$$\varphi_r\left(\overline{r}_{ij}\right) = \frac{k_{ij} r_{0ij}^2}{(B-2)} \left(\frac{1}{B} e^{B\left(1-\overline{r}_{ij}\right)} - \overline{r}_{ij}^{-1}\right), \quad \overline{r}_{ij} < 1$$

$$\varphi_a\left(\overline{r}_{ij}\right) = \frac{1}{2} k_{ij} r_{0ij}^2 \left(\overline{r}_{ij}^2 - 2\overline{r}_{ij}\right), \quad \overline{r}_{ij} \geq 1 \tag{9}$$

where k_{ij} is link stiffness in tension, $\overline{r}_{ij} = r_{ij}/r_{0ij}$, while the fitting parameter B defines the slope (steepness) of the repulsive wall and can be deduced, for example, from the equation of state. The potential Eq. 9_2 has a tension cutoff to simulate elastic-perfectly brittle behavior. This interparticle potential was developed by Mastilovic and Krajcinovic (1999a) to capture several important features of the deformation process typical of quasibrittle materials, namely, brittle behavior in tension, increase of shock wave velocity, and the decrease of compressibility with increasing pressure.

The critical modeling problem is how to pass from a given atomic/molecular potential or a set of continuum properties to an interparticle potential, which is common key issue of all numerical approaches in computational mechanics of discontinua.

The quenched disorder can be introduced into the particle model to describe brittle materials with random microstructure. The particle network disorder may be topological (unequal coordination number), geometrical (unequal length of bonds), or structural (unequal stiffness and/or strength of bonds). The disorder is further enhanced by damage evolution. Due to the initial randomness of the microstructure, the nature of the damage evolution is stochastic.

Evaluation of Stress, Strain, and Stiffness Components

The general expression for the stress and stiffness tensors at a given particle in a system in equilibrium can be obtained by expanding the elastic strain energy density of the discrete system into a Taylor series with respect to strain $\varepsilon_{\alpha\beta}$. In equilibrium state, when the total force Eq. 8_2 acting on any particle is zero, the system must be stable with respect to the application of a small, homogeneous strain tensor $\varepsilon_{\alpha\beta}$. The linear term in the Taylor series for the elastic strain energy represents the stress tensor, which is a general relation of thermodynamics. The quadratic term defines the elastic stiffness tensor. If the interaction of a system of particles can be approximated by the central-force potential, $\varphi = \varphi(r_{ij})$, expressions for stress and stiffness tensors can be obtained (Vitek 1996) in the following forms:

$$\sigma_{\alpha\beta} = \frac{1}{2V} \sum_{\substack{i,j \\ j \neq i}} \frac{d\varphi}{dr_{ij}} \frac{r_{ij}^{\alpha} r_{ij}^{\beta}}{|r_{ij}|}, \qquad C_{\alpha\beta\gamma\delta} = \frac{1}{2V} \sum_{\substack{i,j \\ j \neq i}} \left(\frac{d^2\varphi}{dr_{ij}^2} - \frac{1}{r_{ij}} \frac{d\varphi}{dr_{ij}} \right) \frac{r_{ij}^{\alpha} r_{ij}^{\beta} r_{ij}^{\gamma} r_{ij}^{\delta}}{r_{ij}^2} \quad (10)$$

where V is an averaging area, r_{ij}^{α} an appropriate α projection of $\boldsymbol{r}_{ij}$, etc.

The strain is calculated by comparing particle positions in the current and the reference (initial) configuration. The components of the left Cauchy–Green strain tensor of ith particle are commonly defined by

$$b_{\alpha\beta} = \frac{1}{3} \sum_{j=1}^{6} \frac{r_{ij}^{\alpha} r_{ij}^{\beta}}{r_{0ij}^2} \quad (11)$$

It should be noted that, unlike the stress Eq. 10_1, the virial strain Eq. 11 is valid instantaneously in time and space (Buehler et al. 2003).

Examples of Applications of Particle Models

Krajcinovic and Vujosevic (1998) used a particle model to simulate quasi-static biaxial compression test in order to investigate localization phenomena in heterogeneous quasibrittle materials. The experimental basis for the particle model application was provided by physical test data reported in this chapter. A topologically ordered central-force particle network dual to the Voronoi tessellation of grain boundaries represented a generic rock sample of appropriate aspect ratio. The quenched disorder was introduced by sampling the equilibrium interparticle distances r_{0ij} and rupture strengths from the Gaussian and uniform probability distributions, respectively. The interparticle forces were derived from the Hookean potential Eq. 9_2 for the entire interaction domain.

The particle network loading mimicked the actual physical experimental procedure: "the sample" was subjected initially to the hydrostatic pressure, kept fixed at

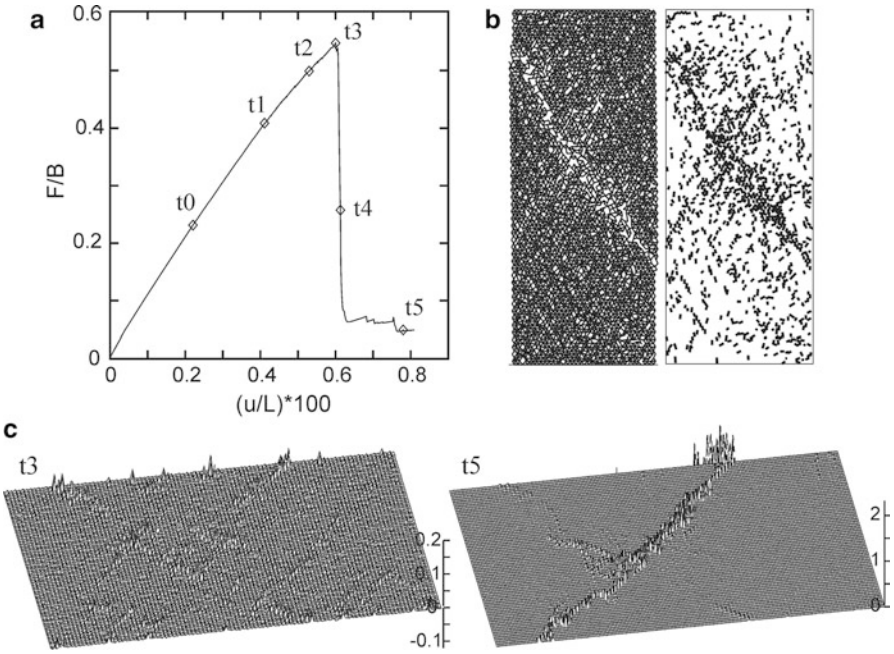


Fig. 8 (**a**) Dimensionless load–displacement curve in the loading direction, (**b**) the damage pattern in state t_5 (remaining and broken links, respectively), (**c**) strain distributions in states t_3 and t_5 (Reprinted from Krajcinovic and Vujosevic (1998) with permission from Elsevier)

certain level defined by the desired lateral confinement, and followed by the monotonically applied displacement-controlled uniaxial compression. After each application of the contraction increment, the local viscous damping was used to dissipate kinetic energy in order to reach equilibrium state more efficiently. The elementary damage events took place by progressive link ruptures whenever the interparticle forces satisfied the link-rupture criterion (microscale tensile strength).

The kinetics of the localization process was related to the rate of increase of the correlation length by recording the distance $\lambda_{i,j}^{\min}$ between two consecutive link ruptures (mimicking acoustic emission signals). The response of the random system was initially homogeneous as the damage originated by defect nucleation. The preponderance of small distances $\lambda_{i,j}^{\min}$ at the load peak and the post-peak snapshots (t_3 and t_5) indicate clustering illustrated by Fig. 8c. The authors concluded that "the softening is related to the growing role of the auto-catalytic growth of the largest cluster."

Krajcinovic and Vujosevic (1998) discussed various aspects of the localization fault geometry. The angle at which the acoustic tensor for the given state was minimum ($\approx 34^\circ$) was in excellent agreement with continuum prediction, which was underlined by a series of simulation-generated damage patterns. The particle simulations agreed well with experimental results in indicating that faults in rock specimen were characterized by irregular shapes and blurred boundaries.

The elusive fault width was determined to depend strongly on observation details and, hence, offered but a partial description of the fault size. Thus, instead of measuring the geometry – almost always hidden within the specimen – the authors concluded that it was more reasonable to determine the fault "width" by measuring its effect on the effective transport properties of the specimen. This approach was demonstrated by using propagation of elastic primary (pressure, *p*-) waves through a specimen spanned by a fault to estimate the fault width. The gist was that the *p*-wave imparted at the top of faulted specimen would need more time to arrive to the bottom than in the case of pristine specimen ("ant in the labyrinth" concept) and that the time lag was proportional to the fault width. The authors suggested that the fault width, which was proportional to the time lag, admitted the scaling law

$$w \propto \dot{\varepsilon}^{0.35} \tag{12}$$

where $\dot{\varepsilon}$ was the rate of the imparted strain pulse. The scaling relation Eq. 12 implied that the fault width would vanish in the static case, which was justified by the static treatment of the dynamic phenomenon.

Mastilovic and Krajcinovic (1999a) used a particle model to simulate the dynamic cavity expansion in a heterogeneous brittle material to obtain mesoscale data for analytical modeling of high-velocity expansion of a cylindrical cavity. The particle assembly was topologically ordered but geometrically and structurally disordered in the way just described with regard to the model used by Krajcinovic and Vujosevic (1998). On the other hand, each bulk particle was linked to six nearest neighbors by a nonlinear force–displacement relation obtained from the hybrid potential Eq. 9. The link-rupture criterion was defined in terms of the critical link elongation. The healing of a microcrack was prevented by ruling out the establishment of the cohesive interaction between two particles that were not linked initially or were separated earlier in the course of deformation by the rupture of the link that kept them together. However, the repulsive contact force could be established between the two particles that were originally not connected "not being nearest neighbors" or were at one point separated by stretching the link beyond the rupture limit. The cavity was nucleated by removal of a single particle from the middle of the random particle network of circular shape (Fig. 9). The nearest neighbors of the removed particle defined the cavity rim, which was driven radially outward in a displacement-controlled manner at a desired constant expansion rate $\dot{a}$.

The typical damage patterns (Fig. 9) depended on the magnitude of the externally imparted energy. The damage map corresponding to the highest expansion rate (Fig. 9a) was axially symmetric with damage front propagating with the velocity $(0.8–0.9)C_L$, which agreed well with experimental results (Mastilovic and Krajcinovic 1999a and references therein). With reference to Fig. 10a, the circumferential stiffness degraded more rapidly than the radial stiffness since initially most of the damage in the process zone is attributable to the radial microcracks.

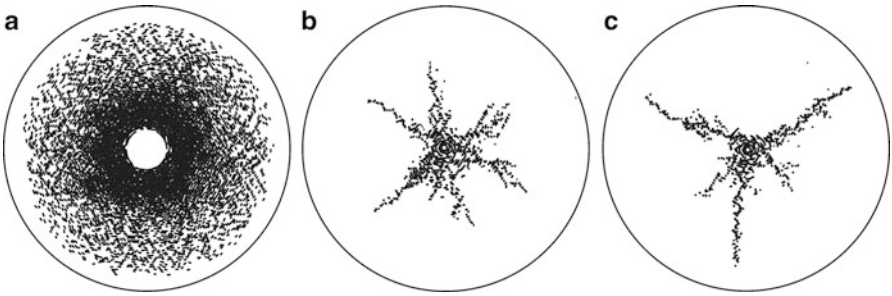


Fig. 9 Typical damage patterns for three cavity expansion rates: (**a**) $\dot{a} = 0.135\,C_L$, (**b**) $\dot{a} = 0.00135\,C_L$, and (**c**) $\dot{a} = 0.000135\,C_L$, where C_L is the velocity of longitudinal elastic waves and each short line represents a ruptured link (Reprinted from Mastilovic and Krajcinovic (1999a) with permission from Elsevier)

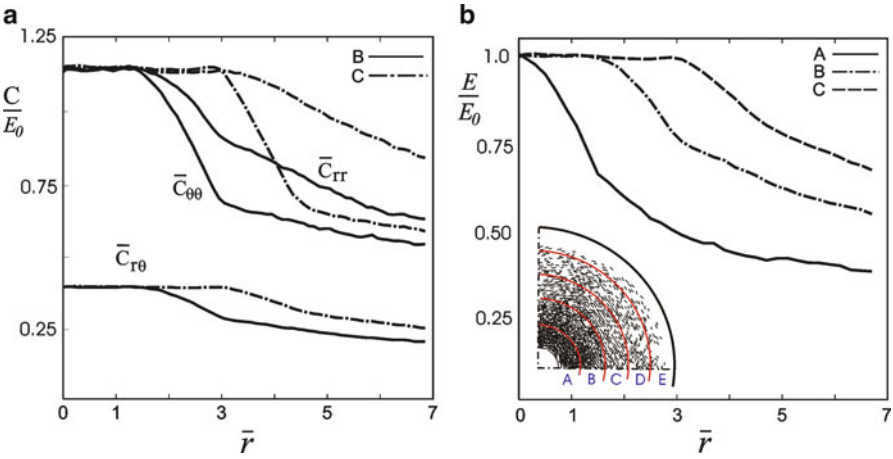


Fig. 10 Time history of the effective material properties: (**a**) stiffness tensor components in annular averaging regions B and C and (**b**) modulus of elasticity in annular regions A, B, and C. The computation domain is divided into five annular regions of equal width, marked by A to E, over which the field parameters or properties are averaged. E_0 is modulus of elasticity of pristine material. Indices r and θ mark the radial and circumferential directions, respectively (Reprinted from Mastilovic and Krajcinovic (1999a) with permission from Elsevier)

The reduction of the radial stiffness rate depended on the radial distance from the cavity, while the degradation of the effective circumferential stiffness was characterized by two distinctive rates that appear to be independent of the radial coordinate. The peak radial traction at the cavity rim, crucial for penetration mechanics modeling, was found to be approximately equal to the radial stress at the elastic wave front for which the analytical solution was available. The parabolic and bilinear simulation data fits of radial traction at the cavity surface as functions of

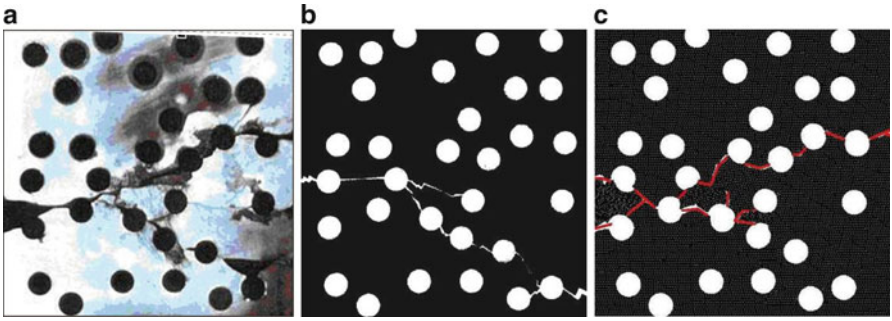


Fig. 11 Experimental and computational simulation fracture patterns in a randomly perforated epoxy plate loaded in tension: (**a**) experiment, (**b**) finite element method, and (**c**) particle method (Reprinted from Wang et al. (2010) with permission from the authors)

the cavity expansion velocity were used successfully by Mastilovic and Krajcinovic (1999b) to model penetration depth of an ogive-nosed projectile into limestone and concrete targets.

Wang and Ostoja-Starzewski (2005) and Wang et al. (2010) adopted particle modeling to simulation of dynamic fracture phenomena in homogeneous and heterogeneous materials. The lattice-type particle model had the same functional as traditional MD models, including the use of the interparticle potential inspired by the classic Lennard–Jones type, yet on centimeter length scales (refer to Kale and Ostoja–Starzewski in this handbook). The typical particle modeling issue – crucially important for the appropriateness of this model to simulate the complex mechanical response of the material – was the transition from a given empirical interatomic potential to an interparticle potential. In order to determine four unknown variables of the interparticle potential, Wang and Ostoja-Starzewski (2005) established four equality conditions (of mass, elastic strain energy, modulus of elasticity, and tensile strength) between the particle model and the MD model. They derived the equations for four unknown variables and carried out a parametric study to investigate the differing effects they had on mechanics response.

This particle model was applied in a series of papers to simulate dynamic fragmentation, thermal effects on ore breakage, random crack growth in epoxy plates (Fig. 11), polymeric material indentation, wave propagation-induced fracture, and impact of a rigid indenter (refer to Wang et al. (2010) for a list of references).

Mastilovic et al. (2008) and Mastilovic (2011, 2013) used particle modeling for simulation of dynamic uniaxial tensile test of disordered material with a low fracture energy. The hybrid interparticle potential Eq. 9 governed interaction of continuum particles with their nearest neighbors. The model was geometrically and structurally disordered. The link-rupture criterion was defined in terms of the critical link elongation $\varepsilon_{cr} = const.$ The problems of the loading at extremely high rates, including the uniform load distribution, were solved by imposing an instantaneous initial velocity field to all particles in the loading direction, $\dot{x}_1(t=0) = \dot{\varepsilon}_1 \; x_1$, and perpendicular to it, $\dot{x}_2(t=0) = -v_0^{(\varepsilon)} \dot{\varepsilon}_1 \; x_2$,

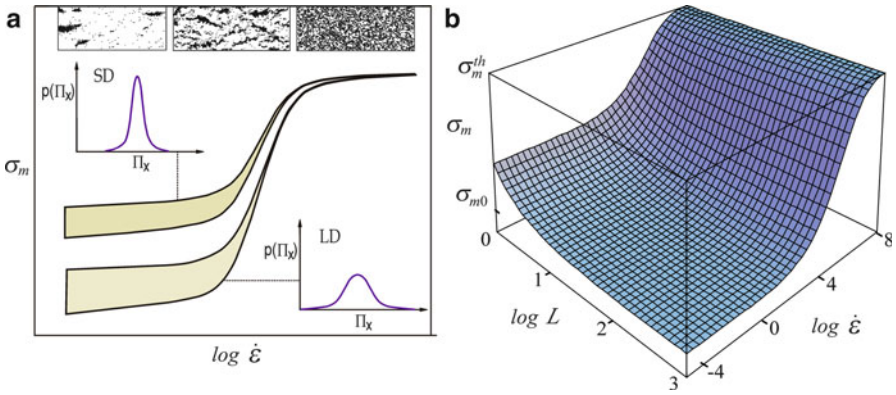


Fig. 12 (**a**) The tensile strength (σ_m) dependence on the strain rate indicating the ordering effect of kinetic energy and the effect of geometrical and structural disorder (SD and LD stand for small and large disorder, respectively; typical damage patterns are illustrated at the *top*) (Mastilovic 2011); and (**b**) schematic representation of the tensile strength dependence on the strain rate and the representative sample size (L) (Mastilovic 2013)

defined in terms of the prescribed strain rate, $\dot{\varepsilon}_1 = \dot{L}/L$. (The coordinates refer to the centroidal coordinate system and $v_0^{(\varepsilon)}$ is the apparent plane-strain Poisson's ratio). Subsequently, at $t > 0$, only velocity of the particles located at longitudinal boundaries was controlled, $\dot{x}_1 = \pm\dot{\varepsilon}_1 \; L/2$, while motions of all other particles were governed by Newton's equation of motion Eq. 7.

The shaded areas in Fig. 12a depicted schematically the scatter of strength data. The large scatter, characteristic of the low strain rates, was reduced to a single line at the extreme rates (roughly, $\dot{\varepsilon} \geq 1 \times 10^7 \; \mathrm{s}^{-1}$), which was indicative of the substantial reduction of the tensile-strength scatter close to the "upper-plateau" loading-rate range. The evident transition from the stochastic to the deterministic behavior – reflected by the reduction of the strength dispersion and change of damage-evolution patterns discussed by Mastilovic et al. (2008) – was more pronounced in the case of the large disorder. Figure 12b illustrates an empirical expression proposed by Mastilovic (2013) to model the strain rate effect on the dynamic tensile strength including the dependence upon the representative sample size.

The observed linearity of the rate dependence of the stress-peak macroscopic response parameters, time-to-failure, and damage energy rate (t_m, $\dot{\mathrm{E}}_{Dm}$) was expressed as follows:

$$t_m \;\; \dot{\varepsilon} = \text{Const.}, \quad \dot{\mathrm{E}}_{Dm} \;\; \dot{\varepsilon}^{-1} = \text{Const.} \tag{13}$$

Furthermore, the computer simulation results obtained from this particle model offered connection between the macroscopic response parameters at the stress peak (t_m, $\dot{\mathrm{E}}_{Dm}$) and the microscopic failure criterion (ε_{cr})

$$t_m \;\; \dot{\varepsilon} \propto \varepsilon_{cr}^{-1}, \quad \dot{\mathrm{E}}_{Dm} \;\; \dot{\varepsilon}^{-1} = \varepsilon_{cr}^{-1/6} \tag{14}$$

for loading within the wide range of the strain rates that encompasses variety of damage mechanisms. As noted by Mastilovic (2011), the scaling relation Eq. 13_1 is identical to

the empirical relation between creep rate and time to rupture for the constant-stress quasi-static loading and the strain-controlled brittle creep fracture. It is also similar to the time-to-failure derived by Mishnaevsky (1998) by combining the main ideas of continuum damage mechanics and statistical and kinetic theories of strength.

Failure Size Effects and Fractal Theories

Brittle (or embrittled) and quasibrittle microstructural systems have the tendency to fail catastrophically with little or no early warning. Failure modeling and prediction for these systems is of utmost importance and has proven to be a formidable task of damage mechanics. A major complication is the sample size dependence of both the onset of strain localization and the consequent damage evolution. Consequently, it is hard to predict behavior of large structures based on laboratory tests on similarly shaped samples unless a size effect model (scaling law) can be established. If a scaling law is available, knowledge of the statistics of a process on one scale allows inferring the statistics of the same process on any other scale.

The problem has been under investigation for centuries and many researchers have attempted a number of different strategies. Some modern approaches have originated from fractal theory and lattice models. While this remains an open and promising research topic, current models are prone to criticism and are far from design codes in structural engineering.

Within the context of continuum models, several research groups attempted to prove the fractal nature of the random failure patterns and the self-affine roughness on crack mechanics in elastic solids. For instance, Mishnaevsky, Jr., (1996) monitored the surface roughness of crack and the specific surface energy needed to form a crack by the mechanism of microcrack coalescence and concluded that the fractal dimension of crack may be monitored during the crack formation process to compute the time-to-fracture in heterogeneous solids. Also, in a number of papers, another group (Cherepanov et al. 1995; Balankin et al. 1996) suggested that the usual LEFM expressions for stress concentration at the crack tip could be replaced by a fractal version based on a roughness-related power law exponent α and a fractal stress intensity factor K_f as

$$\sigma_{ij} \propto K r^{-0.5} \Rightarrow \sigma_{ij} \propto K_f \frac{r^{-\alpha}}{l_0} \tag{15}$$

when crack length l falls between a lower cut off l_0 and a self-affine correlation length ς, $l_0 < l < \varsigma$. Similar continuum-based approaches have followed (Borodich 1997) with some noteworthy contributions that include, for example, the quantized fracture mechanics for fractal cracks (Pugno and Ruoff 2004; Wnuk and Yavari 2008) or the fractional continuum framework of fracture and damage discussed by Tarassov (2013) and Ostoja-Starzewski et al. (2013). Wnuk treats in this handbook the topic of fracture mechanics with fractal cracks.

On the other hand, the scope of this section is restricted to discrete damage models, succinct commentaries of some finite-size scaling ideas, and relevant

literature. For quite a few decades, lattice models have been focused on the investigation of finite-size scaling and on the formulation of physical/rational models of damage. They have appealed specially to the community of physicists and mathematicians active in statistical physics who have seized the opportunity to investigate failure in heterogeneous systems by the same approaches developed for phase transitions and chaos.

Percolation Theory of Damage in Discrete Models

Percolation theory is one simple approach to investigate phase transitions in statistical physics (e.g., Stauffer and Aharony 1994) and has been applied with some success to geometrical and transport properties of mechanical lattices. Since a damaged lattice can be regarded as a random graph of connected clusters, it can be studied by means of percolation theory. In that view failure is treated as a phase transition that occurs at percolation condition, i.e., when the correlation length ξ, associated to the connected/interacting clusters of microcracks, spans the entire finite-size lattice L (or diverges for an infinite lattice as $L \to \infty$). The percolation threshold p_c is defined as the occupation probability p at which an infinite cluster appears in the lattice according to a power law fractal exponent ν

$$\xi \propto |p_c - p|^{-\nu} \tag{16}$$

For a mechanical network, p roughly corresponds to the density of unbroken springs and p_c is critical point associated to failure. The threshold p_c is defined with respect to an infinite lattice and approached asymptotically in the limit of $L \to \infty$. The application of the results to finite-size systems happens by renormalization group approaches, such as coarse graining techniques (e.g., Christensen 2002).

Similar scaling laws were sought for many other and diverse parameters (e.g., connectivity, number of microcracks, etc.) and network transport properties (e.g., conductivity, stiffness, etc.) related to the "failure transition" and thereby exhibiting a singularity.

As an example, Sen et al. (1985) studied the percolation model for the central-force elastic lattice, finding that bulk modulus (K) and shear modulus (G) scaled as

$$K, G \propto (p - p_c)^{\beta} \tag{17}$$

with the following numerical estimates:

$$\begin{aligned} p_c &= 0.58, \quad \beta = 2.4 \pm 0.4 \quad \text{2D triangular lattices} \\ p_c &= 0.42, \quad \beta = 4.4 \pm 0.6 \quad \text{3D FCC lattices} \end{aligned} \tag{18}$$

The same group proposed an effective medium theory of α models, mapping the percolation property of the central-force lattice to a continuum scale, also exploring importance of the coordination number of lattice sites on the scaling (Feng et al. 1985).

Many authors (Chelidze 1982; Roux and Guyon 1985; Ostoja-Starzewski 1989, etc.) have reported similar results, but, despite the apparent simplicity, the application of percolation ideas to damage has proved to be not straightforward.

Krajcinovic (1996) offered a clear and detailed essay on this subject, stressing the importance of percolation theory in damage mechanics and its limitations. Percolation theory ought to be regarded as complementary to mean-field theories of continuum mechanics (e.g., dilute concentration models of damage), offering a way to approach size-scaling issues by means of relations that are universal and supposedly independent of microstructural details. Hansen et al. (1989), among others, investigated the universality problem for central-force lattices. However, one main problem is the estimation of fractal exponents associated to asymptotic behaviors, which require large computations. In time, successive reports have modified earlier accounts, and larger simulations have indicated that fracture damage may not comply with basic (uncorrelated) percolation process (Nukala et al. 2006). Another drawback inherent to some percolation studies resides in the preservation of isotropy during the percolation process that proceeds by either random suppression or strengthening of links (e.g., Garcia-Molina et al. 1988), which makes them ill posed to study damage-induced anisotropy that immediately arises in quasibrittle (vectorial) systems (Rinaldi 2009). A critical review by Guyon et al. (1990) represents a relevant and insightful reading on the subject.

Fractal Scaling Laws of Damage in Discrete Models

Besides percolation models, lattice models represent a fertile playground for the application of many other methods of statistical physics. The fuse lattice by De Arcangelis et al. (1985) illustrated in Fig. 13 is one of the first attempts to depart from percolation ideas and introduce damage by a more realist mechanism of fuse burnout caused by quenched or annealed disorder as opposed to random link suppression (Krajcinovic 1996).

These fuse models drew great attention immediately (e.g., Duxbury et al. 1986; Alava et al. 2006) as simple scalar models of failure in heterogeneous solids, but

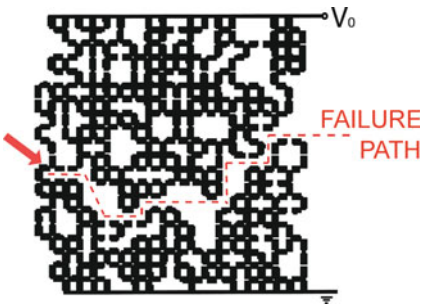


Fig. 13 Example of fuse lattice at the onset of failure, where the suppression of the last fuse pointed by the *arrow* leads to failure (zero conductance) (Redrawn from De Arcangelis et al. 1985)

were sided by actual "vectorial" mechanical models such as beam and central–force lattices. The latter are indeed significantly more complex and realistic, especially as far as the damage-induced elastic anisotropy and failure patterns are concerned.

The research scope also expanded to consider not just the scaling of one critical point corresponding to the failure threshold, like in percolation, but the entire response of the system during the damage process, particularly after strain localization. The objective was to establish fractal-based transformations that succeed in reconciling the mechanical response of samples of any size by mapping their mechanical response into one scale invariant curve, thus yielding a scaling law for the given damage process.

This idea is illustrated here for one specific method called the Family–Vicsek scaling (Family and Vicsek 1991; Barabasi and Stanley 1995), which was first used for growth of advancing solidification fronts at the liquid–solid interface. Let us consider the generic function $y(x, L)$ as dependent on a variable x defined over a network domain but also on the size L of the network itself, as depicted in Fig. 14a. The *knee shape* typically marks a phase transition at the critical point location (x^*, y^*).

If the Family–Vicsek scaling holds, then the data $y(x, L)$ shall map onto one universal scale invariant curve such that $f\left(x/L^{\beta}, L\right) = y/L^{\alpha}$ for any L, as indicated in Fig. 14b, according to the following scaling relation:

$$y(x, L) = L^{\alpha}\, f\left(\frac{x}{L^{\beta}}\right) \tag{19}$$

Three conditions must be met for this scaling procedure to be feasible:

1. At the transition the y value must be a fractal such that $y(x^*, L) \propto L^{\alpha}$.
2. The location of the transition must be a fractal such that $x^* \propto L^{\beta}$.
3. Before the transition the data must follow a power law $y(x, L) \propto x^{\gamma}$.

Consequently only two out of the three exponents $\{\alpha, \beta, \gamma\}$ are independent due to the constraint $\gamma = \alpha/\beta$.

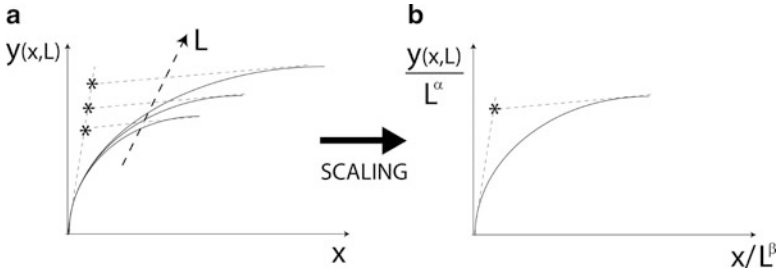


Fig. 14 Illustration of the scaling procedure applicable to some nonlinear systems that experience a transition governed by a universal law. The response y depends on the controlled variable x but also on the system size L, which controls the occurrence of the transition. A scaling law exists if $y(x, L)$ maps into a scale invariant curve upon normalizing y and x by L^{α} and L^{β}, respectively (Rinaldi 2011)

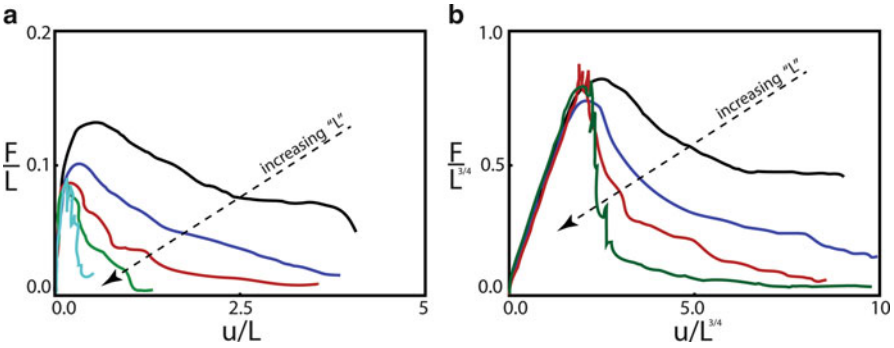


Fig. 15 Results of Family–Vicsek scaling (Data from Herrmann et al. 1989)

This empirical scaling procedure, borrowed from phase transitions and clusters theory, has proved useful to investigate microcracks cooperation and damage localization in multisite cracking and fracture processes, where lattice models had already revealed the existence of several fractals quantities. Hermann, Hansen, and Roux sought to apply the scaling to data from numerical experiments on central–force lattice (Hansen et al. 1989) and beam lattice (Hermann et al. 1989), finding satisfactory results only over certain portions of the damage process. Figure 15 displays original force–displacement data $F(u)$ for the beam lattice versus scaled data $F = L^{\alpha} f(\lambda L^{-\beta})$ with scaling exponents $\alpha = \beta = 0.75$. They also sought a scaling relation for other quantities, such as the number of broken links $n = L^{\gamma}\, \Psi(\lambda L^{-\beta})$. The results of the scaling displayed in Fig. 15 exemplify that the scaling could be used only up to the force peak.

Krajcinovic and coworkers (Krajcinovic and Basista 1989; Krajcinovic and Rinaldi 2005a) reexamined that approach, resolved some discrepancies, and reattempted the procedure on larger simulation data. More importantly, they introduced a significant modification to the original approach in two respects:

1. The force peak was acknowledged to be a transition point (distinct from failure) between two different damage mechanism (i.e., microcrack nucleation versus crack propagation) controlled by different fractal quantities.
2. The Family–Vicsek scaling was applied to the number of microcracks versus displacement, as opposed to force versus displacement.

Hence, in what is called here the "Krajcinovic approach," the Family–Vicsek is essentially applied twice, according to a two-step scheme entailing the sequential application of Eq. 19 to the number of ruptured links in hardening and softening regime (i.e., after the force peak) separately.

The procedure was developed for the α model (Krajcinovic and Rinaldi 2005b; Rinaldi et al. 2006) and was fully described in its final form by Rinaldi (2011), containing the resulting scaling relations. Figure 16 reports the comparison between the established scaling and the original simulation data over a large size range,

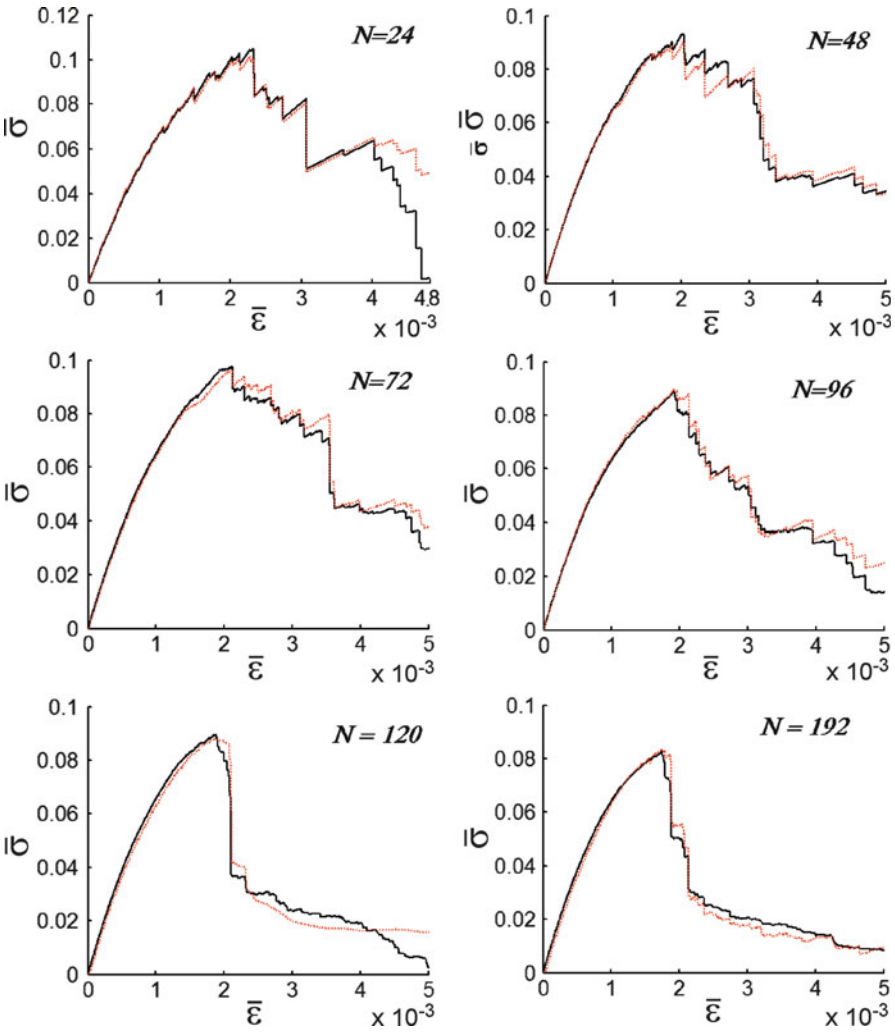


Fig. 16 Responses of scaling relations from the "Krajcinovic approach" (*full line*) versus original simulation data (*dotted line*) for six random replicates over a large size range $L = \{24, 48, 96, 120, 192\}$ (L = N = number nodes per side)

demonstrating a remarkable agreement throughout the stochastic damage process, also after the force peak.

The selection of results in this overview is by no means comprehensive and ever-growing body of literature is available to expand on the topic. Several other groups (e.g., Meisner and Frantziskonis 1996) contributed significant effort in this area and addressed the possibility to use also multifractal framework (i.e., beyond the simple fractal exponent obtained from a "box counting" procedure) to cast scaling relations for failure in quasibrittle lattice. The detailed overviews by Hansen and Roux (2000)

and Alava et al. (2006) on statistical methods of fracture in heterogeneous lattices remain precious references for the interested audience.

It is noteworthy that although physical models of fractures are considered a favorite research tool for physicists and mathematicians, the subject has appealed significantly to the engineering community (Ince et al. 2003; Rinaldi et al. 2007).

Conclusion

With the advent of affordable computers of ever-increasing power and feasibility of experimental observation at smaller spatial and temporal scales, discrete element models have already offered powerful solutions to many complex problems addressing both research and industry needs. The computational mechanics of discontinua has emerged in the process as an important and fast-growing branch of computational mechanics, firmly established, nowadays, as an integral part of cutting edge research in diverse fields such as nanotechnology, stem cell research, medical engineering, space propulsion, mining, milling, pharmaceuticals, powders, ceramics, composites, blasting, and construction.

The concise survey of 2D discrete damage models featured in the two chapters of this handbook represents a selection of topics and contributions – as representative as possible yet incomplete – in boasting a diverse range of research areas and a list of many influential references. While attempting to be as comprehensive and objective as possible, the overview admittedly and unavoidably suffers from usual shortcomings of similar endeavors such as an unintentional bias of authors from their personal interests, research backgrounds, and limited possibilities to abridge a vast amount of relevant work in the allotted space. Nonetheless, it may, hopefully, serve equally well as a critical compendium to the experienced scientist and as a primer to beginners.

References

M.J. Alava, P.K.V.V. Nukala, S. Zapperi, Statistical models of fracture. Adv. Phys. **55**(3–4), 349–476 (2006)

M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987)

F. Alonso-Marroquín, H.J. Herrmann, The incremental response of soils. An investigation using a discrete-element model. J. Eng. Math. **52**, 11–34 (2005)

A.S. Balankin, A. Bravo-Ortega, M.A. Galicia-Cortes, O. Susarey, The effect of self-affine roughness on crack mechanics in elastic solids. Int. J. Fract. **79**(4), R63–R68 (1996)

A.L. Barabasi, H.E. Stanley, *Fractal Concepts in Surface Growth* (Cambridge University Press, Cambridge, 1995)

N. Bicanic, Discrete Element Methods, in *Encyclopedia of Computational Mechanics: Fundamentals*, ed. by E. Stein, R. De Borst, T. Hughes (Wiley, New York, 2004), pp. 311–337

F.M. Borodich, Some fractal models of fracture. J. Mech. Phys. Solids. **45**(2), 239–259 (1997)

M.J. Buehler, F.F. Abraham, H. Gao, Hyperelasticity governs dynamic fracture at a critical length scale. Nature **426**, 141–146 (2003)

T.L. Chelidze, Percolation and fracture, physics of the earth. Planet. Inter. **28**, 93 (1982)
Y.P. Cheng, Y. Nakata, M.D. Bolton, Discrete element simulation of crushable soil. Geotechnique **53**(7), 633–641 (2003)
G.P. Cherepanov, A.S. Balankin, V.S. Ivanova, Fractal fracture mechanics. Eng. Fract. Mech. **51**(6), 997–1033 (1995)
K. Christensen, *Percolation Theory (ebook)* (MIT, Cambridge, 2002)
P.A. Cundall, A computer model for simulating progressive large scale movements in blocky rock systems, in *Proceedings of the Symposium of International Society of Rock Mechanics*, vol. 1, Paper No II-8. Nancy, France, 1971
P.A. Cundall, *UDEC – A Generalized Distinct Element Program for Modelling Jointed Rock*. Report PCAR-1-80, Peter Cundall Associates, European Research Office, US Army Corps of Engineers, 1980
P.A. Cundall, Formulation of a three-dimensional distinct element model – part I: a scheme to detect and represent contacts in a system composed of many polyhedral blocks. Int. J. Rock. Mech. Min. Sci. Geomech. Abstr. **25**(3), 107–116 (1988)
P.A. Cundall, R. Hart, Numerical modeling of discontinua. J. Eng. Comp. **9**, 101–113 (1992)
P.A. Cundall, O.D.L. Strack, A discrete numerical model for granular assemblies. Geotechnique **29**(1), 47–65 (1979)
G.A. D'Addetta, F. Kun, E. Ramm, H.J. Herrmann, in *From Solids to Granulates - Discrete Element Simulations of Fracture and Fragmentation Processes in Geomaterials,* In: Continuous and Discontinuous Modelling of Cohesive-Frictional Materials, Lecture Notes in Physics, 568, ed. by P.A. Vermeer et al. (eds.) (Springer, Berlin Heidelberg, 2001), pp. 231–258
G.A. D'Addetta, F. Kun, E. Ramm, On the application of a discrete model to the fracture process of cohesive granular materials. Granul. Matter **4**, 77–90 (2002)
L. De Arcangelis, S. Redner, H.J. Hermann, A random fuse model for breaking processes. J. Phys. Lett. **46**, 585–590 (1985)
F.V. Donze, V. Richefeu, S.-A. Magnier, Advances in discrete element method applied to soil, rock and concrete mechanics. Electr. J. Geotech. Eng. **08**, 1–44 (2008)
P.M. Duxbury, P.D. Beale, P.L. Leath, Size effects of electrical breakdown in quenched random media. Phys. Rev. Lett. **57**(8), 1052–1055 (1986)
F. Family, T. Vicsek, *Dynamics of Fractal Surfaces* (World Scientific, Singapore, 1991)
S. Feng, M.F. Thorpe, E. Garboczi, Effective-medium theory of percolation on central-force elastic networks. Phys. Rev. B. **31**(1), 276–280 (1985)
R. Garcia-Molina, F. Guinea, E. Louis, Percolation in isotropic elastic media. Phys. Rev. Lett. **60**, 124–127 (1988)
D. Greenspan, *Particle Modeling* (Birkhäuser Publishing, Boston, 1997)
E. Guyon, S. Roux, A. Hansen, D. Bideaull, J.P. Troadec, H. Crapon, Non-local and non-linear problems in the mechanics of disordered systems: application to granular media and rigidity problems. Rep. Prog. Phys. **53**, 373–419 (1990)
A. Hansen, S. Roux, Statistics Toolbox for Damage and Fracture, in *Damage and Fracture of Disordered Materials*, ed. by D. Krajcinovic, J.G.M. Van Mier (Springer, Berlin/Heidelberg/New York, 2000)
A. Hansen, S. Roux, H.J. Herrmann, Rupture of central-force lattices. J. Phys. France **50**, 733–744 (1989)
H.J. Herrmann, A. Hansen, S. Roux, Fracture of disordered, elastic lattices in two dimensions. Phys. Rev. B. **39**(1), 637–648 (1989)
R. Ince, A. Arslan, B.L. Karihaloo, Lattice modeling of size effect in concrete strength. Eng. Fract. Mech. **70**(16), 2307–2320 (2003)
R.P. Jensen, P.J. Bosscher, M.E. Plesha, T.B. Edil, DEM simulation of granular media – structure interface: effects of surface roughness and particle shape. Int. J. Numer. Anal. Method Geomech. **23**, 531–547 (1999)
R.P. Jensen, M.E. Plesha, T.B. Edil, P.J. Bosscher, N.B. Kahla, DEM simulation of particle damage in granular media – structure interfaces. Int. J. Geomech. **1**(1), 21–39 (2001)

L. Jing, A review of techniques, advances and outstanding issues in numerical modelling for rock mechanics and rock engineering. Int. J. Rock Mech. Min. Sci. **40**, 283–353 (2003)
H. Kim, W.G. Buttlar, Discrete fracture modeling of asphalt concrete. Int. J. Solids Struct. **46**, 2593–2604 (2009)
D. Krajcinovic, *Damage Mechanics* (Elsevier, Amsterdam, 1996)
D. Krajcinovic, M. Basista, Rupture of central-force lattices. J. Phys. France **50**, 733–744 (1989)
D. Krajcinovic, A. Rinaldi, Thermodynamics and statistical physics of damage processes in quasi-ductile solids. Mech. Mater. **37**, 299–315 (2005a)
D. Krajcinovic, A. Rinaldi, Statistical damage mechanics – 1. Theory. J. Appl. Mech. **72**, 76–85 (2005b)
D. Krajcinovic, M. Vujosevic, Strain localization – short to long correlation length transition. Int. J. Solids. Struct. **35**(31–32), 4147–4166 (1998)
N.P. Kruyt, L. Rothenburg, A micro-mechanical definition of the strain tensor for two dimensional assemblies of particles. J. Appl. Mech. **63**, 706–711 (1996)
N.P. Kruyt, L. Rothenburg, Statistical theories for the elastic moduli of two-dimensional assemblies of granular materials. Int. J. Eng. Sci. **36**, 1127–1142 (1998)
F. Kun, H. Herrmann, A study of fragmentation processes using a discrete element method. Comput. Methods. Appl. Mech. Eng. **138**, 3–18 (1996)
F. Kun, G.A. D'Addetta, H. Herrmann, E. Ramm, Two-dimensional dynamic simulation of fracture and fragmentation of solids. Comput. Assist. Mech. Eng. Sci. **6**, 385–402 (1999)
S. Mastilovic, Some observations regarding stochasticity of dynamic response of 2D disordered brittle lattices. Int. J. Damage Mech. **20**, 267–277 (2011)
S. Mastilovic, On strain-rate sensitivity and size effect of brittle solids: transition from cooperative phenomena to microcrack nucleation. Contin. Mech. Thermodyn. **25**, 489–501 (2013)
S. Mastilovic, K. Krajcinovic, High-velocity expansion of a cavity within a brittle material. J. Mech. Phys. Solids. **47**, 577–610 (1999a)
S. Mastilovic, D. Krajcinovic, Penetration of rigid projectiles through quasi-brittle material. J. Appl. Mech. **66**, 585–592 (1999b)
S. Mastilovic, A. Rinaldi, D. Krajcinovic, Ordering effect of kinetic energy on dynamic deformation of brittle solids. Mech. Mater. **40**(4–5), 407–417 (2008)
M.J. Meisner, G.N. Frantziskonis, Multifractal fracture-toughness properties of brittle heterogeneous materials. J. Phys. B. **29**(11), 2657–2670 (1996)
L.L. Mishnaevsky Jr., Determination for the time-to-fracture of solids. Int. J. Fract. **79**(4), 341–350 (1996)
L.L. Mishnaevsky Jr., *Damage and Fracture of Heterogeneous Materials* (AA Balkema, Rotterdam, 1998)
A.A. Munjiza, E.E. Knight, E. Rougier, *Computational Mechanics of Discontinua* (Wiley, New York, 2011)
P.K.V.V. Nukala, S. Simunovic, R.T. Mills, Statistical physics of fracture: scientific discovery through high-performance computing. J. Phys. **46**, 278–291 (2006)
M. Ostoja-Starzewski, Damage in Random Microstructure: Size Effects, Fractals and Entropy Maximization, in *Mechanics Pan-America 1989*, ed. by C.R. Steele et al. (ASME Press, New York, 1989), pp. 202–213
M. Ostoja-Starzewski, *Microstructural Randomness and Scaling in Mechanics of Materials* (Taylor & Francis Group, Boca Raton, 2007)
M. Ostoja-Starzewski, J. Li, H. Joumaa, P.N. Demmie, From fractal media to continuum mechanics. Zeit. Angew. Math. Mech. (ZAMM) **93**, 1–29 (2013)
M.E. Plesha, E.C. Aifantis, On the modeling of rocks with microstructure, in *Proceedings of 24th US Symposium on Rock Mechanics*, Texas A&M University, College Station, Texas, 1983, pp. 27–39
D.O. Potyondy, P.A. Cundall, A bonded-particle model for rock. Int. J. Rock. Mech. Min. Sci. **41**, 1329–1364 (2004)
N.M. Pugno, R.S. Ruoff, Quantized fracture mechanics. Philos. Mag. **84**, 2829 (2004)

A. Rinaldi, A rational model for 2D disordered lattices under uniaxial loading. Int. J. Damage. Mech. **18**, 233–257 (2009)
A. Rinaldi, Advances in Statistical Damage Mechanics: New Modelling Strategies, in *Damage Mechanics and Micromechanics of Localized Fracture Phenomena in Inelastic Solids*, ed. by G. Voyiadjis. CISM Course Series, vol. 525 (Springer, Berlin/Heidelberg/New York, 2011)
A. Rinaldi, S. Mastilovic, D. Krajcinovic, Statistical damage mechanics – 2. Constitutive relations. J. Theor. Appl. Mech. **44**(3), 585–602 (2006)
A. Rinaldi, D. Krajcinovic, S. Mastilovic, Statistical damage mechanics and extreme value theory. Int. J. Damage. Mech. **16**(1), 57–76 (2007)
S. Roux, E. Guyon, Mechanical percolation: a small beam lattice study. J. Phys. Lett. **46**, L999–L1004 (1985)
S. Van Baars, Discrete element modelling of granular materials. Heron **41**(2), 139–157 (1996)
P.N. Sen, S. Feng, B.I. Halperin, M.F. Thorpe, Elastic Properties of Depleted Networks and Continua, in *Physics of Finely Divided Matter*, ed. by N. Boccara, M. Daoud (Springer, Berlin/Heidelberg/New York, 1985), pp. 171–179
D. Stauffer, A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, London, 1994)
V.E. Tarasov, Review of some promising fractional physical models. Int. J. Modern. Phys. **27**(9), 1330005 (2013)
J.M. Ting, A robust algorithm for ellipse-based discrete element modelling of granular materials. Comput. Geotech. **13**(3), 175–186 (1992)
V. Topin, J.-Y. Delenne, F. Radjaï, L. Brendel, F. Mabille, Strength and failure of cemented granular matter. Eur. Phys. J. E. **23**, 413–429 (2007)
V. Vitek, Pair Potentials in Atomistic Computer Simulations, in *Interatomic Potentials for Atomistic Simulations*, ed. by A.F. Voter. MRS Bulletin, vol. 21, 1996, pp. 20–23
G. Wang, M. Ostoja-Starzewski, Particle modeling of dynamic fragmentation-I: theoretical considerations. Comput. Mater. Sci. **33**, 429–442 (2005)
G. Wang, A.H.-D. Cheng, M. Ostoja-Starzewski, A. Al-Ostaz, P. Radziszewski, Hybrid lattice particle modelling approach for polymeric materials subject to high strain rate loads. Polymers **2**, 3–30 (2010)
M. Wnuk, A. Yavari, Discrete fractal fracture mechanics. Eng. Fract. Mech. **75**, 1127–1142 (2008)
J. Xiang, A. Munjiza, J.-P. Latham, R. Guises, On the validation of DEM and FEM/DEM models in 2D and 3D. Eng. Comput. **26**(6), 673–687 (2009)
S.C. Yang, S.S. Hsiau, The simulation of powders with liquid bridges in a 2D vibrated bed. Chem. Eng. Sci. **56**, 6837–6849 (2001)
R. Zhang, J. Li, Simulation on mechanical behavior of cohesive soil by distinct element method. J. Terramech. **43**, 303–316 (2006)
A. Zubelewicz, Z. Mroz, Numerical simulation of rockburst processes treated as problems of dynamic instability. Rock. Mech. Eng. **16**, 253–274 (1983)

Two-Dimensional Discrete Damage Models: Lattice and Rational Models

10

Antonio Rinaldi and Sreten Mastilovic

Contents

Abstract

Many materials exhibit a discontinuous and inhomogeneous nature on various spatial scales that can lead to complex mechanical behaviors difficult to reproduce with continuum-based models. "Among these complex phenomena is damage evolution with nucleation, propagation, interaction, and coalescence

A. Rinaldi (✉)
Materials Technical Unit, ENEA, C.R. Casaccia, Rome, Italy

Center for Mathematics and Mechanics of Complex System (MEMOCS), University of L'Aquila, Cisterna di Latina (LT) Italy
e-mail: antonio.rinaldi@enea.it

S. Mastilovic
Faculty of Construction Management, Union–Nikola Tesla University, Belgrade, Serbia
e-mail: smastilovic@fgm.edu.rs

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_22

of cracks that can result in a plethora of macroscale deformation forms." Discontinua-based models are computational methods that represent material as an assemblage of distinct elements interacting with one another. The mesoscale methods of computational mechanics of discontinua presented in this our two essays can be, arguably, divided into three broad and intervening categories: spring network (lattice) models, discrete/distinct-element methods (DEM), and particle models. The distinct-element computational methods such as molecular dynamics and smoothed-particle hydrodynamics are outside the scope of the present overview. The objective of this chapter is to briefly survey the spring network models and their main applications. The discrete-based models have been extensively applied in the last decade to three-dimensional configurations. However, since the scope of this article is limited to two-dimensional (2D) models for practical purposes, these important advances are ignored. Likewise, one-dimensional (1D) fiber bundle models are also excluded from this account.

Introduction

During the 1970s, the remarkably rapid growth of computer capabilities and the corresponding advance of numerical algorithms enabled researchers to start developing computational methods that used distinct elements such as molecules, particles, or trusses to model various problems of scientific or engineering interest. Computer simulation modeling is more flexible in application than analytical modeling and has that advantage over experimental modeling of having data accessible at any stage of the "virtual experiment." This flexibility extends to loading configurations and modeling of topological, geometrical, and structural disorder of material texture (Fig. 1). Furthermore, all discrete-element models offer some common advantages in damage analyses when compared with conventional continuum-based counterparts. Damage and its evolution are represented explicitly as broken bonds or disengaged contacts; no empirical relations are needed to define damage or to quantify its effect on material behavior. Microcracks nucleate, propagate, and coalesce into macroscopic fractures without the need for numerical artifices such as re-meshing or grid reformulation. It is unnecessary to develop constitutive laws to represent complex nonlinear behaviors since they emerge naturally through collective behavior of discrete elements governed by simple constitutive rules.

The lattice (spring network) models are the simplest models of discontinua used to simulate complex response features and fracture phenomenology of various classes of materials. As the name suggests, they are comprised of 1D discrete rheological or structural elements given geometrical, structural, and failure properties that enable them to mimic the elastic, inelastic, and failure behavior of certain class of materials. Their comparative advantages are exhibited with most distinction when the material may be naturally represented by a system of discrete elements interacting by way of rheological elements (in its basic form – springs).

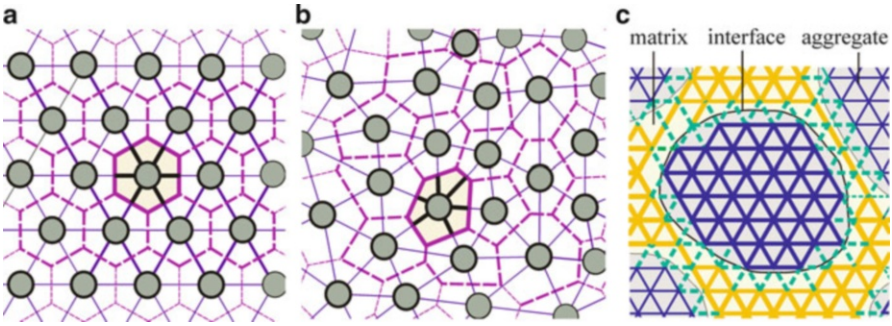


Fig. 1 (**a**) Regular and (**b**) irregular triangular Delaney network dual to the Voronoi (Wigner-Seitz) tessellation of grain boundaries. (**c**) Mesostructure of a three-phase composite projected onto a regular triangular lattice

Hence, it is not unexpected that spatial trusses and frameworks have been the primary material systems thus modeled – in engineering-mechanics applications the idea dates back, at least, to the pioneering work of Hrennikoff (1941). Comprehensive reviews of lattice models in micromechanics are presented by Ostoja-Starzewski (2002, 2007).

Lattices with Central Interactions (α-Models)

In the development of a lattice model, it is necessary to establish relationships between lattice parameters and material properties. Various approaches were proposed in that regard. Cusatis et al. (2003, 2006) used Delaunay triangulation to determine lattice connections and assign their effective cross-sectional areas. Kozicki and Tejchman (2008) derived normal and shear stiffness by using experimental coefficients. However, the prevailing approach, used herein to obtain the lattice parameters, is based on equivalence of strain energies of the unit lattice cell and its continuum counterpart (Ostoja-Starzewski 2002; Wang et al. 2009a). This approach is also presented by Kale and Ostoja-Starzewski in this handbook.

The basic idea is to ensure equivalence of the strain energy stored in a unit cell of the lattice with its associated continuum structure:

$$U_{cell} = U_{continuum}. \tag{1}$$

The energy of spatially linear displacements of the effective continuum system is given by the familiar expression

$$U_{continuum} = \frac{V}{2}\,\boldsymbol{\varepsilon}\!:\!\boldsymbol{C}\!:\!\boldsymbol{\varepsilon} = \frac{V}{2}\; C_{ijkm}\,\varepsilon_{ij}\,\varepsilon_{km}. \tag{2}$$

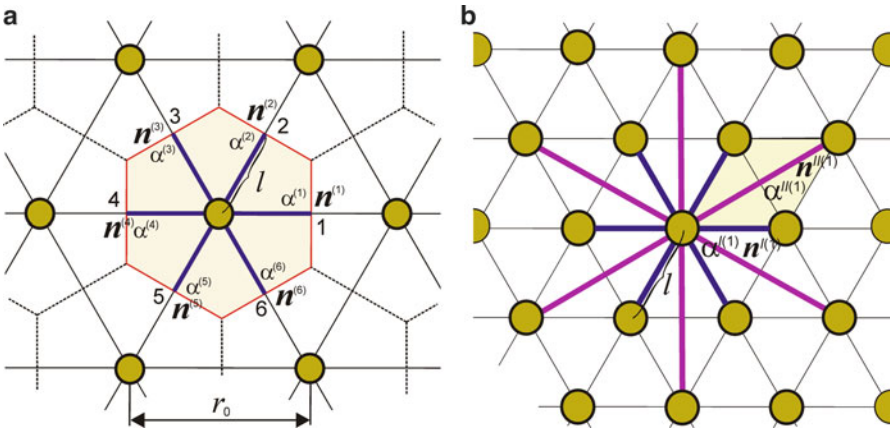


Fig. 2 An ideal triangular lattice with central interactions among (**a**) first and (**b**) first and second neighbors

In order to obtain the strain energy stored in the lattice unit cell, it is necessary to take into account its specific periodic particle arrangements and interactions (Ostoja-Starzewski 2002).

Triangular Lattice with Central Interactions

The first lattice to be considered for its simplicity is the α-model comprised of equilateral triangular lattice with central-force interactions among first-neighbor particles. The mesh illustrated in Fig. 2a is based on a spring of length l, equal to the half-length of the equilibrium interparticle distance r_0, which defines the equilibrium lattice spacing. The area of the hexagonal unit cell is $V = 2\sqrt{3}\,l^2$. Each bond b that belongs to the given unit cell is characterized by a spring constant $\alpha^{(b)}$ and bond unit vectors $\boldsymbol{n}^{(b)}$ along respective directions $\theta^{(b)} = (b - 1)\pi/3$.

The elastic strain energy stored in the hexagonal unit cell that consists of six uniformly stretched bonds that transmit only axial forces is

$$U_{cell} = \frac{1}{2}\sum_{b-1}^{6}(\alpha \boldsymbol{u}\cdot\boldsymbol{u})^{(b)} = \frac{l^2}{2}\sum_{b=1}^{6}\alpha^{(b)} n_i^{(b)} n_j^{(b)} n_k^{(b)} n_m^{(b)}\,\varepsilon_{ij}\,\varepsilon_{km}. \tag{3}$$

The crucial step in this procedure is making a connection between $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$, which, in general, depends on the particular geometry of the lattice cell and particular model of interparticular interaction.

By Eq. 1, the stiffness tensor component can be derived as

$$C_{ijkm} = \frac{1}{2\sqrt{3}}\sum_{b=1}^{6}\alpha^{(b)} n_i^{(b)} n_j^{(b)} n_k^{(b)} n_m^{(b)} \tag{4}$$

which in the case of equal spring constants, $\alpha^{(b)} = \alpha$ ($b = 1, \ldots, 6$), results in

$$C_{1111} = C_{2222} = \frac{9}{8\sqrt{3}}\,\alpha = \frac{E}{1-\upsilon^2}, \quad C_{1122} = C_{2211} = \frac{3}{8\sqrt{3}}\alpha = \frac{E\upsilon}{1-\upsilon^2},$$

$$C_{1212} = \frac{3}{8\sqrt{3}}\,\alpha = \frac{E}{2\,(1+\upsilon)}. \tag{5}$$

It should be noticed that the condition for material isotropy

$$C_{1212} = (C_{1111} - C_{1122})/2 \tag{6}$$

is satisfied. Since the Poisson's ratio is fixed, the spring constant α defines only the planar modulus of elasticity of the unit cell of this lattice model

$$E = \alpha/\sqrt{3}, \quad \upsilon = C_{1122}/C_{1111} = 1/3. \tag{7}$$

Triangular Lattice with First and Second Neighbor Central Interactions

The previous triangular central-force lattice can be upgraded by superposing an additional central-force structure (Fig. 2b). The original structure (I) is now represented by three triangular networks with unit defined as

$$\alpha^{I(b)} = \alpha^I, \;\; \theta^{I(b)} = (b-1)\frac{\pi}{3}, \quad \boldsymbol{n}^{I(b)} = \left(cos\,\theta^{I(b)}, sin\,\theta^{I(b)}\right) \quad b = 1,2,3 \tag{8}$$

and the lattice spacing $r_0^I = 2l$. The superposed structure (II) is represented by three triangular networks with the following spring constants:

$$\alpha^{II(b)} = \alpha^{II}, \;\; \theta^{II(b)} = (2b-1)\frac{\pi}{6}, \quad \boldsymbol{n}^{II(b)} = \left(cos\,\theta^{II(b)}, sin\,\theta^{II(b)}\right), \quad b = 1,2,3 \tag{9}$$

and the lattice spacing $r_0^{II} = 2\sqrt{3}\,l$. In the resulting system, each particle communicates with six first neighbors by means of structure I and with six second neighbors by means of structure II. The unit cell area is $V = 2\sqrt{3}\,l^2$.

Under the condition of uniform strain, the equivalence of the strain energy stored in the lattice unit cell and the corresponding effective continuum model results in

$$C_{ijkm} = \frac{2}{\sqrt{3}}\,\alpha^I \sum_{b=1}^{3} n_i^{I(b)}\, n_j^{I(b)}\, n_k^{I(b)}\, n_m^{I(b)} \;+\; \frac{6}{\sqrt{3}}\,\alpha^{II} \sum_{b=1}^{3} n_i^{II(b)}\, n_j^{II(b)}\, n_k^{II(b)}\, n_m^{II(b)}. \tag{10}$$

Thus, the nonzero stiffness components are

$$C_{1111} = C_{2222} = \frac{3}{4\sqrt{3}}\left(3\alpha^I + 9\alpha^{II}\right) = \frac{E}{1-\upsilon^2}, \;\; C_{1212} = \frac{3}{4\sqrt{3}}\left(\alpha^I + 3\alpha^{II}\right) = \frac{E}{2\,(1+\upsilon)}$$

$$C_{1122} = C_{2211} = \frac{3}{4\sqrt{3}}\left(\alpha^I + 3\alpha^{II}\right) = \frac{E\upsilon}{1-\upsilon^2}. \tag{11}$$

The Poisson's ratio is again independent of the spring constant

$$E = 2\left(\alpha^{I} + 3\alpha^{II}\right)/\sqrt{3}, \quad \upsilon = C_{1122}/C_{1111} = 1/3. \tag{12}$$

The expression $(12)_1$ is reduced to $E = 8\alpha/\sqrt{3}$ if $\alpha^{I} = \alpha^{II} = \alpha$.

Examples of Applications of α-Model

Bažant et al. (1990) employed a random α-model for brittle heterogeneous materials with the aim to study the effects of specimen size on the maximum load, the post-peak softening behavior, and the progressive damage spread of the microcracking zone. The model, accounting for particle interactions and random geometry, may be classified as a particle model, but since the shear and bending interaction among particles in contact are neglected, it represents an illustrative example of an application of random lattice with central interactions. This model was very influential for further developments of DEM techniques for this class of materials. A more refined model of Zubelewicz and Bažant (1987) accounted for shear interactions as well.

The model was based on the central interaction of elastic circular particles (aggregates) embedded randomly in a softer matrix (cement paste). The matrix was initially elastic with modulus of elasticity E_m, while the modulus of elasticity of the aggregates was designated by E_a. The particles (denoted as i and j) interacted by means of a truss connecting particle centers whose stiffness, $S = (S_i^{-1} + S_m^{-1} + S_j^{-1})^{-1}$, was defined by the serial connection of the three truss segments (Fig. 3a). The individual segment stiffness was determined by the standard truss theory (e.g., $S_m = E_m A_m/L_m$) taking into account the empirically modifies truss segment lengths corresponding to aggregate particles in contacts. The middle truss segment (mimicking cement paste), L_m, represented the contact region of the matrix and was assumed to exhibit softening behavior depicted by the triangular constitutive law (Fig. 3b). This softening behavior

$$E_s = -\frac{f_t^m}{\varepsilon_f - \varepsilon_p}, \quad \varepsilon_f = \frac{2G_f^m}{L_m\, f_t^m} \tag{13}$$

was determined based on the fracture energy of the interparticle layer, G_f^m, which was assumed to be an intrinsic material property. The geometrical lattice disorder (random L_m) necessitated corresponding changes of the softening modulus E_s in order to preserve the fracture energy.

The simulation results of this model revealed pronounced size effect on the failure load, which is a salient consequence of heterogeneity that has both fracture mechanics and probabilistic aspects. The load–displacement curves, representing the response of unnotched samples under uniaxial tension, revealed that the softening was captured realistically since the slope got steeper with the sample size increase. In addition to pronounced size effect, the simulation results exhibited substantial data scatter as well. The revealed size effect on the maximum normal

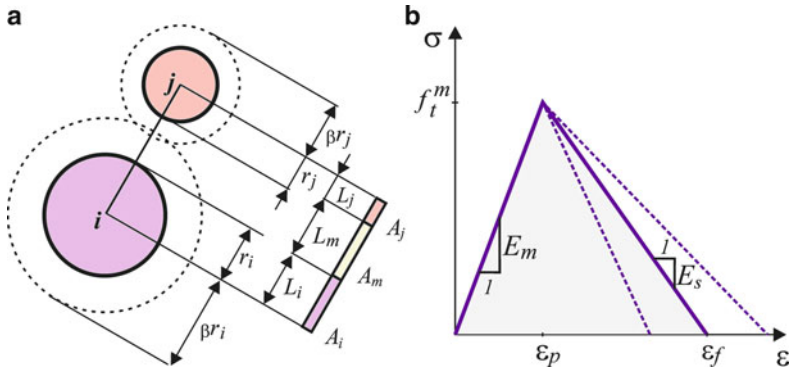


Fig. 3 The α-model of Bažant et al. (1990): (**a**) Particles and truss parameters and (**b**) constitutive law for the binding matrix

stress (nominal strength) was in contrast to predictions of local continuum models. The corresponding simulation data were fitted with the sample diameter, d, in accordance with size effect law proposed by Bažant, $\sigma_N \propto d^{-1/2}$. The same effect was observed for notched specimen as well. It was also demonstrated that the size effect observed in simulations is intermediate between the strength criterion and the linear elastic fracture mechanics. In agreement with laboratory tests, the results of uniaxial tension simulations revealed development of asymmetric response in the softening region. The spread of cracking and its localization observed experimentally in quasibrittle materials was captured reasonably well.

The extension of this simulation technique was used by Jirásek and Bažant (1995) to determine relationship between the macroscopic failure properties (the fracture energy and the size of the effective process zone) and the statistics of microscopic properties (such as microstrength, microductility, and average interparticle distance of particle links). These simulation results revealed that realistic modeling, especially under far-field compression loading conditions, required the lattice elements to be capable of not only central interactions but also shear (angular) interactions.

Vogel et al. (2005) utilized a topologically and geometrically ordered lattice with the interparticle interactions based on Hookean springs with randomly sampled finite tensile strength. Thus, each node of triangular lattice was connected to all six first neighbors (the coordination number $z = 6$), each link had an equal equilibrium length and spring constant, while the microstrength was defined by random sampling of the link rupture strains from a normal distribution. The model aimed to capture the underlying physical processes involved in the clay soil crack formation in the course of desiccation. The corresponding load was due to slow contraction of the sample as a result of water evaporation.

The shrinkage of a pseudo 2D clay surface due to desiccation was simulated by successive reduction of the natural spring length that caused contracting forces and increase of total energy in the lattice. As soon as the strain between two consecutive

nodes reached the critical value, the spring ruptured and the corresponding released strain energy had to be redistributed among the adjacent links in the system transition toward the new equilibrium state. The change in nodal position depended on the total force exerted on the node but the node moved only if the nodal force exceeded a static adhesion limit. Heterogeneity was introduced into the system through the random sampling of the spring rupture strain from a Gaussian probability distribution $N(\bar{\varepsilon}_{cr}, \sigma^2)$. Thus, the model parameters were the mean critical strain $\bar{\varepsilon}_{cr}$, its variance σ^2, and the friction μ.

An additional parameter results from iterative relaxation of the lattice, which was the maximum number of iterations, n_{it}, performed after each spring rupture. Depending on the n_{it}, the lattice might not be relaxed completely prior to the breakage of the next spring. Thus, the relaxation parameter was more than just an arbitrarily selected simulation parameter since it may be related to the speed of desiccation. Vogel et al. (2005) claimed that n_{it} might be interpreted as a dimensionless quantity relating the characteristic times of external forcing, t_{ext}, to the characteristic times of the internal dynamics t_{int} by $n_{it} = t_{ext}/t_{int}$.

The presented triangular α-model reproduced prominent features of the nonlinear dynamics of the desiccated clay crack network development observable in nature such as characteristic aggregate shapes and angles of bifurcations (Fig. 4). The authors quantitatively verified that the model reproduced both the characteristic features of natural crack patterns and the characteristic pattern evolution dynamics. The model parameters could be related to physical properties of the material and to the boundary conditions during shrinkage by desiccation and crack formation.

Topin et al. (2007) used discretization based on a sub-particle triangular α-model to analyze strength and damage behavior of cemented granular materials as a function of the matrix volume fraction (structural parameter) and the particle–matrix adhesion (material parameter). The objective was to elucidate roles of those parameters for breaking characteristics (stiffness, tensile strength), damage growth (stiffness degradation, particle fracture), and stress transmission (statistical distributions, phase stresses).

This α-model consisted of linear elastic–brittle springs defined by a spring constant and a rupture-force threshold. The high connectivity of the lattice nodes ensured global resistance in shear and distortion since the springs transmitted only normal forces. There were five distinct link types representing three bulk phases (granule, matrix, and void) and two interface phases (granule–granule and granule–matrix). The lattice with free lateral boundaries was loaded alternatively in uniaxial tension and compression by displacement application on the upper sample edge.

The stress–strain curves revealed the expected asymmetry between tension and compression, related to existence of preexisting "fabrication" damage (reflected by non-cohesive interparticle contacts – bare contacts). The post-peak behavior was characterized by nonlinear propagation of the main crack and the progressive reduction of the effective stiffness due to the damage accumulation in the

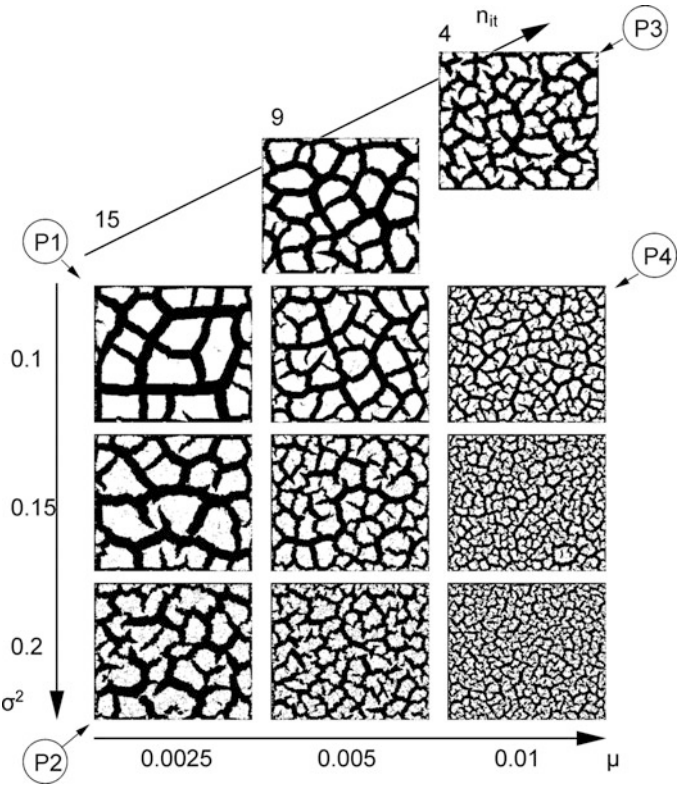


Fig. 4 Damage patterns obtained for various critical microstrain data dispersions (σ^2), nodal frictions (μ), and relaxation intensities (n_{it}) (Reprinted from Vogel et al. 2005 with permission from Elsevier)

heterogeneous material. Maps of vertical stress fields in tension and compression were used to study the effect of jamming of the particles that resulted in stress concentration along particle chains (jammed backbone). Influence of the volume fraction of binding matrix on the effective stiffness in tension and compression was discussed and the model results were compared with the Mori–Tanaka theoretical predictions for the effective stiffness of three-phase composite. It was shown that effective stiffness in tension increases linearly with the matrix volume fraction due to the space-filling role and disappearance of bare contacts (voids). The surface and bulk effects of binding matrix (reflecting the reduction of bare contacts and porosity, respectively) were discussed.

The crack patterns in tension and compression (Fig. 5) in the neighborhood of the homogeneous–heterogeneous damage phase transition revealed diffuse and localized cracking occurring mostly in pre-peak and post-peak regimes, respectively. Since the particle–matrix interfacial strength was inferior in comparison with the individual particle and matrix strengths, the cracking almost traced particle

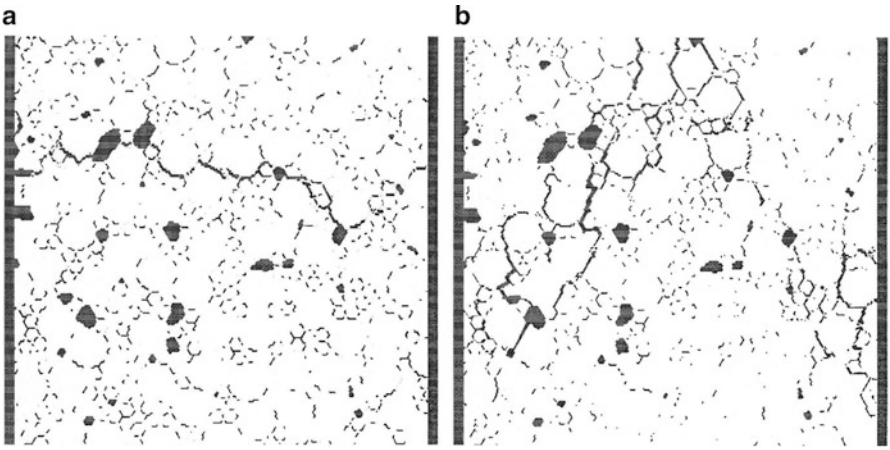


Fig. 5 Damage pattern: (**a**) tension and (**b**) compression (Topin et al. 2007)

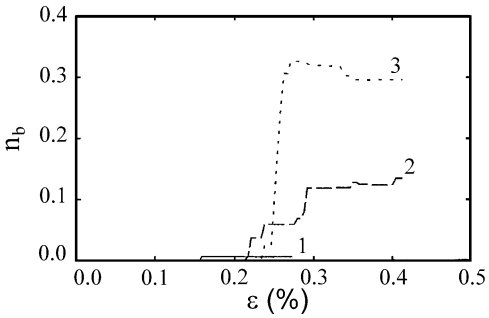


Fig. 6 Evolution of the fraction of broken bonds in the particle phase for (*1*) $\sigma^{pm}/\sigma^{[p]} = 0.6$ and $\rho^m = 0.08$; (*2*) $\sigma^{pm}/\sigma^{[p]} = 1$and $\rho^m = 0.08$; and (*3*) $\sigma^{pm}/\sigma^{[p]} = 1$ and $\rho^m = 0.18$ (Topin et al. 2007)

contours. The evolution of the main crack was abrupt in tension and practically perpendicular to the applied load direction; in compression the main cracking paths were inclined and thicker and involve secondary crack branches. It can be observed in Fig. 5 that the lattice geometry controlled crack propagation and the damage maps reflected the regular structure of the underlying lattice. The damage evolution was accompanied by stiffness degradation that could be observed in stress–strain curves.

The abrupt effective stiffness degradation in tension was vividly contrasted to the more progressive nature of damage evolution in compression, which reflected distinction in the nature of damage accumulation and cracking patterns. The authors found that the matrix volume fraction and the particle–matrix adherence played nearly the same role in the tensile strength of cemented granular materials. On the other hand, the two parameters controlled differently the damage characteristics as reflected by the fraction of broken bonds in the particle phase just after the sample failure (Fig. 6).

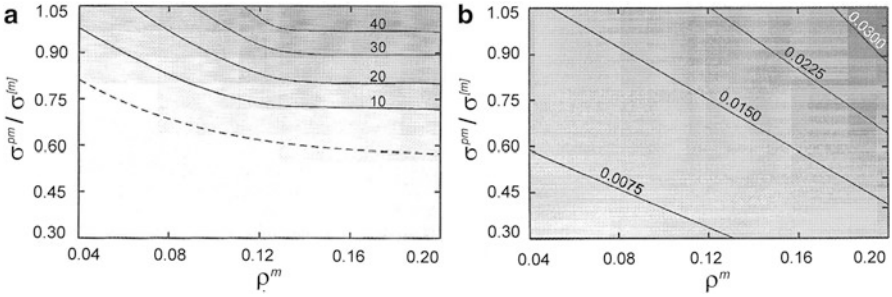


Fig. 7 Parametric maps of: (**a**) tensile strength and (**b**) fraction of broken particle bonds at macrofailure in the particle–matrix adherence versus the matrix volume fraction space (Topin et al. 2007)

The authors observed that a bilinear boundary limited the parametric space in which particle damage occurred. Thus, for that range of parametric values, the cracks propagated either in matrix or along interface. According to Fig. 7b, the particle damage was more sensitive to interphase strength than to the matrix volume fraction. It was also noted that that the matrix volume fraction $\rho_m = 0.12$, representing the limit of influence on n_b, reflected the percolation threshold of the binding matrix, with particles covered entirely except at the bare contacts. Finally, the authors suggested that the particle damage limit was controlled by a single parameter: the relative toughness of the particle–matrix interface.

Hou (2007) improved the triangular spring network with central interactions and hexagonal unit cell by introducing large-strain elasticity into the modeling framework. The large-strain α-model was used to simulate several representative problems of large-strain elasticity: a square planar sample under uniform uniaxial tension, a wedge loaded in tension by a force, and a planar sample with a preexisting crack under mode I loading. The comparison between analytical and lattice simulation results revealed an exceptional agreement and also demonstrated that the large-strain lattice model can capture large deformation singularity rather well.

Assumed failure criterion was used to describe the fracture process of large-strain elasticity and large-strain composite. As the lattice deformation increased, the individual bond extensions increased as well until rupture criterion was met in one or a few springs, which was/were then removed from the network. The spring rupture criterion could be defined in many ways, e.g., in terms of the critical values

$$f_b = f_{cr}, \quad \varepsilon_b = \varepsilon_{cr}, \quad E_b = E_{cr} \tag{14}$$

of spring: force, extension, and strain energy, respectively. The author opted for the link strength criterion $(14)_1$. This process progressed until the lattice loosed completely the load carrying capability (cracked percolation). Two edge-crack mode I loading configurations were simulated until global large-strain failure.

Rational Models of Brittle Materials

Rational models of damage that connect microstructural material properties to overall macroscale properties by means of "exact" constitutive models (as closed-form solutions, where possible) are very desirable for scientific and technological purposes, albeit usually available only for 1D mechanical systems such as fiber bundle models (e.g., Rinaldi 2011a; Phoenix and Beyerlein 2000). Lattices provide a powerful option to address rational approaches to damage in real materials with heterogeneous microstructures, which represent an inherently more complex and higher-dimensional problem. One such 2D α-model lattice is discussed next in some details (after Rinaldi and Lai (2007) and Rinaldi (2009)).

Let us consider the perfect central-force triangular lattice (Fig. 2a) with links of equal stiffness k and length ℓ_o. To introduce some mechanical disorder, suppose further that each b-th link breaks irreversibly at a tensile critical strain $\varepsilon^{*(b)} = \boldsymbol{u}^{(b)}/\ell_o$ (Eq. 14)$_2$. For the results reported herein, simulation parameters were $k = 100$, $\ell_o = 1$, and $\varepsilon^{*(b)}$ randomly sampled from a uniform distribution in the interval $[0, 10^{-2}]$.

On the macroscale, the stress–strain response of this quasibrittle system under tensile loading (or equivalently uniaxial compression) is expressed after L. Kachanov's relation in scalar form as

$$\overline{\sigma} = \overline{K}_0\left[1 - \overline{D}(\overline{\varepsilon}_{MAX})\right]\overline{\varepsilon} \tag{15}$$

which accounts for the (permanent) loss of secant stiffness $\Delta\overline{K}_0 = \overline{K}_0 \cdot \overline{D}$ associated with the cracking process. The parameter $\overline{K}_0$ is the secant stiffness in the pristine state and the damage process is measured by the scalar macroscopic damage parameter $\overline{D}$ ranging from 0 (pristine state) to 1 (failure). Assuming no damage healing, $\overline{D}$ is a nondecreasing function of applied strain and, thus, depends on the maximum tensile strain $\overline{\varepsilon}_{MAX}(t)$ reached at time t, such that

$$\overline{\varepsilon}_{MAX}(t) = \max\{\overline{\varepsilon}(t_0),\ \forall t_0 \le t\} \quad \text{and} \quad \overline{D} = \int_0^{\overline{\varepsilon}_{MAX}} d\overline{D}(\overline{\varepsilon}). \tag{16}$$

Correspondingly, the strain energy computed from the "top" is

$$U(\overline{\varepsilon}) = \frac{1}{2}\overline{\sigma}\,\overline{\varepsilon} = \frac{1}{2}\overline{K}_0\left[1 - \overline{D}\,(\overline{\varepsilon}_{MAX})\right]\overline{\varepsilon}^2. \tag{17}$$

To solve this problem for a given loading history, $\overline{D}$ must be computed and linked to the microcracking process in the microstructure. From the knowledge of microstructure and full field microstrains, the strain energy of the lattice can in fact be recomputed from the "bottom" by summing all unbroken links over all unit cells as

$$U(\overline{\varepsilon}) = \sum^{Cells}\left(\frac{1}{2}\sum_{b}^{N_b}(k\,\boldsymbol{u}\cdot\boldsymbol{u})^{(b)}\right). \tag{18}$$

Brittle damage evolves from two types of dissipative events, either an individual rupture or an avalanche, i.e., a cascade of distinct individual ruptures initiated at a

random value $\overline{\varepsilon}$ (Rinaldi and Lai 2007). When one spring is suppressed, it entirely releases its stored energy ΔU_1

$$\Delta U_1 = \frac{1}{2} k (\varepsilon^*)^2 \tag{19}$$

and causes a macroscale loss of strain energy

$$\Delta U(\overline{\varepsilon}) = \frac{1}{2} \overline{K}_0 \Delta \overline{D}(\overline{\varepsilon}) \overline{\varepsilon}^2 = \frac{1}{2} \Delta \overline{K}(\overline{\varepsilon}) \overline{\varepsilon}^2 \tag{20}$$

that is equal or greater than ΔU_1 and reflects cooperative phenomena, snapback instability (for an avalanche), and load redistribution within the lattice. Such an effect can be conveyed by a "redistribution parameter"

$$\eta_p = \frac{\Delta U - \Delta U_1}{\Delta U_1} \tag{21}$$

which is always $\eta_p > -1$ and is null only when there is no redistribution effect (as in many 1D models). Then $\overline{D}(\overline{\varepsilon})$ is obtained from Eqs. 16, 17, 18, 19, and 20 by summing normalized stiffness decrements $\Delta \overline{K}_p$ from each microcrack

$$\overline{D}(\overline{\varepsilon}) = \frac{\sum_{p=1}^{n(\overline{\varepsilon})} \Delta \overline{K}_p}{\overline{K}_0} = \frac{k}{\overline{K}_0} \left(\frac{\ell_0}{L} \right)^2 \sum_{p=1}^{n(\overline{\varepsilon})} (1 + \eta_p) \left(\frac{\varepsilon_p^*}{\overline{\varepsilon}^*} \right)^2 . \tag{22}$$

This *stochastic* model requires the three random input parameters $\{\varepsilon_p^*, \eta_p, n_p\}$ – three distinct sources of variability:

- $\varepsilon_p^*(\overline{\varepsilon})$, the critical strain of the spring failing at $\overline{\varepsilon}$, linking the macroscale kinematics to the microscale kinematics and depending on the chosen sampling distribution
- $\eta_p(\overline{\varepsilon})$, the redistribution parameter related to the local load redistribution capability of the microstructure and to snapback effects (for an avalanche)
- $n_p(\overline{\varepsilon})$, the number of broken links

This rational model renders an "exact" estimate of the lattice response to a uniaxial loading but requires $\{\varepsilon_p^*, \eta_p, n_p\}$ to be known.

If the springs are partitioned by orientation $\theta = \{0, 60°, -60°\}$, damage can be conveniently broken down in a "spectral form"

$$\overline{D}(\overline{\varepsilon}) = \overline{D}_1(\overline{\varepsilon}) + \overline{D}_2(\overline{\varepsilon}) + \overline{D}_3(\overline{\varepsilon}) = \frac{k}{\overline{K}_0} \left(\frac{\ell_0}{L} \right)^2 \sum_{j=1}^{3} \sum_{p=1}^{nj(\overline{\varepsilon})} (1 + \eta_p) \left(\frac{\varepsilon_p^*}{\overline{\varepsilon}^*} \right)^2 . \tag{23}$$

For validation sake, Fig. 8 displays input and output data, i.e., $\{\varepsilon_p^*, \eta_p, n_p\}$ and $\{\overline{\sigma}, \overline{D}\}$ respectively, from the simulation of a random lattice under tensile test,

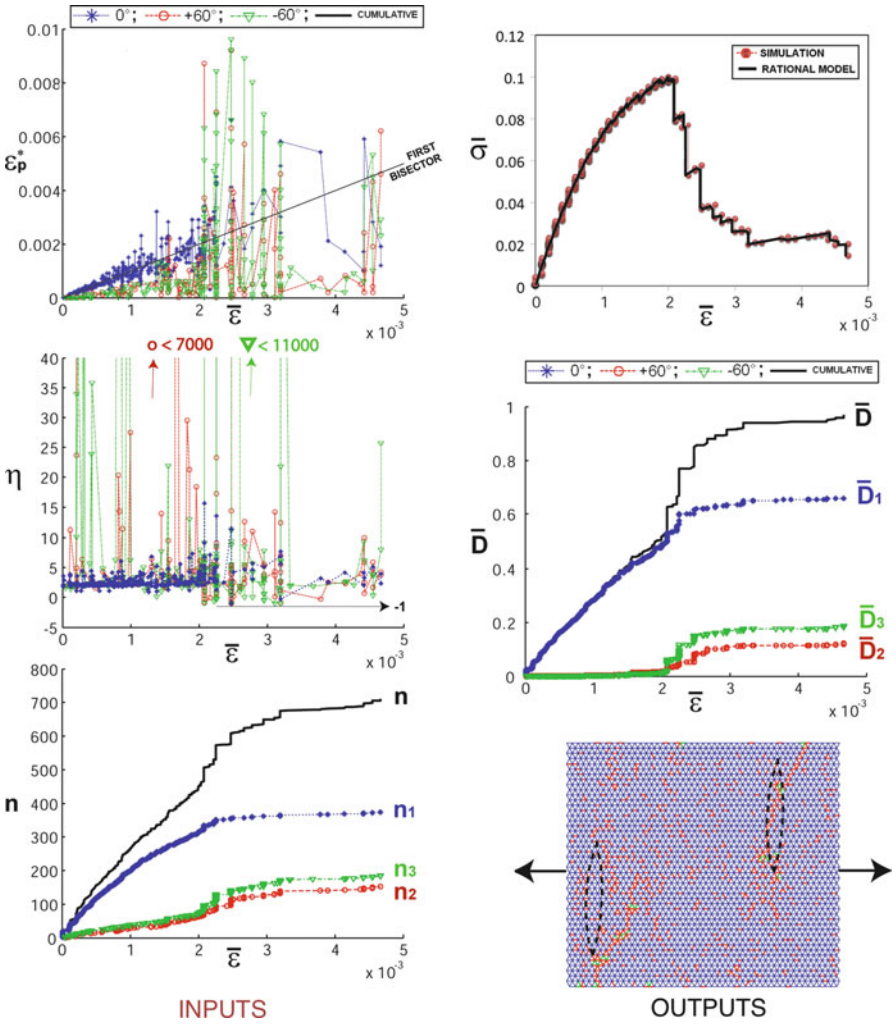


Fig. 8 Input (*left*) and output (*right*) data from a lattice simulation under tension (Rinaldi and Placidi 2013)

making self-evident the random nature of input data in Eq. 23. The data markers $\{*, \mathrm{O}, \nabla\}$ for $\theta = \{0, 60°, -60°\}$ highlight differences in damage process between different orientations.

The output data plotted aside on the right-hand side display the stress response estimated from Eq. 15 marked by "*", which overlaps the actual stress–strain curve from simulation (solid bold line) such that the two data series are indeed indistinguishable (i.e., zero error as expected in a rational theory). The "staircase" damage function $\overline{D}(\overline{\varepsilon})$ (solid bold line) with components $\overline{D}_i(\overline{\varepsilon})$ from solid line with markers is also shown.

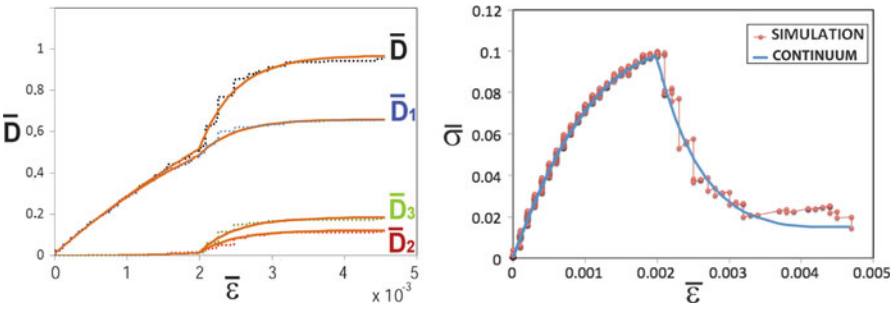


Fig. 9 (*left*) Regression approximation of exact damage components from spectral decomposition in Fig. 8 (*right*). Comparison of resulting equivalent first-order damage model versus simulation data (Rinaldi 2013)

This example clarifies that $\overline{D}(\overline{\varepsilon})$ (the measure of the *global effect of microcracks*) and $n(\overline{\varepsilon})$ (the *number of microcracks*) are related in an intricate manner in the lattice and most real materials, such that their increments are linked as

$$\Delta\overline{D}_p(\overline{\varepsilon}) = w_p(\overline{\varepsilon})\,\Delta n(\overline{\varepsilon}) \quad (\text{read } \Delta n = 1) \tag{24}$$

by means of a stochastic weight function

$$w_p(\overline{\varepsilon}) = c\left(1+\eta_p\right)\left(\frac{\varepsilon_p^*}{\overline{\varepsilon}^*}\right)^2, \quad c = k/\overline{K}_0(\ell_0/L)^2. \tag{25}$$

When comparing the 2D lattice to the corresponding 1D parallel fiber bundle model, the constitutive model differs only in the definition of $\overline{D}(\overline{\varepsilon})$

$$\overline{D}(\overline{\varepsilon}) = \int_0^{\overline{\varepsilon}} p_f(\overline{\varepsilon})d\overline{\varepsilon} = \frac{n(\overline{\varepsilon})}{N} \tag{26}$$

where $\overline{D} \propto n(\overline{\varepsilon})$ and the weight function $(25)_1$ reduces to a constant $w_p = 1/N$ due to the absence of between-links interaction that allows analytical computation of $\overline{D}$ from the sole knowledge of the distribution of critical strains $p_f(\varepsilon)$ (Rinaldi 2011a, b).

In the lattice case, the knowledge of its rational damage model (22) is a significant advance as it enables the derivation of several physically based approximate continuum solutions from the analysis of the input random fields. The potential of this bottom-up approach has been discussed in the original papers as well as in subsequent work dedicated to equivalent continuum models, both of first order (Rinaldi 2013) and second order (Misra and Chang 1993; Alibert et al. 2003; Rinaldi and Placidi 2013). Figure 9 displays the first-order continuum model for the case in Fig. 8, obtained by estimating the damage parameter with regression functions and deducing the (macro) Helmholtz function matching the microscale physics.

The theory can be generalized from the 2D α-model to real quasibrittle materials as it fully captures the fundamental physics of the damage nucleation and evolution of real systems. Statistical models such as Eq. 22 have the potential for engineering application, for example, in conjunction to acoustic emission tests for health monitoring. A connection with the diverse approaches to deal with size effects (e.g., Rinaldi et al. 2006, 2007) from the next chapter also has to be developed for this purpose.

Lattices with Central and Angular Interactions (α–β Models)

The lattice with central and angular interactions is developed by augmenting of the α-model by angular springs acting between contiguous bonds incident into the same particle (Ostoja-Starzewski 2002; Wang et al. 2009a).

Triangular α–β model is discussed in detail by Kale and Ostoja-Starzewski in this handbook.

Square Lattice with Central and Angular Interactions

For the square α–β model, the particle at the center of the square unit cell interacts with four first neighbors and four second (diagonal) neighbors, whose half-length spring constants are α^I and α^{II}, respectively. The volume of square unit cell is $V = 4l^2$. The bond angles $\theta^{(b)}$ and corresponding unit vectors $\boldsymbol{n}^{(b)}$ are

$$\alpha^{I(b)} = \alpha^I, \quad \theta^{I(b)} = (b-1)\frac{\pi}{2}, \quad \boldsymbol{n}^{I(b)} = \left(cos\,\theta^{I(b)}, sin\,\theta^{I(b)}\right), \quad b = 1,2,3,4$$

$$\alpha^{I(b)} = \alpha^I, \quad \theta^{II(b)} = (2b-9)\frac{\pi}{4}, \quad \boldsymbol{n}^{II(b)} = \left(cos\,\theta^{II(b)}, sin\,\theta^{II(b)}\right), \quad b = 5,6,7,8. \quad (27)$$

For the sake of simplicity the spring constants of all bonds are, henceforth, assumed to be $\alpha^{I(b)} = \alpha^I$, $\alpha^{II(b)} = \alpha^{II}$, and $\beta^{(b)} = \beta$.

After performing the derivation process analogous to the one outlined in section "Triangular Lattice with First and Second Neighbor Central Interactions," the nonzero components of the effective stiffness tensor are obtained as follows:

$$C_{1111} = C_{2222} = \frac{1}{2}\alpha^I + \alpha^{II} = \frac{E}{1-\upsilon^2}$$

$$C_{1122} = C_{2211} = \alpha^{II} = \frac{E\upsilon}{1-\upsilon^2}, \quad C_{1212} = \alpha^{II} + \frac{\beta}{l^2} = \frac{E}{2(1+\upsilon)}. \quad (28)$$

Expressions (28) indicate that the angular interactions affect only the shear modulus (C_{1212}). The Poisson's ratio range is, consequently, extended due to the effect of angular springs on the shear modulus. Thus, the spring constants are

$$\alpha^I = \frac{2E}{1+\upsilon}, \quad \alpha^{II} = \frac{E\upsilon}{1-\upsilon^2}, \quad \beta = \frac{(1-3\upsilon)El^2}{2(1-\upsilon^2)}. \tag{29}$$

By definition, the plane-strain elasticity coefficients are

$$E = \frac{\alpha^I}{2}\left(\frac{\alpha^I + 4\alpha^{II}}{\alpha^I + 2\alpha^{II}}\right), \quad \upsilon = \frac{C_{1122}}{C_{1111}} = \frac{2\alpha^{II}}{\alpha^I + 2\alpha^{II}}. \tag{30}$$

Expressions (28) and (29) indicate the complete range of Poisson's ratio $-1 < \upsilon \leq 1/3$ is identical to that obtained for the triangular α–β model. Substitution of Eqs. 29 and 23 into the preceding inequality yields a restricted range of the axial and diagonal α-spring ratios $\alpha^{II}/\alpha^I \leq 1/4$, which is an extension of the corresponding requirement physically imposed on the rectangular lattices with central interactions.

Examples of Applications of Lattices with Central and Angular Interactions

Grah et al. (1996) provided an illustrative example of thoughtful application of the triangular lattice with central and angular interactions by simulating brittle intergranular fracture of gallium embrittled sheets of polycrystalline aluminum. The authors both manufactured aluminum sheets and performed quasistatic biaxial tension test in order to have experimental data for lattice model validation. The computation domain was discretized by a geometrically ordered triangular lattice on the sub-grain scale. The axial-spring constants and strengths were assigned depending on their location within the sample, that is, according to the grain that given bonds were associated with in comparison with actual micrographs. The interface bonds (that straddle the boundary of two crystals) have their axial-spring constants assigned in accordance with the serial spring connection rule, weighted by their respective partial lengths

$$\alpha = \left(\frac{l_1}{l\alpha_1} + \frac{l_2}{l\alpha_2}\right)^{-1}, \quad l = l_1 + l_2. \tag{31}$$

The strength reduction factor of 0.01 was applied to each interface α-spring to account for the preferential gallium embrittlement along the grain boundaries. Assignment of angular spring constants naturally did not present such complexities.

The rectangular lattices was loaded quasistatically in biaxial tension by applying the controlled displacement $u_i = \overline{\varepsilon}_{ij}\, x_j = \overline{\varepsilon}\delta_{ij}\, x_j$ on each boundary node, defined in terms of macroscopic strain $\overline{\varepsilon}_{ij} = \overline{\varepsilon}\delta_{ij}$ (δ is the Kronecker symbol). Thus, all boundaries experienced only mode I loading prior to sample failure. As the lattice deformation increased, the individual link extensions increased as well until critical link force criterion $(14)_1$ was met in some link, which was then removed from the network. The load increase by $\Delta\overline{\varepsilon}_{ij}$ was conducted first by unloading the entire

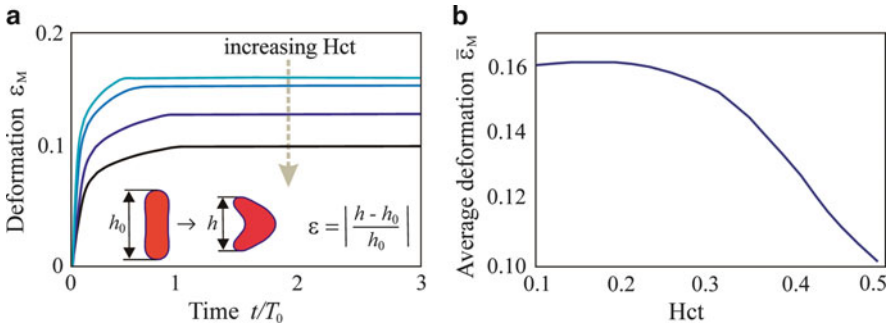


Fig. 10 Schematics of change of (**a**) the deformation index in middle part of flow channel, ε_M, for four Hct values (0.1, 0.3, 0.4, 0.49), and (**b**) the average value of deformation index, $\bar{\varepsilon}_M$, over the time interval $t/T_0 \in [1, 3]$ (i.e., the upper plateau) as function of Hct (Sketched after Tsubota et al. 2006). The characteristic time used for normalization was $T_0 = v_0/L$, where $v_0 = 0.011$ m/s was a constant uniform velocity at the capillary channel inlet, while $L = 90$ μm was the capillary length

lattice and then reloading it by $\bar{\varepsilon}_{ij} + \Delta\bar{\varepsilon}_{ij}$. The process was repeated until the damage percolated the lattice and global failure took place (cracked percolation).

Computer simulations of biaxial loading of 2D sample of a brittle polycrystal with inferior grain boundary strength produced cracking patterns that matched very well to the one obtained from the actual experiments. The departures from the experimental results were discussed in detail.

Tsubota et al. (2006) used computer simulations of discontinua to investigate the red blood cell (RBC) microcirculation in viscous plasma. The triangular spring network with central and angular interactions was used to model deformable RBC membrane. The α–β model (extended subsequently to 3D problems by Tsubota and Wada (2010)) consisted of RBC membrane particles with central and angular interaction representing elastic response of the membrane structure during incompressible viscous flow. The relevant parameter whose effect on the blood flow in microcirculation was investigated was hematocrit (Hct), defined as the volumetric ratio of RBCs to whole blood, which influences greatly the blood rheological characteristics. The Hct effect on blood flow resistance was of obvious interest as well as the mechanical factors, such as the deformation and shape of RBC, and mechanical interaction of RBCs and plasma.

The total elastic strain energy of the RBC membrane was

$$E = E_\alpha + E_\beta = \frac{1}{2}\alpha\sum(\lambda - 1)^2 + \frac{1}{2}\beta\sum \tan^2\left(\frac{\theta - \theta_0}{2}\right) \tag{32}$$

where α and β were the axial and angular spring constants, respectively, $\lambda = l/l_0$ the bond stretch ratio, and θ and θ_0 the current and initial angles between consecutive bonds. Based on the principle of virtual work, the force acting on the RBC membrane particle i was $\boldsymbol{F}_i = -\partial E/\partial \boldsymbol{r}_i$ where $\boldsymbol{r}_i$ designated the particle position vector.

Figure 10a presents the definition of a deformation index, ε, and schematics of time history of its mean value, ε_M, for four different Hct values. The curves show

that ε_M monotonically increased until an upper plateau was reached at a Hct-dependent saturation time. The effect of Hct on the average ε_M value over the time interval $t/T_0 \in [1, 3]$ (illustrated schematically by the upper plateaus in Fig. 10a) is depicted by Fig. 10b. The results obtained by the spring network model agreed very well with in vitro experimental observation. The authors illustrated the time history of RBC membrane shape change as function of Hct.

A similar elastic spring model was applied by Wang et al. (2009b) to investigate the skeletal structure of the RBC membrane and to study the dynamical behaviors of the RBC aggregates in microchannels.

Lattices with Beam Interactions

The beam network model is an upgrade of the spring network α-model obtained by substituting 1D structural elements capable of transferring only axial force with structural elements that can also transfer shear forces and bending moments. The following abbreviated discussion of beam networks is based primarily on papers by Ostoja-Starzewski (2002), Karihaloo et al. (2003), and Liu et al. (2008).

Triangular Bernoulli–Euler Beam Lattice

Bernoulli–Euler beams that transfer normal forces, shear forces, and bending moments are employed conventionally in beam lattice models to simulate the fracture process in concrete (e.g., Schlangen and Garboczi 1997; van Mier 1997; Lilliu and van Mier 2003). The kinematics of a beam network is described by three linear functions defining two displacement components and rotations at the network nodes. Highlights of the detailed analysis available in Ostoja-Starzewski (2002) are presented herein.

The elementary beam theory implies that the force–displacement and moment–rotation relations for each beam (b) are

$$F^{(b)} = E^{(b)} A^{(b)} \gamma^{(b)}, \quad Q^{(b)} = \frac{12 E^{(b)} I^{(b)}}{\left(L^{(b)}\right)^2} \widetilde{\gamma}^{(b)}, \quad M^{(b)} = E^{(b)} I^{(b)} \kappa^{(b)} \tag{33}$$

where the unit cell beams have the same geometrical properties: length $L^{(b)}$ and rectangular cross section $h^{(b)} \times t^{(b)}$, characterized by the area $A^{(b)}$ and the centroidal moment of inertia $I^{(b)}$. The average axial strain, $\gamma^{(b)}$, and the difference between the rotation angle of the beam chord and the rotation of its end node, $\widetilde{\gamma}^{(b)}$, are kinematic parameters defined with respect to the average axial strain in the half-beam in the lattice cell. The difference between the angles of rotation of beam ends, $\kappa^{(b)}$, is defined by its curvature.

This triangular Bernoulli–Euler beam lattice is an isotropic micropolar continuum with strain energy

$$U_{continuum} = \frac{V}{2}\gamma_{ij} C_{ijkl}\, \gamma_{kl} + \frac{V}{2}\kappa_i D_{ij}\, \kappa_j. \tag{34}$$

The equivalence of the strain energies in Eq. 1 for the triangular beam lattice made of equal beams ($L^{(b)} = L$, $t^{(b)} = t$, $h^{(b)} = h$, $A^{(b)} = A$, $I^{(b)} = I$) leads to

$$\begin{aligned} &C_{1111} = C_{2222} = \frac{3}{8}\left(3R + \tilde{R}\right), \quad C_{1122} = C_{2211} = \frac{3}{8}\left(R - \tilde{R}\right), \quad C_{1212} = \frac{3}{8}\left(R + 3\tilde{R}\right), \\ &C_{1221} = C_{2112} = \frac{3}{8}\left(R - \tilde{R}\right), \quad C_{2121} = \frac{3}{8}\left(R + 3\tilde{R}\right), \quad D_{11} = D_{22} = \frac{3}{8}S \end{aligned} \tag{35}$$

where zero stiffness components are omitted while

$$R = \frac{2E^{(b)}A}{L\sqrt{3}}, \quad \tilde{R} = \frac{24E^{(b)}I}{L^3\sqrt{3}}, \quad S = \frac{2E^{(b)}I}{L\sqrt{3}}. \tag{36}$$

Consequently, the effective modulus of elasticity and Poisson's ratio are written as

$$E = 2\sqrt{3}E^{(b)}\, t^{(b)} \left(\frac{h}{L}\right) \left[\frac{1 + (h/L)^2}{3 + (h/L)^2}\right], \quad \upsilon = \frac{1 - (h/L)^2}{3 + (h/L)^2} \tag{37}$$

The similar derivation for the square Bernoulli–Euler beam lattice reveals that it represents an orthotropic micropolar continuum unsuitable to model isotropic continuum (Ostoja-Starzewski 2002).

Triangular Timoshenko Beam Lattice

In the Timoshenko beam theory only the shear force

$$Q^{(b)} = \frac{12E^{(b)}I^{(b)}}{(1+\varsigma)\left(L^{(b)}\right)^3} L^{(b)}\tilde{\gamma}^{(b)}, \quad \varsigma = \frac{12E^{(b)}I^{(b)}}{G^{(b)}\tilde{A}^{(b)}\left(L^{(b)}\right)^2} = \frac{E^{(b)}}{G^{(b)}}\,\overline{h} \tag{38}$$

and the corresponding displacement $L^{(b)}\tilde{\gamma}^{(b)}$ differ from the Bernoulli–Euler formulation, while the normal force–displacement and moment–rotation relations $(33)_{1,3}$ are identical. In Eq. 38, ζ is the dimensionless parameter defined by the ratio of bending to shear stiffness. Bernoulli–Euler beam is recovered for very large shear stiffness and the slender beam since $\zeta \to 0$.

The equivalence of the strain energies in Eq. 1 for the triangular lattice yields the same Bernoulli–Euler beam expressions (35, 36) except that

$$\tilde{R}^{(b)} = \frac{24E^{(b)}I}{L^3\sqrt{3}} \frac{1}{(1+\varsigma)} \tag{39}$$

assuming all unit cell beams have the same dimensions ($L^{(b)} = L$, etc.).

Following the same steps, the nonzero stiffness components are

$$C_{1111} = C_{2222} = \frac{3}{8}\left(3R + \tilde{R}\right), \quad C_{1212} = \frac{3}{8}\left(R + 3\tilde{R}\right), \quad C_{1122} = C_{2211} = \frac{3}{8}\left(R - \tilde{R}\right),$$

$$C_{1221} = C_{2112} = \frac{3}{8}\left(R - \tilde{R}\right), \quad C_{2121} = \frac{3}{8}\left(R + 3\tilde{R}\right), \quad D_{11} = D_{22} = \frac{3}{8}S \tag{40}$$

while the effective modulus of elasticity and the Poisson's ratio are

$$E = 2\sqrt{3}E^{(b)}t^{(b)}\left(\frac{h}{L}\right)\left(\frac{1+(h/L)^2/(1+\zeta)}{3+(h/L)^2/(1+\zeta)}\right), \quad \upsilon = \frac{1-(h/L)^2/(1+\zeta)}{3+(h/L)^2/(1+\zeta)}. \tag{41}$$

Computer Implementation Procedure for Beam Lattices

Fracture process in beam lattice models is simulated by performing a linear elastic analysis under prescribed loading and removing from the network all beam elements that satisfy a predefined rupture criterion. Normal forces, shear forces, and moments are calculated using one of the beam theories. The global stiffness matrix is constructed for the entire lattice; its inverse matrix is calculated and then multiplied with the load vector to obtain the displacement vector. The heterogeneity of the material is taken into account by assigning different strengths to beams (e.g., using a Gaussian or Weibull distribution) or by assuming random dimensions of beams and random geometry of the lattice mesh or by mapping of different material properties to beams corresponding to the cement matrix, aggregate, and interface zones, respectively, in the case of concrete. To obtain aggregate overlay in the lattice, a Fuller curve is usually chosen for the distribution of grains. The beam length in concrete should be less than $l_b < d_{min}$ (where d_{min} is the minimum aggregate diameter).

Beam models can reproduce complex macroscopic damage patterns by cumulative effect of microcracking, crack branching, crack tortuosity, and bridging. They can also capture a size effect (e.g., Vidya Sagar 2004). The advantages of this approach are simplicity and a direct insight in the fracture process on the level of the microstructure. By applying an elastic-purely brittle local fracture law at the beam level, global softening behavior is observed. The main disadvantages of the conventional beam lattice model are the following: the results depend on the beam

size and direction of loading, the response of the material is too brittle (due to the assumed brittleness of single beams), the compressed beam elements overlap each other, and an extreme computational effort on the macrostructural level is needed. The first disadvantage can be removed by assuming a heterogeneous structure (Schlangen and Garboczi 1997). In turn, the second drawback can be mollified by 3D calculations and consideration of very small particles (Lilliu and van Mier 2003) and by applying a nonlocal approach in calculations of beam deformations (Schlangen and Garboczi 1997).

Examples of Applications of Lattices with Beam Interactions

Schlangen and Garboczi (1996, 1997) compared model techniques used in lattice simulations of random heterogeneous materials. The crack patterns obtained from a double-edge-crack concrete specimen loaded in shear (Fig. 11a) were compared to those obtained by numerical simulations with various lattice interaction types and lattice orientations. A selection of results is reproduced in Fig. 11b–d.

It is important to note that geometrical disorder (heterogeneity) was not implemented in the lattice models to emphasize the ability of the particular element type to describe continuum fracture. With reference to Fig. 11, the beam elements (Fig. 11d) were clearly superior to the two spring network interactions in capturing the experimentally observed complex crack pattern (Fig. 11a) in the *absence* of the model geometrical disorder. Nonetheless, it was obvious that the cracked patterns even in this case (Fig. 11d) revealed the mesh bias unavoidable in geometrically ordered lattices.

The comparison of simulated crack patterns, with four homogeneous lattices developed with square mesh, two differently oriented triangular meshes, and random triangular mesh, demonstrated expected superiority of the last type to capture the crack shape. These results, as well as those of Jirásek and Bažant (1995), emphasized the importance of lattice geometrical disorder for realistic simulation of crack propagation. However, geometrically disordered lattices were generally not homogeneous under uniform straining (Jagota and Bennison 1994). Schlangen and Garboczi (1996) suggested an approach to obtain an elastically uniform random network, which involved the iterative refinement of the lattice element properties. However, the authors expressed the expectation that for the problems characterized with inherent material randomness directly implemented in the lattice, the geometrically ordered lattice should exhibit similar crack pattern resemblance as long as the beam length was small compared to the textural length scale. The effect of the lattice resolution was also investigated and it was found that while the crack pattern was not strongly affected by the size of the beams, the load–crack opening curve was affected in much the same way as the effect of mesh refinement in local strain-softening models: the finer the lattice, the smaller the inelastic displacement and the dissipated energy (Cusatis et al. 2003).

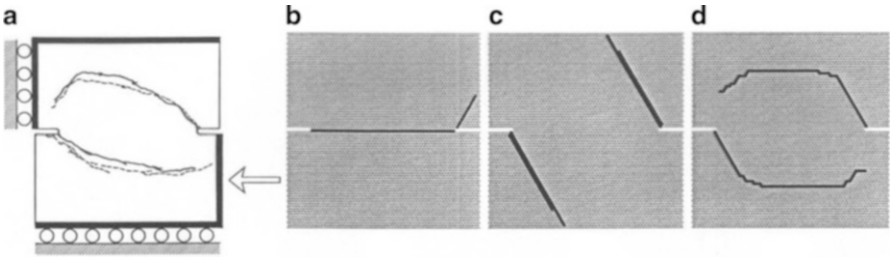


Fig. 11 (**a**) Geometry and crack pattern of concrete plate loaded in shear; crack patterns obtained by simulation on geometrically ordered triangular lattices with (**b**) central interactions (see section on "Triangular Lattice with Central Interactions"), (**c**) central and angular interaction (see section on "Square Lattice with Central and Angular Interactions"), and (**d**) beam interactions (see section on "Triangular Bernoulli–Euler Beam Lattice") (Reprinted from Schlangen and Garboczi (1996) with permission from Elsevier)

Schlangen and Garboczi proposed a new fracture law that used the maximum tensile stress in each *node* – instead of each beam – to address the problem of the directionality dependence of the lattice fracture stress. The nodal stress was determined based on the axial and shear contributions of each nodal bond and then used to determine the maximum normal force plane and corresponding beam area projections. The effective nodal stress was then determined as this normal force divided by the projected area and then used as a fracture criterion for each beam.

The methodology had been developed in the same study to implement the material heterogeneity in a direct way by using scanning electron micrographs and digital image processing of the microstructure to map different properties to lattice elements. The beam lattice in which this methodology was applied was subjected to a few basic loading configurations and realistic crack patterns were obtained.

Bolander and co-workers (1998, 2000, Bolander and Sukumar 2005) developed an elastically homogeneous lattice of random geometry to address the cracking direction bias caused by element-breaking low-energy pathways of regular lattices. Lattice elements were defined on the edges of the Delaunay tessellation of an irregular set of points generated in the material domain (Fig. 12a). The dual Voronoi tessellation was used to scale the elemental stiffness terms, in a manner that rendered the lattice model elastically homogeneous. This random lattice model could be viewed as an assemblage of rigid polygonal particles interconnected along their boundaries by a flexible interface inspired by the rigid-body-spring network developed by Kawai (1978).

The flexible interface, modeled by a set of discrete springs positioned midway along each boundary segment, is illustrated in Fig. 12b. Each spring set consisted of normal, tangential, and rotational springs with their respective stiffness k_n, k_t, and k_ϕ assigned in such a way to approximate the elastic properties of uniform continuum

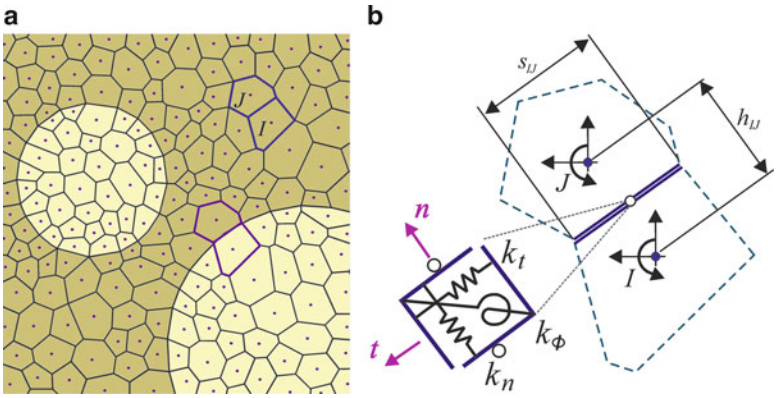


Fig. 12 (**a**) Irregular Voronoi tessellation of multiphase material and related nodes of dual Delaney network. (**b**) Two rigid particles joint by a flexible interface

$$k_n = EA_{IJ}/h_{IJ}, \quad k_t = k_n, \quad k_\phi = k_n s_{IJ}^2/12 \tag{42}$$

where $A_{IJ} = s_{IJ}t$, with t being the thickness of the planar model and E the modulus of elasticity of material. The assignment of spring constants Eq. 42 was uniquely defined by the Voronoi diagram to ensure elastically uniform lattice response since beam cross-sectional areas were scaled in proportion to the length of the common boundary segment. Notably, the equality of normal and tangential stiffness, Eq. $(42)_2$, necessary to ensure elastically uniform lattice response (Bolander and Sukumar 2005), resulted in unrealistic Poisson's effect.

The elastic formulation of the model and the mesh generation techniques were described in detail in original papers as well as a number of different fracture models. The use of fracture model based on the Bažant's crack-band approach (e.g., Bažant and Oh 1983) involved an incremental softening of the lattice elements in accordance to a predefined traction–displacement rule. Consequently, in contrast to conventional lattice approach, element rupture was gradual and governed by rules that provided an energy-conserving representation of fracture through the irregular lattice. This fracture model was objective with respect to the irregular lattice geometry: uniform fracture energy was consumed along the crack path regardless of the mesh geometry (Bolander and Sukumar 2005).

The irregular lattice models of Bolander and co-workers demonstrated ability to capture cracking patterns quite realistically. The distributed flexural cracking, the mixed-mode cracking of double-edge-notched plate, the bridge pier cracking under various design loads, and the fracture of three-point bend specimen were notable examples of successful applications of this modeling approach as illustrated by a selection of crack patters presented in Fig. 13. The ability of the model to simulate load–deformation response in fracture-sensitive simulations (challenging because of their complex cracks pathways) is illustrated in Fig. 14.

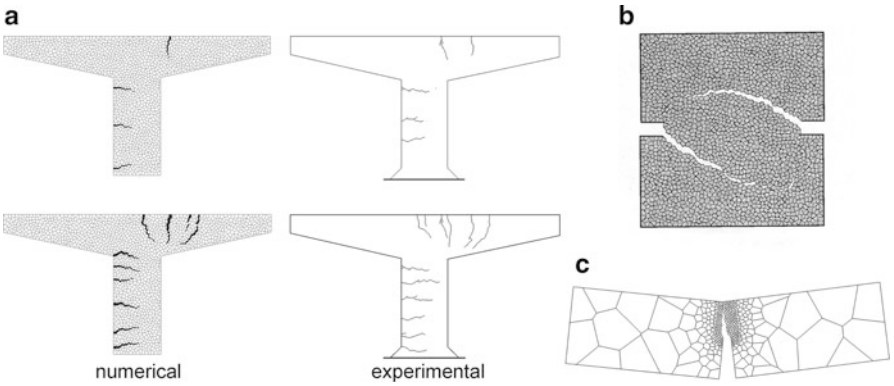


Fig. 13 (**a**) Experimentally and numerically obtained crack patterns in a bridge pear (Reprinted from Bolander et al. (2000) with permission from John Wiley and Sons). (**b**) Crack pattern for double-edge-notched plate loaded in shear (Reprinted from Bolander and Saito (1998) with permission from Elsevier). (**c**) Irregular mesh for simulation fracture in three-point bending test (reprinted figure with permission from Bolander and Sukumar (2005). Copyright (2005) by the American Physical Society. http://prb.aps.org/abstract/PRB/v71/i9/e094106)

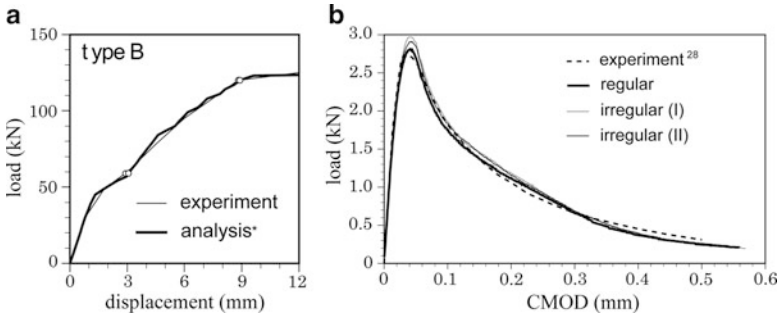


Fig. 14 Experimentally and numerically obtained (**a**) load–displacement curves for the bridge pear test (Fig. 13a) (Reprinted from Bolander et al. (2000) with permission from John Wiley and Sons) and (**b**) load–CMOD curves for the three-point bending test (Fig. 13c) (Reprinted figure with permission from Bolander and Sukumar (2005). Copyright (2005) by the American Physical Society. http://prb.aps.org/abstract/PRB/v71/i9/e094106)

Van Mier et al. (2002) investigated the effect of material's microstructural stochasticity on load–deformation response and crack patterns under uniaxial tension. They used the regular triangular lattice comprised of Bernoulli–Euler beams and the particle overlay method to mimic geometry of three-phase concrete to estimate the effect of strength and stiffness contrast in the composite. The particle overlay method appeared to be very successful in realistically capturing complex crack patterns. The progressive damage accumulation was simulated by subsequent removal of beam elements that satisfy a tensile-strength fracture criterion

$$\sigma_{eff}^{(b)} = \frac{F^{(b)}}{A^{(b)}} \pm \zeta \frac{\left(\left|M_i^{(b)}\right|, \left|M_j^{(b)}\right|\right)_{\max}}{W^{(b)}} = f_{cr} \tag{43}$$

where $F^{(b)}$ was the normal force in the considered beam, $M^{(b)}$ was the bending moments in the beam element nodes, and $A^{(b)}$ and $W^{(b)}$ were the area of cross section and the section modulus of the beam, respectively. The scaling coefficient ζ (selected in their study to be 0.05) was a fitting parameter that regulated which part of the bending moment was considered (effectively to match the experimentally observed response). This implied that an energy package $U_e^{(b)} = f_{cr}^2/2E$ was released at each bond rupture, which resembled the dissipated energy measured by acoustic emission monitoring. Three different tensile strengths were specified for the three material phases (aggregate, matrix, and interface; Fig. 1c) whose relative values were of importance.

The numerical simulation results had been compared to the outcome of simulations where the effects of microstructure were mimicked by assigning random strength values drawn from Weibull or Gaussian distribution to the regular triangular lattice. The results indicated that strength contrast was more pronounced than stiffness contrast when the maximum global force was considered and that global behavior was dominantly driven by damage percolation in the weakest material phase. The strength of the aggregate–matrix bonding interface and the connectivity of elements belonging to that phase were determined to be the decisive factor for the global strength under uniaxial tension. The results from the different Weibull distribution simulations resembled to a certain extent the failure mode observed in the more realistic three-phase particle overlay (notably the salient bridging phenomenon). In contrast, the correct cracking response could not be achieved with Gaussian strength distribution. The failure modes from Gaussian distribution simulations did not resemble the real fracture behavior observed experimentally in concrete, regardless of the fact that a large variety in force–deformation curves could be simulated depending on the choice of the distribution parameters.

In conclusion, the authors advised against the use of statistical strength distributions for simulating concrete response. They also cautioned that the force–deformation curves cannot be used as a single indicator to judge ability of a model to capture the fracture behavior of heterogeneous materials. The crack mechanisms and the ensuing crack patterns were considered salient elements in such judgments.

Arslan et al. (2002) and Karihaloo et al. (2003) improved in several ways the regular triangular beam lattice aimed at modeling fracture in particle composites. First, the aggregate phase remained linearly elastic–ideally brittle, but tension softening was allowed for the matrix and interface bonds by following the Bažant's bilinear stress–strain model outlined in section "Examples of Applications of α Model" (Fig. 3b). Second, the Timoshenko beam formulation was used to improve accuracy of interparticle interactions. Finally, the displacement-control simulation setup, able to account for finite deformations, was applied in attempt to capture the large deformations and rotations involved in the damage evolution. These changes

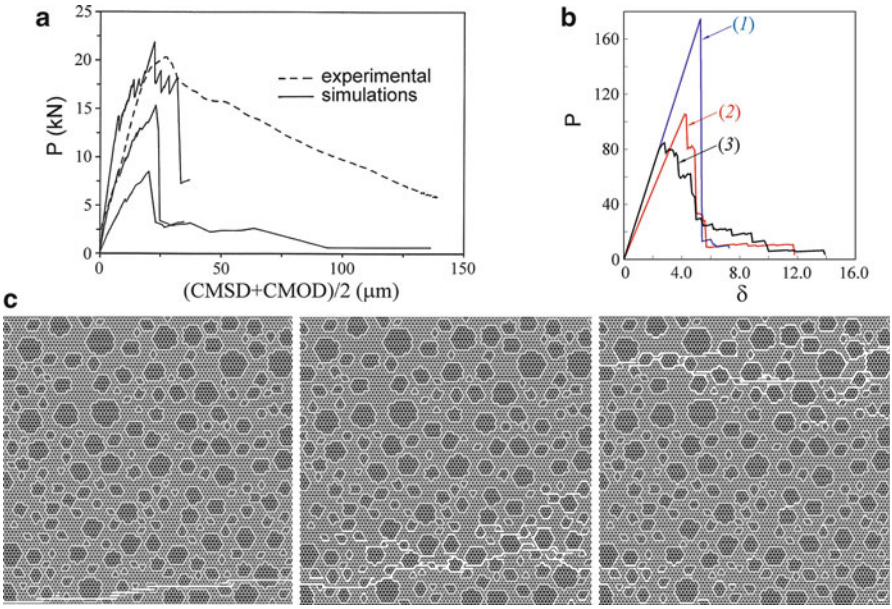


Fig. 15 (**a**) Experimental and numerical P–δ curves for four-point-shear beams of three different sizes (Reprinted from Schlangen and van Mier (1992) with permission from Elsevier). (**b**) P–δ curves for three cases: (*1*) no tension softening phase, (*2*) only matrix is tension softening, and (*3*) both matrix and interface are tension softening. (**c**) Crack patterns corresponding to the end of softening for the three *curves* above (*1–3*), respectively (Reprinted from Karihaloo et al. (2003) with permission from Elsevier)

were introduced in effort to address the primary deficiency of the beam lattice modeling of concrete at that time: the tendency of load–displacement (P–δ) curve to deviate substantially – regardless of the quite realistic crack patterns – from the experimentally observed response. Specifically, numerically simulated response was too discontinuous and brittle as shown in Fig. 15a. (Lilliu and van Mier (2003) developed a 3D lattice model for concrete to address these outstanding issues, which resulted in substantial increase of computational effort even for a relatively small model with pronounced boundary effects.)

Karihaloo and co-workers used an incremental iterative procedure based on the current secant modulus to account for the tension softening and other nonlinearities. The longitudinal strain of each beam element was calculated from nodal displacement and checked against the corresponding limit strain

$$\varepsilon = (1/L)\left[\left(u_1^j - u_1^i\right)cos\,\theta + \left(u_2^j - u_2^i\right)sin\,\theta + \left|\phi_j - \phi_i\right|\alpha_s\,(h/2)\right] = \varepsilon_{cr} \qquad (44)$$

where u_1, u_2, and ϕ correspond to three nodal degrees of freedom, h is the depth of nodal element, and α_s is the scaling parameter. If a phase is ideally brittle, the

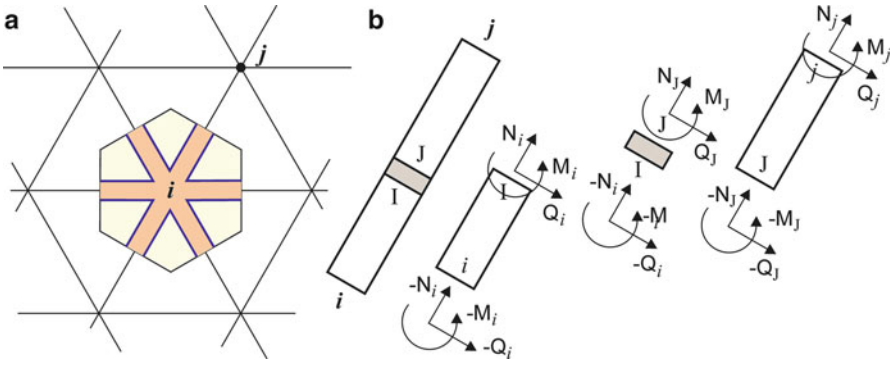


Fig. 16 The geometry of a GB lattice representing three-phase composite structure: (**a**) an particle overlay on GB lattice and a hexagonal unit cell of a triangular aggregate/matrix GB lattice, and (**b**) a composite GB element formed by the aggregate beam (*i*–*I*), the interface beam (*I*–*J*), and the matrix beam (*J*–*j*)

critical strain ε_{cr} corresponds to the tensile strength. When a bilinear constitutive relation is adopted for a phase (e.g., Fig. 3b), $\varepsilon_{cr} = \varepsilon_f$.

The typical simulation observations suggested that the difference between phase properties reduced steepness of load–displacement curve in the softening region. If the interface was the weakest phase, the deviation from the initially linear P–δ response marked the onset of substantial interface debonding, commonly related to stress concentrations. The softening threshold corresponded to the onset of matrix failure and the dominant macrocrack(s) formation as demonstrated experimentally by Van Mier and Nooru-Mohamed (1990) by using photoelastic coating.

The introduction of tension softening in the matrix and interface constitutive laws lead to more diffuse crack patterns (Fig. 15c) and, accordingly, to the more ductile response (Fig. 15b) in better agreement to the experimental observation for particle composites.

The inclusion of shear deformation based on Timoshenko beam theory apparently affected the P–δ response only slightly in contrast to the marked influence on the crack patterns. The finite deformations influenced substantially both softening behavior and the crack patterns. It exhibited a tendency of reduction of the softening response steepness as well as the number of the ruptured beam element.

Liu and co-workers (2007, 2008) developed a generalized-beam (GB) lattice model to mollify the problem of computational effort in the standard beam lattice approaches. The three-phase material structure was projected directly on top of the regular GB triangular lattice (Fig. 16a) with aggregate centers lying on lattice nodes and GB elements being a two-node and three-phase elements as depicted in Fig. 16. Every phase of the GB element is represented by a Timoshenko beam of equivalent properties; if a hexagonal unit cell did not contain an aggregate, it was still formally composed of three beam types with identical equivalent properties. The three beams of the GB element were assumed to cling firmly together so that the

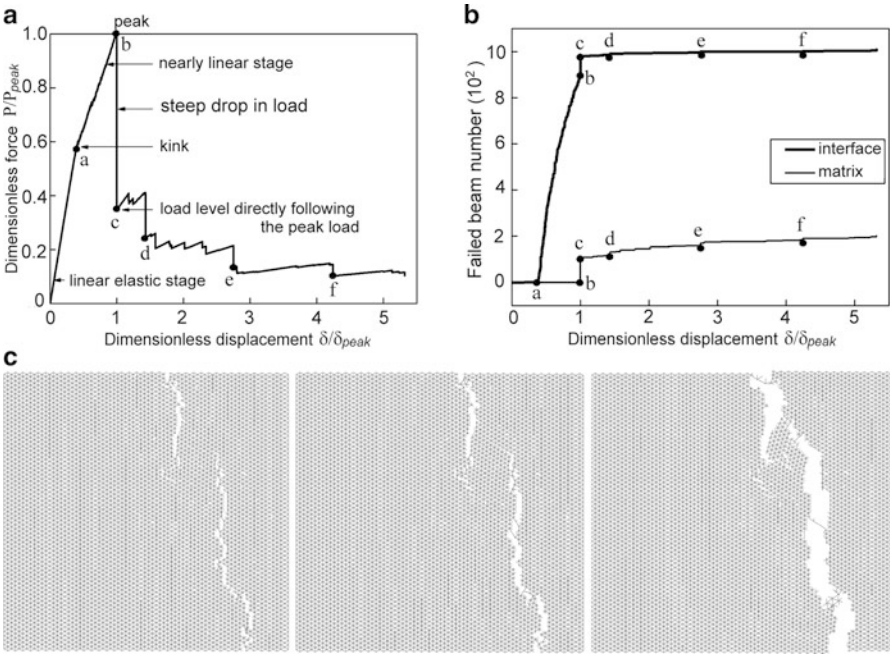


Fig. 17 (**a**) The P–δ curve and (**b**) the cumulative number of failed beams versus dimensionless displacement. (**c**) Three post-peak crack patterns at points *c*, *d*, and *f*, respectively (Reprinted from Liu et al. (2007) with permission from Elsevier)

displacements of two ends of the middle beam were completely determined by the two nodes of the GB element.

Matrix and aggregate beams were described by Timoshenko beam theory while the interface beams were defined in accordance with Appendix A of Bolander and Saito (1998). The composition of stiffness matrix of the GB element and parameter calibration of the corresponding lattice were presented in detail by Liu et al. (2007, 2008). The algorithm of quasistatic problem solution was essentially unchanged: fracture was simulated by subsequent removal of GB elements that satisfied the rupture criterion bearing in mind that rupture criticalities of three beams in each element were judged independently.

The rupture criterion used by Liu et al. (2007, 2008) was based on the maximum tensile stress. On the other hand, numerical representation of element rupture events in planar networks simulating the failure of paper (Liu et al. 2010) included also the maximum shearing stress criterion:

$$|\tau| = |q|/A_b = \tau_{cr} \tag{45}$$

where A_b is the intersecting (bond) area of two consecutive fibers and q is the corresponding sheering interaction.

The representative set of simulation results corresponding to the uniaxial tension test were presented in Fig. 17. The P–δ curve revealed two response regimes and a number of typical states along the loading path. The corresponding time histories of ruptured matrix and interface beams were depicted in Fig. 17b, while two characteristic snapshots corresponding to softening phase were illustrated in Fig. 17c, d.

Liu et al. (2008) were among first to propose dynamic beam lattice analysis using the central difference technique. The corresponding dynamic approach to the brittle fracture simulation was described and illustrated for four loading rates in the range $\log\dot{\varepsilon} \in [-3, 0]$ and compared with the quasistatic simulation results. This fracture process simulation by the dynamic procedure required a substantial increase of computational effort. The obtained simulation results demonstrated that while the inertial effects abated progressively with the loading rate reduction, the inertial effects due to the unstable crack propagation remained considerable even at the lowest loading rates. This observation raised the issue of the appropriateness of the common disregard of inertial effects in quasistatic analyses.

Khoei and Pourmatin (2011) developed a dynamic mesoscale model based on Timoshenko beam theory to investigate the dynamic response of three-phase inhomogeneous material. The dynamic analysis was performed using the Newmark's average acceleration technique. The aggregate distribution within the lattice model was generated based on their distribution within the network in a decelerating manner. The axial strain criterion, Eq. 44, was utilized for the beam element rupture.

Two experiments were used to illustrate the model capabilities in simulating the crack propagation in concrete: the simple uniaxial tensile (mode I) test and the double-edge-notched (mixed I and II modes) panel specimen. The crack patterns and P–δ curves were presented and discussed.

Conclusion

The survey of about 30 years of work, from the mid-1980s to present, highlights that lattice and DEM models (the later reviewed briefly in the next essay of this handbook) have evolved into a significant companion, if not alternative, to continuum models and traditional micromechanics. There is no reason to believe that this trend will stop anytime soon considering an increasing reliance on large computations to perform "virtual experiments" in structural and materials research and engineering.

These types of "discontinua-based models" enable the investigation of fracture problems by a unique bottom-up perspective, i.e., starting from physical behavior on a finer scale, to deduce the macroscale response as an outcome. In this regard, they do not supersede top-down continuum approaches or analytical modeling, but rather offer a complementary strategy to be used in a synergy with them for the purpose of controlling complex behaviors and complex materials. The rational model presented for the spring lattice is emblematic in this respect, with potentially far reaching consequences in addressing random fracture initiation and self-organized propagation problems typical of ceramics and ceramic matrix composites.

The possibility to use springs, trusses, beams or any special and fancy elements makes lattice models a tool to be customized at will and in virtually unlimited ways. Such tailorability is an unparalleled advantage for modeling and designing exotic and complex systems necessitated by modern engineering. Sure enough, discrete models have yet a long path ahead to become full-blown quantitative engineering tools but will certainly benefit from the boost due to innovative experimental techniques (discussed in detailed elsewhere in this handbook) and emerging novel size-confined materials for nanotechnology.

References

J. Alibert, P. Seppecher, F. Dell'Isola, Truss modular beams with deformation energy depending on higher displacement gradients. Math. Mech. Sol. **8**, 51–73 (2003)

A. Arslan, R. Ince, B.L. Karihaloo, Improved lattice model for concrete fracture. J. Eng. Mech. **128**(1), 57–65 (2002)

Z.P. Bažant, B.H. Oh, Crack band theory for fracture of concrete. Mater. Struct. (RILEM) **16**(93), 155–177 (1983)

Z.P. Bažant, M.R. Tabbara, M.T. Kazemi, G. Pyaudier-Cabot, Random particle model for fracture of aggregate and fiber composites. J. Eng. Mech. **116**(8), 1686–1705 (1990)

J.E. Bolander Jr., S. Saito, Fracture analyses using spring networks with random geometry. Eng. Fract. Mech. **61**, 569–591 (1998)

J.E. Bolander, N. Sukumar, Irregular lattice model for quasistatic crack propagation. Phys. Rev. B **71**, 094106 (2005)

J.E. Bolander Jr., G.S. Hong, K. Yoshitake, Structural concrete analysis using rigid-body-spring networks. Comput. Aided Civ. Infrast. Eng. **15**, 120 (2000)

G. Cusatis, Z.P. Bažant, L. Cedolin, Confinement-shear lattice model for concrete damage in tension and compression I. Theory. J. Eng. Mech. **129**(12), 1439–1448 (2003)

G. Cusatis, Z.P. Bazant, L. Cedolin, Confinement-shear lattice CSL model for fracture propagation in concrete. Comput. Methods Appl. Mech. Eng. **195**, 7154–7171 (2006)

M. Grah, K. Alzebdeh, P.Y. Sheng, M.D. Vaudin, K.J. Bowman, M. Ostoja-Starzewski, Brittle intergranular failure in 2D microstructures: experiments and computer simulations. Acta Mater. **44**(10), 4003–4018 (1996)

P. Hou, Lattice model applied to the fracture of large strain composite. Theory Appl. Fract. Mech. **47**, 233–243 (2007)

A. Hrennikoff, Solution of problems of elasticity by the framework method. J. Appl. Mech. **8**, A619–A715 (1941)

A. Jagota, S.J. Bennison, Spring-network and finite element models for elasticity and fracture, in *Proceedings of a Workshop on Breakdown and Non-linearity in Soft Condensed Matter*, ed. by K.K. Bardhan, B.K. Chakrabarti, A. Hansen (Springer, Berlin/Heidelberg/New York, 1994), pp. 186–201

M. Jirásek, Z.P. Bažant, Microscopic fracture characteristics of random particle system. Int. J. Fract. **69**, 201–228 (1995)

B.L. Karihaloo, P.F. Shao, Q.Z. Xiao, Lattice modelling of the failure of particle composites. Eng. Fract. Mech. **70**, 2385–2406 (2003)

T. Kawai, New discrete models and their application to seismic response analysis of structures. Nucl. Eng. Des. **48**, 207–229 (1978)

A.R. Khoei, M.H. Pourmatin, A dynamic lattice model for heterogeneous materials. Comput. Methods Civ. Eng. **2**, 1–20 (2011)

J. Kozicki, J. Tejchman, Modelling of fracture process in concrete using a novel lattice model. Granul. Matter **10**, 377–388 (2008)

G. Lilliu, J.G.M. van Mier, 3D lattice type fracture model for concrete. Eng. Fract. Mech. **70**, 927–941 (2003)
J.X. Liu, S.C. Deng, J. Zhang, N.G. Liang, Lattice type of fracture model for concrete. Theory Appl. Fract. Mech. **48**, 269–284 (2007)
J.X. Liu, S.C. Deng, N.G. Liang, Comparison of the quasi-static method and the dynamic method for simulating fracture processes in concrete. Comput. Mech. **41**, 647–660 (2008)
J.X. Liu, Z.T. Chen, K.C. Li, A 2-D lattice model for simulating the failure of paper. Theory Appl. Fract. Mech. **54**, 1–10 (2010)
A. Misra, C.S. Chang, Effective elastic moduli of heterogeneous granular solids. Int. J. Sol. Struct. **30**(18), 2547–2566 (1993)
M. Ostoja-Starzewski, Lattice models in micromechanics. Appl. Mech. Rev. **55**(1), 35–60 (2002)
M. Ostoja-Starzewski, *Microstructural Randomness and Scaling in Mechanics of Materials* (Taylor & Francis Group, Boca Raton, 2007)
S.L. Phoenix, I.J. Beyerlein, Statistical strength theory for fibrous composite materials, in *Comprehensive Composite Materials*, ed. by A. Kelly, vol. 1 (Pergamon, Oxford, 2000), pp. 559–639
A. Rinaldi, A rational model for 2D disordered lattices under uniaxial loading. Int. J. Damage Mech. **18**, 233–257 (2009)
A. Rinaldi, Statistical model with two order parameters for ductile and soft fiber bundles in nanoscience and biomaterials. Phys. Rev. E **83**(4–2), 046126 (2011a)
A. Rinaldi, Advances in statistical damage mechanics: new modelling strategies, in *Damage Mechanics and Micromechanics of Localized Fracture Phenomena in Inelastic Solids*, ed. by G. Voyiadjis. CISM Course Series, vol. 525 (Springer, Berlin/Heidelberg/New York, 2011b)
A. Rinaldi, Bottom-up modeling of damage in heterogeneous quasi-brittle solids. Continuum Mech. Thermodyn. **25**(2–4), 359–373 (2013)
A. Rinaldi, Y.C. Lai, Damage theory of 2D disordered lattices: energetics and physical foundations of damage parameter. Int. J. Plast. **23**, 1796–1825 (2007)
A. Rinaldi, L. Placidi, A microscale second gradient approximation of the damage parameter of quasi-brittle heterogeneous lattices. Z. Angew. Math. Mech. (ZAMM) (2013). doi:10.1002/zamm.201300028
A. Rinaldi, S. Mastilovic, D. Krajcinovic, Statistical damage mechanics – 2. Constitutive relations. J. Theory Appl. Mech. **44**(3), 585–602 (2006)
A. Rinaldi, D. Krajcinovic, S. Mastilovic, Statistical damage mechanics and extreme value theory. Int. J. Damage Mech. **16**(1), 57–76 (2007)
E. Schlangen, E.J. Garboczi, New method for simulating fracture using an elastically uniform random geometry lattice. Int. J. Eng. Sci. **34**, 1131–1144 (1996)
E. Schlangen, E.J. Garboczi, Fracture simulations of concrete using lattice models: computational aspects. Eng. Fract. Mech. **57**, 319–332 (1997)
E. Schlangen, J.G.M. Van Mier, Micromechanical analysis of fracture of concrete. Int. J. Damage Mech. **1**, 435–454 (1992)
V. Topin, J.-Y. Delenne, F. Radjaï, L. Brendel, F. Mabille, Strength and failure of cemented granular matter. Eur. Phys. J. E **23**, 413–429 (2007)
K. Tsubota, S. Wada, Elastic force of red blood cell membrane during tank-treading motion: consideration of the membrane's natural state. Int. J. Mech. Sci. **52**, 356–364 (2010)
K. Tsubota, S. Wada, T. Yamaguchi, Simulation study on effects of hematocrit on blood flow properties using particle method. J. Biomech. Sci. Eng. **1**(1), 159–170 (2006)
J.G.M. Van Mier, *Fracture Processes of Concrete* (CRC Press, New York, 1997)
J.G.M. Van Mier, M.B. Nooru-Mohamed, Geometrical and structural aspects of concrete fracture. Eng. Fract. Mech. **35**(4/5), 617–628 (1990)
J.G.M. Van Mier, M.R.A. van Vliet, T.K. Wang, Fracture mechanisms in particle composites: statistical aspects in lattice type analysis. Mech. Mater. **34**, 705–724 (2002)
R. Vidya Sagar, Size effect in tensile fracture of concrete – a study based on lattice model applied to CT-specimen, in *Proceedings of the 21th International Congress of Theoretical and Applied*

Mechanics ICTAM04, ed. by W. Gutkowski, T.A. Kowaleski (Springer, Berlin/Heidelberg/New York, 2004), pp. 1–3

H.-J. Vogel, H. Hoffmann, A. Leopold, K. Roth, Studies of crack dynamics in clay soil II. A physically based model for crack formation. Geoderma **125**, 213–223 (2005)

G. Wang, A. Al-Ostaz, A.H.-D. Cheng, P.R. Mantena, Hybrid lattice particle modeling: theoretical considerations for a 2D elastic spring network for dynamic fracture simulations. Comput. Mater. Sci. **44**, 1126–1134 (2009a)

T. Wang, T.-W. Pan, Z.W. Xing, R. Glowinski, Numerical simulation of rheology of red blood cell rouleaux in microchannels. Phys. Rev. E **79**, 041916 (2009b)

A. Zubelewicz, Z.P. Bažant, Interface element modeling of fracture in aggregate composites. J. Eng. Mech. **113**(11), 1619–1630 (1987)

Section III

Damage in Crystalline Metals and Alloys

Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials

11

Akrum Abdul-Latif

Contents

Abstract

This chapter provides the basic and necessary elements concerning the anisotropy induced by damage. Moreover, local mechanism effect on the damaged behavior will be demonstrated leading to exceed the scope of this chapter. So, it is convenient to present, discuss, and analyze these subjects. Understanding some physical phenomena related to plasticity and damage is important in modeling. Hence, the main objective is to describe accurately the overall cyclic plasticity behavior coupled with damage via constitutive relations.

A. Abdul-Latif (✉)
Université Paris 8, Laboratoire d'Ingénierie des Systémes Mécaniques et des Matériaux (LISMMA), Saint Ouen Cedex, France
e-mail: aabdul@iu2t.univ-paris8.fr

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_5

A micromechanical model of damage initiation in low-cycle fatigue will be presented for describing notably the damage deactivation effect. Actually, it is considered that the damage is active only if microcracks are open, while damage affects the mechanical properties of polycrystals during its closure (inactive phase) differently. With the small strain assumption, the plastic strain and local damage variables are examined at the crystallographic slip system scale. The anisotropic damaged behavior, induced by activation/deactivation phenomenon, is modeled using a fourth-order damage tensor at the overall scale. Accordingly, the overall nonlinear behavior, notably the deactivation phase due to the microcrack closure under complex cyclic loadings, is of particular interest in this chapter.

Introduction

The excellent knowledge of used polycrystalline materials is always required. Consequently, it is necessary to understand their mechanical behavior through the related microstructures and their impact on such a behavior. The nonlinearity of material behavior is generally induced by plasticity and damage mechanics. Ductile polycrystalline materials usually fail as a result of nucleation, growth, and coalescence of microcracks and/or microvoids. Experimental observations show that the accumulation of microcracks and/or microvoids has a tendency to form a localized damage, due to plastic strain localization up to the final failure of structure. In fact, in several metallic materials, the kinematic plastic strengthening is related to the creation of slip bands. The setting of these bands in the material induces undoubtedly an internal back stress in grains leading accordingly to an anisotropic behavior.

In mechanical engineering applications, fatigue is one of the major considerations in engineering structures and machine design since numerous of the constituent parts, in-service, are generally subjected to various complex loading paths. Metallurgical changes in deformation behavior are directly involved in the initiation and accumulation of fatigue damage. These occur at highly stressed (or deformed) and localized weak areas (e.g., slip bands, grain boundaries, second-phase particles, inclusions, and other local heterogeneities). Fatigue damage is basically associated with these regions where the plastic deformation is highly localized due to the inhomogeneous nature of polycrystalline metals at the local level. This consequently gives local heterogeneous deterioration in their deformation resistance. Some connections between local and global responses represent an appropriate context to thoroughly understand the importance of the substructure heterogeneity and its effect on the strain field and then on the fatigue life. Besides, the effects of principal parameters (plastic strain amplitude and accumulative plastic strain dependence of the cyclic plasticity behavior) represent key issues for the multiaxial fatigue behavior.

In general, fatigue damage consists of two main stages: microcrack initiation and their subsequent propagation. Actually, the microcrack initiation is an important stage in fatigue life. Several studies show that the complexity of the local

mechanisms of this stage has an important impact on the fatigue life. In fact, the initiation of microcracks of an intragranular type takes place in the intensive slip bands which are the natural site for this type. TEM (transmission electron microscopy) observations demonstrated by several research programs reveal strain localization in slip bands taking place during loading in which an important dislocation density occurs. Microstructural observations show that crack initiation occurs in some slip bands as in Waspaloy. Thus, these slip bands together with microcracks seem to be important factors inducing an anisotropic behavior.

From the local fatigue damage point of view, one can summarize the following key points:

- Intragranular fatigue crack initiation is a local phenomenon localizing on the free surfaces of the structure components.
- The site of intragranular fatigue crack initiation varies depending on the involved material microstructure as well as the applied loading condition.
- For relatively pure metals subjected to low cyclic straining, intragranular fatigue crack initiation takes place preferentially at sites of surface roughness associated with emerging planar slip bands called persistent slip bands (PSB). Generally, this roughness is due to extrusion and intrusion mechanisms.
- Whatever the metallic materials, intragranular fatigue crack initiation is governed by strain localization inside narrow bands due to local heterogeneities related to the material microstructure such as inclusions, porosities, second phases, precipitates, heterogeneous distribution of dislocations, etc.
- In the case of FCC and BCC crystalline materials, there is a widespread belief that fatigue crack nuclei are strongly governed by a continuous irreversible slip within PSB. All microcrack nuclei (less than 5 μm deep) are almost parallel to the primary slip plane.

In LCF, it has been observed that the resulting lives are highly sensitive to cyclic plastic deformation governed by the applied loading amplitude and also to loading path complexity. Due to these factors, metallurgical changes arise at locally high deformed regions leading naturally to the fatigue failure.

Under constant cyclic amplitude, metallic single-phase, especially FCC, polycrystals show distinct substructural behavior evolutions. Moreover, under out-of-phase or nonproportional cyclic loading paths, it has been shown that the principal stress and strain axes rotate during loading, often inducing, in general, additional cyclic hardening for many metallic materials such as stainless steel 316L, Waspaloy, aluminum alloy 2024, etc. The reason that low-cycle out-of-phase loading is more damaging than low-cycle in-phase loading is attributed to this additional cyclic hardening.

Concerning the modeling issue, most of the proposed models based on the formulation of extrusions and/or intrusion use the concept of dislocation movement along different paths on the slip band forward and reverse loading. The typical example is the one-dimensional model proposed by Mura and his co-workers (Mura 1982; Tanaka and Mura 1981). This model adopts the concept of dislocation dipole by considering

two adjacent layers of dislocation pileups of opposite sign. Thus, the forward and reverse plastic flows within the slip band are modeled by dislocations with different signs moving on two closely located layers under the assumption of the irreversible dislocation motion. Another modeling approach has also been developed for fatigue life of materials and structures subjected to various multiaxial loading paths. This research led to the development of several models for high- and low-cycle fatigue lives. The widely used approach for modeling these two subjects is the macroscopic one based on some physical considerations or on purely phenomenological ones using a thermodynamic framework of the continuum damage mechanics (CDM) describing the isotropic and anisotropic damaged-plastic (or viscoplastic) behavior of materials. The phenomenological (macroscopic) approach is extensively used in many research laboratories and industrial developments.

Moreover, ductile damage, due to microvoids evolution, occurs especially in well plastically deformed zones where the stress triaxiality is high. The latter has a significant effect on the voids growth rate. Voids initiation and growth have been extensively studied by means of micromechanics analysis. In 1977, Gurson proposed a pioneered model of damage by cavitation based on an approximation analysis of spherical voids with only one yield function for porous ductile perfectly plastic matrix. The initial Gurson's model shows some limitations. In fact, it overpredicts the evolution of microvoids under monotonic loading conditions. Any type of ratcheting under cyclic loading paths cannot be predicted since the yield function depends only on a single yield function. Therefore, several extensions have been made. The most important ones are either based on improving predictions at low volume fraction of voids (Tvergaard 1982) or the modification of its yield function in order to describe the rate sensitivity, necking instabilities, and better description of the final voids coalescence.

Alternatively, the micromechanical approach is currently considered one of the approaches progressively used due to the enormous progress in computer science. Fewer assumptions are required and the elegance of the solutions makes micromechanical models interesting. They use almost some "physical variables" in order to appropriately reproduce the principle cyclic features. These models are based on localization–homogenization method (Germain et al. 1983) as initially developed by Dang Van (1973) since many years ago. In the case of LCF, the micromechanical approaches describe appropriately the principle cyclic features such as Bauschinger effect, additional hardening, and many other phenomena under different cyclic loading paths without damage effect. Moreover, as a theoretical attempt, an elasto-inelastic-damage model has been already proposed within the team work of the author since 20 years ago. It describes the cyclic behavior of polycrystals under complex loading paths. This model assumes that the microcracks initiate at the crystallographic slip system level neglecting all types of damaging related to grain boundaries. It has correctly predicted the fatigue life emphasizing the advantages to demonstrate the loading path complexity effect on such a life, i.e., the greater the complexity of the loading path, the greater is the additional hardening and the shorter the fatigue life. Recently, two theoretical issues have been developed as an important extension to this mode. The first one relates to the

damage induced-oriented anisotropy behavior of the material due to damage deactivation. In fact, the microcracks may open or close depending on the loading path types, thus giving different responses in compression and tension. The second point concerns the predicted difference in fatigue lives for a given polycrystal under several cyclic loading situations having the same equivalent strain. As a fundamental development, these two extensions will largely be presented in this chapter.

The problem of damage deactivation seems to be incompletely resolved, even though numerous approaches have been developed since the end of the last century. It is well known that anisotropy induced by damage deactivation in polycrystals is highly complex notably when it is coupled with plasticity leading to a solution that is not straightforward. One of the principal difficulties is the discontinuous stress–strain relation when the damage deactivation condition takes place. To remedy this theoretical difficulty, several solutions were given. Among these, three solutions are cited here. The first one considers the so-called principal directions of damage. The damage deactivation condition only modifies the diagonal principal terms of the stiffness or compliance operators (Chaboche 1993). These were evaluated in a system consistent with respect to the principal direction of damage. Moreover, this solution also describes the anisotropy induced by damage. Another possible answer is to express the damage deactivation character in terms of energy rather than to distinguish between the tension and compression behaviors using stresses and strains. These two approaches consider that the damage kinematics is limited to the case of the elastic behavior, i.e., neglected the inelastic strain behavior. Concerning the third one, discontinuities in the response are avoided by introducing smooth functions to guarantee properly the transition from the active to the inactive state (Hansen and Schreyer 1995). This one describes likewise either the damaged-elastic or damaged-elastoplastic behavior. The above-suggested approaches are of macroscopic nature.

In this chapter, a micromechanical model of damage-inelastic behavior, which has been initially proposed for polycrystalline structures and recently extended by Abdul-Latif and Mounounga (2009), will be presented, describing the damage deactivation effect in plastic fatigue with the small strain assumption. The damage activation/deactivation is formulated and treated only at the macroscopic scale using the mathematical projection operators. These operators allow to define a fourth-order damage tensor which is capable to take into account the damage deactivation effect in the case of multiaxial cyclic loading paths and to naturally describe the related phenomenon of the anisotropy induced by damage.

Some Physical Considerations Related to Fatigue Failure

Reliable prediction of fatigue failure can be obtained only by thorough understanding of physical mechanisms involved. Three main stages of low-cycle fatigue are experimentally observed. These are accommodation, crack initiation, and crack growth. Since the developed model can describe the first two stages, only the accommodation and crack initiation stages will be presented.

In LCF, the first stage (accommodation) which can represent from 10 % to 50 % of the fatigue life attaints a stabilized regime when the structure of the material evolves in a manner to fit the imposed cyclic strain (in strain-controlled situation) and to reach stress saturation. The dislocation mechanisms are responsible for the stress saturation. In fact, either the same dislocations move to and fro or a dynamic equilibrium between dislocation multiplication and annihilation can interpret the existence of a cycle saturation regime (steady state). Generally, fatigue properties of pure metals show an enormous change of cyclic hardening with respect to critical shear stress to the saturation stress. Fatigue phenomena depend strongly on how fast the cyclic hardening is taking place. Due to the Bauschinger effect (partial reversibility of plastic strain), the cyclic hardening is much slower than the hardening obtained via a monotonic tensile test. Because of the absence of long-range internal stresses in the fatigued state, the differences between the cyclically and unidirectionally hardened states are impressive. The important phenomenon related to this stage is the intensive slip band formation. It can be expressed by a heterogeneous and localized strain on the slip bands. In these bands, it is observed that the understructure of dislocations is different to that of the matrix. In the case of low strains, the dislocations form certain arrangements (bands) in the matrix. One of the most commonly intensive slip bands is the persistent slip bands (PSB). Such bands develop parallel to the primary slip plane across the whole cross section in lamellae, which are softer than the cyclically hardened matrix. Pure single and polycrystalline FCC structures exhibit a considerable cyclic hardening due to the dislocation multiplication on primary slip system producing strong latent hardening for the secondary slip system. The dislocation structure, after fatigue, can be defined by regions of high dislocation density (veins) separated by almost dislocation-poor area (channels). The veins consist predominantly of primary edge dislocations. This primary slip system remains the most highly stressed slip throughout the history. Channel's size is comparable to that of the veins. The dislocation density in the channels is about three orders of magnitude smaller than the dislocation density in the veins. It is obvious that during cyclic loading, the continuous forward and reverse plastic flow is governed by dislocations with different signs. The dislocations trap each other and stop moving over large distances forming dislocation dipole. Moreover, in the case of FCC metals and alloys, the stacking fault energy plays an important role on the cyclic behavior of materials (e.g., copper alloys). In general, alloying leads to smaller cyclic hardening rate when the stacking fault energy is low.

It is well known that the definition of the microcrack initiation in LCF depends on the measurement device. This means that there is not yet a well precise and unified definition of this phenomenon (generally measured by a number of cycles). Moreover, several microcrack sites can usually be observed in LCF. Note that this phenomenon can be determined when a microcrack reaches a certain size comparable to grain size (a length of 100 μm). In fact, this dimension can easily be detected. In many fatigue cases, as soon as a microcrack reaches a depth not far from the above dimension, its propagation takes place across the section. The number of cycles needed to reach such a length represents a non-neglected proportion of the fatigue life. Consequently, the damage deactivation effect is quite clear

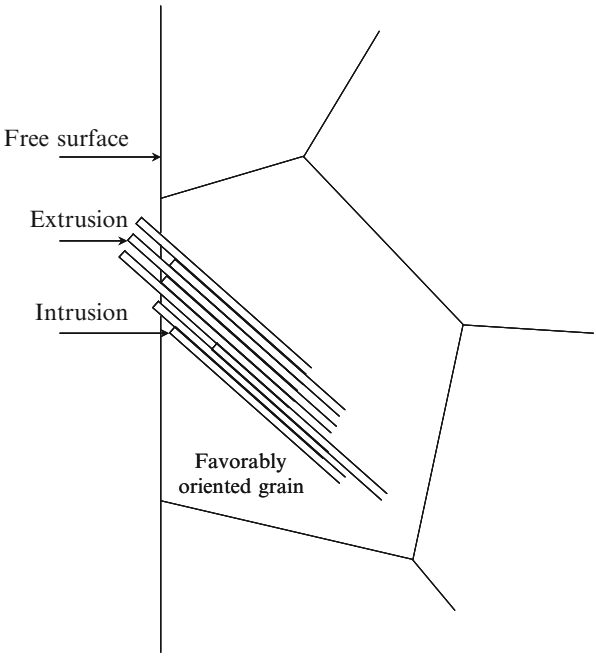


Fig. 1 Schematic plot of the extrusion and intrusion phenomenon occurring during cyclic loading at the free surface of polycrystalline metals (From Abdul-Latif et al. (1999), with permission from ASME Publications)

at the end of stage I of fatigue and becomes clearer during the microcrack propagation (stage II) when the microcracks are open.

The nature of microcrack sites seems to be similar to both fatigue types (low- and high-cycle fatigues). Thus, it is observed that the microcrack nucleates, in general, in some grains located at the free surface or at the interfaces (matrix–inclusion interfaces, grain boundaries). As a matter of fact, the grain boundaries can constitute substantial sites of microcrack initiation. For high strain amplitudes, the initiation mechanism can, in fact, be purely geometric. Under these conditions, slip bands invade completely the grains and the free surface of the specimen taking a tormented aspect with form changes, especially in the grain boundary level cutting the free surface. Hence, deep intrusions develop at grain boundaries consequently leading to microcrack initiation as in the case of Cu and Al.

When the initiation is of an intragranular type, the intensive slip bands represent a natural site of microcrack nucleation. The nucleation mechanisms are highly complicated; therefore, they are not yet perfectly defined. In the case of alloys containing a second phase, the interaction nature between dislocations and the particles is important. However, intragranular cracks initiate most commonly at sites of surface roughness associated with planar slip bands or PSB for pure metals and alloys. This roughness is related to extrusions and intrusions. The plastic strain concentration in these bands gives the formation of extrusions and intrusions (Fig. 1). Hence, the microcrack initiation "embryo" takes place in the bands or at the interfaces between these plastic bands and the matrix.

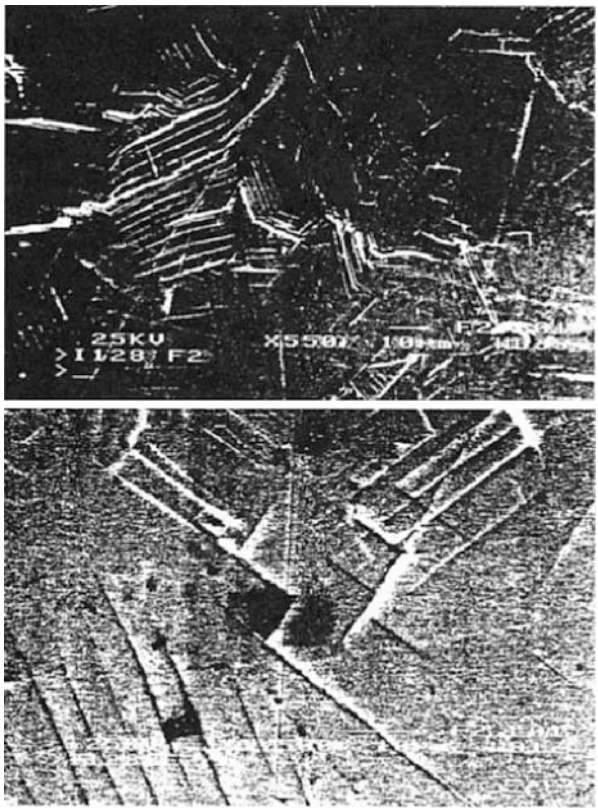


Fig. 2 Waspaloy (under aged state) under tension–compression (specimen axis is vertical), SEM showing the microcrack initiation on the crystallographic slip bands (From Abdul-Latif et al. (1999), with permission from ASME Publications)

Concentrated on the intragranular cracking, Fig. 2 points out via scanning electron microscope that crack initiation occurs in some slip bands of specimen outer surfaces (case of Waspaloy). These results are in accordance with those of austenitic stainless steel (AISI304) (Parsons and Pascoe 1976) and of Waspaloy (Lerch et al. 1984). It appears that cracking does not occur simultaneously along the entire length of a slip band, but may be confined to one part of the band. After initiation in slip bands, cracks zigzag from one slip band to another within a grain. This observation (cracks zigzagging) could therefore be interpreted by the interaction phenomenon among these microcracks within the same grain. This gives certain validity to introducing the damage interaction matrix D_{rs} in presented model as it will be shown later.

Modeling Motivation

In the fatigue of a smooth structure of polycrystalline materials, damages at the slip band level are expected to be nucleated in some grains located at the free surface having a high value of resolved cyclic shear stress τ^s on the slip plane in the slip

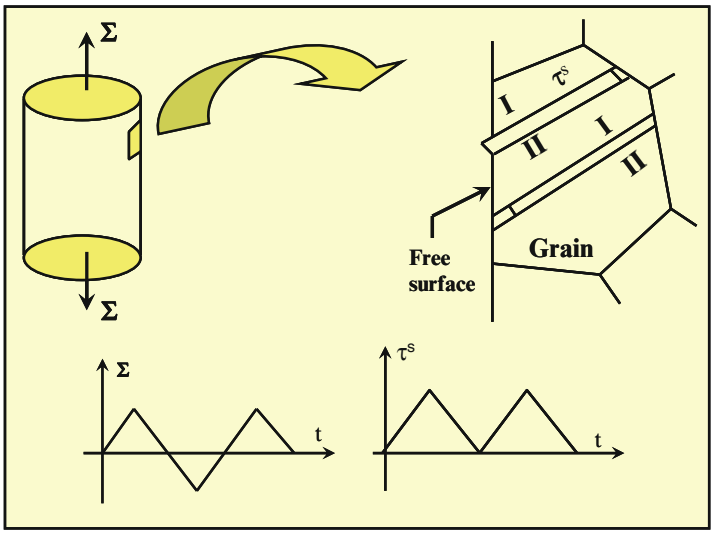


Fig. 3 Principle of the intragranular fatigue crack initiation (From Abdul-Latif et al. (1999), with permission from ASME Publications)

direction. The forward and reverse plastic flows within the slip band are supposed to be governed by dislocations with different signs moving irreversibly in two closely located layers. Instead of modeling explicitly the dislocation movement to describe the formation of extrusions and intrusions as in Mura's model, the micromechanical approach is used to represent the extrusions and intrusions by an internal variable. This variable aims to describe the damage initiated due to the formation of these extrusions and intrusions and their accumulations during cyclic loading. The schematic plot of this idea is given in Fig. 3, where a smooth specimen is subjected to an overall uniaxial cyclic stress (Σ). This figure shows also a section perpendicular to the specimen surface with a most favorably oriented grain. The slip is consequently submitted to resolved cyclic shear stress τ^s.

Since the lowest level of the microstructure considered in the presented model is the crystallographic slip system (CSS), it is suggested here that all phenomena related to those levels lower than the CSS (i.e., dislocations, molecules, lattice defects, atoms) are globally modeled on the CSS level. Hence, the crack "embryo" represented by the intrusions, extrusions, or vacancy dipoles is globally described by a damage internal state variable (d^s) at the CSS level. So, for a system *s*, the internal state variable d^s represents these microdefects localized on the system, i.e., vacancy dipoles and extrusions, or interstitial dipoles and intrusions which are due to the dislocation motions as indicated above.

On the other hand, modeling of the spatial localization of cyclic fatigue damage on the free surface of the structure is not a trivial task. This difficult problem is not explicitly treated in the present chapter, but one can indicate hereafter an approximate numerical method to model this phenomenon through the finite elements

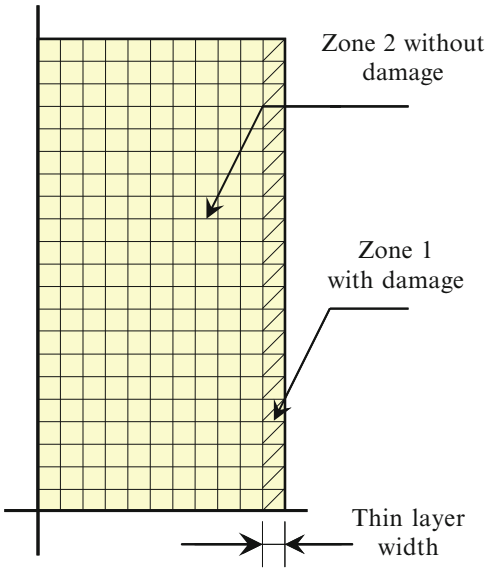


Fig. 4 Schematic representation of the quarter of a round tensile specimen (From Abdul-Latif et al. (1999), with permission from ASME Publications)

method. Actually, Fig. 4 illustrates the numerical modeling of the quarter of the specimen defined above (Fig. 3). Two zones are introduced: the first one is a thin layer located at the surface of the specimen having constitutive equations coupled with intragranular damage. The second zone is the remaining (volumic) part of the specimen having constitutive equations without damage effect. This allows the localization of the fatigue damage on the finite elements within the zone 1 (near to the specimen free surface). Naturally, this method of spatial localization modeling of the fatigue damage at the free surface is approximate, and a more straightforward approach should be used in the framework of the nonlocal mechanics. In fact, modeling the fatigue damage localization at the free surface via the finite elements methodology can suitably be performed by this framework where the internal length concept is incorporated to solve the problem of mesh dependency and then to model suitably the damaging process. From the computational point of view, the nonlocal models appear to be straightforward to implement. Until now, the nonlocal formulation is separately used either in brittle damage models or in ductile plastic damage models.

The present model which will be considered in this chapter does not take into account any surface effect, and particularly, it does not give $\sigma^{g}{}_{11} = \sigma^{g}{}_{22} = \sigma^{g}{}_{33} = 0$ at the free surface. The model is then devoted to describe the low-cycle fatigue crack initiation. In fatigue, microcracks may open or close depending on the applied loading direction. Thus, different responses can experimentally be observed under compression and tension loads, which lead to damage deactivation behavior for an aluminum alloy (Fig. 5). This leads consequently to an induced-oriented anisotropy phenomenon. It is important to keep in mind that fatigue damage behaves almost

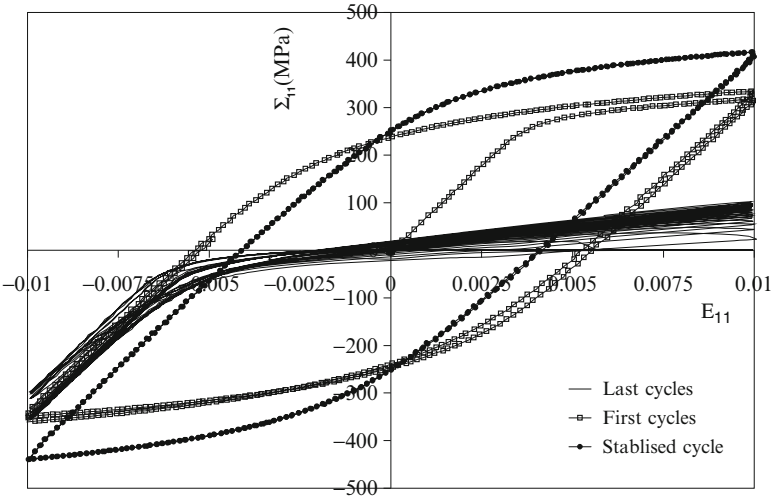


Fig. 5 Cyclic experimental response of aluminum 2024 under tension–compression (Abdul-Latif and Chadli 2007) (From Abdul-Latif et al. (1999), with permission from ASME Publications)

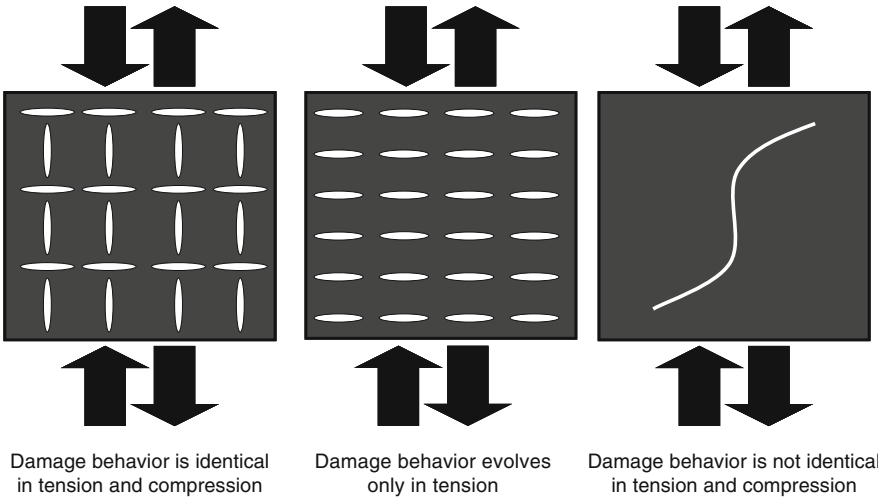


Fig. 6 Schematic representation of damage activation/deactivation (From Abdul-Latif et al. (1999), with permission from ASME Publications)

in a partial deactivation fashion (see Fig. 6). It is well known that this phenomenon is quite clear at the end of stage I of fatigue. But, it becomes remarkable during the microcrack propagation (stage II) when the microcracks are open.

An attempt is made to model such a phenomenon, especially at the end of stage I of the fatigue. Generally speaking, if the loading path reverses, the effect of damage

may become inactive (microcrack closing), and with a further evolution along the original loading path, the effect of damage may become reactive (microcrack opening). This means that the damage is active only if microcracks are open, while damage affects the mechanical properties of polycrystals during its closure (inactive phase) differently. To model this induced anisotropy due to damage under multiaxial cyclic loading conditions, a fourth-order damage tensor is proposed at the overall level.

Formulation of Micromechanical Modeling

A micromechanical model of the early fatigue damage initiation has been proposed by the author and his co-workers. Then, several applications have been conducted under different loading situations. This model considers that the intragranular damage evolves in bilateral manner whatever the applied loading direction (i.e., tension and compression). Its rate is exclusively governed by the inelastic part of the energy at the CSS level. Therefore, there is no damage if there is no intragranular isotropic hardening. Phenomena observed at these levels lower than the CSS are globally modeled at the CSS level. Moreover, the inelastic deformation of single crystal is defined by the motion of dislocations. Hence, for each single crystal, there is a set of slip systems, which can be activated depending on the loading condition as well as on the grain orientation with respect to the overall loading reference. As a starting point, slip is considered as a dominant deformation mechanism, and other mechanisms like twinning, grain boundary sliding, etc., are neglected. The constitutive equations of the inelastic strain are examined at the slip system scale. It is well known that the crystal plasticity deals with different material unit cells such as body-centered cubic (BCC), face-centered cubic (FCC), hexagonal close packed (HCP), etc. This means that mechanical properties of crystals depend on the crystal structure via the crystallographic planes and directions of a given grain containing a large concentration of atoms. The type of the unit cell should be defined to determine the number of slip systems that can be activated in plasticity. In this chapter, the FCC metallic structure will be considered only. Therefore, these slip systems are defined by $\{111\}$ crystallographic planes and $<110>$ directions. Only 12 octahedral slips are present, while, for other materials like nickel-base single crystal at elevated temperature, both cube and octahedral slips are taken into account. Due to the heterogeneous deformation pattern at this scale, the microstructure heterogeneities profoundly affect the overall material strength response. The resolved shear stresses τ^s are essentially defined by the resistance to dislocation motion due to the presence of microstructural inhomogeneities. Locally, in addition to the intragranular damage variables, only the intragranular isotopic hardening is defined. Two sources of kinematics hardening are physically observed at two different levels; the first one (intergranular) comes from the plastic strain incompatibility between grains giving nonuniform distribution of stresses at granular

level. The second source (intragranular) represents the long-range interactions inside the grain. In general, the intergranular kinematic hardening is rather predominant compared to the intragranular one for several materials. Thus, only the intergranular kinematic hardening variable is considered, i.e., neglecting the intragranular one. The granular inelastic strain rate is deduced as the sum of the contribution from all activated slip systems. The elastic part is supposed to be homogenous and isotropic and kept at the macroscopic level.

Recently, a new extension of the above approach has been developed by the author and his co-workers for describing the damage activation/deactivation effect in plastic fatigue with the small strain assumption. The definition of a microcrack initiation in LCF is still without a consensus answer; however, there is a general trend which considers that as soon as a microcrack attains a depth around 100 μm (i.e., comparable to the usual grain size), its propagation occurs across the section. Hence, the effect of damage activation/deactivation is quite clear at the end of stage I of fatigue and becomes clearer during the microcrack propagation (stage II) with crack opening phenomenon. It is worth emphasizing that the damage activation/deactivation effect is assumed to be described only at the RVE level. The model can therefore take into account the damage deactivation effect on the behavior of polycrystals under uniaxial and multiaxial cyclic loading conditions. As a final goal, the related phenomenon of the anisotropy induced by damage will be naturally described.

Choice of State Variables

The small strain assumption is adopted in this theoretical development. Thus, the overall total strain tensor $\underline{\underline{E}}$ is classically partitioned into elastic $\underline{\underline{E}}_e$ and inelastic $\underline{\underline{E}}_{in}$ parts:

$$\underline{\underline{E}} = \underline{\underline{E}}_e + \underline{\underline{E}}_{in} \tag{1}$$

At the macroscopic level (RVE, representative volume element), the elastic strain tensor $\underline{\underline{E}}_e$ represents an internal state variable associated with the Cauchy stress tensor $\underline{\underline{\Sigma}}$. It is assumed that the overall elastic deformation is homogenous. This assumption considers that all the grains of the RVE have the same elastic properties. The overall inelastic tensor $\underline{\underline{E}}_{in}$ is obtained through the localization–homogenization concept, i.e., micro–macro methodology.

At the granular level, the nonlinear intergranular kinematic hardening tensor $\underline{\underline{\beta}}^g$ is proposed as an internal state variable associated with the internal stress tensor $\underline{\underline{x}}^g$. It is worth noting that there is no damage variable associated to the grain boundaries (no creep damage) in this model.

As far as the CSS level is concerned, an internal state variable of isotropic hardening is introduced by the couple (q^s, R^s) for each octahedral slip system. This

variable describes the expansion of the elastic domain on the system(s). The intragranular damage variable (d^s, Y^s) is also proposed at this level. In fact, the intragranular damage will initiate and then evolve when the accumulated slip reaches a certain critical value. The dual variable (Y^s) is the thermodynamic force associated to the damage internal variable (d^s).

Based on the local damages, a scalar damage parameter (D^T) at the RVE level is determined by averaging procedure. It varies from zero for a virgin material to one for a totally damaged one (zero macro-stress). As a matter of fact, when $D^T = 1$, the macroscopic crack initiation stage is terminated (end of the stage I). Based on this parameter, a fourth-order damage tensor is defined to describe the concept of induced-oriented anisotropy.

Throughout this chapter, the index $s \in \{1,2,\ldots,n\}$ is associated to the system rank, with n being the maximum number of octahedral systems in the grain (e.g., $n = 12$ for FCC and $n = 24$ for BCC). Similarly, the index $g \in \{1,2,\ldots,N^g\}$ describes the grain rank, with Ng being the maximum number of grains contained in the RVE. It is important to note that slip which is the main phenomenon in plastic deformation usually occurs on these planes and directions with highest density of atoms. For FCC metallic materials, 24 possible slip systems can be activated. Based on the symmetry, only 12 possible slip systems could be activated, while in BCC case, 48 possible slip systems can be activated. Due to the symmetry, only 24 possibilities are defined. Note that the same concept can be entirely applicable for other different material unit cells such as hexagonal close packed.

Effective State Variables

Damage can be defined as the progressive deterioration of material prior to failure. From the modeling point of view, it is supposed that damage occurs at the CSS level where the plastic deformation is highly localized. When a local plastic strain attains a certain value (criterion depends explicitly on the accumulated slip as well as on the applied loading), the intragranular damage is therefore initiated ($d^s > 0$). The final stage of microcrack occurs when d^s has a critical value d^s_{cr}.

The effective state variables are defined by using the hypothesis of energy equivalence.

At the CSS level

$$\tilde{R}^s = \frac{R^s}{\sqrt{1-d^s}} \qquad \text{and} \qquad \tilde{q}^s = q^s\sqrt{1-d^s} \tag{2}$$

For the overall effective state variables, the damage activation and deactivation phases are modeled only at the macroscopic level through a tensorial approach. Hence, the effective overall stress and elastic strain tensors are defined by

$$\underline{\underline{\widetilde{E}}}e = \underline{\underline{E}}_e : \left(\underline{\underline{\underline{\underline{I}}}} - \underline{\underline{\underline{\underline{D}}}}\right)^{\frac{1}{2}} \quad \text{and} \quad \underline{\underline{\widetilde{\Sigma}}} = \left(\underline{\underline{\underline{\underline{\delta}}}}\right)^{\frac{1}{2}} : \underline{\underline{\Sigma}} \tag{3a}$$

with

$$\underline{\underline{\underline{\underline{\delta}}}} = \left(\underline{\underline{\underline{\underline{I}}}} - \underline{\underline{\underline{\underline{D}}}}\right)^{-1} \tag{3b}$$

where $\underline{\underline{\underline{\underline{I}}}}$ and $\underline{\underline{\underline{\underline{D}}}}$ are, respectively, the fourth-order identity and overall damage tensors. $\underline{\underline{\underline{\underline{D}}}}$ will be defined later in Eq. 14 depending on the overall scalar damage measure D^T (determined by the contribution of all intragranular damages in the RVE) and $\underline{\underline{\underline{\underline{I}}}}$.

Localization Process

The interaction law of a polycrystalline structure has its own role in defining the relation between the variables in each grain of the aggregate (RVE) and the overall variables. It is assumed that a polycrystalline metallic material is viewed as an aggregate of single crystals having various orientations. In the actual model, an appropriate self-consistent approach is used representing the grain-to-grain interaction. When the macroscopic Cauchy stress tensor $\underline{\underline{\Sigma}}$ is applied to the RVE, the granular Cauchy stress tensor $\underline{\underline{\sigma}}^g$ is consequently determined through the well-known self-consistent approach of Berveiller and Zaoui (1979) modified by Pilvin (1990). Thus, the used interaction law is given by

$$\underline{\underline{\sigma}}^g = \underline{\underline{\Sigma}} + C^g \left\{ \sum_{h=1}^{Ng} v^h \underline{\underline{\beta}}^h - \underline{\underline{\beta}}^g \right\} \tag{4}$$

where C^g is the material coefficient representing the intergranular hardening modulus and v^h the volume fraction of the same oriented grains.

The nonlinear accommodation can be obtained by means of Pilvin's modification, where Kröner solution with the difference of the local and global plastic strains is replaced by the difference between a local and a global nonlinear hardening variable. The concept of introducing the intergranular kinematic hardening variable ($\underline{\underline{\beta}}^g$) and its volume average on the whole aggregate presents a nonlinear evolution with respect to plastic strain.

Two important points require further discussion. The first one is related to the uniformity of the plastic strain concept within a grain. It represents one of the theoretical hypotheses related to the interaction laws (Eq. 4). The second point concerns the damage and its effect on the interaction law. In fact, the adopted

assumption considers that there is no damage variable associated to the grain boundaries in this model. An examination of Eq. 4 shows that the intergranular kinematic hardening variable ($\underline{\underline{\beta}}^{g}$) is implicitly affected by damage. As a matter of fact, since this variable is function of granular strain rate ($\dot{\underline{\underline{\varepsilon}}}^{g}_{in}$), therefore a direct coupling will be illustrated between $\underline{\underline{\varepsilon}}^{g}_{in}$ and the local damage d^{s} in Eq. 42. Thus, $\underline{\underline{\beta}}^{g}$ is implicitly affected by the local damage within a given grain. In spite of these two theoretical remarks, impressive capacities of the self-consistent approach are obtained where several cyclic hardening phenomena coupled with damage are faithfully described. Therefore, it is considered that the present approach is a pragmatic solution justified by the quality of recorded predictions.

After determining the stress tensor at each grain (Eq. 4), the resolved shear stress τ^{s} on each system for each grain can be expressed as a function of the granular stress $\underline{\underline{\sigma}}^{g}$ by means of the Schmid orientation tensor $\underline{\underline{\underline{\underline{m}}}}^{s}$, i.e., by the twice-contracted tensorial product between $\underline{\underline{\sigma}}^{g}$ and $\underline{\underline{\underline{\underline{m}}}}^{s}$:

$$\tau^{s} = \underline{\underline{\sigma}}^{g} : \underline{\underline{\underline{\underline{m}}}}^{s} \tag{5}$$

$$\underline{\underline{\underline{\underline{m}}}}^{s} = \frac{1}{2}\left[n^{s} \otimes g^{s} + g^{s} \otimes n^{s}\right] \tag{6}$$

where g^{s} is the unit vector in the slip direction and n^{s} being the vector normal to the slip plane.

Anisotropic Damage Modeling

The state of the system can be described by its free energy (state potential). This free energy (ψ) represents the sum of the reversible (ψ_{e}) and irreversible (ψ_{in}) energies per unit volume under small strains and isothermal conditions:

$$\rho\psi = \rho\psi_{e} + \rho\psi_{in} \tag{7}$$

where ρ is the density of the material.

To model the elastic behavior coupled with damage for a given metallic polycrystalline material, the following approach is developed. Actually, the attention is focused on the formulation of the elastic part which is regarded by two potentials at the RVE level. One corresponds to virgin (undamaged) state ψ^{e}_{o} and the other to damaged state ψ^{e}_{d}. The undamaged state potential is classically defined by

$$\rho\psi^{e}_{o} = \frac{1}{2}\,\underline{\underline{\Sigma}} \;:\; \underline{\underline{E}}_{e} = \frac{1}{2}\,\underline{\underline{\underline{\underline{R}}}}^{o} :\; \underline{\underline{E}}_{e} : \underline{\underline{E}}_{e} \tag{8}$$

where $\underline{\underline{\underline{\underline{R}}}}^{o}$ is the classical fourth-order rigidity tensor.

In the damaged state, the elastic potential is expressed by

$$\rho\psi_d^e = \frac{1}{2}\,\underline{\underline{\underline{\underline{R}}}}^{d} :: \underline{\underline{E}}_e : \underline{\underline{E}}_e \tag{9}$$

where $\underline{\underline{\underline{\underline{R}}}}^{d}$ represents the rigidity tensor for damaged material, defined by

$$\underline{\underline{\underline{\underline{R}}}}^{d} = \left(\underline{\underline{\underline{\underline{I}}}} - \underline{\underline{\underline{\underline{D}}}} \right) :: \underline{\underline{\underline{\underline{R}}}}^{o} \tag{10}$$

Thus, Eq. 9 can be rewritten as follows:

$$\rho\psi_d^e = \frac{1}{2}\left(\underline{\underline{\underline{\underline{I}}}} - \underline{\underline{\underline{\underline{D}}}} \right) :: \underline{\underline{\underline{\underline{R}}}}^{o} :: \underline{\underline{E}}_e : \underline{\underline{E}}_e \tag{11}$$

To determine the overall fourth-order damage tensor $\underline{\underline{\underline{\underline{D}}}}$, the projection operators are used. In the present formulation, stress-based (or strain-based) projection operators are adopted. This concept has been proposed to introduce a mode I microcrack opening and closure model based on a strain-based projection operator (Ju 1989). By considering the spectral decomposition of all macroscopic stress (or strain) tensors, one can consequently obtain

$$\underline{\underline{\Sigma}} = \sum_{i=1}^{3} \Sigma_i^* \, \underline{p}_i \otimes \underline{p}_i \tag{12}$$

where Σ_i^* is the *i*th principal stress and $\underline{p}_i$ the corresponding *i*th to eigenvalue and eigenvector of Σ^*, and the symbol $\otimes$ represents the tensor product. The fourth-order positive spectral projection tensor is thereafter expressed as follows:

$$P_{ijkl}^{+} = Q_{ia}^{+} \, Q_{jb}^{+} \, Q_{ka} \, Q_{lb} \tag{13}$$

The operator $\underline{\underline{\underline{\underline{P}}}}^{+}$ allows verifying naturally the complex phenomenon of activation/deactivation of damage and the active/passive passage criterion, especially under multiaxial loading. Indeed, it is capable to select the damage activation and deactivation phases in a given loading axis whatever the loading path complexity.

Taking into account the remarks made previously, the definition of the damage tensor $\underline{\underline{\underline{\underline{D}}}}$ can be deduced by the following equation:

$$\underline{\underline{\underline{\underline{D}}}} = D^T \underline{\underline{\underline{\underline{P}}}}^{+} \tag{14}$$

where D^T is the macroscopic damage measurement which will be determined later via the homogenization process of the local damages. By substituting the new form of the damage tensor in the overall elastic potential expression of damaged state given in Eq. 11, one thus gets

$$\rho\psi_d^e = \frac{1}{2}\left(\underline{\underline{\underline{\underline{I}}}} - D^T\underline{\underline{\underline{\underline{P}}}}^+\right) : \underline{\underline{\underline{\underline{R}}}}^o : \underline{\underline{E}}_e : \underline{\underline{E}}_e \tag{15}$$

This equation shows in the absence of damage that the potential of virgin material is effectively retrieved. This situation is also found when the damage is completely deactivated, i.e., $\underline{\underline{\underline{\underline{P}}}}^+ = \underline{\underline{\underline{\underline{0}}}}$.

Based on the thermodynamic concept, the overall stress tensor can be determined via the overall elastic potential function of $\underline{\underline{E}}_e$. In fact, this can be defined by deriving the overall stress tensor from the overall elastic potential with respect to the overall elastic strain tensor $\underline{\underline{E}}_e$ as shown by

$$\underline{\underline{\Sigma}} = \rho\frac{\partial\psi_d^e}{\partial\underline{\underline{E}}_e} = \underbrace{(\underline{\underline{\underline{\underline{I}}}} - D^T\underline{\underline{\underline{\underline{P}}}}^+) : \underline{\underline{\underline{\underline{R}}}}^o : \underline{\underline{E}}_e}_{❶} \underbrace{- \frac{1}{2}D^T\frac{\partial\underline{\underline{\underline{\underline{P}}}}^+}{\partial\underline{\underline{E}}_e} : \underline{\underline{\underline{\underline{R}}}}^o : \underline{\underline{E}}_e : \underline{\underline{E}}_e}_{❷} \tag{16}$$

The components of the tensor $\underline{\underline{\underline{\underline{P}}}}^+$ are functions of the eigenvectors. The second term in the right hand of Eq. 16 depends explicitly on the eigenvectors variation in the course of loading. In fact, when the loading is applied according to laboratory reference axes, the principal vectors coincide with the latter. In this case, these vectors are constant, i.e., their characteristics vary neither with respect to time nor with deformation. Thus, the second term in the right hand of Eq. 16 vanishes. Consequently, the constitutive equation of overall stress can be deduced as follows:

$$\underline{\underline{\underline{\underline{\Sigma}}}} = \left(\underline{\underline{\underline{\underline{I}}}} - D^T\underline{\underline{\underline{\underline{P}}}}^+\right) : \underline{\underline{\underline{\underline{R}}}}^o : \underline{\underline{E}}_e \tag{17}$$

The rigidity tensor coupled with damage (Eqs. 10 and 14) is written by

$$\underline{\underline{\underline{\underline{\Sigma}}}} = \underline{\underline{\underline{\underline{R}}}}^d : \underline{\underline{E}}_e \tag{18}$$

The overall stress tensor can be rewritten in a tensorial manner as

$$\Sigma_{ij} = R^d_{ijkl}\ E_{ekl} \tag{19}$$

However, when the eigenvectors vary with respect to time (e.g., loading condition having a shearing component), the two terms of Eq. 16 should completely be taken into account; thus

$$\underline{\underline{\underline{\underline{\Sigma}}}} = \underline{\underline{\underline{\underline{R}}}}^d : \underline{\underline{E}}_e + \underline{\underline{\underline{\underline{M}}}} \tag{20}$$

where

$$\underline{\underline{M}} = -\frac{1}{2} D^T \frac{\partial \underline{\underline{\underline{\underline{P}}}}^+}{\partial \underline{\underline{E}}_e} : \underline{\underline{\underline{\underline{R}}}}^o : \underline{\underline{E}}_e : \underline{\underline{E}}_e \tag{21}$$

In this case, the components of the overall stress tensor are defined by

$$\Sigma_{ij} = R^d_{ijkl} E_{ekl} + M_{ij} \tag{22}$$

It is clear that with the new term (M_{ij}), the determination of the overall stress is not standard. The presence of this term is undoubtedly due to the fact that, in the general loading case (especially with shear components), the principal vectors do not coincide with the laboratory reference axes. Therefore, its characteristics vary with time and deformation. Hence, the second term in the right hand of Eq. 16 becomes important to define correctly the overall constitutive equations of stress coupled with damage.

Now, Eq. 22 constitutes obviously a nonlinear transformation which can be physically interpreted by the nonlinearity effect induced by damage.

The rate of change of the overall stress tensor is described by

$$\dot{\underline{\underline{\Sigma}}} = \dot{\underline{\underline{\underline{\underline{R}}}}}^d : \underline{\underline{E}}_e + \underline{\underline{\underline{\underline{R}}}}^d : \dot{\underline{\underline{E}}}_e + \dot{\underline{\underline{M}}} \tag{23}$$

with

$$\dot{\underline{\underline{\underline{\underline{R}}}}}^d = \left(\dot{D}^T \underline{\underline{\underline{\underline{P}}}}^+ + D^T \dot{\underline{\underline{\underline{\underline{P}}}}}^+ \right) : \underline{\underline{\underline{\underline{R}}}}^o \tag{24}$$

and

$$\dot{\underline{\underline{M}}} = -\frac{1}{2} \dot{D}^T \frac{\partial \underline{\underline{\underline{\underline{P}}}}^+}{\partial \underline{\underline{E}}_e} : \underline{\underline{\underline{\underline{R}}}}^o : \underline{\underline{E}}_e : \underline{\underline{E}}_e - \frac{1}{2} D^T \left[\frac{\partial \dot{\underline{\underline{\underline{\underline{P}}}}}^+}{\partial \underline{\underline{E}}_e} \right] : \underline{\underline{\underline{\underline{R}}}}^o : \underline{\underline{E}}_e : \underline{\underline{E}}_e - D^T \frac{\partial \underline{\underline{\underline{\underline{P}}}}^+}{\partial \underline{\underline{E}}_e} : \underline{\underline{\underline{\underline{R}}}}^o : \dot{\underline{\underline{E}}}_e : \underline{\underline{E}}_e \tag{25}$$

Equation 23 emphasizes the fact that the presence of the damage and loading path complexity induces a high nonlinearity in the material behavior. This tensorial equation, although complex, has the advantage of generalizing for multiaxial loading paths.

Plastic State Potential

For a metallic polycrystalline material, the inelastic part of the macroscopic specific free energy ψ_{in} can be written as a sum of all granular inelastic potentials ψ^g_{in} in the aggregate given by

$$\psi_{in} = \sum_{g=1}^{Ng} \psi_{in}^{g} \tag{26}$$

The granular inelastic part of the state potential $\rho\psi_{in}^{g}$ is expressed as a quadratic function of the internal state variables of intergranular kinematic hardening and intragranular isotropic hardening:

$$\rho\psi_{in}^{g} = \frac{1}{3} C^{g} \underline{\underline{\beta}}^{g} \underline{\underline{\beta}}^{g} + \frac{1}{2} \sum_{r=1}^{n} \sum_{s=1}^{n} H_{rs} Q^{s} \tilde{q}^{r} \tilde{q}^{s} \tag{27}$$

where Q^{s} is the intragranular isotropic hardening modulus of the crystallographic slip system. The hardening interaction matrix H_{rs} is supposed to describe dislocation–dislocation interaction allowing the introduction of the cross influence of the slip of the system *s* on the hardening of the system *r*, belonging to the same family or not. Furthermore, for some metallic materials of FCC, only the octahedral slips (12 × 12 matrix) are taken into account. However, for other materials like nickel-base single crystals at elevated temperature, both cube and octahedral slips should be considered, i.e., 18 × 18 matrix. On the other hand, in the case of BCC, the H_{rs} matrix is of 24 × 24 as discussed above. All the presented applications of the model will be illustrated later considering only the FCC metallic polycrystalline materials. Therefore, a simple 12 × 12 matrix is chosen, neglecting the cube slips.

The dual variables (thermodynamic associated force variables) $\underline{\underline{x}}^{g}$, R^{s}, and Y_{in}^{s} (state laws) can be derived from Eq. 27 as follows:

$$\underline{\underline{x}}^{g} = \rho \frac{\partial \psi_{in}^{g}}{\partial \underline{\underline{\beta}}^{g}} = \frac{2}{3} C^{g} \underline{\underline{\beta}}^{g} \tag{28}$$

Note that in the actual micromechanical model, the internal variable $\underline{\underline{\beta}}^{g}$ is not directly affected by damage as discussed above.

The intragranular isotropic hardening variable coupled with damage can be deuced as follows:

$$R^{s} = \rho \frac{\partial \psi_{in}^{g}}{\partial q^{s}} = Q^{s} \sqrt{1 - d^{s}} \sum_{r=1}^{n} H_{rs} q^{r} \sqrt{1 - d^{r}} \tag{29}$$

The thermodynamic force Y_{in}^{s} (local damage inelastic energy) associated with the damage variable d^{s} at the CSS level is defined, considering only the intragranular isotropic hardening effect:

$$\overline{Y}_{in}^{s} = -Y_{in}^{s} = \rho \frac{\partial \psi_{in}^{g}}{\partial d^{s}} = \frac{\tilde{R}^{s} \tilde{q}^{s}}{2(1 - d^{s})} \tag{30}$$

New Local Damage Criterion

In this paragraph, a new formulation of local damage criterion is presented to describe the influence of the loading path complexity upon the fatigue life of metallic polycrystals. It is assumed that the damage initiation occurs at the local level when the accumulated slip attained a certain threshold value. Some of our numerical simulations performed previously display a certain limitation of the model in describing the loading path effect on the fatigue life for several random crystal distributions (grain aggregates), except a special distribution of 48 grains generated in a particular manner respecting the cubic symmetry, i.e., their grains are situated in the position of simple slip. It demonstrates, during loading, a phenomenon of competition between the damage inelastic energy release rate Y^s_{in}, accumulated slip for each system λ^s, and the interaction damage matrix D_{rs}. Thus, the loading path complexity has a considerable incidence on the hardening and Y^s_{in} evolutions. In other words, when the applied loading path becomes more complex, the intragranular isotropic hardening increases proportionally with the damage energy Y^s_{in}. Nevertheless, an obvious reduction in the accumulated slip λ^s occurs since the material is further work-hardened. In this case, the threshold concept γ^s_{th} already proposed is not enough to correctly take into account the influence of the loading path complexity on the fatigue lives. As a typical example, the difference in predicted fatigue lives for several random crystal distributions (grain aggregates) in tension–compression (uniaxial loading) and in tension–torsion with 90° out-of-phase angle (biaxial loading) is about 25 %. However, experimentally it is more than 600 % in the Waspaloy case.

In order to overcome such a theoretical shortcoming, a new definition of intragranular damage initiation is thus required. To accomplish this task, a new intragranular damage criterion is developed. It depends explicitly on λ^s as well as on the applied loading complexity. Therefore, the number of activated slip systems within a given aggregate of grains is considered. To give certain legitimacy to this development, it is experimentally recognized that the number of activated slip systems increases proportionally with loading path complexity for many engineering metals and alloys used in diverse industries like the nickel-based alloy, stainless steel 316L, etc. As a result, a new ratio (N_{sp}/N_{st}) is introduced, where N_{sp} is the number of activated slip systems and N_{st} represents the total number of systems likely to be activated (e.g., for BCC, $N_{st} = 24\times$ number of grains in the RVE; in the case of FCC, $N_{st} = 12\times$ number of grains in the RVE). This leads to the fact that the new intragranular damage nucleation d^s in fatigue can be started up by applying the following condition:

$$\left(1 - \frac{N_{sp}}{N_{st}}\right)^{-\alpha} \gamma^s_o \leq \lambda^s \tag{31}$$

It can be also written as

$$\gamma^s_{th} \leq \lambda^s \tag{32}$$

with

$$\gamma_{th}^{s} = \left(1 - \frac{N_{sp}}{N_{st}}\right)^{-\alpha} \gamma_{o}^{s} \tag{33}$$

where γ_{th}^{s} is the new damage criterion depending on the accumulated slip as well as on the applied loading path complexity, under which the local damage does not take place. γ_{o}^{s} is the slip reference supposed to be the same for all slip systems within the aggregate. α is a model characteristic parameter ensuring the coherence of the variations of recorded fatigue lives among the different applied loading paths. It is important to underline once again that all types of damage occurring at grain boundaries are totally neglected in this approach.

For a given aggregate of grains and loading type (x), the identification of the parameter α can be conducted by the following procedure: the intragranular damage within a grain initiates as soon as the accumulated plastic strain $\gamma^{s}(x)$ attains its threshold value γ_{th}^{s} defined by

$$\gamma^{s}(x) = \left(1 - \frac{N_{sp}(x)}{N_{st}}\right)^{-\alpha} \gamma_{o}^{s} \tag{34}$$

where $N_{sp}(x)$ is the number of activated slip systems within the aggregate under loading type x. By linearizing the Eq. 34, the following relation is thus deduced:

$$\ln(\gamma^{s}(x)) = -\alpha \ln\left(1 - \frac{N_{sp}(x)}{N_{st}}\right) + \ln \gamma_{o}^{s}(x) \tag{35}$$

The number of cycles needed to initiate the d^{s} is an important issue. Based on the experimental observation, this question can accordingly be determined. After identification of the parameters related to the elasto-inelastic behavior, some numerical simulations should be performed using the available experimental data. Such simulations permit to evaluate the accumulated slip $\gamma^{s}(x)$ and $N_{sp}(x)$ for each cyclic loading at the moment of the damage initiation. The values of $\gamma^{s}(x)$, associated to each loading x, are therefore defined. The coefficient α is thus determined directly as the slope of the linear relationship of a set of points where the coordinates of each point are given by $\left(\ln\left(1 - \frac{N_{sp}(x)}{N_{st}}\right), \ln \gamma_{o}^{s}(x)\right)$ and $\ln \gamma_{o}^{s}(x)$ is the ordinate into 0 of this line.

Dissipation Potential

After determining the associated force variables of the intergranular kinematic ($\underline{\underline{x}}^{g}$) and intragranular isotropic hardening (R^{s}) and damage inelastic energy release (Y_{in}^{s}), the formulation is completed by the rate equations of these internal state variables $\dot{\underline{\underline{\beta}}}^{g}$, $\dot{q}^{s}$ and $\dot{d}^{s}$, respectively. From the local inelastic flow point of view, the determination of the local inelastic flow is made by adopting a threshold concept.

Hence, the rate of change of these state variables can be obtained by the introduction of an elastic domain $\tilde{f}^s$ (local yield surface) for each slip system as well as a dissipation potential $\tilde{F}^s$ considering the non-associated plasticity case. Note that the nonlinearity of the intergranular kinematic and intragranular isotropic hardening can be introduced by the inelastic potential $\tilde{F}^s$. For the local inelastic flow, a slip system becomes active, once the absolute value of its resolved shear stress $|\tau^s|$ is greater than the actual flow surface radius $(\tilde{R}^s + k_o^s)$. The slip rate can be determined provided that the stress and the hardening variables are known. Thus, the local yield surface for each slip system with the presence of damage can be expressed as follows:

$$\tilde{f}^s = |\tau^s| - \tilde{R}^s - k_o^s \tag{36}$$

where k_o^s is the initial value of the critical resolved shear stress (friction stress).

The intragranular damage-inelastic dissipation potential can be written for each system as

$$\tilde{F}^s = \tilde{f}^s + \frac{3\,a^g}{4\,C^g}\,\underline{\underline{x}}^g : \underline{\underline{x}}^g + b^s q^s \tilde{R}^s + \sum_{r=1}^{n} D_{rs}\,\frac{S^s}{S_o^s + 1}\left(\frac{Y^s}{S^s}\right)^{S_0^s+1}\frac{H\left(\lambda^s - \gamma_{th}^s\right)}{\left(1 - d^s\right)^{ws}}\left(\frac{Y^r}{S^r}\right)^{s_o^s+1}\frac{H\left(\lambda^r - \gamma_{th}^r\right)}{\left(1 - d^r\right)^{ws}} \tag{37}$$

where a^g and b^s are material parameters describing the nonlinearity of the intergranular kinematic and intragranular isotropic hardenings, respectively. The coefficients S^s, S_o^s, w^s and γ_{th}^s are material constants characterizing the damage mechanism at the CSS level. $H(\lambda^s - \gamma_{th}^s)$ is the Heaviside function. It is equal to zero if $\lambda^s < \gamma_{th}^s$ and equal to 1 if $\lambda^s \geq \gamma_{th}^s$. The parameter γ_{th}^s represents, as discussed above, the new damage criterion given in Eq. 33. It measures the accumulation of the dislocation pileups on the CSS. The cross influence of the damage state of the system *s* on the damage evolution of the neighboring systems *s* of the same grain is described by the damage interaction matrix D_{rs}. In the case of FCC, this damage matrix is of 12 × 12, as in the hardening interaction matrix H_{rs}. Concentrated only on the intragranular cracking observed in Fig. 2 (case of Waspaloy) using scanning electron microscope and replicas observations of specimen outer surfaces, crack initiation occurs in some slip bands. It appears that a microcrack does not occur simultaneously along the entire length of a slip band, but may be confined to one part of the band. After initiation in slip bands, cracks zigzag from one slip band to another within a grain. This observation (cracks zigzagging) could be interpreted by microcrack interaction phenomenon. This gives legitimacy to introduce the damage interaction matrix D_{rs} in the intragranular damage evolution. For the sake of simplicity, it is assumed that D_{rs} has only two different parameters: the diagonal terms representing the self-damage interaction (equal to 1) and the non-diagonal

terms describing the interaction between the damage evolutions at the different system levels in the same grain.

The 12×12 damage interaction matrix D_{rs} is defined as follows:

$$D_{rs} = \begin{bmatrix} d_1 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 \\ d_2 & d_1 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 \\ d_2 & d_2 & d_1 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 \\ d_2 & d_2 & d_2 & d_1 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 \\ d_2 & d_2 & d_2 & d_2 & d_1 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 \\ d_2 & d_2 & d_2 & d_2 & d_2 & d_1 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 \\ d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_1 & d_2 & d_2 & d_2 & d_2 & d_2 \\ d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_1 & d_2 & d_2 & d_2 & d_2 \\ d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_1 & d_2 & d_2 & d_2 \\ d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_1 & d_2 & d_2 \\ d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_1 & d_2 \\ d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_2 & d_1 \end{bmatrix} \tag{38}$$

Using the generalized normality rule, the evolution laws are given:

- At the granular level

The granular inelastic strain can be deduced as follows:

$$\underline{\underline{\dot{\varepsilon}}}^{g}_{in} = \sum_{s=1}^{n} \dot{\lambda}^{s} \frac{\partial F^{s}}{\partial \underline{\underline{\sigma}}^{g}} = \sum_{s=1}^{n} \frac{\dot{\lambda}^{s}}{\sqrt{1-d^{s}}} \mathrm{sign}(\tau^{s}) \underline{\underline{m}}^{s} \tag{39}$$

with

$$\dot{\gamma}^{s} = \dot{\lambda}^{s} \mathrm{sign}(\tau^{s}) \tag{40}$$

$$\underline{\underline{\dot{\varepsilon}}}^{g}_{in} = \sum_{s=1}^{n} \frac{\dot{\gamma}^{s}}{\sqrt{1-d^{s}}} \underline{\underline{m}}^{s} \tag{41}$$

The intergranular kinematic hardening is expressed by

$$\underline{\underline{\dot{\beta}}}^{g} = -\sum_{r=1}^{n} \dot{\lambda}^{s} \frac{\partial F^{s}}{\partial \underline{\underline{x}}^{g}} = \underline{\underline{\dot{\varepsilon}}}^{g}_{in} - a^{g} \underline{\underline{\beta}}^{g} \sum_{s=1}^{n} \dot{\lambda}^{s} \tag{42}$$

where $\dot{\gamma}^{s}$ is the slip rate. Note that when $a^{g} \neq 0$ Eq. 42, this gives a nonlinear evolution rule of $\underline{\underline{\beta}}^{g}$; however when $a^{g} = 0$, i.e., $\underline{\underline{\dot{\beta}}}^{g} = \underline{\underline{\dot{\varepsilon}}}^{g}_{in}$, this leads to the Kröner rule interaction laws:

- At the CSS level

The rate of change of the intragranular isotropic hardening variable coupled with damage is derived as follows:

$$\dot{q}^s = -\sum_{r=1}^{n} \dot{\lambda}^r \frac{\partial \tilde{F}^r}{\partial R^r} = \frac{\dot{\lambda}^s}{\sqrt{1-d^s}}\left(1 - b^s q^s\right) \tag{43}$$

The intragranular damage evolution is derived as shown by

$$\dot{d}^s = \dot{\lambda}^s \frac{\partial \tilde{F}^s}{\partial Y_{in}^s} = \dot{\lambda}^s \left(\frac{\overline{Y}_{in}^s}{S^s}\right)^{S_0^s} \frac{H\left(\lambda^s - \gamma_{th}^s\right)}{\left(1-d^s\right)^{ws}} \sum_{r=1}^{n} D_{rs} \left(\frac{\overline{Y}_{in}^r}{S^r}\right)^{s_o^s+1} \frac{H\left(\lambda^r - \gamma_{th}^r\right)}{\left(1-d^r\right)^{ws}} \tag{44}$$

In the framework of viscoplasticity, the value of pseudo-multiplier $\dot{\lambda}^s$ for each slip system is a power function of the distance to the yield point defined by the criterion $\tilde{f}^s$:

$$\dot{\lambda}^s = \left\langle \frac{\tilde{f}^s}{K^s} \right\rangle^{z^s} = \left\langle \frac{|\tau^s| - \tilde{R}^s - k_o}{K^s} \right\rangle^{z^s} \tag{45}$$

where K^s and z^s are material constants describing the local viscous effect of the material. It is worth emphasizing that since the rate-independent (plastic) models (considering that plastic flow at the slip system level is a rate independent) do not possess the uniqueness in the numerical applications, rate-dependent slip is thus adopted to resolve such numerical difficulties used previously by several researchers. Although the developed model is a rate-dependent (viscoplastic) type, the rate-independent case can be practically obtained by choosing a high value of viscous exponent z^s and a low value of the coefficient K^s. The viscosity effect can be minimized, i.e., a low and constant viscous stress σ_v^s can be obtained by this relation:

$$\sigma_v^s = K^s |\dot{\gamma}^s|^{1/Z^s} \tag{46}$$

Positivity of the Intrinsic Dissipation

The volumetric intrinsic dissipation $\mathfrak{I}_{Dis}^s$ for each activated slip system should be positive according to the second law of thermodynamics. By assuming the partitioning between the damage dissipation $\mathfrak{I}_d^s$ and the inelastic one $\mathfrak{I}_{in}^s$, thus the two quantities $\mathfrak{I}_{Dis}^s = \mathfrak{I}_d^s + \mathfrak{I}_{in}^s$ can be expressed as follows:

$$\mathfrak{I}_d^s = \overline{Y}_{in}^s \dot{d}^s \geq 0 \tag{47}$$

$$\mathfrak{I}_{in}^s = \tau^s \dot{\gamma}^s + R^s \dot{q}^s \geq 0 \tag{48}$$

According to the fact that $\overline{Y}_{in}^s$ is always positive, the inequality Eq. 47 implies that $\dot{d}^s \geq 0$, which means that the intragranular damage $\dot{d}^s$ cannot decrease.

The second inequality Eq. 48 is verified as long as the constants b^s Q^s and k_o^s are positive.

Homogenization

It is obvious that the overall properties are function of grain properties. Thus, each grain is assumed to be heterogeneous embedded in a homogenous equivalent medium having an average response of all grains. In this context, the overall inelastic strain rate tensor is obtained through the micro–macro methodology, i.e., there are no state variables at this level. Hence, the macroscopic Cauchy stress tensor $\underline{\underline{\Sigma}}$ is also deduced by the homogenization process. After determining the granular inelastic strain rate as the sum of the contribution from all activated slip systems, the transition from the single to polycrystal response is performed by the well-known averaging procedure depending on the granular inelastic strain rate. It is worth noting that, for homogenous elastic media, it has been demonstrated (Mandel 1965; Bui 1969) that the overall stresses $\underline{\underline{\Sigma}}$ are simple averages of granular stresses $\underline{\underline{\sigma}}^{g}$. However, this is not the case for the overall inelastic strain where the averaging procedure is usually not simple but involves localization tensors (Mandel 1971). Nevertheless, in the case of elasto-inelastic behavior with homogeneous elasticity, the overall inelastic strain rate $\dot{\underline{\underline{E}}}_{in}$ can be calculated as an average of the granular inelastic strain rate $\dot{\underline{\underline{\varepsilon}}}^{g}_{in}$ (Bui 1969). In a more precise manner, in the special case of a single-phase polycrystal, as in this chapter, the overall inelastic strain rate is equal to the average of granular rate $\dot{\underline{\underline{\varepsilon}}}^{g}_{in}$ (Mandel 1971). The rate of change of the overall inelastic strain is therefore determined by the following homogenization procedure:

$$\dot{\underline{\underline{E}}}_{in} = \sum_{g=1}^{Ng} v^{g} \dot{\underline{\underline{\varepsilon}}}^{g}_{in} \tag{49}$$

where v^{g} represents the volume fraction of the same oriented grains.

On the other hand, the macroscopic measure of the damage D^{T} should be used, since the coupled elastic behavior is calculated at the overall level as mentioned above. This can be achieved by using the notion of damaged grains $N_{D}^{g'}$ and their volume fraction of the same oriented damaged grains $v_{D}^{g'}$. This leads always to $v_{D}^{g'} \leq v^{g}$. The same concept is also used over only these slip systems n' where the local damage takes place at their level:

$$\dot{D}^{T} = \sum_{g=1}^{N_{D}^{g'}} v_{D}^{g'} \sum_{s=1}^{n'} \frac{\dot{d}^{s}}{n'} \tag{50}$$

The complexity of the overall damage initiation criterion (i.e., totally damaged aggregate) is undeniable, particularly in the case of micromechanical approach. Hence, the notion of damaged grains and damaged system seems to be a reasonable manner to define a macrocrack initiation criterion.

Conclusion

The anisotropy-induced behavior motivated by damage is of particular interest in this chapter. Based on the low-cycle fatigue description, the micromechanical model is expressed in the rate-independent plasticity with the small strain assumption. A new criterion depending on the accumulated slips as well as on the applied loading path is demonstrated. After introducing the damage variable at the slip level, the coupling elasto-inelastic damage is then accomplished using the concept of the effective state variables defined by the hypothesis of energy equivalence. The damage activation/deactivation phenomenon is formulated and treated at the macroscopic scale using the mathematical operators of projection. These operators allow defining a fourth-order damage tensor. These operators are capable to take into account the damage deactivation effect in the case of the multiaxial cyclic loadings describing naturally the oriented anisotropy phenomenon induced by damage.

Acknowledgments It is certainly not a classical grateful, nonetheless it is in memory of a friend and ex-Ph. D. student Blondin Mounounga, since a part of the present effort has been developed with him.

References

A. Abdul-Latif, M. Chadli, Modeling of the heterogeneous damage evolution at the granular scale in polycrystals under complex cyclic loadings. Int. J. Damage Mech. **16**, 133–158 (2007)

A. Abdul-Latif, B.S.T. Mounounga, Damage deactivation modeling under multiaxial cyclic loadings for polycrystals. Int. J. Damage Mech. **18**, 177–198 (2009)

A. Abdul-Latif, V. Ferney, K. Saanouni, Fatigue damage of the Waspaloy under complex loading. ASME J. Eng. Mat. Tech. **121**, 278–285 (1999)

M. Berveiller, A. Zaoui, An extension of the self-consistent scheme to plasticity flowing polycrystals. J. Mech. Phys. Solids **26**, 325 (1979)

H.D. Bui, Étude de l'Évolution de la Frontière du Domaine Élastique avec l'Écrouissage et Relations de Comportement Elastoplastique de Métaux Cubiques, Thèse de Doctorat ès Sciences, Paris, 1969

J.L. Chaboche, Development of continuum damage mechanics for elastic solids sustaining anisotropic and unilateral damage. Int. J. Damage Mech. **3**, 311 (1993)

K. Dang Van, Sur la Resistance à la Fatigue des Materiaux. Sciences et Techniques de l'Armement, 47, Mémorial de l'Artillerie Française, 3ème Fascicule, 1973

P. Germain, Q.S. Nguyen, P. Suquet, Continuum thermodynamics. J. Appl. Mech. **105**, 1010 (1983)

A.L. Gurson, Continuum theory of ductile rupture by void nucleation and growth- part I. Yield criteria and flow rules for porous ductile media. J. Eng. Mat. Technol. **99**, 2–15 (1977)

N.R. Hansen, H.L. Schreyer, Damage deactivation. ASME J. Appl. Mech. **62**, 450 (1995)

J.W. Ju, On energy-based coupled elastoplastic damage theories: constitutive modeling and computational aspects. Int. J. Solids Struct. **25**, 803 (1989)

B.A. Lerch, N. Jayaraman, S.D. Antolovich, A study of fatigue damage mechanisms in Waspaloy from 25 to 800 °C. Mater. Sci. Eng. **66**, 151 (1984)

J. Mandel, Une Généralisation de la Théorie de la Plasticité de W. T. Koiter. Int. J. Solids Struct. **1**, 273 (1965)

J. Mandel, *Plasticité Classique et Viscoplasticité*. Cours CISM, vol. 97 (Springer, Udine, 1971)
T. Mura, *Micromechanics of Defects in Solids* (Martinus Nijhoff, 1982)
M.W. Parsons, K.L. Pascoe, Observations of surface deformation, crack initiation and crack growth in low-cycle fatigue under biaxial stress. Mat. Sci. Eng. **22**, 31 (1976)
P. Pilvin, Approches Multiéchelles pour la Prévision du Comportement Anélastique des Métaux, Ph. D. Thesis, Université Paris VI, Paris, France, 1990
K. Tanaka, T. Mura, A dislocation model for fatigue crack initiation. ASME J. Appl. Mech. **48** (1981)
V. Tvergaard, On localization in ductile materials containing spherical voids. Int. J. Fract. **18**, 237–252 (1982)

From Single Crystal to Polycrystal Plasticity: Overview of Main Approaches

12

Esteban P. Busso

Contents

Abstract

This chapter provides a brief overview of the different continuum mechanics approaches used to describe the deformation behavior of either single crystals or individual grains in polycrystalline metallic materials. The crucial role that physics-based crystal plasticity approaches may play in understanding the mechanisms of damage initiation and growth is addressed. This includes a discussion of the main strain gradient constitutive approaches used to describe size effects in crystalline solids. Finally, representative examples are given about the effect of the local stress and strain fields in the mechanisms of intergranular damage initiation and growth in FCC polycrystal materials.

E.P. Busso (✉)
Materials and Structures Branch, ONERA, Palaiseau, Cedex, France
e-mail: esteban.busso@onera.fr; esteban.busso@ensmp.fr

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_7

Introduction

It is well understood that the macroscopic phenomena which control the physical and mechanical properties of materials originate from the underlying microstructure. The chemical and phase compositions, microstructural morphology, and characteristic length scales, such as grain size or mean dislocation spacing, have a significant effect on the material's properties and behavior. Furthermore, the evolution of the material's microstructure subject to thermal and mechanical loads typical of service is equally important as it determines its long-term properties. This includes the development and growth of internal defects (or damage) such as voids, intergranular or transgranular cracks which generally lead to a life-limiting microstructural state. The identification of physics-based relationships between microstructure and macroscopic behavior is one of the crucial issues which engineers, materials scientists, and physicists alike face and which have been at the center of most recent research efforts. A common goal is the development of physics-based analytical and computational material modelling tools to underpin scientific investigations and complement traditional theoretical and experimental approaches. The power of analytical theories lies in their ability to reduce the complex collective behavior of the basic ingredients of a solid (e.g., electrons, atoms, lattice defects, single crystal grains) into insightful relationships between cause and effect. Computational approaches such as those based on multiscale material modelling techniques are required to complement continuum and atomistic analysis methods. At transitional (or microstructural) scales, such as those in-between continuum and atomistic, continuum approaches begin to break down, and atomistic methods reach inherent time and length-scale limitations (Ghoniem et al. 2003).

Transitional theoretical frameworks and modelling techniques are being developed to bridge the gap between length-scale extremes. For example, the description of deformation beyond the elastic regime is usually described by appropriate *constitutive equations*, and the implementation of such relationships within continuum mechanics generally relies on the inherent assumption that material properties vary continuously throughout the solid. However, certain heterogeneities linked to either the microstructure, such as dislocation patterns or the deformation per se cannot be readily described within the framework provided by continuum mechanics. New promising application areas require novel and sophisticated physically based approaches for design and performance prediction. Thus, theory and modelling are playing an ever increasing role to reduce development costs and manufacturing times as well as to underpin computational material design. In the last decade or so, there has been a shift away from reproducing known properties of known materials and toward simulating the behavior of possible alloys as a forerunner to finding real materials with these properties.

In high-payoff, high-risk technologies such as those required in the design of large structures in the aerospace and nuclear industries, the effects of aging and environment on failure mechanisms cannot be left to conservative approaches. Increasing efforts are now focused on developing multiscale materials modelling

approaches to develop new alloys and material systems in these areas. Appropriate validation experiments are also crucial to verify that the models predict the correct behavior at each length scale, ensuring that the linkages between approaches are directly enforced.

As the material dimensions become smaller, its resistance to deformation is increasingly determined by internal or external discontinuities (e.g., surfaces, grain boundaries, dislocation cell walls). The Hall–Petch relationship has been widely used to explain grain size effects, although the basis of the relationship is strictly related to dislocation pileups at grain boundaries. Recent experimental observations on nanocrystalline materials with grains of the order of 10–70 nm indicate that the material is weaker than what would be expected from the Hall–Petch relationship. Thus, the interplay between interfacial or grain boundary effects and slip mechanisms within a single crystal grain may result in either strength or weakness, depending on their relative sizes. Although experimental observations of plastic deformation heterogeneities are not new, the significance of these observations has not been addressed till very recently. The length scales associated with these deformation patterns (e.g., typically the size of dislocation cells, the ladder spacing in persistent slip bands (PSBs), or the spacing between coarse shear bands) control the material strength and ductility. As it may not be possible to homogenize such types of microstructures in an average sense using either atomistic simulations or continuum theories, new intermediate approaches are needed.

The issues discussed above, in addition to the ever increasingly powerful and sophisticated computer hardware and software available, are driving the development of materials modelling approaches. New concepts, theories, and computational tools are being continuously developed to enable the prediction of deformation phenomena at different microstructural scales to be linked. This chapter is aimed at providing a brief overview of the different approaches that are being used to deal with the continuum mechanics modelling of plasticity at the grain/single crystal level. Special emphasis is placed on highlighting the crucial role that physics-based crystal plasticity approaches play in developing an understanding of the local stress and strain fields known to be the precursors to damage initiation and growth at the scale of the grain in polycrystal metallic materials. Some representative examples are also given about the use of single crystal theories to predict polycrystal behavior.

Continuum Discretization of a Boundary Value Problem

In this section, an overview will first be given of the main continuum mechanics-based framework used to describe the nonlinear deformation behavior of materials at the local (e.g., single phase or grain level) scale. Emphasis will be placed on recent progress made in crystal plasticity and strain gradient plasticity.

Standard tensorial notation will be used throughout. Unless otherwise specified, vectors will be described by boldface lowercase letters, second-order tensors by boldface uppercase letters, and fourth-order tensors by italic uppercase letters.

In a generic boundary value problem (BVP), the deformation of a body subjected to external forces and prescribed displacements is governed by the (i) equilibrium equations, (ii) constitutive equations, (iii) boundary conditions, and (iv) initial conditions. The "weak" form of the boundary value problem is obtained when the equilibrium equations and the boundary conditions are combined into the "principle of virtual work." Such "weak form" constitutes the basis for obtaining a numerical solution of the deformation problem via, e.g., the finite element method. Thus, in a continuum mechanics Lagrangian formulation of a quasi-static BVP, the principle of virtual work is the vehicle by which the global equilibrium equations are obtained.

The basic features of a generic Galerkin-type discretization framework are given next. Consider a structure occupying a domain V in the deformed configuration which is subjected to external forces and displacements on its boundary, Γ_b. In the absence of body forces and inertial effects, the principle of virtual work for the structure, in its rate form, satisfies the following equation:

$$\int_V \boldsymbol{\sigma} : \delta\dot{\boldsymbol{\varepsilon}} \mathrm{d}V - \int_{\Gamma_b} \mathbf{t} \cdot \delta\mathbf{v} \mathrm{d}\Gamma_b = 0 \tag{1}$$

for any arbitrary virtual velocity vector field $\delta\mathbf{v}$ compatible with all kinematics constraints. In the above equation, $\mathbf{t} = \boldsymbol{\sigma}\mathbf{n}_s$ represents the boundary traction forces; $\boldsymbol{\sigma}$, the Cauchy stress; $\mathbf{n}_s$, the normal to the surface on which the tractions act; and, $\delta\mathbf{v}$ the virtual strain rate associated with the velocity field.

To solve a complex BVP numerically, the discretization of the principle of virtual work is generally performed using the finite element method. Let $\mathbf{v}$ be approximated at a material point within an element by

$$\mathbf{v} = \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{N_{max}}} N^i \, \hat{\mathbf{v}}^{\mathbf{i}} \equiv \boldsymbol{N}\,\hat{\mathbf{v}} \tag{2}$$

where $\hat{\mathbf{v}}$ denotes the nodal values of the element velocity field and $\mathbf{N}$ are the isoparametric shape functions. Substituting Eq. 2 into Eq. 1 leads to the discretized version of the principle of virtual work on the finite element, V_e:

$$\mathbf{r}\{\hat{\mathbf{v}}\} \equiv \mathbf{f}^{\mathbf{int}} - \mathbf{f}^{\mathbf{ext}} = 0, \tag{3}$$

where

$$\mathbf{f}^{\mathrm{int}} = \int_{V_e} \mathbf{B}^{\mathrm{T}} \boldsymbol{\sigma} \mathrm{d}V_e, \quad \mathbf{f}^{\mathrm{ext}} = \int_{\Gamma_e} \mathbf{N}^{\mathrm{T}} \mathbf{t} \, \mathrm{d}\Gamma_e \tag{4}$$

are the internal and external global force vectors, respectively, and $\mathbf{B}$ relates the symmetric strain rate tensor with $\hat{\mathbf{v}}$. The global equilibrium relations (Eq. 3) represent a set of implicit nonlinear equations which may be solved incrementally

using a Newton-type algorithm. In a Newton–Raphson iterative scheme, the nonlinear system (Eq. 3) is typically expanded using Taylor series in the neighborhood of $\hat{\mathbf{v}}$:

$$\mathbf{r}\left\{\hat{\mathbf{v}}^{\mathrm{k}} - \delta\hat{\mathbf{v}}^{\mathrm{k}}\right\} = \mathbf{r}\left\{\hat{\mathbf{v}}^{\mathrm{k}}\right\} + \frac{\partial \mathbf{r}\left\{\hat{\mathbf{v}}^{\mathrm{k}}\right\}}{\partial \hat{\mathbf{v}}^{\mathrm{k}}}\delta\hat{\mathbf{v}}^{\mathrm{k}} + O\left\{\hat{v}^{k^2}\right\}, \tag{5}$$

where k represents a generic iteration and $\partial\mathbf{r}/\partial\hat{\mathbf{v}}$ is the global tangent stiffness or Jacobian matrix of the nonlinear system of equations. The formulation of accurate estimates of the global Jacobian is at the heart of most numerical schemes developed to provide robust algorithms for the use of complex constitutive models with continuum approaches (e.g., see Crisfield (1997), Busso et al. (2000), and Meissonnier et al. (2001)).

Single Crystal Plasticity

Constitutive models developed to predict the anisotropic behavior of single crystal materials generally follow either a Hill-type or a crystallographic approach. As a common feature, they treat the material as a continuum in order to describe properly plastic or viscoplastic effects. Hill-type approaches (e.g., Schubert et al. 2000) are based on a generalization of the Mises yield criterion proposed by Hill (1950) to account for the non-smooth yield or flow potential surface required to describe the anisotropic flow stress behavior of single crystals. By modelling polycrystal structures with an appropriate crystallographic formulation based on microstructural internal state variables (e.g., dislocation densities), greater insight into the grain interaction and deformation behavior of polycrystals can be achieved. In constitutive formulations based on crystallographic slip, the macroscopic stress state is resolved onto each slip system following the Schmid law. Internal state variables are generally introduced in both formulations to represent the evolution of the microstructural state during the deformation process. Although recent developments in these two approaches have now reached an advanced stage, the major improvements have been made by crystallographic models due to their ability to incorporate complex micromechanisms of slip within the flow and evolutionary equations of the single crystal models. Typically, in dislocation density-based models, the evolution of the dislocation structure is described by processes of dislocation multiplication and annihilation as well as by the trapping of dislocations (Peeters et al. 2001; Zikry and Kao 1996). Further discretization into pure edge and screw types enables their individual roles to be more clearly distinguished (Arselins and Parks 2001). For example, edge and screw dislocations are associated with different dynamic recovery processes (i.e., climb for edges and cross-slip for screws), combining to influence the evolving dislocation structure of a deforming material. Moreover, it is now possible through X-ray profile analyses to quantify the edge and screw dislocation densities in deformed metals (Kysar et al. 2010; Dunne et al. 2012). Therefore, the ability to make quantitative rather than just qualitative

comparisons between predicted and measured dislocation densities constitutes a powerful tool. However, the roles of edge and screw dislocations in determining the nonuniform distribution of plastic strain in polycrystals as a result of intergranular and intragranular interactions are still not well understood.

A brief outline of the salient features of local and nonlocal crystal plasticity approaches is given below.

Local Single Crystal Approaches

A generic internal variable-based crystallographic framework is said to be a local one when the evolution of its internal variables can be fully determined by the local microstructural state at the material point. The description of the kinematics of most crystal plasticity theories follows that originally proposed in Asaro and Rice (1977), which has been widely reported in the computational mechanics literature (e.g., Kalidindi et al. 1992; Busso and McClintock 1996; Hatem and Zikry 2009; Busso et al. 2000; Abrivard et al. 2012). It relies on the multiplicative decomposition of the total deformation gradient, $\mathbf{F}$, into an inelastic, $\mathbf{F^p}$, and an elastic, $\mathbf{F^e}$ component. Thus, under isothermal conditions,

$$\mathbf{F} = \mathbf{F^e}\,\mathbf{F^p}. \tag{6}$$

Although single crystal laws can be formulated in a corrotational frame, i.e., the stress evolution is computed on axes which rotate with the crystallographic lattice, the most widely used approach is to assume that the material's response is hyperelastic, that is, its behavior can be derived from a potential (i.e., free energy) function. Such potential function may be expressed in terms of the elastic Green–Lagrange tensorial strain measure,

$$\mathbf{E^e} = \frac{1}{2}\left(\mathbf{F^{e^T}}\mathbf{F^e} - 1\right), \tag{7}$$

and the corresponding objective work conjugate (symmetric) stress, or second Piola–Kirchhoff stress, $\mathbf{T}$. Note that the Cauchy stress is related to $\mathbf{T}$ by

$$\sigma = \det\{\mathbf{F^e}\}^{-1}\mathbf{F^e}\mathbf{T}\,\mathbf{F^{e^T}}. \tag{8}$$

The hyperelastic response of the single crystal is governed by

$$\mathbf{T} = \frac{\partial \Phi\{\mathbf{E^e}\}}{\partial \mathbf{E^e}}, \tag{9}$$

where $\partial\Phi/\partial\mathbf{E}^{\mathrm{e}}$ represents the Helmholtz potential energy of the lattice per unit reference volume. Differentiation of Eq. 9, and assuming small elastic stretches, yields

$$\mathbf{T} \cong \boldsymbol{L}:\mathbf{E}^{\mathbf{e}}, \tag{10}$$

where $\boldsymbol{L}$ is the anisotropic linear elastic moduli. In rate-dependent formulations, the time rate of change of the inelastic deformation gradient, $\mathbf{F}^{\mathbf{p}}$, is related to the slipping rates on each slip system (Asaro and Rice 1977) as

$$\dot{\mathbf{F}}^{\mathrm{p}} = \left(\sum_{\alpha=1}^{n_\alpha} \dot{\gamma}^\alpha \mathbf{P}^\alpha\right)\mathbf{F}^{\mathrm{p}}, \qquad \text{with} \qquad \mathbf{P}^\alpha \equiv \mathbf{m}^\alpha \otimes \mathbf{n}^\alpha. \tag{11}$$

Here, $\mathbf{m}^\alpha$ and $\mathbf{n}^\alpha$ are unit vectors defining the slip direction and the slip plane normal to the slip system.

In rate-independent formulations, in contrast, flow rules are based on the well-known Schmid law and a critical resolved shear stress, τ_c^α, whereby the rate of slip is related to the time rate of change of the resolved shear stress, $\tau^\alpha (= \mathbf{T}:\mathbf{P}^\alpha)$. Then,

$$\dot{\tau}^\alpha = \dot{\tau}_c^\alpha = \sum_{\beta=1}^{n_\alpha} h^{\alpha\beta}\dot{\gamma}^\alpha, \text{ if } \qquad \dot{\gamma}^\alpha > 0. \tag{12}$$

In the above equation, $h^{\alpha\beta}$, the slip hardening matrix coefficients, incorporate latent hardening effects. Due to the severe restrictions placed on material properties, such as latent hardening, to ensure uniqueness in the mode of slip (e.g., Anand and Kothari 1996; Busso and Cailletaud 2005), and the associated difficulties in its numerical implementation, the use of rate-independent formulations has been somehow restricted and much more limited than rate-dependent ones. This has been compounded by the fact that, by calibrating their strain rate sensitivity response accordingly, rate-dependent models have been successfully used in quasi-rate-independent regimes. Thus, the focus of the discussions will henceforth be on rate-dependent approaches.

The slip rate in Eq. 11 can functionally be expressed as

$$\dot{\gamma}^\alpha = \hat{\dot{\gamma}}^\alpha\left\{\tau^\alpha, \mathrm{S}_1^\alpha, \ldots, \mathrm{S}_{m_s}^\alpha, \theta\right\}, \tag{13}$$

where S_i^α (for $i = 1,\ldots,m_s$) denotes a set of internal state variables for the slip system α and θ is the absolute temperature. A useful and generic expression for the overall flow stress in the slip system can be conveniently found by inverting Eq. 13. Let us, for instance, consider a case with three slip resistances ($m_s = 3$). Then,

$$\tau^\alpha = \pm\hat{f}_v^\alpha\left\{\dot{\gamma}^\alpha, \mathrm{S}_3^\alpha, \theta\right\} \pm c_{\mathrm{dis}}\mathrm{S}_1^\alpha \pm c_{\mathrm{ss}}\mathrm{S}_2^\alpha, \tag{14}$$

where c_{dis} and c_{ss} are scaling parameters, S_1^α and S_2^α represent additive slip resistances, and S_3^α represents a multiplicative component. Here the distinction between the additive (S_1^α and S_2^α) and the multiplicative (S_3^α) slip resistances is motivated by the additive and multiplicative use of nondirectional hardening variables rather than on mechanistic considerations. By expressing the flow stress in the slip system α, as in Eq. 14, the contributions from viscous effects (first term in Eq. 14), and

dissipative (e.g., hardening, recovery) mechanisms arising from, for instance, forest dislocation and solid solution strengthening (second and third terms), can be clearly identified. The majority of formulations rely on power law functions for Eq. 13, where the resolved shear stress is normalized by a slip resistance or hardening function, which corresponds to $S_3^\alpha \neq 0$ and $S_1^\alpha = S_2^\alpha = 0$ in Eq. 14. This introduces a coupling between the viscous term and microstructure which is inconsistent with most strengthening mechanisms. Works such as that of Busso and McClintock (1996) and Cheong and Busso (2004) have proposed flow stress relations with $S_1^\alpha \neq 0$ and $S_2^\alpha = S_3^\alpha = 0$ which allows a more physically meaningful interpretation of strengthening phenomena controlled by the dislocation structure. The particular application for FCC polycrystals to be discussed in the next section assumes that $S_1^\alpha \neq 0$, $S_2^\alpha \neq 0$, and $S_3^\alpha = 0$. For a more detailed discussion of these issues, see also Busso and Cailletaud (2005).

The relation between the overall slip resistance associated with statistically stored dislocation forest type of obstacles and the individual dislocation densities is defined by

$$S_i^\alpha = \lambda\, \mu\, b^\alpha \left\{ \sum_\beta h^{\alpha\beta} \rho_i^\beta \right\}^{1/2} \qquad (\text{for } i = 1, \ldots, n_s) \tag{15}$$

Here, λ is a statistical coefficient which accounts for the deviation from regular spatial arrangements of the dislocations, b^α represents the magnitude of the Burgers vector, and $h^{\alpha\beta}$ is a dislocation interaction matrix defined as

$$h^{\alpha\beta} = \omega_1 + (1 - \omega_2)\delta^{\alpha\beta} \tag{16}$$

The terms ω_1 and ω_2 in Eq. 16 are the interaction coefficients and $\delta^{\alpha\beta}$ is the Kronecker delta. The corresponding total athermal slip resistance due to forest dislocations can then be expressed according to

$$S_{\text{dis}}^\alpha = \left\{ \left(S_1^\alpha\right)^r + \left(S_2^\alpha\right)^r + ... + \left(S_{n_s}^\alpha\right)^r \right\}^{1/r} \tag{17}$$

with $r = 1$ being used when a linear sum of the slip resistances is desired and $r = 2$ for a mean square value.

To complete the set of constitutive relations, separate evolutionary equations need to be formulated for the individual dislocation densities, with dislocation multiplication and annihilation forming the bases of their evolutionary behavior. The time rate of change of each internal slip system variable can, in its most general form, be expressed as

$$\begin{aligned} \dot{\rho}_1^\alpha &= \hat{\dot{\rho}}_1^\alpha \left\{ \dot{\gamma}^\alpha, \rho_1^\alpha, \rho_2^\alpha, ..., \rho_{n_s}^\alpha, \theta \right\}, \dot{\rho}_2^\alpha = \hat{\dot{\rho}}_2^\alpha \left\{ \dot{\gamma}^\alpha, \rho_1^\alpha, \rho_2^\alpha, ..., \rho_{n_s}^\alpha, \theta \right\}, \begin{matrix} \cdot\cdot \\ \cdot\cdot \\ \cdot\cdot \end{matrix} \dot{\rho}_{n_s}^\alpha \\ &= \hat{\dot{\rho}}_{n_s}^\alpha \left\{ \dot{\gamma}^\alpha, \rho_1^\alpha, \rho_2^\alpha, ..., \rho_{n_s}^\alpha, \theta \right\}. \end{aligned} \tag{18}$$

An Application of Crystal Plasticity to the Study of Intergranular Damage in an FCC Alloy

In this section, the typical single crystal framework described in the previous section will be applied to the study of intergranular cracking in a typical FCC Al alloy. The work to be described here is based on that by Pouillier et al. (2012) which studied the effects of plasticity on the mechanism of intergranular cracking assisted by hydrogen-induced embrittlement in an Al-5%Mg alloy. As one of the main strengthening mechanism arises from the presence of Mg in solid solution, it is a suitable example for the use of crystal plasticity to study intergranular cracking phenomena.

Aluminum alloys, strengthened by elements in solid solution, can be sensitive to intergranular stress corrosion cracking in some specific microstructural states. In such alloys, precipitation of the Al3Mg2 phase at grain boundaries strongly favors intergranular fracture. When hydrogen is absorbed from the environment into the material, it diffuses along grain boundaries, weakening the matrix-precipitate interfaces which may, under high local stress–strain conditions, lead to a true intergranular decohesion mechanism. The main objective of this work was thus to study the effects of plasticity on the mechanism of intergranular cracking assisted by hydrogen-induced embrittlement in an Al-5%Mg alloy.

Single Crystal Formulation for the FCC Material

The single crystal model is based on the original formulation proposed by Cheong and Busso (2004) for Cu and on the recent work by Pouillier et al. (2012). The generic form of the slip rate, $\dot{\gamma}^{\alpha}$, given in Eq. 13, is assumed to be dominated by the thermally activated glide of dislocations over obstacles (i.e., mainly forest dislocations as the heat treatment used in the alloy of interest did not lead to full precipitation of Mg within the grains). The slip rate is related to the resolved shear stress, τ^{α}, through the exponential function proposed by Busso and McClintock (1996) and Busso et al. (2000):

$$\dot{\gamma}^{\alpha} = \dot{\gamma}_o \exp\left[-\frac{F_o}{k\theta}\left\{1 - \left\langle \frac{|\tau^{\alpha}| - S_T^{\alpha}\mu/\mu_o}{\hat{\tau}} \right\rangle^{p}\right\}^{q}\right] \operatorname{sign}(\tau^{\alpha}), \tag{19}$$

which accounts for the absolute temperature (θ, K) and the stress dependence of the activation energy. In Eq. 19, F_0 represents the Helmholtz free energy of activation at 0 K, k the Boltzmann constant, $\dot{\gamma}_o^{\alpha}$ a reference slip rate, and $\hat{\tau}$ the maximum glide resistance at which dislocations can be mobilized without thermal activation. Furthermore, μ and μ_0 are the shear moduli at θ and 0 K, respectively. The exponents p and q describe the shape of the energy barrier vs. stress profile associated with interactions between dislocations and obstacles.

The main contributions to the overall slip resistance to plastic flow, S_T^{α}, are due to the friction stress induced by the Mg atoms in solid solution, S_{ss}^{α}, and to the athermal slip resistance, S_{dis}^{α}. As the slip resistance contributions are assumed to be additive (see Eq. 14), then

$$S_T^{\alpha} = S_{dis}^{\alpha} + S_{ss}^{\alpha}. \tag{20}$$

Note that the scaling parameters c_{dis} and c_{ss} in the general expression given by Eq. 14 are, for this particular application, taken to be equal to one. Since only a very small part of the Mg atoms precipitate during the heat treatment, the concentration of Mg in solid solution is set to be equal to the average concentration in the material. Thus, the friction stress due to Mg atoms in solid solution, S_{ss}^{α}, is calculated based on the atomic size and concentration of Mg in the alloy (see Pouillier et al. 2012 for details), as proposed by Saada (1968). The athermal slip resistance is expressed as

$$S_{dis}^{\alpha} = \lambda \mu b^{\alpha} \sqrt{\left(\sum_{\beta=1}^{N} h^{\alpha\beta} \rho_T^{\beta} \right)}, \tag{21}$$

where the overall dislocation density for a given slip system β, ρ_T^{β}, is obtained from a discretization of the dislocation structure into pure edge and pure screw types, of densities ρ_e^{β} and ρ_s^{β}, respectively. Thus,

$$\rho_T^{\beta} = \rho_e^{\beta} + \rho_s^{\beta}. \tag{22}$$

In Eq. 21, λ and $h^{\alpha\beta}$ were as previously defined in Eqs. 15 and 16. The evolutionary equations of the individual dislocation densities account for the competing dislocation storage-dynamic recovery processes in FCC metals. They can be expressed as (Cheong et al. 2004; Cheong and Busso 2006)

$$\dot{\rho}_e^{\alpha} = \frac{C_e}{b^{\alpha}} \left[K_e \sqrt{\sum_{\beta=1}^{N} \rho_T^{\beta}} - 2 d_e \rho_e^{\alpha} \right] |\dot{\gamma}^{\alpha}|, \tag{23}$$

and

$$\dot{\rho}_s^{\alpha} = \frac{C_s}{b^{\alpha}} \left[K_s \sqrt{\sum_{\beta=1}^{N} \rho_T^{\beta}} - \rho_s^{\alpha} \left(\pi d_s^2 K_s \sqrt{\sum_{\beta=1}^{N} \rho_T^{\beta}} + 2 d_s \right) \right] |\dot{\gamma}^{\alpha}|. \tag{24}$$

Here, the parameters C_e and C_s describe the relative contributions to the overall slip from edge and screw dislocations, while K_e and K_s are mobility constants associated with their respective mean free paths. Recovery processes are associated with the parameters d_e and d_s, which represent critical annihilation distances between dislocations of opposite Burgers vectors for both edge and screw types.

The calibration of the model's parameters was inspired by those reported for pure aluminum by Cheong and Busso (2006), except for the additional term which accounts for the Mg solid solution effect. A comparison between the uniaxial stress–strain tensile curve obtained from a smooth uniaxial specimen of the polycrystal and that predicted using a 100 grain aggregate and the single crystal model

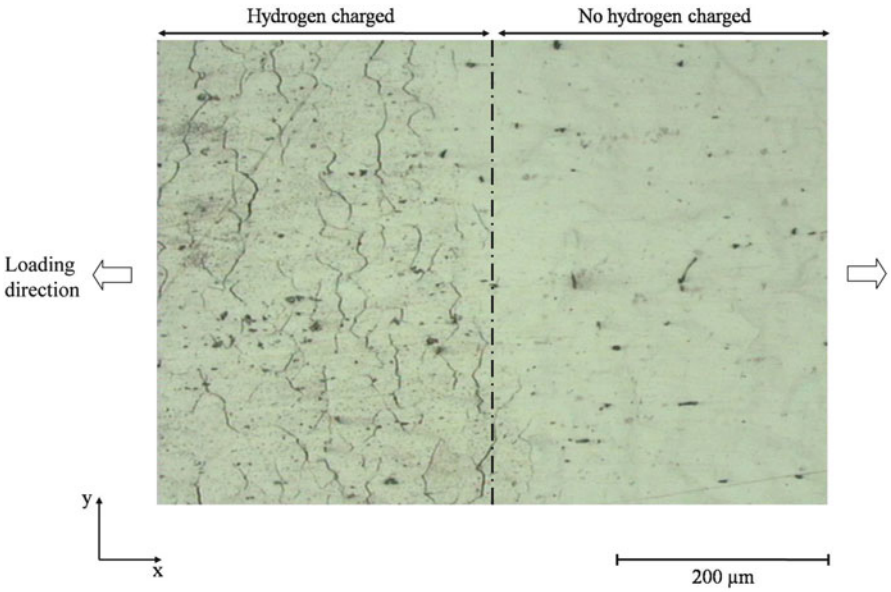


Fig. 1 Optical micrograph of intergranular fracture after 10 % axial strain along the x-axis (Pouillier et al. 2011)

was relied upon as a polycrystal validation. Details of the implicit numerical implementation of the above constitutive theory into the finite element method can be seen in Busso et al. (2000).

Intergranular Crack Observations, Strain Field Measurements, and Predicted Local Stress and Strain Fields

Here, tensile specimens charged with hydrogen were used to investigate quantitatively the effect of plastic deformation on the mechanism of intergranular crack initiation at the scale of the individual grains. An experimental procedure was set up to monitor the evolution of surface strain fields on in situ tested SEM notched specimens using digital image correlation techniques. In addition, measurements of the associated crystal orientation evolution at the micron scale were carried out using electron backscatter diffraction (EBSD). These measurements were then compared with finite element predictions of the local strain fields on the observed regions of the in situ specimen. The crystallographic grain orientations of the region of interest were discretized for the finite element analyses from EBSD maps.

Figure 1 shows the surface of one of the specimens tested in tension up to a 10 % elongation. It can be seen that intergranular cracking occurred on grain boundaries perpendicular to the tensile direction after a few percent of plastic strain only on the half of the specimen which had been previously charged with hydrogen.

An optimized digital image correlation technique was used to compute surface strain fields from the recorded images. The field of view for the different regions

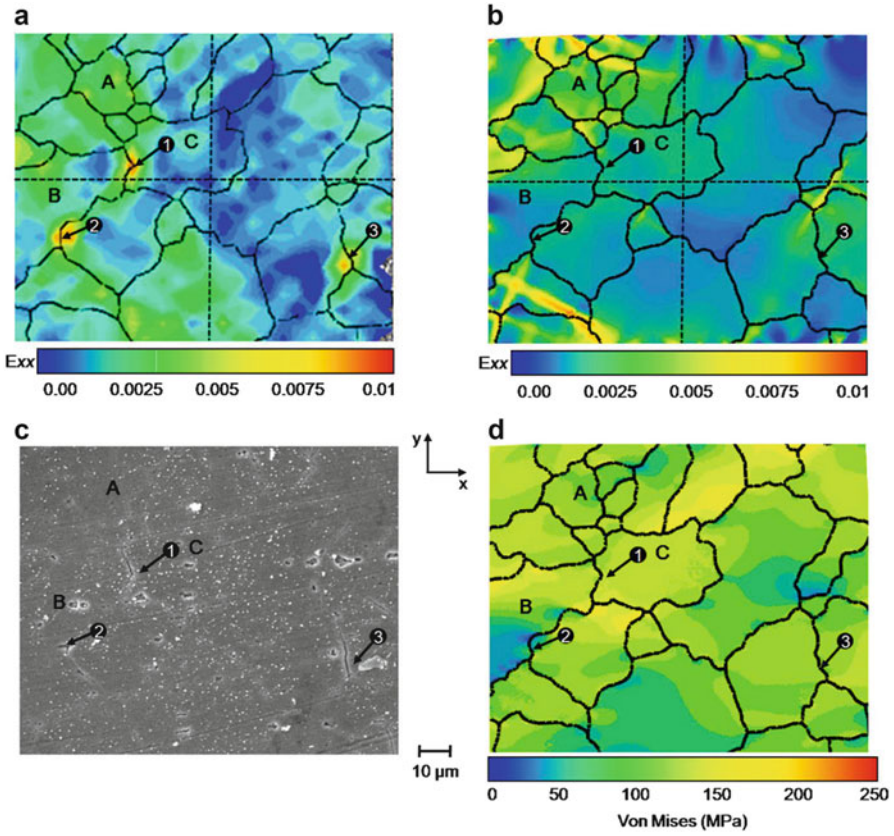


Fig. 2 (**a**) Measured and (**b**) predicted axial strain fields after 0.45 % applied macroscopic strain, (**c**) micrograph of the same region as (**a**) and (**b**) at the end of the test (4.5 % applied strain), and (**d**) predicted Mises stresses after 0.45 % applied strain. *Arrows* 1, 2, and 3 indicate cracked grain boundaries (loading direction is parallel to the x-axis; Pouillier et al. 2011)

analyzed was chosen so that it provided a range of fine-scale strain field measurements as well as the boundary conditions applied to the crystal aggregate region used in the modelling part of the study.

Figure 2a, b shows typically measured and predicted strain fields after 0.45 % applied macroscopic strain, respectively. Figure 2c is a SEM micrograph of Fig. 2a's region, and Fig. 2d gives the predicted Mises stresses at the time when the first grain boundary failure was observed. It should be noted that at the cracked locations 1, 2, and 3 indicated in Fig. 2, the high levels of axial strain measured are to a great extent different from the real values due to the distortion introduced by the opening of the cracks during the digital image correlation measurements.

In this study, it was found that grain boundary cracking occurs on boundaries normally oriented to the applied tensile stress when the average local axial strain is as low as 0.45 %. In addition, failed grain boundaries were found to be between

grains that undergo very limited plastic deformation despite being embedded in large localized deformation regions. Local grain boundary tractions were predicted analytically from an Eshelby-type model and compared to the local normal tractions obtained from the finite element simulations of the polycrystal aggregate specimen region using crystal plasticity. The analytical traction of 175 MPa was consistent with the numerical predictions of 170 $\pm$ 35 MPa obtained from the finite element model of the polycrystal.

From this work, it can be seen that crystal plasticity concepts can provide an accurate insight into the local stress and strain fields responsible for intergranular damage.

Nonlocal Single Crystal Approaches

The study of experimentally observed size effects in a wide range of mechanics and materials problems has received a great deal of attention recently. Most continuum approaches and formulations dealing with these problems are based on strain gradient concepts and are known as nonlocal theories since the material behavior at a given material point depends not only on the local state but also on the deformation of neighboring regions. Examples of such phenomena include particle size effects on composite behavior (e.g., Nan and Clarke 1996), precipitate size in two-phase single crystal materials (Busso et al. 2000), increase in measured microhardness with decreasing indentor size (e.g., Swadener et al. 2002), and decreasing film thickness (e.g., Huber and Tsakmakis 1999), among others. The dependence of mechanical properties on length scales can in most cases be linked to features of either the microstructure, boundary conditions, or type of loading, which give rise to localized strain gradients. In general, the local material flow stress is controlled by the actual gradients of strain when the dominant geometric or microstructural length scales force the deformation to develop within regions of less than approximately 5–10 μm wide in polycrystalline materials and of the order of 0.1–1.0 μm in single crystal materials. Thus, gradient-dependent behavior is expected to become important once the length scale associated with the local deformation gradients becomes sufficiently large when compared with the controlling microstructural feature (e.g., average grain size in polycrystal materials). In such cases, the conventional crystallographic framework discussed in the previous sections will be unable to predict properly the evolution of the local material flow stress.

The modelling of size effects observed in crystalline solids has been addressed by adding strain gradient variables into the constitutive framework, either in an explicit way in the flow rule (e.g., Aifantis 1984, 1987), or in the evolutionary equations of the internal slip system variables (e.g., Acharya and Beaudoin 2000; Busso et al. 2000; Bassani 2001), or by means of additional degrees of freedom associated with higher-order boundary and interface conditions (e.g., Shu 1998). Motivations for introducing strain gradients in continuum modelling stem from the multiscale analysis of micromechanics, as reviewed in Ghoniem et al. (2003).

The resulting strain gradient components are related to the dislocation density tensor introduced by Nye (1953). As it will be shown later in the text, the dislocation density tensor is computed from the rotational part of the gradient of plastic deformation so that the resulting partial differential equations to be solved are generally of higher order than those used in classical mechanics.

Nonlocal Models Based on Internal Strain Gradient Variables

The more physically intuitive continuum approaches to describe strain gradient effects are constitutive theories (e.g., Arsenlis and Parks 2001; Busso et al. 2000; Acharya and Bassani 2000; Bassani 2001; Cheong et al. 2004; Dunne et al. 2007) which rely on internal state variables to describe the evolution of the obstacle or dislocation network within the material and generally introduce the strain gradient effects directly in the evolutionary laws of the slip system internal variables without the need for higher-order stresses. This requires that the overall slip resistance arising from the dislocation network, S^{α}_{dis} (see Eq. 21), incorporates contributions from both statistically stored (SS) and geometrically necessary (GN) forest dislocations.

The general form for the functional dependency of the evolutionary laws slip system internal variables given in Eq. 18, extended to include the additional dependency on the GNDs and the gradient of the slip rates, $\nabla\dot{\gamma}^{\alpha}$, is

$$\begin{aligned}
\dot{\rho}^{\alpha}_{1} &= \hat{\dot{\rho}}^{\alpha}_{1}\left\{\dot{\gamma}^{\alpha}, \rho^{\alpha}_{1}, \ldots, \rho^{\alpha}_{n_s+n_G}, \theta\right\},\\
&\vdots\\
\dot{\rho}^{\alpha}_{n_s} &= \hat{\dot{\rho}}^{\alpha}_{n_s}\left\{\dot{\gamma}^{\alpha}, \rho^{\alpha}_{1}, \ldots, \rho^{\alpha}_{n_s+n_G}, \theta\right\},\\
\dot{\rho}^{\alpha}_{n_s+1} &= \hat{\dot{\rho}}^{\alpha}_{n_s+1}\left\{\dot{\gamma}^{\alpha}, \rho^{\alpha}_{n_s+1}, \ldots, \rho^{\alpha}_{n_s+n_G}, \nabla\dot{\gamma}^{\alpha}, \theta\right\},\\
&\vdots\\
\dot{\rho}^{\alpha}_{n_s+n_G} &= \hat{\dot{\rho}}^{\alpha}_{n_s+n_G}\left\{\dot{\gamma}^{\alpha}, \rho^{\alpha}_{n_s+1}, \ldots, \rho^{\alpha}_{n_s+n_G}, \nabla\dot{\gamma}^{\alpha}, \theta\right\},
\end{aligned} \tag{25}$$

where n_s and n_G denote the number of SSD and GND types, respectively.

Consider the particular case where $n_s = 2$ and $n_G = 3$ in Eq. 25. Then, the total dislocation density on an arbitrary slip system can be defined by

$$\rho^{\alpha}_{\text{T}} = \left(\rho^{\alpha}_{\text{e}} + \rho^{\alpha}_{\text{s}}\right) + \left(\rho^{\alpha}_{\text{Gs}} + \rho^{\alpha}_{\text{Get}} + \rho^{\alpha}_{\text{Gen}}\right), \tag{26}$$

where (ρ^{α}_{e}, ρ^{α}_{s}) are the SS densities introduced in Eq. 22 and ($\rho^{\alpha}_{\text{Gs}}$, $\rho^{\alpha}_{\text{Get}}$, $\rho^{\alpha}_{\text{Gen}}$) the GND densities. Here, the GNDs have, in addition, been discretized into pure edge and screw components based on a mathematically equivalent GND vector, ρ^{α}_{G}, projected into a local orthogonal reference system where $\rho^{\alpha}_{\text{Gs}}$ represents a set of screw GNDs parallel to the slip direction, $\mathbf{m}^{\text{a}}$, and $\rho^{\alpha}_{\text{Gen}}$ and $\rho^{\alpha}_{\text{Get}}$ edge GND components oriented parallel to the slip system normal, $\mathbf{n}^{\text{a}}$, and to $\mathbf{t}^{\text{a}} = \mathbf{m}^{\text{a}} \times \mathbf{n}^{\text{a}}$, respectively.

The evolution of the GNDs can be expressed in terms of a mathematically equivalent GND density vector, $\dot{\boldsymbol{\rho}}^{\alpha}_{G}$, defined so that its projection into the local

($\mathbf{m}^\alpha$, $\mathbf{n}^\alpha$, $\mathbf{t}^\alpha$) orthogonal reference system is as follows (Busso et al. 2000; Cheong et al. 2004):

$$\dot{\boldsymbol{\rho}}_{\mathrm{G}}^{\alpha} = \dot{\rho}_{\mathrm{Gs}}^{\alpha}\mathbf{m}^{\alpha} + \dot{\rho}_{\mathrm{Get}}^{\alpha}\mathbf{t}^{\alpha} + \dot{\rho}_{\mathrm{Gen}}^{\alpha}\mathbf{n}^{\alpha}. \tag{27}$$

Subsequently, the evolutionary law for each set of GNDs is determined from Nye's dislocation density tensor, $\boldsymbol{\Gamma}$ (Nye 1953), in terms of the spatial gradient of the slip rate:

$$\dot{\boldsymbol{\Gamma}} = \mathrm{curl}\,(\dot{\gamma}^{\alpha}\mathbf{n}^{\alpha}\mathbf{F}^{\mathrm{p}}) = \mathrm{b}^{\alpha}\left(\dot{\rho}_{\mathrm{Gs}}^{\alpha}\mathbf{m}^{\alpha} + \dot{\rho}_{\mathrm{Get}}^{\alpha}\mathbf{t}^{\alpha} + \dot{\rho}_{\mathrm{Gen}}^{\alpha}\mathbf{n}^{\alpha}\right). \tag{28}$$

Under small strains and rotations, Eq. 28 simplifies to

$$\dot{\rho}_{\mathrm{Gs}}^{\alpha} = \frac{1}{\mathrm{b}^{\alpha}}\nabla\dot{\gamma}^{\alpha}\cdot\ \mathbf{t}^{\alpha},\ \dot{\rho}_{\mathrm{Get}}^{\alpha} = \frac{1}{\mathrm{b}^{\alpha}}\nabla\dot{\gamma}^{\alpha}\cdot\ \mathbf{m}^{\alpha},\ \dot{\rho}_{\mathrm{Gen}}^{\alpha} = 0. \tag{29}$$

The slip resistance contributions from the SSDs and GNDs can then be determined from Eq. 21 using the definition of the overall dislocation density given by Eq. 26.

This class of theories has been shown capable of providing great physical insight into the effects of microstructure on the observed macroscopic phenomena, including rate-independent plastic deformation and viscoplasticity in both single crystal and polycrystalline materials (e.g., Arsenlis and Parks 2001; Busso et al. 2000; Acharya and Bassani 2000). One additional attractive aspect of these theories is that they are relatively easy to implement numerically and do not require higher-order stresses and additional boundary conditions or independent degrees of freedom. However, some of the limitations of these types of theories is that they are unable to describe problems which may require nonstandard boundary conditions, as discussed in Cheong et al. (2004), such as the boundary layer problem modeled by Shu et al. (2001), and that they may exhibit a mesh sensitivity in cases where there is a predominance of geometrically necessary dislocations relative to statistically stored dislocations (Cheong et al. 2004).

Nonlocal Models Based on the Mechanics of Generalized Continua

Approaches based on the so-called mechanics of *generalized continua* incorporate as a common feature extra-hardening effects associated with the dislocation density tensor. Generalized crystal plasticity models developed in the past 40 years can be classified into two main groups.

In the first one, strain gradient plasticity models involve either the rotational part of the plastic distortion (i.e., the plastic rotation), its full gradient, or just the gradient of its symmetric part (Steinmann 1996; Fleck and Hutchinson 1997; Gurtin 2002; Gurtin and Anand 2009).

The second group involves generalized continuum theories with additional degrees of freedom accounting for either the rotation or the full deformation of a triad of crystal directors, and the effect of their gradients on hardening, such as

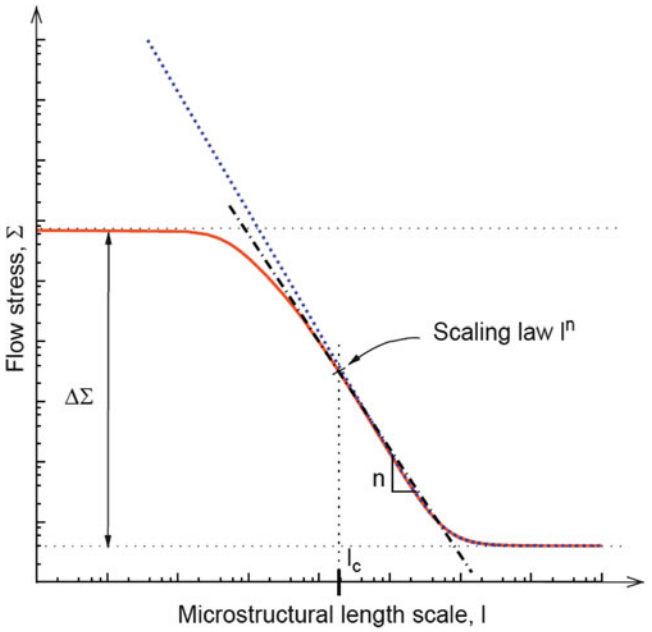


Fig. 3 Effect of the dominant microstructural length scale, l, on the material flow stress, Σ, predicted by different types of models such as those exhibiting two asymptotic regimes (*solid line*), and others which exhibited an unbounded flow stress for small length scales (*dotted line*). Also included is the scaling law in the transition domain (*dot-dashed line*) (Cordero et al. 2010)

Cosserat-type models (Forest et al. 2001; Clayton et al. 2006), and those based on the micromorphic theory (Eringen and Claus 1970; Bammann 2001; Cordero et al. 2010, 2012a, b).

Most of these theories have been shown to capture size effects, at least in a qualitative way. However, a clear demonstration that they can reproduce the scaling laws expected in precipitate hardening or grain size effect has not been fully provided yet. The additional hardening effects inherent in generalized continuum crystal plasticity models can be summarized by the main features identified in Fig. 3. Here, the effect of the dominant microstructural length scale, l, such as grain or precipitate size, on the material flow stress is shown schematically in a log–log diagram. The curve can be characterized by three main features: the stress range, $\Delta\Sigma$; the characteristic length, l_c; and the slope of the intermediate region, defined by a scaling law of the form, $\Sigma \propto l^n$ at $l = l_c$. Here, $\Delta\Sigma$ corresponds to the maximum increase in strength due to size effects relative to the size-independent level. Figure 3 shows that when the characteristic size of the microstructure decreases, the material strengthens. For large values of l, the asymptotic behavior corresponds to the size-independent response of conventional crystal plasticity models reviewed in the previous section. In contrast, for small values of l, a

bounded or unbounded asymptotic behavior can be obtained, depending on the type of model considered. Cosserat-type crystal plasticity models (e.g., Forest et al. 2001), for instance, predict an asymptotically saturated overstress $\Delta\Sigma$ as in Fig. 3. In the intermediate region, when l is closed to the characteristic length, l_c, the size-dependent response is characterized by the scaling law, $\Sigma \propto l^n$. The parameters $\Delta\Sigma$, l_c, and n can be derived explicitly for the different classes of generalized material models described above. However, an analytic description of the size-dependent behavior of materials is possible only for some specially simplified geometrical situations. Examples are the shearing of a single crystal layer under single (or double) slip for strain gradient plasticity models considered in Shu et al. (2001), Bittencourt et al. (2003), Hunter and Koslowski (2008), and Cordero et al. (2010, 2012a, b) and the single slip in a two-phase laminate microstructure by Forest and Sedlacek (2003). The plastic slip distributions were compared with those obtained from the reference continuous dislocation line tension model, Cosserat, and strain gradient plasticity models.

When crystal plasticity is considered under small strain assumptions, the gradient of the velocity field can be decomposed into the elastic and plastic distortion rates:

$$\dot{\mathbf{H}} = \dot{\mathbf{u}} \otimes \nabla = \dot{\mathbf{H}}^{\mathrm{e}} + \dot{\mathbf{H}}^{\mathrm{p}}, \tag{30}$$

where

$$\dot{\mathbf{H}}^{\mathrm{p}} = \sum_{\alpha} \dot{\gamma}^{\alpha} \mathbf{P}^{\alpha}, \tag{31}$$

with $\mathbf{u}$, the displacement field; α, the number of slip systems; $\dot{\gamma}^{\alpha}$, the slip rate for the slip system α; and $\mathbf{P}^{\alpha}$ as defined in Eq. 11. The elastic distortion tensor, $\mathbf{H}^{\mathrm{e}}$, which represents the stretch and rotation of the lattice, links the compatible total deformation, $\mathbf{H}$, with the incompatible plastic deformation, $\mathbf{H}^{\mathrm{p}}$, which describes the local lattice deformation due to the flow of dislocations. On account of Eq. 30 and since applying the curl operator to the compatible field represented by $\mathbf{H}$ is equal to zero, it follows that

$$\mathrm{curl}\ \dot{\mathbf{H}} = 0 = \mathrm{curl}\ \dot{\mathbf{H}}^{\mathrm{e}} + \mathrm{curl}\ \dot{\mathbf{H}}^{\mathrm{p}}. \tag{32}$$

The incompatibility of the plastic distortion is characterized by its curl part, also known as the dislocation density tensor or Nye's tensor, $\boldsymbol{\Gamma}$ (Nye 1953; Steinmann 1996; Acharya and Bassani 2000), defined as

$$\boldsymbol{\Gamma} = -\mathrm{curl}\ \mathbf{H}^{\mathrm{p}} = \mathrm{curl}\ \mathbf{H}^{\mathrm{e}}. \tag{33}$$

The tensors $\mathbf{H}$, $\mathbf{H}^{\mathrm{e}}$, and $\mathbf{H}^{\mathrm{p}}$ are generally nonsymmetric; thus, they can be decomposed into their symmetric and skew-symmetric parts:

$$\mathbf{H} = \mathbf{E} + \mathbf{W}, \ \mathbf{H}^e = \mathbf{E}^e + \mathbf{W}^e, \mathbf{H}^p = \mathbf{E}^p + \mathbf{W}^p, \tag{34}$$

Combining Eqs. 32 and 34 leaves

$$0 = \text{curl } \mathbf{E}^e + \text{curl } \mathbf{W}^e + \text{curl } \mathbf{H}^p. \tag{35}$$

Neglecting the curl part of the elastic strain, $\mathbf{E}^e$, leads to the following approximation to the dislocation density tensor derived by Nye:

$$\boldsymbol{\Gamma} = \text{curl } \mathbf{H}^e = \text{curl } \mathbf{E}^e + \text{curl } \mathbf{W}^e \approx \text{curl } \mathbf{W}^e. \tag{36}$$

Thus, Nye's formula sets a linear relationship between the dislocation density tensor and the lattice curvature defined by $\mathbf{W}^e$. The Cosserat crystal plasticity theory accounts for the effect of lattice curvature on the crystal hardening behavior by incorporating the three additional independent degrees of freedom associated with the components of the lattice rotation, $\mathbf{W}^e$. In contrast, theories such as those proposed by Gurtin (2002) and Svendsen (2002), for example, include the full curl of the plastic distortion, $\mathbf{H}^p$, as an independent internal variable of the constitutive model. This requires in general nine additional degrees of freedom associated with the generally nonsymmetric plastic distortion tensor, $\mathbf{H}^p$. This subclass of models is sometimes referred to as "curl $\mathbf{H}^p$" type (Cordero et al. 2010).

A consequence of neglecting the curl of the elastic strain tensor in Cosserat-type models is that Cosserat effects can arise even in the elastic regime as soon as a gradient of "elastic" rotation exists (i.e., curl $\mathbf{W}^e \neq 0$). This implies that as soon as the curl $\mathbf{E}^e \neq 0$, the curl $\mathbf{W}^e \neq 0$. In contrast, in the curl $\mathbf{H}^p$-type theories, strain gradient effects can only arise when plastic deformation has developed. As has been shown in Cordero et al. (2010), this can lead to discontinuities in the generalized tractions at the interface between elastic and plastic regions. For the curl $\mathbf{H}^p$-type models, it is necessary to identify numerically higher-order boundary conditions at the elastoplastic boundaries which poses difficulties in the numerical implementation of this type of formulations, as discussed in Cordero et al. (2010).

To overcome the limitations of both the Cosserat and curl $\mathbf{H}^p$-type theories, a new regularization method has recently been proposed by Cordero et al. (2010) (see also Cordero et al. 2012a, b). Their model, which they have called *microcurl*, falls into the class of generalized continua with additional degrees of freedom. Here, the effect of the dislocation density tensor is introduced into the classical crystal plasticity framework by means of the micromorphic theory of single crystals. It relies on the introduction of an additional plastic micro-deformation variable, χ^p, a second-rank generally nonsymmetric tensor. It is distinct from the plastic distortion tensor $\mathbf{H}^p$, which is still treated as an internal variable of the problem in the same way as in curl $\mathbf{H}^p$-type theories. For the general three-dimension case, the nine components of χ^p are introduced as independent degrees of freedom. The *microcurl* theory will be briefly summarized next.

Microcurl Model: Balance and Constitutive Equations

If it is assumed that only the curl part of the gradient of plastic micro-deformation plays a role in the power of internal forces, $p^{(i)}$, then

$$p^{(i)} = \sigma : \dot{\mathbf{H}}^{\mathrm{e}} + \mathrm{s} : \dot{\chi}^{\mathrm{p}} + \mathbf{M} \ : \ \mathrm{curl}\, \dot{\chi}^{\mathrm{p}}, \tag{37}$$

where $\mathbf{s}$ and $\mathbf{M}$ are the generally asymmetric micro-stress and double or hyper-stress tensors, respectively, work conjugates to the plastic micro-deformation and its curl. The curl operator is defined in a Cartesian basis as

$$(\mathrm{curl}\, \chi^{\mathrm{p}})_{\mathrm{ij}} = \varepsilon_{\mathrm{jkl}} \chi^{\mathrm{p}}_{\mathrm{ik},\mathrm{l}}. \tag{38}$$

Using the method of virtual power to derive the generalized balance of momentum equations and assuming no volume forces for simplicity, one finds the following balance equations:

$$\mathrm{div}\, \sigma = 0, \quad \mathrm{curl}\, \mathbf{M} + \mathbf{s} = 0. \tag{39}$$

The corresponding boundary conditions are

$$\mathbf{t} = \sigma \cdot \mathbf{n}_{\mathrm{e}}, \quad \mathbf{m}_{\mathrm{e}} = \mathbf{M} \cdot \breve{\boldsymbol{\varepsilon}} \cdot \mathbf{n}_{\mathrm{e}}, \tag{40}$$

where $\mathbf{t}$ and $\mathbf{m_e}$ are the simple and double tractions at the boundary and $\breve{\boldsymbol{\varepsilon}}$ is the third-order permutation tensor.

The free energy function is assumed to depend on the elastic strain tensor, $\mathbf{E}^e$, the curl of χ^{p}, and on a relative plastic strain, $\mathbf{e}^p$, defined as the difference between the plastic distortion and the plastic micro-variable, $\mathbf{e}^p = \mathbf{H}^{\mathrm{p}} - \chi^{\mathrm{p}}$. Then,

$$\psi = \hat{\psi}(\mathbf{E}^{\mathrm{e}}, \mathrm{curl}\, \chi^{\mathrm{p}}, \mathbf{e}^{\mathrm{p}}). \tag{41}$$

Furthermore, considering the following state laws,

$$\sigma = \rho \frac{\partial \psi}{\partial \mathbf{E}^{\mathrm{e}}}, \quad \mathbf{s} = -\rho \frac{\partial \psi}{\partial \mathbf{e}^{\mathrm{p}}}, \quad \mathbf{M} = \rho \frac{\partial \psi}{\partial \boldsymbol{\Gamma}_{\chi}}, \tag{42}$$

where $\boldsymbol{\Gamma}_{\chi} = \mathrm{curl}\, \chi^{\mathrm{p}}$ and assuming a quadratic function for the potential function ψ in Eq. 42, one obtains

$$\sigma = \boldsymbol{L}\, \mathbf{E}^{\mathrm{e}}, \quad \mathbf{s} = -\mathrm{H}_{\chi}\, \mathbf{e}^{\mathrm{p}}, \quad \mathbf{M} = \mathrm{A}\, \boldsymbol{\Gamma}_{\chi}. \tag{43}$$

Here, H_{χ} and A are the generalized moduli, which define an intrinsic length scale associated with the size effect exhibited by the solution of the boundary value problem,

$$l_w = \sqrt{\frac{A}{H_\chi}}. \tag{44}$$

The flow rule can be derived from a viscoplastic potential, $\Omega(\boldsymbol{\sigma} + \mathbf{s})$, expressed in terms of the effective stress, $(\boldsymbol{\sigma} + \mathbf{s})$, that intervenes in the dissipation rate equation. Then,

$$\dot{\mathbf{H}}^p = \frac{\partial \Omega}{\partial(\boldsymbol{\sigma} + \mathbf{s})}. \tag{45}$$

For a single crystal having N potentially active slip systems, the kinematics of plastic deformation is dictated by Eq. 31. It is worth mentioning that, in a similar way as in the Cosserat- and curl $\mathbf{H}^p$-type theories, a back stress component arises naturally from the formulation. Here, the back stress is $\mathbf{x} = -\mathbf{s} : (\mathbf{l} \otimes \mathbf{n})$.

Note also that the modulus H_χ in Eq. 43 introduces a coupling between the macro- and micro-variables. This can be interpreted as a penalty factor that constrains the relative plastic deformation, $\mathbf{e}^p$, to remain sufficiently small. Equivalently, a high value of H_χ forces the plastic micro-deformation to be as close as possible to the macroscopic plastic distortion tensor, $\mathbf{H}^p$. In the limit, the use of a Lagrange multiplier instead of the penalty factor, H_χ, is necessary to enforce the following internal constraint:

$$\chi^p \equiv \mathbf{H}^p. \tag{46}$$

When Eq. 46 is satisfied, then

$$\text{curl}\, \chi^p \equiv \text{curl}\, \mathbf{H}^p = \boldsymbol{\Gamma}_\chi. \tag{47}$$

Note also that when the internal constraint (Eq. 46) is enforced, the *microcurl* model reduces to the curl $\mathbf{H}^p$-type theories (e.g., Gurtin 2002). As a general case, the selection of H_χ should be made so that the micro-deformation χ^p does not depart too much from $\mathbf{H}^p$ and retains the physical meaning of the dislocation density tensor.

Application of the Microcurl Model to Study the Deformation Behavior of a Polycrystalline Aggregate

The microcurl model was applied to study the global and local responses of two-dimensional polycrystalline aggregates with grain sizes ranging from 1 to 200 microns. (For full details about this work, refer to Cordero et al. 2012b.) A typical result concerning the effect of grain size on the way plastic deformation in polycrystals evolves is shown in Fig. 4 for a 52-grain aggregate. These contour plots show the field of equivalent plastic deformation, $\widetilde{\varepsilon}^p$, defined as the time-integrated value of

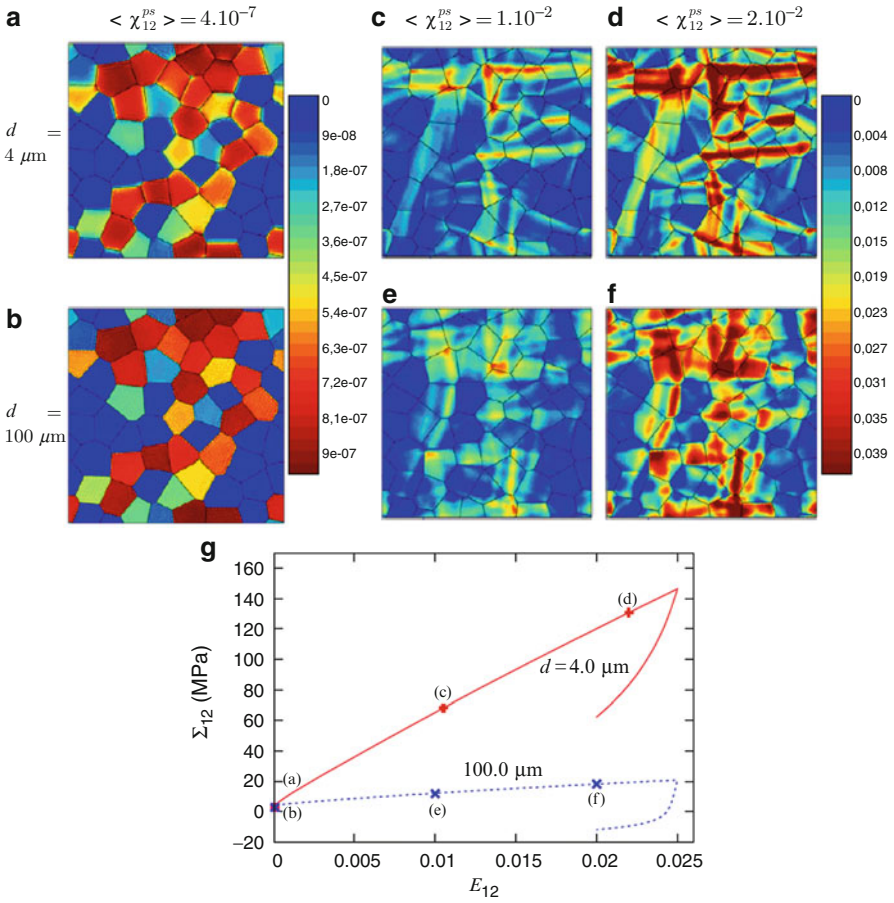


Fig. 4 (**a**–**f**) Contour plots of the accumulated plastic strain $\widetilde{\varepsilon}^p$ for two grain sizes, $d = 100$ and 4 μm, and different mean values of the plastic strain: $\chi_{12}^{ps} \approx 0.0$, 0.01, and 0.02, *o*btained with a 2D 55-grain aggregate under simple shear, (**g**) macroscopic stress–strain response of the corresponding aggregate, with the letters indicating the different loading steps corresponding to the (**a**–**f**) contour plots (Cordero et al. 2010)

$$\dot{\widetilde{\varepsilon}}^p = \sqrt{\frac{2}{3}\,\dot{\mathbf{H}}^{\mathrm{p}} : \dot{\mathbf{H}}^{\mathrm{p}}} \tag{48}$$

From Fig. 4a, b, it can be seen that, at the onset of plastic deformation, plasticity starts in the same grains and at the same locations in 100-μm grains as in 1-μm grains. This is due to the fact that the same critical resolved shear stress is adopted for both grain sizes, that is, the same initial dislocation densities are assumed in both cases. In contrast, at higher mean plastic strain levels, the strongly different values of the plastic micro-deformation gradients lead to significantly different plastic strain fields. Two main features are evidenced in Fig. 4c–f. Firstly, a

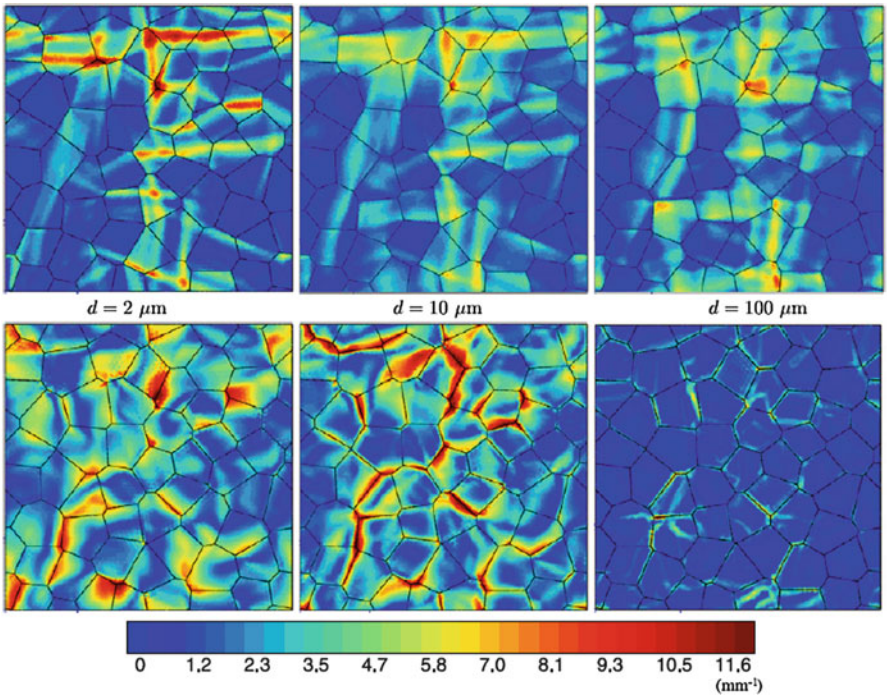


Fig. 5 Grain size effect on the accumulated plastic strain, $\widetilde{\varepsilon}^{p}$ (*top* figures), and on the norm of the dislocation density tensor, $\|\Gamma\|$ (*bottom* figures). These contour plots are obtained with the 2D 55-grain aggregate for the same mean value of $\chi_{12}^{ps} = 0.01$. The color scale for the plastic strain field of the top figures is the same as that of Fig. 4 on the right. The color scale at the bottom is that for the dislocation density tensor fields (Cordero et al. 2010)

tendency to strain localization in bands is observed for small grain sizes. The strain localization bands cross several grains, whereas plastic strain becomes more diffuse at larger grain sizes, something which had already been seen in the simulations presented in Cordero et al. (2012a). Secondly, a consequence of this localization is that some small grains are significantly less deformed than the larger ones. These features are also visible on the plastic deformation maps of Fig. 5 for the same aggregate but different grain sizes. This figure also shows the field of the norm of the dislocation density tensor:

$$\left\|\boldsymbol{\Gamma}_{\chi}\right\| := \sqrt{\boldsymbol{\Gamma}_{\chi} : \boldsymbol{\Gamma}_{\chi}}. \tag{49}$$

This scalar quantity indicates the presence of GNDs and has the physical dimension of lattice curvature (mm^{-1}). For large grains, GNDs are mainly located close to grain boundaries. At smaller grain sizes, the GND densities become significantly greater and spread over larger zones within the grains. Note also that pileups-like structures close to grain boundaries are clearly visible in the 10-μm grain aggregate. It should be noted that strain gradient plasticity models may be

prone to strain localization when plasticity is confined in small regions. The reason for such behavior is that acute slip bands that exhibit a strong gradient of plastic slip perpendicular to the slip plane are not associated with GND formation. In contrast, regions of high lattice curvature or kink bands lead to an energy increase. This explains why, at small scales, intense slip bands are preferred to strongly curved regions and pileups. This has been confirmed by the observation of the equivalent plastic deformation contour plots in Cordero et al. (2012b)'s Fig. 8 for three aggregates with different mean grain sizes. Here, the zones of intense plastic deformation were found to be systematically parallel to slip plane traces, thus indicating the formation of slip bands during deformation.

In summary, the microcurl model was found to naturally predict a size-dependent kinematic hardening behavior which is responsible for the observed strong size effects. Furthermore, the results showed that the flow stress attained at a given averaged plastic strain follows a power law scaling relation with the grain size, for grain sizes larger than a critical value. Likewise, the predicted plastic deformation fields were found to be strongly affected by grain size, with micron-size grain regions exhibiting the formation of intense slip bands crossing several grains. Finally, the dislocation density tensor, $\boldsymbol{\Gamma}_{\chi}$, was found to not only impact the overall polycrystal behavior but also control the way plastic deformation develops within the grains.

In Cordero et al. (2010), it was shown that the microcurl approach could also be successfully used to predict experimentally observed precipitate size effects in two-phase single crystal nickel-based superalloys. The results are shown in Fig. 6, where a comparison between experimental data, in the form of precipitate size vs. size effect strengthening from a two-phase superalloy material (γ phase matrix with an embedded 68 % γ' precipitates), the predictions of Busso et al. (2000), and that obtained using the *microcurl* model (Cordero et al. 2012b) are shown. It can be seen that the microcurl model is able to simulate a precipitate size effect naturally. Moreover, the identified characteristic length, $l_c = 200$ nm, is approximately the matrix channel width in Ni-based superalloys.

Concluding Remarks

The different constitutive modelling approaches which address a broad range of phenomena at either the single crystal or the polycrystalline levels have been critically discussed. This review has also highlighted the rich variety of physical, computational, and technological issues within the broad area of micromechanics which have been successfully addressed and has identified some theoretical and computational difficulties and challenges for future developments. In the future, crystallographic approaches for single crystal behavior which rely on internal slip system variables will continue to provide the most powerful framework to incorporate basic mechanistic understanding in continuum models. However, further development of 3D measurement and microstructure characterization techniques, such as X-ray tomography and high-resolution EBSD, will require new challenges

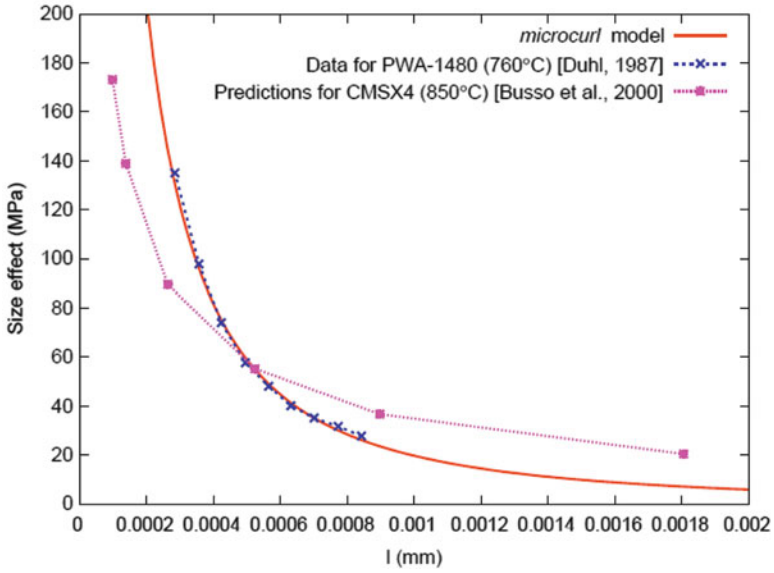


Fig. 6 Comparison between experimental data, in the form of precipitate size vs. size effect strengthening from a two-phase material (elasto-viscoplastic γ phase matrix with an embedded 68 % quasi-elastic γ' precipitates) from Duhl (1987), the prediction of Busso et al. (2000), and that obtained using the *microcurl* model by Cordero et al. (2012b)

to be overcome. Novel and more efficient computational techniques for processing and visualizing the enormous amount of data generated when studying real 3D polycrystalline materials will be required. The ever increasing need to address coupled multi-physics phenomena, such as the microstructural evolution driven by diffusion processes, is driving new multidisciplinary research and providing crystal plasticity with new and exciting challenges.

References

G. Abrivard, E.P. Busso, S. Forest, B. Appolaire, Phase field modelling of grain boundary motion driven by curvature and stored energy gradient. Part I – theory and numerical implementation. Philos. Mag. **92**(28–30), 3618–3642 (2012)

A. Acharya, J.L. Bassani, Lattice incompatibility and a gradient theory of crystal plasticity. J. Mech. Phys. Solids **48**, 1565–1595 (2000)

A. Acharya, A.J. Beaudoin, Grain size effects in viscoplastic polycrystals at moderate strains. J. Mech. Phys. Solids **48**, 2213–2230 (2000)

E.C. Aifantis, On the microstructural origin of certain inelastic models. J. Eng. Mater. Technol. **106**, 326–330 (1984)

E.C. Aifantis, The physics of plastic deformation. Int. J. Plast. **3**, 211–248 (1987)

L. Anand, M. Kothari, A computational procedure for rate independent crystal plasticity. J. Mech. Phys. Solids **44**, 525–558 (1996)

A. Arsenlis, D. Parks, Modeling the evolution of crystallographic dislocation density in crystal plasticity. J. Mech. Phys. Solids **50**, 1979–2009 (2001)
R.J. Asaro, J.R. Rice, Strain localization in ductile single crystals. J. Mech. Phys. Solids **25**, 309–338 (1977)
D.J. Bammann, A model of crystal plasticity containing a natural length scale. Mater. Sci. Eng. A **309–310**, 406–410 (2001)
J.L. Bassani, Incompatibility and a simple gradient theory of plasticity. J. Mech. Phys. Solids **49**, 1983–1996 (2001)
E. Bittencourt, A. Needleman, M. Gurtin, E. Van der Giessen, A comparison of nonlocal continuum and discrete dislocation plasticity predictions. J. Mech. Phys. Solids **51**(2), 281–310 (2003)
E.P. Busso, G. Cailletaud, On the selection of active slip systems in crystal plasticity. Int. J. Plast. **21**, 2212–2231 (2005)
E.P. Busso, F. McClintock, A dislocation mechanics-based crystallographic model of a B2-type intermetallic alloy. Int. J. Plast. **12**, 1–28 (1996)
E.P. Busso, F.T. Meissonnier, N.P. O'Dowd, Gradient-dependent deformation of two-phase single crystals. J. Mech. Phys. Solids **48**, 2333–2361 (2000)
K.S. Cheong, E.P. Busso, Discrete dislocation density modelling of single phase FCC polycrystal aggregates. Acta Mater. **52**, 5665–5675 (2004)
K.S. Cheong, E.P. Busso, Effects of lattice misorientations on strain heterogeneities in FCC polycrystals. J. Mech. Phys. Solids **54**(4), 671–689 (2006)
K. Cheong, E. Busso, A. Arsenlis, A study of microstructural length scale effects on the behavior of FCC polycrystals using strain gradient concepts. Int. J. Plast. **21**, 1797–1814 (2004)
J.D. Clayton, D.L. McDowell, D.J. Bammann, Modeling dislocations and disclinations with finite micropolar elastoplasticity. Int. J. Plast. **22**, 210–256 (2006)
N.M. Cordero, A. Gaubert, S. Forest, E.P. Busso, F. Gallerneau, S. Kruch, Size effects in generalised continuum crystal plasticity for two-phase laminates. J. Mech. Phys. Solids **58**, 1963–1994 (2010)
N.M. Cordero, S. Forest, E.P. Busso, S. Berbenni, M. Cherkaoui, Grain size effects on plastic strain and dislocation density tensor fields in metal polycrystals. Comput. Mater. Sci. **52**, 7–13 (2012a)
N.M. Cordero, S. Forest, E.P. Busso, Generalised continuum modelling of grain size effects in polycrystals. Comptes Rendus Mécanique **340**, 261–264 (2012b)
M.A. Crisfield, *Non-linear Finite Element Analysis of Solids and Structures, vols. 1 & 2*, 4th edn. (Wiley, New York, 1997)
D.N. Duhl, Directionally solidified superalloys, in Superalloys II – High Temperature Materials for Aerospace and Industrial Power, ed. by C.T. Sims, N.S. Stoloff, W.C. Hagel (Wiley, Toronto, 1987), pp. 189–214
F.P.E. Dunne, D. Rugg, A. Walker, Length scale-dependent, elastically anisotropic, physically-based hcp crystal plasticity: Application to cold-dwell fatigue in Ti alloys. Int. J. Plast. **23**, 1061–1083 (2007)
F.P.E. Dunne, R. Kiwanuka, A.J. Wilkinson, Crystal plasticity analysis of micro-deformation, lattice rotation and geometrically necessary dislocation density. Proc. R. Soc. A-Math. Phys. Eng. Sci. **468**, 2509–2531 (2012)
A.C. Eringen, W.D. Claus, A micromorphic approach to dislocation theory and its relation to several existing theories, in *Fundamental Aspects of Dislocation Theory*, ed. by J.A. Simmons, R. de Wit, R. Bullough. National bureau of standards (US) special publication 317, II, 1970, pp. 1023–1062
N.A. Fleck, J.W. Hutchinson, Strain gradient plasticity. Adv. Appl. Mech. **33**, 295–361 (1997)
S. Forest, R. Sedlacek, Plastic slip distribution in two-phase laminate microstructures: dislocation-based vs. generalized-continuum approaches. Philos. Mag. A **83**, 245–276 (2003)
S. Forest, F. Pradel, K. Sab, Asymptotic analysis of heterogeneous Cosserat media. Int. J. Solids Struct. **38**, 4585–4608 (2001)

N.M. Ghoniem, E.P. Busso, H. Huang, N. Kioussis, Multiscale modelling of nanomechanics and micromechanics: an overview. Philos. Mag. **83**, 3475–3528 (2003)
M.E. Gurtin, A gradient theory of single-crystal viscoplasticity that accounts for geometrically necessary dislocations. J. Mech. Phys. Solids **50**, 5–32 (2002)
M.E. Gurtin, L. Anand, Thermodynamics applied to gradient theories involving the accumulated plastic strain: the theories of Aifantis and Fleck & Hutchinson and their generalization. J. Mech. Phys. Solids **57**, 405–421 (2009)
T.M. Hatem, M.A. Zikry, Dislocation density crystalline plasticity modeling of lath martensitic microstructures in steel alloys. Philos. Mag. **89**(33), 3087–3109 (2009)
R. Hill, *The Mathematical Theory of Plasticity*, 4th edn. (Clarendon, Oxford, UK, 1950)
N. Huber, C. Tsakmakis, Determination of constitutive properties from spherical indentation data using neural networks. Part II: plasticity with nonlinear isotropic and kinematic hardening. J. Mech. Phys. Solids **47**, 1589–1607 (1999)
A. Hunter, M. Koslowski, Direct calculations of material parameters for gradient plasticity. J. Mech. Phys. Solids **56**(11), 3181–3190 (2008)
S. Kalidindi, C. Bronkhorst, L. Anand, Crystallographic texture theory in bulk deformation processing of fcc metals. J. Mech. Phys. Solids **40**, 537 (1992)
J.W. Kysar, Y. Saito, M.S. Oztop, D. Lee, W.T. Huh, Experimental lower bounds on geometrically necessary dislocation density. Int. J. Plast. **26**(8), 1097–1123 (2010)
F. Meissonnier, E.P. Busso, N.P. O'Dowd, Finite element implementation of a generalised non-local rate-dependent crystallographic formulation for finite strains. Int. J. Plast. **17**(4), 601–640 (2001)
C.-W. Nan, D. Clarke, The influence of particle size and particle fracture on the elastic–plastic deformation of metal matrix composites. Acta Mater. **44**, 3801–3811 (1996)
J.F. Nye, Some geometrical relations in dislocated crystals. Acta Metall. **1**, 153–162 (1953)
B. Peeters, M. Seefeldt, C. Teodosiu, S.R. Kalidindi, P. VanHoutte, E. Aernoudt, Work-hardening/softening behaviour of B.C.C. polycrystals during changing strain paths: I. an integrated model based on substructure and texture evolution, and its prediction of the stress–strain behaviour of an IF steel during two-stage strain paths. Acta Mater. **49**, 1607–1619 (2001)
E. Pouillier, A.F. Gourgues, D. Tanguy, E.P. Busso, A study of intergranular fracture in an aluminium alloy due to hydrogen embrittlement. Int. J. Plast. **34**, 139–153 (2012)
G. Saada, Limite élastique et durcissement dessolutions solides. Pont à Mousson **16**, 255–269 (1968)
F. Schubert, G. Fleury, T. Steinhaus, Modelling of the mechanical behaviour of the SC Alloy CMSX-4 during thermomechanical loading. Model. Simul. Sci. Eng. **8**, 947–957 (2000)
J.Y. Shu, Scale-dependent deformation of porous single crystals. Int. J. Plast. **14**, 1085–1107 (1998)
J.Y. Shu, N.A. Fleck, E. Van der Giessen, and A. Needleman, Boundary layers in constrained plasticflow: com parison of non-local and discrete dislocation plasticity. J. Mech. Phys. Solids 49, 1361–1395 (2001)
P. Steinmann, Views on multiplicative elastoplasticity and the continuum theory of dislocations. Int. J. Eng. Sci. **34**, 1717–1735 (1996)
B. Svendsen, Continuum thermodynamic models for crystal plasticity including the effects of geometrically-necessary dislocations. J. Mech. Phys. Solids **50**, 1297–1329 (2002)
J. Swadener, A. Misra, R. Hoagland, M. Nastasi, A mechanistic description of combined hardening and size effects. Scripta Met. **47**, 343–348 (2002)
M.A. Zikry, M. Kao, Inelastic microstructural failure mechanisms in crystalline materials with high angle grain boundaries. J. Mech. Phys. Solids V **44**(11), 1765–98 (1996)

Micromechanics for Heterogeneous Material Property Estimation

13

Muneo Hori

Contents

M. Hori (✉)
Earthquake Research Institute, The University of Tokyo, Bunkyo, Tokyo, Japan
e-mail: hori@eri.u-tokyo.ac.jp

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_42

Abstract

There are two well-established theories in micromechanics for analytical estimation of overall property of a heterogeneous material, instead of experimental estimation; a heterogeneous material includes a partially damaged or plastically deformed material. These two theories, namely, the average field theory and the homogenization theory, are explained in this article. The average field theory is based on physical treatment of the heterogeneous material in the sense that it mimics a material sample test, and it derives a closed-form expression of the overall property in terms of the average strain and stress. The homogenization theory is based on mathematical treatment in the sense that it applies the singular perturbation expansion to the governing equations and obtains numerical solution for the overall property. In this article, the following three advanced topics are also explained: (1) strain energy consideration to obtain the consistent overall property, (2) the Hashin–Shtrikman variational principle to obtain bounds for the overall property, and (3) the extension to the overall property estimation at dynamic state from that at quasi-static state.

Introduction

Overall Property of Heterogeneous Material

It is intuitively clear that there is a certain overall property for a heterogeneous material. The overall property is understood as the property of a material sample the size of which is sufficiently larger than the size of the heterogeneity that is included in the heterogeneous material. When a structure which consists of a heterogeneous material is analyzed, it is a standard practice to model the structure as being made of a fictitious but uniform material that has the overall property of the heterogeneous material. The presence of heterogeneities in the original material is ignored in such analysis of the structure.

The overall property is usually considered for a heterogeneous material which is in elastic regime. The overall property could be considered even when the heterogeneous material reaches a plastic elastoplastic regime. In particular, when a uniform material is damaged and numerous cracks of small sizes are initiated, the overall property of the material at this damaged state could be considered by regarding these cracks as material heterogeneities.

Average Field Theory and Homogenization Theory

For a given heterogeneous material, its overall property is usually measured by carrying out a material sample test. However, it is sometimes necessary to analytically (or numerically) estimate the overall property, in particular, for an expensive

material such as composites or metal alloy. In this case, micromechanics plays a key role in estimating the overall property based on the microstructure of the material. It provides two fundamental theories of analytically estimating the overall properties, namely, the *average field theory* (or the *mean field theory*) and the *homogenization theory*. The basic characteristics of these two theories are summarized as follows:

Average Field Theory: The average field theory is based on the fact that what is measured in a material sample test is the volume average of field variables in the sample. It seeks to compute the volume average of the field variables, considering a microstructure of a target heterogeneous material, and to estimate the overall property as the relation between the volume average of the field variables.

Homogenization Theory: The homogenization theory solves a governing equation for displacement in a heterogeneous material by applying a singular perturbation expansion (or often called a multi-scale or two-scale analysis). Hence, this theory is purely mathematical. The overall property naturally emerges as the consequences of numerically solving the expanded term of the displacement.

Nemat-Nasser and Hori (1993) presents a concise list of references related to the average field theory and the homogenization theory; see also Hill (1963), Mura (1987) for the average field theory and Sanchez-Palencia (1981), Bakhvalov and Panasenko (1984), Francfort and Murat (1986) for the homogenization theory. For relatively recent works, recommended are Hornung (1996), Kevorkina and Cole (1996), Ammari et al. (2006), Gao and Ma (2012), Le Quang et al. (2008), Liu (2008), Wang and Xu (2005), Wang and Gao (2011), Zheng and Du (2001), Zou et al. (2010) and Terada et al. (1996).

The average field theory and the homogenization theory deal with the overall property in an utterly different manner. For instance, modeling the microstructure of a heterogeneous material is different. The average field theory uses a simple model of an isolated inclusion which is embedded in an infinitely extended body, whereas the homogenization theory usually uses a periodic microstructure. It gives an impression that the two theories are basically different. However, it is possible to establish a common platform on which both the theories are explained in a unified manner.

This article is primarily concerned with explaining the average field theory and the homogenization theory in a unified manner, so that the applicability of these two theories to damaged materials can be seen. As advanced topics, brief explanation is given on the average strain energy consideration and the Hashin–Shtrikman variational principle Hashin and Shtrikman (1962); these two are interesting subjects of the average field theory. The two theories are easily extended to estimate the overall property at dynamic state, and the analytic estimation of dynamic overall property is explained in the end of this article.

Field Equations

Symbolic and index notations are used in this article; for instance, stress tensor is denoted by either $\boldsymbol{\sigma}$ or σ_{ij} (the index corresponds to the Cartesian coordinate system, x_i for $i = 1, 2, 3$). In the symbolic notation, $\cdot$ and: stand for the first- and second-order contractions and $\otimes$ for the tensor product. In the index notation, summation convention is employed. For simplicity, assumed are linear elasticity, infinitesimally small deformation, and quasi-static state with the absence of body forces. This setting is easily extended to nonlinear elastoplasticity or finite deformation state if incremental behavior is considered. The variable elasticity tensor is denoted by $\boldsymbol{c}$, and displacement, strain, and stress fields are denoted by $\boldsymbol{u}$, $\boldsymbol{\varepsilon}$, and $\boldsymbol{\sigma}$, respectively. These fields satisfy

$$\boldsymbol{\varepsilon}(\boldsymbol{x}) = \mathrm{sym}\{\nabla \boldsymbol{u}(\boldsymbol{x})\}, \tag{1}$$

$$\nabla \cdot \boldsymbol{\sigma}(\boldsymbol{x}) = \boldsymbol{0}, \tag{2}$$

$$\boldsymbol{\sigma}(\boldsymbol{x}) = \mathbf{c}(\boldsymbol{x}) : \boldsymbol{\varepsilon}(\boldsymbol{x}), \tag{3}$$

where sym stands for the symmetric part ($\mathrm{sym}\{()_{ij}\} = ()_{ij} + ()_{ji})/2$), and ∇ is the differential operator ($(\nabla \boldsymbol{u})_{ij} = \partial \mathrm{u}_j / \partial \mathrm{x}_i$). Note that $\boldsymbol{x}$ stands for a point. This set of the three equations lead to

$$\nabla \cdot (\boldsymbol{c}(\boldsymbol{x}) : \nabla \boldsymbol{u}(\boldsymbol{x})) = \boldsymbol{0}. \tag{4}$$

This is the governing equation for $\boldsymbol{u}$.

Average Field Theory

Averaging Scheme

The average field theory starts by introducing a *representative volume element* (RVE), denoted by V, as a body which models the microstructure of a given heterogeneous material; see Fig. 1. While various definitions can be made in this chapter, the RVE is regarded as a model of a material test sample which is used to actually estimate the overall property.

In a material sample test, the overall property is estimated by assuming the uniform distribution of strain and stress in the sample, and the strain and stress are measured from the sample's surface displacement and traction. For instance, when a cubic sample is subjected to the uniform loading of T and the resulting displacement is U, the uniaxial strain and stress are computed as

$$\varepsilon = \frac{U}{A}, \text{and } \sigma = \frac{T}{A^2},$$

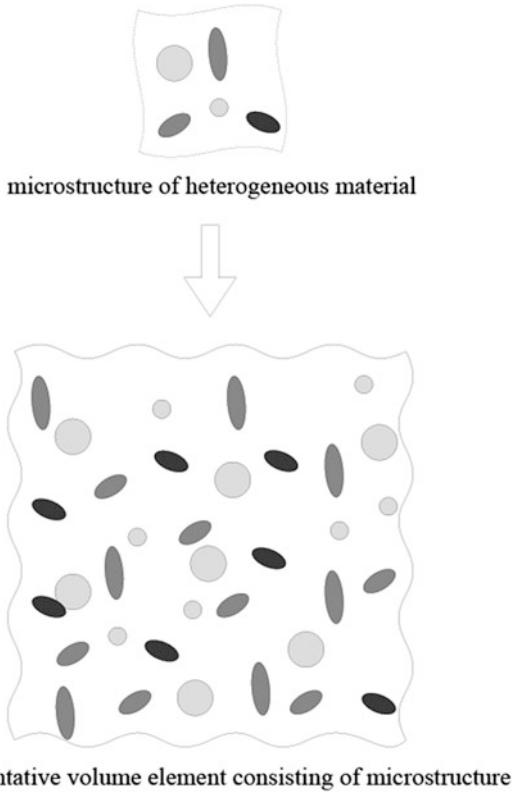


Fig. 1 Representative volume element for estimating overall property of heterogeneous material

where A is the edge length of the cubic sample. It is easily shown that the strain and stress measured in this way are actually the volume average of strain and stress within the sample, which, of course, are not uniform. Indeed, the volume average of strain and stress, denoted by $\langle \boldsymbol{\varepsilon} \rangle_V$ and $\langle \boldsymbol{\sigma} \rangle_V$, respectively, is given as

$$\langle \boldsymbol{\varepsilon} \rangle_V = \frac{1}{V} \int_{\partial V} \mathrm{sym}\{\boldsymbol{\nu} \otimes \boldsymbol{u}\} \mathrm{d}s, \tag{5}$$

$$\langle \boldsymbol{\sigma} \rangle_V = \frac{1}{V} \int_{\partial V} \{\boldsymbol{t} \otimes \boldsymbol{x}\} \mathrm{d}s, \tag{6}$$

where $\boldsymbol{\nu}$ is the outer unit normal on the boundary ∂V and $\boldsymbol{u}$ and $\boldsymbol{t}$ are the surface displacement and traction, respectively. The sample strain and stress, $\varepsilon = \frac{U}{A}$ and $\sigma = \frac{T}{A^2}$, are easily derived from these two equations. In this article, evaluation of the volume average in terms of the surface integration is called the averaging scheme; the averaging schemes presented in this article are readily proved by applying the integral by part and the Gauss theorem to the computation of the volume average of field variables. The averaging scheme is the consequences of Eqs. 1 and 2, and hence

it holds for any arbitrary material with any constitutive properties. Note that the symmetry of $\boldsymbol{\sigma}$ ($\sigma_{ij} = \sigma_{ji}$) is related to the moment equilibrium, and the averaging scheme leads to the symmetry of $\int \{\boldsymbol{t} \otimes \boldsymbol{x}\} \mathrm{d}s$ ($\int \{\boldsymbol{t} \otimes \boldsymbol{x}\} \mathrm{d}s = \int \{\boldsymbol{x} \otimes \boldsymbol{t}\} \mathrm{d}s$).

Average Field

In view of Eqs. 5 and 6, the average field theory introduces average fields, which are defined as the weighted average of the corresponding field variables. The weight function, denoted by φ_V, satisfies $\int \varphi_V \, \mathrm{d}v = 1$ and takes on a constant value of $1/V$ within V, except for a thin layer near ∂V where it decays smoothly from $1/V$ to 0 and vanishes on ∂V. Now, a body B which is larger than V is considered. The same symbols, $\{\boldsymbol{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}\}$, are used for the field variables of B, and the average fields are expressed in terms of φ_V and $\{\boldsymbol{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}\}$ as

$$\begin{Bmatrix} \boldsymbol{U} \\ \mathbf{E} \\ \boldsymbol{\Sigma} \end{Bmatrix} (\mathbf{X}) = \int_B \varphi_V(\boldsymbol{X} - \boldsymbol{x}) \begin{Bmatrix} \boldsymbol{u} \\ \boldsymbol{\varepsilon} \\ \boldsymbol{\sigma} \end{Bmatrix} (\boldsymbol{x}) \mathrm{d}v. \tag{7}$$

Here, $\boldsymbol{X}$ is a point in B. Due to the nature of φ_V, $\{\boldsymbol{U}, \mathbf{E}, \boldsymbol{\Sigma}\}$ is smoother than the original $\{\boldsymbol{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}\}$. Furthermore, it is seen that taking weighted average of φ_V is commutable with operating ∇, and hence the average fields satisfy

$$\mathbf{E}(\boldsymbol{X}) = \mathrm{sym}\{\nabla \boldsymbol{U}(\boldsymbol{X})\}, \tag{8}$$

$$\nabla \cdot \boldsymbol{\Sigma}(\boldsymbol{X}) = \mathbf{0}, \tag{9}$$

These two equations are the consequence of Eqs. 1 and 2.

As is seen, Eqs. 8 and 9 serve as field equations for $\{\boldsymbol{U}, \mathbf{E}, \boldsymbol{\Sigma}\}$, just as Eqs. 1 and 2 do for $\{\boldsymbol{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}\}$. If there is another equation which corresponds to Eq. 3, a governing equation for $\boldsymbol{U}$ can be derived. For instance, if this equation is given as

$$\boldsymbol{\Sigma}(\boldsymbol{X}) = \overline{\boldsymbol{C}} : \mathbf{E}(\boldsymbol{X}),$$

then it follows that $\nabla \cdot \left(\overline{\boldsymbol{C}} : \nabla \boldsymbol{U}\right) = \mathbf{0}$ is derived as the governing equation for $\boldsymbol{U}$. Note that $\overline{\boldsymbol{C}}$ is assumed to be uniform in B. Or it can be assumed that B consists of the fictitious but homogeneous material of $\overline{\boldsymbol{C}}$. If suitable boundary conditions are posed on the boundary ∂B, the averaged displacement $\boldsymbol{U}$ can be obtained by solving the resulting boundary value problem. As expected, this $\overline{\boldsymbol{C}}$ is the overall elasticity of the heterogeneous material. By definition, $\mathbf{E}$ and $\boldsymbol{\Sigma}$ correspond to $\langle \boldsymbol{\varepsilon} \rangle_V$ and $\langle \boldsymbol{\sigma} \rangle_V$, i.e., the volume average $\boldsymbol{\varepsilon}$ and $\boldsymbol{\sigma}$ taken over the RVE, V, of the target heterogeneous modulus. The average field theory, therefore, seeks to estimate $\overline{\boldsymbol{C}}$, which is now given as

$$\langle \boldsymbol{\sigma} \rangle_V = \overline{\boldsymbol{C}} : \langle \boldsymbol{\varepsilon} \rangle_V. \tag{10}$$

Note that while $\overline{\boldsymbol{C}}$ is defined in terms of $\langle\boldsymbol{\varepsilon}\rangle_V$ and $\langle\boldsymbol{\sigma}\rangle_V$, it does not guarantee that the presence of the unique $\overline{\boldsymbol{C}}$ that precisely relates $\mathbf{E}$ to $\boldsymbol{\Sigma}$ as $\boldsymbol{\Sigma} = \overline{\boldsymbol{C}} : \mathbf{E}$ for any point in B. However, it is naively understood that $\boldsymbol{\Sigma} = \overline{\boldsymbol{C}} : \mathbf{E}$ approximately holds if $\overline{\boldsymbol{C}}$ is computed for a sufficiently large V, and such V is used to define the average strain and stress fields, $\mathbf{E}$ and $\boldsymbol{\Sigma}$, of B. As will be explained later, it is indeed true that $\boldsymbol{\Sigma} = \overline{\boldsymbol{C}} : \mathbf{E}$ approximately holds since $\overline{\boldsymbol{C}}$ of V changes depending on the boundary conditions.

Explicit Expression of Overall Elasticity in Terms of Strain Concentration Tensor

The average field theory provides several schemes to estimate $\overline{\boldsymbol{C}}$ defined by Eq. 10. The simplest example of a two-phase composite (consisting of a matrix phase and an inclusion phase) is considered in order to explain these schemes. A common target of these schemes is the estimate of strain concentration tensor, denote by $\boldsymbol{A}$, which is defined as

$$\langle\boldsymbol{\varepsilon}\rangle_I = \boldsymbol{A} : \langle\boldsymbol{\varepsilon}\rangle_V. \tag{11}$$

Here, $\langle\rangle_I$ is the volume average of the inclusion phase of the composite. If the volume fraction of the inclusion phase is f, it holds $\langle\rangle_V = f\langle\rangle_I + (1-f)\langle\rangle_M$, where $\langle\rangle_M$ is the volume average of the matrix phase. After simple manipulations, $\langle\boldsymbol{\sigma}\rangle_V$ is evaluated as

$$\langle\boldsymbol{\sigma}\rangle_V = f\langle\boldsymbol{\sigma}\rangle_I + (1-f)\langle\boldsymbol{\sigma}\rangle_M = f\boldsymbol{C}^I : \boldsymbol{A} : \langle\boldsymbol{\varepsilon}\rangle_V + \boldsymbol{C}^M : (\boldsymbol{I} - f\boldsymbol{A}) : \langle\boldsymbol{\varepsilon}\rangle_V,$$

where $\boldsymbol{C}^M$ and $\boldsymbol{C}^I$ are the elasticity of the matrix and inclusion phases, and $\boldsymbol{I}$ is the identity tensor. In terms of $\boldsymbol{A}$, therefore, $\overline{\boldsymbol{C}}$ is expressed as

$$\overline{\boldsymbol{C}} = \boldsymbol{C}^M + f\left(\boldsymbol{C}^I - \boldsymbol{C}^M\right) : \boldsymbol{A}. \tag{12}$$

Note that no assumption is made in deriving this equation, except for the presence of the strain concentration tensor $\boldsymbol{A}$ of Eq. 11.

Use of Eshelby's Tensor for Evaluation of Strain Concentration Tensor

A model of an infinitely extended body which includes one inclusion is usually used, since a closed-form analytic solution, called Eshelby's solution, is available for this model when the inclusion is ellipsoidal form. There are various schemes which take advantage of Eshelby's solution (Eshelby 1957) to $\boldsymbol{A}$ and to estimate $\overline{\boldsymbol{C}}$ using Eq. 12. Representative schemes are the dilute

distribution assumption, the self-consistent method, and the differential scheme; see Nemat-Nasser and Hori (1993).

The original problem to which Eshelby's solution is found is an infinitely extended body problem; the body is homogeneous and linearly elastic and includes a certain uniform strain distributed in an ellipsoidal domain (Eshelby 1957). The strain is due to phase transition or thermal deformation and is usually called *eigen-strain*. It is shown that this uniform eigen-strain produces uniform strain in the ellipsoidal domain, although the strain due to the eigen-strain smoothly decays to zero outside of the ellipsoidal domain. Moreover, the strain in the ellipsoidal domain is analytically computed for the given elastic tensor of the infinite body and the configuration of the ellipsoidal domain. This is Eshelby's solution. The strain in the ellipsoidal domain is expressed as the second-order contraction of a certain fourth-order tensor and the eigen-tensor, and this fourth-order tensor is called Eshelby's tensor.

Since a closed-form expression is available for Eshelby's tensor, it is easy to analytically compute the strain concentration tensor, $\boldsymbol{A}$, in terms of Eshelby's solution. This is the reason that numerous researches have been made to evaluate the overall elasticity tensor in a closed form, using Eshelby's tensor; see Kachanov et al. (1994), Markenscoff (1998), Nozaki and Taya (2001), Kawashita and Nozaki (2001), Onaka et al. (2002), and Ru (2003). It is remarkable to note that eigen-strain which is uniformly generated in an ellipsoidal domain generates a uniform strain in the domain. This is because it implies that when an ellipsoidal inclusion of a different material is embedded in an infinitely extended and homogeneous body and the body is subjected to far-field loading, strain and stress of the inclusion become uniform; see Tanaka and Mori (1972) and Hori and Nemat-Nasser (1993).

Homogenization Theory

Singular Perturbation Expansion

The homogenization theory focuses the governing equation for $\boldsymbol{u}$, Eq. 4, by considering the nature of $\boldsymbol{c}$ in it. That is, $\boldsymbol{c}$ changes spatially in the length scale of material heterogeneity. To this end, the homogenization theory introduces two length scales, the one for the material heterogeneity and the other for a target structure, denoted by l and L, respectively. For simplicity, l and L are called the micro- and macro-length scales. The ratio of l and L is denoted by

$$\epsilon = \frac{l}{L}. \tag{13}$$

An insight is given to the length scale, if a finite element method analysis is considered. The dimension of a target body is in the order of L, and the dimension of an element is in the order of l. When the number of elements is $(10^3)^3$, the ratio is

$\epsilon \sim 10^{-3}$. In terms of this small ϵ, it is natural to define a slowly changing spatial coordinates, $\boldsymbol{X}$, by

$$\boldsymbol{X} = \epsilon \boldsymbol{x}. \tag{14}$$

By definition, $\boldsymbol{X}$ is the coordinate in the macro-length scale, if $\boldsymbol{x}$ is regarded as the coordinate in the micro-length scale.

The homogenization theory takes the following *singular* perturbation expansion of $\boldsymbol{u}$:

$$\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{u}^0(\boldsymbol{X}, \boldsymbol{x}) + \epsilon \boldsymbol{u}^1(\boldsymbol{X}, \boldsymbol{x}) + \cdots. \tag{15}$$

This expansion is called singular since the terms in the right side are a function of $\boldsymbol{X}$ and $\boldsymbol{x}$. Note that ϵ defined by Eq. 14 appears as the coefficient of $\boldsymbol{u}^1$ as well as the slow spatial variable $\boldsymbol{X}$. A *regular* perturbation expansion considers the change in $\mathbf{c}$ only. That is, $\mathbf{c}$ is expressed as $\boldsymbol{c} = \boldsymbol{c}^0 + \epsilon \boldsymbol{c}^1$ with $\boldsymbol{c}^0$ being constant and $\boldsymbol{c}^1$ spatially varying, and this $\epsilon \boldsymbol{c}^1$ is regarded as the small change in $\mathbf{c}$ from the uniform state of $\boldsymbol{c}^0$. In view of this $\boldsymbol{c}$, the regular perturbation expansion is readily applied to $\boldsymbol{u}$, as

$$\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{u}^0(\boldsymbol{x}) + \epsilon \boldsymbol{u}^1(\boldsymbol{x}) + \cdots,$$

and the first term $\boldsymbol{u}^0$ is the solution of the homogeneous body problem, and the later terms correct $\boldsymbol{u}^0$; $\boldsymbol{u}^0 + \epsilon \boldsymbol{u}^1$ is a good approximate solution of a body which consists of $\boldsymbol{c}^0 + \epsilon \boldsymbol{c}^1$ if ϵ is small.

The target of the singular perturbation expansion is the first term and hence $\nabla_{\boldsymbol{x}}$ for $\boldsymbol{x}$ is replaced with $\nabla_{\boldsymbol{X}}$ for $\boldsymbol{X}$, which is now evaluated in terms of $\boldsymbol{X}$ and $\boldsymbol{x}$ as

$$\nabla_{\boldsymbol{x}} = \nabla_{\boldsymbol{x}} + \frac{1}{\epsilon} \nabla_{\boldsymbol{X}},$$

where subscript $\boldsymbol{X}$ or $\boldsymbol{x}$ emphasizes the operator for $\boldsymbol{X}$ or $\boldsymbol{x}$, respectively. Substitution of Eq. 15 into Eq. 4 together with the above differential operator, the homogenization theory yields

$$\begin{aligned} &\epsilon^{-2} \left\{ \nabla_{\boldsymbol{x}} \cdot \left(\boldsymbol{c} : \nabla_{\boldsymbol{x}} \boldsymbol{u}^0 \right) \right\} \\ &+ \epsilon^{-1} \left\{ \nabla_{\boldsymbol{X}} \cdot \left(\boldsymbol{c} : \nabla_{\boldsymbol{x}} \boldsymbol{u}^0 \right) + \nabla_{\boldsymbol{x}} \cdot \left(\boldsymbol{c} : \left(\nabla_{\boldsymbol{X}} \boldsymbol{u}^0 + \nabla_{\boldsymbol{x}} \boldsymbol{u}^1 \right) \right) \right\} \\ &+ \epsilon^0 \left\{ \nabla_{\boldsymbol{X}} \cdot \left(\boldsymbol{c} : \left(\nabla_{\boldsymbol{X}} \boldsymbol{u}^0 + \nabla_{\boldsymbol{x}} \boldsymbol{u}^1 \right) \right) + \cdots \right\} + \cdots = 0. \end{aligned} \tag{16}$$

For the term of ϵ^{-2} to vanish, $\nabla_{\boldsymbol{x}} \boldsymbol{u}^0$ must be 0. That is, $\boldsymbol{u}^0$ is a function of $\boldsymbol{X}$ only. For the terms of ϵ^{-1} to vanish, the form of $\boldsymbol{u}^1$ is assumed as

$$\boldsymbol{u}^1(\boldsymbol{X}, \boldsymbol{x}) = \boldsymbol{\chi}(\boldsymbol{x}) : \left(\nabla_{\boldsymbol{X}} \boldsymbol{u}^0(\boldsymbol{X}) \right),$$

and the term becomes $\nabla_{\boldsymbol{x}} \cdot \left(\boldsymbol{c} : \left(\nabla_{\boldsymbol{x}} \boldsymbol{\chi} + \boldsymbol{I} \right) : \left(\nabla_{\boldsymbol{X}} \boldsymbol{u}^0 \right) \right)$. This term vanishes if $\boldsymbol{\chi}$ satisfies

$$\nabla_{\boldsymbol{x}} \cdot \left(\boldsymbol{c}(\boldsymbol{x}) : \left(\nabla_{\boldsymbol{x}} \boldsymbol{\chi}(\boldsymbol{x}) + \boldsymbol{I} \right) \right) = \mathbf{0}. \tag{17}$$

The term of ϵ^0 is now rewritten as $\nabla_{\boldsymbol{X}} \cdot (\boldsymbol{c} : (\nabla_{\boldsymbol{x}} \boldsymbol{\chi} + \boldsymbol{I}) : \nabla_{\boldsymbol{X}} \boldsymbol{u}^0)$. It is this $\boldsymbol{c} : (\nabla_{\boldsymbol{x}} \boldsymbol{\chi} + \boldsymbol{I})$ that plays a role of an overall elasticity for $\boldsymbol{u}^0$, even though it is a function of $\boldsymbol{x}$, not $\boldsymbol{X}$. Hence, the homogenization theory replaces $\boldsymbol{c} : (\nabla_{\boldsymbol{x}} \boldsymbol{\chi} + \boldsymbol{I})$ with its volume average taken over a suitable volume, i.e.,

$$\overline{\boldsymbol{C}} = \boldsymbol{c} : (\nabla_{\boldsymbol{x}} \boldsymbol{\chi}(\boldsymbol{x}) + \boldsymbol{I})_U. \tag{18}$$

Here, U is a certain domain in which the average is taken; this U must be in the micro-length scale so that $\langle \rangle_U$ is independent from $\boldsymbol{x}$.

Use of Periodic Structure as Microstructure Model

The two tasks remain to estimate $\overline{\boldsymbol{C}}$ by computing Eq. 18. Namely, solving Eq. 17 for $\boldsymbol{\chi}$ and determining U. As shown in Fig. 2, the homogenization theory usually uses a periodic structure, assuming the microstructure of the target heterogeneous material is more or less the same; see, for instance, Nuna and Keller (1984) and Walker et al. (1991); Oleinik et al. (1992) for the homogenization theory of nonperiodic media. $\boldsymbol{\chi}$ is thus computed for this periodic structure, and a unit cell of the periodic structure is chosen for U.

Unlike the average field theory, the homogenization theory is based on the mathematical approach of the singular perturbation expansion, in solving the governing equation of $\boldsymbol{u}$ when $\boldsymbol{c}$ changes in the micro-length scale of l. An advantage of this theory is to analyze only the leading term in the expansion, $\boldsymbol{u}^0$, which is a function of $\boldsymbol{X}$ or changes in the macro-length scale of L. Note that $\boldsymbol{u}^0$ accompanies strain given by

$$\mathrm{sym}\{(\nabla_{\boldsymbol{x}} \boldsymbol{\chi} + \boldsymbol{I}) : (\nabla_{\boldsymbol{X}} \boldsymbol{u}^0)\} = \mathrm{sym}\{\nabla_{\boldsymbol{x}} \boldsymbol{\chi} : \nabla_{\boldsymbol{X}} \boldsymbol{u}^0\} + \mathrm{sym}\{\nabla_{\boldsymbol{X}} \boldsymbol{u}^0\}.$$

It is the first term of the right side that reflects the spatial change in $\boldsymbol{c}$ in the micro-length scale, while the second term is strain which changes in the macro-length scale. By definition, Eq. 18, the effect of the first term on stress which changes in the macro-length scale is included in $\overline{\boldsymbol{C}}$ by taking volume average over U.

It is of interest to take the limit as ϵ goes to 0 in the homogenization theory. Numerous researches have been made for this limit. On the viewpoint of the perturbation expansion, however, it is standard to use a finite value for ϵ, using the definition of Eq. 13. It is thus possible to separately treat the homogenization theory from the singular perturbation expansion when the theory focuses on the special case of ϵ approaches 0.

Comparison of Average Field Theory and Homogenization Theory

In Table 1, the comparison of the average field theory and the homogenization theory is summarized. The field variables used in these theories are different; the

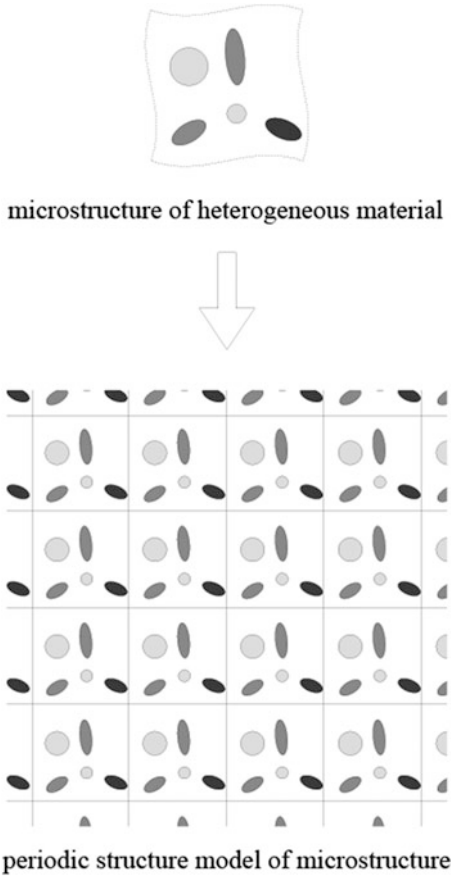


Fig. 2 Periodic structure for estimating overall property of heterogeneous material

Table 1 Comparison of average field theory and homogenization theory

	Average field theory	Homogenization theory
Displacement	$\boldsymbol{u}$	$\boldsymbol{u}^0 + \epsilon\boldsymbol{\chi} : (\nabla \boldsymbol{u}^0)$
Strain	$\boldsymbol{\varepsilon} = \text{sym}\{\nabla \boldsymbol{u}\}$	$\boldsymbol{\varepsilon}^0 = (\nabla\boldsymbol{\chi} + \boldsymbol{I}) : (\nabla \boldsymbol{u}^0)$
Stress	$\boldsymbol{\sigma} = \boldsymbol{c} : \boldsymbol{\varepsilon}$	$\boldsymbol{\sigma}^0 = \boldsymbol{c} : \boldsymbol{\varepsilon}^0$
Overall displacement	$\boldsymbol{U} = \int \varphi_V \boldsymbol{u} \mathrm{d}v$	$\boldsymbol{u}^0$
Overall strain	$\boldsymbol{E} = \text{sym}\{\nabla \boldsymbol{U}\}$	$\langle \boldsymbol{\varepsilon}^0 \rangle = \text{sym}\{\nabla \boldsymbol{u}^0\}$
Overall stress	$\boldsymbol{\Sigma} = \overline{\boldsymbol{C}} : \boldsymbol{E}$	$\langle \boldsymbol{\sigma}^0 \rangle = \overline{\boldsymbol{C}} : \langle \boldsymbol{\varepsilon}^0 \rangle$
Overall elasticity	$\overline{\boldsymbol{C}} = \boldsymbol{C}^M + f\left(\boldsymbol{C}^I - \boldsymbol{C}^M\right) : \boldsymbol{A}$	$\overline{\boldsymbol{C}} = \langle \boldsymbol{c} : (\nabla\boldsymbol{\chi} + \boldsymbol{I}) \rangle$

displacement function is different, and hence the strain and stress become different. The overall field variables are defined as the volume average of the field variables, even though the domain in which the average is taken is different.

It is of interest to note that while the homogenization theory is based on the singular perturbation, the first term, $\boldsymbol{u}^0$, is actually computed by taking the volume

average of the strain due to the second term, $\boldsymbol{u}^1$, even though the domain in which the average is taken is a unit cell. While a certain similarity is found for the two theories, there are two major differences between them. The first difference is the modeling of the microstructure: the homogenization theory uses a unit cell of the periodic structure, while the average field theory considers an RVE. The second difference is that the homogenization theory is able to compute higher-order terms, if necessary, but the average field theory does not have systematic procedures to increase the accuracy of the estimation.

The abovementioned differences are not essential. And these two theories are unified to construct a theory of analytically estimating overall property; the singular perturbation expansion that is employed by the homogenization theory is applied to a nonperiodic microstructure, and higher-order terms are computed by taking the volume average that is a core concept of the average field theory. Indeed, the essential procedures of the unified theory are stated as follows: (1) to use an RVE instead of the periodic structure, (2) to apply the singular perturbation expansion to the field in the RVE, and (3) to take the volume average of the terms in the expansion to compute the overall elasticity.

In the unified theory, setting of the boundary conditions of the target RVE needs some consideration. As will be explained later, it is sufficient to choose either linear displacement boundary conditions or uniform traction boundary conditions, respectively, if the resulting boundary tractions or displacement do not change wildly. A natural choice of the RVE configuration is cubic, and a cubic RVE is regarded as a unit cell if periodic boundary conditions are used instead of linear displacement boundary conditions or uniform traction boundary conditions. As will be explained later again, least dependence of the overall property on the boundary conditions is important to make the property consistent (i.e., the overall property is able to relate strain and stress as well as strain and strain energy density).

Strain Energy Consideration

Consistency of Overall Elasticity

Besides relating strain to stress, elasticity relates strain to strain energy density. The average field theory extensively studies this dual role of the elasticity; see Hill (1963). That is, an overall elasticity of an RVE is required to relate the average strain to the average strain energy density as follows:

$$\langle e \rangle_V = \frac{1}{2} \langle \boldsymbol{\varepsilon} \rangle_V : \overline{\boldsymbol{C}} : \langle \boldsymbol{\varepsilon} \rangle_V, \tag{19}$$

where $e = \frac{1}{2} \boldsymbol{\varepsilon} : \mathbf{c} : \boldsymbol{\varepsilon}$ or $e = \frac{1}{2} \boldsymbol{\sigma} : \boldsymbol{\varepsilon}$ is strain energy density. A question arises regarding the dual role of the overall elasticity, i.e., whether the two $\overline{\boldsymbol{C}}$'s of Eqs. 10 and 19 are the same or whether the overall property that relates the average

strain to the average stress does relate the average strain to the average strain energy density. This is called the consistency of $\overline{\boldsymbol{C}}$ in this article. It is easily understood that the condition for $\overline{\boldsymbol{C}}$ to be consistent is that the average of strain energy density coincides half of the product of the average stress and the average strain, $\langle e \rangle_V = \frac{1}{2}\langle \boldsymbol{\sigma} \rangle_V : \langle \boldsymbol{\varepsilon} \rangle_V$ or $\frac{1}{2}\langle \boldsymbol{\sigma} : \boldsymbol{\varepsilon} \rangle_V = \frac{1}{2}\langle \boldsymbol{\sigma} \rangle_V : \langle \boldsymbol{\varepsilon} \rangle_V$.

An averaging scheme for the product of strain and stress is readily derived from the two field equations, Eqs. 1 and 2, as

$$\langle \boldsymbol{\sigma} : \boldsymbol{\varepsilon} \rangle_V = \frac{1}{V} \int_{\partial V} \boldsymbol{t} \cdot \boldsymbol{u} \ \mathrm{d}s; \tag{20}$$

recall that $\boldsymbol{t}$ is the surface traction. Using the averaging schemes for strain and stress, Eqs. 5 and 6, the following equation is derived from Eq. 20:

$$\langle \boldsymbol{\sigma} : \boldsymbol{\varepsilon} \rangle_V - \langle \boldsymbol{\sigma} \rangle_V : \langle \boldsymbol{\varepsilon} \rangle_V = \frac{1}{V} \int_{\partial V} \left(\boldsymbol{t} - \boldsymbol{\nu} \cdot \langle \boldsymbol{\sigma} \rangle_V \right) \cdot \left(\boldsymbol{u} - \boldsymbol{x} \cdot \langle \boldsymbol{\varepsilon} \rangle_V \right) \mathrm{d}s. \tag{21}$$

As is seen, the left side, the difference between the average of strain and stress product, $\langle \boldsymbol{\sigma} : \boldsymbol{\varepsilon} \rangle_V$ and the product of average strain and stress, $\langle \boldsymbol{\sigma} \rangle_V : \langle \boldsymbol{\varepsilon} \rangle_V$, is given as the surface integration of the product of $\boldsymbol{t} - \boldsymbol{\nu} \cdot \langle \boldsymbol{\sigma} \rangle_V$ and $\boldsymbol{u} - \boldsymbol{x} \cdot \langle \boldsymbol{\varepsilon} \rangle_V$, which are the deviation of the traction from that computed in terms of the average stress ($\boldsymbol{\nu} \cdot \langle \boldsymbol{\sigma} \rangle_V$) and the deviation of the displacement from that computed in terms of the average strain ($\boldsymbol{x} \cdot \langle \boldsymbol{\varepsilon} \rangle_V$), respectively.

Condition for Consistent Overall Elasticity

In view of Eq. 20, it is clear that $\overline{\boldsymbol{C}}$'s of Eqs. 10 and 19 are the same if the RVE is subjected to either uniform traction boundary conditions or linear displacement boundary conditions, so that $\boldsymbol{t} - \boldsymbol{\nu} \cdot \langle \boldsymbol{\sigma} \rangle_V$ or $\boldsymbol{u} - \boldsymbol{x} \cdot \langle \boldsymbol{\varepsilon} \rangle_V$ identically vanishes on ∂V, respectively. Furthermore, if boundary conditions are chosen so that

$$\left| \left(\boldsymbol{t} - \boldsymbol{\nu} \cdot \langle \boldsymbol{\sigma} \rangle_V \right) \cdot \left(\boldsymbol{u} - \boldsymbol{x} \cdot \langle \boldsymbol{\varepsilon} \rangle_V \right) \right| < w$$

with w being a constant, then, Eq. 20 leads to

$$\left| \langle \boldsymbol{\sigma} : \boldsymbol{\varepsilon} \rangle_V - \langle \boldsymbol{\sigma} \rangle_V : \langle \boldsymbol{\varepsilon} \rangle_V \right| < \frac{1}{V} \int_{\partial V} \left| \left(\boldsymbol{t} - \boldsymbol{\nu} \cdot \boldsymbol{\sigma}_V \right) \cdot \left(\boldsymbol{u} - \boldsymbol{x} \cdot \boldsymbol{\varepsilon}_V \right) \right| \mathrm{d}s < \frac{Sw}{V},$$

where S and V are the area of ∂V and the volume of V; the same symbol is used for the volume of V. If Sw/V is sufficiently small, $\overline{\boldsymbol{C}}$'s of Eqs. 10 and 19 are regarded as being approximately consistent. Note that when w is fixed, Sw/V decreases as the size of RVE increases since S/V is of the order of the inverse of the RVE's size.

Dependence of Overall Elasticity on Loading Condition

Another question arises regarding the property of $\overline{\boldsymbol{C}}$. The consistent overall elasticity of $\overline{\boldsymbol{C}}$ is defined by using field variables of an RVE which is subjected to certain boundary conditions, namely, linear displacement boundary conditions or uniform traction boundary conditions (which are typical examples of boundary conditions satisfying $|(\boldsymbol{t} - \boldsymbol{\nu} \cdot \langle \boldsymbol{\sigma} \rangle_V) \cdot (\boldsymbol{u} - \boldsymbol{x} \cdot \langle \boldsymbol{\varepsilon} \rangle_V)| < w$). So, the question is whether $\overline{\boldsymbol{C}}$ is independent from the boundary conditions or from the loading conditions of the RVE. The answer is negative. $\overline{\boldsymbol{C}}$ changes depending on the boundary conditions. Putting superscripts S, G, and E for field variables due to uniform traction, general, linear displacement boundary conditions, respectively, the following averaging scheme holds for the average strain energy density of the RVE subjected to different boundary conditions:

$$\begin{cases} \langle e^S \rangle_V \le \langle e^G \rangle_V & \text{for } \langle \boldsymbol{\varepsilon}^G \rangle_V = \langle \boldsymbol{\varepsilon}^S \rangle_V, \\ \langle e^E \rangle_V \le \langle e^G \rangle_V & \text{for } \langle \boldsymbol{\sigma}^G \rangle_V = \langle \boldsymbol{\sigma}^E \rangle_V; \end{cases} \tag{22}$$

see Fig. 3 (Nemat-Nasser and Hori 1995). The proof is simple, just by computing the right side of the following inequalities:

$$0 \le \left\langle \left(\boldsymbol{\varepsilon}^G - \boldsymbol{\varepsilon}^S\right) : \boldsymbol{c} : \left(\boldsymbol{\varepsilon}^G - \boldsymbol{\varepsilon}^S\right) \right\rangle_V \text{ and } 0 \le \left\langle \left(\boldsymbol{\varepsilon}^G - \boldsymbol{\varepsilon}^E\right) : \boldsymbol{c} : \left(\boldsymbol{\varepsilon}^G - \boldsymbol{\varepsilon}^E\right) \right\rangle_V.$$

The meaning of Eq. 22 is clear, if it is interpreted that the average strain energy density is increased by putting more wildly varying surface displacement or traction on the RVE's boundary and that the average strain energy density attains the lowest value if there is no varying surface displacement or traction. While not proved, it is intuitively clear that if surface displacement or traction does not change wildly (just as an inequality of $|(\boldsymbol{t} - \boldsymbol{\nu} \cdot \langle \boldsymbol{\sigma} \rangle_V) \cdot (\boldsymbol{u} - \mathbf{x} \cdot \langle \boldsymbol{\varepsilon} \rangle_V)| < w$ holds), $\langle e^G \rangle_V$ would be close to both $\langle e^S \rangle_V$ and $\langle e^E \rangle_V$; the difference, $\langle e^G \rangle_V - \langle e^S \rangle_V$ or $\langle e^G \rangle_V - \langle e^E \rangle_V$, would

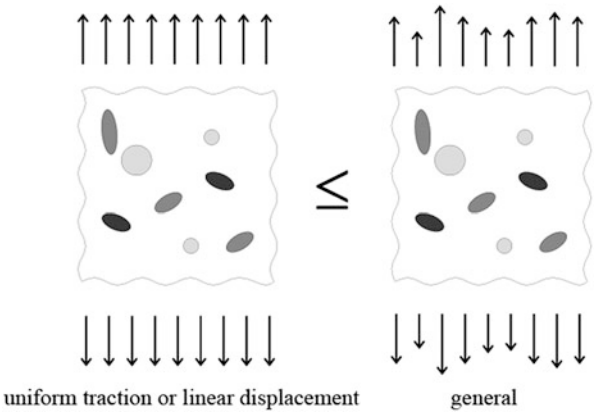


Fig. 3 Dependence of strain energy on boundary conditions

vanish as the size of the RVE increases. Therefore, the dependence of $\overline{\boldsymbol{C}}$ on the boundary conditions would be negligible for a sufficiently large RVE.

It is of interest to note that Eq. 22, the inequalities of the average strain energy density, can be derived from following functional:

$$I^E(\boldsymbol{u}, \bar{\boldsymbol{t}}) = \left\langle \frac{1}{2}\boldsymbol{\varepsilon} : \boldsymbol{c} : \boldsymbol{\varepsilon} \right\rangle_V + \frac{1}{V}\int_{\partial V} \boldsymbol{\lambda}(\boldsymbol{x}) \cdot (\boldsymbol{\nu}(\boldsymbol{x}) \cdot \boldsymbol{\sigma}(\boldsymbol{x}) - \bar{\boldsymbol{t}}(\boldsymbol{x}))\mathrm{d}s + \boldsymbol{\mu} : \left(\langle \boldsymbol{\sigma} \rangle_V - \overline{\boldsymbol{\sigma}}\right).$$

Here, $\bar{\boldsymbol{t}}$ and $\overline{\boldsymbol{\sigma}}$ are the traction described on $\overline{\boldsymbol{\sigma}}$ and a given average stress, respectively, and $\boldsymbol{\sigma}$ is a stress field which is associated with $\boldsymbol{u}$; $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are the Lagrange multipliers enforcing the boundary condition, $\boldsymbol{\nu}(\boldsymbol{x}) \cdot \boldsymbol{\sigma}(\boldsymbol{x}) = \bar{\boldsymbol{t}}(\boldsymbol{x})$, and the average stress condition, $\langle \boldsymbol{\sigma} \rangle_V = \overline{\boldsymbol{\sigma}}$ ($\boldsymbol{\lambda}$ is a vector function defined on ∂V, and $\boldsymbol{\mu}$ is a second-order tensor defined on ∂V). A stationary value of I^E or $\left\langle \frac{1}{2}\boldsymbol{\varepsilon} : \boldsymbol{c} : \boldsymbol{\varepsilon} \right\rangle = \langle e \rangle_V$ is found by taking the variation with respect to both $\boldsymbol{u}$ and $\overline{\boldsymbol{u}}$. For instance, the variation of I^E with respect to $\overline{\boldsymbol{u}}$ is

$$\delta I^E(\boldsymbol{u}, \bar{\boldsymbol{t}}) = \frac{1}{V}\int_{\partial V} \delta\bar{\boldsymbol{t}}(\boldsymbol{x}) \cdot (\boldsymbol{u}(\boldsymbol{x}) - \boldsymbol{x} \cdot \boldsymbol{\mu})\mathrm{d}s,$$

Since $\boldsymbol{\mu}$ is a constant second-order tensor, the Euler equation shown in the above means that I^E is stationarized for $\bar{\boldsymbol{t}}$ that makes the associated displacement $\boldsymbol{u}$ on ∂V linear with respect to $\boldsymbol{x}$. That is, the linear displacement boundary conditions minimize I^E, and the inequality $\langle e^E \rangle_V \leq \langle e^G \rangle_V$ holds for any $\boldsymbol{u}^G$ which accompanies $\boldsymbol{\sigma}^G$ that satisfies $\langle \boldsymbol{\sigma}^G \rangle_V = \langle \boldsymbol{\sigma}^E \rangle_V$. It is similarly shown that the inequality $\langle e^S \rangle_V \leq \langle e^G \rangle_V$ holds for any displacement field $\boldsymbol{u}^G$ which accompanies $\boldsymbol{\varepsilon}^G$ that satisfies $\langle \boldsymbol{\varepsilon}^G \rangle_V = \langle \boldsymbol{\varepsilon}^S \rangle_V$ from the following functional:

$$I^S(\boldsymbol{u}, \overline{\boldsymbol{u}}) = \left\langle \frac{1}{2}\boldsymbol{\varepsilon} : \boldsymbol{c} : \boldsymbol{\varepsilon} \right\rangle_V + \frac{1}{V}\int_{\partial V} \boldsymbol{\lambda}(\boldsymbol{x}) \cdot (\boldsymbol{u}(\boldsymbol{x}) - \overline{\boldsymbol{u}}(\boldsymbol{x}))\mathrm{d}s + \boldsymbol{\mu} : \left(\langle \boldsymbol{\varepsilon} \rangle_V - \overline{\boldsymbol{\varepsilon}}\right),$$

where $\overline{\boldsymbol{u}}$ and $\overline{\boldsymbol{\varepsilon}}$ are a displacement field which is defined on ∂V and a given average strain and $\boldsymbol{\varepsilon}$ is a strain field associated with $\boldsymbol{u}$.

Hashin–Shtrikman Variational Principle

Fictitious Uniform RVE

A unique approach of estimating the overall property of an RVE is the use of the variational principle. This is called the Hashin–Shtrikman variational principle (Hashin and Shtrikman 1962); see Walpole (1969), Willis (1977), Milton and Kohn (1988), Torquato (1991), and Munashinghe et al. (1996). It uses a fictitious but homogeneous RVE and seeks to find a distribution of *eigen-stress* so that the field variables in the homogeneous RVE coincide with those in the original heterogeneous RVE. Even if the exact distribution is not found, a suitable

distribution of the eigen-stress provides an upper or lower bound for the strain energy of the original heterogeneous RVE.

The formulation of the Hashin–Shtrikman variational principle starts from eigen-stress, denoted by $\boldsymbol{\sigma}^*$, and the following stress–strain relation holds in the homogeneous RVE, denoted by V^{o}:

$$\boldsymbol{\sigma}(\mathbf{x}) = \boldsymbol{c}^{\mathrm{o}} : \boldsymbol{\varepsilon}(\boldsymbol{x}) + \boldsymbol{\sigma}^*(\boldsymbol{x}), \tag{23}$$

where $\boldsymbol{c}^{\mathrm{o}}$ is the elasticity of the homogeneous RVE. The other two field equations, Eqs. 1 and 2, hold in the homogeneous RVE. The governing equation for $\boldsymbol{u}$ is thus given as

$$\nabla \cdot (\boldsymbol{c}^{\mathrm{o}} : \nabla \boldsymbol{u}(\boldsymbol{x})) + \nabla \cdot \boldsymbol{\sigma}^*(\boldsymbol{x}) = 0. \tag{24}$$

Regarding $\nabla \cdot \boldsymbol{\sigma}^*$ as a body force, Eq. 24 can be solved for $\boldsymbol{u}$ by using the fundamental solution of the homogeneous body, $\boldsymbol{G}^{\mathrm{o}}$; this $\boldsymbol{G}^{\mathrm{o}}$ is often called Green's function of an infinitely extended body, instead of the fundamental solution of the partial differential equation. The contribution of $\boldsymbol{\sigma}^*$ on $\boldsymbol{u}$ is given as $\int \boldsymbol{G}^{\mathrm{o}}(\boldsymbol{x} - \boldsymbol{y}) \cdot (\nabla \cdot \boldsymbol{\sigma}^*(\boldsymbol{y}))\mathrm{d}v$, where the integration is taken with respect to $\boldsymbol{y}$.

Now, suppose the linear displacement boundary conditions of the original RVE are posed for the homogeneous RVE. If $\boldsymbol{\sigma}^*$ satisfies

$$\boldsymbol{c}^{\mathrm{o}} : \boldsymbol{\varepsilon}(\boldsymbol{x}) + \boldsymbol{\sigma}^*(\boldsymbol{x}) = \boldsymbol{c}(\boldsymbol{x}) : \boldsymbol{\varepsilon}(\boldsymbol{x}), \tag{25}$$

the field variables of the homogeneous RVE coincide with those of the original RVE. Note that $\boldsymbol{\varepsilon}$ in both sides of Eq. 25 are the strain in the homogeneous RVE.

Hashin–Shtrikman Functional for Eigen-stress

A functional for $\boldsymbol{\sigma}^*$ is readily defined so that its Euler equation becomes Eq. 25. That is,

$$J^E(\boldsymbol{\sigma}^*) = \int_{V^{\mathrm{o}}} \frac{1}{2}\boldsymbol{\sigma}^*(\boldsymbol{x}) : (\boldsymbol{c}(\boldsymbol{x}) - \boldsymbol{c}^{\mathrm{o}})^{-1} : \boldsymbol{\sigma}^*(\boldsymbol{x}) - \frac{1}{2}\boldsymbol{\sigma}^*(\boldsymbol{x}) : \boldsymbol{\varepsilon}^{\mathrm{d}}(\boldsymbol{x}) - \boldsymbol{\sigma}^*(\boldsymbol{x}) : \boldsymbol{\varepsilon}^{\mathrm{o}}(\boldsymbol{x})\mathrm{d}v; \tag{26}$$

$\boldsymbol{\varepsilon}^{\mathrm{d}}$ is strain caused by $\boldsymbol{\sigma}^*$ and can be expressed in terms of $\boldsymbol{G}^{\mathrm{o}}$, and $\boldsymbol{\varepsilon}^{\mathrm{o}}$ is a strain field caused by the given displacement boundary conditions in the absence of $\boldsymbol{\sigma}^*$. Since $\boldsymbol{\varepsilon}^{\mathrm{d}}$ and $\boldsymbol{\sigma}^*$ satisfy

$$\int_{V^{\mathrm{o}}} \boldsymbol{\sigma}^*(\boldsymbol{x}) : \boldsymbol{\varepsilon}^{\mathrm{d}}(\boldsymbol{x})\mathrm{d}v = -\int_{V^{\mathrm{o}}} \boldsymbol{\varepsilon}^{\mathrm{d}}(\boldsymbol{x}) : c^{\mathrm{o}} : \boldsymbol{\varepsilon}^{\mathrm{d}}(\boldsymbol{x})\mathrm{d}v,$$

and hence the variation of J^E leads to Eq. 25; note that the above equation is derived from a condition of $\boldsymbol{u}^{\mathrm{d}} = \boldsymbol{0}$ on ∂V^{o}.

When $\boldsymbol{c}^{\mathrm{o}}$ is chosen so that $\boldsymbol{c} - \boldsymbol{c}^{\mathrm{o}}$ becomes negative definite in V^{o}, the stationary value of J^E is actually the minimum value. That is, for any $\boldsymbol{\sigma}^*$, the following inequality holds:

$$J^E(\boldsymbol{\sigma}^*) > \int_{V^{\mathrm{o}}} -\frac{1}{2}((\mathbf{c}(\mathbf{x}) - \mathbf{c}^{\mathrm{o}}) : \boldsymbol{\varepsilon}(\mathbf{x})) : \boldsymbol{\varepsilon}(\mathbf{x})\mathrm{d}v = \left\langle \frac{1}{2}\boldsymbol{\varepsilon} : \mathbf{c}^{\mathrm{o}} : \boldsymbol{\varepsilon} \right\rangle_{V^{\mathrm{o}}} - \langle e \rangle_V.$$

Here, $\boldsymbol{\varepsilon}$ is the strain of the strain in the original RVE subjected to the same displacement boundary conditions (or the strain that is computed for $\boldsymbol{\sigma}^*$ satisfying Eq. 25), and $\langle e \rangle_V$ is the average strain energy density that corresponds to this $\boldsymbol{\varepsilon}$. Hence, a lower bound is obtained for $\langle e \rangle_V$.

$$\langle e \rangle_V > \left\langle \frac{1}{2}\boldsymbol{\varepsilon} : \mathbf{c}^{\mathrm{o}} : \boldsymbol{\varepsilon} \right\rangle_{V^{\mathrm{o}}} - J^E(\boldsymbol{\sigma}^*). \tag{27}$$

The first term in the right side of this inequality is $\frac{1}{2}\langle \boldsymbol{\varepsilon} \rangle_V : \mathbf{c}^{\mathrm{o}} : \langle \boldsymbol{\varepsilon} \rangle_V$, since $\boldsymbol{\varepsilon}$ is that strain which corresponds to the linear displacement boundary conditions. The second term is evaluated by using a piecewise constant distribution of $\boldsymbol{\sigma}^*$ and Eshelby's solution. Thus, the lower bound for $\langle e \rangle_V$ provides a lower bound for the overall property.

In the above discussion, linear displacement boundary conditions are assumed in computing the lower bound for $\langle e \rangle_V$ of V. However, the principle holds if any displacement boundary conditions are posed on for ∂V. This is because a key condition of the principle is $\int \boldsymbol{\sigma}^* : \boldsymbol{\varepsilon}^{\mathrm{d}}\mathrm{d}v = -\int \boldsymbol{\varepsilon}^{\mathrm{d}} : \boldsymbol{c}^{\mathrm{o}} : \boldsymbol{\varepsilon}^{\mathrm{d}}\mathrm{d}v$, which is derived from the condition that displacement caused by $\boldsymbol{\sigma}^*$ vanishes on ∂V.

In a similar manner, an upper bound for the overall property can be computed by considering the case when traction boundary conditions are given, and $\boldsymbol{c}^{\mathrm{o}}$ is chosen so that $\boldsymbol{c} - \boldsymbol{c}^{\mathrm{o}}$ becomes positive definite. That is, the following functional is considered:

$$J^S(\boldsymbol{\sigma}^*) = \int_{V^{\mathrm{o}}} \frac{1}{2}\boldsymbol{\varepsilon}^*(\boldsymbol{x}) : (\boldsymbol{d}(\boldsymbol{x}) - \boldsymbol{d}^{\mathrm{o}})^{-1} : \boldsymbol{\varepsilon}^*(\boldsymbol{x}) - \frac{1}{2}\boldsymbol{\varepsilon}^*(\boldsymbol{x}) : \boldsymbol{\sigma}^{\mathrm{d}}(\boldsymbol{x}) - \boldsymbol{\varepsilon}^*(\boldsymbol{x}) : \boldsymbol{\sigma}^{\mathrm{o}}(\boldsymbol{x})\mathrm{d}v;$$

where $\boldsymbol{\varepsilon}^*(\boldsymbol{x})$ is defined as $\boldsymbol{\varepsilon}^*(\boldsymbol{x}) = (\boldsymbol{c}^{\mathrm{o}})^{-1} : \boldsymbol{\sigma}^*(\boldsymbol{x})$; $\boldsymbol{d}^{\mathrm{o}}$ and $\boldsymbol{d}$ are the inverse tensor of $\boldsymbol{c}^{\mathrm{o}}$ and $\boldsymbol{c}$ (usually called compliance tensor), respectively; $\boldsymbol{\sigma}^{\mathrm{d}}$ is the stress caused by $\boldsymbol{\sigma}^*$; and $\boldsymbol{\sigma}^{\mathrm{o}}$ is a stress field caused by the given traction boundary conditions in the absence of $\boldsymbol{\sigma}^*$. Note that just like $\int \boldsymbol{\sigma}^* : \boldsymbol{\varepsilon}^{\mathrm{d}}\mathrm{d}v$ in J^E is negative definite, $\int \boldsymbol{\varepsilon}^* : \boldsymbol{\sigma}^{\mathrm{d}}\mathrm{d}v$ in J^S is negative definite and that the traction boundary conditions do not have to be uniform; the principle holds for any traction boundary conditions.

Application of Hashin–Shtrikman Variational Principle to Periodic Structure

A homogeneous RVE is introduced for the original heterogeneous RVE, in order to formulate the Hashin–Shtrikman variational principle. It gives an impression that this principle holds only in the framework of the average field theory. However, the principle can be applied to a periodic structure of the homogenization theory, if the following identity holds:

$$\int \boldsymbol{\sigma}^*(\boldsymbol{x}) : \boldsymbol{\varepsilon}^{\mathrm{d}}(\boldsymbol{x})\mathrm{d}v = -\int \boldsymbol{\varepsilon}^{\mathrm{d}}(\boldsymbol{x}) : \boldsymbol{c}^{\mathrm{o}} : \boldsymbol{\varepsilon}^{\mathrm{d}}(\boldsymbol{x})\mathrm{d}v.$$

Here, $\boldsymbol{\varepsilon}^{\mathrm{d}}$ is a strain field produced in the unit cell U when $\boldsymbol{\sigma}^*$ is given to the homogeneous periodic structure. Actually, the periodic boundary conditions which are posed for U satisfy the above equality; $\boldsymbol{\sigma}^*$ and $\boldsymbol{\varepsilon}^{\mathrm{d}}$ are the eigen-stress and the strain caused by $\boldsymbol{\sigma}^*$ in U.

Overall Property at Dynamics State

Averaging Scheme at Dynamic State

In this article, quasi-static state has been assumed in analyzing field variables to estimate overall property of a heterogeneous material. It is of interest to consider the estimation of overall property at dynamic state, applying the two micromechanics theories. At dynamic state, Eq. 2 is replaced by

$$\nabla \cdot \boldsymbol{\sigma}(\boldsymbol{x}, t) - \mathrm{D}\boldsymbol{p}(\boldsymbol{x}, t) = \boldsymbol{0}, \tag{28}$$

where t is time, D is the differential operator ($\mathrm{D}(\ ) = \partial(\ )/\partial t$), and $\boldsymbol{p}$ is momentum which is defined as

$$\boldsymbol{p}(\boldsymbol{x}, t) = \rho(\boldsymbol{x})\boldsymbol{v}(\boldsymbol{x}, t). \tag{29}$$

Here, ρ is density and $\boldsymbol{v}$ is velocity given as $\boldsymbol{v} = \mathrm{D}\boldsymbol{u}$.

According to the average field theory, an RVE at dynamic state is considered to estimate the overall property at dynamic state, and the volume average of field variables is computed. While the averaging scheme of strain, Eq. 5, holds, that of stress, Eq. 6, is replaced by

$$\langle \boldsymbol{\sigma} \rangle = \frac{1}{V}\int_{\partial V} \mathrm{sym}\{\boldsymbol{t} \otimes \boldsymbol{x}\}\mathrm{d}s - \langle \boldsymbol{p} \otimes \boldsymbol{x} \rangle, \tag{30}$$

together with

$$\langle \mathrm{D}\boldsymbol{p} \rangle = \frac{1}{V} \int_{\partial V} \boldsymbol{t} \ \mathrm{d}s. \tag{31}$$

Here, for simplicity, subscript V is excluded in $\langle\rangle$, which is a function of t. Note that Eq. 30 includes $\langle \boldsymbol{p} \otimes \boldsymbol{x} \rangle$; unlike Eq. 6 at quasi-static state, the average stress at dynamic state is not determined by the surface integration only. It is important to know that the average strain and stress of a material sample are *measurable* at quasi-static state. However, only the average strain is measurable at dynamic state; computing the average stress of the sample needs a distribution of $\boldsymbol{p} \otimes \boldsymbol{x}$ within the sample.

Fictitious Uniform RVE at Dynamic State

In order to estimate the overall property, the average field theory seeks to relate the volume averages which are computed by the averaging schemes. It is useful to analyze a homogeneous RVE, V^{o}, to which the fundamental solution at dynamic state is applicable, as explained in the previous section. Eigen-stress and eigen-momentum, denoted by $\boldsymbol{\sigma}^*$ and $\boldsymbol{p}^*$, respectively, disturb stress and momentum in V^{o} as

$$\boldsymbol{\sigma}(\boldsymbol{x}, t) = \boldsymbol{c}^{\mathrm{o}} : \boldsymbol{\varepsilon}(\boldsymbol{x}, t) + \boldsymbol{\sigma}^*(\boldsymbol{x}, t), \tag{32}$$

$$\boldsymbol{p}(\boldsymbol{x}, t) = \rho^{\mathrm{o}} \boldsymbol{v}(\boldsymbol{x}, t) + \boldsymbol{p}^*(\boldsymbol{x}, t), \tag{33}$$

from which the governing equation of $\boldsymbol{u}$ at dynamic state is derived as

$$\nabla \cdot (\boldsymbol{c}^{\mathrm{o}} : \nabla \boldsymbol{u}(\boldsymbol{x}, t)) - \rho^{\mathrm{o}} \mathrm{D}^2 \boldsymbol{u}(\boldsymbol{x}, t) + \nabla \cdot \boldsymbol{\sigma}^*(\boldsymbol{x}, t) - \mathrm{D}\boldsymbol{p}^*(\boldsymbol{x}, t) = \boldsymbol{0}. \tag{34}$$

As is seen, $\nabla \cdot \boldsymbol{\sigma}^* - \mathrm{D}\boldsymbol{p}^*$ plays a role of body force, and its contribution on $\boldsymbol{u}$ is computed by the spatial and temporal integration of the fundamental solution, which is expressed as the convolution form

$$\iint \boldsymbol{G}^{\mathrm{o}}(\boldsymbol{x} - \boldsymbol{y}, t - s) \cdot (-\nabla \cdot \boldsymbol{\sigma}^*(\boldsymbol{y}, s) + \mathrm{D}\boldsymbol{p}^*(\boldsymbol{y}, s)) \mathrm{d}v \mathrm{d}s = \boldsymbol{G}^{\mathrm{o}} * (-\nabla \cdot \boldsymbol{\sigma}^* + \mathrm{D}^*),$$

where $\boldsymbol{G}^{\mathrm{o}}$ is the fundamental solution at dynamic state, and differentiation and integration are taken with respect to $\boldsymbol{y}$ and s. After careful manipulations are carried out in substituting of $\boldsymbol{G}^{\mathrm{o}} * (-\nabla \cdot \boldsymbol{\sigma}^* + \mathrm{D}\boldsymbol{p}^*)$ into Eqs. 30 and 31, the volume average of stress and momentum is estimated as follows:

$$\langle \boldsymbol{\sigma} \rangle = \overline{\boldsymbol{C}} * \langle \boldsymbol{\varepsilon} \rangle + \overline{\boldsymbol{S}} * \langle \boldsymbol{v} \rangle, \tag{35}$$

$$\langle \mathbf{p} \rangle = \overline{\mathbf{S}}^t * \langle \boldsymbol{\varepsilon} \rangle + \overline{\Upsilon} * \langle \boldsymbol{v} \rangle, \tag{36}$$

where $\overline{\boldsymbol{C}}$, $\overline{\Upsilon}$, and $\overline{\boldsymbol{S}}$ are convolution operators and superscript t stands for the transpose; there is symmetry for Eqs. 35 and 36 in the sense that $\overline{\boldsymbol{S}}$ and $\overline{\mathbf{S}}^t$ express the contribution of $\langle \boldsymbol{v} \rangle$ and $\langle \boldsymbol{\varepsilon} \rangle$ to $\langle \boldsymbol{\sigma} \rangle$ and $\langle \mathbf{p} \rangle$, respectively. The presence of these operators is proved, but the explicit form of the operators in terms of $\boldsymbol{G}^{\mathrm{o}}$ is not obtained.

It is of interest to note that the presence of the average velocity, $\langle \boldsymbol{v} \rangle$, in Eqs. 35 and 36 appears odd. The identical material properties ought to be measured by two observers who move in constant but different speed. However, $\langle \boldsymbol{v} \rangle$ would be different for these two observers, and hence the presence of $\langle \boldsymbol{v} \rangle$ in Eqs. 35 and 36 appears odd. Actually, contribution of $\langle \boldsymbol{v} \rangle$ would be replaced by $\langle \mathrm{D}\boldsymbol{\varepsilon} \rangle$, without breaking the *symmetry* of the operators and the averages, if temporal average is taken for Eqs. 35 and 36.

In closing the average field theory for dynamic state, the averaging scheme for strain energy is summarized as follows:

$$\langle \boldsymbol{\sigma} : \boldsymbol{\varepsilon} \rangle = \frac{1}{V} \int_{\partial V} \boldsymbol{t}(\boldsymbol{x}, t) \cdot \boldsymbol{u}(\boldsymbol{x}, t) \mathrm{d}s - \langle \mathrm{D}\boldsymbol{p} \cdot \boldsymbol{p} \rangle, \tag{37}$$

$$\begin{aligned} \langle \boldsymbol{\sigma} : \boldsymbol{\varepsilon} \rangle - \langle \boldsymbol{\sigma} \rangle : \langle \boldsymbol{\varepsilon} \rangle = & \frac{1}{V} \int_{\partial V} (\boldsymbol{t}(\boldsymbol{x}, t) - \boldsymbol{\nu} \cdot \langle \boldsymbol{\sigma} \rangle) \cdot (\boldsymbol{u}(\boldsymbol{x}, t) - \boldsymbol{x} \cdot \langle \boldsymbol{\varepsilon} \rangle) \mathrm{d}s - \langle \mathrm{D}\boldsymbol{p} \cdot \boldsymbol{u} \rangle \\ & + \langle \boldsymbol{x} \otimes \mathrm{D}\boldsymbol{p} \rangle : \langle \boldsymbol{\varepsilon} \rangle. \end{aligned} \tag{38}$$

These two equations are the dynamic version of Eqs. 20 and 21. Just like the quasi-static version, Eqs. 37 and 38 are the consequences of the two field equations, Eqs. 1 and 28, and hold for any arbitrary materials. Note that unlike quasi-static state, $\langle \boldsymbol{\sigma} : \boldsymbol{\varepsilon} \rangle$ and $\langle \boldsymbol{\sigma} : \boldsymbol{\varepsilon} \rangle - \langle \boldsymbol{\sigma} \rangle : \langle \boldsymbol{\varepsilon} \rangle$ are not determined by the surface integral only, which implies that measurement of field variables within V will be needed to evaluate these terms exactly.

Application of Singular Perturbation Expansion

Like the quasi-static state, the homogenization theory seeks to apply the singular perturbation expansion to the governing equation of $\mathbf{u}$ at dynamic state, i.e.,

$$\mathrm{D}(\rho(\boldsymbol{x}) \mathrm{D}\boldsymbol{u}(\boldsymbol{x}, t)) - \nabla \cdot (\boldsymbol{c}(\boldsymbol{x}) : \nabla \boldsymbol{u}(\boldsymbol{x}, t)) = \mathbf{0}.$$

If the expansion of $\boldsymbol{u}$ is made in the same manner as Eq. 15, i.e.,

$$\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{u}^0(\boldsymbol{X}, \boldsymbol{x}, t) + \epsilon \boldsymbol{u}^1(\boldsymbol{X}, \boldsymbol{x}, t) + \cdots,$$

then assuming that $\boldsymbol{u}^0$ is a function of $\boldsymbol{X}$ and t and that $\boldsymbol{u}^1$ is the form of $\boldsymbol{u}^1 = \boldsymbol{\chi} : \boldsymbol{u}^0$ with $\boldsymbol{\chi}$ being a function of $\boldsymbol{x}$ only, the overall elasticity is identical with $\overline{\boldsymbol{C}}$ of Eq. 18,

and the overall density, $\overline{\Upsilon}$, is the volume average of ρ taken over a unit cell U. That is,

$$\overline{\boldsymbol{C}} = \langle \boldsymbol{c}(\boldsymbol{x}) : \boldsymbol{\chi}(\boldsymbol{x}) + \boldsymbol{I} \rangle_U \text{ and } \overline{\Upsilon} = \langle \rho \rangle_U.$$

This estimate of $\overline{\boldsymbol{C}}$ and $\overline{\Upsilon}$ is intuitively acceptable.

While the dynamic overall property estimated by the average field theory is different from the quasi-static overall property, the homogenization theory produces the same estimate of the overall elasticity and a well-expected estimate of the overall density. This difference comes from the form of the expansion. The different phases will have different elastic wave velocity (which is determined by the ratio of the elasticity and the density). At dynamic state, therefore, each phase will have a different timescale for its elastic wave to travel in the phase, provided that all phases have more or less the same spatial size. Thus, an alternative (probably more realistic) expansion of $\boldsymbol{u}$ will be

$$\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{u}^0(\boldsymbol{X}, \boldsymbol{x}, T, t) + \epsilon \boldsymbol{u}^1(\boldsymbol{X}, \boldsymbol{x}, T, t) + \cdots,$$

where

$$T = \frac{1}{\epsilon} t$$

is a temporal variable slower than t, just as $\boldsymbol{X}$ is a spatial variable changing more slowly than $\boldsymbol{x}$. The results of estimating the overall elasticity and density will be different if the alternative expansion is taken.

Conclusion

The two micromechanics theories explained in this article are well established. It is not difficult at all to derive a closed-form expression or a numerical solution for the overall property according to these theories. The accuracy of the estimation is, in general, limited; the analytical estimation does not reach the level of an alternative of the experimental estimation, and the numerical solution, which is better than the closed-form solution, is not often used in practice. However, due to its suppleness, the analytical estimation could be useful to obtain a rough estimate of the overall property.

Besides practical use, the two theories of micromechanics give a clear concept of modeling the microstructure of a heterogeneous material. We can take advantage of this concept in order to estimate nonmechanical properties, such as electromagnetic property, thermal conductivity property, or coupling property among them. The Hashin–Shtrikman variational principle is worth being extended for this purpose, since it provides both upper and lower bounds for the overall property.

References

H. Ammari, H. Kang, M. Lim, Effective parameters of elastic composites. Indiana Univ. Math. J. **55**(3), 903–922 (2006)

N. Bakhvalov, G. Panasenko, *Homogenization: Averaging Processes in Periodic Media* (Kluwer, New York, 1984)

J.D. Eshelby, The determination of the elastic field of an ellipsoidal inclusion, and related problems. Proc. R. Soc. A **A241**, 376–396 (1957)

G.A. Francfort, F. Murat, Homogenization and optimal bounds in linear elasticity. Arch Ration. Mech. Anal. **94**, 307–334 (1986)

X.L. Gao, H.M. Ma, Strain gradient solution for the Eshelby-type anti-plane strain inclusion problem. Acta Mech. **223**, 1067–1080 (2012)

Z. Hashin, S. Shtrikman, On some variational principles in anisotropic and nonhomogeneous elasticity. J. Mech. Phys. Solid **10**, 335–342 (1962)

R. Hill, Elastic properties of reinforced solids: some theoretical principles. J. Mech. Phys. Solid **11**, 357–372 (1963)

M. Hori, S. Nemat-Nasser, Double-Inclusion model and overall moduli of multi-phase composites. Mech. Mater. **14**, 189–206 (1993)

U. Hornung (ed.), *Homogenization and Porous Media* (Springer, Berlin, 1996)

M. Kachanov, I. Tsukrov, B. Shafiro, Effective modulus of solids with cavities of various shapes. Appl. Mech. Rev. **47**, 151–174 (1994)

M. Kawashita, H. Nozaki, Eshelby tensor of a polygonal inclusion and its special properties. J. Elast. **74**(2), 71–84 (2001)

J. Kevorkina, J.D. Cole, *Multiple Scale and Singular Perturbation Methods* (Springer, Berlin, 1996)

H. Le Quang, Q.C. He, Q.S. Zheng, Some general properties of Eshelby's tensor fields in transport phenomena and anti-plane elasticity. Int. J. Solid Struct. **45**(13), 3845–3857 (2008)

L.P. Liu, Solutions to the Eshelby conjectures. Proc. R. Soc. A **464**, 573–594 (2008)

X. Markenscoff, Inclusions with constant eigenstress. J. Mech. Phys. Solid **46**(2), 2297–2301 (1998)

G.W. Milton, R. Kohn, Variational bounds on the effective moduli of anisotropic composites. J. Mech. Phys. Solid **43**, 63–125 (1988)

H.M.S. Munashinghe, M. Hori, Y. Enoki. Application of Hashin-Shtrikman Variational Principle for Computing Upper and Lower Approximate Solutions of Elasto-Plastic Problems, in *Proceedings of the International Conference on Urban Engineering in Asian Cities*, 1996, pp. 1–6

T. Mura, *Micromechanics of Defects in Solids* (Martinus Nijhoff Publisher, New York, 1987)

S. Nemat-Nasser, M. Hori, *Micromechanics: Overall Properties of Heterogeneous Materials* (North-Holland, London, 1993)

S. Nemat-Nasser, M. Hori, Universal bounds for overall properties of linear and nonlinear heterogeneous solids. Trans. ASME **117**, 412–422 (1995)

H. Nozaki, M. Taya, Elastic fields in a polyhedral inclusion with uniform eigenstrains and related problems. ASME J. Appl. Mech. **68**, 441–452 (2001)

K.C. Nuna, J.B. Keller, Effective elasticity tensor of a periodic composite. J. Mech. Phys. Solid **32**, 259–280 (1984)

O.A. Oleinik, A.S. Shamaev, G.A. Yosifian, *Mathematical Problems in Elasticity and Homogenization* (North-Holland, New York, 1992)

S. Onaka, N. Kabayashi, M. Kato, Two-dimensional analysis on elastic strain energy due to a uniformly eigenstrained supercircular inclusion in an elastically anisotropic material. Mech. Mater. **34**, 117–125 (2002)

C.Q. Ru, Eshelby inclusion of arbitrary shape in an anisotropic plane or half-plane. Acta Mech. **160**, 219–234 (2003)

E. Sanchez-Palencia, *Non-homogeneous Media and Vibration Theory*. Lecture Note in Physics, No. 127 (Springer, Berlin, 1981)

K. Tanaka, T. Mori, Note on volume integrals of the elastic field around an ellipsoidal inclusion. J. Elast. **2**, 199–200 (1972)

K. Terada, T. Miura, N. Kikuchi, Digital image-based modeling applied to the homogenization analysis of composite materials. Comput. Mech. **20**, 188–202 (1996)

S. Torquato, Random heterogeneous media: microstructure and improved bounds on effective properties. Appl. Mech. Rev. **42**(2), 37–76 (1991)

K.P. Walker, A.D. Freed, E.H. Jordan, Microstress analysis of periodic composites. Compos. Eng. **1**, 29–40 (1991)

L.J. Walpole, On the overall elastic moduli of composite materials. J. Mech. Phys. Solid **17**, 235–251 (1969)

X. Wang, X.L. Gao, On the uniform stress state inside an inclusion of arbitrary shape in a three-phase composite. Z. Angew. Math. Phys. **62**, 1101–1116 (2011)

M.Z. Wang, B.X. Xu, The arithmetic mean theorem of Eshelby tensor for a rotational symmetrical inclusion. J. Elast. **77**, 12–23 (2005)

J.R. Willis, Bounds and self-consistent estimates for the overall properties of anisotropic composites. J. Mech. Phys. Solid **25**, 185–202 (1977)

Q.S. Zheng, D.X. Du, An explicit and universally applicable estimate for the effective properties of multiphase composites which accounts for inclusion distribution. J. Mech. Phys. Solid **49**, 2765–2788 (2001)

W.N. Zou, Q.C. He, M.J. Huang, Q.S. Zheng, Eshelby's problem of non-elliptical inclusions. J. Mech. Phys. Solid **58**, 346–372 (2010)

Microstructural Behavior and Fracture in Crystalline Materials: Overview

14

Pratheek Shanthraj and Mohammed A. Zikry

Contents

Abstract

A dislocation-density-based multiple-slip crystalline plasticity framework, which accounts for variant morphologies and orientation relationships (ORs) that are uniquely inherent to lath martensitic microstructures, and a dislocation-density

P. Shanthraj (✉)
Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC, USA

Department of Microstructure–Physics and Alloy Design, Max Planck Institut ür Eisenforschung, Düsseldorf, Germany
e-mail: p.shanthraj@mpie.de

M.A. Zikry
Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC, USA
e-mail: zikry@ncsu.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_8

grain-boundary (GB) interaction scheme, which is based on dislocation-density transmission and blockage at variant boundaries, are developed and used to predict stress accumulation or relaxation at the variant interfaces. A microstructural failure criterion, which is based on resolving these stresses on martensitic cleavage planes, and specialized finite-element (FE) methodologies using overlapping elements to represent evolving fracture surfaces are used for a detailed analysis of fracture nucleation and intergranular and transgranular crack growth in martensitic steels. The effects of block and packet boundaries are investigated, and the results indicate that the orientation of the cleavage planes in relation to the slip planes and the lath morphology are the dominant factors that characterize specific failure modes. The block and packet sizes along the lath long direction are the key microstructural features that affect toughening mechanisms, such as crack arrest and deflection, and these mechanisms can be used to control the nucleation and propagation of different failure modes.

Introduction

In this chapter, a recently developed dislocation-density crystalline plasticity formulation is coupled with a new fracture methodology to investigate large strain inelastic modes and associated ductile crack nucleation and evolution in crystalline materials. The methodology is applied to martensitic steels. Fracture behavior in crystalline materials is inherently complex due to the microstructural effects, which can affect behavior on scales that range from the nano to the macro. The overarching challenge is to identify dominant microstructural effects on behavior, such as failure initiation and evolution in ductile crystalline materials. Lath martensitic steels offer a unique system, since the microstructure at different scales can include different crystalline structures (b.c.c. and f.c.c.), dislocation-density evolution and interactions, variant orientations and distributions, grain morphologies, grain-boundary distributions and orientations, and dispersed particles and precipitates.

Lath martensitic steels have a myriad of military and civilian applications due to their high strength wear resistance and toughness. These properties are a result of the unique microstructure inherent to martensitic steels, which have been characterized extensively as lath, block, and packet substructures (Morito et al. 2003, 2006). Depending on the processing and chemical composition, martensitic steels can offer a variety of microstructures and properties (Takaki et al. 2001; Tsuji et al. 2004; Song et al. 2005). Superior combinations of mechanical properties have been achieved through microalloying elements and thermomechanical treatment followed by tempering and aging (Ayada et al. 1998; Barani et al. 2007; Kimura et al. 2008). Failure in martensitic steels and its relation to the microstructure has been experimentally studied, and various failure modes, such as the formation of intensely localized shear bands (Minaar and Zhou 1998) and transgranular and intergranular fracture (Krauss 1999; Inoue et al. 1970; Matsuda et al. 1972), evolve as a function of the interrelated effects of martensitic structure, the ORs, and strain rate.

The size of cleavage facets in transgranular fracture modes has been related to the packet size (Inoue et al. 1970; Matsuda et al. 1972) and the block size (Hughes et al. 2011). Refinement of the block and packet sizes can reduce the coherence interface length on the {110} and {112} slip planes, which improves strength by resisting dislocation motion, and on the {100} cleavage planes, which improves toughness by suppressing crack propagation modes (Morris 2011; Guo et al. 2004; Morris et al. 2011). While block and packet size refinement through intercritical heat treatments (Jin et al. 1975; Kim et al. 1998) and tempforming (Kimura et al. 2008) has been used to obtain high strength and toughness, it has generally been observed to result in high strength with a significant loss in ductility (Howe 2000; Tsuji et al. 2002). The relative roles of the block and packet boundaries in the strengthening and toughening mechanisms are unclear in these investigations, as the effect of processing on the refinement of both packet and block sizes is not considered, which can be significant as noted by Kawata et al. (2006) who have observed that block sizes can be increased by packet size refinement by changing the processing conditions. Using microbending experiments, Shibata et al. (2010) have shown the significant contribution of block boundaries relative to that of subblock boundaries to strengthening, which has been attributed to dislocation pileups at high-angle block boundaries. Ohmura et al. (2004) have observed dislocation absorption into the block boundary with no indication of pileups, and the observed hardening due to block size refinement has been attributed to the decoration of the boundary with carbides (Ohmura and Tsuzaki 2007).

All of these investigations clearly indicate that the morphology and crystallography of the blocks and packets have a significant influence on the strength and toughness of martensitic microstructures, through the complex interactions of the prior austenite grain boundaries, the packet and the block boundaries with the evolving dislocation microstructure, and propagating cracks. However, what is lacking is a systematic investigation of the relationship between the microstructure and the material behavior, which is not well established. The objective of the present work, therefore, is to develop an integrated framework that incorporates material descriptions, which are sensitive to dominant martensitic microstructural features, with specialized computational representations of evolving failure surfaces and microstructurally based failure criteria, to accurately model the initiation and evolution of failure in martensitic steels. A physically based dislocation-density GB interaction scheme that is representative of the resistance to dislocation-density transmission across block and packet boundaries is developed, and it is incorporated into a multiple-slip dislocation-density-based constitutive formulation. The formulation accounts for variant morphologies and ORs that are uniquely inherent to lath martensitic microstructures. The disadvantages of existing crack propagation methods are overcome through the use of a specialized FE methodology using overlapping elements to represent failure surfaces (Hansbo and Hansbo 2004), and a failure criterion based on the evolving orientation of the cleavage planes in different martensitic variants is developed. This framework is then used to perform large-scale FE simulations to characterize the dominant dislocation-density mechanisms for the localization of plastic strains and the initiation and propagation of failure in martensitic microstructures.

This chapter is organized as follows: the dislocation-density crystalline plasticity formulation, the derivation of the dislocation-density GB scheme, and the martensitic microstructure representation are presented in section "Dislocation-Density-Based Multiple Slip Formulation"; the numerical implementation of the failure surface representation and microstructural failure criterion, which is based on resolving stresses along fracture planes, is outlined in section "Computational Representation of Failure Surfaces and Microstructural Failure Criterion"; the results are presented and discussed in section "Results and Discussion"; and a summary of the results and conclusions are given in section "Conclusion."

Dislocation-Density-Based Multiple Slip Formulation

In this section, the multiple-slip crystal plasticity rate-dependent constitutive formulation and the derivation of the evolution equations for the mobile and immobile dislocation densities, which are coupled to the constitutive formulation, are presented.

Multiple-Slip Crystal Plasticity Formulation

The crystal plasticity constitutive framework used in this study is based on the formulation developed by Asaro and Rice (1977) and Zikry (1994). It is assumed that the velocity gradient is decomposed into a symmetric deformation rate tensor D_{ij} and an antisymmetric spin tensor W_{ij}. D_{ij} and W_{ij} are then additively decomposed into elastic and inelastic components as

$$D_{ij} = D^*_{ij} + D^p_{ij}, W_{ij} = W^*_{ij} + W^p_{ij} \tag{1}$$

The inelastic components are defined in terms of the crystallographic slip rates as

$$D^p_{ij} = \sum P^{(\alpha)}_{ij} \dot{\gamma}^{(\alpha)}, W^p_{ij} = \sum \omega^{(\alpha)}_{ij} \dot{\gamma}^{(\alpha)} \tag{2}$$

where α is summed over all slip systems and $P^{(\alpha)}_{ij}$ and $\omega^{(\alpha)}_{ij}$ are the symmetric and antisymmetric parts of the Schmid tensor in the current configuration, respectively.

A power law relation can characterize the rate-dependent constitutive description on each slip system as

$$\dot{\gamma}^{(\alpha)} = \dot{\gamma}^{(\alpha)}_{\text{ref}} \left[\frac{\tau^{(\alpha)}}{\tau^{(\alpha)}_{\text{ref}}}\right] \left[\frac{\tau^{(\alpha)}}{\tau^{(\alpha)}_{\text{ref}}}\right]^{\frac{1}{m}-1} \tag{3}$$

where $\dot{\gamma}^{(\alpha)}_{\text{ref}}$ is the reference shear strain rate which corresponds to a reference shear stress $\tau^{(\alpha)}_{\text{ref}}$ and m is the rate sensitivity parameter. $\tau^{(\alpha)}$ is the resolved shear stress on

slip system α. The reference stress used is a modification of widely used classical forms (Franciosi et al. 1980) that relate reference stress to immobile dislocation density $\rho_{\text{im}}^{(\alpha)}$ as

$$\tau_{\text{ref}}^{(\alpha)} = \left(\tau_y^{(\alpha)} + G \sum_{\beta=1}^{\text{nss}} b^{(\beta)} \sqrt{a_{\alpha\beta} \rho_{\text{im}}^{(\beta)}} \right) \left(\frac{T}{T_0} \right)^{-\xi} \tag{4}$$

where $\tau_y^{(\alpha)}$ is the static yield stress on slip system α, G is the shear modulus, nss is the number of slip systems, $b^{(\beta)}$ is the magnitude of the Burgers vector, and $a_{\alpha\beta}$ are Taylor coefficients which are related to the strength of interactions between slip systems (Devincre et al. 2008; Kubin et al. 2008a, b). T is the temperature, T_0 is the reference temperature, and ξ is the thermal softening exponent.

Mobile and Immobile Dislocation-Density Evolution Equations

Following the approach of Zikry and Kao (1996), it is assumed that, for a given deformed state of the material, the total dislocation density, $\rho^{(\alpha)}$, can be additively decomposed into a mobile and an immobile dislocation density, $\rho_m^{(\alpha)}$ and $\rho_{\text{im}}^{(\alpha)}$, respectively. During an increment of strain on a slip system, a mobile dislocation-density rate is generated and an immobile dislocation-density rate is annihilated. Furthermore, the mobile and immobile dislocation-density rates can be coupled through the formation and destruction of junctions as the stored immobile dislocations act as obstacles for evolving mobile dislocations. This is the basis for taking the evolution of mobile and immobile dislocation densities as

$$\frac{d\rho_m^{(\alpha)}}{dt} = \left| \dot{\gamma}^{(\alpha)} \right| \left(\frac{g_{\text{sour}}^{\alpha}}{b^{(\alpha)2}} \frac{\rho_{\text{im}}^{(\alpha)}}{\rho_m^{(\alpha)}} - g_{\text{mnter}-}^{\alpha} \rho_m^{(\alpha)} - \frac{g_{\text{immob}-}^{\alpha}}{b^{(\alpha)}} \sqrt{\rho_{\text{im}}^{(\alpha)}} \right) \tag{5}$$

$$\frac{d\rho_{\text{im}}^{(\alpha)}}{dt} = \left| \dot{\gamma}^{(\alpha)} \right| \left(g_{\text{mnter}+}^{\alpha} \rho_m^{(\alpha)} + \frac{g_{\text{immob}+}^{\alpha}}{b^{(\alpha)}} \sqrt{\rho_{\text{im}}^{(\alpha)}} - g_{\text{recov}}^{\alpha} \rho_{\text{im}}^{(\alpha)} \right) \tag{6}$$

where g_{sour} is a coefficient pertaining to an increase in the mobile dislocation density due to dislocation sources; g_{mnter} are coefficients related to the trapping of mobile dislocations due to forest intersections, cross-slip around obstacles, or dislocation interactions; g_{recov} is a coefficient related to the rearrangement and annihilation of immobile dislocations; and g_{immob} are coefficients related to the immobilization of mobile dislocations.

Determination of Dislocation-Density Evolution Coefficients

To couple the evolution equations for mobile and immobile dislocation densities to the crystal plasticity formulation, the nondimensional coefficients in Eqs. 5 and 6

Table 1 g coefficients in Eqs. 5 and 6

g coefficient	Expression
g^{α}_{sour}	$b^{\alpha}\varphi\sum_{\beta}\sqrt{\rho^{\beta}_{\text{im}}}$
$g^{\alpha}_{\text{mnter}}$	$l_c f_0 \sum_{\beta}\sqrt{a_{\alpha\beta}}\left[\frac{\rho^{\beta}_{m}}{\rho^{\alpha}_{m}b^{\alpha}}+\frac{\dot{\gamma}^{\beta}}{\dot{\gamma}^{\alpha}b^{\beta}}\right]$
$g^{\alpha}_{\text{immob}-}$	$\frac{l_c f_0}{\sqrt{\rho^{\alpha}_{\text{im}}}}\sum_{\beta}\sqrt{a_{\alpha\beta}}\rho^{\beta}_{\text{im}}$
$g^{\alpha}_{\text{mnter}+}$	$\frac{l_c f_0}{\dot{\gamma}^{\alpha}\rho^{\alpha}_{\text{im}}}\sum_{\beta,\gamma} n^{\beta\gamma}_{\alpha}\sqrt{a_{\beta\gamma}}\left[\frac{\rho^{\gamma}_{m}\dot{\gamma}^{\beta}}{b^{\beta}}+\frac{\rho^{\beta}_{m}\dot{\gamma}^{\gamma}}{b^{\gamma}}\right]$
$g^{\alpha}_{\text{immob}+}$	$\frac{l_c f_0}{\dot{\gamma}^{\alpha}\sqrt{\rho^{\alpha}_{\text{im}}}}\sum_{\beta,\gamma} n^{\beta\gamma}_{\alpha}\sqrt{a_{\beta\gamma}}\rho^{\gamma}_{\text{im}}\dot{\gamma}^{\beta}$
$g^{\alpha}_{\text{recov}}$	$\frac{l_c f_0}{\dot{\gamma}^{\alpha}}\left(\sum_{\beta}\sqrt{a_{\alpha\beta}}\frac{\dot{\gamma}^{\beta}}{b^{\beta}}\right)e^{\left(\frac{-H_0\left(1-\sqrt{\frac{\rho^{\alpha}_{\text{im}}}{\rho_s}}\right)}{kT}\right)}$

were determined as functions of the crystallography and deformation mode of the material, by considering the generation, interaction, and recovery of dislocation densities in Shanthraj and Zikry (2011). These expressions are summarized in Table 1, where f_0 and ϕ are geometric parameters. H_0 is the reference activation enthalpy, ρ_s is the saturation density and the average junction length, l_c, can be approximated as

$$l_c = \frac{1}{\sum_{\beta}\sqrt{\rho^{(\beta)}_{\text{im}}}} \tag{7}$$

The interaction tensor, $n^{\beta\gamma}_{\alpha}$, is defined as having a value of 1 if dislocations on slip systems β and γ interact to form an energetically favorable junction on slip system α and a value of 0 if there are no interactions.

The Taylor interaction coefficients, $a_{\alpha\beta}$, for the slip-system interactions in BCC crystalline materials, which are required to determine the reference shear stress (Eq. 4) and the evolution of mobile and immobile dislocation densities (Eqs. 5 and 6), have been calculated in Shanthraj and Zikry (2012a, b) and listed in Table 2.

Dislocation-Density GB Interaction Scheme

In this section, a dislocation-density GB interaction scheme is presented. It is assumed that the dislocation-density interactions occur between slip systems on each side of the GB. Following Ma et al. (2006), the dislocation-density transmission

Table 2 Interaction coefficient values for the types of reactions between slip systems in b.c.c. crystals and comparison with values from literature (Madec and Kubin 2008; Queyreau et al. 2009)

Interaction type	Dissipation $\left(\propto \sqrt{a_{ij}}\right)$	a_{ij}	a_{ij} from literature
Self, colinear	1.5 kGb^2	0.6	0.550.72
Binary junction	0.5 kGb^2	0.067	0.0450.09
Ternary junction	kGb^2	0.267	0.12250.3364

is modeled as an activational event, and the constitutive relation (Eq. 3) has been modified at the GB through the introduction of a transmission factor $P^{(\alpha)}$:

$$\dot{\gamma}^{(\alpha)} = \dot{\gamma}_{\text{ref}}^{(\alpha)} \left[\frac{\left|\tau^{(\alpha)}\right|}{\tau_{\text{ref}}^{(\alpha)}}\right] \left[\frac{\tau^{(\alpha)}}{\tau_{\text{ref}}^{(\alpha)}}\right]^{\frac{1}{m}-1} P^{(\alpha)} = \dot{\gamma}_{\text{ref}}^{(\alpha)} \left[\frac{\left|\tau^{(\alpha)}\right|}{\tau_{\text{ref}}^{(\alpha)}}\right] \left[\frac{\tau^{(\alpha)}}{\tau_{\text{ref}}^{(\alpha)}}\right]^{\frac{1}{m}-1} e^{\left(\frac{-U_{\text{GB}}^{(\alpha)}}{kT}\right)} \tag{8}$$

where the line tension model for the activation of a Frank–Read source in the presence of a GB developed in de Koning et al. (2002) is used to obtain the energy required for dislocation-density transmission across a GB, $U_{\text{GB}}^{(\alpha)}$. The energy of such a dislocation configuration (Fig. 1), for incoming and outgoing slip systems α and β, is given by

$$\begin{aligned} U_{\text{GB}}^{(\alpha\beta)} &= 2\text{Gb}^{(\alpha)2} l_1 + 2\text{Gb}^{(\beta)2} l_2 + \text{Gb}^{(\alpha)2}(\Delta_1 - \Delta_2) + G\Delta b_{\text{eff}}^2 \Delta_2 \\ &\quad - \tau^{(\alpha)} b^{(\alpha)} A_{\text{sw},1} - \tau^{(\beta)} b^{(\beta)} A_{\text{sw},2} \end{aligned} \tag{9}$$

The magnitude of the effective residual Burger's vector, Δb_{eff}, is related to the residual Burger's vector, $\Delta\vec{b} = \vec{b}^{(\alpha)} - \vec{b}^{(\beta)}$, by

$$\left(\frac{\Delta b_{\text{eff}}}{\Delta b}\right)^2 = 1 + \left(\psi^2 - 2\psi\cos\upsilon + 1\right)^{\frac{1}{2}} \tag{10}$$

where

$$\psi = \frac{\Delta_2}{\Delta_1} = \cos\upsilon - \frac{\sin\upsilon}{\tan B}, \text{and } B = \cos^{-1}\left(\frac{\Delta b^2}{2b^{(\alpha)2}}\right) \tag{11}$$

For the critical configuration of a Frank–Read source, it is assumed that the geometric parameters are constant. The energy required to drive the system to the critical configuration, and thus initiate dislocation-density glide across a GB, which is taken here to be the energy required for dislocation-density transmission across a GB, for incoming and outgoing slip systems α and β, can then be simplified as

$$U_{\text{GB}}^{(\alpha\beta)} = c_1 \text{Gb}^3 \left[1 + c_2(1-\psi) + c_2\psi\left(\frac{\Delta b_{\text{eff}}}{b}\right)^2 - c_3\frac{\tau^{(\alpha)}}{\tau_{\text{ref}}^{(\alpha)}} - c_4\frac{\tau^{(\beta)}}{\tau_{\text{ref}}^{(\beta)}}\right] \tag{12}$$

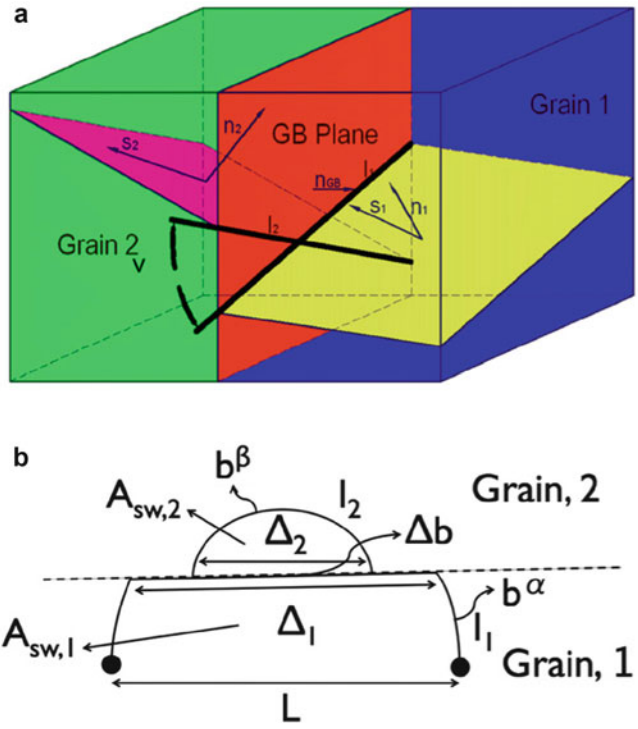


Fig. 1 (**a**) GB schematic and (**b**) dislocation configuration of a Frank–Read source in the vicinity of a GB viewed along the GB plane

where c_1c_4 are constants related to the geometric parameters at the critical configuration. They are given by $c_1 b = 2l_1 + l_2$ and $c_1c_2 b = \Delta_2$, and using the condition that for activated boundaries with no misorientation, $U_{\mathrm{GB}}^{(\alpha\beta)} = 0$, $c_3 = 0.5$ and $c_4 = 0.5$ is obtained. Dislocation-density transmission is assumed to be for the most energetically favorable outgoing slip system

$$U_{\mathrm{GB}}^{(\alpha)} = \min_{\beta} U_{\mathrm{GB}}^{(\alpha\beta)} \tag{13}$$

In the FE implementation, the modified constitutive relation (Eq. 8) is used for elements in the vicinity of the GB, which implies that the motion of dislocation densities within the element width, L_e, is constrained by the GB. However, only the motion of dislocation densities within the GB region, $L_{\mathrm{GB}} << L_e$, should be affected by the GB, and the dislocation-density motion in the FE implementation at the GB is overconstrained. To overcome this spurious constraint, the activation energy for dislocation-density transmission $U_{\mathrm{GB}}^{(\alpha)}$ is relaxed by a scaling factor L_{GB}/L_e, which is absorbed into the constant c_1. Through this formulation,

the criteria for dislocation-density transmission proposed by (Lee et al. 1990) have been incorporated into a physical model for dislocation-density GB interaction:

1. The misorientation between the slip planes must be minimum.
2. The magnitude of the residual Burger's vector at the GB must be minimum.
3. The shear stress on the outgoing slip system must be maximum.

Furthermore, the flux of dislocation densities across the GB is given by

$$\dot{\rho}_{\text{flux}}^{(\beta)} = \frac{\left|\dot{\gamma}^{(\alpha)}\right|}{\text{wb}^{(\alpha)}} \tag{14}$$

where β is the most favorable outgoing slip system and w is the boundary width, and the accumulation of residual GB dislocation densities is given by

$$\dot{\rho}_{\text{GB}}^{(\alpha)} = \sqrt{\frac{\Delta b}{b^{(\alpha)}}} \frac{\left|\dot{\gamma}^{(\alpha)}\right|}{\sqrt{\rho_m^{(\alpha)} b^{(\alpha)}}} \tag{15}$$

Martensitic Microstructural Representation

Following Hatem and Zikry (2009) the martensitic lath structure is related to the global coordinates through the parent austenite grain orientation and variant orientations. Commonly accepted ORs for lath martensitic steels are Kurdjumov–Sachs (KS) and Nishiyama–Wassermann (NW) ORs. KS ORs are based on a γ to α' martensitic transformation as $(111)_\gamma || (011)_{\alpha'}$, $\left[\overline{1}01\right]_\gamma || \left[\overline{1}\,\overline{1}\,1\right]_{\alpha'}$. The NW OR is a KS OR with a 5.12° rotation around the $[011]_\gamma$ direction. The 24 variants obtained from a KS OR are tabulated in Table 3.

To relate the martensitic local orientation to the global orientation, three transformations are needed. The first transformation, $[T]_1$, relates an observed OR to a theoretical OR, such as KS and NW ORs. The second transformation, $[T]_2$, relates a martensite OR to the parent austenite grain orientation. The third transformation, $[T]_3$, relates the austenite grain orientation to the global coordinates. These transformations are given by $[X]_{\text{Global}} = [T]_3[T]_2[T]_1[X]_{\alpha'}$.

To classify the martensitic microstructure, the characterization scheme of Morito et al. (2003) is followed. We designate a block as a group of laths with low-angle misorientation and a packet as a collection of blocks with the same habit plane. The lath long directions are oriented along $[011]_\gamma$ as shown in Fig. 2. Blocks will have the largest dimension aligned with the constituent lath long direction. Using this methodology, the ORs and original austenite orientations are used to model different variant orientations for different block and packet arrangements.

Table 3 The 24 variants corresponding to the KS OR

Variant no.	Parallel planes	Parallel directions	Variant no.	Parallel planes	Parallel directions
1	$(111)_\gamma \| \| (011)_{\alpha'}$	$[\bar{1}01]_\gamma \| \| [\bar{1}\,\bar{1}\,1]_{\alpha'}$	13	$(\bar{1}11)_\gamma \| \| (011)_{\alpha'}$	$[0\bar{1}\,1]_\gamma \| \| [\bar{1}\,\bar{1}\,1]_{\alpha'}$
2		$[\bar{1}01]_\gamma \| \| [\bar{1}1\bar{1}]_{\alpha'}$	14		$[0\bar{1}\,1]_\gamma \| \| [\bar{1}1\bar{1}]_{\alpha'}$
3		$[0\bar{1}1]_\gamma \| \| [\bar{1}\,\bar{1}\,1]_{\alpha'}$	15		$[\bar{1}0\bar{1}]_\gamma \| \| [\bar{1}\,\bar{1}\,1]_{\alpha'}$
4		$[0\bar{1}1]_\gamma \| \| [\bar{1}1\bar{1}]_{\alpha'}$	16		$[\bar{1}0\bar{1}]_\gamma \| \| [\bar{1}1\bar{1}]_{\alpha'}$
5		$[1\bar{1}0]_\gamma \| \| [\bar{1}\,\bar{1}\,1]_{\alpha'}$	17		$[110]_\gamma \| \| [\bar{1}\,\bar{1}\,1]_{\alpha'}$
6		$[1\bar{1}0]_\gamma \| \| [\bar{1}1\bar{1}]_{\alpha'}$	18		$[110]_\gamma \| \| [\bar{1}1\bar{1}]_{\alpha'}$
7	$(1\bar{1}\bar{1})_\gamma \| \| (011)_{\alpha'}$	$[10\bar{1}]_\gamma \| \| [\bar{1}\,\bar{1}\,1]_{\alpha'}$	19	$(11\bar{1})_\gamma \| \| (011)_{\alpha'}$	$[\bar{1}10]_\gamma \| \| [\bar{1}\,\bar{1}\,1]_{\alpha'}$
8		$[10\bar{1}]_\gamma \| \| [\bar{1}1\bar{1}]_{\alpha'}$	20		$[\bar{1}10]_\gamma \| \| [\bar{1}1\bar{1}]_{\alpha'}$
9		$[\bar{1}\,\bar{1}0]_\gamma \| \| [\bar{1}\,\bar{1}\,1]_{\alpha'}$	21		$[0\bar{1}\,\bar{1}]_\gamma \| \| [\bar{1}\,\bar{1}\,1]_{\alpha'}$
10		$[\bar{1}\,\bar{1}0]_\gamma \| \| [\bar{1}1\bar{1}]_{\alpha'}$	22		$[0\bar{1}\,\bar{1}]_\gamma \| \| [\bar{1}1\bar{1}]_{\alpha'}$
11		$[011]_\gamma \| \| [\bar{1}\,\bar{1}\,1]_{\alpha'}$	23		$[101]_\gamma \| \| [\bar{1}\,\bar{1}\,1]_{\alpha'}$
12		$[011]_\gamma \| \| [\bar{1}1\bar{1}]_{\alpha'}$	24		$[101]_\gamma \| \| [\bar{1}1\bar{1}]_{\alpha'}$

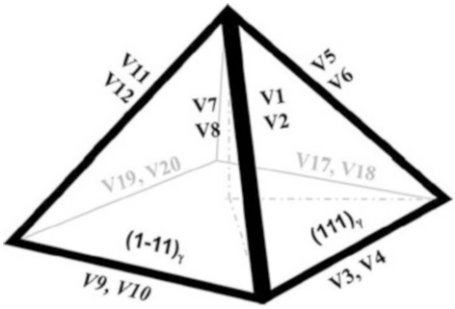


Fig. 2 The alignment of lath long directions of variants V1-24 in Table 3

Computational Representation of Failure Surfaces and Microstructural Failure Criterion

In this section, a method for the representation of the initiation and evolution of failure surfaces using overlapping elements is presented. Following the work by Hansbo and Hansbo (2004), an element e in an FE mesh with area A_0 is considered, which is crossed by a crack, dividing the element domain into two subdomains: elements e_1 and e_2 having areas A_{e1} and A_{e2}, respectively (Fig. 3). Adding an overlapping element on top of the existing element represents the displacement discontinuity due to the crack surface. The connectivity of the overlapping element is defined such that the two elements do not share nodes and therefore have independent displacement fields. This method can address some of the shortcomings associated with the application of cohesive fracture and extended finite-elements (XFEM) techniques to ductile fracture. The proposed fracture criteria do not assume a priori a fracture criteria through enrichment functions (XFEM) or on assumed fracture energies curve for unloaded surface (cohesive fracture). The proposed fracture approach resolves and nucleates cracks on cleavage planes based on the evolving inelastic microstructural behavior due to effects, such as dislocation-density evolution and interaction and different variant distributions. The approach is general in that it can be applied to a broad class of elements.

For a 4-node quadrilateral element with reduced integration and hourglass control, the internal nodal force vector of the cracked element is given by (Song et al. 2006)

$$\vec{f}_e^{\,\text{int}} = \vec{f}_{e1}^{\,\text{int}} + \vec{f}_{e2}^{\,\text{int}} \tag{16}$$

where $\vec{f}_{e1}^{\,\text{int}}$ and $\vec{f}_{e2}^{\,\text{int}}$ are the internal nodal force vectors of the overlapping elements representing the cracked element, which is given by

$$\vec{f}_{(e1/e2)}^{\,\text{int}} = \frac{A_{(e1/e2)}}{A_0} \int \left[B^T \sigma_{(e1/e2)} \right] dA_e \tag{17}$$

To improve numerical performance, a penalty force, which decays with time, is added to resist large initial relative displacements between overlapping

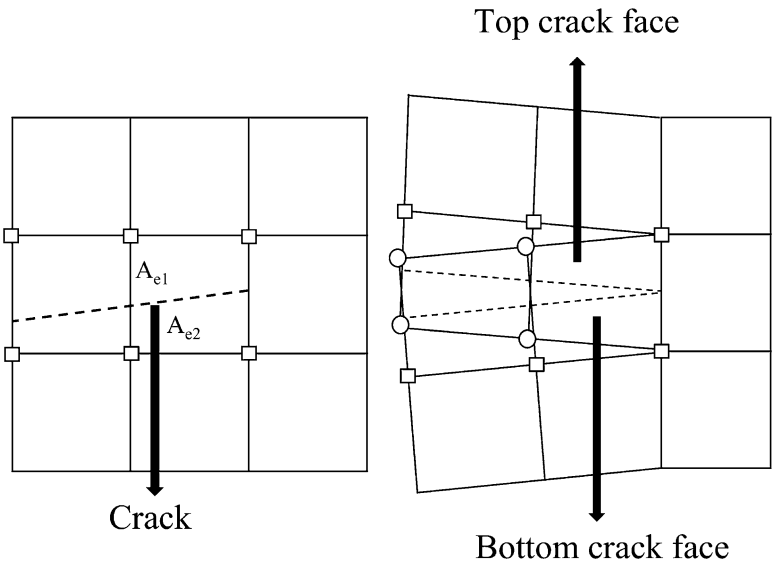


Fig. 3 Representation of a crack using overlapping elements

elements due to the sudden loss in stiffness of a cracked element. The penalty force is given by

$$\vec{f}^{\,\text{penalty}}_{(e1/e2)} = \alpha(0.995)^{\text{nstep}}\left[\vec{u}_{e1} - \vec{u}_{e2}\right] \tag{18}$$

where α is a penalty parameter, which is taken to be of the order of the material stiffness. A simulation time-step size is chosen, based on convergence studies, so that the penalty force decays at faster timescale than the crack opening displacement.

The inherent fracture mode in martensitic steel is cleavage on $\{100\}_{\alpha'}$ planes in the microstructure (Guo et al. 2004). To formulate this into a microstructural failure criterion, the orientation of the cleavage planes for each variant in the global coordinate system is obtained by applying the series of transformations outlined in section "Martensitic Microstructural Representation":

$$n_{\text{cleave}} = [T]_3[T]_2[T]_1 n_{\text{cleave},\alpha'} \tag{19}$$

The global orientation of the cleavage planes in the current configuration is then obtained by updating at every time-step due to the lattice rotations as $\dot{n}_{\text{cleave}} = W^* n_{\text{cleave}}$. The normal component of the traction acting on each cleavage plane has a direct influence on fracture along that plane. The maximum, over all the $\{100\}_{\alpha'}$ cleavage planes, of the normal component of the traction on these planes is, therefore, used as the failure criterion for mode I fracture loading conditions

$$t_{\text{cleave}} = \max_{\{100\}_{\alpha'}\text{planes}} \left\langle n_{\text{cleave}}^{T}[\sigma] n_{\text{cleave}} \right\rangle \tag{20}$$

A crack is assumed to nucleate when $t_{\text{cleave}} > \sigma_{\text{frac}}$, and the crack is orientated along the most favorable cleavage plane. The 3D cleavage model is implemented in a 2D setting by projecting the 3D crack path onto the 2D plane.

The total deformation rate tensor, D_{ij}, and the plastic deformation rate tensor, D_{ij}^{p}, are needed to update the material stress state. The method used here is the one developed by Zikry (1994) for rate-dependent crystalline plasticity formulations.

Results and Discussion

The multiple-slip dislocation-density-based crystal plasticity formulation, the dislocation-density GB interaction scheme, and the representation of cracks using overlapping elements were used to investigate the microstructural failure behavior of martensitic steel.

Martensitic Block Size

The microstructure morphology associated with a single parent austenite grain of dimensions 3×6 mm^2 in low- and high-carbon martensitic steels on the $(100)_{\gamma}$ plane is shown in Fig. 4, where the block morphology is obtained by projecting the constituent lath variant long direction (Fig. 2) onto the plane. For the low-carbon steel microstructure, 19 blocks are distributed within three packets with an average packet size of 250 μm (Fig. 4a). For the high-carbon steel microstructure, due to the packet and block size refinement (Maki et al. 1980), for the same parent austenite grain, 48 blocks are distributed within eight packets resulting in an average packet size of 100 μm (Fig. 4b). The parent austenite grain was assumed to have a cube orientation, and the KS relation was adopted as the martensite OR with $\{111\}_{\gamma}$ as the habit plane. A convergent plane strain FE mesh with approximately 9,000 elements was subjected to tensile loading along the (001) direction at nominal strain rates of 10^{-4} s^{-1}, 500 s^{-1}, and 2,500 s^{-1} with symmetry boundary conditions applied on the left and bottom edges. The material properties (Table 4) that are used for the constituent crystals are representative of low nickel alloy steel (Hatem and Zikry 2009).

Low-Carbon Steel

The normalized mobile dislocation densities corresponding to the two most active slip systems at a nominal strain of 9.1 %, which is just before the onset of crack nucleation, are shown in Fig. 5a, b. The maximum normalized mobile dislocation densities are 0.47 for slip system $\left(\bar{1}01\right)\left[1\bar{1}1\right]$ and 0.49 for slip system $\left(101\right)\left[11\bar{1}\right]$. The normalized interaction density, which is the increase in immobile dislocation density due to junction formation relative to the decrease of mobile dislocation

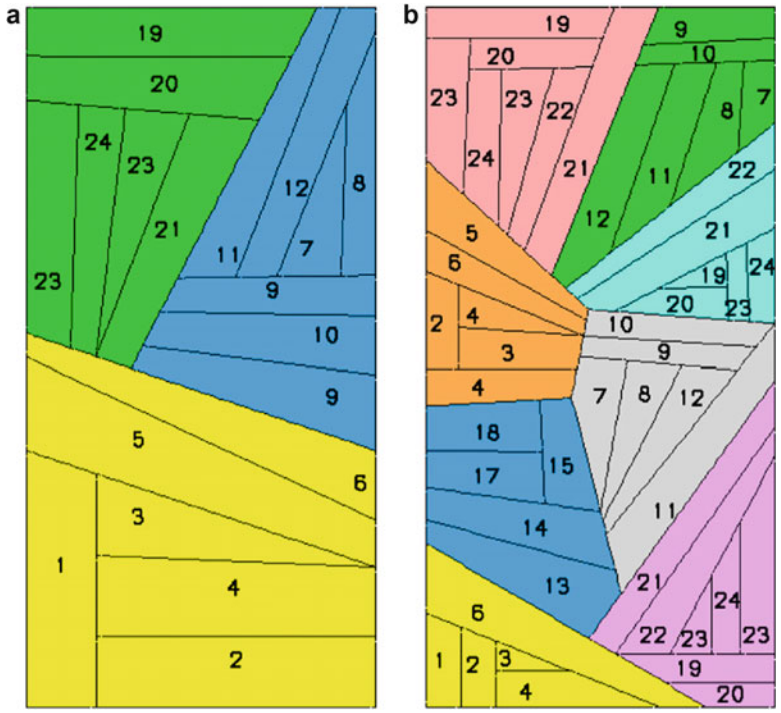


Fig. 4 Block and packet distribution in (**a**) low-carbon steel with an average packet size of 250 μm and (**b**) high-carbon steel with an average packet size of 100 μm. Packets are represented by distinct colors, and numbers in blocks indicate the constituent lath variant (Table 3)

Table 4 Material properties

Properties	Value
Young's modulus, E	228 GPa
Static yield stress, τ_y	517 MPa
Poisson's ratio, ν	0.3
Rate sensitivity parameter, m	0.01
Reference strain rate, $\dot{\gamma}_{\text{ref}}$	0.001 s^{-1}
Burger's vector, b	3.0×10^{-10} m
Fracture stress, σ_{frac}	$7\tau_y$
c_1	0.15
c_2	0.9

density, is shown in Fig. 5c. Negative values, with a minimum of −0.35, indicate that the annihilation of dislocation junctions through self and colinear dislocation interactions is dominant. Positive values, with a maximum of 0.5, indicate that the formation of dislocation junctions through binary and ternary dislocation interactions is dominant (Table 2). The dominant interaction type is determined by the

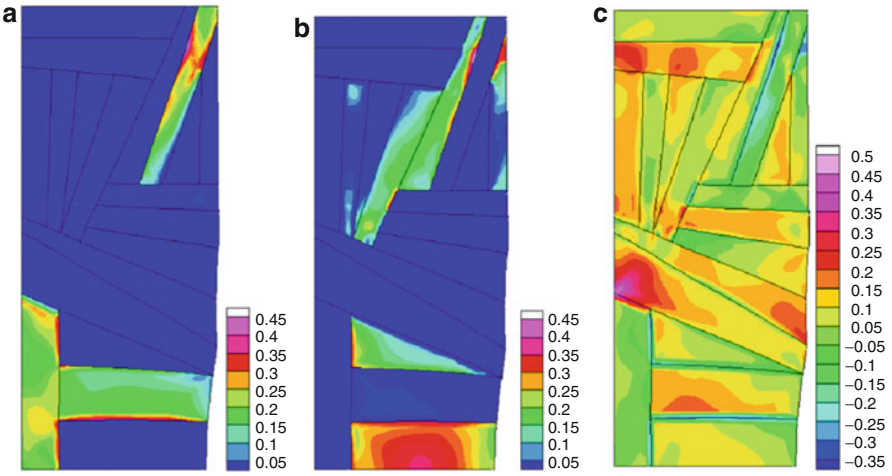


Fig. 5 Normalized mobile dislocation densities at a nominal strain of 9.1 % on (**a**) slip system $(\bar{1}01)[1\bar{1}1]$, (**b**) slip system $(101)[11\bar{1}]$, and (**c**) normalized interaction density at a strain rate of $10^{-4}\ s^{-1}$

active slip systems in each variant and using the interaction tensor (section "Determination of Dislocation-Density Evolution Coefficients").

The accumulated plastic slip at a nominal strain of 9.1 % is shown in Fig. 6a. The maximum accumulated slip is 0.14. The loading is aligned along the $[001]_\gamma$ direction, which results in a maximum resolved shear stress along the $[011]_\gamma$ directions. The $[011]_\gamma$ directions are also parallel to the long direction of the laths and blocks and to the slip direction $[111]_{\alpha'}$ based on the KS OR. This configuration, which aligns the slip systems with the maximum resolved shear stress, along with local material softening mechanisms due to the annihilation of dislocations in the blocks corresponding to a negative interaction density (Fig. 5c), results in the localization of shear strain. Mobile dislocation densities are transmitted across the block boundaries between the active $(\bar{1}01)[1\bar{1}1]$ and $(101)[11\bar{1}]$ slip systems through the compatibility of the slip systems, which is associated with a low activation energy (Eqs. 12 and 13), and result in shear pipes for the formation of shear bands (Hatem and Zikry 2009). While no special coherency exists across packet boundaries, slip transmission can be observed, and therefore, these packet boundaries can behave similarly to block boundaries (Shanthraj and Zikry 2012a, b). The accumulation of plastic slip is observed along high-angle boundaries as a result of dislocation-density blockage due to slip-system incompatibility, which is exacerbated by lattice rotations (Fig. 6b), and is also observed experimentally (Morito et al. 2003).

The total normalized GB dislocation density due to all active slip systems at a nominal strain of 9.1 % is shown in Fig. 6c. The normalized GB dislocation density attains a maximum value of 0.8, which is due to the pileup of dislocation densities at the block and packet boundaries. This happens along block and packet boundaries

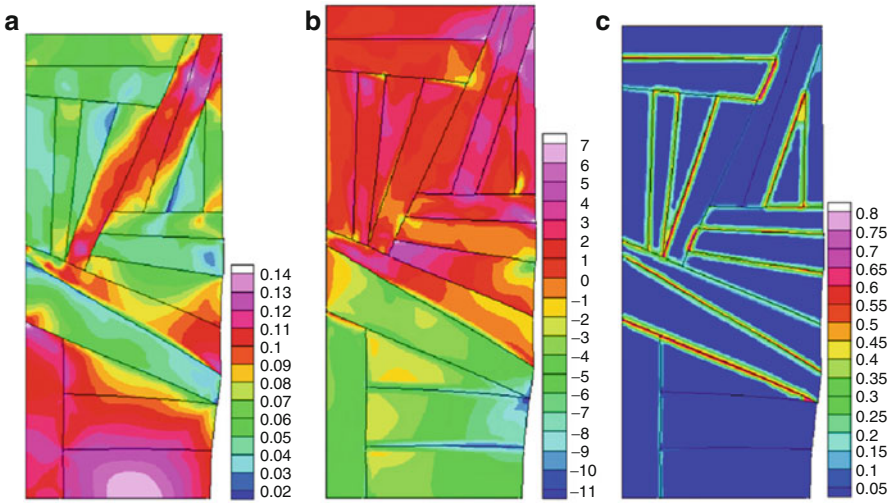


Fig. 6 (**a**) Plastic slip, (**b**) lattice rotations, and (**b**) GB dislocation density at 9.1 % nominal strain indicating dislocation-density pileup and plastic slip accumulation at a strain rate of 10^{-4} s^{-1}

between variant pairs having large incompatibilities between the active slip systems and results in large local stress concentrations along high-angle block and packet boundaries, with a maximum normalized stress (by static yield stress) of 8.0, as shown in Fig. 7a. The nucleation of a crack occurs at 9.2 % nominal strain at a triple junction between variants 11, 20, and 21 (Fig. 7b), which is separated by a packet boundary, as a result of the local stress concentrations (Fig. 7a) due to GB dislocation-density accumulation at the triple junction (Fig. 6c). The crack initially grows across the packet boundary (Fig. 7c), as dislocation-density transmission relaxes the stress concentrations along the packet boundary and the favorable orientation of the cleavage planes in contiguous blocks, which is accommodated by the lattice rotations (Fig. 6b). The neighboring block morphologies are such that the lath long directions are normal to the crack propagation path, which results in resistance to crack propagation, defection of the crack path, and debonding along the variant boundaries due to the GB dislocation-density accumulation and cleavage plane incompatibilities (Fig. 7d). These intergranular and transgranular fracture modes are consistent with experimental observations (Krauss 1999; Inoue et al. 1970; Matsuda et al. 1972), as well as crack path deflection, crack arrest, and debonding (Hughes et al. 2011).

High-Carbon Steel

To further elucidate the role of block and packet morphology in failure, the results are compared with the failure behavior of a refined variant distribution, corresponding to high-carbon steels (Fig. 4b). The normalized mobile dislocation densities corresponding to the two most active slip systems at a nominal strain of 5.9 %, which is just before the onset of crack nucleation, are shown in Fig. 8a, b.

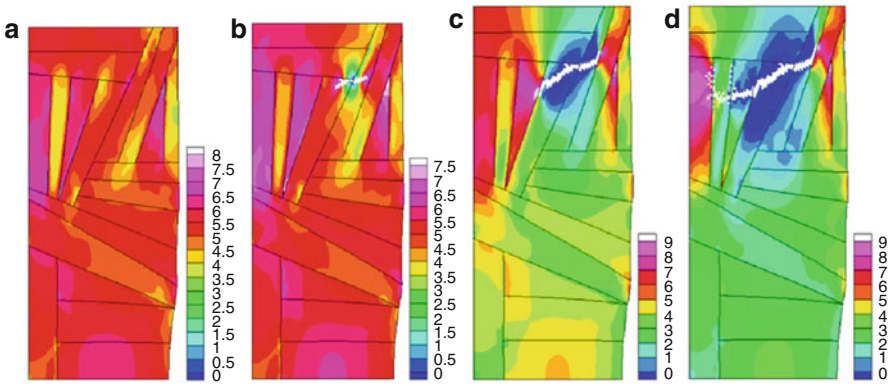


Fig. 7 Axial stress at (**a**) 9.1 % nominal strain, (**b**) 9.2 % nominal strain showing crack nucleation and (**c**) 9.3 % nominal strain, and (**d**) 9.6 % nominal strain showing crack propagation at a strain rate of $10^{-4}\ s^{-1}$

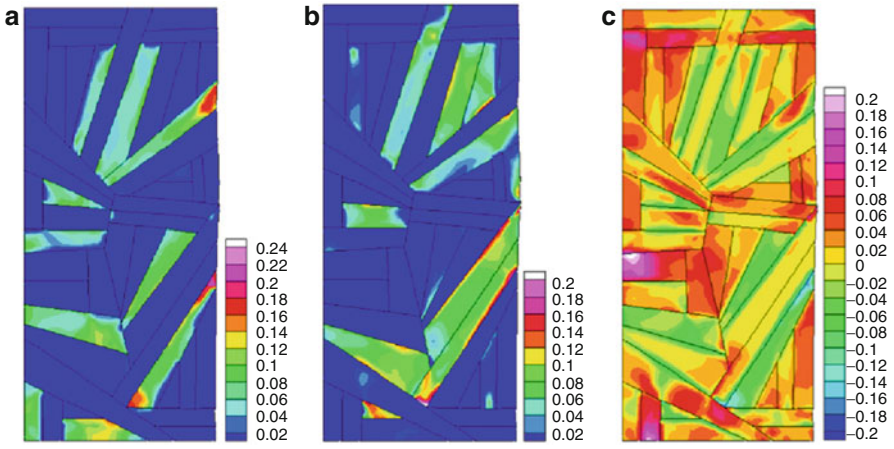


Fig. 8 Normalized mobile dislocation densities for a refined variant arrangement at a nominal strain of 5.9 % on (**a**) slip system $(\bar{1}01)[1\bar{1}1]$, (**b**) slip system $(101)[11\bar{1}]$, and (**c**) normalized interaction density at a strain rate of $10^{-4}\ s^{-1}$

The maximum normalized mobile dislocation densities are 0.24 for slip system $(\bar{1}01)[1\bar{1}1]$ and 0.20 for slip system $(101)[11\bar{1}]$. The mobile dislocation-density activity on these slip systems corresponds to a negative interaction density with a minimum value of −0.2 (Fig. 8c). The accumulated plastic slip, and the resulting lattice rotation at a nominal strain of 5.9 %, is shown in Fig. 9a, b. The maximum accumulated slip is 0.1, with lattice rotations ranging from ±8°. The localization of shear strain in relation to the lath orientation, interaction density, and lattice rotations (Fig. 9b) occurs as discussed in section "Martensitic Block Size." The plastic slip is constrained to flow along the lath long directions, and the low

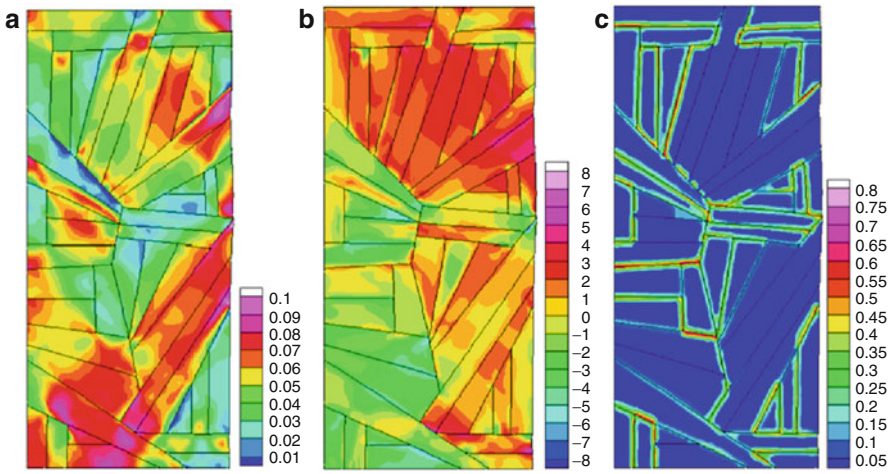


Fig. 9 (**a**) Plastic slip, (**b**) lattice rotations, and (**c**) GB dislocation density for a refined variant arrangement at 5.9 % nominal strain indicating dislocation-density pileup and plastic slip accumulation at a strain rate of 10^{-4} s^{-1}

activation energy associated with the transmission of dislocation densities between the active slip systems in neighboring blocks results in a dominant shear band, which is of comparable size to the coarser microstructure. The accumulation of shear slip is observed when there is a blockage of dislocation densities, with a maximum normalized GB dislocation density of 0.8 (Fig. 9c).

The refinement of the block and packet sizes increases the number of incompatible variant interfaces, which serve as sites for stress concentrations due to the accumulation of GB dislocation densities, with a maximum normalized stress of 6.5 (Fig. 10a). This results in the nucleation of a crack at a nominal strain of 6.1 %, which is lower than that corresponding to the coarse block and packet microstructure, at a triple junction between variants 9, 11, and 12 (Fig. 10b). Variants 11 and 12 belong to different Bain groups and thus have a large misorientation in {100} cleavage planes (Guo et al. 2004). This results in a resistance to the crack propagation across the variant boundary and forces the crack path to deflect along the lath long direction (Fig. 10c). On encountering the neighboring packet boundary, the crack is arrested as a result of the incompatibility in cleavage planes and a change in the neighboring block morphology. A new crack is then nucleated in the neighboring block ahead of the arrested crack tip, which is constrained to propagate along the neighboring block morphology (Fig. 10d).

In contrast to coarse blocks, where the crack propagates across the block width (Fig. 7b–d), block size refinement in high-carbon steels constrains crack propagation along the block morphology resulting in greater fracture resistance, with an increase in load-bearing capacity after crack nucleation of a nominal strain of 0.4 %. However, the stress–strain curve indicates that there is no increase in strength due to refinement (Fig. 11). Block size refinement is more effective in

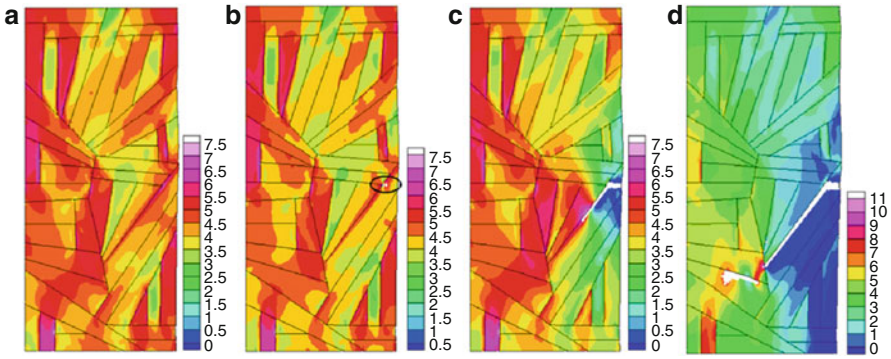


Fig. 10 Axial stress for a refined variant arrangement at (**a**) 5.9 % nominal strain, (**b**) 6.1 % nominal strain showing crack nucleation and (**c**) 6.4 % nominal strain, and (**d**) 6.75 % nominal strain showing crack propagation at a strain rate of 10^{-4} s^{-1}

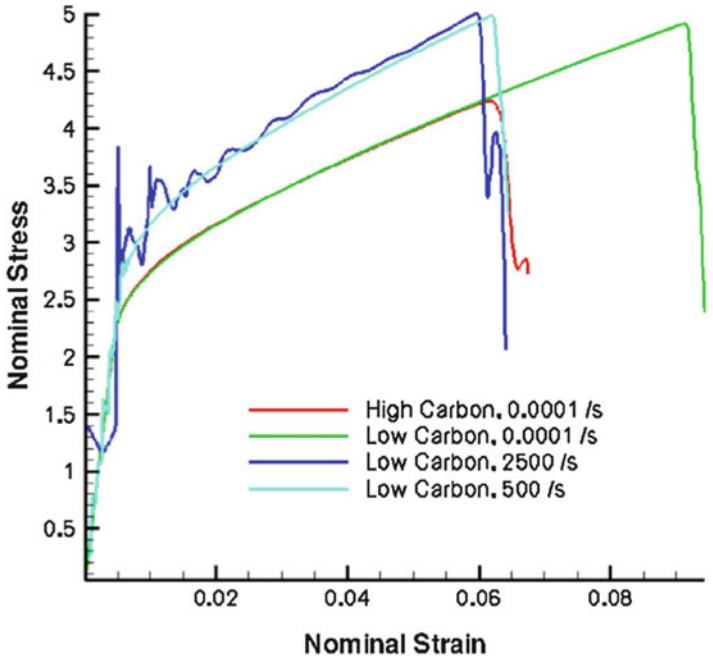


Fig. 11 Nominal stress–strain curves at strain rates of 10^{-4} s^{-1}, 500 s^{-1}, and 2,500 s^{-1}

improving fracture resistance than strength as the variant boundaries offer relatively less resistance to slip transmission onto the 24 possible slip systems, on the {110} and {112} crystallographic planes, compared to the transmission of cracks onto only three possible {100} cleavage planes. This results in the formation of shear bands of comparable sizes and in both microstructures and similar strengths since

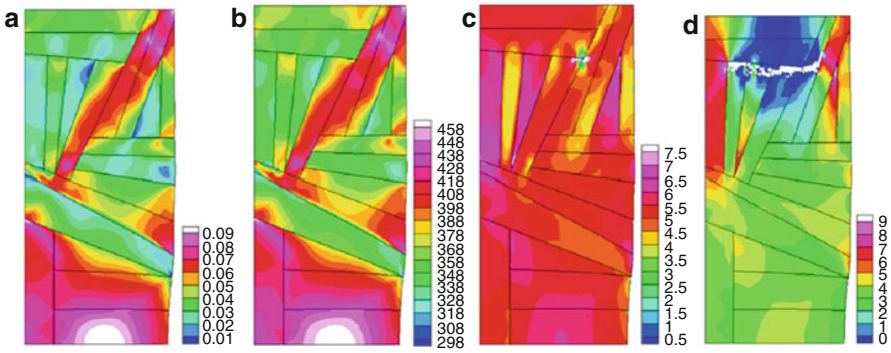


Fig. 12 (**a**) Plastic slip and (**b**) temperature at 6.0 % nominal strain. Normal stress at (**c**) 6.1 % nominal strain showing crack nucleation and (**d**) 6.4 % nominal strain at a strain rate of 2,500 s^{-1}

the microstructural strength is associated with the resistance to plastic slip. As the plastic slip is constrained to flow along the lath long directions, block refinement in this dimension is most effective in strengthening, while refinement of the block width is most effective in improving fracture resistance by constraining the crack to propagate along the lath long direction. However, refinement also increases the number of incompatible triple junctions, which can be sites for crack nucleation, and therefore also reduces the ductility. Therefore, the effects of both block and packet refinement should be considered in relation to the lath morphology in determining fracture behavior.

Dynamic Behavior

In this section, the dynamic fracture of low-carbon martensitic steels at loading rates of 500–2,500 s^{-1} is investigated. The nominal stress–strain curves over the range of loading conditions are shown in Fig. 11. The oscillations at high strain rates occur due to stress wave reflections along the free and fixed boundary, which is dampened due to plasticity. The lower failure strains of 6.2 % at a strain rate of 500 s^{-1} 6.0 % at a strain rate of 2,500 s^{-1} are a result of dynamic strain-rate hardening (Fig. 11). The accumulated plastic slip at a nominal strain of 5.9 % and strain rate of 2,500 s^{-1} is shown in Fig. 12a. The maximum accumulated slip is 0.09, and the shear strain localization is narrower at higher strain rates, which is a result of material and thermal softening mechanisms, with a maximum temperature of 458 K, as well as the dynamic strain rates, which prevent the accumulation of shear strains over wide regions. This is consistent with the experimental observations of Dodd and Bai (1985). At this strain rate, crack nucleation is at a nominal strain of 6.0 %, as shown in Fig. 12c. Crack deviation is observed at block boundaries with large incompatibilities in the cleavage planes (Fig. 12d). However, a lower resistance to the crack propagation path across block boundaries due to lower accumulation of GB dislocation densities, with a maximum value of 0.7, and stress concentrations along the variant boundary is observed, as a result of the strain-rate hardening at high strain rates. Block size refinement is less effective in

improving fracture resistance at high strain rates, and these high strain-rate characteristics result in fracture modes that are different from quasi-static strain rates.

Martensitic Block Distribution

Ductility in martensitic steels is directly related to the transmission of plastic slip across variant boundaries (Tsuji et al. 2008; Guo et al. 2004; Morito et al. 2006). The slip transmission factor (section "Dislocation-Density GB Interaction Scheme")

$$P^{(\alpha\beta)} = e^{\left(\frac{-U_{\mathrm{GB}}^{(\alpha\beta)}}{kT}\right)} \tag{21}$$

between the most active slip systems α and β across a variant boundary can therefore be used as a measure of the ductility. The activation energy, $U_{\mathrm{GB}}^{(\alpha\beta)}$, is given by Eq. 12. Taking τ/τ_{ref} as approximate unity for the most active slip systems, the transmission factor can then be reduced to a function of the orientation of the interaction slip planes, slip directions, and GB. The orientation of the slip plane and slip direction of the most active slip system in a variant are determined from the variant OR (Table 3) for a cube-oriented parent austenite grain. The transmission factor across the variant boundaries for the 24 KS variants is calculated using Eq. 12 for different GB orientations. The variants are grouped based on the transmission factors in Table 5. The boundaries between variants belonging to group I and II have large transmission factors (>0.26), which is desirable for ductility, while variant boundaries in group III and IV have low transmission factors (<0.01), which can result in slip blockage, stress accumulation, and crack nucleation. In addition, fracture resistance in martensitic steels is related to the misorientation between the cleavage planes across variant boundaries (section "Computational Representation of Failure Surfaces and Microstructural Failure Criterion"). The KS variants can be categorized into three Bain groups – A, B, and C, – as indicated in Table 5, where variants belonging to each Bain group have low misorientations in the {100} cleavage planes. Boundaries between variants belonging to different Bain groups can have large misorientations in the {100} cleavage planes (Guo et al. 2004), which can be desirable for fracture resistance as it can result in crack deflection and crack arrest. A microstructural distribution, which increases the boundaries between variants belonging to different Bain groups and having large slip transmission factors (Table 5), therefore optimizes ductility and fracture toughness in martensitic steels.

Based on these guidelines, the microstructure associated with a single parent austenite grain for a random and optimized variant distribution on the $(100)_{\gamma}$ plane is shown in Fig. 13, where 68 blocks are distributed within two packets and the block morphology is obtained by projecting the constituent lath variant long direction (Fig. 2) onto the plane. For the optimized variant distribution, only variant boundaries belonging to group I are used, and the use of boundaries between

Table 5 Slip transmission factor across variant boundaries indicating Bain groups A, B, and C

Variant group	Slip transmission factor
Group I: V1(C), V3(B), V4(C), V6(B), V13(C), V14(B), V16(C), V17(B)	0.26–1.0
Group II: V8(C), V9(B), V11(C), V12(B), V19(B), V21(C), V22(B), V24(C)	0.26–1.0
Group III: V2(A), V7(A), V5(A), V10(A)	0.00–0.01
Group IV: V15(A), V18(A), V20(A), V23(A)	0.00–0.01

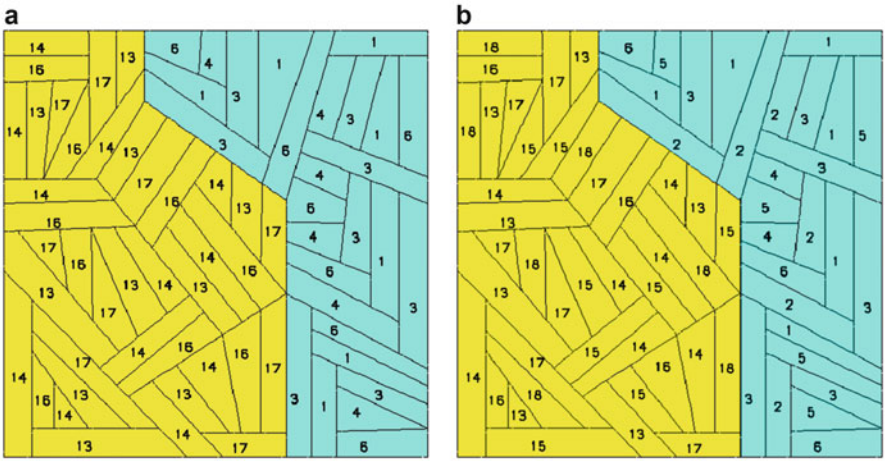


Fig. 13 Block and packet distribution in (**a**) random variant distribution and (**b**) optimized variant distribution. Packets are represented by distinct colors, and numbers in blocks indicate the constituent lath variant (Table 3)

variants belonging to the same Bain group is restricted to be along the less effective lath lateral direction (Fig. 13b). The variant orientation is represented as outlined in section "Martensitic Microstructural Representation." The parent austenite grain is assumed to have a cube orientation, and the KS relation is adopted as the martensite OR with $\{111\}_{\gamma}$ as the habit plane.

A convergent plane strain FE mesh with approximately 9,000 elements was subjected to tensile loading along the (001) direction at nominal strain rate of 10^{-4} s^{-1} and 5,000 s^{-1}. The material properties (Table 4) are used for the constituent crystals.

Random Variant Distribution

The normalized (by the saturation dislocation density) mobile dislocation densities corresponding to the two most active slip systems at a nominal strain of 4.8 %, which is just before the onset of crack nucleation, are shown in Fig. 14a, b.

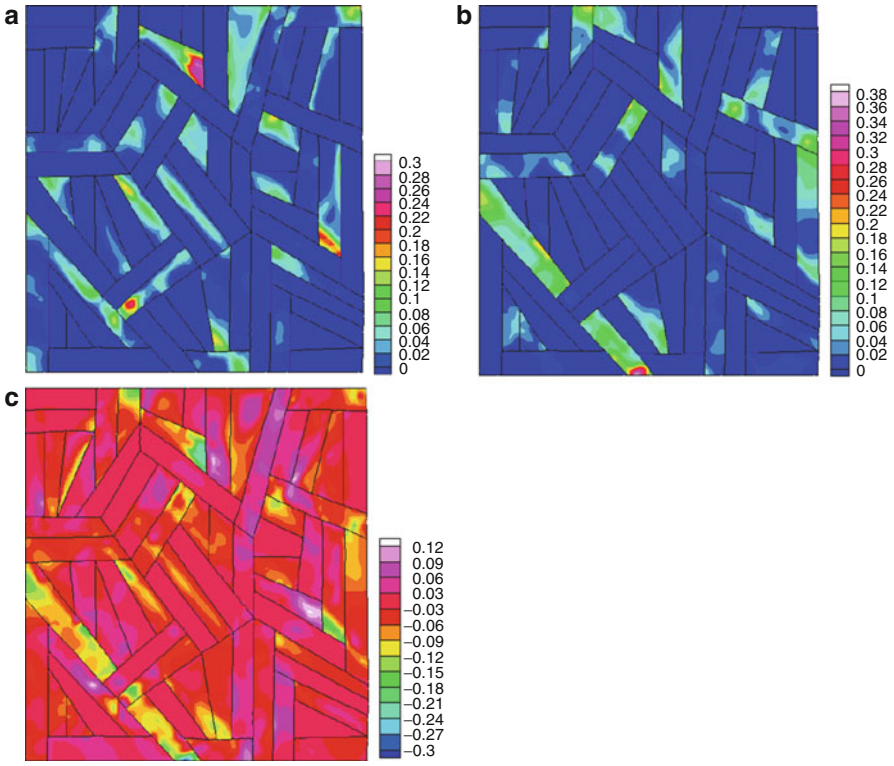


Fig. 14 Mobile dislocation densities, normalized by the saturation density, for a random variant distribution at a nominal strain of 4.8 % on (**a**) slip system $(\bar{1}12)[1\bar{1}1]$, (**b**) slip system $(112)[11\bar{1}]$, and (**c**) normalized interaction density at a strain rate of 10^{-4} s^{-1}

The maximum normalized mobile dislocation densities are 0.30 for slip system $(\bar{1}12)[1\bar{1}1]$ and 0.38 for slip system $(112)[11\bar{1}]$. The normalized interaction density

$$\rho_{\text{int}} = \sum_{\alpha} \int \dot{\gamma}^{\alpha} \left(g^{\alpha}_{\text{mnter}+} \rho^{\alpha}_{m} + \frac{g^{\alpha}_{\text{mnter}+}}{b} \sqrt{\rho^{\alpha}_{\text{im}}} - g^{\alpha}_{\text{mnter}-} \rho^{\alpha}_{m} - \frac{g^{\alpha}_{\text{mnter}-}}{b} \sqrt{\rho^{\alpha}_{\text{im}}} \right) dt \quad (22)$$

which is the increase in immobile dislocation density due to junction formation relative to the decrease of mobile dislocation density, is shown in Fig. 14c. Negative values, with a minimum of −0.35, indicate that the annihilation of dislocation junctions through self and colinear dislocation interactions is dominant. Positive values, with a maximum of 0.12, indicate that the formation of dislocation junctions through binary and ternary dislocation interactions is dominant (Table 2). The dominant interaction type is determined by the active slip systems in each variant and using the interaction tensor (section "Determination of Dislocation-Density Evolution Coefficients").

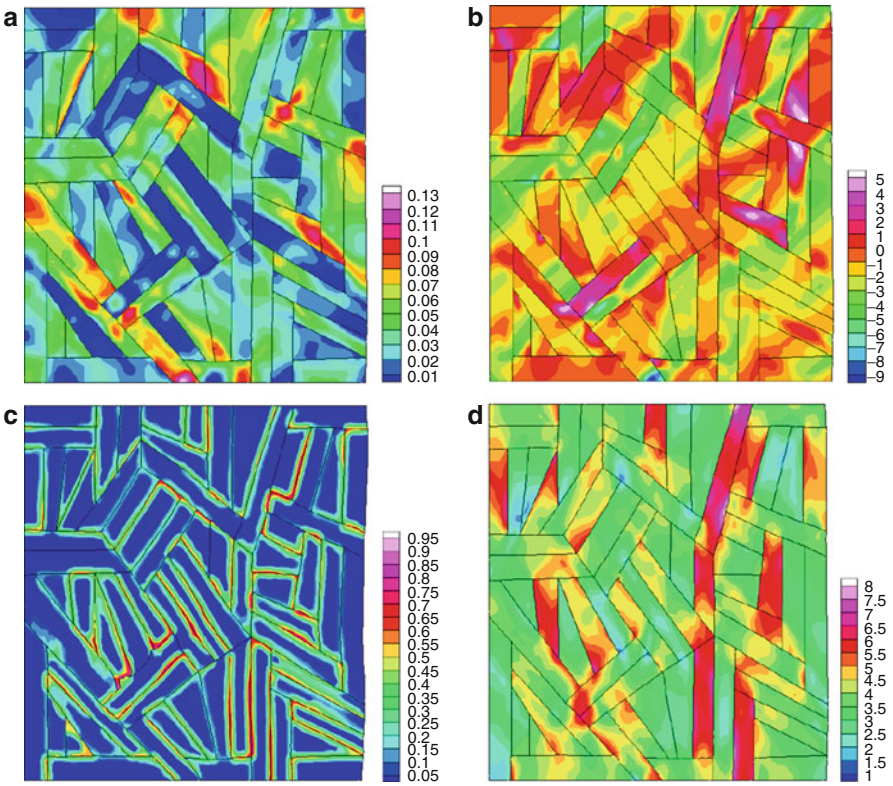


Fig. 15 (**a**) Plastic slip, (**b**) lattice rotations (in degrees), (**c**) GB dislocation density, normalized by the saturated density, and (**d**) normal stress, normalized by the yield stress for a random variant distribution at 4.8 % nominal strain indicating dislocation-density pileup and plastic slip accumulation at a strain rate of 10^{-4} s^{-1}

The accumulated plastic slip at a nominal strain of 4.8 % is shown in Fig. 15a. The maximum accumulated slip is 0.13. The loading is aligned along the $[001]_\gamma$ direction, which results in a maximum resolved shear stress along the $[011]_\gamma$ directions. The $[011]_\gamma$ directions are also parallel to the long direction of the laths and blocks and to the slip direction $[111]_{\alpha'}$ based on the KS OR. This configuration, which aligns the slip systems with the maximum resolved shear stress, along with local material softening mechanisms due to the annihilation of dislocations in the blocks corresponding to a negative interaction density (Fig. 14c), results in the localization of shear strain. Mobile dislocation densities are transmitted across the block boundaries between the active $\left(\bar{1}12\right)\left[1\bar{1}1\right]$ and $(112)\left[11\bar{1}\right]$ slip systems through the compatibility of the slip systems, which is associated with a low activation energy (Eq. 12), and result in shear pipes for the formation of shear bands (Hatem and Zikry 2009). The accumulation of plastic slip is observed along high-angle boundaries as a result of dislocation-density blockage due to slip-system

incompatibility, which is exacerbated by lattice rotations (Fig. 15b), and is also observed experimentally (Morito et al. 2003).

The total normalized GB dislocation density due to all active slip systems at a nominal strain of 4.8 % is shown in Fig. 15c. The normalized GB dislocation density attains a maximum value of 0.95, which is due to the pileup of dislocation densities at the block and packet boundaries. This occurs along block boundaries between variant pairs belonging to group III and IV in Table 5, which have large incompatibilities between the active slip systems and low slip transmission factors. This results in large local stress concentrations along high-angle block and packet boundaries, with a maximum normalized stress (by static yield stress) of 8.0, as shown in Fig. 15d. The nucleation of a crack occurs at 5.0 % nominal strain at a triple junction between variants 13, 17, and 18 (Fig. 16a), as a result of the local stress concentrations (Fig. 15d) due to GB dislocation-density accumulation at the triple junction (Fig. 15c). The crack initially grows across the block width in variant 13, but is deflected at the boundaries with variant 17 and is constrained to propagate along the block morphology (Fig. 16b), because variant 17 is not favorably oriented for fracture as it belongs to a different Bain group resulting in a large misorientation in the cleavage planes along which the crack propagates. A new crack is then nucleated in a neighboring block due to the large stress buildup ahead of the arrested crack tip (Fig. 16b). The neighboring block morphologies are such that the lath long directions are normal to the crack propagation path, which results in resistance to crack propagation, defection of the crack path along the variant boundaries due to the GB dislocation-density accumulation and cleavage plane incompatibilities, and the nucleation of microcracks ahead of the arrested crack tip (Fig. 16c, d). These intergranular and transgranular fracture modes are consistent with experimental observations (Inoue et al. 1970; Matsuda et al. 1972; Krauss 1999) as well as crack path deflection and crack arrest (Hughes et al. 2011). The nominal stress (normalized by the static yield stress)–strain curve indicating the toughening due to crack arrest and deflection at incompatible variant boundaries, and the stress drop due to the subsequent nucleation of secondary cracks during the different stages of crack propagation (Fig. 16a–d), is shown in Fig. 17.

Optimized Variant Distribution

These results are compared with the failure behavior of a variant distribution, optimized for slip transmission (Fig. 14b). The normalized mobile dislocation densities corresponding to the two most active slip systems at a nominal strain of 20 % are shown in Fig. 18a, b. The maximum normalized mobile dislocation density is 0.7 for slip system $(\bar{1}12)[1\bar{1}1]$ and 0.75 for slip system $(112)[11\bar{1}]$, compared to 0.3 and 0.38, respectively, for the random variant distribution. The mobile dislocation-density activity on these slip systems corresponds to a negative interaction density with a minimum value of -0.55 (Fig. 18c) compared to -0.35 for the random variant distribution, indicating a greater degree of dislocation annihilation activity. The accumulated plastic slip and the resulting lattice rotation at a nominal strain of 20 % are shown in Fig. 19a, b. Larger accumulated slip is

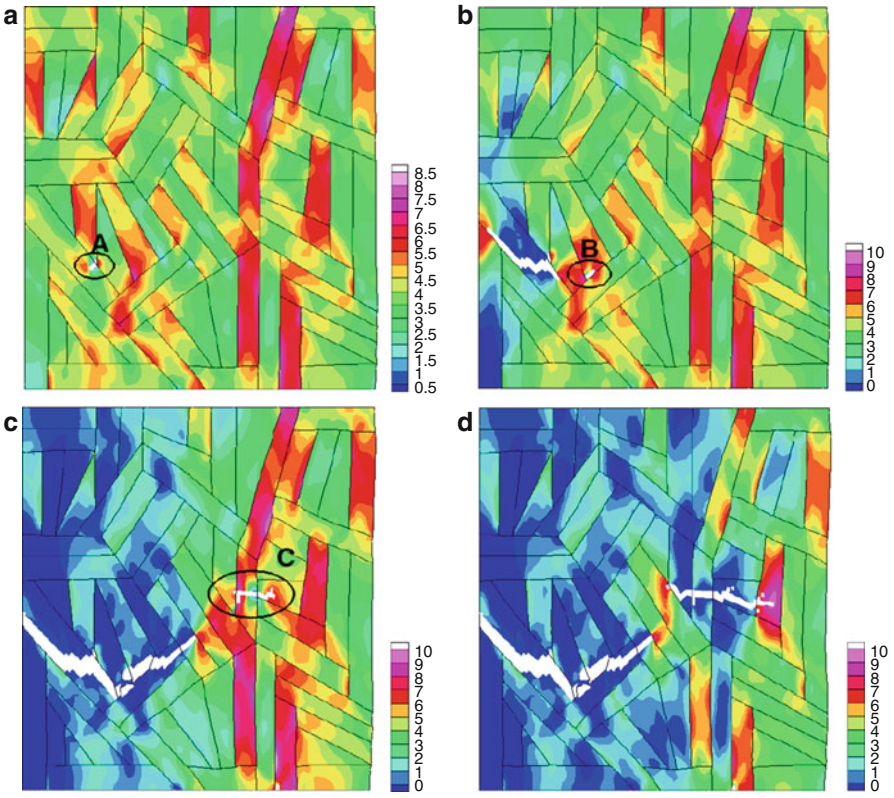


Fig. 16 Normal stress for a random variant distribution at (**a**) 5.0 % nominal strain, (**b**) 6.8 % nominal strain showing crack nucleation and (**c**) 8.0 % nominal strain, and (**d**) 8.8 % nominal strain showing crack propagation at a strain rate of 10^{-4} s^{-1}

observed, with a maximum of 0.19 compared to 0.13 for the random variant distribution, with lattice rotations ranging from -11° to 6°. The localization of shear strain in relation to the lath orientation, interaction density, and lattice rotations (Fig. 19b) occurs as discussed in section "Random Variant Distribution." The plastic slip is not constrained to flow along the lath long directions as in the random variant distribution, as the low activation energy associated with the transmission of dislocation densities between the active slip systems in neighboring blocks results in extensive plastic deformation, and lower accumulated GB dislocation densities with a maximum normalized GB dislocation density of 0.5 (Fig. 19c) compared to 0.95 in the random variant distribution. This also results in a more homogeneous normal stress distribution (Fig. 19d), with a maximum normal stress level of 6.5, as a result of the relaxation of stress concentrations at the block and packet boundaries due to dislocation transmission.

In contrast to the random variant distribution, where the refinement of the block and packet sizes increases the number of incompatible variant triple junctions,

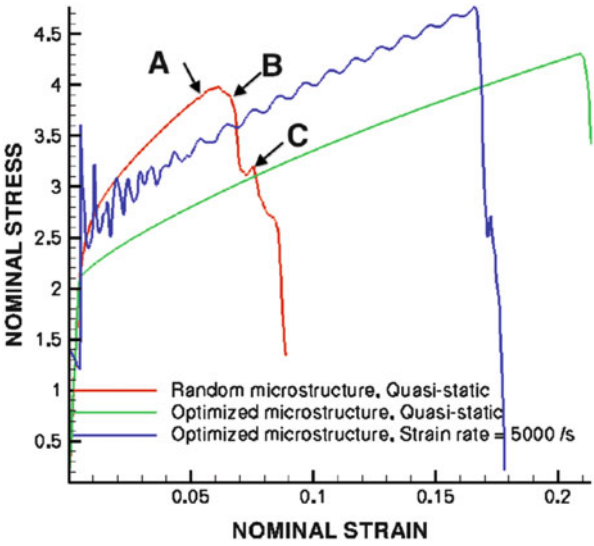


Fig. 17 Nominal stress–strain curves

which serve as sites for stress concentrations due to the accumulation of GB dislocations, and results in the nucleation of a crack at a lower nominal strain of 5 %, the large slip transmission factors associated with boundaries between variants belonging to group I in the optimized variant distribution decrease the number of incompatible variant triple junctions and result in the delayed nucleation of cracks, at a nominal strain of 21 %, and increased ductility. The dislocation-density transmission and relaxation of stress at the boundaries also result in a reduction in strength of approximately 350 MPa (Fig. 17), since the microstructural strength is associated with the resistance to plastic slip. Block size refinement is effective in improving fracture resistance by arresting the crack to propagation across variants belonging to different Bain groups. Since the crack propagation path is constrained along the lath long direction, the effects of both block and packet refinement should be considered in relation to the lath morphology in determining fracture behavior.

Dynamic Behavior

In this section, the dynamic fracture of the optimized microstructure at a loading rate of 5,000 s^{-1} is investigated. The nominal stress–strain curves over the range of loading conditions are shown in Fig. 17. The oscillations at high strain rates occur due to stress wave reflections along the free and fixed boundary, which is dampened due to the dissipative plastic mechanisms. The lower failure strain of 16 % at a strain rate of 5,000 s^{-1} is a result of dynamic strain-rate hardening (Fig. 17). The accumulated plastic slip at a nominal strain of 16 % at a strain rate of 5,000 s^{-1} is shown in Fig. 20a. The maximum accumulated slip is 0.175 compared to 0.19 under quasi-static loading, and the shear strain localization is narrower at higher strain rates, which is a result of material and thermal softening mechanisms, with a maximum temperature of 518 K, which is approximately

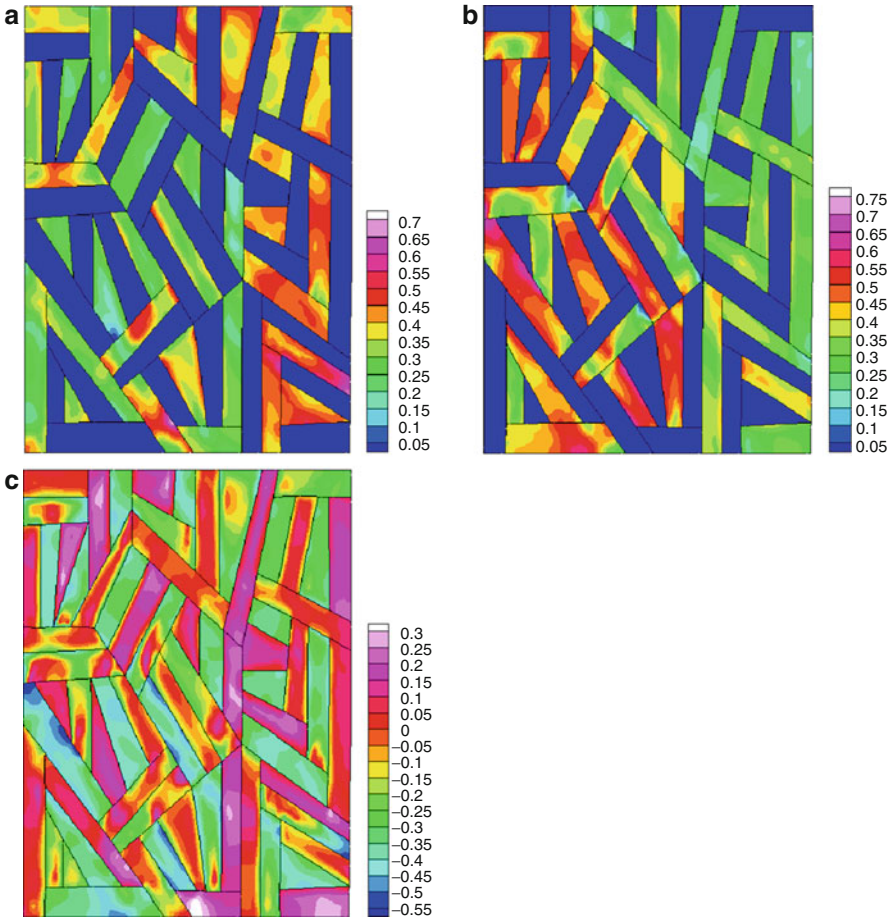


Fig. 18 Normalized mobile dislocation densities for an optimized variant distribution at a nominal strain of 20 % on (**a**) slip system $(\bar{1}12)[1\bar{1}1]$, (**b**) slip system $(112)[11\bar{1}]$, and (**c**) normalized interaction density at a strain rate of 10^{-4} s^{-1}

$0.5T_{\text{solidus}}$, as well as the dynamic strain rates, which prevent the accumulation of shear strains over wide regions. Narrow shear strain localization at higher strain rates is consistent with the experimental observations of Dodd and Bai (1985). A lower accumulation of GB dislocation densities, with a maximum value of 0.45 compared to 0.5 under quasi-static loading, occurs as a result of the strain-rate hardening at high strain rates (Fig. 20c). At this strain rate, crack nucleation occurs at a nominal strain of 16 %, as shown in Fig. 21a. Crack deviation, crack arrest, and secondary crack nucleation occur due to boundaries between variants belonging to different Bain groups, with large incompatibilities in the cleavage planes, along the variant long directions (Fig. 21b–d). This results in a

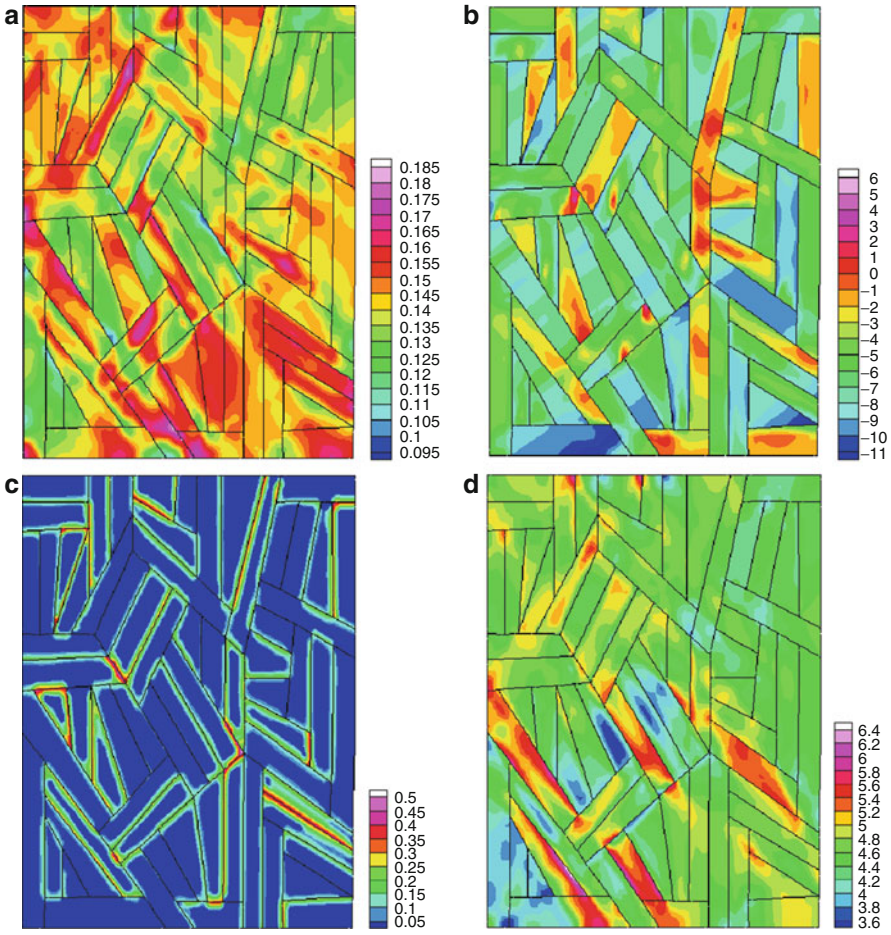


Fig. 19 (**a**) Plastic slip, (**b**) lattice rotations, (**c**) GB dislocation density, and (**d**) normal stress for an optimized variant distribution at 20 % nominal strain indicating dislocation-density pileup and plastic slip accumulation at a strain rate of 10^{-4} s^{-1}

greater fracture resistance upon block size refinement. The high strain-rate characteristics result in fracture modes that are different from quasi-static strain rates.

Conclusion

A physically based dislocation-density GB interaction scheme that is representative of the resistance to dislocation-density transmission across block and packet boundaries in martensitic steel has been developed and was incorporated into a dislocation-density-based multiple-slip crystalline plasticity framework, which

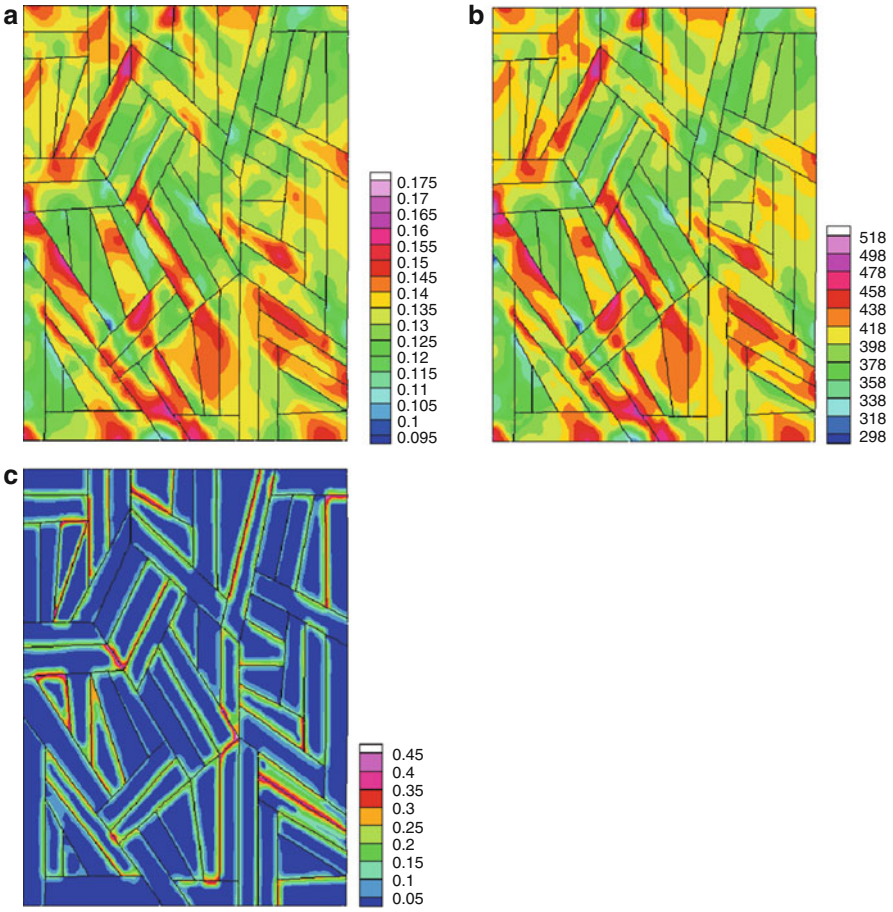


Fig. 20 (**a**) Plastic slip, (**b**) temperature, and (**c**) GB dislocation density for an optimized variant distribution at 16 % nominal strain indicating dislocation-density pileup and plastic slip accumulation at a strain rate of 2,500 s^{-1}

accounts for variant morphologies and orientation relationships that are uniquely inherent to lath martensitic microstructures.

Specialized FE methodologies to represent evolving failure surfaces and a microstructurally based failure criterion for cleavage were then used to perform large-scale FE simulations to characterize and predict the dominant dislocation-density mechanisms for the localization of plastic strains and the initiation and propagation of failure in martensitic microstructures over a range of loading conditions. The proposed fracture criteria do not assume a priori fracture criteria but are based on incorporating inelastic effects such as variant boundaries and dislocation-density evolution and interactions to resolve and nucleate cracks on cleavage planes based on evolving microstructural behavior at different physical scales.

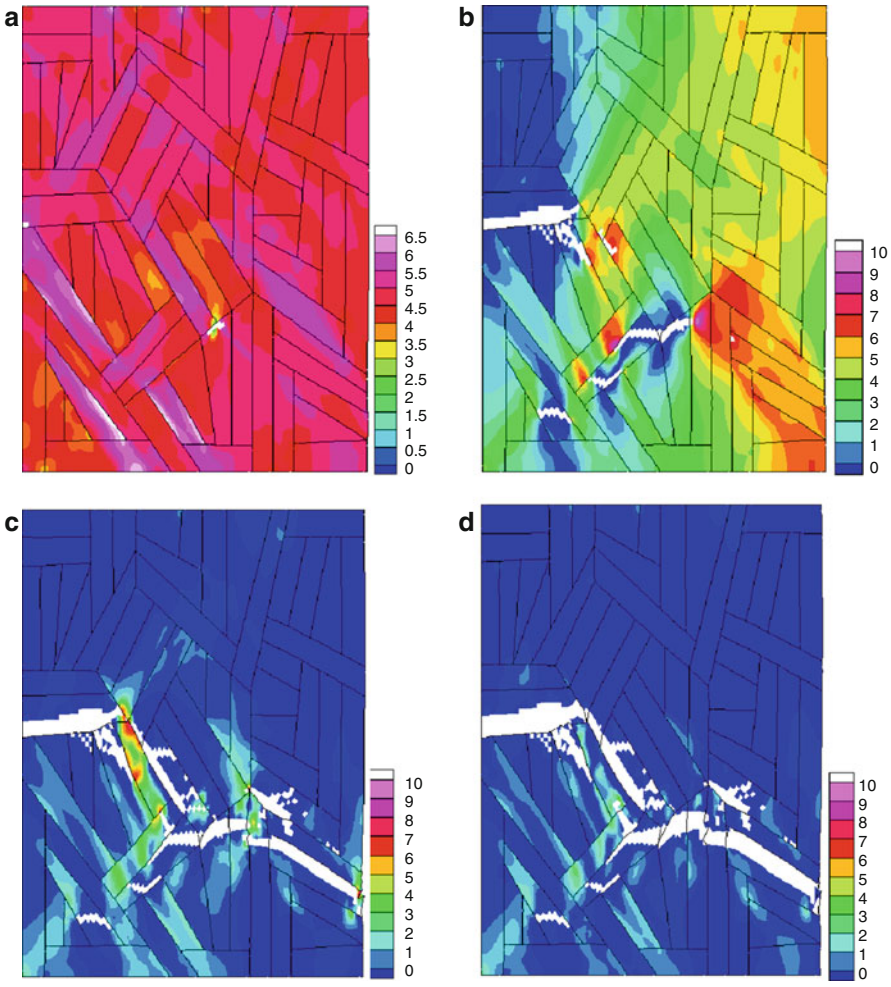


Fig. 21 Normal stress for an optimized variant distribution at (**a**) 16 % nominal strain, (**b**) 17 % nominal strain showing crack nucleation and (**c**) 18 % nominal strain, and (**d**) 19 % nominal strain showing crack propagation at a strain rate of 5,000 s^{-1}

The role of packet and block boundaries in the failure behavior of low- and high-carbon steels was investigated, and the results indicated that variant interfaces along the lath long direction were dominant. Shear strain localization was constrained by these interfaces to be along the lath long direction, and slip transmission to neighboring blocks was accommodated through compatible slip systems or the formation of low-energy residual GB dislocations. The effective block size refinement for strengthening is along the lath long direction. These interfaces also result in resistance to crack propagation and deviation of the crack path along the variant boundaries due to the GB dislocation-density accumulation and cleavage

plane incompatibilities, resulting in an effective block size for fracture resistance along the block width. It was observed that refinement is more effective in improving the fracture resistance than in improving the strength as the variant boundaries offer relatively less resistance to slip transmission onto the 24 possible slip systems, on the {110} and {112} crystallographic planes, compared to crack propagation onto only three possible {100} cleavage planes. However, refinement also increases the number of incompatible triple junctions, which serve as sites for stress concentrations, and results in the nucleation of a crack at a lower nominal strain. At high strain rates, and as a result of increased strain-rate hardening, thermal softening and dynamic interactions resulted in crack formation that was different from quasi-static loading conditions in that the failure surfaces nucleate earlier and form in regions where strain-hardening has been surmounted by thermal and geometrical softening. These validated approaches offer a predictive framework that can be used to tailor heterogeneous crystalline systems for failure-resistant and damage-tolerant applications.

Acknowledgments Support from both the US Office of Naval Research Multi-Disciplinary University Research Initiative on Sound and Electromagnetic Interacting Waves under grant number N00014-10-1-0958 and from the Office of Naval Research under grant number10848631 is gratefully acknowledged.

References

R.J. Asaro, J.R. Rice, Strain localization in ductile single-crystals. J. Mech. Phys. Solids **25**, 309–338 (1977)

M. Ayada, M. Yuga, N. Tsuji, Y. Saito, A. Yoneguti, Effect of vanadium and niobium on restoration behavior after hot deformation in medium carbon spring steels. ISIJ Int. **38**, 1022–1031 (1998)

A.A. Barani, F. Li, P. Romano, D. Ponge, D. Raabe, Design of high-strength steels by microalloying and thermomechanical treatment. Mater. Sci. Eng. A **463**, 138–146 (2007)

M. de Koning, R. Miller, V.V. Bulatov, F. Abraham, Modelling grain-boundary resistance in intergranular slip transmission. Philos. Mag. A **82**, 2511–2527 (2002)

B. Devincre, T. Hoc, L. Kubin, Dislocation mean free paths and strain hardening of crystals. Science **320**, 1745–1748 (2008)

B. Dodd, Y. Bai, Width of adiabatic shear bands. Mater. Sci. Tech. **1**, 38–40 (1985)

P. Franciosi, M. Berveiller, A. Zaoui, Latent hardening in copper and aluminum single-crystals. Acta Metall. **28**, 273–283 (1980)

Z. Guo, C.S. Lee, J.W. Morris, On coherent transformations in steel. Acta Mater. **52**, 5511–5518 (2004)

A. Hansbo, P. Hansbo, A finite element method for the simulation of strong and weak discontinuities in solid mechanics. Comput. Methods Appl. Mech. Eng. **193**, 3523–3540 (2004)

T. Hatem, M.A. Zikry, Shear pipe effects and dynamic shear–strain localization in martensitic steels. Acta Mater. **57**, 4558–4567 (2009)

A.A. Howe, Ultrafine grained steels: industrial prospects. Mater. Sci. Tech. **16**, 1264–1266 (2000)

G.M. Hughes, G.E. Smith, A.G. Crocker, P.E.J. Flewitt, An experimental and modelling study of brittle cleavage crack propagation in transformable ferritic steel. Mater. Sci. Tech. **27**, 767–773 (2011)

T. Inoue, S. Matsuda, Y. Okamura, K. Aoki, Fracture of a low carbon tempered martensite. Trans. Jpn. Inst. Metals **11**, 36–43 (1970)
S. Jin, J.W. Morris, V.F. Zackay, Grain refinement through thermal cycling in an Fe–Ni–Ti cryogenic alloy. Met Trans. **6A**, 141–149 (1975)
H. Kawata, K. Sakamoto, T. Moritani, S. Morito, T. Furuhara, T. Maki, Crystallography of ausformed upper bainite structure in Fe–9Ni–C alloys. Mater. Sci. Eng. A **438**, 140–144 (2006)
H.J. Kim, Y.H. Kim, J.W. Morris, Thermal mechanisms of grain and packet refinement in a lath martensitic steel. ISIJ Int. **38**, 1277–1285 (1998)
Y. Kimura, T. Inoue, F. Yin, K. Tsuzaki, Inverse temperature dependence of toughness in an ultrafine grain-structure steel. Science **320**, 1057–1060 (2008)
G. Krauss, Martensite in steel: strength and structure. Mater. Sci. Eng. A **273–275**, 40–57 (1999)
L. Kubin, B. Devincre, T. Hoc, Towards a physical model for strain hardening in fcc crystals. Mater. Sci. Eng. A **483–484**, 19–24 (2008a)
L. Kubin, B. Devincre, T. Hoc, Modeling dislocation storage rates and mean free paths in face-centered cubic crystals. Acta Mater. **56**, 6040–6049 (2008b)
T.C. Lee, I.M. Robertson, H.K. Birnbaim, An in situ transmission electron-microscope deformation study of the slip transfer mechanisms in metals. Metall. Trans. A **21**, 2437–2447 (1990)
A. Ma, F. Roter, D. Raabe, Studying the effect of grain boundaries in dislocation density based crystal-plasticity finite element simulations. Int. J. Solids Struct. **43**, 7287–7303 (2006)
R. Madec, L.P. Kubin, Second order junctions and strain hardening in bcc and fcc crystals. Scripta Mater. **58**, 767–770 (2008)
T. Maki, K. Tsuzaki, I. Tamura, The morphology of microstructure composed of lath martensites in steels. Trans. Iron Steel Inst. Jpn. **20**, 207 (1980)
S. Matsuda, Y. Okamura, T. Inoue, H. Mimura, Toughness and effective grain-size in heat-treated low-alloy high-strength steels. Trans. Iron Steel Inst. Jpn. **12**, 325–333 (1972)
K. Minaar, M. Zhou, An analysis of the dynamic shear failure resistance of structural metals. J. Mech. Phys. Solids **46**, 2155–2170 (1998)
S. Morito, H. Tanaka, R. Konoshi, T. Furuhara, T. Maki, The morphology and crystallography of lath martensite in Fe–C alloys. Acta Mater. **51**, 1789–1799 (2003)
S. Morito, X. Huang, T. Furuhara, T. Maki, N. Hansen, The morphology and crystallography of lath martensite in alloy steels. Acta Mater. **54**, 5323–5331 (2006)
J.W. Morris, On the ductile–brittle transition in lath martensitic steel. ISIJ Int. **51**, 1569–1575 (2011)
J.W. Morris, Z. Guo, C.R. Krenn, Y.H. Kim, The limits of strength and toughness in steel. ISIJ Int. **41**, 599–611 (2011)
T. Ohmura, K. Tsuzaki, Plasticity initiation and subsequent deformation behavior in the vicinity of single grain boundary investigated through nanoindentation technique. J. Mater. Res. **42**, 1728–1732 (2007)
T. Ohmura, A.M. Minor, E.A. Starch, J.W. Morris, Dislocation-grain boundary interactions in martensitic steel observed through in situ nanoindentation in a transmission electron microscope. J. Mater. Res. **12**, 3626–3632 (2004)
S. Queyreau, G. Monnet, B. Devincre, Slip systems interactions in alpha-iron determined by dislocation dynamics simulations. Int. J. Plast. **25**, 361–377 (2009)
P. Shanthraj, M.A. Zikry, Dislocation density evolution and interactions in crystalline materials. Acta Mater. **59**, 7695–7702 (2011)
P. Shanthraj, M.A. Zikry, Dislocation-density mechanisms for void interactions in crystalline materials. Int. J. Plast. **34**, 154–163 (2012a)
P. Shanthraj, M.A. Zikry, Optimal microstructures for martensitic steels. J. Mater. Res. **27**, 1598–1611 (2012b)
A. Shibata, T. Nagoshi, M. Sone, S. Morito, Y. Higo, Evaluation of the block boundary and sub-block boundary strengths of ferrous lath martensite using a micro-bending test. Mater. Sci. Eng. A **29**, 7538–7544 (2010)

R. Song, D. Ponge, D. Raabe, Mechanical properties of an ultrafine grained C–Mn steel processed by warm deformation and annealing. Acta Mater. **53**, 4881–4892 (2005)

J.H. Song, M.A. Areias Pedro, T. Belytschko, A method for dynamic crack and shear band propagation with phantom nodes. Int. J. Numer. Methods Eng. **67**, 868–893 (2006)

S. Takaki, K. Kawasaki, Y. Kimura, Mechanical properties of ultra fine grains steels. J. Mater. Process. Technol. **117**, 359–363 (2001)

N. Tsuji, Y. Ito, Y. Saito, Y. Minamino, Strength and ductility of ultrafine grained aluminum and iron produced by ARB and annealing. Scripta Mater. **47**, 893–899 (2002)

N. Tsuji, Y. Ito, Y. Saito, Y. Minamino, Toughness of ultrafine grained ferritic steels fabricated by ARB and annealing process. Mater. Trans. **45**, 2272–2281 (2004)

N. Tsuji, N. Kamikawa, R. Ueji, N. Takata, H. Koyama, D. Terada, Managing both strength and ductility in ultrafine grained steels. ISIJ int. **48**, 1114–1121 (2008)

M.A. Zikry, An accurate and stable algorithm for high strain-rate finite strain plasticity. Comput. Struct. **50**, 337–350 (1994)

M.A. Zikry, M. Kao, Inelastic microstructural failure mechanisms in crystalline materials with high angle grain boundaries. J. Mech. Phys. Solids **44**, 1765–1798 (1996)

Molecular Dynamics Simulations of Plastic Damage in Metals

15

Shijing Lu, Dong Li, and Donald W. Brenner

Contents

Abstract

Presented in this chapter is an overview of molecular dynamics simulation (MDS) as applied to modeling damage in metals. This is followed by some examples that illustrate how this technique is being used in the engineering

S. Lu • D. Li • D.W. Brenner (✉)
Department of Materials Science and Engineering, North Carolina State University, Raleigh, NC, USA
e-mail: slu3@ncsu.edu; dli18@ncsu.edu; brenner@ncsu.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_6

community to help understand how new materials can be made that have targeted mechanical properties.

Introduction and Historical Perspective

Presented in this chapter is a relatively brief overview of molecular dynamics simulation (MDS) as applied to modeling damage in metals along with some examples that illustrate how this technique is being used in the engineering community to help understand how new materials can be made that have targeted mechanical properties. MDS is a relatively straightforward computational method in which atom trajectories are calculated and analyzed for different sets of conditions (Allen and Tildesley 1989). The fundamental assumption in MDS is that atom motion can be adequately treated by classical mechanics and therefore their trajectories can be calculated by numerically integrating a set of classical equations of motion that are coupled through the interatomic forces. For MDS there are no expressions such as constitutive relations that are typical to modeling methods at higher length scales. Instead all of the properties of a material are set by the expression used to model the interatomic forces.

Interpreting an MDS typically starts by following the atom motion and then using different techniques to analyze structures, potential energies, and forces as a simulation progresses. As discussed in the following section, deriving and fitting the mathematical expressions used for the interatomic forces that accurately describe particular systems and analyzing simulations to both obtain detailed information and glean new understandings of materials properties at the atomic level are part science (Brenner et al. 1998) and also often part art (Brenner 2000).

In contrast to its relatively recent emergence as an engineering tool, MDS has a long history in the physical sciences. The first study of a chemical process using classical trajectories was published in 1936 by Hirschfelder, Eyring, and Topley, who followed one trajectory associated with the reaction $H + H_2 \rightarrow H_2 + H$ (Hirschfelder et al. 1936). Unfortunately the potential energy surface that they used was flawed in that it yielded a stable H_3 molecule. A consequence of this flaw (and the resources that were limited to hand calculation) is that the reaction never completed. In the following two decades, classical trajectories were used to probe some of the fundamental principles of statistical mechanics, including the rate at which equilibrium is established for relatively simple systems, and the conditions under which ensemble and time averages agree with one another (Alder and Wainwright 1957; Berman and Izrailev 2005). Some of the conclusions from these simulations are that some small and simple systems (e.g., short harmonic chains with weak anharmonic interactions) may never fully equilibrate to any practical degree and that approximations that are inherent to numerically solving classical equations of motion in other systems can ensure an approach to equilibrium in a manner that is similar to how a small disturbance to a physical system can help drive the system to equilibrium.

In 1959, Vineyard and coworkers at Brookhaven National Laboratory arguably carried out first contribution of MDS to materials science. They reported on a study that used classical dynamics to model ion irradiation and damage in metals (Gibson et al. 1960). In subsequent work by Rahman at Argonne National Laboratory, classical trajectories were used to characterize properties of liquids (Rahman 1964). Over the following decades, computing power continued to increased, which allowed Rahman and others to extend the initial work to more complicated fluids (including water) and biological systems (McCammon et al. 1977; Stillinger 1974).

Several common themes emerged from these early MDS studies. First, the development of MDS and the systems that could be studied paralleled the availability of increasingly powerful computing resources, particularly those at the U.S. National Laboratories. Second, MDS was rapidly becoming a powerful tool in chemistry and physics that could complement both experiment and theory. For example, not only could MDS provide data but it could also be used to test (and refine when needed) theories under exceedingly well-controlled conditions. Finally, it became clear that the requirements put on a potential energy function (from which interatomic forces are obtained) are very strongly dependent on the application. In chemistry applications, for example, typically well-refined potential energy surfaces are needed to provide qualitative data, but even highly approximate or generic forces can yield useful data for examining general phenomena. Similarly simple potentials can yield very useful information in statistical mechanics. In the case of materials, modeling simplified interatomic force expressions can be used to model defects such a grain boundaries and dislocations; however, these expressions must be detailed enough with regard to particular materials to yield useful, quantitative data.

The increasing importance of MDS as an engineering tool can be correlated to several factors. First, an exponential increase in the speed of computer processors together with relatively inexpensive platforms for parallel computing, visualization, and data storage has significantly decreased the cost of resources needed to carry out large-scale simulations. Consequently computing platforms have come to the point that MDS is affordable for research groups with even modest modeling budgets. Second, due to advances in experimental processing and characterization techniques, the structure of some materials can be manipulated to scales approaching the atomic level. This is well below the scales where traditional continuum modeling approaches like finite element analysis are appropriate. Finally, interatomic potential energy functions, which as mentioned above play a central role in MDS, have become available that provide good descriptions of the structure and elastic and mechanical properties of a wide range of materials while at the same time being sufficiently efficient from a computing viewpoint that they can be used in large-scale simulations. These advances in potential energy functions have resulted from new theoretical analyses that results in mathematical expressions that can more accurately capture quantum bonding effects, a better understanding of how to fit these expressions to the properties of specific materials, and readily available simulation codes that allow researchers outside of the original

developers to vet and improve interatomic potentials as they are used (Brenner et al. 1998; Sinnott and Brenner 2012).

Given in the following section are brief details presented at a tutorial level on some of the assumptions, numerical methods, and potential energy functions that are essential parts of an MDS to study damage in materials. Also mentioned are some currently available codes and user communities that are useful for carrying out MDS. This is followed by a section in which some applications of MDS for understanding damage in crystalline metals are discussed. This is not a comprehensive literature review but rather a survey of results that illustrate the various ways in which MDS is contributing to this area. The final section includes a brief discussion of new capabilities being developed for MDS of metals and the challenges facing this field in the short and long term.

Molecular Dynamics Simulations

In comparison to other modeling techniques, carrying out an MDS is relatively straight forward. The initial step is to initialize coordinates and the velocities for each atom and to establish any system conditions such as periodic boundaries. Using these positions, the interatomic forces are calculated, followed by the determination of constraints from thermostats or other conditions imposed on the system. As discussed more below, these constraints can be in form of additional forces, velocity rescaling, changes in periodic box size and shape, application of shock waves, etc. Atom positions are then moved ahead one time step according to forces, velocities, and accelerations (and higher derivatives of position with time depending on the numerical integration) using one of many possible numerical integration schemes. The atom positions, velocities, and forces can be analyzed during the course of a simulation, and post-simulation analysis can be carried out. The former can be done at the same time that the simulation is carried out or done on a file that saves atomic properties at some specific interval. The choice depends on considerations like system size, the total number of time steps, computing and visualization facilities, and whether there are user constraints applied to a system as a simulation progresses, including terminating a simulation if desired. The process is then repeated with a new set of forces and constraints calculated from the updated atom positions. These steps are described in more detail in the following subsections.

Initial Conditions

For crystals, atom positions are usually generated from a given lattice and basis. To mitigate edge effects when modeling a bulk material, periodic boundaries can be applied to the system such that an atom leaving from one side of the simulations reenters from the complementary side, and pairs of atoms interact with their nearest image in the repeating boxes. Figure 1, for example, illustrates a two-dimensional

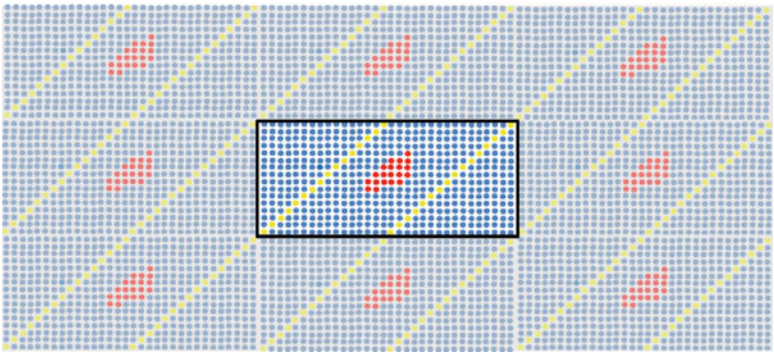


Fig. 1 Illustration of a localized defect (cluster of *red spheres* representing atoms) and an extended defect (*line* of *yellow spheres* representing atoms) in a *central box* surrounded by eight replicas that are defined by the periodic boundaries

computational box surrounded by eight replicas, where the circles represent atoms. For localized defects, represented by the cluster of red atoms in Fig. 1, periodic boundaries create a periodic array of the defect. For an extended defect that passes over the cell boundaries, which is represented by the yellow atoms in Fig. 1, the periodic boundaries create an infinite structure that is constrained to a superperiodicity defined by that of the periodic boundaries. This can constrain the types and number of extended defects that can be formed in a simulated lattice. Similarly, periodic boundaries can constrain stress fields that would otherwise be larger than a box dimension, which can influence plastic dynamics in a nonphysical manner.

Coordinates for initial point defects like voids and interstitial atoms are generally trivial to generate; however, creating coordinates for more complicated microstructure elements like grain boundaries and dislocations can be a challenge. There are presently commercial codes that can generate structures for a number of common defects and free tools like the gosam website that can generate coordinates for bicrystals with periodic boundaries based on coincident site lattices (Wojdyr et al. 2010). To generate velocities for what is supposed to be an initially equilibrium simulation, initial velocities can be generated randomly from a Boltzmann distribution that corresponds to a particle temperature. Other possible initial velocity choices may be based on attempting to mimic particular nonequilibrium situations such as an applied shock, a collision with a surface, or the initiation of a collision cascade.

Interatomic Force Expressions

The atom dynamics that mimic the behavior of a particular material are mainly determined from the expression used to describe the interatomic forces. This expression is therefore a key feature of the simulation in terms of reproducing properties

that would correspond to an experiment on the true system. Calculating the interatomic forces is also typically the most computationally expensive part of an MDS and the feature that is least well known in terms of the appropriate equations.

In principle interatomic forces can be obtained from the quantum forces due to the electrons for a given set of atom positions and then adding these forces to those from the internuclear repulsions. This is done in an especially efficient manner through Car–Parrinello dynamics, a method that defines an extended Lagrangian containing nuclear and electronic degrees of freedom from which a set of coupled equations of motion are derived for both the ions and electrons (Car and Parrinello 1985). While mitigating much of the uncertainty associated with deriving and fitting analytic force expressions like those discussed below, this method remains computationally expensive in terms of evaluating the expression. In addition, because electrons are explicitly treated, the time steps are typically smaller when numerically integrating the electron–ion equations of motion compared to those for atoms only. Furthermore, the electronic degrees of freedom should be periodically "quenched" to the Born–Oppenheimer surface, which adds some additional computational overhead. There are also simplified electronic structure methods (e.g., tight binding Hamiltonians) that can be used to include electronic degrees of freedom into MDS forces. These approaches have significantly reduced computational demands compared to, e.g., the Car–Parrinello method, but they remain computationally burdensome for large systems and can require a significant degree of parameterization compared to fully first-principles methods.

Because of the desire to treat systems that are sufficiently large to model extended defects such as dislocations, an analytic expression that gives the potential energy of a collection of atoms as a function of the relative coordinates is typically used to mimic explicit electronic degrees of freedom (Sinnott and Brenner 2012). The interatomic forces are then calculated as the gradient of this potential energy function (PEF). The challenge with this approach of using an effective PEF is to come up with a function that not only accurately reproduces some set of fitting data but is also capable of producing physically reasonable (if not accurate) energies for atomic arrangements that are not considered in a database used to fit the PEF parameters. This property of a PEF is called the transferability. Currently there is no single standard functional form that is used for all systems; instead there are a plethora of forms and approaches that work to varying degrees of success. Developing a PEF requires both developing a convenient mathematical expression and parameterizing that expression to a particular system.

Probably the simplest approach to developing a PEF with continuous forces is to assume that the atoms interact in pairs, with other neighboring atoms not having any influence on the interactions between an atom pair. With this approximation, the total potential energy U is given by

$$U = \Sigma_{i=1}^{n-1} \Sigma_{j=i+1}^{n} V\left(r_{ij}\right) \tag{1}$$

where n is the number of atoms, r_{ij} is the scalar distance between atom i and atom j, and V(r) is a function such as a Morse or Lennard-Jones potential that has

parameters that can be fit to a set of physical properties (Brenner and Garrison 1989). For a pair interaction with a single well, pair potentials tend to produce stable close-packed lattices as lowest energy structures, so for metals pair potentials can sometimes produce acceptable results depending on the collection of properties that are trying to be reproduced. The pair potential approximation, however, is limited for reproducing elastic properties, and in general, it does not produce accurate energies for defects such as surfaces and vacancies.

One way to build on the pair potential approximation is to use it as the lead term in a Taylor series expansion of potential energy in relative atomic coordinates. The many-body expansion can be given as

$$U = \Sigma_{i=1}^{n-1}\Sigma_{j=i+1}^{n} V_2(r_{ij}) + \Sigma_{i=1}^{n}\Sigma_{j(\neq i,k)=1}^{n}\Sigma_{k(\neq i,j)=1}^{n} V_2(r_{ij}) + \ldots \quad (2)$$

where the subscripts give the order of the interaction. The first term in Eq. 2 represents a bond stretch, the second term represents a bond bend, and higher-order terms correspond to dihedral angles and combinations of bends and stretches. This expansion combined with electrostatic and other nonbonded interactions has been extensively used in biological and polymeric systems. This form of potential function, however, has found limited use for simulating metals.

Experience suggests that the most transferable PEFs are those that are based on expressions derived from principles of quantum mechanical bonding. There are several classes of PEFs that fall under this category that are widely used for modeling metals. These are the embedded-atom method (EAM) (Daw and Baskes 1984; Daw et al. 1993), effective medium theory (Jacobsen et al. 1987), the glue model (Ercolessi et al. 1986), and the Finnis–Sinclair (Finnis and Sinclair 1984) potential. While the functional forms of each of these classes of PEFs are similar, the derivation of each is different. The first three models come from the concept of "embedding" atoms into an electron gas so that the energy of an atom depends on the density of the electron gas near the embedded atom plus a pair interaction between the atomic cores. In the glue and EAM models, the electron density is modeled as a pairwise sum of densities from surrounding atoms at the site of the atom whose energy is being determined. This leads to three central components of these PEFs: the contribution of the electron density from a neighboring atom to a given site (which depends on the distance between the atom and site), a pairwise-additive interatomic force, and an embedding function that relates the electron density at a given site to the energy of an atom at that site. Each contribution can be adjusted by fitting to properties such as the crystal cohesive energy, lattice constant, elastic moduli, and vacancy formation energy. Effective medium theory, in contrast, averages the electron density in the vicinity of an atom whose energy is being calculated and uses a less empirical relationship between this electron density and the energy of an atom (Jacobsen et al. 1987).

The Finnis–Sinclair form of the PEF is similar to other metal PEFs. However, the derivation of these expressions from quantum bonding principles is different. The Finnis–Sinclair functional expression is derived from a "second moment" approximation that relates the spread in the local electronic density of states due

to bonding to the potential energy of an atom and to its local coordination. The derivation is too long to be included here (Brenner et al. 1998), but the result is that the Finnis–Sinclair embedded energy is a square root of the coordination, while the form of the EAM embedding function is part of the parameter fitting process. Another historical difference is that the EAM and related potentials were originally developed for face-centered cubic metals, while the Finnis–Sinclair potentials originally focused on describing body-centered cubic metals.

A major strength of these PEFs is that by using pair terms for the interatomic repulsion and for the electron densities, their evaluation scales like a pair potential while a many-body aspect of the PEF is introduced by an embedding function. There have been many attempts to go beyond these basic PEFs; higher moments of the electrons and angular interactions in the electron densities can be introduced, for example (Dongare et al. 2012; Lee and Baskes 2000). These approaches, however, typically add computational burden that must be balanced against any improved accuracy in reproducing the fitting database and in the overall PEF transferability.

Incorporation of charge transfer terms into a PEF that allow for different degrees of ionic bonding is a development that has produced good results for modeling metal oxides and metal interfaces. The two leading formalisms for metals that include charge transfer are the ReaxFF (Van Duin et al. 2001) and charge-optimized many-body (COMB) expressions (Shan et al. 2010). Both of these PEFs take advantage of a bond-order formalism that includes angular terms to model local bonding interactions plus Coulomb terms that use atom-centered partial charges in the form of a Gaussian function. The magnitude and sign of the charge on each atom are determined by minimizing the electrostatic energy with respect to the potential energy from the Coulomb interactions between partially charged atoms and energy associated with forming the charges. The latter can be expressed as a Taylor series in charge with coefficients determined from the atomic ionization and electron affinity energies. In practice, however, these coefficients are typically taken as parameters that can be fit along with parameters in the short-range bonding interactions.

In addition to the functional form of the PEF, the training set and the method used to parameterize the function are also important. A minimum database for parameter fitting for metals traditionally includes cohesive energy, lattice constant, elastic properties, and vacancy formation energy of the lowest energy crystal structure. It should also include the difference in cohesive energies between the lowest energy structure and other competing crystal structures. With this type of fitting database, a PEF can be validated against quantities such as melting points and surface energies. In the past, a fitting database was usually developed from experimental measurements. However, first-principles data is increasingly being used to fit parameters in PEFs. This approach has the major advantage that very large data sets can be generated (e.g., transition states for activated processes) with comparable levels of error for each quantity evaluated. A related approach is to carry out a simulation with quantum forces, for example, Car–Parrinello dynamics, and use the atomic configurations that are generated to create a best-fit set of

parameters (and even the functional form) for an analytic PEF (Ercolessi and Adams 1994). As the number of configurations increases, the analytic PEF can be substituted for the quantum forces for larger systems and longer time simulations.

Classical Equations of Motion, Numerical Integrators, and Thermostats

Calculating atomic trajectories involves approximating differential classical equations of motion by finite difference approximations that are solved stepwise using one of several standard methods employed in MDS. These numerical integration methods include Gear, predictor–corrector, and leapfrog algorithms. Which is used depends on whether constraints are operating on the system, the desired accuracy of the solution, and whether time steps with varying sizes are used. The time step size depends on the integration method and the magnitude of the forces acting on the atoms. A general rule of thumb is that the time step should be no larger than about 1/20 of the shortest vibrational period in the system. For systems with hydrogen, an appropriate time step is roughly one femtosecond or less; for metals a larger time step can typically be used. An initial determination of the largest allowable time step can be done by monitoring energy conservation both step-to-step and over a given period of time. The latter evaluates energy drift, which is a measure of the accumulation rather than cancellation of errors.

Solving classical equations of motion without additional constraints leads to constant energy trajectories; at equilibrium this should yield properties that correspond to a microcanonical ensemble. The Nosé thermo- and barostats add constraints to the equations of motion that produce trajectories consistent with other thermodynamic ensembles (Nosé 1984). Similar methods have also been introduced for describing other system states. Hugoniostats, for example, produce equilibrium states for shocked matter (Ravelo et al. 2004).

Two other common ways used to control temperature are the Langevin and Berendsen thermostats. Both methods couple system dynamics to an external bath, although with different coupling methods. The Langevin approach, which comes from Brownian dynamics, adds to the interatomic force, friction, and random force terms that balance one another to produce a desired temperature. In the Berendsen approach, only a friction term is added, but the friction coefficient is proportional to the difference between the current and desired temperatures so that the friction coefficient can be positive or negative (Berendsen et al. 1984). The Langevin thermostat therefore couples each atom locally to a system bath, while the Berendsen thermostat is a global coupling between the total and desired temperatures.

Although not thermostats in the same sense as those discussed above, Zhigilei and coworkers and Brenner and coworkers introduced methods that modulate temperature in an MDS by coupling atom dynamics to a continuum reservoir (Schall et al. 2005; Padgett and Brenner 2005; Schäfer et al. 2002). In the Zhigilei approach, the reservoir represents an electron gas that exchanges energy with the nuclear degrees of freedom. With the two-temperature model, processes such as the

interaction of lasers with metals can be modeled. In this case, the laser initially excites the electronic degrees of freedom, and this energy is then appropriately channeled into the atom motion. The Brenner approach coarse grains a simulation into regions for which a temperature is defined from the average velocities of the atoms in each region. A continuum heat transport equation with parameters fit to the thermal conductivity and heat capacity of the bulk material is numerically solved for one time step on the grid defined by the coarse graining, and the atom velocities are scaled to match the resulting grid temperatures. This ad hoc coupling between grid temperatures and the atom velocities is continued as the MDS continues. In addition to controlling heat flow (and not just temperature), this approach has been extended to include electrical current flow and associated Joule heating (Crill et al. 2010; Irving et al. 2009a, b; Padgett and Brenner 2005).

The thermostats discussed here require some parameterization. For example, the Berendsen thermostat requires a decay constant between the temperature gradient and the difference between the desired and current temperature. The Nosé thermostat requires a thermal inertia parameter that determines the rate of the heat transfer in the system.

Parrinello and Rahman introduced a classical Lagrangian that allows a repeating periodic cell to change shape in response to the stress state (Parrinello and Rahman 1981). With this method, solid phase transitions can be conveniently simulated as a function of stress state and temperature. Other constraints have also been applied to classical equations of motion that produce rigid bond distances and angles, although these are more commonly applied to polymers and biological molecules (Ryckaert et al. 1977).

Other constraints can be added to simulations to model particular conditions. Compressive shock loading, for example, can be modeled with a flyer plate impacting an open end of a material. Shocks can also be generated by shrinking the periodic boundaries at each step of a simulation. Similarly, a constant tensile strain can be applied by expanding the periodic box dimensions in one or more directions, or shear strain can be applied by shearing the periodic box. As another example, heat flow can be modeled by creating hot and cold regions of a system, and quantities such as thermal diffusivity can be estimated by monitoring the temperature distribution as it evolves during the dynamics (Müller-Plathe 1997).

A recent creative example of an ad hoc condition driving simulation dynamics is the work of Foiles and coworkers, who introduced a way to drive a grain boundary to move through a crystal (Janssens et al. 2006). A parameter is defined in this method for each atom that is determined by the orientation of surrounding atoms (see Fig. 2a). If the atom orientation does not match some given reference structure, a relatively small potential energy is added to the cohesive energy of that atom. This procedure drives the system toward a particular grain orientation, which at sufficiently high temperatures can drive a grain boundary from a preferred to a non-preferred grain orientation. Similarly, an applied shear strain can induce motion of particular grain boundaries and dislocations depending on the lattice orientation with respect to the shear. Simulations using these driving conditions are discussed in more detail below.

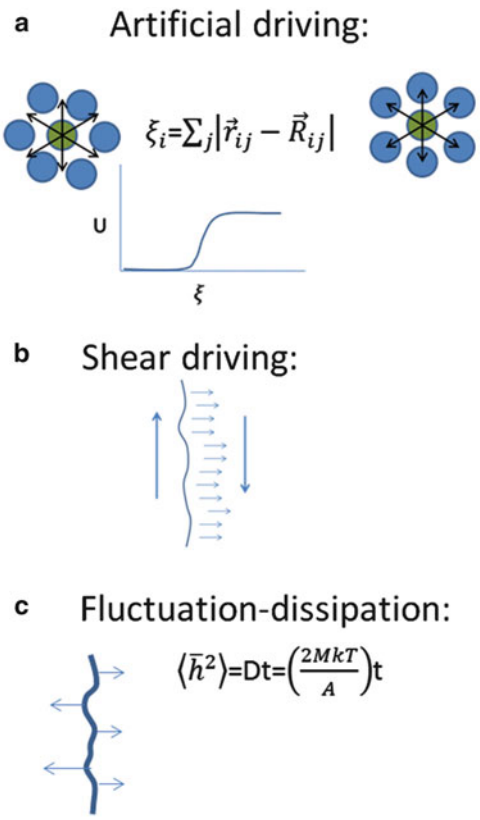


Fig. 2 Methods for determining grain boundary mobilities. (**a**) An extra energy term is added to atoms with an orientation that does not match that of the preferred grain orientation. (**b**) Grain boundary motion driven by a shear forces. (**c**) Using fluctuations in position to calculated grain boundary mobility

Analyzing Atomic Simulations

An MDS provides only atom motion, and therefore, extracting useful data and insights from a simulation requires further analysis of the trajectories. For many equilibrium properties, Green–Kubo correlation functions of statistical mechanics can be used to extract important data. For example, vibrational modes, thermal conductivity, and dispersion relations for longitudinal vibrational modes can be evaluated in this way. Similar methods have also been suggested for determining grain boundary mobilities from position fluctuations (Fig. 2c) (Trautt et al. 2006).

Other methods are needed to identify and understand the origin of damage. One of the simplest methods is to represent atoms as spheres and create movies of the trajectories (Figs. 3 and 4). Codes that produce visualizations of this type are freely available as given below. While powerful, experience shows that additional data beyond atom positions are often needed to analyze complex dynamics and to identify structures such as dislocations that correspond to well-established damage elements. With a PEF like one of those discussed above, each atom can be assigned a cohesive energy and this cohesive energy can often be used to discern structures and understand the atomic-scale origin of the energy of a particular defect. Although strictly speaking

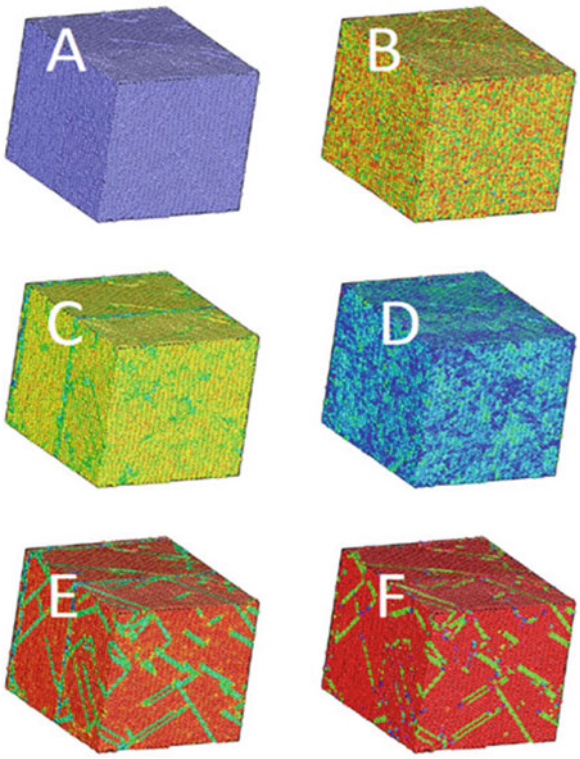


Fig. 3 Atomic system with 481,680 atoms at a temperature of 300 K containing two tilt grain boundaries and stacking faults created by dislocation motion during an applied strain. An embedded-atom method potential for copper was used. Periodic boundaries were applied and the virtual atoms are visualized as spheres. (**a**) All atoms the same color; (**b**) atoms colored by kinetic energy; (**c**) atoms colored by potential energy; (**d**) atoms colored according to local hydrostatic stress; (**e**) atoms colored according to their centrosymmetry parameter; (**f**) atoms colored according to a common-neighbor analysis

stress is a continuum concept, stresses can be derived from the interatomic forces. Like the cohesive energy, stresses can be assigned to atoms to understand subtle contributions to the stability of various defects. There are also other complementary quantities that have been proposed that are valuable for damage characterization in crystals. For example, a centrosymmetry parameter is a quantitative measure of the local bonding structure of a given atom (Kelchner et al. 1998). It has values that can be traced to particular damage configurations such as bulk crystals, dislocation cores, stacking faults, and surfaces. Similarly, a common-neighbor analysis assigns a set of four integers to each bond that is unique for certain bonding structures (Honeycutt and Andersen 1987). There are also more automated methods that not only identify dislocations but also characterize Burger's vector (Stukowski and Albe 2010). These methods greatly enhance the ability to fully analyze simulated damage, especially when a system is so large that post-simulation analysis becomes burdensome compared to analysis as simulation configurations are generated.

Multiscale Modeling

The small time step size used in MDS ($\sim 10^{-15}$ s) and system sizes of typically thousands to billions of atoms (depending on available computing resources) are

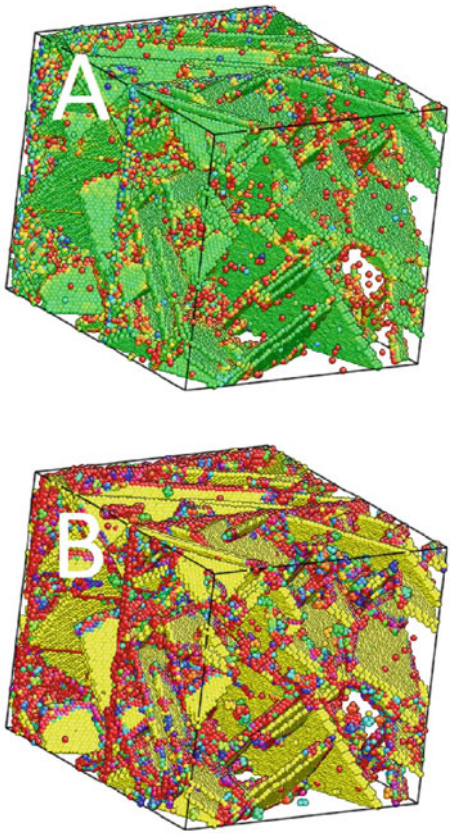


Fig. 4 Same system as depicted in Fig. 3e, f except that atoms with a bulk centrosymmetry parameter (*top panel*) and bulk common-neighbor analysis (*bottom panel*) are removed. The grain boundaries and stacking faults are now apparent in the figures

both major constraints for applying MDS to engineering-scale materials phenomena. The increasing number of processors available through different computing resources that can be dedicated to a single simulation is able to overcome some of the size limitations. However, because the trajectories are determined stepwise, the number of time steps that can be processed is more closely tied to processor speeds, which are predicted to deviate substantially from the exponential increase given by Moore's law curve over the past decades.

Different multiscale modeling approaches have been developed to overcome some of the discrepancy between the time and size scales inherent to MDS and engineering applications. These approaches can be conveniently separated into concurrent and serial methods. The former refers to approaches in which different scales are treated simultaneously in a given simulation. Serial multiscale modeling refers to approaches in which parameters and modeling conditions are passed between modeling methods that treat different scales of a given system. Brenner has recently reviewed some of the challenges and opportunities in multiscale modeling of these types (Brenner 2013).

Concurrent multiscale modeling approaches typically look to embed a region in which atoms are explicitly treated into a set of continuum boundary conditions. The problem that has been the most examined with this approach is the propagation of cracks in brittle solids, where dynamics at the crack tip are resolved to the atom scale, while surrounding stresses are modeled with finite element or other continuum approaches. These types of simulations can be partitioned to even finer scales. For example, in a region that includes a crack tip, the interatomic forces can be calculated from explicit electronic states. A region further from the tip can be modeled with an analytic PEF, and further regions can be treated as a continuum.

Miller and Tadmor have compiled a detailed comparison of the accuracy and efficiency of 14 concurrent atom–continuum approaches (which they call "partition-domain multiscale modeling") for an example problem of dislocation dipole motion in a face-centered cubic aluminum lattice (Miller and Tadmor 2009). Challenges to this approach, which are discussed in detail in the Miller–Tadmor and Brenner reviews (Brenner 2013), include matching forces at the atom–continuum boundaries (and the boundaries where different interatomic forces are used), phonon motion across boundaries, and being able to switch between continuum and atomistic descriptions of a region of the system as damage evolves and stresses change.

Voter and others have proposed different schemes to overcome the time limitations of MDS when modeling rare events. In a straightforward and elegant scheme called parallel replica dynamics, multiple trajectories are followed from the same initial atomic positions but with sufficiently different initial equilibrium velocities that they explore different phase space (Voter 1998). When a rare event reaction is detected in any of the systems, all of the runs are discontinued. For first-order kinetics, the total time between reactions can be taken as the sum of the times for each system replica. The systems for which no reactions were observed are eliminated, and the process is started again using the one system that did show a reaction. Because no communication is needed between replicas as they run in parallel, this approach is highly parallel in time.

In a different approach developed by Voter called hyper-dynamics (Voter 1997), a bias potential is used to raise the potential energy in regions of the potential energy hypersurface between saddle points that correspond to various reactions. With this approach, an effective time step can be defined that depends on the bias potential and that can be orders of magnitude larger than the actual time step used for following the trajectory on the new potential surface. When the system approaches a configuration corresponding to a reaction, the bias potentials are removed, and the dynamics of the system is followed using the small time step typical of an MDS. As a system moves away from the region of the reaction barrier, the bias potential and effective time step are reestablished. One drawback to this approach is that it does not scale well with system size. This is because multiple rare events can occur that bring the time step for the entire system to the small size inherent to systems with non-biased potentials. Because they are based on accelerating rare event dynamics, these types of time-scaling approaches are not generally used to model large damage in crystalline solids, which typically involve creation of extended defect structures.

Available Codes

Several freely available codes for executing and analyzing MDS of damage in crystalline metals are summarized in Table 1. Probably the most widely used at this point is LAMMPS (the acronym for Large-scale Atomic/Molecular Massively Parallel Simulator) (Plimpton 1995). Maintained by researchers at Sandia National Laboratories, this code has a large and still growing number of capabilities that are often contributed by an active user base. This code's capabilities include a high degree of parallelism, multiple types of interatomic potentials and thermostats, advanced multiscale modeling capabilities, and different ways of post-processing and analyzing simulations.

Another very useful suite of MDS codes is maintained by the Theoretical and Computational Biophysics Group at the University of Illinois at Urbana-Champaign. Although the capabilities of this code emphasize biological systems, the visualizer VMD can be useful for analyzing damage in crystalline metals generated from an MDS, especially when used in combination with one of the methods mentioned above (e.g., the Central Symmetry Parameter) to color atoms during visualization (Huang and Humphreys 2000). Other examples of available and very useful codes for MDS and analysis include GULP(Gale and Rohl 2003) and AtomEye (Li 2003).

Example Simulations of Metal Dynamics

Examples from the literature and from work within our research group are discussed below in which MDS was used to characterize damage in crystalline metals and their alloys. The discussion is not intended as an extensive literature review; instead these were chosen as examples because they demonstrate new insights and unique data that can be obtained from MDS in this area.

Table 1 Summary of some useful MDS and related codes for metal simulations

Code name	System emphasis	Example capabilities	Website
LAMMPS	All	MDS and serial multiscale dynamics; efficient processor scaling; large number of routines for thermostating, etc.	lammps.sandia.gov
NAMD/VMD	Biological	Molecular mechanics; efficient processor scaling; independent MDS visualization code	www.ks.uiuc.edu
GULP	All	Lattice dynamics	projects.ivec.org/gulp/
AtomEye	All	Advanced MDS visualization	http://li.mit.edu/Archive/Graphics/A/

Simulations of Shock-Loaded Crystals

Shock fronts travel at speeds that exceed the sound speed of the medium through which they are traveling. For metals, the sound speed is typically 1–6 km per second, or in molecular terms ~1–6 nm/ps. While the transient properties of shock fronts make them difficult to study experimentally, this is an ideal time and length regime for MDS. MDS has therefore been used extensively to study shock phenomena in solids.

Early MDS by Holian of close-packed solids subject to shock loading suggested that for strong shocks, shear damage (e.g., dislocation motion) can occur very near the shock front, while for intermediate-strength shocks, an elastic compression at the front is a precursor to a steady plastic wave in which shear stress is relaxed (Holian 1998). At lower shock strengths, only an elastic wave was observed, with the transition between the two occurring when the ratio of the piston speed to the long wavelength sound speed exceeds about 0.25. This is roughly the point at which the shear pressure equals the theoretical shear strength of the solid (~1/10 of the shear modulus for slippage of close-packed planes).

Holian's study used periodic boundaries and was restricted to thousands of atoms. These restrictions can influence the plastic damage observed in an MDS because any line and planar defects that are formed must repeat within the boundary box in the two directions perpendicular to the shock propagation (see Fig. 1). For Holian's MDS, this resulted in slip being restricted to a limited number of close-packed planes with spacing between shear bands dictated by the periodic boundaries. Whether this influenced the observed sudden onset of plastic damage is not clear. Because of this and related considerations, Holian and Lomdahl examined shock loading of larger systems containing up to 10 million atoms in a close-packed crystal lattice (Holian and Lomdahl 1998). In the larger system, significant slippage along multiple planes was observed, and the interaction between active slip systems reduced the planarity of the shock front. Distinct patterns of plastic damage (termed "randomly placed plaid pattern") were observed in the MDS. This observation suggested that the periodic boundaries did not significantly influence plastic damage mechanisms, which included a sudden onset of plastic damage depending on the shock loading strength. To examine the influence of extended defects on the onset of plastic damage, the planar front driving the shock was replaced with a curved interface that fit within the periodic boundaries. This curved interface was intended to model the influence of existing extended defects on an initially planar shock fronts as it moves through a crystal. Plastic damage at significantly lower shock loadings was observed in the MDS compared to the driving force comparable to a flat piston (e.g., heterogeneous versus homogeneous nucleation of plastic damage under shock loading).

Holian and coworkers also used MDS to characterize shock-induced transformation of iron from a body-centered cubic to a hexagonal close-packed structure (Kadau et al. 2002). At low piston velocities, there was no evidence for plastic deformation in the initial structure. Instead a split shock structure was observed that was composed of an elastic precursor under uniaxial compression followed by a

slower phase-transition front to the close-packed structure. At higher pressures the strain within the compressed precursor was relieved by a homogeneous nucleation of close-packed grains. Because only a few grains were initially formed, the transformation front was rough. Increasing the velocity of the piston driving the shock loading increased the nucleation rate of close-packed grains. This in turn produced a smoother transition front until at very high shock loadings a single front was observed that separated the body-centered cubic from the close-packed structure (which contained multiple twins and related defects).

In a following study, diffraction signals were generated from the structures created in the MDS and compared to in situ experimental x-ray measurements of laser-driven shocked iron. The observation in the MDS of a transformation to a hexagonal close-packed structure was confirmed by the experiments. The presence of a uniaxially compressed precursor to a phase transformation was also confirmed.

A large number of other MDS studies of shocked single-crystal metals have been carried out that are similar to the work of Holian and coworkers. For example, Bringa et al. reported results from a combined experimental and MDS study of the shock loading of single-crystal copper (Bringa et al. 2005). The MDS used sample sizes of two million atoms and two different EAM potentials that differed in their stacking fault energies. A light-gas gun was used in the experimental studies to shock load the samples. The relationship between shock velocity and the velocity of the driving piston (the Hugoniot relation) was found to depend strongly on the orientation of the shock loading, as did the agreement between the experiment and MDS. Consistent with Holian's studies, an elastic precursor followed by plastic deformation was observed in the simulations, as well as melting at high shock loading pressures. Dupont and Germann simulated the uniaxial strain of single-crystal copper at different strain rates and temperatures along different orientations (Dupont and Germann 2012). The yield strength and its dependence on strain rate were found to be a function of orientation but not temperature. With increasing strain rate, the plastic dynamics progressed from dislocation nucleation to atom disordering/dislocation nucleation to creation of an amorphous structure.

Shehadeh et al. used the results of an MDS of shock-loaded copper to validate and parameterize a larger length-scale model involving discrete dislocation dynamics coupled to finite element analysis (Shehadeh et al. 2006). This concurrent multiscale approach allowed the analysis of dislocation loop formation and related plastic damage to be carried out to much longer time and length scales than would be possible with MDS alone. This approach also provides a more accurate representation of the damage dynamics than might otherwise be used in the finite element or discrete dislocation approach.

A steady-state two-zone elastic–plastic model was proposed based on the results of MDS studies of the shock loading of single crystals by White and coworkers. In this model, the wave establishing the elastic precursor and the wave front at which shear stress is relieved via plastic deformation travel at the same speed (Zhakhovsky et al. 2011). The coordination of the speed of the two fronts occurs via local elastic pulses that originate from plastic deformation through the elastic precursor to the leading shock front. The ability to observe this rich dynamics and

its influence on reaching steady-state shock fronts was the result of a modeling technique in which the dynamics of the shocked material is followed in a window that travels with the shock front. The window allows long timescale phenomena to be observed compared to a simulation in which the reference frame is the undisturbed material in front of the initial shock compression.

Using the same simulation methodology, White and coworkers examined the evolution of plastic dynamics behind the second shock front in single-crystal aluminum that was shocked along different orientations (Zhakhovsky et al. 2012). It was found that despite direction-dependent properties such as melting dynamics, the systems evolved to the same point on the Hugoniot relations. This result helped to clarify what had been an apparent discrepancy between MDS, which was predicting orientation-dependent plastic dynamics, and experiment which was measuring Hugoniot relations that are orientation independent. Together with the other papers mentioned above, this work is an excellent example of how when properly set up and interpreted a modeling technique as basic as MDS can explain and predict a wider range of phenomena. It also provides an example of where insight and data generated by one technique can help inform other models.

The influence of preexisting defects on plastic deformation mechanisms and mechanical properties of metals under extremely high loading rates have been examined by numerous MDS. For example, in nanocrystalline solids, grain sizes are of the order of nanometers so that they contain a very large density of grain boundaries. This results in unique plastic deformation mechanisms (compared to systems with conventional grain sizes) in which grain boundary sliding and grain rotation accompanied by diffusion can play roles comparable to dislocation slip (this is discussed more below). The rapid temperature and pressure rise associated with a shock front can effectively remove deformation mechanisms requiring diffusion while at the same time increasing the shear modulus. This increases the threshold stress for inter-planar sliding. In addition, increasing pressure increases the threshold for dislocation plasticity. Considering these effects together with the assumption that the maximum in hardness as a function of grain size occurs where stress for sliding and dislocation plasticity become equal, Bringa and coworkers suggested that ultrahigh hardness can be achieved by shock loading nanostructured crystals (Bringa et al. 2005).

Bringa et al. modeled the shock loading of nanostructured copper with different grain sizes using MDS to validate their analysis (Bringa et al. 2005). At low shock loadings, grain boundary sliding occurred, which leads to a relatively low value of hardness that increases with increasing grain size. For intermediate shock loadings, they observed that hardness increased with increasing shock strength for each grain size, with the maximum hardness shifting toward smaller grain sizes compared to lower strain rate deformation. This results in an increase in the overall maximum hardness of the metal. At still higher shock loadings, the MDS produced a drop in strength due to an increase in the rate of dislocation nucleation and motion resulting from a temperature increase. At the highest shock pressures used in the MDS (150–220 GPa), shock-induced melting was observed with an associated low shear stress and further reduction in hardness.

In a recent set of studies, Dongare and coworkers modeled high strain rate deformation of nanocrystalline copper using MDS (Dongare et al. 2009, 2010). Copper with an average grain size of 6 nm was subject to uni- and triaxial tensile strain at a strain rate of 10^8 s^{-1} (Dongare et al. 2009) Random nucleation of nonspherical voids occurred at grain boundaries and/or triple junctions by creation of a shell of disordered atoms that surround the voids rather than by dislocation nucleation from the void surfaces. Void coalescence as the system was further strained was observed to occur by shearing of the disordered regions between individual voids. The fraction of voids was found to vary linearly with plastic strain in two distinct stages. In the first region, voids formed along the grain boundaries and triple junctions. In the second stage, which began after a minimum in tensile stress was reached, the disordered regions around the voids begin to recrystallize due to the higher temperatures compared to the first stage that is accompanied by void coalescence.

In another study, Dongare and coworkers used MDS to simulate shock loading of nanocrystalline copper with the same grain size containing a free surface. The system was initially compressed with an impact piston traveling with velocities ranging from 0.25 to 1 km/s (Dongare et al. 2010). This was followed in the simulation by unloading and spalls as the compression wave met the free surface. The same two regions of void growth and coalescence behavior were observed as in the high tensile strain rate studies. The first region where voids initially nucleate at grain boundaries and triple junctions was found to be strongly dependent on the initial sample structure and the piston impact velocity. Higher strain rates lead to creation of a greater number of voids that nucleate over a shorter time. In contrast, dynamics associated with the second regime depended only on the number of voids created during the first regime.

In another recent study, MDS was used to characterize the influence of twin spacing on shock-induced plasticity and spall behavior for twin boundaries inserted into single-crystal copper and twin boundaries inserted into nanocrystalline copper with an average grain size of 10 nm (Yuan and Wu 2012). Twin spacings between 0.6 and about 4 nm were used. It was found for the nanocrystalline system that the average flow stress increased with increasing twin spacing up until about 1 nm, after which the flow stress decreased with increasing spacing between the twins. This trend resulted from two competing mechanisms of dislocation dynamics under shock loading. The first mechanism involves the intersection of the dislocation–twin boundary by inclined dislocations; the second mechanism was twin-boundary migration via parallel dislocations. In contrast, the twins were not found to influence the spall behavior of the nanocrystalline system. This is because voids nucleate and grow along the grain boundaries, while for the other system the twins provide void nucleation sites so that twinning density influences spall behavior.

Deformed Nanocrystalline Metals

Discussed in this subsection are representative MDS of the homogenous deformation (instead of shock loading) of metals with grain sizes on the nanometer scale.

In metals that have conventional grain sizes, plastic deformation involves primarily the motion of collections of dislocations (i.e., slip). Because in a given lattice slip systems will change orientation across a grain boundary, these boundaries tend to inhibited dislocation motion. This leads to an increase in strength with decreasing grain size. Both experiment and modeling suggest, however, that there is a grain size below which materials begin to get softer with decreasing grain size. This "inverse Hall–Petch" behavior can be attributed to a change from dislocation mediated plasticity to grain boundary sliding. Such a transition has been observed in MDS, which predict a critical grain size of about 10–15 nm in good agreement with experimental results. While agreeing with experiment, the MDS have also revealed rich and unanticipated dynamics near the threshold region that can apparently be related to the inherent properties of the bulk material. This dynamics includes an enhanced role of grain rotation, cooperative inter-grain dynamics, and stacking fault formation across grains that occurs by motion of partial dislocations.

MDS of strained nanoscrystalline copper with average grain sizes of ~5 nm was carried out by Jacobsen and coworkers. There studies showed softening for small grain sizes, in qualitative agreement with experimental results (Schiøtz et al. 1999). These MDS studies revealed that in the inverse Hall–Petch regime, plastic deformation can be chiefly attributed to grain boundary sliding and a reduced influence of dislocation motion on the deformation. In related MDS carried out by Van Swygenhoven and coworkers, the deformation mechanisms of nanostructured nickel and copper with grain sizes ranging between 3.5 and 12 nm were characterized (Van Swygenhoven et al. 1999). For grain sizes less than ~10 nm, deformation occurred mainly by grain boundary sliding, while at larger grain sizes, deformation occurred by a combined sliding and dislocation motion. Mechanisms of strain accommodation later identified by MDS included single atom motion and correlated motion of several atoms, as well as stress-assisted free volume migration.

Similar studies of the deformation of nanocrystalline aluminum with a columnar structure carried out by Wolf and coworkers showed the emission of partial dislocations originating from grain boundaries and triple junctions (Yamakov et al. 2001). An interesting observation from these simulations is that partial dislocations may be reabsorbed by a grain boundary after the applied stress is removed. It was suggested that this may be responsible for the lack of dislocations observed in experiments after release of external stresses.

Based on MDS studies, Van Swygenhoven and coworkers suggested that atom rearrangements at grain boundaries at points where nucleation and emission of partial dislocation occur lower the grain boundary energy. Emission of a trailing partial dislocation is therefore not always needed for further relaxation. Based on MDS of aluminum with a columnar nanostructure, Yamakov et al. suggested that the width of stacking faults (hence the intrinsic stacking fault energy) determined by the distance separating two partial dislocations is the main quantity defining the transition from full to partial dislocation emission as grain sizes approach the critical size where inverse Hall–Petch behavior is first observed (Haslam et al. 2004). In subsequent work, Van Swygenhoven and coworkers pointed out that stacking fault energies for nickel, copper, and aluminum do not show a strong

correlation with partial dislocation emission (Van Swygenhoven et al. 2004). It was instead suggested that full dynamics associated with partial dislocation nucleation from a grain boundary should be considered. This further suggests that the full planar fault energy, including the stable and unstable stacking fault energy and twin fault energies, must be considered to understand and predict relations between grain size and plastic deformation.

Using MDS, Farkas and Curtin characterized detailed mechanisms for the emission of dislocations within a columnar nickel nanostructure with grain sizes of 4–20 nm (Farkas and Curtin 2005). Apparent from these simulations were dislocations being emitted from grain boundaries. The simulations also indicated that the number of emitted dislocations per unit length of grain boundary saturates with increasing grain sizes. Assuming that dislocation emission occurs only at some small distance from triple junctions, they reproduced the dislocation densities observed in the MDS. Complicating these observations with respect to experiment is the timescale of the mechanisms needed to nucleate dislocation emission relative to the short timescale of the simulations.

"Dimples" have been observed experimentally on the fracture surfaces of nanocrystalline metals that have grain sizes less than about 100 nm (Wang et al. 2002). This observation is somewhat surprising because they appear to be fully compact materials and hence the dimples are not the result of voids or similar structural artifacts. The scale of the dimples is the size of multiple grains, which suggests that their formation mechanism involves inter-grain cooperative motion. Derlet et al. used MDS to characterize the fracture dynamics of nanostructured nickel with a mean grain size of 6 nm and a narrow grain size distribution (Derlet et al. 2003). The MDS was carried out at almost one half of the melting temperature of pure nickel to accelerate grain boundary sliding and diffusion, and a tensile load of 1.5 GPa was applied for 350 ps. Formation of local shear bands was observed. These shear bands extended through multiple grains and accounted for the cooperative dynamics. Three mechanisms were identified that contributed to the shear band formation. These were reorientation of neighboring grains separated by low-angle grain boundaries, migration of grains by boundary sliding that lined up complementary shear planes, and intragranular slip. Distributed within the microstructure were interfaces such as twins that are particularly resistant to sliding. These interfaces created pinning points around which surrounding grains deform and form the shear bands. This was proposed as the mechanism that leads to the dimples observed experimentally.

Different groups have used MDS to characterize the possible role of dopants at grain boundaries on plastic deformation mechanisms. Saxens and coworkers, for example, modeled the influence of antimony dopants in nanocrystalline copper on plastic dynamics under uniaxial load (Rajgarhia et al. 2010). The simulations revealed that relatively small amounts of antimony can increase the flow stress; this was attributed to an increase in the stress needed for grain boundary sliding. It was also observed that antimony does not appear to have a large influence on dislocation nucleation at grain boundaries. In contrast, results from MDS by Jang et al. suggested that lead atoms that segregate to grain boundaries in nanostructured aluminum can suppress the nucleation of partial dislocations from grain boundaries

that are under uniaxial tension (Jang et al. 2008a, b). In the pure system, dislocations were observed to nucleate from sites at which aluminum atoms are already under tensile stress. Segregation to grain boundaries for this system arises mainly from stress effects that drive the larger lead atoms (compared to aluminum) to atom sites that in pure aluminum would be under tensile stress. Hence, the Pb relieves the stress that contributes to dislocation emission. It was also observed that lead addition disordered the atoms at the grain boundaries. This allowed more strain to be accommodated by the grain boundaries compared to the pure system that had dislocations travel through the grains.

In ductile materials, crack propagation involves dislocation emission from the crack tip. This emission can increase the tip radius and decrease the stress concentration that drives the tip forward under an applied load. While the stresses away from the crack can be well modeled by continuum-based methods, atomic theory and modeling is needed to understand the processes occurring at the crack tip. In principle MDS can be effectively used for this purpose. In practice, however, modeling the scale of the plastic dynamics leading to ductile fracture and its coupling to the surrounding stress field has many challenges (Brenner 2013).

Holian and coworkers used MDS to simulate crack propagation in a copper crystal that had interatomic forces modeled by Morse or EAM potentials (Zhou et al. 1997). An initial crack was introduced into the material, and it was strained to just below the Griffith criterion for crack propagation. The long-range character of the various stress and strain fields was incorporated into the simulation using periodic boundaries in two directions. Atoms were displaced according to the continuum elastic solution for a finite crack in an unbounded solid. This large MDS simulation – it involved up to 35 million atoms – was the first to unambiguously observe dislocation loops emitting from a crack tip that lead to blunting and jogging. The MDS revealed crack tip propagation of about two lattice spacings that was followed by distinct elastic rounding of the tip just prior to dislocation emission. Interestingly, the initial partial dislocation was emitted backward from the crack tip. This was followed by a second forward emission. This behavior can be explained by the orientation of Burger's vectors for the dislocations with respect to the surrounding lattice and stresses. The details of crack blunting and plasticity revealed by this MDS simulation are beyond what typically comes from analytic theory and continuum-based modeling.

Hess et al. subsequently used MDS to examine the role of temperature, interatomic potential, strain rate, system size, and boundary conditions on dislocation nucleation from a crack tip in a system modeling nickel (Hess et al. 2005). The primary factor in the mechanism of dislocation emission revealed by the MDS is the crack orientation with respect to the surrounding lattice and temperature, and the interatomic force expression played a relatively minor role in terms of the qualitative behavior of plastic dynamics. Dislocation nucleation in response to local stresses that is thermally activated was observed, which explains the temperature dependence of the results.

MDS has also been used to characterize mechanisms associated with mode I crack propagation from a notch placed in a nanostructured system with grain sizes

between 5 and 12 nm and an interatomic PEF intended to model nickel (Farkas et al. 2002). Incremental loading of the system normal to the notch orientation followed by relaxation resulted in propagation of the notch through the nanostructure. For the smaller grain size, the crack propagated almost exclusively along the grain boundaries before being blocked at a triple junction. Adding strain resulted in blunting but not crack propagation. Nanovoids also formed along the grain boundaries in front of the crack tip. The voids eventually coalesced with the crack, which resulted in the crack front moving to the next triple junction, where this process was repeated. The same general behavior occurred for the larger grain sizes, except that the simulation was initiated with part of the crack tip located in a grain. In this simulation, the tip propagation through the grains was accompanied by emission of dislocations similar to crystalline materials, with the exception that the dislocations could be re-adsorbed by surrounding grain boundaries. After the crack tip went through an initial grain, the propagation mechanism was the same as that for the smaller-grained system. Comparison of the energy released to the Griffith criterion for fracture suggested that this dynamics corresponded to a brittle material.

Simulations of Grain Boundary Migration

Grain boundary motion plays a critical role in processes such as grain growth during recrystallization after a metal has been work hardened. How complex atomic dynamics is coupled to stresses and external applied strains as grain boundaries move remains one of the major unresolved problems in materials science (Molodov et al. 2007) (Gottstein and Molodov 1998). In most annealing experiments, grain growth represents an average behavior of many types of grain boundaries. However, different grain boundaries can behave very differently, and therefore, grain growth experiments are limited in terms of characterizing the dynamics of individual grain boundaries. Even in bicrystals it is a formidable task to characterize migration over a wide range of interface structure (Furtkamp et al. 1998; Gorkaya et al. 2009). Within the scale and interatomic potential constraints discussed above, MDS can provide unique insight and atom-resolved properties of specific grain boundaries that complement experimental capabilities (Olmsted et al. 2009).

Mobility is a fundamental property of interest in characterizing grain boundary dynamics. Mobility is defined as the linear coefficient relating the grain boundary normal velocity v_{GB} to the driving force (pressure) P, i.e., $M \equiv v_{GB}/P$. For low-angle tilt grain boundaries, the mobility can be analytically approximated by the number of dislocations per unit length times the Peach–Koehler force associated with each dislocation. This results in an analytic mobility expression that is inversely proportional to tilt angle and that has an Arrhenius temperature dependence. While activated mobility is typically observed, the inverse relation between mobility and tilt angle is not a feature of typical experimental or MDS results. At higher tilt angles, assuming that grain boundary motion is closely related to atom diffusion leads to the Burke–Turnbull expression for grain boundary velocity of the form

$$v = \frac{\Omega}{a^2}\left[\eta e^{-\frac{Q}{kT}} - \eta e^{\frac{Q+p\Omega}{kT}}\right] \tag{3}$$

In this expression, a is the lattice spacing, η is an attempt frequency, Ω is the volume per atom, Q is an activation energy, k is the Boltzmann constant, T is temperature, and p is pressure. In the limit that $p\Omega >> kT$, velocity and pressure are proportional to one another with proportionality constant (i.e., mobility):

$$M = \frac{\eta a^4}{kT} e^{-\frac{Q}{kT}} \tag{4}$$

In this expression, the mobility is an activated process but is independent of tilt angle.

Based on results from MDS and experiment, it appears that grain boundary mobilities can depend on many factors, but with a dependence that is not well understood. However, several trends have become apparent. First, mobility typically shows Arrhenius behavior with respect to temperature; this agrees with the expression above. Second, low-angle grain boundaries tend to be less mobile compared to high-angle structures, a trend that in general appears valid for both twist and tilt structures (Huang and Humphreys 2000). In addition, in contrast to the analytic theory mentioned above, the mobility of low-angle grain boundaries tends to increase with increasing misorientation angle in a power law relationship $M = k\theta^{\alpha}$ (Huang and Humphreys 2000). Third, the mobility of twist structures tends to be higher than tilt grain boundaries, while activation energies are similar (Godiksen et al. 2008). Finally, mobilities are different for planar and curved grain boundaries and very sensitive to impurity and vacancy concentrations (Gottstein 2009).

Foiles and coworkers used the artificial driving force method mentioned above in an MDS (Janssens et al. 2006) to calculate energies and mobilities for 388 grain boundaries in nickel. The interfaces included ⟨111⟩, ⟨100⟩ twist and ⟨110⟩, ⟨111⟩, ⟨100⟩ symmetric tilt and coherent twin grain boundaries (Olmsted et al. 2009). Greater than 25 % of the grain boundaries were reported to migrate using a mechanism that couples shear stress to motion. This mechanism included most of the non- Σ3 structures with the highest mobilities. Among the rest of the structures, the incoherent Σ3 twins had an anomalously high mobility, while other boundaries remained static on the timescale of the simulations. Thermal activation energies also varied widely, with some migration mechanisms being temperature independent. Thermal roughening of the grain boundaries was also reported, which resulted in relatively large increases in grain boundary mobility above the roughening temperatures. Despite the large number of structures studied and the wide range of different properties examined, no correlations were uncovered between mobility and scalar quantities such as grain boundary energy, misorientation angle, Σ value, or excess volume.

The definition of grain boundary mobility given above assumes a linear relationship between the normal velocity of a grain boundary and the driving force. However, both MDS and analytic theory (cf. Eq. 3) predict nonlinear behavior

(Zhou and Mohles 2011). Various authors have different explanations for this dependence. Godiksen et al. (2008) found in MDS that grain boundary velocity and driving force ($v_{GB} \propto P$) are proportional for twist grain boundary dynamics, but nonlinear ($v_{GB} \propto P^2$) for tilt grain boundaries. It was proposed that this result is due to local interactions between the grain boundaries and nearby dislocations. Another explanation proposed by Zhang et al. (2004) attributed the nonlinearity to an increase in effective activation barrier with increasing applied driving force.

Zhou and Mohles (2011) proposed a mechanism based on Eq. 3 that the approximation leading to Eq. 4 is invalid for high driving forces needed to observe grain boundary motion in MDS. They proposed that a low driving force limit can be achieved by extrapolating data from high driving force simulations. Zhou and Mohles used this idea to determine misorientation-angle-dependent grain boundary mobilities and migration activation energies by MDS with the artificial driving force method (Janssens et al. 2006). These simulations used a series of flat twist $\langle 110 \rangle$ grain boundaries with different Σ values and misorientations. The resultant mobilities of small- ($\leq 25°$) and large-angle misorientation grain boundaries were reported to be about $10^{-9} m^4 J^{-1} s^{-1}$ and $10^{-8} m^4 J^{-1} s^{-1}$, respectively.

To avoid the nonlinearity resulting from a high driving force, Trautt et al. performed equilibrium MDS and obtained the mobility according to the fluctuation dissipation theorem by assuming that mobility is linearly dependent on the fluctuation of the mean interface position (Fig. 2c) (Trautt et al. 2006). Based on their results, they concluded that mobilities of planar grain boundaries at the low driving force limit (i.e., true mobilities) are an order of magnitude higher than the mobilities measured with high driving force.

Like mobility, there is no simple rule to determine grain boundary migration activation energies because the energies depend on too many variables (e.g., grain boundary type, impurity concentration, and shape). For example, activation energies computed from MDS are often lower than those determined from experiment (Schönfelder et al. 2006; Zhou and Mohles 2011). One factor contributing to this observation is likely the impurities in the experimental systems that are not present in the simulations (Olmsted et al. 2009).

A distinct change in activation energy between low-angle and high-angle grain boundaries (irrespective of planar or curved type) was detected numerically (Zhou and Mohles 2011) and experimentally (Winning and Gottstein 2002) at approximately the same misorientation. The transition angle measured is about 15° for both $\langle 110 \rangle$ twist (Zhou and Mohles 2011) and $\langle 100 \rangle$ twist boundaries (Schönfelder et al. 2006), 14.1° for $\langle 111 \rangle$ symmetric tilt boundaries (Winning amd Gottstein 2002), 8.6° for $\langle 100 \rangle$ symmetric tilt boundaries (Winning 2003), and 13.6° for $\langle 112 \rangle$ symmetric tilt (Winning and Rollett 2005) grain boundaries. The former two transition angles were determined from MDS, while the latter three were measured experimentally. The agreement suggests that MDS is able to capture the important aspects of this transition.

Equation 3 assumes that activation energies for grain boundary migration in the high-angle regime does not depend on misorientation angle (Winning and Rollett 2005; Winning 2003). This is different from grain boundary energies, which show a

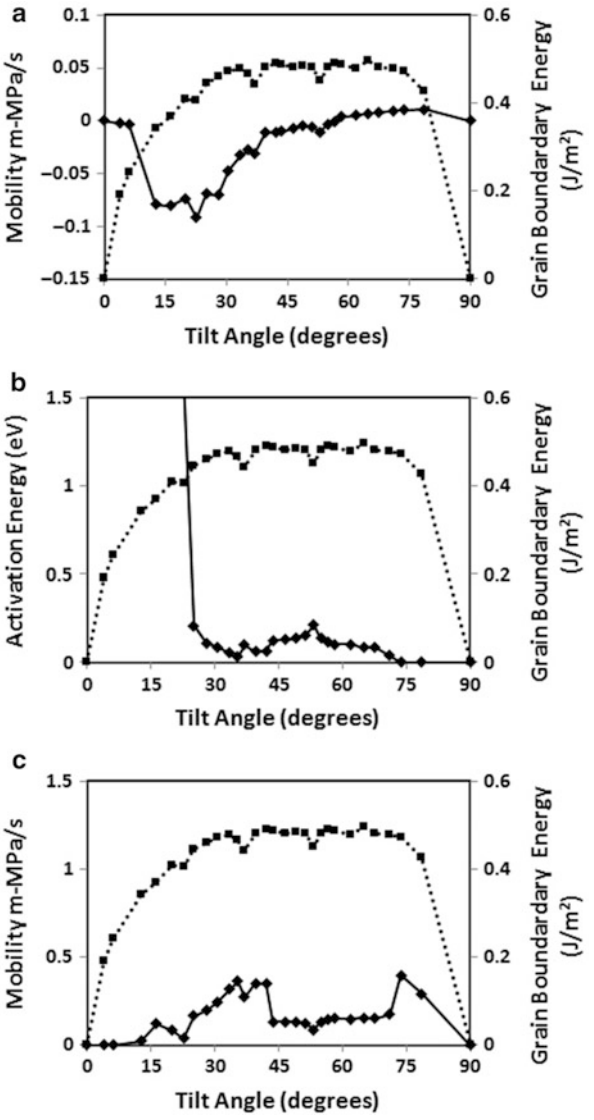


Fig. 5 Dependence of various grain boundary properties on tilt angle from an MDS of ⟨100⟩ symmetric tilt grain boundaries in copper at temperature ≥800 K and a virtual driving force of 10^8 Pa. In each plot the dashed line indicates grain boundary energy at 0 K (the scale is indicated by the label along the right ordinate of each plot). (**a**) mobility; (**b**) activation energy; (**c**) shear strain. There appear to be correlations between energy and the dynamic properties of the grain boundaries, but the origin of the correlations is unknown

distinct dependence on angle. This includes energy cusps at low sigma structures in symmetric tilt boundaries. Plotted in Fig. 5 are grain boundary energy and migration activation energy as a function of tilt angle from MDS using an embedded-atom potential for copper. In this unpublished study, a correlation between grain boundary energy and activation energy is apparent, especially with respect to the cusps. However, whether cusps are local maxima or minima does not appear to correlate. More work to try to understand this relationship is underway.

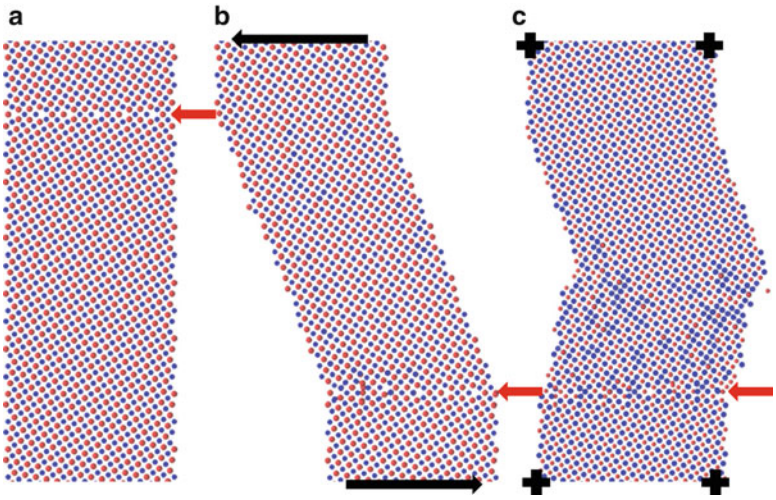


Fig. 6 Illustration of atom motion from an unpublished MDS of a tilt grain boundary migration. The image on the *left* is the initial atom configuration. The *right-side* two images illustrate the final atom configurations when the grain boundary is driven to move downward (indicated by the *red arrows*) by shear stress (**b**) or normal direction driving force (**c**). The *black arrows* indicate the shear direction and the *black crosses* imply that shear motion on the two ends is prohibited. *Blue* and *red colors* specify atoms on (100) and (200) planes, respectively

A recent experiment indicated that that grain boundary migration in nanocrystalline aluminum thin films is governed by shear stresses that produce distortion work. This is in contrast to the normal stresses assumed in conventional theory (Rupert et al. 2009). Shear stress driving grain boundary motion has also been reported in other experimental work (Molodov et al. 2007; Winning and Rollett 2005; Winning 2003). MDS results agree with these experimental studies (Cahn et al. 2006; Cahn and Taylor 2004). MDS for symmetric $\langle 100 \rangle$ tilt grain boundary systems in copper by Cahn et al. confirmed that normal motion of planar grain boundaries can be driven by shear stress (Fig. 2b). This work also showed that the ratio of normal to tangential translation of a grain boundary is a constant that is independent of temperature or of the magnitude of the applied shear stress.

Olmsted et al. (2009) reported that they found that even under a normal driving force significant shear can be built up during normal motion of planar grain boundaries. They further suggested that a traditional diffusion-controlled mechanism exists in all grain boundaries, but this mechanism can be overshadowed by the faster shear-coupled mechanism if the latter is allowed geometrically (e.g., lateral translation is unconstrained).

A subset of atoms near a tilt grain boundary from an unpublished MDS study is illustrated in Fig. 6. An embedded-atom potential for copper was used, and the interface is a symmetric tilt grain boundary with a tilt angle of 22.62°. The position of the grain boundary is indicated by the red arrows. The left panel

illustrates the grain boundary region before a driving force for migration is applied. The middle panel illustrates the same subset of atoms after the grain boundary has moved due to a shear stress that is applied at the boundaries with an orientation given by the black arrows. Applying this stress precludes the use of periodic boundaries in the direction normal to the grain boundary interface (but they are used in the other two directions). The right panel illustrates the same atoms after the grain boundary has moved due to a driving force normal to the interface from the application of an artificial energy term as illustrated by Fig. 1a (Janssens et al. 2006). For this case, periodic boundaries are applied in all directions. In both cases, a shear strain in the plane of the grain boundary is created during the grain boundary migration. For the conditions corresponding to Fig. 6b, this shear strain is smoothly and continuously accommodated by nonperiodic boundary condition, and the grain boundary moves continuously (Cahn et al. 2006). For Fig. 6c where the boundary conditions do not allow shear strain accumulation (indicated by the black crosses), shear strain starts to build up as the grain boundary begins to move. The grain boundary motion, however, is stopped after about 100 time steps due to the accumulation of shear stress in the direction opposite of the strain. This corresponds to the inflection point of the strain in panel c. If the normal direction driving force is large enough, the grain boundary can overcome this shear stress accumulation and continue moving, but through a different mechanism involving a combination of dislocation glide and climb. Further analysis of this mechanism is ongoing.

Current Challenges

Given in this chapter was an overview of MDS at a tutorial level, followed by some examples where MDS is providing new insights into the formation, stability, and structure of various types of damage in metals. The intent of this chapter was not to be a comprehensive review, but rather a snapshot of the scientific literature in this area. Within these limited goals, it should be clear that MDS has been used successfully to study a wide variety of different types of damage in crystalline metals. Even with the advances in computing resources and methodology, however, challenges remain to using MDS to fully understand plastic damage. Some of these challenges, and steps toward meeting these challenges, are briefly described in this section.

Interatomic Forces

PEFs commonly used in MDS were briefly discussed above. The most successful functional forms for large-scale MDS are generally derived from approximations to quantum mechanical bonding, with parameters in the functions fit to a set of physical properties. In some cases, researchers favor functional forms with many parameters and then fit these parameters to as many properties as they can. The other extreme is to create a PEF with as much physics as possible built into a

functional form that has fewer parameters and then fit these parameters to a fewer, carefully chosen properties. Both approaches have advantages and limitations that can become apparent in analyzing simulation results and comparing these results to experimental observations.

There are currently improvements being made to the derivation, fitting, and implementation of PEFs. For example, one active area is incorporating effects related to charge transfer (Mathieu 2007; Nistor et al. 2006; Nistor and Müser 2009). While unimportant for pure metals for most situations, including charge transfer improves the description of bonding for intermetallic phases and for interfaces between metals and other types of materials (e.g., oxides). There have also been recent advances in incorporating magnetic effects into PEFs, although much more work is needed (Ackland 2006; Dudarev and Derlet 2005; Ma et al. 2012). Replacing predefined functions with neural networks and similar approaches is showing promise in enhancing the accuracy of interatomic forces for MDS (Behler 2011; Jos et al. 2012).

In the longer term, the preferred approach from a fundamental science viewpoint is to calculate interatomic forces directly from quantum mechanics. Depending on the approximations made within density functional theory, for example, a single calculation that includes electrons can be carried out for between tens and millions of atoms with the current level of computing resources. With better approximations within quantum mechanics, increasingly clever scaling algorithms, and simply larger and faster computers, the system sizes for which forces can be calculated with explicit electronic states will continue to dramatically increase. In addition to interatomic forces that are transferable compared to analytic PEFs, using forces directly from quantum mechanics should provide an efficient pathway for relating plastic damage in metals to the properties and behavior of the electrons.

Length Scales

System sizes scale naturally with the number of processors (rather than processor speed) dedicated to a simulation. In the last two decades the size scales that can be attained in an MDS have increased by orders of magnitude. This is especially true for researchers with access to large computers like those at the national laboratories. In addition, the cost of parallel processors has decreased to the point where research groups with modest resources can model damage of the size associated with extended structures like grain boundaries and dislocations. More researchers working on this scale ensure more studies that can probe behavior and structure with atomic resolution.

Timescales

Because the classical equations of motion are numerically solved stepwise in a serial fashion, the prospects for simulating significantly longer times from a

computational resources viewpoint are strongly coupled to increases in processor speed. Maintaining Moore's law far into the future is unclear, and so other methods will be needed to increase accessible timescales. For rare events or well-defined kinetics, the methods mentioned above have been successful in this regard, and various ways of combining kinetic Monte Carlo and molecular dynamics simulations will also contribute to increased accessible timescales. Overall this remains a major challenge to connecting atomic and engineering scales.

Quantum Dynamics

Using classical mechanics to model atom motion is a more reasonable approximation for heavy atoms (i.e., most metals) than light atoms like hydrogen, which would have a much higher probability for quantum tunneling and larger zero point energies. For metals accurate heat capacities and thermal transport is inhibited by the assumption of classical mechanics as well as the inability to explicitly treat of electronic degrees of freedom that can carry heat. Compared to other challenges, this remains an unmet but otherwise low priority challenge for MDS as applied to metal damage.

Interpretation of MDS Results

New algorithms have recently been proposed with which damage in crystalline lattices can be easily identified. This includes dislocations and Burger's vectors (Stukowski and Albe 2010). There still remains, however, a crucial need for automating the detection and interpretation of damage mechanisms in MDS and passing this information to larger length-scale analyses. As an example, the methods of advanced statistics (Bayesian analysis) could be used to analyze simulations so that different types of damage gleaned from multiple simulations can be passed to higher scale analyses. Other critical and innovative thinking is needed in this area.

Conclusion

This chapter provided an overview of MDS as applied to modeling plastic damage in metals. This overview included a tutorial-level explanation of how a traditional MDS is implemented and analyzed and some examples of where MDS had provided new insights into the dynamics of damage that cannot typically be obtained from continuum modeling that is more traditionally used by engineers. This chapter concluded with a list of challenges that remain in expanding the use of MDS for modeling damage. These challenges included improved interatomic force expressions, numerical and computational routes for increasing time and length scales, incorporating quantum dynamics where needed, and new methods for interpreting MDS, including using modern tools of statistical analysis.

Acknowledgments Unpublished work reported from our group was supported by the Office of Naval Research. DL is supported by the National Science Foundation.

References

G.J. Ackland, Two-band second moment model for transition metals and alloys. J. Nucl. Mater. **351**(1–3), 20–27 (2006)

B.J. Alder, T.E. Wainwright, Phase transition for a hard sphere system. J. Chem. Phys. **27**(5), 1208 (1957)

M.P. Allen and D.J. Tildesley, Computer Simulation of Liquids (Oxford University Press, New York, 1989) ISBN-10: 0198556454

J. Behler, Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. Phys. Chem. Chem Phy. PCCP **13**(40), 17930–17955 (2011)

H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath. J. Chem. Phys. **81**(8), 3684 (1984)

G.P. Berman, F.M. Izrailev, The Fermi-Pasta-Ulam problem: fifty years of progress. Chaos. (Woodbury) **15**(1), 15104 (2005)

D.W. Brenner, The art and science of an analytic potential. Phys. Status Solidi B **217**(1), 23–40 (2000)

D.W. Brenner, Challenges to marrying atomic and continuum modeling of materials. Curr. Opinion Solid State Mater. Sci. **17**(6), 257–262 (2013)

D.W. Brenner and B.J. Garrison, Gas-Surface Reactions: Molecular Dynamics Simulations of Real Systems, in Adv. Chem. Phys, (Wiley, New York, K.P. Lawley, Ed.) Vol. 76, pp. 281–333 (1989)

D.W. Brenner, O.A. Shenderova, D.A. Areshkin, Quantum-based analytic interatomic forces and materials simulation. Rev. Comput. Chem. **12**, 207–239 (1998)

E.M. Bringa et al., Ultrahigh strength in nanocrystalline materials under shock loading. Science (New York) **309**(5742), 1838–1841 (2005)

J.W. Cahn, J.E. Taylor, A unified approach to motion of grain boundaries, relative tangential translation along grain boundaries, and grain rotation. Acta Mater. **52**(16), 4887–4898 (2004)

J.W. Cahn, Y. Mishin, A. Suzuki, Coupling grain boundary motion to shear deformation. Acta Mater. **54**(19), 4953–4975 (2006)

R. Car, M. Parrinello, Unified approach for molecular dynamics and density-functional theory. Phys. Rev. Lett. **55**(22), 2471–2474 (1985)

J.W. Crill, X. Ji, D.L. Irving, D.W. Brenner, C.W. Padgett, Atomic and multi-scale modeling of non-equilibrium dynamics at metal–metal contacts. Model. Simul. Mater. Sci. Eng. **18**(3), 034001 (2010)

J. D. Schall, C.W. Padgett, D.W. Brenner, Ad hoc continuum-atomistic thermostat for modeling heat flow in molecular dynamics simulations. Mol. Simul. **31**(4), 283–288 (2005)

M. Daw, M. Baskes, Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals. Phys. Rev. B **29**(12), 6443–6453 (1984)

M.S. Daw, S.M. Foiles, M.I. Baskes, The EAM is reviewed in: mater. Sci. Rep. **9**, 251 (1993)

P. Derlet, A. Hasnaoui, H. Van Swygenhoven, Atomistic simulations as guidance to experiments. Scr. Mater. **49**(7), 629–635 (2003)

A. Dongare, A. Rajendran, B. LaMattina, M. Zikry, D. Brenner, Atomic scale simulations of ductile failure micromechanisms in nanocrystalline Cu at high strain rates. Phys. Rev. B **80** (10), 104108 (2009)

A.M. Dongare, A.M. Rajendran, B. LaMattina, M.A. Zikry, D.W. Brenner, Atomic scale studies of spall behavior in nanocrystalline Cu. J. Appl. Phys. **108**(11), 113518 (2010)

A.M. Dongare et al., An angular-dependent embedded atom method (A-EAM) interatomic potential to model thermodynamic and mechanical behavior of Al/Si composite materials. Model. Simul. Mater. Sci. Eng. **20**(3), 035007 (2012)

S.L. Dudarev, P.M. Derlet, A 'magnetic' interatomic potential for molecular dynamics simulations. J. Phys. Condens. Matter **17**(44), 7097–7118 (2005)
V. Duin, C.T. Adri, S. Dasgupta, F. Lorant, W.A. Goddard, ReaxFF: a reactive force field for hydrocarbons. J. Phys. Chem. A **105**(41), 9396–9409 (2001)
V. Dupont, T.C. Germann, Strain rate and orientation dependencies of the strength of single crystalline copper under compression. Phys. Rev. B **86**(13), 134111 (2012)
F. Ercolessi, J.B. Adams, Interatomic potentials from first-principles calculations: the force-matching method. Europhys. Lett. (EPL) **26**(8), 583–588 (1994)
F. Ercolessi, E. Tosatti, M. Parrinello, Au (100) surface reconstruction. Phys. Rev. Lett. **57**(6), 719–722 (1986)
D. Farkas, W.A. Curtin, Plastic deformation mechanisms in nanocrystalline columnar grain structures. Mater. Sci. Eng. A **412**(1–2), 316–322 (2005)
D. Farkas, H. Van Swygenhoven, P. Derlet, Intergranular fracture in nanocrystalline metals. Phys. Rev. B **66**(6), 060101 (2002)
M.W. Finnis, J.E. Sinclair, A simple empirical N -body potential for transition metals. Philos. Mag. A **50**(1), 45–55 (1984)
M. Furtkamp, G. Gottstein, D.A. Molodov, V.N. Semenov, L.S. Shvindlerman, Grain boundary migration in Fe–3.5 % Si bicrystals with [001] tilt boundaries. Acta Mater. **46**(12), 4103–4110 (1998)
J.D. Gale, A.L. Rohl, The general utility lattice program (GULP). Mol. Simul. **29**(5), 291–341 (2003)
J. Gibson, A. Goland, M. Milgram, G. Vineyard, Dynamics of radiation damage. Phys. Rev. **120** (4), 1229–1253 (1960)
R.B.N. Godiksen, S. Schmidt, D. Juul Jensen, Molecular dynamics simulations of grain boundary migration during recrystallization employing tilt and twist dislocation boundaries to provide the driving pressure. Model. Simul. Mater. Sci. Eng. **16**(6), 065002 (2008)
T. Gorkaya, D.A. Molodov, G. Gottstein, Stress-driven migration of symmetrical ⟨100⟩ tilt grain boundaries in Al bicrystals. Acta Mater. **57**(18), 5396–5405 (2009)
G. Gottstein, D.A. Molodov, Grain boundary migration in metals: recent developments. Inter. Sci. **22**, 7–22 (1998)
G. Gottstein, L.S. Shvindlerman, *Grain Boundary Migration in Metals: Thermodynamics, Kinetics, Applications*. Materials Science & Technology, 2nd edn. (CRC Press, Boca Rotan, 2009)
A.J. Haslam et al., Effects of grain growth on grain-boundary diffusion creep by molecular-dynamics simulation. Acta Mater. **52**(7), 1971–1987 (2004)
B. Hess, B. Thijsse, E. Van der Giessen, Molecular dynamics study of dislocation nucleation from a crack tip. Phys. Rev. B **71**(5), 054111 (2005)
J. Hirschfelder, H. Eyring, B. Topley, Reactions involving hydrogen molecules and atoms. J. Chem. Phys. **4**(3), 170 (1936)
B.L. Holian, P.S. Lomdahl, Plasticity induced by shock waves in nonequilibrium molecular-dynamics simulations. Science **280**(5372), 2085–2088 (1998)
J.D. Honeycutt, H.C. Andersen, Molecular dynamics study of melting and freezing of small Lennard-Jones clusters. J. Phys. Chem. **91**(19), 4950–4963 (1987)
Y. Huang, F.J. Humphreys, Subgrain growth and low angle boundary mobility in aluminium crystals of orientation {110}⟨001⟩. Acta Mater. **48**(8), 2017–2030 (2000)
D.L. Irving, C.W. Padgett, D.W. Brenner, Coupled molecular dynamics/continuum simulations of Joule heating and melting of isolated copper–aluminum asperity contacts. Model. Simul. Mater. Sci. Eng. **17**(1), 015004 (2009a)
D.L. Irving, C.W. Padgett, J.W. Mintmire, D.W. Brenner, Multiscale modeling of metal–metal contact dynamics under high electromagnetic stress: timescales and mechanisms for joule melting of Al–Cu asperities. IEEE Trans. Magn. **45**(1), 331–335 (2009b)
K. Jacobsen, J. Norskov, M. Puska, Interatomic interactions in the effective-medium theory. Phys. Rev. B **35**(14), 7423–7442 (1987)
S. Jang, Y. Purohit, D.L. Irving et al., Influence of Pb segregation on the deformation of nanocrystalline Al: insights from molecular simulations. Acta Mater. **56**(17), 4750–4761 (2008a)

S. Jang, Y. Purohit, D. Irving et al., Molecular dynamics simulations of deformation in nanocrystalline Al–Pb alloys. Mater. Sci. Eng. A **493**(1–2), 53–57 (2008b)
K.G.F. Janssens et al., Computing the mobility of grain boundaries. Nat. Mater. **5**(2), 124–127 (2006)
K.V. Jose, N.A. Jovan, J. Behler, Construction of high-dimensional neural network potentials using environment-dependent atom pairs. J. Chem. Phys. **136**(19), 194111 (2012)
K. Kadau, T.C. Germann, P.S. Lomdahl, B.L. Holian, Microscopic view of structural phase transitions induced by shock waves. Science (New York) **296**(5573), 1681–1684 (2002)
C. Kelchner, S. Plimpton, J. Hamilton, Dislocation nucleation and defect structure during surface indentation. Phys. Rev. B **58**(17), 11085–11088 (1998)
B.-J. Lee, M. Baskes, Second nearest-neighbor modified embedded-atom-method potential. Phys. Rev. B **62**(13), 8564–8567 (2000)
J. Li, AtomEye: an efficient atomistic configuration viewer. Model. Simul. Mater. Sci. Eng. **11**(2), 173–177 (2003)
P.-W. Ma, S.L. Dudarev, C.H. Woo, Spin–lattice-electron dynamics simulations of magnetic materials. Phys. Rev. B **85**(18), 184301 (2012)
D. Mathieu, Split charge equilibration method with correct dissociation limits. J. Chem. Phys. **127**(22), 224103 (2007)
J.A. McCammon, B.R. Gelin, M. Karplus, Dynamics of folded proteins. Nature **267**(5612), 585–590 (1977)
R.E. Miller, E.B. Tadmor, A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods. Model. Simul. Mater. Sci. Eng. **17**(5), 053001 (2009)
D. Molodov, V. Ivanov, G. Gottstein, Low angle tilt boundary migration coupled to shear deformation. Acta Mater. **55**(5), 1843–1848 (2007)
F. Müller-Plathe, A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity. J. Chem. Phys. **106**(14), 6082 (1997)
R. Nistor, M. Müser, Dielectric properties of solids in the regular and split-charge equilibration formalisms. Phys. Rev. B **79**(10), 104303 (2009)
R.A. Nistor, J.G. Polihronov, M.H. Müser, N.J. Mosey, A generalization of the charge equilibration method for nonmetallic materials. J. Chem. Phys. **125**(9), 094108 (2006)
S. Nosé, A unified formulation of the constant temperature molecular dynamics methods. J. Chem. Phys. **81**(1), 511 (1984)
D.L. Olmsted, E.A. Holm, S.M. Foiles, Survey of computed grain boundary properties in face-centered cubic metals – II: grain boundary mobility. Acta Mater. **57**(13), 3704–3713 (2009)
C.W. Padgett, D.W. Brenner, A continuum-atomistic method for incorporating Joule heating into classical molecular dynamics simulations. Mol. Simul. **31**(11), 749–757 (2005)
M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method. J. Appl. Phys. **52**(12), 7182 (1981)
S. Plimpton, Fast parallel algorithms for short-range molecular dynamics. J. Comput. Phys. **117**(1), 1–19 (1995)
A. Rahman, Correlations in the motion of atoms in liquid argon. Phys. Rev. **136**(2A), A405–A411 (1964)
R.K. Rajgarhia, D.E. Spearot, A. Saxena, Molecular dynamics simulations of dislocation activity in single-crystal and nanocrystalline copper doped with antimony. Metall. Mater. Trans. A **41**(4), 854–860 (2010)
R. Ravelo, B. Holian, T. Germann, P. Lomdahl, Constant-stress Hugoniostat method for following the dynamical evolution of shocked matter. Phys. Rev. B **70**(1), 014103 (2004)
T.J. Rupert, D.S. Gianola, Y. Gan, K.J. Hemker, Experimental observations of stress-driven grain boundary migration. Science (New York) **326**(5960), 1686–1690 (2009)
J.-P. Ryckaert, G. Ciccotti, H.J. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. **23**(3), 327–341 (1977)

C. Schäfer, H. Urbassek, L. Zhigilei, Metal ablation by picosecond laser pulses: a hybrid simulation. Phys. Rev. B **66**(11), 115404 (2002)
J. Schiøtz, T. Vegge, F. Di Tolla, K. Jacobsen, Atomic-scale simulations of the mechanical deformation of nanocrystalline metals. Phys. Rev. B **60**(17), 11971–11983 (1999)
B. Schönfelder, G. Gottstein, L.S. Shvindlerman, Atomistic simulations of grain boundary migration in copper. Metall. Mater. Trans. A **37**(6), 1757–1771 (2006)
T.-R. Shan et al., Second-generation charge-optimized many-body potential for Si/SiO_2 and amorphous silica. Phys. Rev. B **82**(23), 235302 (2010)
M.A. Shehadeh, E.M. Bringa, H.M. Zbib, J.M. McNaney, B.A. Remington, Simulation of shock-induced plasticity including homogeneous and heterogeneous dislocation nucleations. Appl. Phys. Lett. **89**(17), 171918 (2006)
S.B. Sinnott, D.W. Brenner, Three decades of many-body potentials in materials research. MRS Bull. **37**(05), 469–473 (2012)
F.H. Stillinger, Improved simulation of liquid water by molecular dynamics. J. Chem. Phys. **60**(4), 1545 (1974)
A. Stukowski, K. Albe, Dislocation detection algorithm for atomistic simulations. Model. Simul. Mater. Sci. Eng. **18**(2), 025016 (2010)
Z.T. Trautt, M. Upmanyu, A. Karma, Interface mobility from interface random walk. Science (New York) **314**(5799), 632–635 (2006)
H. Van Swygenhoven, M. Spaczer, A. Caro, Microscopic description of plasticity in computer generated metallic nanophase samples: a comparison between Cu and Ni. Acta Mater. **47**(10), 3117–3126 (1999)
H. Van Swygenhoven, P.M. Derlet, A.G. Frøseth, Stacking fault energies and slip in nanocrystalline metals. Nat. Mater. **3**(6), 399–403 (2004)
A.F. Voter, A method for accelerating the molecular dynamics simulation of infrequent events. J. Chem. Phys. **106**(11), 4665 (1997)
A. Voter, Parallel replica method for dynamics of infrequent events. Phys. Rev. B **57**(22), R13985–R13988 (1998)
Y.M. Wang, E. Ma, M.W. Chen, Enhanced tensile ductility and toughness in nanostructured Cu. Appl. Phys. Lett. **80**(13), 2395 (2002)
M. Winning, Motion of ⟨100⟩-tilt grain boundaries. Acta Mater. **51**(20), 6465–6475 (2003)
M. Winning, G. Gottstein, On the mechanisms of grain boundary migration. Acta Mater. **50**, 353–363 (2002)
M. Winning, A.D. Rollett, Transition between low and high angle grain boundaries. Acta Mater. **53**(10), 2901–2907 (2005)
M. Wojdyr, S. Khalil, Y. Liu, I. Szlufarska, Energetics and structure of ⟨001⟩ tilt grain boundaries in SiC. Model. Simul. Mater. Sci. Eng. **18**(7), 075009 (2010)
V. Yamakov, D. Wolf, M. Salazar, S.R. Phillpot, H. Gleiter, Length-scale effects in the nucleation of extended dislocations in nanocrystalline Al by molecular-dynamics simulation. Acta Mater. **49**(14), 2713–2722 (2001)
F. Yuan, X. Wu, Shock response of nanotwinned copper from large-scale molecular dynamics simulations. Phys. Rev. B **86**(13), 134108 (2012)
V.V. Zhakhovsky, M.M. Budzevich, N.A. Inogamov, I.I. Oleynik, C.T. White, Two-zone elastic–plastic single shock waves in solids. Phys. Rev. Lett. **107**(13), 135502 (2011)
V.V. Zhakhovsky, M.M. Budzevich, N. Inogamov, C.T. White, I.I. Oleynik, *Single Two-Zone Elastic–Plastic Shock Waves in Solids* (2012), pp. 1227–32
H. Zhang, M.I. Mendelev, D.J. Srolovitz, Computer simulation of the elastically driven migration of a flat grain boundary. Acta Mater. **52**(9), 2569–2576 (2004)
J. Zhou, V. Mohles, Mobility evaluation of <110> twist grain boundary motion from molecular dynamics simulation. Steel Res. Int. **82**(2), 114–118 (2011)
S. Zhou, D. Beazley, P. Lomdahl, B. Holian, Large-scale molecular dynamics simulations of three-dimensional ductile failure. Phys. Rev. Lett. **78**(3), 479–482 (1997)

Numerical Applications in Damage Induced Anisotropy in Low-Cycle Fatigue Modeling for Metals

16

Akrum Abdul-Latif

Contents

Abstract

Qualitative and quantitative studies are made emphasizing the concept of damage-induced anisotropy using the micromechanical model proposed shown in ▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials." The model deals with the plastic strain and local damage variables. In fact, they are examined at the crystallographic slip scale for FCC metallic polycrystals. The elastic behavior is initially assumed to be compressible and isotropic determined at the macroscopic scale. Due to the activation/deactivation concept, the anisotropic damaged behavior is adopted using a fourth-order damage tensor at the overall scale. Consequently, the

A. Abdul-Latif (✉)
Université Paris 8, Laboratoire d'Ingénierie des Systèmes Mécaniques et des Matériaux (LISMMA), Saint Ouen Cedex, France
e-mail: aabdul@iu2t.univ-paris8.fr

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_43

overall nonlinear behavior, notably the deactivation phase due to microcracks closure under complex cyclic loadings, is of particular interest in this chapter.

The model ability is demonstrated by a host of plastic predicted damaged behaviors of metallic polycrystals focusing on the unilateral damage and loading path effects on the multiaxial low-cycle fatigue (LCF) behavior. Actually, the model is tested under strain- and stress-controlled conditions describing the effects of the loading path complexity and the mean stress on the polycrystal LCF behavior. Finally, the model can successfully describe the LCF behavior of the Waspaloy at room temperature.

Introduction

Cyclic behavior of polycrystals is one of the major considerations in engineering structures and machine design since many of the constituent parts are subjected to repeated loading. Machine components are generally subjected to various complex cyclic loading paths leading to many research investigations. Currently, the elasto-inelastic behavior of metals shows an increasing maturity especially in the engineering theory of plasticity. Damage mechanics notably under cyclic loading with the self-consistent approaches represents a field of challenging for several theoretical developments. Some attempts have been conducted in describing the damaged-elasto-inelastic behavior of material under different loading paths (simple and complex) using micromechanical approach (Abdul-Latif and Saanouni 1994, 1996; Saanouni and Abdul-Latif 1996; Abdul-Latif et al. 1999; Chadli and Abdul-Latif 2005; Abdul-Latif and Chadli 2007). As given previously in ▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials," various damage categories have been described in the literature, such as creep damage, low-cycle fatigue, high-cycle fatigue, and brittle damage (Kachanov 1986; Lemaitre and Chaboche 1990; Lemaitre 1992; Voyiadjis and Kattan 1999 and many others). For polycrystalline metals, metallographic studies demonstrate that the damage is basically characterized by a progressive degradation due to the initiation and coalescence of microcracks.

The nonlinearity of material behavior is generally induced by plasticity and damage mechanics. It is well known that ductile polycrystalline metals usually fail as a result of nucleation, growth, and coalescence of microdamages. It is experimentally shown that the accumulation of microdamages has a tendency to form a localized damage due to plastic strain localization up to final structure failure. As shown in ▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Poly crystalline Metallic Materials," in several metallic materials, the kinematic strengthening is related to the creation of slipbands. The setting of these bands in the material induces undoubtedly an internal back stress in the grains leading accordingly to anisotropic behavior. Besides, TEM observations reveal strain localized in slipbands during cycling leading to an important dislocations density in these bands. Microstructural observations related to specimen outer surfaces show that crack initiation occurs in some slipbands as in Waspaloy. Thus, these

slipbands together with microcracks seem to be important factors leading to anisotropic behavior concerning the elastic and plastic strains.

An induced-oriented anisotropy phenomenon can experimentally be observed, in fatigue. In fact, microcracks may open or close depending on the applied loading direction. Thus, different responses can be observed for compression and tension loads which lead to damage deactivation behavior as in aluminum alloy (see Fig. 5, ▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials"). Theoretically, many approaches have been proposed since the last decade (e.g., Krajcinovic and Fonseka 1981; Ladevèze and Lemaitre 1984; Ortiz 1985; Yazdani and Schreyer 1990; Mazars and Pijaudier-Cabot 1989; Ju 1989; Ramtani 1990; Chaboche 1992, 1993; Hansen and Schreyer 1995; Halm and Dragon 1996; Yazdani and Karnawat 1997; Abdul-Latif and Mounounga 2009). Based on the discussion in ▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials," the micromechanical model of damaged-elasto-inelastic behavior for FCC polycrystals will be tested. Using the small strains assumption, the damage activation/deactivation effect on polycrystalline metals behavior in LCF will be described.

The aim of this chapter is to show the model ability in describing the deactivation phenomenon due to microcracks closure and its effect on the metals behavior. Hence, a host of cyclic plastic damaged behavior of polycrystalline metals is predicted illustrating the damage activation/deactivation and loading path effects on the multiaxial LCF behavior under different strain- and stress-controlled conditions. The corresponding nonlinearity is appropriately described by the model. Note that special emphasis will be focused on the biaxial cyclic loading paths especially the nonproportional ones such as tension-torsion with different out-of-phase angles showing the additional hardening and damage evolution. Quantitatively, the model can successfully describe the LCF behavior of the Waspaloy at room temperature.

Identification of the Model

This paragraph deals with the identification of the model constitutive equations. The model identification can be conducted based on two main steps: (i) determining the model constants and (ii) determining the choice of the microstructure. In this chapter, all the numerical simulations made use a random orientation distribution of 300 grains. Figure 1 demonstrates the standard inverse pole figure of this 300-grain aggregate with two preselected grains (No. 17 and 218). The choice of these two gains is based on their reasonable plastic deformation and then their damage. Moreover, their behavior points out a local heterogeneity which is, in general, far from the representative behavior of the polycrystal. Such heterogeneity is considered as an important and interesting feature of this type of modeling. The 12 octahedral slip systems with their number are presented in Table 1. Assumed to be a single-phase FCC type, the microstructure of this aggregate (i.e., the number and the orientation of the grains) is determined by the well-known Euler angles. Macroscopically, the initial isotropic elastic behavior of this distribution is proved

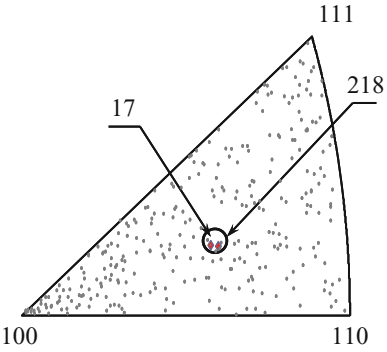


Fig. 1 Standard inverse pole figure of the used 300-grain aggregate (representative volume element) showing the two selected grains: 17 and 218 (From Mounounga et al. (2011), with permission from Elsevier)

Table 1 Definition of the 12 slip systems for FCC structure

Slip system	1	2	3	4	5	6	7	8	9	10	11	12
$\sqrt{3}n_1$	1	1	1	1	1	1	−1	−1	−1	1	1	1
$\sqrt{3}n_2$	1	1	1	−1	−1	−1	1	1	1	1	1	1
$\sqrt{3}n_3$	1	1	1	1	1	1	1	1	1	−1	−1	−1
$\sqrt{2}g_1$	−1	0	−1	−1	0	1	0	1	1	−1	1	0
$\sqrt{2}g_2$	0	−1	1	0	1	1	−1	1	0	1	0	1
$\sqrt{2}g_3$	1	1	0	1	1	0	1	0	1	0	1	1

by the model response (Fig. 2). This aggregate is actually an appropriate compromise between minimizing the calculation and acceptable description of the damaged polycrystalline microstructure.

In order to minimize the model complexity as well as the number of model constants, all the grains as well as all the slip systems are assumed to have the same material properties. Consequently, this gives that all the grains have the same constants (C^g and a^g) and all the slip systems have also the same plasticity and damage constants (z^s, K^s, k_o^s, Q^s, b^s, S^s, s_o^s, w^s, and γ_o^s). The hardening interaction matrix H_{rs} (considering only the octahedral slips) is defined by its constants. As shown above, only two different parameters (d_1 and d_2) define the damage interaction matrix D_{rs} (▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials," Eq. 38); the diagonal terms (d_1) determine the self-damage interaction (equal to 1) and the non-diagonal terms (d_2) describe the interaction between the local damage at the different slip systems within the same grain. Afterward, the qualitative simulations will be carried out using the identified constants summed up in Table 2 (Mounounga et al. 2011).

Various possibilities offered by the model are shown to describe the LCF behavior of FCC polycrystals, at the local and global scales. Thus, the model ability is demonstrated via the description of several phenomena concerning cyclic plasticity coupled with damage such as cyclic hardening evolution, fatigue life, damage deactivation effect, etc. Among these fundamental phenomena, the loading path complexity and mean stress effects on the damage evolution are particularly

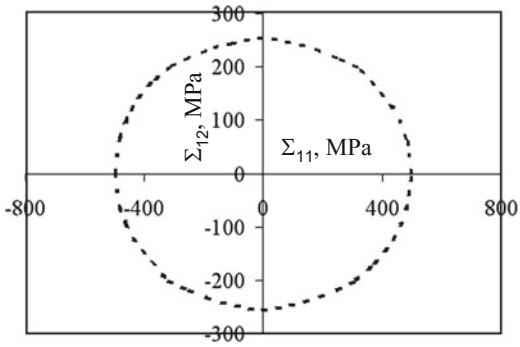


Fig. 2 Initial overall yield surface in the Σ_{11}–Σ_{12} space (From Mounounga et al. (2011), with permission from Elsevier)

Table 2 Constants of the model

E^g (MPa)	215,000	ν	0.3
C^g (MPa)	95,100	a^g	8.9
k_o^s (MPa)	240	z^s	20
K^s	40	Q^s (MPa)	260
b^s	12	h_3	0.8
h_2	0.8	h_5	1.0
h_4	1.5	h_6	3.5
w	1.0	α	3.0
γ_o^s	5.8	d_1	1.0
s_o^s	0.9	d_2	1.4

considered. Numerical tests show the damage deactivation effect (anisotropy induced by damage) especially under multiaxial loadings.

Various cyclic loadings (symmetrical and triangular) are used in this investigation under strain-controlled condition, namely, uniaxial tension-compression (TC); biaxial tension-torsion with various out-of-phase angles: $\Phi = 0°$ (TT00), $\Phi = 30°$ (TT30), $\Phi = 45°$ (TT45), $\Phi = 60°$ (TT60), $\Phi = 90°$ (TT90); and another biaxial cyclic loading of butterfly (Fly) (Fig. 3). Besides, several predictions are conducted under uniaxial tension-compression cyclic loading with stress-controlled condition of different positive mean stresses. They describe the ratcheting phenomenon coupled with damage. A quantitative study is also carried out by comparing the model predictions with experimental data for Waspaloy.

Loading Complexity Effect on the Damaged Behavior

Cyclic loading path effect on the LCF behavior is now examined. Many experimental results reveal the loading effect on the overall and local material behaviors especially for those metallic materials having low stacking fault energy. Indeed, cyclic hardening under nonproportional loading conditions becomes stronger than that under proportional ones. This is due to the rotation of the principal stress and

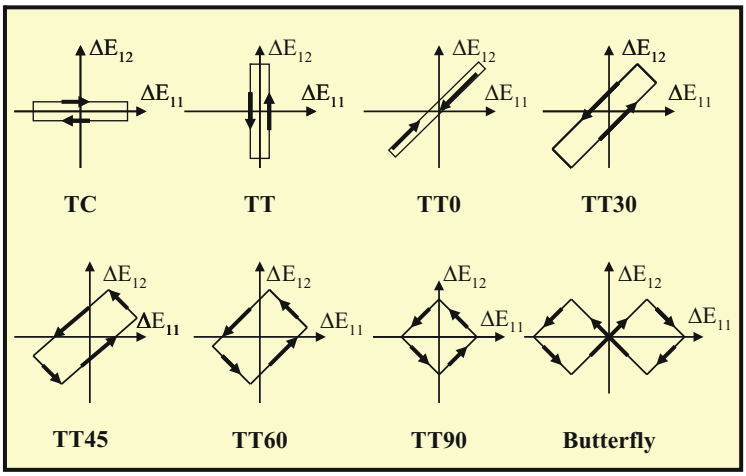


Fig. 3 Schematic representations of the employed cyclic loading paths in strain space (From Mounounga et al. (2011), with permission from Elsevier)

strain axes during nonproportional loading. This phenomenon is the so-called additional hardening. For many metallic materials, such as Waspaloy, different experimental studies show that this phenomenon is the result of slip system multiplication. Interactions between these activated slip systems induce, as a consequence, this additional hardening. Another example concerns to the austenitic stainless steel. In fact, the hardening rate increases significantly, when the material is submitted to multiaxial loading due to the slip multiplication. In cyclic tests of tension-torsion, it has been observed that additional hardening increases proportionally with the out-of-phase angle. This means that TT90 cyclic loading path (circular path with sinusoidal form) gives the maximum cyclic hardening. However, in the butterfly test, Waspaloy has experimentally pointed out that the induced additional hardening is more important than that in TT90 (Clavel et al. 1989).

Overall Behavior

LCF predictions under simple and complex loading paths are recorded and analyzed at both macroscopic and local levels. Numerical simulations are performed to justify the new intragranular damage initiation criterion proposed in Abdul-Latif and Mounounga (2009). This can be performed considering the significant differences between various fatigue lives observed experimentally in some materials. In fact, the main parameters α and γ_o^s of the new intragranular damage criterion (► Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials," Eq. 33) and their effect on the fatigue life are numerically investigated varying these two parameters as summarized in Table 3, where α is varied from 1 to 15 and γ_o^s from 10 to 120. Three different loading paths are used:

Table 3 Adopted scenario in studying the new intragranular damage criterion parameters and their interaction

		γ_o^s				
		10	30	60	90	120
α	1	X	X	X	X	X
	3	X	X	X	X	X
	6	X	X	X	X	X
	9	X	X	X	X	X
	12	X	X	X	X	X
	15	X	X	X	X	X

uniaxial TC with $\Delta E_{11} = 0.85\ \%$, biaxial TT90, and butterfly with $\Delta E_{11} = 0.736\ \%$ and $\Delta E_{12} = 0.37\ \%$ having the same maximum equivalent total strain of $\Delta E_{eq\text{-}vM} = 0.85\ \%$.

Fatigue life simulations are carried out with the two parameters variation. By means of this damage criterion, the model shows a strong dependency of polycrystal fatigue life on these parameters and their interaction (Fig. 4). This figure reveals that the fatigue life decreases when α increases and γ_o^s decreases. In fact, whatever the loading path for a given α, for example, $\alpha = 1$ in TC, the fatigue life changes remarkably between 356 cycles for $\gamma_o^s = 10$ up to 2,430 cycles when $\gamma_o^s = 120$. It is also recognized that for a fixed value of $\gamma_o^s = 10$, the reported fatigue lives in TT90 with $\alpha = 1$, 3, 6, 9, 12, and 15 are 94, 22, 12, 12, 12, and 12 cycles, respectively. Moreover, in Fly test, these are 52, 11, 9, 9, 9, and 9 cycles, respectively. Nonetheless, the model cannot converge, for multiaxial loading paths, for $6 \leq \alpha \leq 15$ for this value of γ_o^s. This means that the fatigue life is not sensitive in the range of $\alpha \geq 6$ whatever the value of γ_o^s. However, the model response becomes remarkably sensitive to α and γ and their interaction in the range of $\alpha \leq 6$ whatever the used value of γ_o^s as shown in Fig. 4.

An examination of Fig. 5 shows the dependence of the polycrystal fatigue life on the loading complexity. By analyzing the maximum von-Mises stress evolution versus the overall accumulated plastic strain up to macroscopic crack initiation, it is obvious that the model can suitably describe the three plastic fatigue stages: accommodation, stabilization, and softening due to damage. The intragranular damage rate $\dot{d}^s$ (▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials," Eq. 44) is principally the function of Y_{in}^s, λ^s, and D_{rs}. Hence, a kind of interaction among these three elements is considered. It plays a decisive role on the local and then on the overall fatigue behavior. Indeed, the greater the complexity of the loading path (increasing the number of activated plastic systems), the greater is the additional hardening (due to the evolution of R^s) and the shorter the fatigue life.

However, this rate is attenuated by the slip evolution λ^s due to the increase in the cyclic hardening which obstructs the damage evolution. It is clear that the butterfly loading (Fly) involves the shortest fatigue life ($N_f = 45$ cycles), because it is the most complex one giving therefore the highest overall cyclic stress (Fig. 5). Its steady state (stabilization) is the shortest due to the highest damage evolution

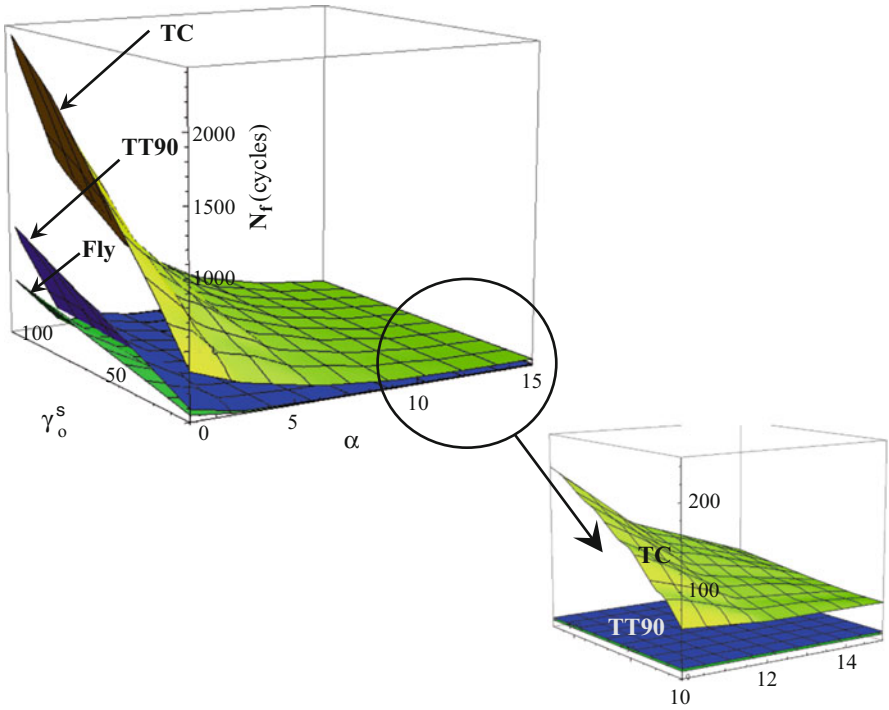


Fig. 4 Influence of α and γ_o^s parameters and their interaction on the evolution of the polycrystal fatigue life under different cyclic loading paths (From Mounounga et al. (2011), with permission from Elsevier)

controlled by the greatest additional hardening (i.e., a strong competition between the damage and hardening occurring).

On the other hand, the highest fatigue lives are obtained in TC and then in TT00 giving fatigue lives N_f of 839 cycles and 689 cycles, respectively. This observation is confirmed by the overall damage evolution analysis with respect to the overall accumulated plastic strain (Fig. 6). Qualitatively, these predictions are in accordance with several published experimental data for many materials.

Local Behaviors

One of the principal advantages of a micromechanical modeling is its ability to produce the principal cyclic plasticity phenomena through local considerations. In this paragraph, some recorded local responses for the different loading paths are illustrated. In fact, responses of the average number of activated slip systems per grain are illustrated in Fig. 7 emphasizing its dependency on the loading complexity. Note that this number is numerically determined by using a threshold in terms of slip ($\gamma^s = 10^{-4}$) before which counting of activated slip systems is no longer

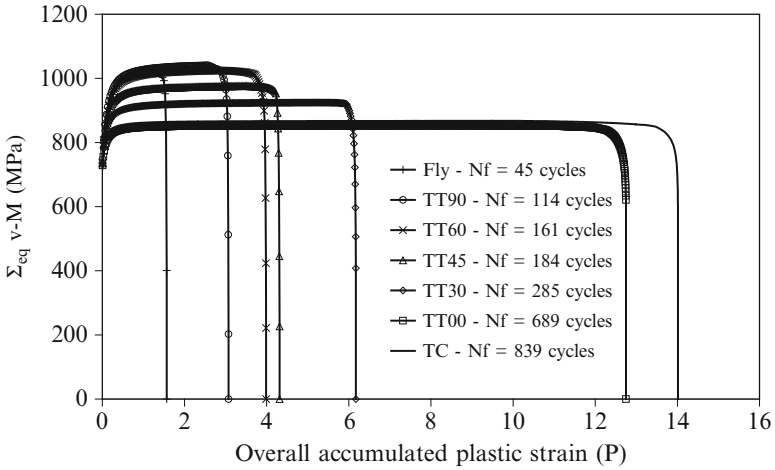


Fig. 5 Overall predicted maximum von-Mises stress evolution per cycle as a function of the accumulated plastic strain for different loading paths (From Mounounga et al. (2011), with permission from Elsevier)

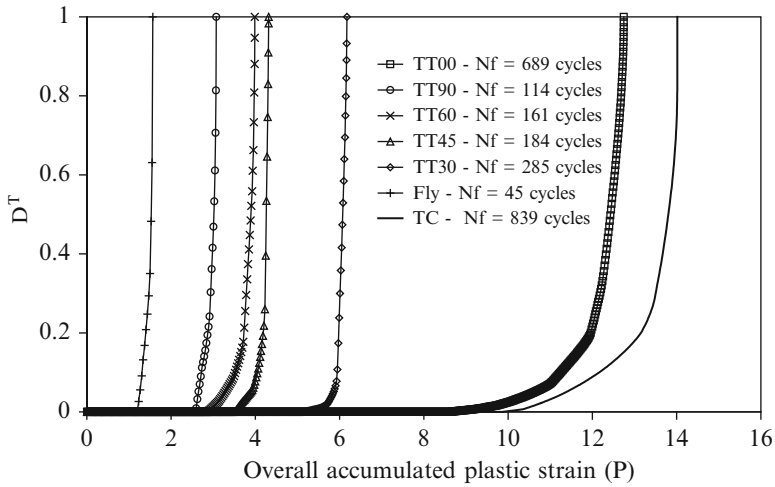


Fig. 6 Overall damage evolutions versus the accumulated plastic strain under different cyclic loading path complexities with the same equivalent strain (From Mounounga et al. (2011), with permission from Elsevier)

made. An important conclusion regarding these numbers is the trend of their evolution as a function of cyclic loading complexity. In fact, it is in a perfect accordance with several experimental results such as in the case of Waspaloy. However, these values are not always in perfect agreement with the experimental data. For example, the recorded average number of activated slips for the simplest load (TC) and for the most complex one (Fly) are respectively of 1.8 and 4.3.

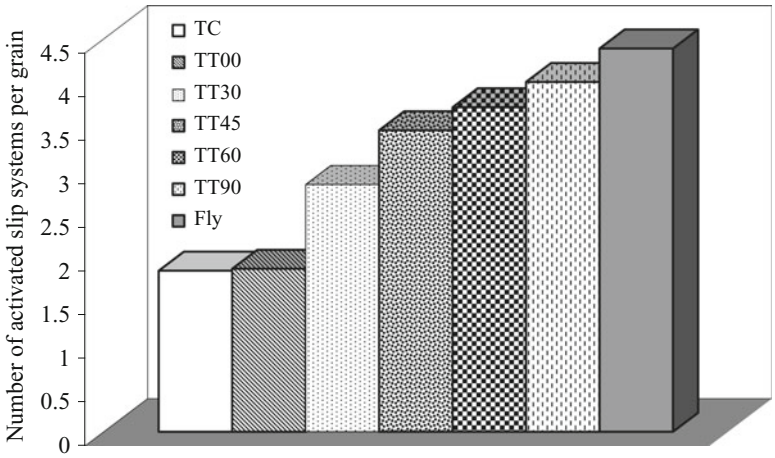


Fig. 7 Predicted evolutions of the activated slip systems for the loading paths TC, TT00, TT30, TT45, TT60, TT90, and Fly (From Mounounga et al. (2011), with permission from Elsevier)

Figure 8 shows distributions of the average number of intragranular damages (NID) per grain for different fatigue life intervals just before the macrocrack initiation. The selected intervals are:

(i) $0.1 \leq d^s < 0.2$
(ii) $0.2 \leq d^s < 0.4$
(iii) $0.4 \leq d^s < 0.6$

In a given fatigue life stage, this figure reveals that the NID varies according to applied loading path whatever the selected damage interval. Obviously, the NID in TC is always dominant with respect to the other loadings. For example, in the damage range of [0.1, 0.2], the NID is approximately 0.13 in TC, while it is 0.02 for Fly (Fig. 8b). This tendency is confirmed for the damage range between 0.2 and 0.4 during which this number becomes almost 0.08 in TC and 0.01 for Fly.

Under biaxial tension-torsion loading, an important variation of the NID with respect to the out-of-phase angle is recorded. It is shown that TT00 has the highest NID. This is due to the fact that, under this loading type (as in TC), almost the same number of grains is systematically loaded within the grain aggregate. This load provokes thus more local damaging zones than in the other more complex loading types. However, a small number of damages reach critical values. As discussed above, the plastic strain localization is governed by the orientation of grains for a given loading path. Therefore, this element has a key role on the variation of damage distribution with respect to the cyclic loading type. Theoretically, a polycrystal is regarded as completely damaged when one or more grains are totally damaged giving at the end $D^T \approx 1$ (▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials," Eq. 50). This is

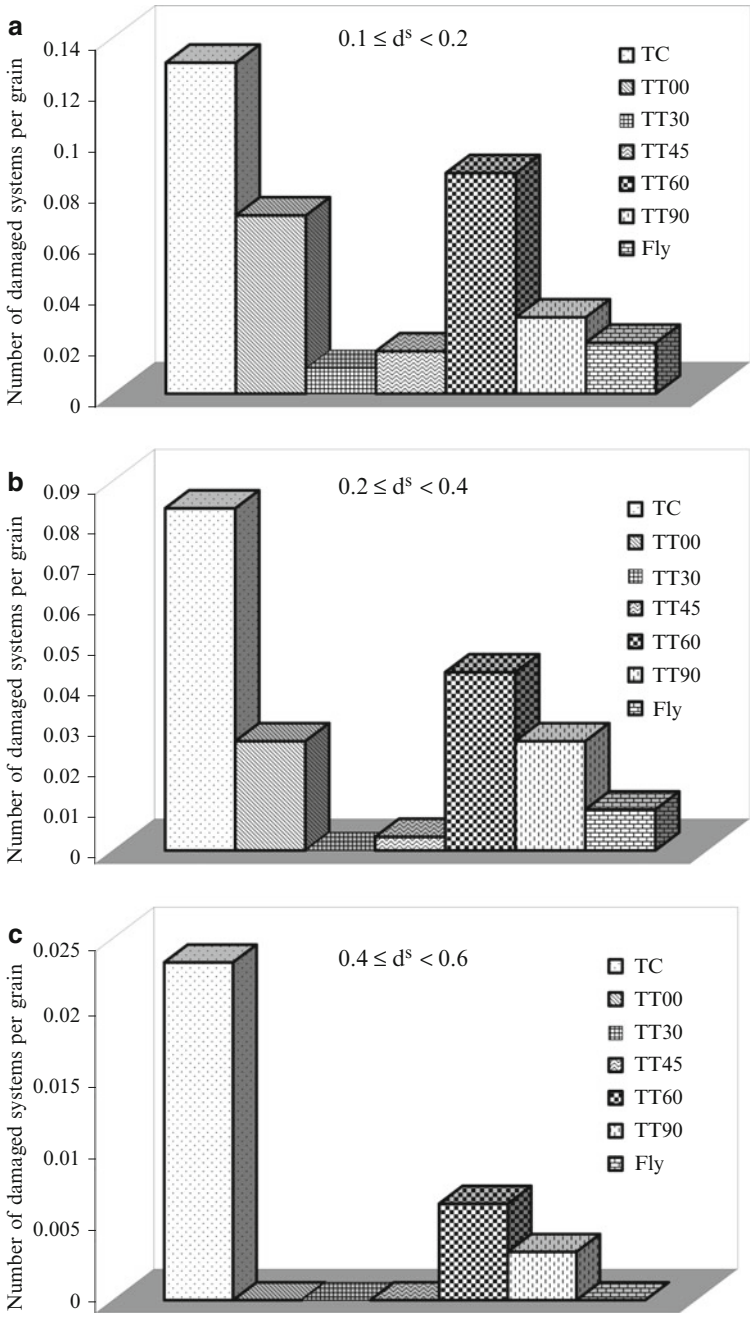


Fig. 8 Variation of the number of intragranular damages under TC, TT00, TT30, TT45, TT60, TT90, and Fly for (**a**) $0.1 > d^s \geq 0.2$, (**b**) $0.2 > d^s \geq 0.4$, and (**c**) $0.4 > d^s \geq 0.6$ (From Mounounga et al. (2011), with permission from Elsevier)

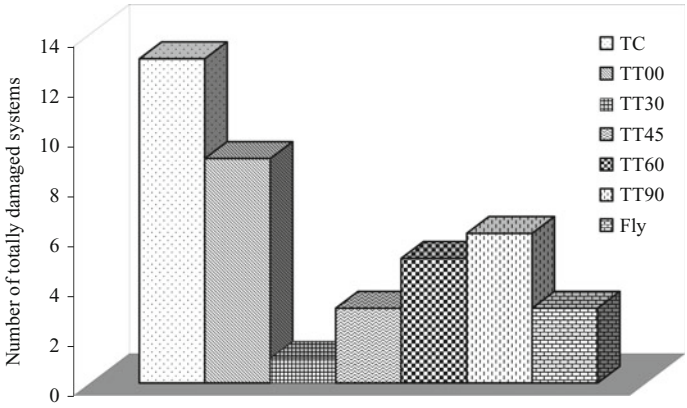


Fig. 9 Comparison of the number of completely damaged slip systems under TC, TT00, TT30, TT45, TT60, TT90, and Fly (From Mounounga et al. (2011), with permission from Elsevier)

exclusively controlled by the concept of damaged grains $N_D^{g\prime}$ and their volume fraction $v_D^{g\prime}$.

Figure 9 shows the variation of the fully damaged systems within the polycrystal under different loading types. Thus, one gets 13 in TC, 9 in TT00, 1 in TT30, 3 in TT45, 5 in TT60, 6 in TT90, and finally 3 for Fly test. By analyzing the polycrystal behavior, the important disparities can find their explanation by the concept of favorably oriented grains with respect to the applied loading direction together with the competition phenomenon of the principal mentioned factors Y_{in}^{s}, λ^{s}, and D_{rs}. Besides, it is recognized that the actual micromechanical approach is unable to take into account the spatial distribution of the grains in the RVE. Thus, the concept of the extrusion-intrusion mechanism cannot be explicitly described as discussed earlier with the initiation of microcracks in these grains at the specimen free surface.

Therefore, Fig. 10 shows a proposed scenario interpreting the distribution of microcracks and its influence by the loading path complexity. As given in Fig. 9, it is concluded that the greater the complexity of the loading path, the lower the microcracks and the shorter the fatigue life. For example, in TC, 13 damaged grains are recorded against three grains in butterfly test. Under TC, this scenario is applied to interpret the obtained results as follows: an important part of these microcracks at the free surface cannot propagate toward their neighbors which have unfavorable orientations where they block consequently themselves at the grain boundaries of the neighbors, except a few ones having a favorably orientation pushing the microcrack propagation easier inside the specimen via neighboring gains. However, under the highest loading complexity as in Fly, microcracks are localized beginning from the free surface toward the interior of the polycrystal initiating therefore a macrocrack (Fig. 10).

To thoroughly understand the local polycrystal behavior described by the model, evolutions of certain local variables are studied. These variables have a key role on the LCF polycrystal behavior. Under different cyclic loading paths, some grains

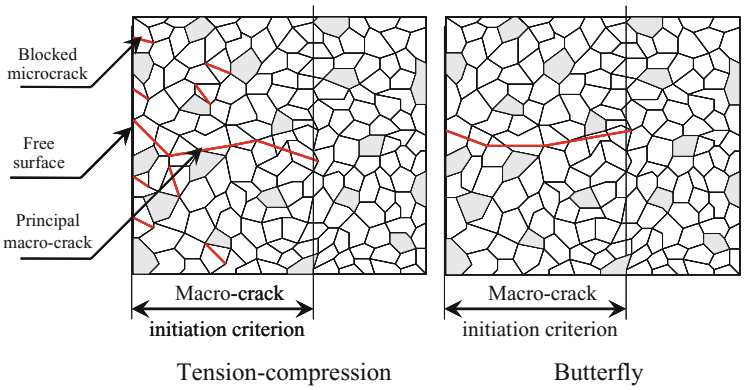


Fig. 10 Scenario showing the microcracks distribution and their influence by the loading path complexity in tension-compression and butterfly showing several microcracks blocked by the grain boundary in TC (From Mounounga et al. (2011), with permission from Elsevier)

and slip systems are previously selected. In fact, Figs. 11 and 12 show some typical examples of local behavior under TC and Fly. In TC, the focus is made on the behavior of the reasonably damaged grain No. 218 (among the 13 damaged grains in TC) with the slip system No. 5, while the system No. 6 of the grain No. 17 is selected in Fly. The intragranular isotropic hardening (R^s) evolution is recorded up to the final grain damaging under TC and Fly (Fig. 11). It is observed that R^s increases relatively slowly during the hardening phase in TC (Fig. 11a). As soon as d^s becomes important, a significant decrease of R^s takes place during the last cycles. However, another scenario is recorded in the case of Fly (Fig. 11b). In fact, R^s evolves rapidly to attain its steady state. In this stage, no damage evolution is observed. Then, the damage initiates and evolves abruptly, as illustrated by the figure, provoking a sudden fall of R^s. This is due to the loading complexity which induces consequently a considerable reduction in fatigue life of the polycrystal.

By comparing the isotropic hardening R^s evolutions in TC and Fly, the highest value is obtained in Fly with almost a value of 110 MPa, whereas it is about 52 MPa in TC. This result is completely foreseeable since the butterfly loading is much more complex than TC.

Predicted evolution of slips within the same grains (i.e., No. 17 and 218) under the same loading paths is considered as typical examples (Fig. 12). In TC, four well activated slip systems are recorded: 1, 3, 5, 7, and 8 (Fig. 12a). Before damaging, it is found that the slip varies in a quasi-linear manner during loading. The accumulated slip (λ^s) is particularly localized on the system no. 5 confirming the strain heterogeneity described by the model. A high increase in the accumulated slip on the system no. 5 (and slower for the system no. 8) induces a significant acceleration of the corresponding damage rate up to the final rupture (Fig. 11a). However, when the latter is completely damaged, the slip within the system no. 7 shows certain acceleration in its rate but without obtaining a totally damaged situation. This can be interpreted by the fact that the energy generated by the external applied loading

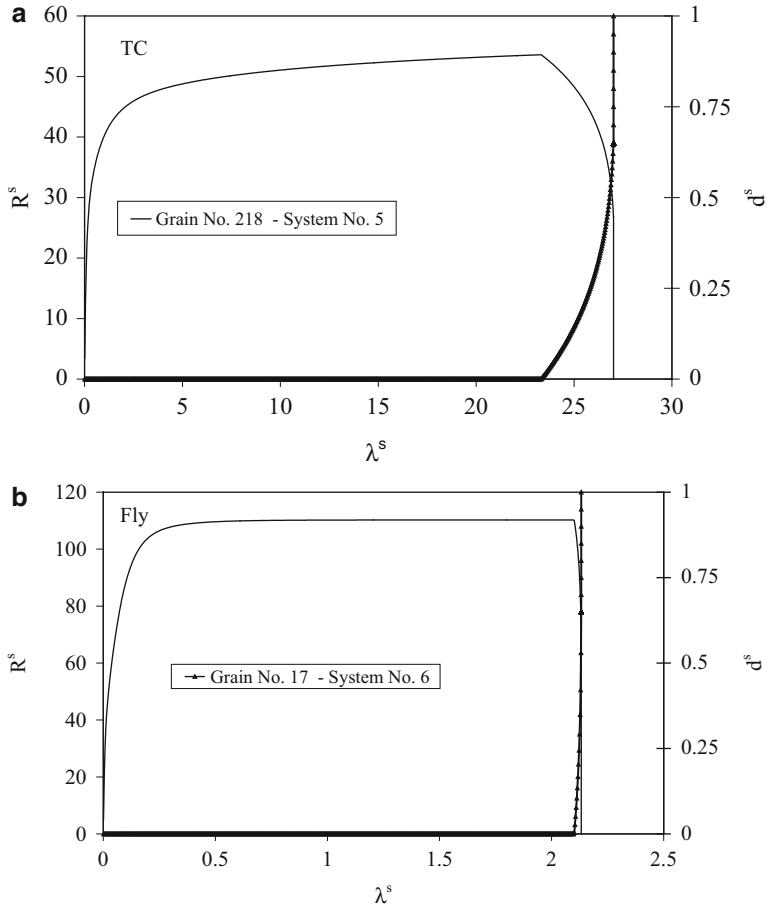


Fig. 11 Predicted evolution of the intragranular isotropic hardening and local damage up to its final value ($d^s \approx 1$) versus accumulated slip in (**a**) TC for the slip system No. 5 for the grain No. 218 and (**b**) Fly for the slip system No. 6 for the grain No. 17 (From Mounounga et al. (2011), with permission from Elsevier)

is transmitted to another activated slip system at the same grain, provoking therefore this acceleration. When the damage reaches an important value, the slip rate undergoes a continuous deceleration up to zero at the instant $d^s = d^s_{cr}$.

In Fly, a large number of activated slip systems are observed with important values for the systems 6, 7, 9, and 10 at the grain No. 17. The most activated slip system, in this case, is the system No. 6. The influence of the new criterion on the slip activation is highlighted via Figs. 11 and 12. Hence, various critical slip values are pointed out (corresponding to intragranular damage initiation) for different applied loading paths. Hence, this damage initiation criterion (▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials," Eq. 33) delays the damage in TC (with an almost threshold accumulated slip value of 23), whereas this

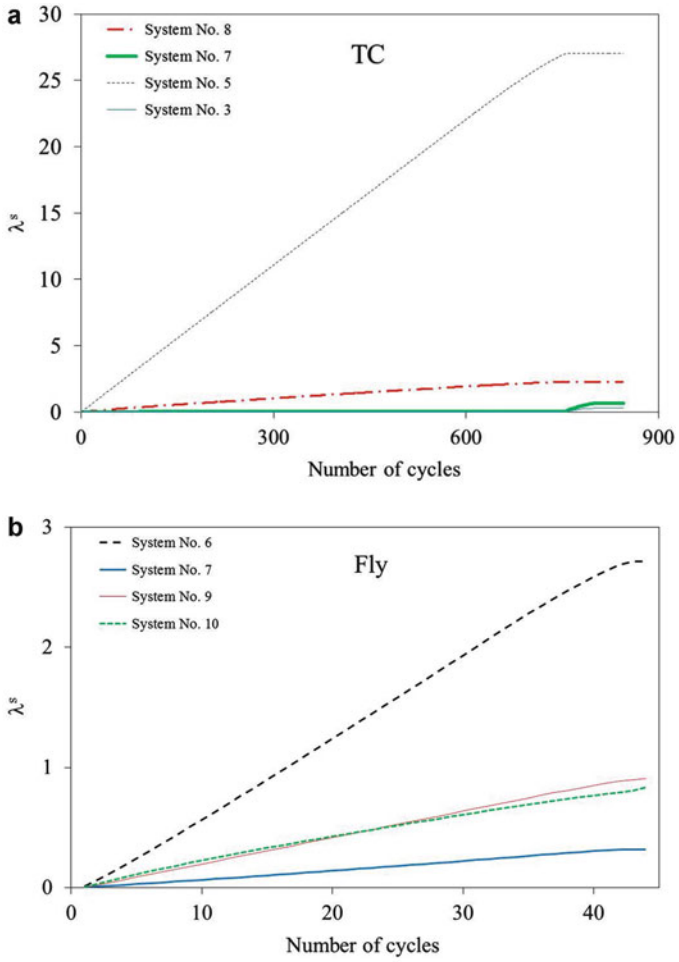


Fig. 12 Predicted evolution of the slip versus number of cycles in (**a**) TC for the grain No. 218 and (**b**) Fly for grain No. 17 (From Mounounga et al. (2011), with permission from Elsevier)

threshold slip value γ_{th}^{s} is approximately 2.1 in Fly (Fig. 11). The effect of the key parameters used in this new criterion (particularly α and γ_o) shows that only one system per grain can activate damage whatever the type of used loading is.

Influence of the Loading Amplitude

In general, the damaged behavior of the metallic polycrystalline materials in LCF has a remarkable sensitivity to the imposed loading amplitude under either strain- or stress-controlled situations. In order to investigate its effect on the fatigue life,

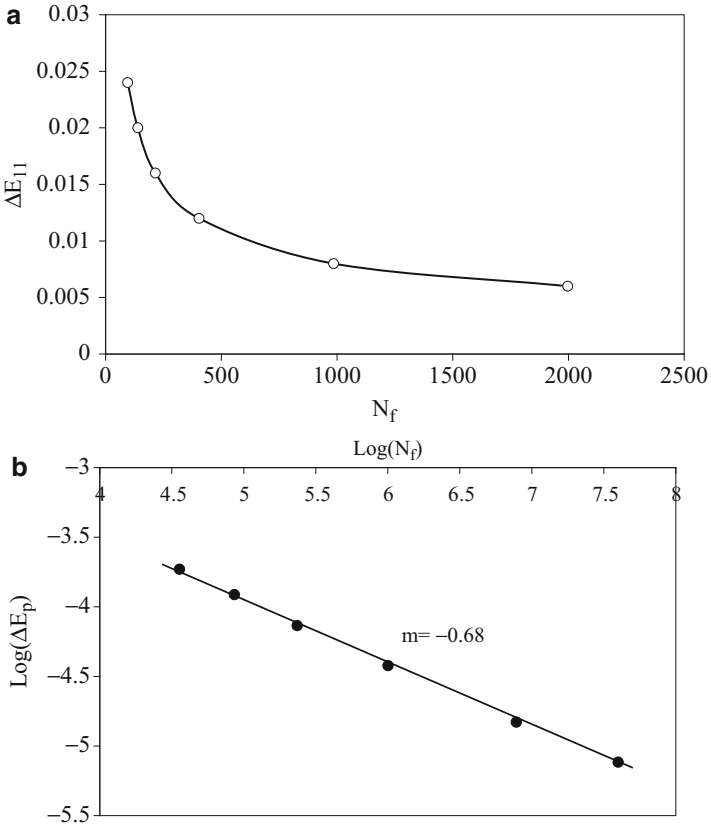


Fig. 13 Effect of applied amplitudes on the fatigue lives in TC (**a**) strain-controlled condition and (**b**) Manson-Coffin diagram prediction (From Mounounga et al. (2011), with permission from Elsevier)

the grain aggregate behavior is simulated under tension-compression using six different strain amplitudes varying from 0.6 % to 2.4 %.

A careful examination of the data in Fig. 13a reveals a clear decrease in fatigue life with the increase of the imposed strain amplitude. Since this theoretical study is focused on LCF behavior of a polycrystal, it can easily consider the Manson-Coffin relation in which the cyclic damage is a function of the stabilized plastic strain amplitude ΔE^p defined by

$$N_f = \left(\frac{\Delta E^p}{C}\right)^m \tag{48}$$

where C and m are material parameters.

Figure 13b displays the variations of overall axial plastic strain during the stabilized phase versus the number of cycles up to final polycrystal damage (Log E_p-Log N_f). A linearization of the predicted results presents a slope of $m = -0.68$. This prediction appears acceptable compared to the usual slope of Manson-Coffin of -0.5. Thus, it can be concluded that the model parameters describe quite well the Manson-Coffin relation. Note that with correctly identified model parameters, a slope of -0.52 is obtained as it will be demonstrated later.

Damage Deactivation Effect

Microcracks open and close especially at the end of the first fatigue stage of polycrystals. The damage activation/deactivation phenomenon translates the influence of these two phases on the overall material behavior during cyclic loading. This model faithfully describes this phenomenon through (Eq. 23, ▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials"), particularly in the case of the complex loading, based on the projection operator technique. Numerical simulations are carried out describing this phenomenon under strain- and stress-controlled conditions.

Damage Deactivation Effect Under Strain-Controlled Condition

Contrary to TC and TT00, the complex cyclic loading path, such as TT60, represents a case in which the eigenvectors do not coincide with the laboratory reference axes. Thus, the eigenvectors vary with respect to time; thus, the three terms of equation (25, ▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials") should be entirely taken into account due to the existence of the shear component. Numerically, the strain amplitudes of $\Delta E_{11} = 1.52\ \%$ (axial strain) and $\Delta E_{12} = 1.32\ \%$ (shear strain) are used. An examination of Fig. 14 shows some selected cycles in the overall stress space (Σ_{11}–Σ_{12}) corresponding to the first three cycles, the stabilized one, and the last cycles of fatigue life. The anisotropy induced by damage is clearly described by the nonsymmetrical evolution of the overall stresses in tension, compression, and torsion phases notably during the softening (damaging) stage. In fact, an important overall nonsymmetrical response is illustrated through the evolution of the two stress fields. This depends exclusively on the damage activation/deactivation and the active/passive passage criterion. Therefore, the operator $\underline{\underline{\underline{\underline{P}}}}^{+}$ (▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials," Eq. 13) verifies it naturally, i.e., it is capable to select the damage activation and deactivation phases even for a complex cyclic loading path such as TT60. In fact, the overall response, in the stress space Σ_{11}–Σ_{12}, is affected in the first quadrant in which Σ_{11} and Σ_{12} are positive (Fig. 14) especially for a significant

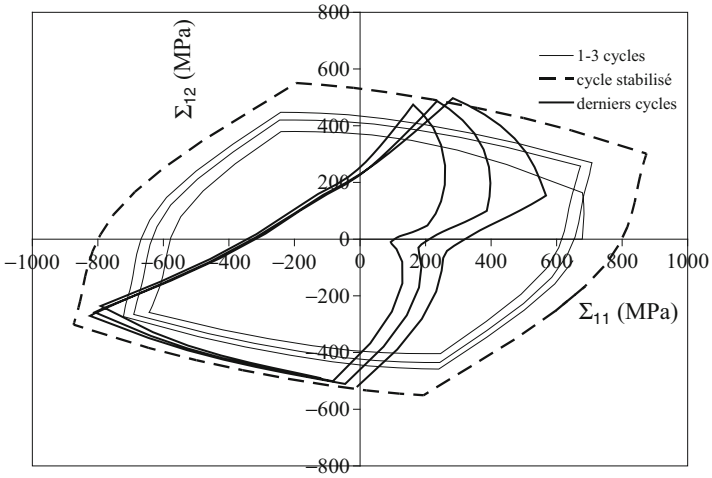


Fig. 14 Evolution of the overall axial stress Σ_{11} versus overall shear stress Σ_{12} in TT60 (From Mounounga et al. (2011), with permission from Elsevier)

damage value, whereas the sound material behavior is partially retrieved in the third quadrant, since the components 11 and 12 are negative. The particular trend of the overall response in the second and fourth quadrants can be explained by the combination effect of the two overall stress components Σ_{11} and Σ_{12}. They have opposed signs because of the applied loading nature. The damage is thus active in a direction and passive in the other.

The most complex cyclic loading path of butterfly gives another evolution related to damage deactivation which has an explicit relation with the eigenvectors and its coincidence with the laboratory reference axes (Fig. 15). As in TT60, the same strain amplitudes (ΔE_{11} and ΔE_{12}) are also employed. The anisotropy induced by damage is also described by the nonsymmetrical overall stress evolution notably during the last cycles (i.e., softening stage). In fact, an important overall distortion of the response is entirely due to the activation and deactivation of the damage controlled by the operator $\underline{\underline{\underline{\underline{P}}}}^{+}$ verifying naturally the two damage phases.

To further demonstrate the anisotropy induced by damage, the evolution of some components of the damage and rigidity tensors is recorded under TT90. Several important features of the induced-oriented anisotropy behavior under TT90 are shown in Fig. 16. In fact, the evolutions of two chosen damage tensor components (D_{1111} and D_{1212}) are studied; one can observe that they do not have the same rate of change. Indeed, D_{1111} evolves up to a value close to 1, while D_{1212} reaches almost a value of 0.25. This anisotropy is clearly confirmed by the rigidities and damage variations for the two given orientations.

Let us examine now the evolution of the two chosen components (R_{1111} and R_{1212}) of the stiffness tensor coupled with damage during fatigue life. It is first noticed that the damage effect in 11 direction is stronger ($a_{11}+$) than that of

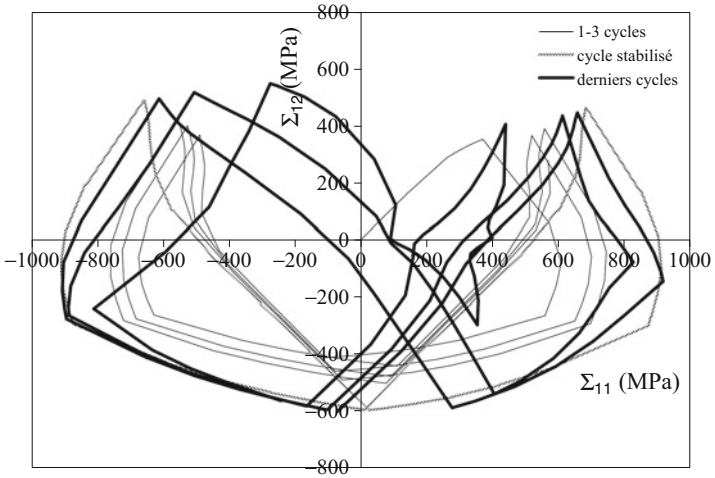


Fig. 15 Evolution of the overall axial stress Σ_{11} versus overall shear stress Σ_{12} in Fly (From Mounounga et al. (2011), with permission from Elsevier)

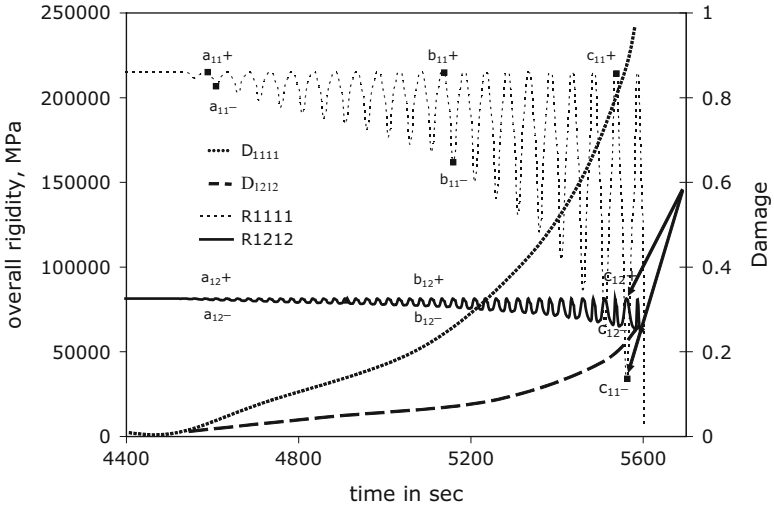


Fig. 16 Evolution of the components R_{1111} and R_{1212} of the stiffness tensor coupled with damage and components D_{1111} and D_{1212} of the damage tensor in TT90 (From Mounounga et al. (2011), with permission from Elsevier)

12 direction ($b_{12}+$). Note that the points a..+, b..+, c.. + correspond to the damage deactivation phases. The activation damage stages are noted by the points a..−, b..−, c..−. For an adequate damage value, the damage activation and deactivation phases are clearly emphasized. For the damaged material at the stage c, an attention is made related to the points $c_{12}+$ and $c_{11}-$. These two points correspond both to the

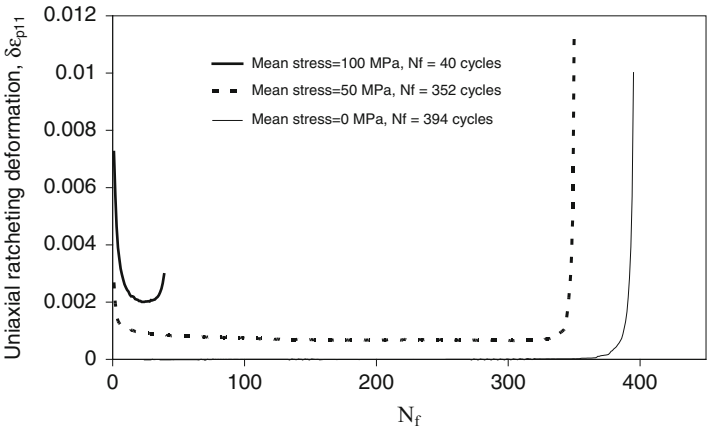


Fig. 17 Predicted evolutions of the uniaxial ratcheting deformation ($\delta\varepsilon_{p11}$) versus number of cycles up to macrocrack initiation under TC with stress-controlled condition with $\Delta\Sigma_{11} = 1{,}760$ MPa for three mean stresses of $\overline{\Sigma}_{11} = 0$, 50 and 100 MPa (From Mounounga et al. (2011), with permission from Elsevier)

same instant. Note that the damage is passive in 12 direction (c_{12}+), whereas it is active in 11 ($c_{11}-$). The total rigidity drop phenomenon, in TT90, is almost captured in 11 direction, whereas it is partial in 12 direction.

Damage Deactivation Effect Under Stress-Controlled Condition

As shown above, the metallic polycrystal behavior in LCF is mainly influenced by the imposed strain amplitude. However, this paragraph deals with the cyclic behavior of polycrystals in a stress-controlled condition with symmetrical and notably nonsymmetrical cases. Numerically, a tension-compression loading is tested under stress-controlled condition having an amplitude of $\Delta\Sigma_{11} = 1{,}760$ MPa. Three different mean stresses of $\overline{\Sigma}_{11} = 0$, 50 and 100 MPa are used. Hence, the effect of mean stress on the uniaxial ratcheting phenomenon coupled with damage is investigated.

The increasing of damage affects significantly the material rigidity as illustrated in Fig. 17. This figure also emphasizes the effect of mean stress on the number of cycles up to damage initiation. Actually, in the nonsymmetrical case of $\overline{\Sigma}_{11} = 50$ MPa, the damage initiates earlier than that in the symmetrical case: one has 230 cycles for $\overline{\Sigma}_{11} = 50$ MPa and 250 cycles when $\overline{\Sigma}_{11} = 0$ MPa. Also, the damage rate in the case of $\overline{\Sigma}_{11} = 50$ MPa is more important than that of $\overline{\Sigma}_{11} = 0$ MPa. When the damage becomes important, the phenomenon of damage activation/deactivation becomes evident as illustrated in Figs. 18 and 20. These results given in Figs. 17, 18, 19, and 20 show the capacity of the model to describe the mean stress effect on the cyclic plasticity and the corresponding damage evolutions with the possibilities to reproduce the ratcheting phenomenon coupled with damage.

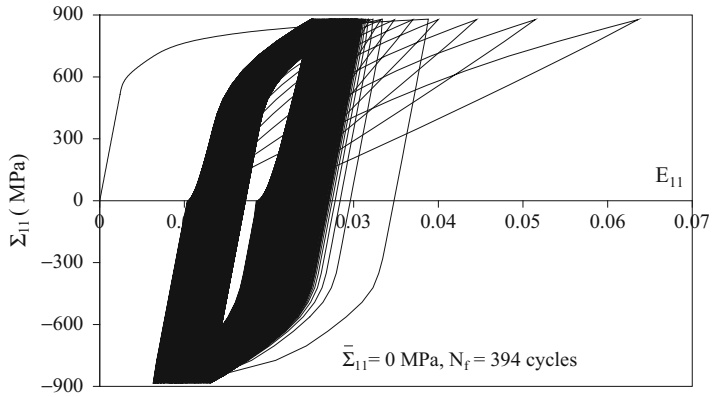


Fig. 18 Evolution of axial overall stress Σ_{11} -strain E_{11} in TC with stress-controlled condition using a mean stress of $\bar{\Sigma}_{11} = 0$ MPa up to the macrocrack initiation (From Mounounga et al. (2011), with permission from Elsevier)

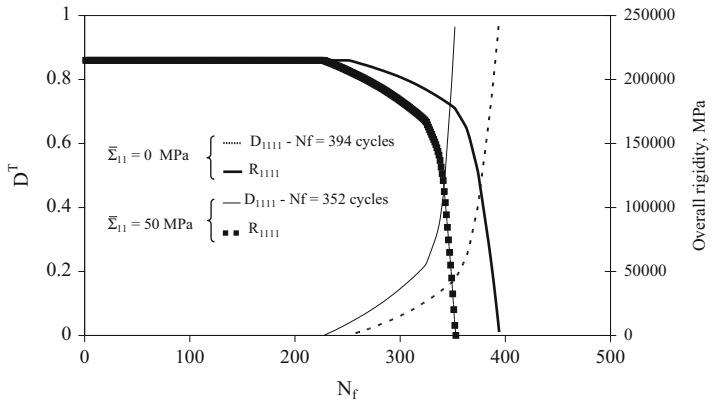


Fig. 19 Evolution of the component R_{1111} of the overall stiffness tensor coupled with the damage component D_{1111} of the overall damage tensor in TC with stress-controlled condition using mean stresses of $\bar{\Sigma}_{11} = 50$ and 50 MPa up to the macrocrack initiation (From Mounounga et al. (2011), with permission from Elsevier)

Quantitative Study

A quantitative study is carried out to compare the model predictions with those experimental data given in Abdul-Latif et al. (1999) for Waspaloy under TC and TT90. The model coefficients already identified are employed. Besides, the model coefficients of the new intragranular damage initiation criterion (α and γ_o^s) are identified based on these available experimental data as well as on the approach

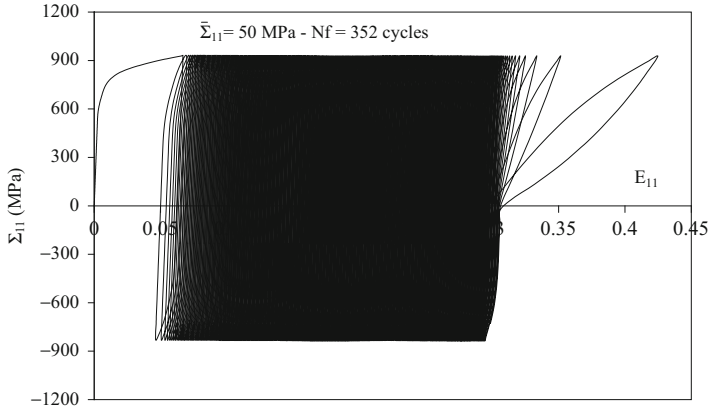


Fig. 20 Overall axial evolution of Σ_{11}-E_{11} in TC with stress-controlled condition using a mean stress of $\bar{\Sigma}_{11} = 50$ MPa up to the macrocrack initiation (From Mounounga et al. (2011), with permission from Elsevier)

given above (▶ Chap. 11, "Ductile Damage Behavior in Low-Cycle Fatigue for Polycrystalline Metallic Materials," Eqs. 34 and 35). In order to conduct this process, the employed databases are tension-compression ($\Delta E^p_{11} = 1$ %, $N_f = 1{,}442$ cycles) and out-of-phase tension-torsion ($\Delta E^p_{11} = 0.8$ % and $\Delta E^p_{11} = 0.52$ %, $N_f = 136$ cycles) with a sinusoidal waveform and a phase lag of 90° between the two sinusoidal signals. In each case, the maximum von-Mises equivalent plastic strain was maintained constant at 0.5 % during the test expressed as follows:

For uniaxial tension-compression test:

$$E^p_{eq-max} = \frac{E^p_{11max} - E^p_{11min}}{2} = 0.5 \ \%$$

For biaxial tension-torsion test:

$$E^p_{eq-max} = \max\left(\sqrt{E^{p^2}_{11} + \frac{\gamma^{p2}}{3}}\right) = 0.5 \ \%$$

with $\gamma^p = 2E^p_{12}$

Note that the used experimental data have been conducted using a thin-walled tube. It is well known that the optimized material coefficients are specified as a best fit between the predicted and experimental results. The identified model coefficients are the same as in (Table 2) except for α and γ^s_o values which are respectively 5 and 400. For the fatigue life, the micromechanical model can suitably describe the fatigue lives under these cyclic loading conditions (N_f (theo) = 1,497 cycles and N_f (theo) = 148 in TT90). These results are not surprising since the identification procedure is conducted using the same experimental data.

Table 4 Experimental and predicted fatigue lives for Waspaloy under TC and TT90 loading paths

Type of loading path	Experimental fatigue life (Nf), cycles	Theoretical fatigue life (Nf), cycles
TC ($\Delta E_{11} = 1$ % – HT)	1,442	1,479
TC ($\Delta E_{11} = 1$ % – ST)	2,328	1,479
TC ($\Delta E_{11} = 1.5$ % – ST)	790	803
TC ($\Delta E_{11} = 2.2$ % – ST)	419	356
TT90 (HT)	136	148

HT, thin-walled tube; *ST*, solid circular section specimen

To validate the calibrated coefficients, several numerical simulations are thus realized showing the model capacity in describing the Waspaloy cyclic behavior. Actually, three TC numerical tests are conducted under plastic strain-controlled situation of different amplitudes: $\Delta E_{11}^{p} = 1$ %, 1.5 %, and 2.2 % simulating the LCF behavior of Waspaloy using solid circular section specimens. Table 4 shows various experimental and predicted fatigue lives obtained under TC and TT90 for two categories of specimens, tubes with solid section (ST) and thin-walled one (HT).

The TC experimental fatigue lives are 2,328, 790, and 419 cycles for $\Delta E_{11}^{p} = 1$ %, 1.5 %, and 2.2 %, respectively. The predicted fatigue lives are 1,479, 803, and 356 cycles, respectively (Table 4). In fact, the identification procedure is performed employing an experimental data of TC and TT90 conducted on a thin-walled tube. Since the employed model cannot take into account the geometrical aspect, therefore, these theoretical predictions are not surprising. As a matter of fact, for a relatively large imposed strain, the numerical and experimental responses are rather close. This phenomenon could be interpreted by the fact that with strain amplitude increasing, the specimen geometry effect (free surface phenomenon) becomes less important compared to the lower strain amplitude, for example, $\Delta E_{11}^{p} = 1$ %, the fatigue lives are 1,442 cycles for a thin-walled tube and 2,328 cycles for a solid one.

An examination of the results given in Fig. 21 confirms that the model can suitably predict the Manson-Coffin relation with a theoretical slope of $m_{sim} = -0.52$ which is almost close to that recorded experimentally $m_{exp} = -0.45$. This difference can be explained by the important difference between the fatigue lives obtained in TC in the case of $\Delta E_{11}^{p} = 1$ % for a specimen having solid section. It is recognized that the sensitivity of this phenomenon decreases considerably once the loading amplitude increases.

Conclusion

Using concept of anisotropy induced by damage, the impact of the new intragranular damage initiation criterion on the fatigue life is obviously demonstrated at the overall and local levels. Hence, a parametric study is conducted

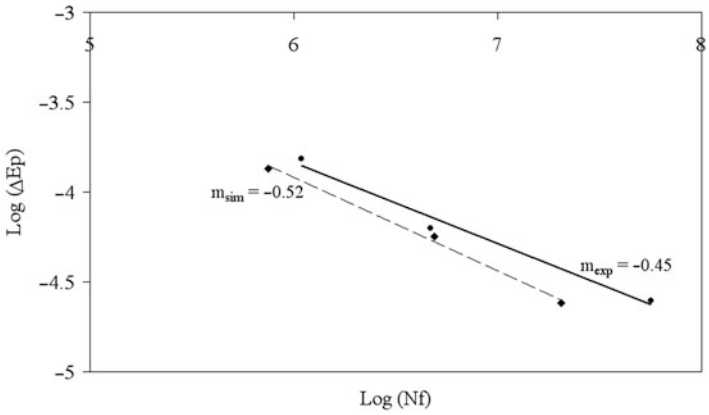


Fig. 21 Manson-Coffin diagram comparing theoretical and experimental results for Waspaloy (From Mounounga et al. (2011), with permission from Elsevier)

showing the influence of the two new model parameters (α and γ_o^s) on the LCF behavior. As a conclusion, that the fatigue life decreases when α increases and γ_o^s decreases. Moreover, the effect of the magnitude and complexity of loading path on the cyclic hardening and on the fatigue life evolutions is numerically studied. Therefore, different complex cyclic loading situations are used under strain-controlled condition showing the ability of the model to describe this induced anisotropy. Indeed, the nonsymmetrical evolution of the overall stresses in tension, compression, and torsion phases especially during the softening stage is correctly described. This depends exclusively on the activation and deactivation of the damage and the active/passive passage criterion governed by the positive spectral projection tensor $\underline{\underline{\underline{\underline{P}}}}^{+}$ notably for complex cyclic loadings.

Under stress-controlled condition, the obtained results illustrate that the proposed model is able to describe the ratcheting phenomenon coupled to damage.

Finally, a quantitative analysis is conducted to compare the results from numerical simulations of the model to the experimental data recorded for Waspaloy.

Acknowledgments This is certainly not a classical acknowledgments, nonetheless it is in memory of a friend and ex-Ph.D. student Blondin Mounounga, since a part of the present effort has been developed with him.

References

A. Abdul-Latif, M. Chadli, Modeling of the heterogeneous damage evolution at the granular scale in polycrystals under complex cyclic loadings. Int. J. Damage Mech. **16**, 133 (2007)

A. Abdul-Latif, T.B.S. Mounounga, Damage deactivation modeling under multiaxial cyclic loadings for polycrystals. Int. J. Damage Mech. **18**, 177–198 (2009)

A. Abdul-Latif, K. Saanouni, Damaged anelastic behavior of FCC polycrystalline metals with micromechanical approach. Int. J. Damage Mech. **3**, 237 (1994)

A. Abdul-Latif, K. Saanouni, Micromechanical modeling of low cyclic fatigue under complex loadings-part II. Applications. Int. J. Plast. **12**, 1123 (1996)
A. Abdul-Latif, V. Ferney, K. Saanouni, Fatigue damage of Waspaloy under complex loading. ASME J. Eng. Mater. Technol. **121**, 278 (1999)
J.L. Chaboche, Une nouvelle condition unilatérale pour décrire le comportement des matériaux avec dommage anisotrope. C. R. Acad. Sci. Paris t. **314**(Série II), 1395 (1992)
J.L. Chaboche, Development of continuum damage mechanics for elastic solids sustaining anisotropic and unilateral damage. Int. J. Damage Mech. **3**, 311 (1993)
M. Chadli, A. Abdul-Latif, Meso-damage evolution in polycrystals. ASME J. Eng. Mat. Tech **127**, 214 (2005)
M. Clavel, P. Pilvin, R. Rahouadj, Analyse microstructurale de la déformation plastique sous sollicitations non proportionnelles dans un alliage base nickel. C. R. Acad. Sci. Paris **309**, 689 (1989)
D. Halm, A. Dragon, A model of anisotropic damage by mesocrack growth; unilateral effect. Int. J. Damage Mech. **5**, 384 (1996)
N.R. Hansen, H.L. Schreyer, Damage deactivation. ASME J. Appl. Mech. **62**, 450 (1995)
J.W. Ju, On energy-based coupled elastoplastic damage theories: constitutive modeling and computational aspects. Int. J. Solids Struct. **25**, 803 (1989)
L.M. Kachanov, *Introduction to Continuum Damage Mechanics* (Martinus Nijhoff, Dordrecht, 1986)
D. Krajcinovic, G.U. Fonseka, Continuous damage theory of brittle materials. J. Mech. **48**, 809 (1981)
P. Ladevèze, J. Lemaitre, *Damage Effective Stress in Quasi-Unilateral Material Conditions* (IUTAM Congress, Lyngby, 1984)
J. Lemaître, *A Course on Damage Mechanics* (Springer, Berlin, 1992)
J. Lemaître, J.L. Chaboche, *Mechanics of Solids Materials* (Cambridge University Press, Cambridge, 1990)
J. Mazars, G. Pijaudier-Cabot, Continuum damage theory-application to concrete. J. Eng. Mech. **115**, 345 (1989)
T.B.S. Mounounga, A. Abdul-Latif, D. Razafindramary, Damage induced-oriented anisotropy behavior of polycrystals under complex cyclic loadings. Int. J. Mech. Sci. **53**(4), 271–280 (2011)
M. Ortiz, A constitutive theory for the inelastic behavior of concrete. Mech. Mater. **4**, 67 (1985)
S. Ramtani, Conribution à la Modélisation du Comportement Multiaxial du Béton Endommagé avec Description un Caractère Unilatéral. Thèse de Doctorat, Université Paris VI, 1990
K. Saanouni, A. Abdul-Latif, Micromechanical modeling of low cycle fatigue under complex loadings-part I. Theoretical formulation. Int. J. Plast. **12**, 1111 (1996)
G.Z. Voyiadjis, P. Kattan, *Advances in Damage Mechanics: Metals and Metal Matrix Composites* (Elsevier, Oxford, 1999). 542 p
S. Yazdani, S. Karnawat, Mode I damage modeling in brittle preloading. Int. J. Damage Mech. **6**, 153 (1997)
S. Yazdani, H.L. Schreyer, Combined plasticity and damage mechanics model for plain concrete. ASCE J. Eng. Mech. **116**, 1435 (1990)

Section IV

Damage in Structures

Damage in Prestressed Concrete Structures due to Creep and Shrinkage of Concrete

17

Zdeněk P. Bažant, Mija H. Hubler, and Qiang Yu

Contents

Z.P. Bažant (✉)
Department of Civil and Environmental Engineering, Northwestern University, Evanston IL, USA
e-mail: z-bazant@northwestern.edu

M.H. Hubler
Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, CO, USA
e-mail: hubler@mit.edu

Q. Yu
Department of Civil and Environmental Engineering, University of Pittsburgh, Swanson School of Engineering, Pittsburgh, PA, USA
e-mail: qiy15@pitt.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_49

Abstract

The theory of concrete creep and shrinkage has not been regarded as part of what has been understood as damage mechanics. However, these inelastic phenomena do result in significant damages to structures, including not only distributed cracking, which is damage in classical sense, but also damage in the form of excessive deflection which puts the structure out of service. It is for this reason that a chapter on creep and shrinkage is included in this handbook. The theory of creep and shrinkage of concrete has become a vast field. In what follows, only a rather brief exposition of this subject will be included. The emphasis will be on the topics of wide recent interest, triggered by the revelation of practical problems that have been synthesized into a coherent picture only during the last few years. The creep effects are particularly important for super-tall buildings and prestressed concrete structures, because of their slenderness and high flexibility, and are paramount for nuclear reactor containments and vessels. At high-temperature exposure, as in fire in tunnels or tall buildings or in postulated nuclear reactor accidents, creep is very large and plays a major role.

Introduction

Creep and shrinkage can cause diverse types of damage in concrete structure, whether prestressed or simply reinforced or unreinforced. Long-term creep may lead to excessive deflections of bridges and other structures which can severely curtail the life span. This is a widespread problem of a magnitude not appreciated until recently, and this chapter will be focused on it.

Nonuniformity of the creep properties and of restrained shrinkage is another phenomenon which leads to redistributions of stresses and bending moments, which in turn may cause deleterious cracking in bridges, buildings, nuclear reactor containments, and other structures. It may promote other processes endangering durability and sustainability, such as corrosion. Creep nonuniformity is caused by differences in the histories of pore humidity and temperature at various points or cross sections of a structure, by differences in age and in concrete type, and by interaction of concrete with steel parts which do not creep. Aside from prestressing steel relaxation, concrete creep is one major cause of the longtime loss of prestress.

Rare though not unheard of, creep in slender structures such as columns, shells, or compressed plates may cause collapse due to longtime creep buckling. Transfer of compressive stresses from creeping and shrinking concrete to steel bars, cables, girders, or other non-creeping parts (e.g., stone cladding or masonry liner) may induce their buckling or compression failure.

Except for leak-tightness endangerment by cracking induced by shrinkage and creep in nuclear containment, creep and shrinkage do not have any significant effect on safety under overload of structures. The main effect of creep and shrinkage is to compromise serviceability and durability of structures and thus sustainability of concrete infrastructure.

Material Models for Creep and Shrinkage of Concrete Structures

To introduce the subject, a brief description of concrete creep and shrinkage behavior is first presented by updating the 2001 exposition in Encyclopedia of Materials on the topic (Bažant 2001). The term creep represents a continued deformation of a material under sustained load. Since concrete also exhibits shrinkage due to the diffusion process of drying or to chemical processes due to hydration, causing a decrease of volume at no external load, the creep, together with the instantaneous (or initial elastic) strain, is defined as the strain difference between identical loaded and load-free specimens. The drying shrinkage strain typically reaches 0.0002–0.0004 and in poor concretes even 0.0008. The shrinkage due to chemical processes is called the autogenous shrinkage, which is negligible for normal concretes with higher water/cement ratios and negligible self-desiccation but large for modern high-strength concretes with very low water/cement ratios (<0.35) and pronounced self-desiccation. After many years, the creep strain typically attains values 2–6 times larger than the initial elastic strain. After unloading, one observes a partial creep recovery. The creep of concrete was discovered by Hatt (1907) at Purdue University, while the shrinkage was discovered much earlier by Le Chatelier at Ecole des Mines in Paris (1905).

The creep during drying normally greatly exceeds the sum of the shrinkage and creep of a sealed specimen. The excess strain, called the drying creep or the Pickett effect, represents a hygro-mechanical coupling between strain and water content changes. Imbibition of water causes swelling which is normally much less than the drying shrinkage or, if concrete has dried, a hysteretic partial reversal of shrinkage. The drying shrinkage (Fig. 1 bottom) is caused by compressive stresses in the microstructure which balance changes in the capillary tension and surface tension on pore walls as well as changes in the disjoining pressure in hindered adsorbed water layers in nanopores.

Creep is caused by slips due to bond ruptures (with bond restorations at adjacent sites) in the atomic nanostructure of the hardened Portland cement paste. The paste is strongly hydrophilic, having a disordered colloidal microstructure, a porosity of about 0.4–0.55, and an enormous internal surface area – about 500 m^2/cm^3. The main component of the paste, in which the creep mechanism resides, is the

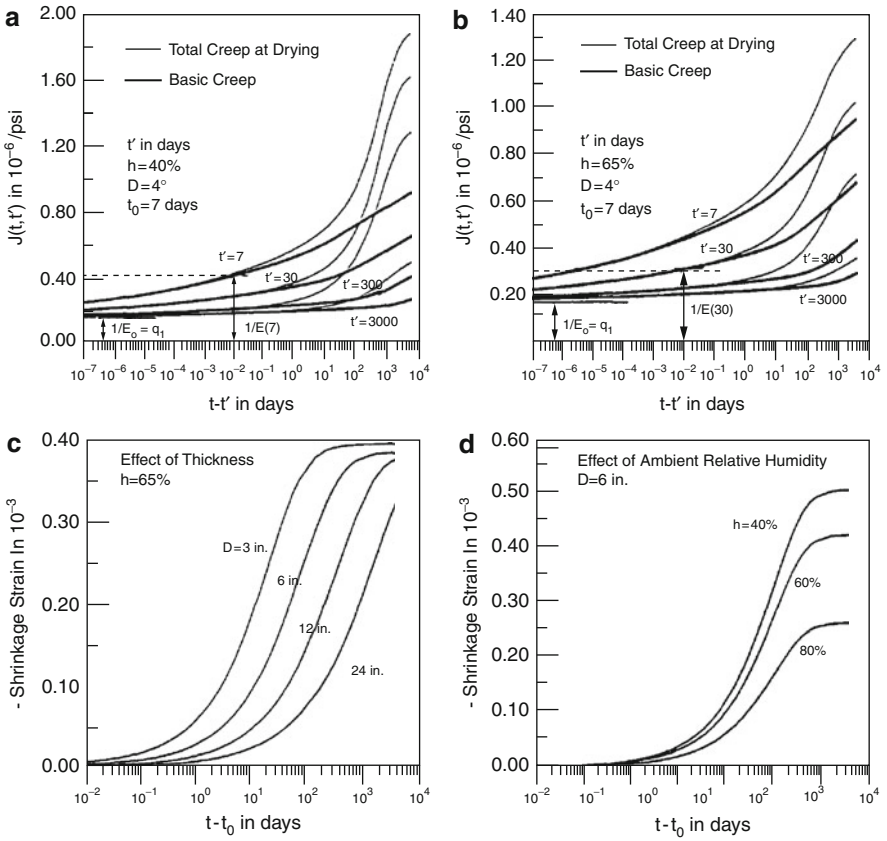


Fig. 1 Typical creep compliance and shrinkage strain curves given by the B3 model (With kind permission from Springer Science+Business Media: Materials and Structures, Creep and shrinkage prediction model for analysis and design of concrete structure – model B3, 28, 1995, 1–83, Bažant, Z.P. and Baweja, S., 1.1)

tricalcium silicate hydrate gel ($3\ CaO \cdot 2\ SiO_3 \cdot H_20$, in short C–S–H). This solid forms crystalline sheets and needles of colloidal dimensions, weakly bound by van der Waals forces. The physical mechanism and modeling are still being debated. The model in Eqs. 1, 2, 3, 4, and 5 which follow is not the only one available but has the deepest theoretical support and fits the test data best.

Constitutive Law: In the absence of cracking and at service stresses in structures (which are generally less than 40 % or 45 % of concrete strength), the creep strain depends on the stress linearly. It may be fully characterized by the compliance function $J(t, t')$ (Fig. 1 top), defined as the strain ϵ at time (or age) t caused by a unit uniaxial stress $\sigma = 1$ applied at age t'. As t' increases, the creep diminishes. This phenomenon, called aging, causes that J depends not only on the time lag t-t' but also on t'. At variable stress $\sigma(t)$, each stress increment $d\sigma(t')$ applied at time t'

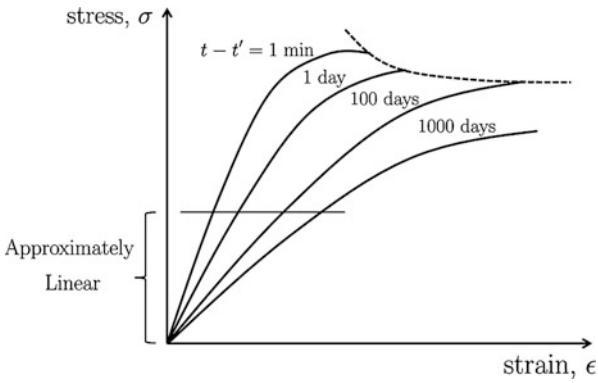


Fig. 2 Typical isochrones of concrete creep at various ages of loading

produces strain history $d\epsilon(t) = J(t, t')d\sigma(t')$. By superposition principle (introduced by Boltzmann and, with aging, by Volterra), one gets the linear (uniaxial) creep law of aging viscoelasticity:

$$\epsilon(t) = \int_{t_1}^{t} J(t, t')d\sigma(t') + \epsilon^0(t) \tag{1}$$

where ϵ^0 denotes the sum of shrinkage strain ϵ_{sh} and thermal expansion (and smeared cracking strain, if any). The integral is the Stieltjes integral, which admits histories $\sigma(t)$ with jumps; for time intervals with no jump, one may set $d\sigma(t') = [d\sigma(t')/dt']/dt'$. When history $\epsilon(t)$ is prescribed, then Eq. 1 represents a Volterra integral equation for $\sigma(t)$ which, for realistic forms of $J(t, t')$, is not integrable analytically, although numerical integration is easy. The solution $\sigma(t)$ for strain $\epsilon = 1$ imposed at age t' (and for $\epsilon^0 = 0$) is called the relaxation function. Generalizing Eq. 1 according to the principle of superposition and taking isotropy into account yields a triaxial stress–strain relation, in which the shear and bulk compliance functions are $J_G(t, t') = 2(1 + \upsilon)J(t, t')$ and $J_K(t, t') = 3(1 - 2\upsilon)J(t, t')$ ($\nu \approx 0.18$ = Poisson ratio, considered as approximately constant). At high stress, the creep law appears to be nonlinear (Fig. 2). Equation 1 remains valid if the strain due to time-dependent growth of distributed microcracks is eluded in $\epsilon^0(t)$.

The value $J(t', t') = q_1$ corresponding to extrapolation of creep curves to zero load duration may be considered as age independent. The conventional Young's elastic modulus, implied as $E(t') = 1/J(t' + \delta, t')$ where usually $\delta \in$ (0.0001 s, 10 min.), increases with age t′. A realistic form of J(t′, t′) (Bazant et al. 1997; bold curves in Fig. 1 top) may conveniently be expressed by its rate:

$$\dot{J}(t', t') = \upsilon^{-1}(t)\dot{C}_g(\theta) + 1/\eta_f \tag{2}$$

$$\upsilon^{-1}(t) = q_2(\lambda_0/t)^m + q_3 \tag{3}$$

$$\dot{C}_g(\theta) = \frac{n\theta^{n-1}}{\lambda_0^n + \theta^n}, \theta = t - t' \tag{4}$$

Here $\dot{o} = \partial o/\partial t$; θ = load duration; $\lambda_0 = 1$ day, $m = 0.5$, $n = 0.1$; q_2, q_3 = dimensionless constants; $C_g(\theta)$ = (age-independent) compliance function for delayed elasticity of the cement gel (hardened cement paste without its pores); $\upsilon(t)$ = volume of gel per unit volume of concrete, growing in time due to hydration; and η_f = effective viscosity for the flow of concrete. By integration, $C_g(\theta) = \ln[1 + (\theta/\lambda_0)^n]$. As for $J(t, t')$, it can be obtained only by numerical integration (however, for computer structural analysis in short time steps, the rate $\dot{J}(t', t')$ suffices and in fact allows a simpler algorithm). For creep of sealed specimens, called the basic creep,

$$1/\eta_f = q_4/t \tag{5}$$

where q_4 = dimensionless constant. Then the flow part of $J(t, t')$ is simply $q_4 \ln(t/t')$. Equations 3, 4, and 5 are the simplest formulas satisfying the asymptotic conditions that $\dot{J}$ for both short and long times θ the aging rate (given by $\mathrm{d}\upsilon^{-1}(t)/\mathrm{d}t$), must be power functions (due to self-similarity conditions, ensuing from the absence of any characteristic time).

Variable Environment: At variable w (mass of water per unit volume of concrete), a physically realistic constitutive relation may be based on the idea of microprestress S, considered to be a dimensionless measure of the stress peaks at the creep sites in the nanostructure. The microprestress is produced by chemical volume changes and by changes in the disjoining pressures in the hindered adsorbed water layers (which can be up to ten water molecules, or 2.7 nm, in thickness) confined between sheets of calcium silicate hydrates. The disjoining pressures must vary with the relative humidity h in the capillary pores, as well as with temperature T, so as to maintain thermodynamic equilibrium (equality of chemical potentials). The rate of bond breakages may be assumed to be a quadratic function of the level of microprestress, which leads to the expression

$$1/\eta_f = q_4 S \tag{6}$$

The microprestress is not appreciably affected by the applied load. It relaxes in time and its evolution at each point of a concrete structure may be solved from the differential equation:

$$\dot{S} + c_0 S^2 = c_1 \left| \dot{T} \ln h + T\dot{h}/h \right| \tag{7}$$

where c_0, c_1 = positive constants (the absolute value ensures that it should never be negative and reflects the fact that not only drying and cooling but also wetting and heating accelerate creep, the latter activating different creep sites than the former). The fact that changes of w or h produce new microprestress peaks and thus activate new creep sites explains the drying creep effect (or Pickett effect). A part of this

effect, however, is caused by the fact that microcracking in a companion load-free specimen causes its overall shrinkage to be less than the shrinkage in an uncracked (compressed) specimen, thus increasing the difference between the two (which is what defines creep).

The concept of microprestress is also needed to explain the stiffening due to aging. One physical cause of aging is that the hydration products gradually fill the pores of hardened cement paste, as reflected in function $v(t)$ in Eq. 3. But hydration ceases after about 1 year, yet the effect of age at loading t' is strong even after many years, even decades. The explanation is that the peak microstress relaxes with age, which reduces the number of creep sites and thus the rate of bond breakages.

At variable environment, time t in Eq. 3 must be replaced by equivalent hydration time $t_e = \int \beta_h \beta_T \mathrm{d}t$ where β_h = decreasing function of h (0 if $h <$ about 0.85) and $\beta_h \propto e^{-Q_h T/R}$, $Q_h/R \approx 2,700$ K. In Eq. 4, $\theta = t - t'$ must be replaced by $t_r - t_r'$ where $t_r = \int \psi_h \psi_T \mathrm{d}t$ = reduced time, capturing the effect of h and T on creep viscosity; ψ_h = function of h decreasing from 1 at $h = 1$ to about 0.1 at $h = 0$; and $\psi_T \propto e^{-Q_h T/R}$, $Q_h/R \approx 5{,}000$ K. The evolution of distributions $h(\mathbf{x}, t)$ ($\mathbf{x}$ = coordinate vector) may be considered uncoupled from the stress and deformation problem and may be solved numerically from the diffusion equation $\dot{h} = \mathrm{div}[C(h)\mathrm{grad}\, h] + \dot{h}_s(t_e)$ where $h_s(t_e)$ = self-desiccation caused by hydration (which is mild in normal concretes but strong in high-strength concretes) and $C(h)$ = diffusivity, which decreases about 20 times as h drops from 100 % to 60 %. The field of free (unrestrained) shrinkage strain rates

$$\dot{\epsilon}_{sh} = k_{sh}\dot{h} \tag{8}$$

where k_{sh} = shrinkage coefficient. Since $\dot{\epsilon}_{sh}$ at various points are incompatible, the calculation of the overall shrinkage of structures as well as test specimens is a stress analysis problem, in which creep and cracking must be taken into account.

For finite element structural analysis in time steps, it is advantageous to convert the constitutive law to a rate-type form. This may be achieved by approximating $C_g(\theta)$ with a Kelvin chain model (or the associated relaxation function with a Maxwell chain model). The history integrals such as Eq. 1 then disappear from the constitutive law, the history being characterized by the current values of the internal state variables, representing the partial strains or stresses of the Maxwell or Kelvin chain.

Conversion to a rate-type form is also convenient for introducing the effect of variable temperature, which affects the Kelvin chain viscosities (according to the Arrhenius law) and also the rate of hydration, as captured by t_e. For a three-dimensional tensorial generalization of Eqs. 2, 3, 4, 5, 6, 7, and 8, see Bazant et al. (1997).

Approximate Cross-Sectional Response at Drying: Although multidimensional finite element calculations of creep and moisture diffusion are nowadays feasible, simplified one-dimensional analysis of concrete beams or girders still reigns in practice. In that approach, one needs to input the average cross-sectional compliance function $\overline{J}(t, t', t_0)$ (Fig. 1 top, light lines) and average shrinkage function $\overline{\epsilon}_{sh}(t, t_0)$ of

the cross section (Fig. 1 bottom) (t_0 = age at start of drying). The algebraic expressions for such average characteristics are considerably more complicated and inaccurate than the constitutive law for a material point because of ignoring the differences due to cross-sectional geometry, reinforcement, and loading (compressive or tensile axial force, bending moment, shear, torque, etc.).

The following approximations (Bažant and Baweja 2000), partly based (under strong simplifications) on the foregoing constitutive relations, have been derived and their coefficients optimized by fitting a large test data bank; for environmental humidity h_e below 98 %,

$$\bar{\epsilon}_{sh}(t, t_0) = -\epsilon_{sh\infty} k_h S(t), k_h = 1 - h_e^3 \tag{9}$$

$$S(t) = \tanh\sqrt{\frac{t - t_0}{\tau_{sh}}}, \tau_{sh} = k_t (k_s D)^2 \tag{10}$$

where $D = 2\ V/S$ = effective thickness, V/S = volume-to-surface ratio, $k_t = 1$ for normal (type I) cement, k_s = shape factor (e.g., 1.0 for a slab, 1.15 for a cylinder), $\epsilon_{sh\ \infty} \approx \epsilon_{s\ \infty} E(607)/(E(t_0 + \tau_{sh})$, $\epsilon_{s\ \infty}$ = constant, and $E(t) \approx E(28)\sqrt{4 + 0.85t}$ = age dependence of Young's modulus (all times are in days). Equations 2, 3, and 4 apply except that $1/\eta_f$ must be replaced by

$$\frac{1}{\overline{\eta_f}} = \frac{q_4}{t} + q_5 \frac{\partial}{\partial t}\sqrt{F(t) - F(t_0')} \tag{11}$$

where $F(t) = \exp\{-8[1 - (1 - h_e)S(t)]\}$ and $t_0' = \max(t', t_0)$. The form of the expression for shrinkage halftime τ_{sh} is based on the diffusion theory. The hyperbolic tangent function in Eq. 9 is the simplest function satisfying two asymptotic conditions ensuing from the diffusion theory: (1) for short times $\bar{\epsilon}_{sh} \propto \sqrt{t - t_0}$, and (2) the final shrinkage must be approached exponentially. Generalizations for the temperature effect have also been made.

In preliminary design of structures, simplified calculations feasible by hand use the creep coefficient:

$$\varphi(t, t') = E(t')J(t, t') - 1 = \frac{\epsilon_{creep}}{\epsilon_{initial}} \tag{12}$$

The change of structure deformation from time t_1 of initial loading to time t can be estimated by elastic analysis in which Young's modulus E is replaced by the so-called age-adjusted effective modulus $E''(t, t_1) = [E(t_1) - R(t, t_1)]/\varphi(t, t_1)$.

Practical Creep and Shrinkage Prediction Model

Analysis and design of structures necessitates a realistic model for predicting the compliance function and shrinkage function of a given concrete. Such a prediction calls for analysis of extensive data from long-term tests conducted on different

concretes and in different environments, generally introducing statistical scatter. A number of prediction models exist which will be briefly introduced here.

Since concrete creep and shrinkage formulations require a prediction of environmental effects, it is natural to begin by considering models for concrete drying. The effects of drying can be taken into account at two different levels of accuracy: the sectional approach only takes into account the mean effects of drying averaged over the cross section of the beam or slab, and the material approach takes into account the time evolution of the distribution of pore humidity point-wise throughout the structure (Jirásek and Bažant 2002). While the material approach allows the researcher to capture the effect of the drying mechanisms on the constitutive behavior of the material, it is too detailed for engineering calculations. For multi-decade predictions of the degradation of structures, a cross-sectional model that incorporates the influence of cross-sectional size and shape on the kinetics of drying in an approximate manner, through semiempirical coefficients, is typically employed.

The prediction models with the deepest theoretical basis are the B3 (Bažant and Baweja 1995, 2000) and the B4 (Bažant et al. 2014) models. They are based on the same theoretical foundation, but the B4 model was developed recently to extend the formulation for modern concretes with autogenous shrinkage and for more accurate multi-decade estimates (see section "Rate-Type Creep" for details). The sectional approach generally defines the compliance function of concrete in the form

$$J(t,t') = q_1 + C_0(t,t') + C_d(t,t',t_o) \tag{13}$$

where $q_1 = 1/E_0$ is the inverse of the asymptotic elastic modulus, $C_0(t,t')$ is the basic creep compliance, and $C_d(t,t',t_0)$ is the drying creep compliance, which is influenced by the time the specimen begins to dry, t_0. The average longitudinal drying shrinkage of a cross section for a long beam or plate may be approximately calculated as

$$\varepsilon_{sh}(t) = -\varepsilon_{sh\infty} k_h S(t - t_0) \tag{14}$$

where $\varepsilon_{sh\infty}$ is the magnitude of the final shrinkage strain, k_h is a coefficient depending on the average environmental humidity h (relative vapor pressure), and $S(\hat{t})$ is an increasing function of the duration of drying. This function describes the evolution of normalized shrinkage strain in a perfectly dry environment. A suitable formula is

$$S(\hat{t}) = \tanh\sqrt{\frac{t}{\tau_{sh}}} \tag{15}$$

where τ_{sh} is called the shrinkage halftime, because it roughly describes the time at which ε_{sh} reaches one-half of its final value. The shrinkage halftime can be estimated as

$$\tau_{sh} = k_t (k_s D)^2 \tag{16}$$

in which k_s is a cross-sectional shape factor based on solutions of the nonlinear diffusion equation for drying of concrete (Bažant and Najjar 1972) and $D = 2\,V/S =$ effective cross-sectional thickness. Factor k_t is approximately the normalized diffusivity of pore water in saturated concrete. The fact that τ_{sh} is proportional to the square of the thickness is a basic property of diffusion processes in general. Another consequence of the diffusion origin of drying shrinkage is that the initial shrinkage curve must evolve as a square root function of time. The final shrinkage value can be estimated empirically in terms of intrinsic and extrinsic parameters of each particular concrete.

According to the B3 model, the basic creep compliance is most conveniently expressed by its time rate

$$\frac{\partial C_0(t,t')}{\partial t} = \frac{n(q_2 t^{-m} + q_3)}{(t-t') + (t-t')^{1-n}} + \frac{q_4}{t} \tag{17}$$

in which t and t' must be in days and m and n are empirical parameters whose values can be taken the same for all normal concretes; q_2, q_3, and q_4 represent empirical relations for constitutive parameters. The total creep compliance is obtained by integrating this rate equation with the initial condition that the compliance after zero duration is zero. Since this integration leads to a binomial integral, the result must be expressed approximately as

$$C_0 = q_2 Q(t,t') + q_3 \ln\left[1 + (t-t')^n\right] + q_4 \ln\left(\frac{t}{t'}\right) \tag{18}$$

where $Q(t,t')$ is a function that can be obtained by numerical integration or by interpolation from a table computed in Bažant and Baweja (1995, 2000). The terms in this form represent the aging viscoelastic compliance, nonaging viscoelastic compliance, and flow compliance, respectively, based on the solidification theory (Bažant and Prasannan 1989; Carol and Bažant 1993; Bažant et al. 1997a, b). The additional mean cross-sectional compliance caused by simultaneous drying, expressing a coupling between creep and shrinkage, can be estimated from the formula (Bažant and Baweja 1995)

$$C_d(t,t',t_0) = q_5\left(e^{-g(t-t_0)} - e^{-g(t'-t_0)}\right) \tag{19}$$

Prediction Models of Engineering Societies

A number of other concrete creep and shrinkage prediction equations are available in practice. However, these are not typically derived from theory, but developed empirically. They include the American Concrete Institute's model (ACI 2008), the

Table 1 Summary of creep and shrinkage model time functions

(a) Creep model		
Creep model	Time function	
B4	Eqs. 6 and 7	
B3	Eqs. 6 and 7	
MC10	$\left(\frac{t}{\beta+t}\right)^{\gamma}$	
MC99	$\left(\frac{t}{\beta+t}\right)^{0.3}$	
GL00	$\beta\left(\frac{t^{0.3}}{14+t^{0.3}}\right)+\gamma\left(\frac{t}{7+t}\right)^{0.5}+\left(\frac{t}{\gamma+t}\right)^{0.5}$	
ACI92	$\frac{1}{\beta}\left(1+\frac{t^{\psi}}{d+t^{\psi}}\right)$	
(b) Shrinkage model		
Shrinkage model	Time function	Autogenous time function
B4	$\tanh\sqrt{t/\tau_{sh}}$	$\left[1+\left(\frac{\tau_{au}}{t}\right)^{\alpha}\right]^{r_t}$
B3	$\tanh\sqrt{t/\tau_{sh}}$	–
MC10	$\sqrt{\frac{t}{\alpha+t}}$	$1-e^{-\alpha\sqrt{t}}$
MC99	$\sqrt{\frac{t}{\alpha+t}}$	$1-e^{-\alpha\sqrt{t}}$
GL00	$\sqrt{\frac{t}{\alpha+t}}$	–
ACI92	$\frac{t^{\alpha}}{f+t^{\alpha}}$	–

European Model Code (FIB 1999, 2010), and the GL model, developed by Gardner in Canada (Gardner 2000; Gardner and Lockman 2001). Design codes in most countries require a prediction model that is based on one of these more empirical equations. Table 1 illustrates how their time functions relate to those of the B3 model that was presented above.

Limitations of Current Practical Creep and Shrinkage Models

The existing linear creep models have a number of limitations that were brought to attention with the release of data from the 1996 collapse of the Koror–Babeldaob prestressed concrete bridge in Palau. Bažant et al. (2012c) have demonstrated that theoretically incorrect and out-of-date prediction equations were responsible for underestimations of the concrete bridge's long-term creep and shrinkage (Bažant et al. 2012c). The standardized compliance functions traditionally approach a final asymptotic bound which no bound actually exists, since the long-term creep is logarithmic. Another problem is that in the ACI (ACI 2008), MC99 (FIB 1999, 2010), and other formulations, the compliance function does not separate the basic and drying creep which have a different dependence on cross-sectional size.

Generally, creep and shrinkage data for concrete is too scattered to empirically extract a time function as seen in Fig. 3. Additionally, the effect of the specimen

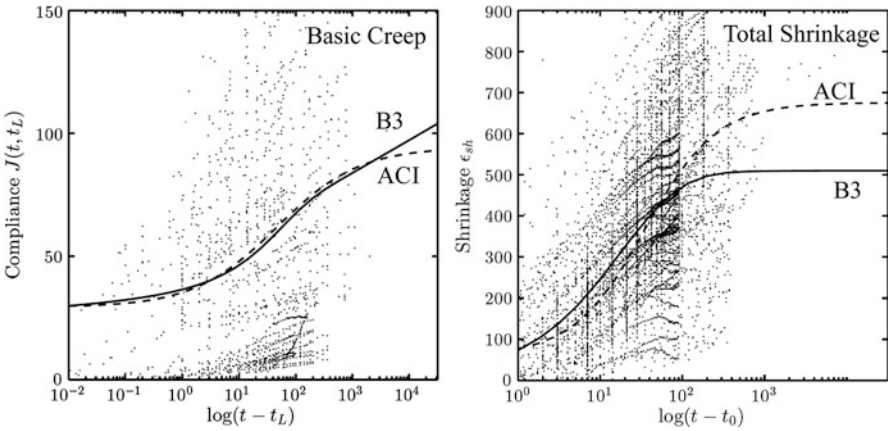


Fig. 3 Creep compliance and shrinkage strain data points from tests taken within a narrow range of composition parameters. The scatter in the data is too large to distinguish between the B3 and ACI functional forms (Wendner et al. 2014a)

size is described by vertical scaling, while in reality the change of size causes a horizontal shift in log-time. Only the B3 (Bažant and Baweja 1995) and GL (Gardner 2000; Gardner and Lockman 2001) formulas (as well as the new model B4) do not have this weakness. The new model B4 which is presented in section "Development of the B4" overcomes a number of these limitations.

Study of Excessive Deflections of Collapsed World-Record K–B Bridge in Palau

K–B Bridge Description and Modeling

The Koror–Babelthuap Bridge (K–B Bridge) was built to connect the islands of Koror and Babelthuap (now Babeldaob) in the Republic of Palau, an island country located in the West Pacific Ocean. This two-lane segmentally constructed bridge had a main span of 241 m (791 ft.), which set a world record for prestressed concrete box girder when it was completed in April 1977 (Yee 1979).

The main span of K–B Bridge consisted of two symmetric cantilevers, each of which consisted of 25 cast-in-place segments, with their depths varying from 14.17 m (46.5 ft.) at the main piers to 3.66 m (12 ft.) at the midspan. The main span was flanked by two side spans, which were partially filled with rock ballast to balance the overturning moment at the main pier. The dimensions of cross sections of the segments can be found in a number of technical reports and post-collapse investigations (Yee 1979; Pilz 1997; McDonald et al. 2003; Bažant et al. 2012a). For the superstructure, type I Portland cement was used. The mix design and the 28-day mean compressive strength were reported in a recent investigation based on the information obtained from the resident engineer at the K–B Bridge construction

(Bažant et al. 2012a). Although the original measurements of the 28-day Young's modulus are not available in literature, estimates can be made based on the core sample tests and truck load tests by two independent investigation teams (JICA 1990; Berger/ABAM Engineers 1995; Bažant et al. 2012a).

There were 316 prestressing tendons above the main pier, densely packed in four layers within the top slab. The jacking force of each tendon was about 0.60 MN, or 135 kips (DRC 1996). Their combined initial prestressing force was about 190 MN, or 42606 kips (Yee 1979; Pilz 1997; McDonald et al. 2003). The same Dywidag threaded alloy bars were used to provide vertical prestress in the webs and horizontal transverse prestress in the top slab. In addition to prestressing tendons, there was also normal (non-prestressed) steel reinforcement (ABAM 1993) in webs and bottom slabs. None of the slabs had cross ties.

The segmental construction of the opposite symmetric cantilevers was almost simultaneous (T. Y. Lin International 1996). Although the construction was closely monitored, the camber planned to offset the anticipated long-term deflections was not met. The creep and shrinkage during the segmental erection caused an initial sag equal to 229 mm (9 in.) at midspan. This initial sag before installation of the midspan hinge is not included in the reported deflection measurements, neither considered in numerical simulations.

The initial deflections for the first 2 years were benign. However, the long-term deflections came as a surprise. In 1990, the midspan deflection since span closing increased to 1.22 m, or 48 in. (JICA 1990), which caused noticeable pavement degradation and partial serviceability loss. By 1993 (ABAM 1993), the deflection grew to 1.32 m (52 in.). In 1995, just before the removal of roadway pavement for the planned retrofit, the midspan deflection reached 1.39 m (54.7 in.) (which represented a deflection of 1.61 m or 63.3 in. compared to the design camber) and was still growing (Berger/ABAM 1995).

To recover the serviceability severely impaired by the unexpectedly excessive deflection, a retrofit was made and completed in 1996, during which remedial prestressing lifted the midspan to the expected level. Unfortunately, 3 months later, the whole bridge suddenly collapsed in a catastrophic manner and with fatalities (Pilz 1997; McDonald et al. 2003).

The collapse was apparently triggered by creep buckling of a slice of previously delaminated top slab subjected to compression by newly added prestressing tendons. This buckling released a huge portion of the prestressing force and emitted a shock wave which overloaded the corner between the pier and the bottom slab and thus triggered a compression shear failure of the box girder (which is a type of failure subject to size effect). In the post-collapse examination, neither the prestressed nor the non-prestressed steel showed any signs of significant corrosion, despite the tropical marine environment (some of the ducts, though, showed mild, yet inconsequential, degradation by corrosion).

Despite its tragic failure, the K–B Bridge provides a real and informative example to study the long-term deflection of the prestressed concrete box girders. Taking advantage of symmetry, only one-half of the bridge needs to be modeled by three-dimensional (3D) finite element (FE) analysis. The modeling

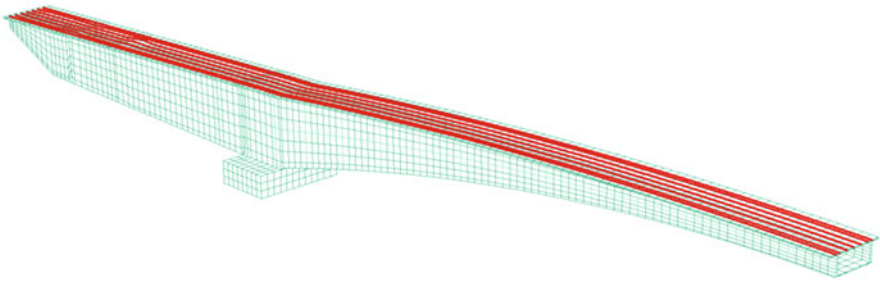


Fig. 4 3D FEM model of K–B Bridge in ABAQUS (With permission from ASCE, Bažant et al. 2012b)

demands a suitable program of the requisite geometric and material modeling features to mimic the construction and service of the K–B Bridge. In the present investigation, the software ABAQUS (SIMULIA, Providence, Rhode Island) has been selected.

Together with the pier, the concrete members are subdivided into 5,036 hexahedral elements. The prestressing tendons and the non-prestressed steel bars are subdivided into further 6,764 bar elements connected rigidly (with no slip) to the nodes of the three-dimensional elements of concrete. The meshed structure is shown in Fig. 4. Since the creep is essentially viscoelastic, not suffering from strain localization instabilities, the sufficiency of the mesh fineness could be, and was, validated by demonstrating a negligible difference in the elastic deflection when compared with a finer mesh of 20,144 hexahedral elements (Bažant et al. 2012b).

As an adequate approximation under service conditions, concrete can be assumed to follow aging linear viscoelasticity with corrections for tensile cracking, variations of humidity and temperature, and drying creep (or Pickett effect). The concrete time-dependent deformation is then fully characterized by one of the existing prediction models for the shrinkage strain and the creep compliance function. The primary prediction models used in practice around the world, namely, the ACI model (ACI 1971, 2008), the CEB (or CEB-FIP, *fib*) model (FIB 1999), the JSCE model (JSCE 1991), the GL model (Gardner 2000; Gardner and Lockman 2001), and the B3 model (Bažant and Baweja 1995, 1996, 2000; Bažant and Prasannan 1988, 1989a, b; Jirásek and Bažant 2002), have been considered in the K–B Bridge analysis.

In the creep prediction models, two types of parameters are employed: extrinsic and intrinsic. The extrinsic parameters characterize the geometry and construction procedures such as the construction sequence, loading, curing, age at loading, relative humidity, and volume–surface ratio, while the intrinsic parameters define the material properties, particularly the compliance and shrinkage functions. For ACI, CEB, and GL models, which are purely empirical, the only important intrinsic parameter is the standard 28-day compressive strength. Contrarily, the B3 model, based on solidification theory, uses more than one intrinsic parameter to capture the main aspects of the concrete properties. In the B3 model, these intrinsic parameters are q_i ($i = 1, 2, \ldots 5$) in the compliance function.

Since the concrete strength in the K–B Bridge is known based on cylinder tests, the input parameters for ACI, CEB, and GL models are almost fixed. Therefore, only one set of input parameters (Set 1) is available for the analysis with these models (Bažant et al. 2012a). On the other hand, for model B3, there are two sets of input parameters, Set 1 and Set 2. In Set 1, the default formulas for predicting material parameters by various models are used (Bažant et al. 2012a).

The purpose of Set 2 is to check whether the observations can be explained with model B3, i.e., whether the observed deflection data and prestress loss data can be fitted with realistic intrinsic parameters of the model. The answer was affirmative, and the fitting parameters were almost the same as those fitting the Brook's (1984, 2005) long-term test data. This showed there was no need to seek the explanation of the excessive deflections in some extraneous hypothetical causes, such as poor-quality control during construction. A successful fit by Set 2 also means that the model is not fundamentally flawed. If it were, it could not capture the long-term deflection curve and its asymptote, no matter how the intrinsic parameters are adjusted.

Computed Deflections and Prestress Loss, Compared to Measurements

The creep models are incorporated in an improved rate-type algorithm to run the three-dimensional (3D) simulations (Yu et al. 2012). Before the creep analysis, the validity of the computational model established in ABAQUS is first examined by making comparison with the in situ truck load test in 1990 (Bažant et al. 2010, 2012a). Under the load exerted by two fully loaded trucks at the cantilever tips, the average downward deflection obtained by the present numerical model agrees well with the recorded measurements.

By utilizing the features provided by ABAQUS, the effects of the prestressing and the construction sequence get reproduced automatically in the simulations. The results of the predicted 19-year deflections are shown in Fig. 5, in both linear and logarithmic time scales. One benefit of plotting the deflection history in the logarithmic time scale is to highlight the asymptotic trend of the long-term deflection, which would be obscured in the linear scale. This long-term asymptote is an essential aspect in creep model evaluation. Indeed, if a creep model cannot capture this asymptotic trend, it means that there is fundamental inadequacy in the model and that the long-term prediction is likely to be misleading.

The predictions based on ACI, CEB, JSCE, GL, and B3 models are plotted in Fig. 5 and compared with the measured deflection, represented by diamond points. After 19 years, the measured deflection at midspan is roughly three times larger than that calculated (by rate-type 3D FEM analysis) with the ACI or CEB model and approximately double of the deflection calculated with the GL model. Furthermore, the ACI, CEB, JSCE, and GL deflection curves have shapes that are rather different from those of the B3 model, as well as that of the observed deflection history. Their

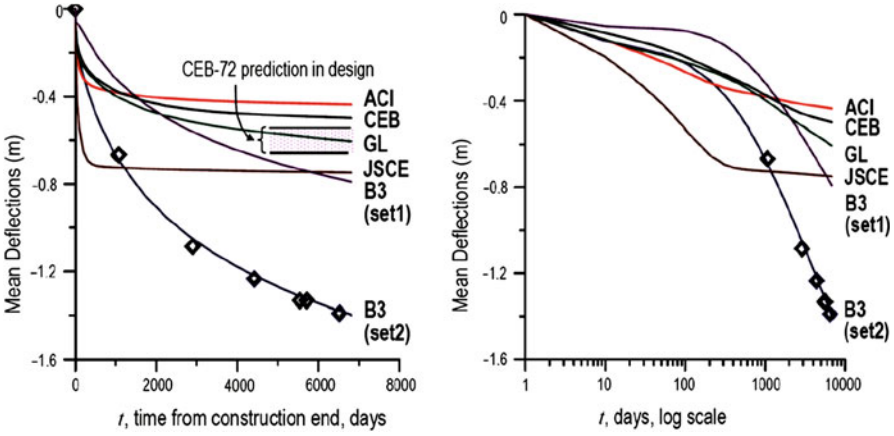


Fig. 5 19-year deflections of K–B Bridge based on different creep models (with permission from ASCE, Bažant et al. 2012a)

long-term asymptotes, which level off in logarithmic time scale, deviate substantially from the inclined straight line shown in Fig. 5. As for model B3, although its prediction based on the default parameters (Set 1) underestimates the midspan deflection, its long-term asymptote agrees with the measured asymptote. According to the updated Set 2, model B3 not only gives a good estimate of the final deflection but also closely approximates the shape of the recorded deflection curve.

The prestress loss predicted by different models is shown in Fig. 6, also in both linear and logarithmic time scales. The 19-year prestress loss was predicted (by - rate-type 3D FEM analysis) to be only 22 % and 24 % when the ACI and CEB models were used, which is substantially lower than the 46 % loss when the model B3, Set 2, was employed (Bažant et al. 2012a). The correctness of the prestress loss predicted by the B3 model is confirmed by the stress relief tests that were made by ABAM on three tendons just before the retrofit (Berger/ABAM 1995). The average residual stress obtained from nine measurements on the tendons was 377 MPa (54.7 ksi), indicating that the average prestress loss over 19 years was approximately 50 %. The prediction of model B3, Set 2, deviates from the measured mean by only 4 %, less than the coefficient of variation of these measurements, which is 12.3 %. Similar tests were also conducted by another investigating company (Wiss, Janney and Elstner, Highland Park, Illinois), and the average measured prestress loss was almost the same.

Creep and shrinkage are notorious for their relatively high random scatter, resulting from fluctuation of the material properties as well as variation of environmental conditions. Therefore, it is more appropriate to design the bridge based on some suitable confidence limits rather than the deterministic mean deflection (Bažant and Liu 1985; Bažant et al. 2010). An efficient way to generate the deflection statistics is to use Latin hypercube sampling of the input parameters (Bažant and Liu 1985; Bažant and Kim 1989). The confidence limits can easily be

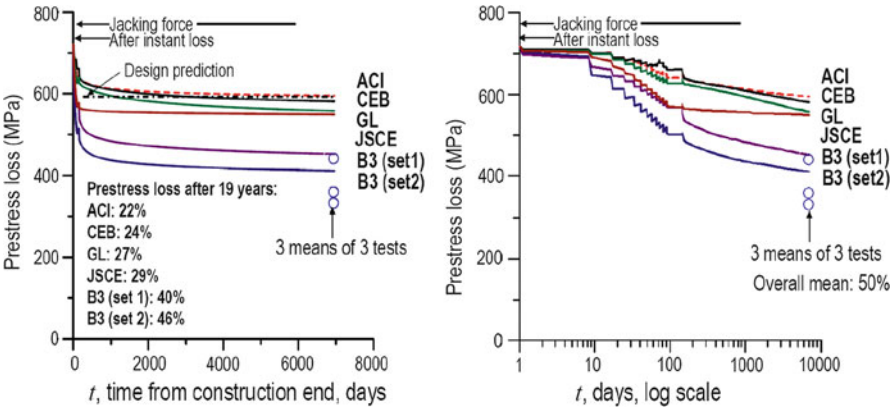


Fig. 6 19-year prestress loss of K–B Bridge based on different creep models (with permission from ASCE, Bažant et al. 2012a)

obtained by repeating the deterministic FEM analysis of a bridge according to the samples of input parameters generated randomly by Latin hypercube sampling (Bažant et al. 2010, 2012a).

The numerical investigation of the K–B Bridge shows that 3D simulations, based on a realistic creep and shrinkage model and utilizing efficient rate-type approach, are crucial for creep structural analysis. Among the models used in practice, the B3 model is demonstrably capable of capturing the long-term asymptote of concrete creep and thus is able to approximate the multi-decade deflection of prestressed concrete bridges, although an update of the intrinsic parameters based on long-term creep tests may be needed. For example, the measured deflections of four Japanese bridges and one bridge in Czech Republic can be closely approximated by 3D simulations based on model B3 with updated intrinsic parameters (Bažant et al. 2012a).

Wake-Up Call: Excessive Long-Term Bridge Deflections

Collection of Excessive Deflection Histories

Prompted by the release of data on the K–B Bridge in Palau and their analysis, an effort to collect data on other bridges was conducted in the Infrastructure Technology Institute of Northwestern University in collaboration with the RILEM Committee TC-MDC. Private communications from other construction firms and the scanning of various papers and reports (Manjure 2001–2002; Burdet and Muttoni 2006; Pfeil 1981; Fernie and Leslie 1975; JICA 1990; Patron-Solares et al. 1996) led to a collection of histories of deflections, excessive or nearly excessive, of 69 bridge spans, of which 56 are shown in Fig. 7. It is likely that hundreds of such cases exist around the world. All the bridges in Fig. 7, except one (the Gladesville Arch), are long-span, segmental prestressed box girders, mostly with midspan

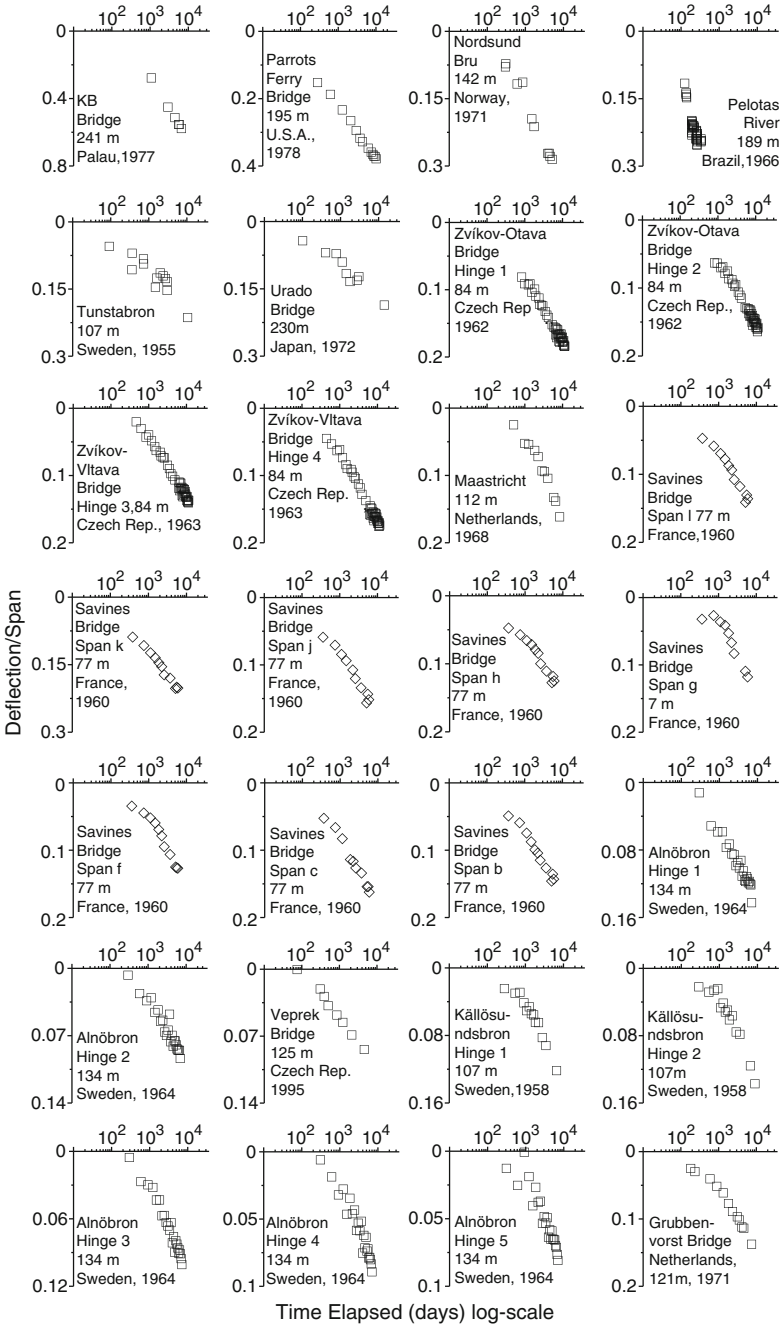


Fig. 7 Histories of excessive deflections of the first 28 among 56 large bridge spans (reprinted with permission from the American Concrete Institute, Bažant et al. 2011)

hinges; however, at least six of them (the Parrotts Ferry, Grubbenvorst, Wessem, Empel, Hetern, and Ravenstein Bridges) are continuous. The elimination of a midspan hinge reduces deflection, but often not enough, as documented in detail by the Labe Bridge in Děčín. Many of the listed sources on bridge deflections provide, unfortunately, only sketches and limited cross-sectional information that does not suffice for finite element modeling.

What is most interesting is that all these deflection histories terminate with an inclined straight line in the logarithmic time scale, which corresponds to a logarithmic curve in the actual time scale. This feature, which was introduced in 1975 in an analysis of nuclear containment (refer to Fig. 4 in Bažant 1975) on the basis of L'Hermite et al.'s (1965, 1969) and L'Hermite and Mamillan's (1968) test data, agrees with the prediction of the theoretically based model B3 and is also supported by other existing longtime laboratory tests (Brooks 1984, 2005; Burg and Orst 1994, Russell and Larson 1989; Browne and Bamforth 1975; Hanson 1953; Harboe 1958, Troxell et al. 1958; Pirtz 1968) (e.g., refer to Figs. 2.2, 2.7, 2.10, 2.24, and 2.28 in RILEM 1988 or Figs. 1, 2, 3, and 4 in Part 2 and Figs. 1, 3, and 4 in Part 2 of the paper by Bažant et al. 1992). Note that, similar to the aforementioned laboratory creep tests, there is no sign of the deflection curves approaching a finite bound.

By contrast (and despite the fact that logarithm terminal curve was proposed to ACI Committee 209 in 1974), the ACI and other existing creep prediction models of engineering societies, including the ACI Committee 209, CEB-FIP, GL, JSCE, and JRA models (CEB-FIP 1990; ACI 1971; ACI 2008; CEB-FIB 1990, 2010; Gardner and Lockman 2001; JSCE 1991; JRA 2002), except the 2012 update of *fib* Model Code 2010, have a form that implies a horizontal asymptote or a finite upper bound on creep. This erroneous assumption has doubtlessly been caused by the habit in most of the engineering literature to plot the creep curve only in the actual time scale with an elongated time axis. When plotted that way, even the logarithmic curve gives an illusion of approaching a bound, although none exists.

The horizontal dashed lines in Fig. 9 represent the deflection equal to 1/800 of the span, which is considered to be the acceptable limit in bridge design specifications (AASHTO 2004). This limit is exceeded within the time range of available measurements by 16 of the 56 analyzed bridge deflection histories and by 26 of them if the straight-line extrapolations to 100 years are considered (100 year lifetime is nowadays the generally required design lifetime). Note that 56 bridges were analyzed but only 35 could be shown in the figure. Based on the data in Fig. 7 and their straight-line extrapolations, the limit of 1/800 is exceeded by 36 spans within 24 years, 39 spans within 40 years, and 50 spans within 100 years among the 56 spans in Fig. 7.

Approximate Multi-Decade Extrapolation of Deflection

According to model B3 (Bažant and Baweja 2000) and the preceding 1991 BPKX model (Bažant et al.1992) developed at Northwestern University, the longtime asymptote of the compliance curve at fixed t' is logarithmic. This feature is supported by the aforementioned laboratory data. The time at which the creep

curve becomes a straight line in the logarithmic time scale depends on many factors – on average, it is approximately 3 years.

It is interesting that, after several years, the bridge deflection curve also becomes a straight line in the logarithmic time scale (Fig. 7). The reason for this must be that the effects of the age differences between segments, the variation of the self-weight bending moment during cantilever construction, the differences in the slab thickness, and the change of the structural system at span closing nearly die out. Also, the transient processes, particularly the drying effect on creep and shrinkage, the gradual filling of capillary pores by cement hydration products, the acceleration of creep by drying, and the prestressing steel relaxation rate, greatly attenuate within a few years.

At earlier times, the drying effects greatly distort the deflection curve. Because the top slab of a segmental box girder is much thinner near the support than the bottom slab, its shrinkage and drying creep become accelerated. This reduces the midspan deflection and may even cause a temporary upward deflection (Bažant 1972). Further complications of the short-term deflection history are caused by the gradual rise of the bending moment at the pier during the segmental erection and the age differences among the segments of the box girder. Therefore, the prediction of deflections during the first few years requires sophisticated FE creep analysis (Bažant et al. 2010, 2012a, b).

Nevertheless, the straight-line trends of longtime deflections in the logarithmic scale suggest that if deflection w_m at time t_m, such as 1,000 days, is known, it could be simply extrapolated to long times by assuming similarity to $J(t,t')$. To keep the extrapolation easy, two simplifications of the regime prior to span closing are needed to be introduced: (1) the age differences among the box girder segments must be ignored, and the age of the concrete must be characterized by one common effective (or average) age t_c at the span closing; (2) instead of the gradual increase of the bending moment in the cantilever segments during the erection, one common effective (or average) age t_a at which the self-weight bending moments are introduced in the erected cantilever must be considered. In the following, the values $t_c = 120$ days and $t_a = 60$ days are considered for all the bridges.

Because of these simplifications and the complexity of the drying and hydration process in the early years, the long-term deflections cannot be assumed to grow in proportion to $J(t,t_a)$. Nevertheless, for the additional deflection w that develops after the span closing time t_c, the errors in approximating the early loading history by t_a and t_c must decay with time and eventually become negligible when $t >> t_c$ – that is, after the lapse of a sufficient time t_m. As shown in the following, the aforementioned time $t_m = 1{,}000$ days (measured from span closing) seems appropriate.

Before the span closing and for a few years afterward, the drying process and the differences in concrete age make the box girder response very complicated. After these effects nearly die out, however – that is, for $t > t_m$ – the box girder begins to behave as a nearly homogenous structure, for which the growth of deflection w should be approximately proportional to the increment of the compliance function that has developed since the closing time t_c – that is, $w = C\,[J(t,t_a) - J(t_c,t_a)]$, where C is a certain stiffness constant. The values of C or w_m can vary widely, and their calculation would require a detailed finite element analysis, considering creep with drying and the construction sequence. Unfortunately, for most of the bridge deflection curves in Fig. 8,

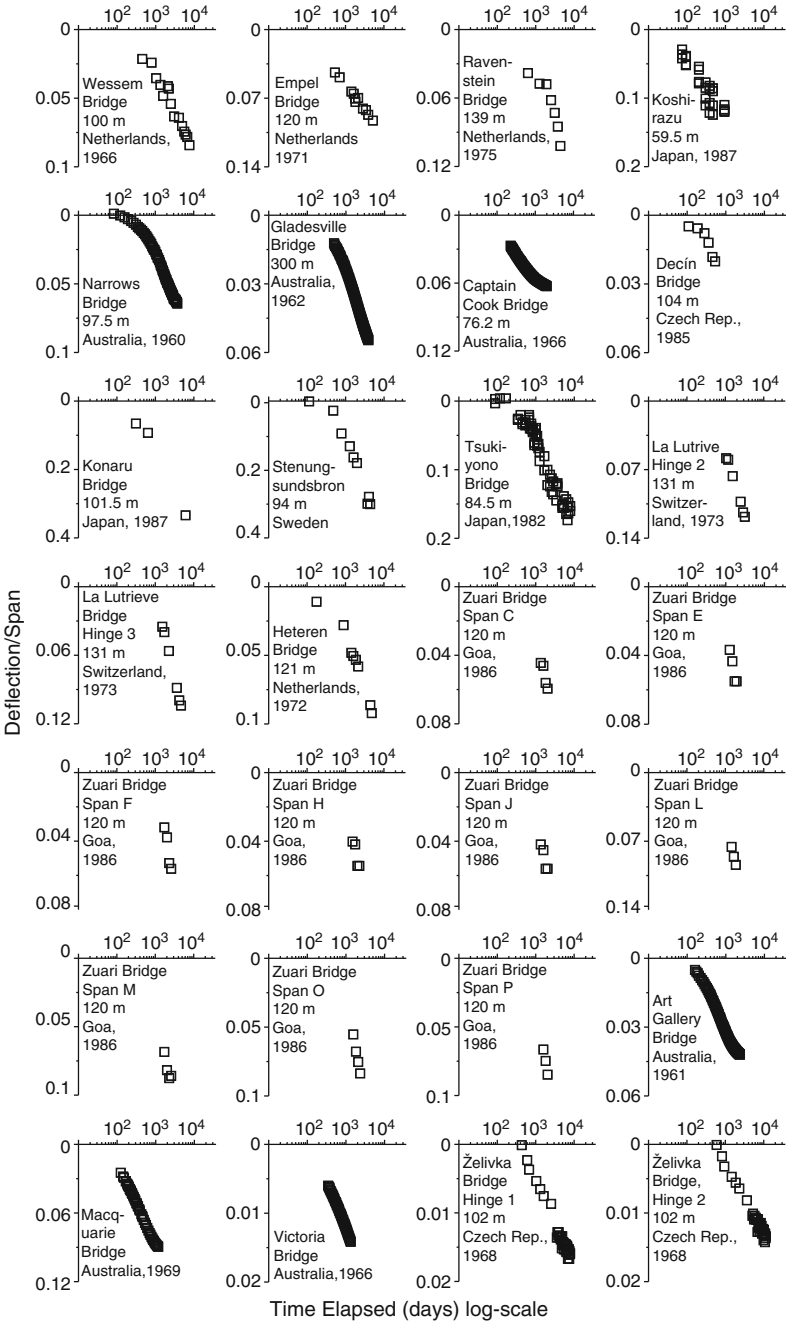


Fig. 8 Histories of excessive deflections of the remaining 28 among 56 large bridge spans (reprinted with permission from the American Concrete Institute, Bažant et al. 2011)

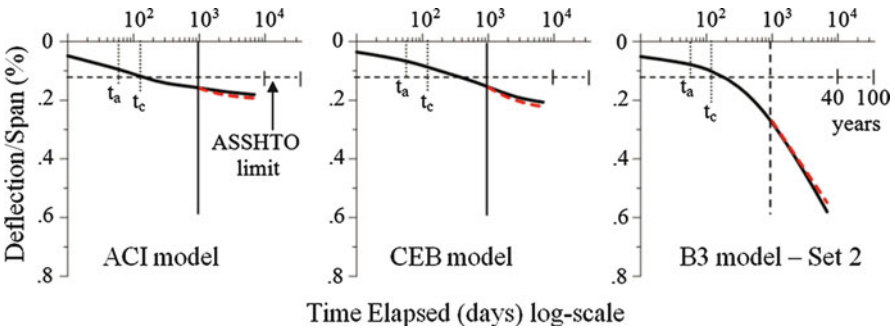


Fig. 9 Comparison of extrapolation formula using the correct concrete strength and composition with the K–B bridge deflections accurately calculated by finite element creep analysis (reprinted with permission from the American Concrete Institute, Bažant et al. 2011)

it turned out to be impossible to obtain the data necessary to calculate C from the material properties, geometry, and construction sequence. Therefore, only the extrapolation from t_m can be examined, assuming that w_m is known.

Therefore, C can be calibrated experimentally from w_m using $C = w_m/[J(t_m, t_a) - J(t_c, t_a)]$. For the extrapolation of deflection after time t_m, the following approximate formula has thus been obtained (Bažant et al. 2011):

$$w(t) = w_m \frac{J(t, t_a) - J(t_c, t_a)}{J(t_m, t_a) - J(t_c, t_a)} \tag{20}$$

To check how good this formula is, the deflection curves accurately calculated by finite elements for the K–B Bridge using the B3, ACI Committee 209, and CEB-FIP material models can be used to an advantage. For each curve, Eq. 20 can be used to extrapolate w_m at 1,000 days from the computed deflection using the same compliance function $J(t,t')$ as that from which the curve was computed. The resulting extrapolations are shown in Fig. 9. It is astonishing how close each extrapolation is to the computed curve for the corresponding model; therefore, it makes sense to compare the extrapolations according to this formula to the observed long-term deflection curves of various bridges.

In theory, Eq. 13 should be applied only if the bending moments caused in the girder after time t_a by the self-weight and the prestress are approximately constant. Because the additional prestress loss after time t_m = 1,000 days is very small, assuming the constancy of the bending moments should be a very good approximation for bridges with a midspan hinge. For a segmental bridge that was made continuous through the midspan, the internal forces redistribute so as to approach the bending moment distribution for an elastic continuous bridge. This redistribution after time t_m could be taken into account by generalizing Eq. 20 according to the age-adjusted effective modulus method; for example, see (Jirásek and Bažant 2002) and (Bažant 1972). However, complete information on the bridge geometry and prestress would be needed for this purpose. It is, unfortunately, unavailable for

most of the bridges with no hinge in Fig. 7, except the Děčín and Vepřek Bridges. Even for these two bridges, however, the degree of redistribution after 1,000 days must have been very small; this can be explained by the relative shallowness and high flexibility of the cross section at the midspan.

Comparison with Deflections Extrapolated from Creep and Shrinkage Models

The input characteristics required by all the creep prediction models are the average compressive strength of concrete $\overline{f_c}$, the environmental relative humidity H, and the effective cross-sectional thickness D. In addition, RILEM model B3 uses the water/cement ratio (w/c), the specific cement content c, and the aggregate/cement ratio (a/c) (the a/c value is implied by the specific weight ρ of concrete) as input, and if these additional input values are unknown, the recommended default values are used. Although the drying creep term of model B3 has little effect on the deflection rate after time t_c, it affects the creep from t_o to t_c; therefore, it must be included in calculating $J(t,t')$ from model B3. Because Eq. 20 cannot take into account the effect of the variation of the slab thickness within box girders, an approximation in which a single average or effective thickness D is used to calculate $J(t,t')$ must be used ($D = 2V/S$, where V/S is the volume–surface ratio of an average cross section).

To apply Eq. 20, the mean concrete strength $\overline{f_c}$ and w/c, c, and ρ for model B3 must be specified. Unfortunately, these parameters are only known for six bridges among the 36 that were analyzed. Therefore, individual comparisons for each bridge are impossible. Nevertheless, a useful comparison, at least in the mean sense for all the bridges combined, can be made.

It is assumed that the concrete design strength in these older bridges was, on average, 31 MPa (4,500 psi), which implies (according to CEB-FIP) (FIB 1999) that the mean strength was at least 39 MPa (5,660 psi). Furthermore, the average effective cross-sectional thickness of $D = 0.25$ m (10 in.) and the environmental humidity of 70 % for the Scandinavian bridges (the NorsundBru, Tunstabron, and Alnöbron Bridges) and 65 % for the other bridges are assumed. For the other parameters, it is assumed that $w/c = 0.5$, $c = 400$ kg/m^3 (25 lb/ft^3), and $\rho = 2{,}300$ kg/m^3 (143 lb/ft^3). Of course, the deflection curve extrapolated in this way from w_m will likely be incorrect for each particular bridge. Nevertheless, because the errors should be of alternating signs, compensating each other, the mean of the extrapolations for all the bridges should still be approximately equal to the mean of the correct extrapolated long-term trend of the deflection curve that would be obtained if the properties of each individual concrete were known.

The last 19 of the 56 bridge spans in Figs. 7 and 8 (counted from the bottom) were omitted from the extrapolation exercise for three reasons: (1) not enough measurements were made, (2) the deflections were not too excessive, and (3) the straight-line regime has not yet been entered at 1,000 days, which means that the drying effects still continued. This occurred for the Konaru, Tsukiyono, Stenungsbron, Želivka, and Victoria Bridges. Moreover, one more figure had to be omitted to

obtain in Fig. 7 a rectangular array, and it was the Savines Bridge span b because its plot is essentially identical to span c. This reduced the number of extrapolations according to Eqs. 8–36.

The extrapolations obtained with the B3, ACI Committee 209, CEB-FIP, and GL models are shown in Fig. 10 by different lines – continuous for model B3; light, dashed lines for the ACI Committee 209 model; dash-dot lines for the CEB-FIP model; and dark, dashed lines for the GL model. None of these models is considered satisfactory because they all systematically and significantly underestimate the measured longtime deflections. Nevertheless, RILEM model B3 does not perform as poorly as the others.

Updating Long-Term Prediction Capabilities

RILEM model B3 has two important advantages:

- The longtime form of model B3 is a logarithmic curve (shown as straight line in the figures), which agrees with the longtime trend of the deflection data, whereas the longtime curves for the ACI Committee 209, CEB-FIP, and GL models (ACI 1971, 2008; FIB 1999; Gardner 2000; Gardner and Lockman 2001) (as well as the JSCE and JRA models from 1996 to 2002, respectively) level off as they approach a horizontal asymptote, and
- Model B3 is the only model that can be updated without compromising the short-time performance because the slope of the straight longtime asymptote can be separately controlled.

From Fig. 8, one can determine for each bridge span i ($i = 1, 2, \ldots N, N = 36$) the ratio of the actual observed terminal slope r_i to the deflection slope extrapolated with model B3. The mean ratio

$$\overline{r} = \sum_{i=1}^{N} r_i / N \tag{21}$$

may then be applied to modify model B3 such that it would not systematically underestimate the longtime extrapolation of creep deflections.

According to RILEM model B3 by (Bažant and Baweja 1995), the terminal asymptotic deflection slope in the $\log(t\text{-}t_c)$ scale is proportional to $q_4 + nq_3$, where n is an exponent of the viscoelastic term equal to 0.1 and q_3 and q_4 are model B3 parameters obtained from empirical formulas as a function of w/c, c, a/c, and mean concrete strength $\overline{f_c}$, for which the aforementioned default values are used.

The parameter values resulting from these formulas have been proposed (Bažant et al. 2011) to be updated by the factor $\overline{r}$, yielding corrected parameters

$$q_3 \leftarrow \overline{r} q_3, \quad q_4 \leftarrow \overline{r} q_4 \quad \text{with } \overline{r} = 1.6 \tag{22}$$

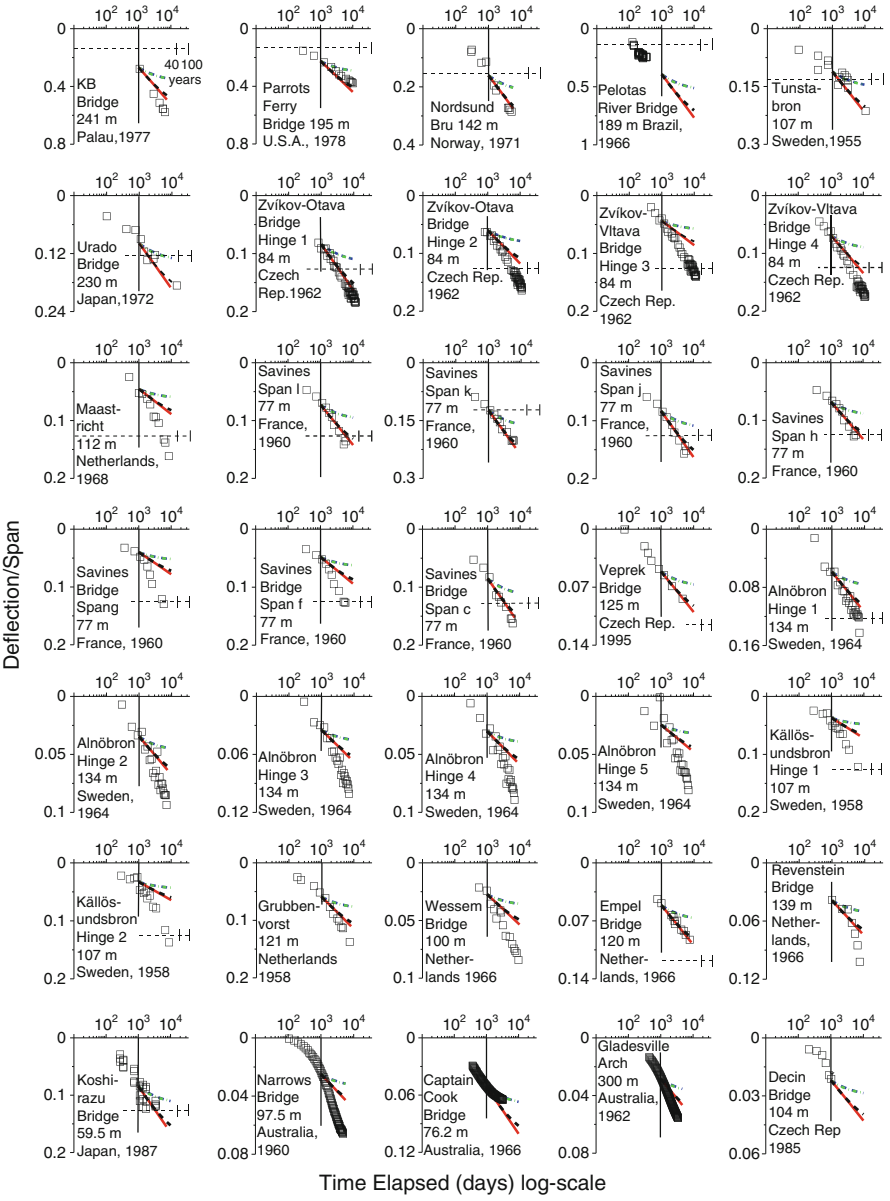


Fig. 10 Extrapolations of creep data for 36 bridges based on formula (Eq. 3) using the estimated average strength and composition of concrete (reprinted with permission from the American Concrete Institute, Bažant et al. 2011)

The coefficient of variation of $\overline{r}$ is $\omega_r = 0.45$ %, but only the average value of $\overline{r}$ can be considered to be realistic because the same mean properties had to be assumed for all the bridges.

Figure 11 compares the lines of the corrected extrapolations with the terminal series of deflection data points. Note that the extrapolation errors are significantly reduced and that the deviations from the measurements now lie nearly equally below and above the measured data point series. An improved model B3 is thus obtained. The other parameters of model B3 have no effect on the longtime bridge deflection slope and thus cannot be improved in this way.

It would hardly be possible to obtain such an improvement of longtime performance by model calibration with the laboratory database alone. Because the database is biased toward short creep durations (Bažant and Panula 1978), large changes in q_3 and q_4 cause only a very small change in the sum of squared deviation from the laboratory data. This causes high uncertainty in the q_3 and q_4 values obtained solely by minimizing the database errors.

Rate-Type Creep Formulation

Coupling Finite Element Analysis with Rate-Type Creep Formulation

A realistic detailed analysis of creep and shrinkage effects necessitates a rate-type formulation consisting of differential rather than integral equations. However, based on the prediction models of engineering societies, concrete creep is characterized by the compliance function $J(t,t')$, which represents the strain at time t induced by a unit sustained uniaxial stress applied at age t'. For time-dependent stress history $\sigma(t)$, the principle of superposition in time is applied to obtain from $J(t,t')$ an integral-type creep formulation (while the cracking model suffices to capture all the nonlinearities, which are due mainly to drying). Thus, the creep is characterized by a Volterra-type integral equation of aging viscoelasticity, having a kernel that, because of aging, is of non-convolution type.

For step-by-step finite element analysis, this integral equation is inadequate. The inadequacies are two: (1) a high computational cost due to memory-demanding history integrals and (2) incompatibility with variable environments as well as the influencing phenomena, such as cracking, causing deviations from linearity, or the principle of superposition. As a result, the integral-type approach is unusable or inaccurate for large-scale structural analysis. It is also inadequate physically, as it is incapable of capturing the concrete cracking, eventual bond slip, and prestressing steel relaxation which is a highly nonlinear phenomenon affected by concrete creep evolution. Inevitably, for realistic analysis of large creep-sensitive structures, the integral-type creep law must be converted to a rate-type formulation with internal variables whose current values account for the previous history of viscoelastic strain.

The key property enabling the rate-type analysis is the fact that any realistic integral-type stress–strain relation of aging viscoelasticity can be approximated with any desired accuracy by a rate-type creep law visualized by a rheological

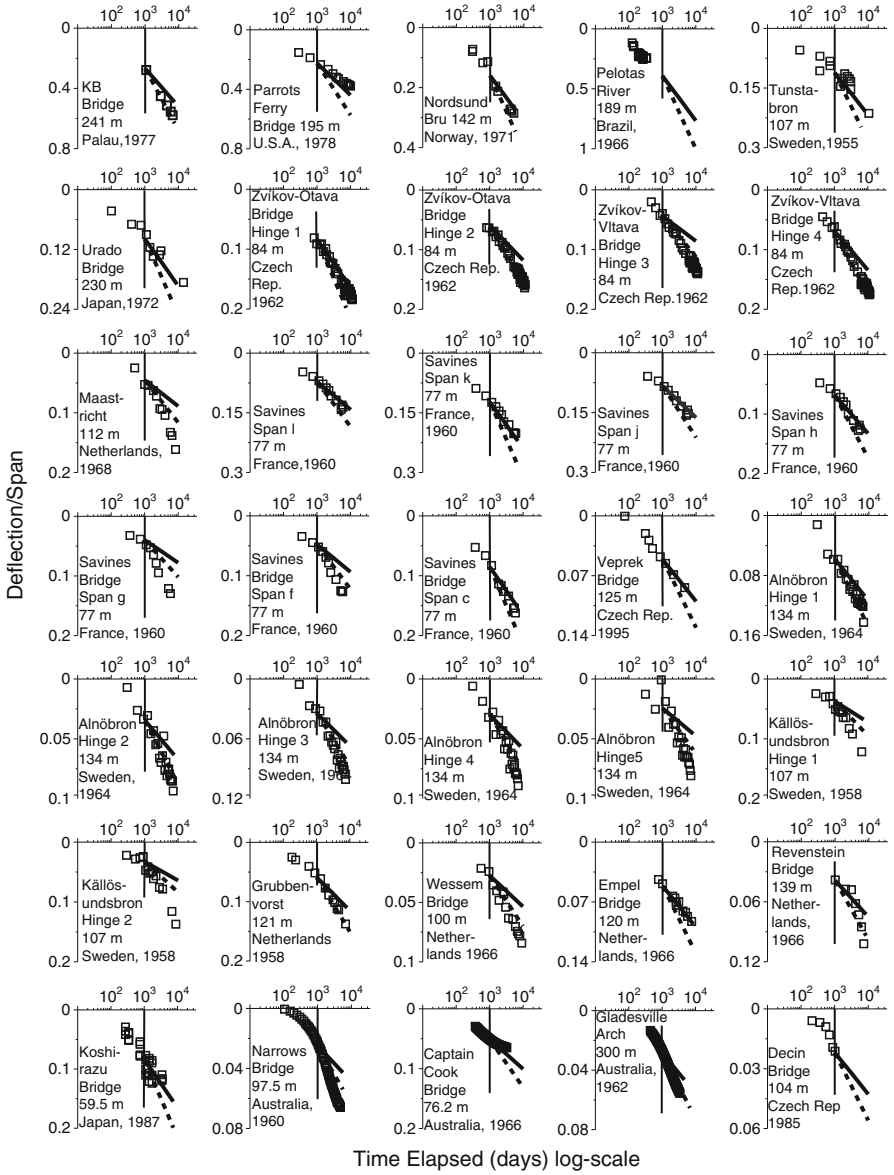


Fig. 11 The long-term extrapolations (*solid lines*) of the deflections of 36 bridges based on the original model B3 and on its update with long-term factor (*dashed lines*) (reprinted with permission from the American Concrete Institute, Bažant et al. 2011)

model, e.g., Kelvin chain model or Maxwell chain model. For creep, the Kelvin chain model is more convenient because its parameters can be related to creep experiments more directly. This model consists of a series of Kelvin units $\mu = 1, 2, 3, \ldots, N$ (Fig. 12), each of which involves a spring of stiffness $E_\mu(t)$ coupled in

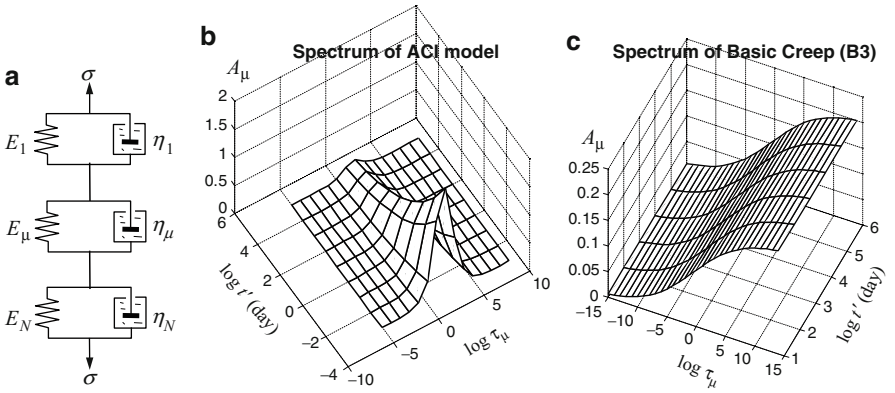


Fig. 12 (**a**) Kelvin chain model, (**b**) continuous spectra based on ACI model, and (**c**) continuous spectra based on B3 basic creep

parallel with a dashpot of viscosity $\eta_\mu(t) = E_\mu(t)\tau_\mu$ where τ_μ are the suitably selected retardation times. For the time step Δt, the internal variable of the Kevin unit is defined as $\gamma_\mu = \tau_\mu \mathrm{d}\varepsilon_\mu/\mathrm{dt}$. The plot of compliances $A(\tau_\mu) = 1/E_\mu$ versus $\log(\tau_\mu)$ is called the discrete retardation spectrum.

Based on the previous experience (Bažant and Prasannan 1989a, b, Bažant and Baweja 2000; Bažant et al. 2012b; Yu et al. 2012), the retardation times for concrete creep are best chosen to be spaced by decades in the logarithmic time scale – that is, $\tau_{\mu+1} = 10\tau_\mu$. For a sparser spacing of τ_μ the representation of the compliance function by the discrete spectrum becomes bumpy and inaccurate, while a denser spacing gives no significant gain in accuracy. For $\tau_\mu \ll \Delta t$ (time increment), the Kevin units will behave like elastic springs, and the dashpot has no effect, while for $\tau_\mu \gg \Delta t$, the Kevin units will perform like rigid connections.

When the rate-type creep law is used, the structural creep problem can be reduced to a system of first-order ordinary differential equations in time, with age-dependent coefficients. However, there is no need to solve this differential equation system in FEM simulation. It is more efficient to convert the incremental stress–strain relation for each time step Δt to a quasi-elastic incremental stress–strain relation. Thus, the structural creep problem gets reduced to a sequence of elasticity problems with initial strains (Bažant 1971, 1975, 1982; Bažant and Wu 1973; RILEM 1988; Jirásek and Bažant 2002).

Continuous Retardation Spectrum

Because of the aging, the spectrum is different for each subsequent time step, although it can be assumed to be constant within each time step that is not too long. Indeed, for the stress changes within a short enough time step, the concrete may be considered to behave as a nonaging, linearly viscoelastic material characterized by the Kelvin chain moduli and viscosities corresponding to age $t_{(n-1/2)}$ at

the middle of the time step (in the logarithmic scale). Therefore, one must identify a nonaging retardation spectrum for the corresponding compliance function (e.g., the ACI, CEB, or GL function) at the average age $t_{(n-1/2)}$ for every time step.

In the original rate-type creep analysis (Bažant and Wu 1973), the discrete spectrum $A(\tau_\mu)$ was determined on the basis of least-squares fitting of the compliance function. However, this approach was found to give nonunique results for $A(\tau_\mu)$ and to be oversensitive to small changes in the compliance. More seriously, the spectrum identified in this way violates, for some periods of time, the thermodynamic constraint, e.g., the moduli of Kelvin units did not increase monotonically as required by aging. Furthermore, implementing the least-squares fitting in each time step makes the FEM programming cumbersome.

As shown in Bažant et al. (2012a), this obstacle can be overcome by calculating the current continuous retardation spectrum of Kelvin chain. A continuous retardation spectrum corresponds to infinitely many Kelvin units with retardation times τ_μ of infinitely close spacing. Thus, it represents a smoothed-out plot of $1/E_\mu$ versus $\log(\tau_\mu)$. Its advantage is that it is unique and can be identified from the given compliance function analytically by means of Laplace transform inversion utilizing Widder's approximate inversion formula (Bažant and Xi 1995; Bažant et al. 2010, 2012a, b).

In Fig. 12, two continuous spectra, one for basic creep of model B3 and the other for ACI model, are demonstrated. The spectrum of ACI model is age dependent, with $A(\tau)$ decreasing with age. In the case of model B3, which is based on the solidification theory (Bažant and Prasannan 1988; Jirásek and Bažant 2002), the aging is taken into account by means of volume growth of the solidifying component and by a gradual increase with age of the flow term viscosity. Thus, it is possible to use a nonaging compliance function for the solidification component. Consequently, the spectrum for that component is nonaging (i.e., age independent), as shown in Fig. 12.

Numerical Procedure

After the Kelvin chain moduli for the current time step Δt and the current integration point are obtained by the continuous spectrum method, the exponential algorithm is implemented to incorporate the creep model into FEM programs. This algorithm, derived as the exact solution for stress varying linearly within the time step, is unconditionally stable (Jirásek and Bažant 2002; Bažant 1971, Bažant and Wu 1973).

In this algorithm, one first calculates the incremental modulus at the middle of the time step:

$$1/E''\left(t_{n-1/2}\right) = 1/E_0 + \sum_{\mu=1}^{N} D_\mu^{-1} = 1/E_0 + \sum_{\mu=1}^{N} A\left(\tau_\mu\right)\left(1-\lambda_\mu\right) \tag{23}$$

where E_0 is the instantaneous modulus, $\lambda_\mu = \tau_\mu\left(1-e^{-\Delta t/\tau_\mu}\right)/\Delta t$. Then the inelastic strain increment, also called the eigenstrain, is obtained as

$$\Delta\varepsilon'' = \sum_{\mu=1}^{N} \left(1 - e^{-\Delta t/\tau_\mu}\right)\gamma_\mu^{(n-1)} \tag{24}$$

where $\gamma_\mu^{(n-1)}$ are the internal variables at the last time step t_{n-1}. In the numerical implementation, the matrix of eigenstrains can now be augmented by the other inelastic strains (e.g., shrinkage and thermal strains), as well as the cracking or damage, and eventually also the cyclic creep eigenstrain. Note that the last three terms in the matrix $\Delta\varepsilon''$ represent the shear creep, which is important for the shear lag in box girders, although normally neglected in the integral-type approach. Based on the isotropic quasi-elastic stress–strain relation, the stress increment at the end of the current step is

$$\Delta\sigma = E''\left(t_{n-1/2}\right)D(\Delta\varepsilon - \Delta\varepsilon'') \tag{25}$$

where D is the elastic stiffness matrix with a unit value of Young's modulus and $\Delta\varepsilon$ is the total strain increment. Taking advantage of this stress–strain relation, the structural creep problem gets reduced to a sequence of incremental elasticity problems, which can be easily solved by FEM programs. After the stress increment is obtained, the internal variables are updated as follows:

$$\gamma_\mu^{(n)} = \lambda_\mu \Delta\sigma D_\mu^{-1} + e^{-\Delta t/\tau_\mu}\gamma_\mu^{(n-1)} \tag{26}$$

Unlike the integral-type approach, here the previous history need not be stored because it is fully characterized by the current values of γ_μ. In FEM, the number N of Kelvin units could actually be reduced to approximately 5, with the first one being computed as the integrated area under the spectrum up to $-\infty$ in the logarithmic time scale (the reason is that when $\tau_\mu \ll \Delta t$, the Kelvin units behave as elastic springs). However, the direct use of $N = 22$ for the B3 model and $N = 13$ for other models, though wasting computer time, made the programming simpler (which was more important).

Unlike the other purely empirical models, the B3 model is based on the solidification theory. In this model, the creep consists of two components: the basic creep and the drying creep. For drying creep, the implementation of its compliance function into FEM is similar to other models, whereas for the basic creep, it includes a nonaging volume constituent and a viscous flow component. Therefore, a simpler exponential algorithm exists for model B3. The spectrum that accounts for the nonaging constituent, which is used to calculate the Kelvin unit stiffness D_μ, is expressed as

$$A^b(\tau_\mu) = \left(\sqrt{1/t_{n-1/2}} + q_3/q_2\right)A(\tau_\mu) = \kappa A(\tau_\mu) \tag{27}$$

Consequently, the total inelastic creep strain is modified as

$$\Delta\varepsilon'' = \Delta\varepsilon_1'' + \Delta\varepsilon_2'' = \left[\sum_{\mu=1}^{N^b} \kappa\left(1 - e^{-\Delta t/\tau_\mu}\right)\gamma_\mu^{(n-1)} + \sum_{\mu=1}^{N^d}\left(1 - e^{-\Delta t/\tau_\mu}\right)\gamma_\mu^{(n-1)}\right] + \left[q_4\sigma^{(n-1)}\Delta t/t_{n-1/2}\right] \tag{28}$$

where N^b and N^d refer to the number of Kelvin units in basic creep and drying creep, respectively. Here only the key steps of numerical implementation are listed; for the detailed information and examples, consult Bažant et al. (2012a, b) and Yu et al. (2012).

Generalization for Cyclic Creep

As demonstrated in Bažant and Hubler (2014), the cyclic creep cannot cause appreciable deflections of concrete bridges, but can cause major inelastic tensile strain on the top face or the bottom face, or both, of the box girder, especially for shorter spans in which the live load is not a negligible portion of the total load. To limit tensile cracking, this effect of cyclic creep needs to be calculated.

Unlike metals and fine-grained ceramics, the microstructure of concrete is disordered on all the scales, from the nanoscale to the macroscale of an RVE, whose size is typically 0.1 m (assuming normal size aggregates). On all scales, the material is full of flaws and preexisting cracks. The growth of cracks larger than the RVE is blocked by reinforcing bars, and cracks much smaller than the RVE are not important for safety. Under fatigue loading, such cracks must be expected to grow but affect only inelastic deformation rather than safety since they are much smaller than the RVE. Experimental results for cyclic compressive loading of concrete within the service stress range (i.e., for stresses less than 40 % of the strength) indicate no degradation of material strength for subsequent short-time loading up to failure and at most only a slight decrease of concrete stiffness (Hellesland and Green 1971).

Within a single RVE of length l_c, the strain increment, $\Delta\varepsilon_N$, of a single penny-shaped microcrack of radius a_0 can be calculated using fracture mechanics (Bažant and Hubler 2014) in terms of the crack size increment after N cycles, Δa_N:

$$\Delta\epsilon_N = 3\gamma_0\frac{\sigma}{E}\left(\frac{a_0}{l_c}\right)^3\frac{\Delta a_N}{a_0} \tag{29}$$

where γ_0 is a nondimensional geometry factor. If the Paris law is applied locally at the RVE level, the increment in crack size may be expressed in terms of the remotely applied stress amplitude $\Delta\sigma$:

$$a_N - a_0 = \lambda\left(\frac{c\Delta\sigma\sqrt{a_0}}{K_c}\right)^m N \tag{30}$$

Note that the prefactor λ and the exponent m are empirical constants that must be obtained through test data fitting. The full relation between the cyclic creep strains per cycle as a function of applied stress is then given as

$$\epsilon_N = C_1 \sigma \left(\frac{\Delta\sigma}{f'_c}\right)^m N \quad \text{where} \quad C_1 = \frac{3\gamma_0}{E}\frac{\lambda}{a_0}\left(\frac{ca_0}{f'_c}\right)^3 \left(\frac{f'_c\sqrt{a_0}}{K_c}\right)^m \tag{31}$$

Here f'_c is the standard compressive strength of concrete. It is noteworthy that ε_N is predicted to depend on both σ and N linearly. This agrees with the available cyclic creep measurements within the service stress range and is convenient for structural analysis. Through dimensional analysis and similitude, it can be shown that this functional form applies to all compressive cyclic crack types including crushing band propagating transversely to compression (Suresh 1998; Suresh et al. 1989; Eliáš and Le 2012), wedge-splitting cracks (Bažant and Xiang 1997) at hard inclusions, parallel to compression, interface cracks at inclusions, and pore-opening cracks parallel to compression (Sammis and Ashby 1986; Kemeny and Cook 1987; Fairhurst and Coment 1981). Because cyclic stress is linearly related to cyclic strain, it is possible to define the cyclic creep compliance $\Delta J_N = \Delta\epsilon_N/\sigma$, i.e.,

$$\Delta J_N = C_1 \sigma \left(\frac{\Delta\sigma}{f'_c}\right)^m N \tag{32}$$

Within a limited time range, this compliance can be applied as an acceleration of basic creep for short-term test data:

$$\Delta t_N = C_t \left(\frac{\Delta\sigma}{f'_c}\right)^m N \quad \text{where} \quad C_t = C_1 \frac{\Delta(\ln J_N)}{J_N} \tag{33}$$

In this case, the factor C_t and exponent m can be obtained by fitting of existing tests. Values of $C_t = 46 \times 10^{-6}$ and $m = 4$ were obtained using data by Whaley and Neville (1973), Kern and Mehmel (1962), and Hirst and Neville (1977).

However, interpreting the cyclic creep as time acceleration gives gross underestimation of cyclic creep strain for multi-decade duration. Correctly, the cyclic creep strain should be considered as additive to other strains.

To estimate the magnitude of cyclic creep effects in structures, the responses of cross sections at the pier of a series of box girder bridges were considered (Bažant and Hubler 2014). The theory of bending may be used to calculate the required residual strain and residual stress ratios at any level in the cross section produced by the cyclic creep, as illustrated in Fig. 13. To evaluate the severity of effects, two dimensionless measures, namely, the inelastic residual strain ratio and the inelastic residual curvature ratio produced by load cycling, have been introduced:

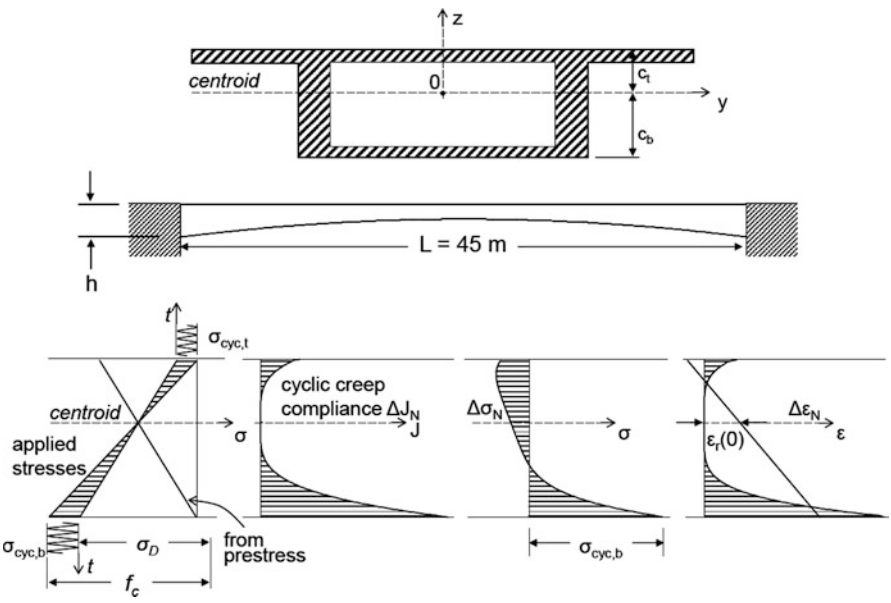


Fig. 13 An example of one of the scaled bridge cross sections at the pier for which the prestressing eccentricity and bridge span have been designed to reach allowable limits. Section profiles of the applied stresses due to dead load, cyclic live load, and prestress, of the computed creep compliance, of the corresponding (free) cyclic creep strain, and of the resulting residual inelastic strain (or eigenstrain) required by cross-sectional planarity (Reprinted from Bažant and Hubler 2014 with permission from Elsevier)

$$\rho = \frac{\varepsilon_\tau^{\max}}{f'_t/E}, \quad r = \frac{\kappa_\tau}{\kappa_{el}} \tag{34}$$

Here $\epsilon_r^{\max} = \max[\epsilon_r(c_b), \epsilon_r(c_t)] =$ maximum strain produced in the pier cross section by stress cycling (occurring at either top or bottom at pier), and κ_{el} $\kappa =$ minimum possible elastic curvature in the pier section.

To appraise the cyclic creep effects in large prestressed segmental box girders, a collection of designs of six large-span bridges has been obtained through the courtesy of Y. Watanabe, Shimizu Construction Co., Tokyo (Japanese bridges Koshirazu, Tsukiyono, Konaru, and Urado); of I. Robert-son, University of Hawaii, Manoa (North Halawa Valley Viaduct); and of G. Klein, WJE, Highland Park, Illinois (the ill-fated world-record bridge in Palau. The data on these bridges are used in Figs. 14 and 15, and for details of these bridges, see Bažant et al. 2012a, b).

For the sake of easier comparability, it is assumed that, in all these bridges, the load and dead load moment values were the maximum allowable for each cross section. Based on the dimensions of each bridge, the live and dead loads

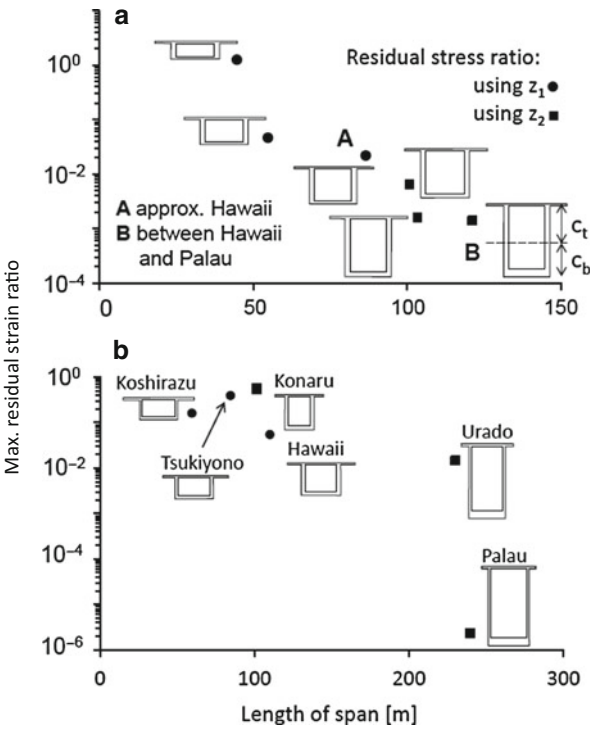


Fig. 14 Maximum residual strain ratios (Eq. 38) calculated for (**a**) a series of scaled cross-sectional design to allowable limits and (**b**) a selection of actual bridge cross sections (Reprinted from Bažant and Hubler 2014 with permission from Elsevier)

(including the self-weight) are determined, and the bending moments at the pier are calculated assuming the girder to be rather flexible at midspan, in which case the moment at the pier is almost as large as it would be for a hinge at midspan. The results for the residual strain ratio ρ at $N = 2$ million are plotted in Fig. 14 (top).

The results for these bridges are rather scattered, which is obviously explained by variability in the cross-sectional shapes. Therefore, another set of scaled bridge cross sections is generated by interpolating and extrapolating trend between the cross sections of the Hawaii bridge (span 110 m) and of the K–B Bridge in Palau (span 241 m). A similar calculation procedure leads to the diagrams for ρ in Fig. 14b.

Much stronger effects of cyclic creep are seen in the plots of the maximum residual strain ratio ρ in Fig. 15. They reveal that, for spans between about 80 and 200 m, the cyclic creep can produce significant tensile strains. These strains can approach or even exceed the maximum elastic strain at the top or bottom faces of the box girder. This strain superposed on strains from differential shrinkage, drying creep and temperature, obviously aggravate distributed tensile cracking, which may in turn lead to corrosion of reinforcement.

Note that whereas the static longtime creep deflections grow in time logarithmically (Bažant et al. 2011, 2012a), the cyclic creep deflections grow linearly

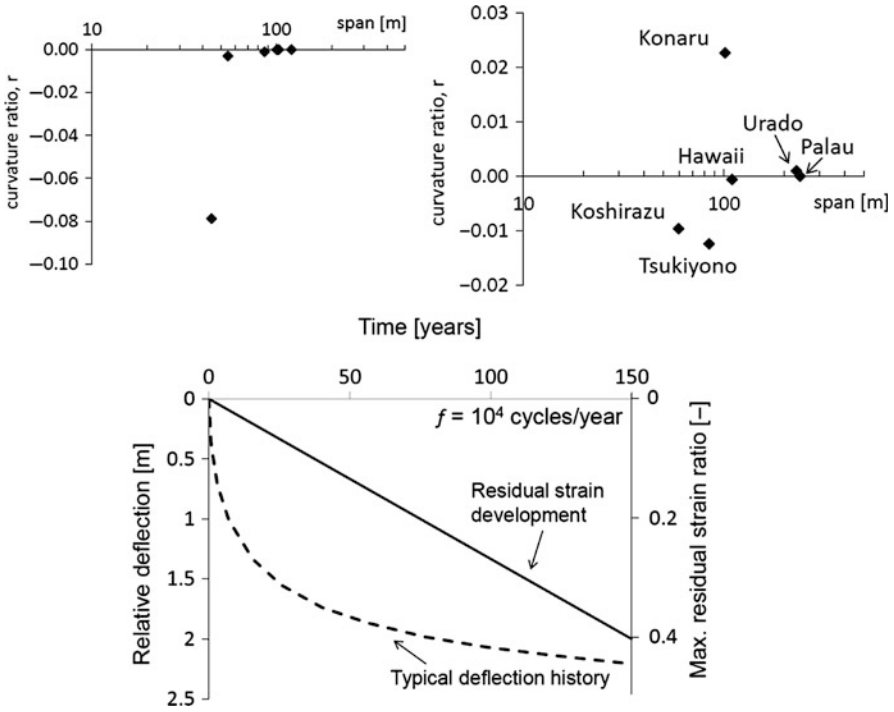


Fig. 15 (*Top*) The calculated curvature ratios for the actual bridge cross sections. The *left* figure shows systematically varying cross sections that reveal a transition from negative to positive curvature after a span of 80 m is reached. The *right* figures illustrate the values that the actual bridge cross sections may reach. (*Bottom*) Schematic comparison of the evolution of creep deflection and of the residual cyclic strain (Reprinted from Bažant and Hubler 2014 with permission from Elsevier)

(provided that the traffic load frequency and amplitude remain constant). This property is verified by experiments and, theoretically, is a consequence of Paris law (Eq. 18), which states that the crack extension is proportional to the number of cycles, N. Consequently, even if the cyclic creep effects are insignificant within the first 20 years of service, they may become significant, compared to creep, for a 100-year lifetime; see Fig. 15 (bottom).

Prestress Relaxation at Variable Strain

In addition to concrete creep and shrinkage, the relaxation of prestressing steel can make a major contribution to the long-term deflection of prestressed box girders. Currently, the practice has been to calculate the stress relaxation from simple empirical formulas that estimate the steel relaxation at constant initial strain ε_0 and constant temperature T_0 (normally the room temperature). However, this is

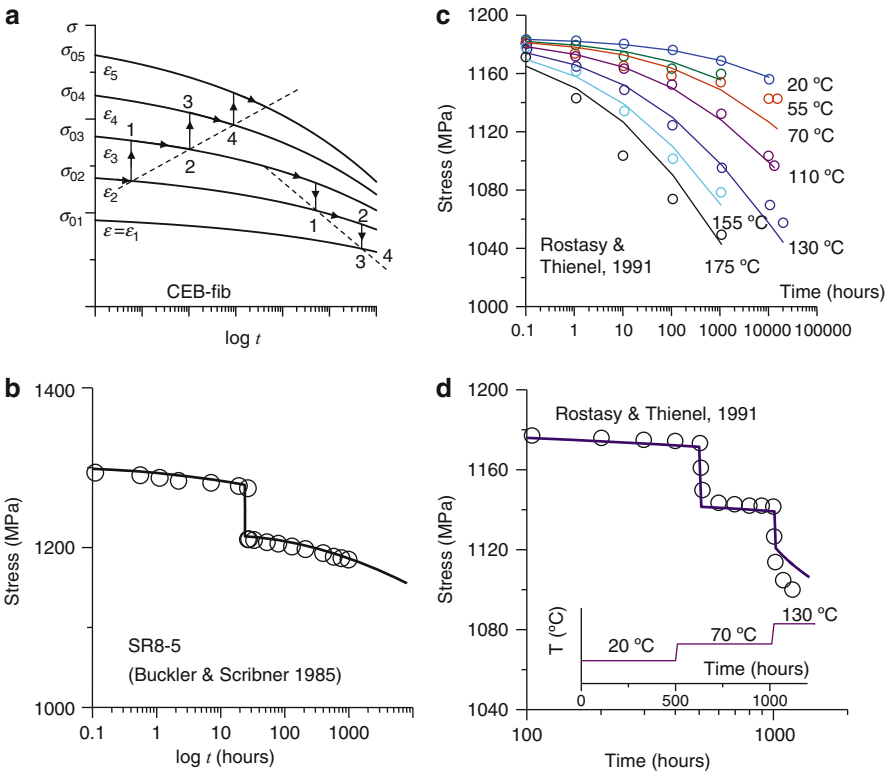


Fig. 16 (**a**) Stress relaxation at various constant levels of strain, with decomposition of stress increment for relaxation during time interval dt at variable strain $\boldsymbol{\varepsilon}(t)$; prediction compared with (**b**) tests of 5 % load drop; (**c**) tests under different temperatures; (**d**) tests under stepwise heating (With permission from ASCE, Bažant and Yu 2013)

valid only when the strain variation in the steel bonded to concrete is negligible, which is not the case for creep-sensitive structures. Also, exposure of the pavement to sun can cause significant heating of tendons embedded in the top slab, especially in tropical areas. Therefore, one needs a general uniaxial viscoplastic constitutive law for prestressing steel.

Based on a Bingham-type viscoplastic model (Jirásek and Bažant 2002), the simplest form of the constitutive equation is to treat the strain increment as the sum of the instantaneous strain increment, the viscoplastic strain increment, and the thermal strain increment due to temperature change. The initial strain ε_0 introduced at the time of prestressing is an initial condition for the differential equation of stress relaxation but does not belong into this differential equation itself. To eliminate ε_0, a realistic hypothesis to be checked by experiment must be made.

For a variable strain history, the stress-relaxation increment $\Delta\sigma$ occurring at strain ε during time interval dt (line 12 in Fig. 16) may be calculated as if the current strain were maintained constant from the beginning. In other words, the hypothesis

means that a small increment ($\Delta\varepsilon$, $\Delta\sigma$) [line 13 in Fig. 16] may be decomposed into stress-relaxation increment $\Delta\sigma$ [line 12 in Fig. 16, imagined to occur at constant strain ε], followed by an instantaneous jump up or down [line 23 in Fig. 16] from the relaxation curve for constant ε [line 12 in Fig. 16] to the relaxation curve for constant $\varepsilon + \Delta\varepsilon$ [line 34 in Fig. 16].

Using this hypothesis, the current relaxation formula used in the CEB-FIP code can be generalized to a rate-type formula to take into account strain variation (Bažant and Yu 2013). The same generalization can be applied to the formula used in the American practice, although this formula cannot be used in general since even at constant strain it is not realistic for short-term relaxation.

On the other hand, the rate-type formula based on the CEB-FIP Model Code, which is realistic for the short-time range at constant strain, has two other limitations: (1) it misses a relaxation threshold that is used in the formula of American practice, and (2) for the same steel, the relaxation curves for different constant strains can cross each other after a long enough period. To remedy these problems, an improvement of the CEB-FIP formula for constant strain has been proposed (Bažant and Yu 2013):

$$\sigma = \min\left(\gamma f'_y, \sigma_0\right) + f'_y\left\langle \sigma_0/f'_y - \gamma \right\rangle \left[1 + \left(\rho t^k\right)/\left(c\lambda^k\right)\right]^{-c} \tag{35}$$

Based on the foregoing hypothesis, and an assumption of viscoplastic behavior, the generalization to variable strain is

$$\dot{\varepsilon} = \dot{\sigma}/E_t + \frac{\left\langle F(\varepsilon) - \gamma f'_c \right\rangle}{E_t} \frac{k\rho^{1/k} c^{1-1/k}}{\lambda \zeta^{1+1/c}} \left(\xi^{1/c} - 1\right)^{1-1/k} \tag{36}$$

where $\xi = [F(\varepsilon) - \gamma f'_y]/[\sigma - \gamma f'_y]$, γ indicates the relaxation threshold, and $\rho = \rho_0 e^{h\xi}$; k, c, ρ_0, h = positive constants for a given steel. Usually, h is a very small value and can be set as 0, thus ensuring that the relaxation curves for different σ_0 would never cross each other. Figure 16b shows the comparisons between the simulation results based on Eq. 36 and the relaxation tests by Buckler and Scribner (1985), in which a sudden load drop (strain change) was introduced. It can be seen that Eq. 36 agrees with the tests satisfactorily.

To take into account an elevated temperature, which accelerates the steel relaxation, the real time may be replaced with the effective time based on the activation-energy theory. The effective time is defined as $dt = A_T d\tau$, where τ is the real time and $A_T = \exp(Q_0/RT_0 - Q_0/RT)$ is the Arrhenius factor (Cottrell 1964), which is equal to 1 at room temperature $T_0 = 20$ °C; here k_B = Boltzmann's constant and Q_0 = activation energy. By doing this, Eq. 36 is extended to time-variable temperature. Furthermore, the thermal strain rate must be added on the right side of Eq. 36.

Figure 16c shows the simulation results compared to Rostasy and Thienel tests (1991). Note that the relaxation parameters are determined by fitting first the data for 20 °C and then predicting the data for higher temperatures. In addition to the

case of constant temperature, the proposed relaxation formula is also compared with the tests of relaxation under stepwise heating (Rostasy and Thienel 1991). The temperature history used in tests is reproduced in Fig. 16d, and the comparison of the prediction with the tests appears to be acceptable.

Comparison with Existing Commercial Software for Creep Effects in Bridges

The existing programs for creep design and creep structural analysis were mostly developed around 1980. The material creep and shrinkage model, typically the CEB model, is embedded in these programs. Neither the nonlinear effects of cracking and cyclic creep nor the effects of drying and diffusion are usually considered. Neither are the environmental variation, heating of tendons, and relaxation under varying strain.

For one-dimensional analysis, these programs, exemplified by SOFiSTik (distributed by the Ingenieur-Software Dlubal GmbH, Germany), usually model the box girder by beam elements, and the creep is represented by linear aging viscoelasticity in the primitive form of an integral-type approach. To avoid the demands on computer time and storage resulting from the memory integrals in the case for 2D and 3D analysis, the software uses a simple quasi-elastic algebraic analysis based on a one-step incremental elastic relation proposed by Trost (1967). This relation uses Trost's coefficient ρ, called the relaxation coefficient, to calculate the incremental Young's modulus for the entire period. Trost's empirical method, however, neglects the major effect of creep aging, an error that was eliminated in 1972 by the development of the age-adjusted effective modulus (AAEM) method (Bažant 1972, 1975; RILEM 1988, endorsed by ACI-209 as well as CEB-FIP). However, even if the programs like SOFiSTiK switched to AAEM, much larger errors due to the use of quasi-elastic analysis for multidimensional finite elements and of pure linear viscoelasticity for one-dimensional beam elements would still remain.

A good example showing the inadequacy of these programs is their application to the ill-fated K–B Bridge. By applying the present rate-type algorithm to 3D analysis, the long-term deflections of the K–B Bridge have been computed for different creep and shrinkage models, as plotted in Fig. 17. Although it is clear from the plot that the CEB creep model performs poorly in predicting the long-term deflections, one must accept this model here to isolate the errors due to numerical algorithm.

Unlike the rate-type analysis using 3D elements, the K–B Bridge deflections are now recalculated by SOFiSTiK using a simplified approach with 45 nodes and 44 beam elements corresponding to the segmental construction; see Fig. 17. In addition to concrete creep, the CEB model is also used for the prestressing steel relaxation (under the assumption of constant strain). The evolution of steel relaxation is handled in SOFiSTiK either by a fixed 1,000-hour relaxation factor or, as used in this example, by the stress-dependent quadratic function prescribed by CEB model (1990). The temperature effect is ignored.

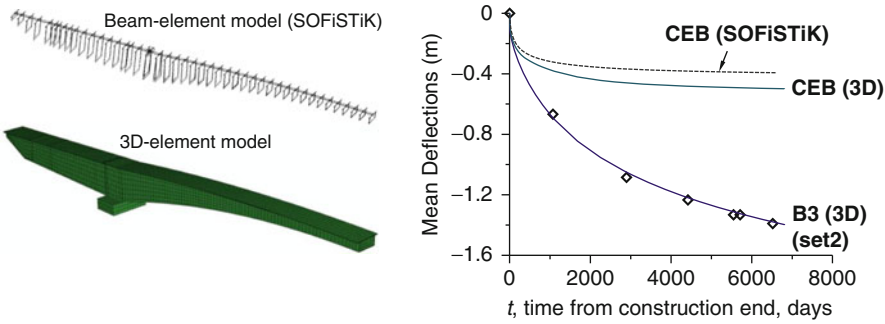


Fig. 17 The comparison of 3D model and simplified model by SOFiSTiK (Yu et al. 2012)

The linear viscoelastic analysis of the K–B Bridge based on SOFiSTiK predicts far smaller deflections than the recorded measurements, as can be seen in Fig. 17. Furthermore, when compared with the present rate-type 3D analysis, which uses the same CEB model, SOFiSTiK still substantially underestimates the creep deflection. The integral-type algorithm used in this calculation gives a deflection that represents 79 % of the deflection predicted by the rate-type 3D analysis, which itself underestimates the measured creep deflection by a factor of approximately 1/3.

Furthermore, the commercial programs using beam elements with a simplified quasi-elastic analysis similar to SOFiSTik cannot realistically simulate the nonlinear evolution of viscoplastic steel relaxation at variable strain, the effects of variations of temperature and humidity, the nonlinear cyclic creep effects, and the nonlinear effects of cracking and cyclic loading. Neither are they able to capture the effect of differential drying in slabs and webs of different thicknesses and environmental exposures. Therefore, the creep structural analyses for large creep-sensitive structures necessitate rate-type 3D analysis, as described in the preceding sections.

Development of the B4 Model

Expanded Database of Laboratory Tests

The new collection of 69 bridge spans from around the world, presented in section "Wake-up Call: Excessive Long-Term Bridge Deflections," confirms the linear behavior in logarithmic time observed in the K–B Bridge in Palau (Table 3). The slope of the deflection curve in the range of the typical structural design life is consistently underestimated in each of these spans. This result provides further evidence that a new creep and shrinkage prediction equation is needed. A preliminary update of the long-term slope of model B3 proves the feasibility of using the shape of deflection records to improve the design equation. The new model B4 makes use of this concept. First, a collection of up-to-date short-term laboratory tests is developed to allow for the calibration of such a model.

The first large worldwide database of creep and shrinkage tests was published at Northwestern University (NU) in 1978. It was followed by several improvements, the last one in 2008. A major expansion, completely restructured and verified, named the NU Database, is presented in Hubler et al. (2014a). It represents by far the most comprehensive collection of creep and shrinkage curves that has ever been published. It covers longer measurement periods and high-strength concretes and encompasses the effects of admixtures in modern concrete mixes. It contains roughly 1,400 creep and 1,800 shrinkage curves, of which approximately 800 creep and 1,050 shrinkage curves contain admixtures. In addition to mix proportions, testing conditions, and specimen geometries, the admixture contents and aggregate types are also indicated. The data are categorized by type – autogenous shrinkage, drying shrinkage, basic creep, drying creep, and total creep. Based on extensive data analyses, various data corrections and curve classifications are proposed.

The new database makes it possible to calibrate and verify improved creep and shrinkage prediction models. Additionally, the statistics of the mix parameters, strength distributions, and scatter of the compliance curves have been extracted for applications in reliability engineering and probabilistic performance assessment. Data analysis brings to light various recommendations and suggests corrections of various oversights to avoid or minimize in the future the testing of concrete creep and shrinkage and improve the reporting of test data. This should make future test data more useful, consistent, complete, and reliable. The NU database is now available for free download at http://www.civil.northwestern.edu/people/bazant/ and http://www.baunat.boku.ac.at/creep.html.

Optimization Using Joint Laboratory and Structural Data

The actual calibration of prediction models ultimately can be reduced to a standard optimization problem if appropriate metrics to counteract the heteroscedastic nature of the data and adequately represent the quality of fit are used. Further complications are introduced by the combination of different data sources (represented here by the shrinkage strains ε, creep compliances J, and bridge deflections δ) (Bažant et al. 2014; Hubler 2014b, Wendner et al. 2014a, b). The data range through different orders of magnitude, exhibit inconsistent sensitivities to environmental and intrinsic material properties, and are associated with varying levels of inherent uncertainty.

On top of that, every data collection shows a certain bias toward specific testing conditions or material compositions, depending on the experimenters' preferences, the ease of sample generation and testing, or simply the preferences given by the engineering design viewpoints. Structural measurements are biased depending on their type and field of application and may lead to conclusions that do not apply to other applications. The bridge deflection data available for this investigation are generally biased toward compositions used in bridge engineering and comparably humid and hot environments that differ significantly from concretes applied in modern high-rise buildings and that are located within a climate-controlled envelope.

One way to deal with this database inherent bias is a multistage optimization strategy where only certain parameters, e.g., those determining the long-term behavior in case of bridge deflection data, are recalibrated. More information on updating strategies based on the Bayesian theorem (Bayes 1964) is given in the following. Although subjective and biased data are generally not advised, a bias toward certain compositions or environmental conditions might actually be intended in the sense of an importance weight. If a perfect model is out of reach, a model that is better representing conditions of practical relevance can be considered preferable over an objective model that introduces the same amount of error in all cases.

One way to deal with heteroscedastic data as well as the inherent bias in the input information is weighting. Although, in a strict statistical sense, weighting introduces bias, it represents a way to counteract the unwanted preexisting biases, e.g., toward short-term data. Li introduces the concept of hyper-box weighting (Bažant and Li 2008) for the calibration of shrinkage and creep models, respectively. The term hyper-box refers to the n-dimensional nature of the optimization problem in which all parameters (time, composition, geometry, environmental conditions, etc.) should be uniformly distributed in order to ensure an unbiased evaluation. The individual weights w_{ij} of curve i and half-decade j counteract any bias stemming from the data point density or preference toward certain data ranges, particularly short ones. The weights W_i (a) remove any bias due to material composition, or preferred testing conditions, and (b) allow the introduction of importance weight for either experimenters that are known to be particularly thorough or types of tests that are particularly relevant for the model development and its future application. Additionally, weights may be assigned due to the smoothness and consistency of the dataset, the quality of the data reporting, and the data completeness (e.g., whether a full curve from the onset of drying creep to its completion is shown). As discussed in Hubler et al. (2014a), a shrinkage strain or creep compliance curve that was re-digitized from a paper where the curve was reported in linear scale may introduce unacceptably high short-term errors of up to 100 %.

Suitable statistics to describe the quality of fit for data of heteroscedastic nature, and thus the basis for the formulation of an objective function, is the coefficient of variation, ω, of the root mean square error, RMSE (Wendner et al. 2014a):

$$\omega = \mathrm{RMSE}/\bar{y} \tag{37}$$

Through the normalization by $\bar{y}$, the dimensionality of the data is removed, and all data are brought to the same order of magnitude. The total contribution of a single data source to the objective function thus is

$$\omega = \sqrt{\sum_{i=1}^{N_y} W_i \sum_{j=1}^{m} w_{ij} \sum_{k=1}^{n} \left(\frac{y_{ijk} - \hat{y}_{ijk}}{\bar{y}} \right)^2} \tag{38}$$

where y_{ijk} = the kth measured value of half-decade j and curve i, $\hat{y}_{ijk}$ = the predicted value of point ijk, and $\bar{y}_i$ = the mean of the measurement data of curve i. It is

important that the total sum of weights of all N_y curves with N points be normalized to 1, as follows:

$$\sum_{i=1}^{N_y} W_i \sum_{j=1}^{m} \sum_{k=1}^{n} w_{ij} = 1 = \sum_{k=1}^{N} w_k \tag{39}$$

For convenience, the weights W_i and w_{ij} may be transformed to independent weights w_k assigned directly to all N data points of any given data source. The optimization problem finally reads

$$\hat{\mathbf{X}} = \begin{cases} \min\left(W_\varepsilon \omega_\varepsilon^2\right) \\ \min\left(W_J \omega_J^2 + W_\delta \omega_\delta^2\right) \end{cases} \tag{40}$$

with $W_\varepsilon = 1 =$ total weight of the shrinkage laboratory database, $W_J =$ total weight of the creep laboratory database, and $W_\delta = 1 - W_J =$ total weight of the bridge deflection database. During optimization of the creep model, the total contributions of the laboratory database and the collection of bridge deflections were weighted 2:1. That means that with N_J laboratory datasets and N_δ bridge deflection records,

$$W_J \sum_{k=1}^{N_J} w_k = 2W_\delta \sum_{k=1}^{N_\delta} w_k = 2/3 \tag{41}$$

The use of bridge deflection records jointly with laboratory compliance data unfortunately is not straightforward. Either the deflections $\delta(t)$ must be converted to compliance $J(t)$ or the predicted compliance function transformed to an evolution of deflections. Only then, residuals can be calculated, and a consistent contribution to the objective function can be formulated. For an accurate conversion, a complicated inverse analysis using 3D finite element simulations is required.

However, if the deflection $\delta(t)$ is known at a certain long enough reference time $t = t_{\text{ref}}$, then the deflection trend can be simply extrapolated to long times by assuming similarity to the compliance function. To apply this extrapolation, the age differences among the box girder segments must be ignored, and the age of the concrete must be characterized by one common effective (or average) age t_{close} at the span closing. Further necessary simplifications include the definition of one common effective age t_a at which the self-weight bending moments are introduced in the erected cantilever, instead of considering the gradual increase of the bending moment during erection. For the purpose of this investigation, the values $t_{\text{close}} = 120$ days and $t_a = 60$ days were used. This approximation yields acceptable results only for sufficiently large reference times, $t_{\text{ref}} \approx 1{,}000$ days, that ensure that the complex initial behavior due to the effects of drying, the construction sequence, and differences in age has almost died out. The extrapolation formula of the deflection from t_{ref} thus reads

$$\frac{\Delta J(t, t_a)}{\Delta J(t_{\text{ref}}, t_a)} \approx \frac{J(t, t_a) - J(t_{\text{close}}, t_a)}{J(t_{\text{ref}}, t_a) - J(t_{\text{close}}, t_a)} \approx \frac{\delta(t)}{\delta(t_{\text{ref}})} \tag{42}$$

This formula was verified using the finite element solution for the deflection of the K–B Bridge.

For statistical comparisons between models, the total unbiased coefficient of variation of the root mean square error of N data points should be correctly written as

$$\text{C.o.V.} = \frac{1}{\sum_{k=1}^{N} w_k y_k} \sqrt{\frac{\sum_{k=1}^{N} w_k}{\left(\sum_{k=1}^{N} w_k\right)^2 - \sum_{k=1}^{N} w_k^2} \sum_{k=1}^{N} w_k (y_k - \hat{y}_k)^2} \tag{43}$$

Although in a strict sense not suitable as indicator for the quality of fit of nonlinear models, the coefficient of correlation (or determination) is still widely used and, thus, introduced for comparison:

$$R^2 = 1 - \frac{\sum_{k=1}^{N} w_k (y_k - \hat{y}_k)^2}{\sum_{k=1}^{N} w_k (y_k - \overline{y}_k)^2} \tag{44}$$

Bayesian Method

The creep and shrinkage are affected by a multitude of influence factors, which include the effects of different mix proportions, cement types, admixtures, aggregate types, environmental conditions, curing and loading times, and geometries. These effects can be studied only under carefully controlled laboratory conditions. The extensive new laboratory databases for creep and shrinkage provide this type of information and ensure a well-calibrated prediction model. Nonetheless, a series of problems remain: the adaptation of a general prediction to a specific concrete mix; the updating of predictions based on, e.g., structural, observations; and the calibration of multi-decade prediction models in the absence of sufficiently long experimental test data. The framework of Bayesian updating provides an answer to these. Assume that a state S occurs with a probability $P(S)$ before any observation M is taken. The probability $P(S)$ is called the prior probability in this context. The posterior probability $P(S|M)$ given an observation can be calculated according to Bayes' rule (Bayes 1964) by:

$$P(S|M) = \frac{P(M|S)P(S)}{P(M)} \tag{45}$$

The probability of observing M given a certain state S is often referred to as the "likelihood"; $P(M)$ is also called "model evidence" or "marginal likelihood."

The concept of Bayesian updating can be applied to the problem of calibrating a multi-decade prediction model for concrete creep in the absence of long-term

laboratory data using bridge deflection data δ as evidence. Recall the functional form of B4 (or B3): the creep compliance function is given by the sum of instantaneous compliance (prefactor q_1), the aging viscoelastic compliance (prefactor q_2), the nonaging viscoelastic compliance (prefactor q_3), the aging flow compliance (prefactor q_4), and the drying creep compliance (prefactor q_5).

As a result of the laboratory database fitting and uncertainty quantification, the prior distributions of q_1 to q_5 can be obtained. Since the long-term behavior and in particular the terminal slope are governed by q_3 and q_4, a simplified procedure of updating solely these two parameters is suggested. In this case, the informed posterior distribution is $f_Q(q_3, q_4|U = \delta)$. The prior $f_Q(q_3, q_4)$ is taken from the uncertainty quantification.

Uncertainty Quantification in Creep and Shrinkage Prediction and Calculation of Confidence Limits

A proper understanding of the uncertainties involved in modeling creep and shrinkage is quintessential for safe and sustainable construction. The gap between the duration of available laboratory data ($<<10$ years) and typical design service lives of at least 50 years and often more than 100 years further aggravates the situation. As a first step, stochastic models for all required input parameters as well as correlation fields are needed in order to properly calibrate partial safety factors for creep and shrinkage as well as provide the basis for a realistic full probabilistic reliability assessment. Stochastic models for environmental conditions, variability in the structural system, and cross-sectional geometry can add the missing information to provide sensitivity factors between model inputs and prediction of creep and shrinkage and thus serve for the determination of reliability profiles at various times.

The broad scope of composition, mechanical, and environmental parameters associated with each creep and shrinkage test in the database allows for a full fit of stochastic models for each. However, the parameters of highest interest are those incorporated in the existing prediction models: the water/cement ratio (w/c), cement content (c), aggregate-to-cement ratio (a/c), 28-day Young's modulus (Eq. 28), and 28-day compressive strength (f_{28}). This parameter set has the most consistent effect on the measured response of the concrete and also represents the most consistently reported values. Gaussian stochastic models were developed for each of these parameters in a subset of all shrinkage tests consisting of midrange 28-day compressive strengths from 30 to 60 MPa. The results are tabulated in Table 2.

To develop a full stochastic framework for creep and shrinkage models of the experimental test data, additional stochastic models are needed for all the input parameters including the test setup parameters such as test duration and age at loading. These values have specific recommendations in the testing guidelines and strong biases toward times that allow for workable consistencies of the mix. As a result, a nonsymmetric distribution such as the lognormal may better capture such input parameters.

Table 3 lists the linear Pearson correlation coefficients between the composition and 28-day measured strength and modulus. While the sign and relative magnitude

Table 2 Stochastic model

Variable	PDF	Mean	C.o.V.	Datasets
Cement content *c*[kg]	LN	377	0.39	1,019
Water/cement ratio *w*/*c*	LN	0.486	0.35	1,101
Aggregate/cement ratio *a*/*c*	LN	5.44	0.43	1,056
Compressive strength *fc* [MPa]	LN	27.4	0.04	617
Young's modulus *cs*[MPa]	LN	28,773	0.12	622
Relative humidity h [−]	Norm	0.75	0.05	
Temperature T [°C]	Norm	20	0.05	

Table 3 Correlation matrix (Hubler et al. 2014a)

	w/*c*	*a*/*c*	*c*	f_{28}	E_{28}
w/*c*	1	0.83	−0.73	−0.19	0.34
a/*c*	–	1	−0.91	−0.13	0.38
c	–	–	1	0.30	−0.41
f_{28}	–	–	–	1	0.06
E_{28}	–	–	–	–	1

of the composition correlations may be estimated from typical requirements for hydration and workability, these values reflect a set of representative distributions across a full spectrum of recent and older cements that may be used to capture standard relationships for design. As is typically assumed, the correlations also reflect a prominent correlation between the water/cement ratio, the resulting strength, and Young's modulus.

The environmental conditions, temperature T, and relative humidity h depend on the geographical zone but also the microclimate. For the purpose of this investigation, a fictitious construction in Central Europe is assumed. The respective stochastic models are listed in Table 2. The variability of the cross section is considered with a Gaussian distribution of 5 % coefficient of variation. The derived stochastic models may be used to determine the scatter in shrinkage strain and creep compliance predictions for concrete members.

Conclusion

The preceding exposition in this chapter attempts to provide a modern overview of the problems of creep and shrinkage which can cause various types of damage to concrete structures, including excessive deflections, excessive cracking, and long-term deterioration. The magnitude and severity of these problems have for a long time been underestimated but recently came to the forefront of attention, partly because of systematic revelation of various long-term damages such as excessive deflections and partly because of the recent nationwide and worldwide emphasis on sustainability of concrete infrastructure.

To mitigate these damages, the design procedure will have to be improved, and this will require modernization of the design codes of engineering societies. Any changes in concrete design codes that have a theoretical underpinning have been notoriously difficult to implement. For example, the ACI-209 design recommendation for creep and shrinkage still features an obsolete 1971 model, whose oversimplifications have been repeatedly pointed out since 1973. One problem is that the code-making committees are dominated by practitioners, many of whom do not seem to worry about structural problems that will not occur during the first 20 years of lifetime.

Another, more serious, problem has been that the information from damages that were legally litigated in courts often ends up sealed in perpetuity. This causes an enormous break on progress, since structural engineering advances occur mainly as a result of clear identification of the damages in structures and their causes.

What is striking is the contrast with commercial aviation where the release of information on any failure, or potential damage that could lead to failure, is required by law as well as international treaties. Attempts to introduce a similar law for structural engineering, or at least to incorporate an article labeling the concealment of such data as unethical into the professional ethical codes of engineering societies, have been met with vehement opposition from engineering consultants and lawyers of engineering firms. Until this opposition is overcome, progress will be difficult.

References

AASHTO, *AASHTO LRFD Bridge Design Specification* (American Association of State Highway and Transportation Officials (AASHTO), Washington, DC, 2004)

ABAM Engineers Inc, *Koror-Babeldaob Bridge Repairs: Basis for Design*. Report submitted to Bureau of Public Works, Republic of Palau, Koror, October 1993)

ACI Committee 209, Prediction of creep, shrinkage and temperature effects in concrete structures, in *Designing for Effects of Creep, Shrinkage and Temperature in Concrete Structures*, *ACI-SP27* (American Concrete Institute, Detroit, 1971), pp. 51–93, reapproved 2008

ACI Committee 209, *Guide for Modeling and Calculating Shrinkage and Creep in Hardened Concrete*, ACI Report 209.2R-08, Farmington Hills, 2008

T. Bayes, *An Essay Toward Solving a Problem in the Doctrine of Chances*, vol. 53 (Philosophical Transactions, London, 1964), pp. 376–398. Also, W. E. Deming, *Fracsimilie of Two Papers by Bayes* (Department of Agriculture, Washington, DC, 1940)

Z.P. Bažant, Numerically stable algorithm with increasing time steps for integral-type aging creep. In *Proceedings, First International Conference on Structural Mechanics in Reactor Technology (SMiRT-1),* West Berling, ed. by T.A. Jaeger, vol. 4, part H (1971), pp. 119–126

Z.P. Bažant, Prediction of concrete creep effects using age-adjusted effective modulus method. Am. Concr. Inst. J. **69**, 212–217 (1972)

Z.P. Bažant, Theory of creep and shrinkage in concrete structures: a precis of recent developments, in *Mechanics Today*, ed. by S. Nemat-Nasser, American Academy of Mechanics, vol. 2 (Pergamon Press, New York, 1975), pp. 1–93

Z.P. Bažant, Mathematical models for creep and shrinkage of concrete, Chapter 7, in Creep and Shrinkage in Concrete Structures, eds. by Z.P. Bažant, F.H. Wittmann (Wiley, London, 1982), pp. 163–256

Z.P. Bažant, Creep of concrete, in *Encyclopedia of Materials: Science and Technology*, ed. by K.H.J. Buschow, vol. 2C (Elsevier, Amsterdam, 2001), pp. 1797–1800

Z.P. Bažant, S. Baweja, In collaboration with RILEM Committee TC 107-GCS, Creep and shrinkage prediction model for analysis and design of concrete structures–model B3 (RILEM Recommendation 107-GSC). Mater. Struct. (RILEM, Paris) **28**, 357–365 (1995); with Errata, vol. 29 (March 1996), p. 126

Z.P. Bažant, S. Baweja, Short form of creep and shrinkage prediction model B3 for structures of medium sensitivity (Addendum to RILEM Recommendation TC 107–GCS). Mater. Struct. (Paris) **29**, 587–593 (1996)

Z.P. Bažant, S. Baweja, Creep and shrinkage prediction model for analysis and design of concrete structures: model B3, in *Adam Neville Symposium: Creep and Shrinkage – Structural Design Effects*, ed. by A. Al-Manaseer. ACI SP-194 (American Concrete Institute, Farmington Hills, 2000), pp. 1–83

Z.P. Bažant, M.H. Hubler, Theory of cyclic creep of concrete based on Paris law for fatigue growth of subcritical microcracks. JMPS **63**, 187–200 (2014)

Z.P. Bažant, J.-K. Kim, Segmental box girder: deflection probability and Bayesian updating. J. Struct. Eng. ASCE **115**(10), 2528–2547 (1989)

Z.P. Bažant, G.–.H. Li, Unbiased statistical comparison of creep and shrinkage prediction models. ACI Mater. J. **105**(6), 610–621 (2008)

Z.P. Bažant, K.-L. Liu, Random creep and shrinkage in structures: sampling. J. Struct. Eng. ASCE **111**, 1113–1134 (1985)

Z.P. Bažant, L.J. Najjar, Nonlinear water diffusion in nonsaturated concrete. Mater. Struct. (RILEM, Paris) **5**, 3–20 (1972) (Reprinted in *Fifty Years of Evolution of Science and Technology of Building Materials and Structures*, ed. by F.H. Wittmann, RILEM (Aedificatio Publishers, Freiburg, 1997), pp. 435–456)

Z.P. Bažant, L. Panula, Practical prediction of time-dependent deformations of concrete. Mater. Struct., RILEM, Paris (11) Part I, "Shrinkage" **11**, 307–316; Part II, "Basic creep" **11**, 317–328. Part III, "Drying creep" **11**, 415–424. Part IV, "Temperature effect on basic creep" **11**, 425–434 (1978)

Z.P. Bažant, S. Prasannan, Solidification theory for aging creep. Cement Concr Res. **18**(6), 923–932 (1988)

Z.P. Bažant, S. Prasannan, Solidification theory for concrete creep: I. Formulation. J. Eng. Mech. ASCE **115**(8), 1691–1703 (1989a)

Z.P. Bažant, S. Prasannan, Solidification theory for concrete creep: II. Verification and application. J. Eng. Mech. ASCE **115**(8), 1704–1725 (1989b)

Z.P. Bažant, S.T. Wu, Dirichlet series creep function for aging concrete. Proc. ASCE J. Eng. Mech. Div. **99**(EM2), 367–387 (1973)

Z.P. Bažant, Y. Xi, Continuous retardation spectrum for solidification theory of concrete creep. J. Eng. Mech. ASCE **121**(2), 281–288 (1995)

Z.P. Bažant, Y. Xiang, Crack growth and lifetime of concrete under long time loading. J. Eng. Mech. ASCE **123**(4), 350–358 (1997)

Z.P. Bažant, Q. Yu, Relaxation of prestressing steel at varying strain and temperature: viscoplastic constitutive relation. ASCE J. Eng. Mech. **139**(7), 814–823 (2013)

Z.P. Bažant, D. Carreira, A. Walser, Creep and shrinkage in reactor containment shells. J. Struct. Div. Am. Soc. Civil. Eng **101**, 2117–2131 (1975)

Z.P. Bažant, L. Panula, J.-K. Kim, Y. Xi, Improved prediction model for time-dependent deformations of concrete: part 6–simplified code-type formulation. Mater. Struct. **25**(148), 219–223 (1992)

Z.P. Bažant, A.B. Hauggaard, S. Baweja, F.-J. Ulm, Microprestress-solidification theory for concrete creep. I. Aging and drying effects. J. Eng. Mech. ASCE **123**(11), 1188–1194 (1997a)

Z.P. Bažant, A.B. Hauggaard, S. Baweja, Microprestress-solidification theory for concrete creep. II. Algorithm and verification. J. Eng. Mech. ASCE **123**(11), 1195–1201 (1997b)

Z.P. Bažant, Q. Yu, G.H. Li, G.J. Klein, V. Křístek, Excessive deflections of record-span prestressed box girder: lessons learned from the collapse of the Koror-Babeldaob Bridge in Palau. ACI Concr. Int. **32**(6), 44–52 (2010)

Z.P. Bažant, M. Hubler, Q. Yu, Pervasiveness of excessive segmental bridge deflections: wake-up call for creep. ACI Struct. J. **108**(6), 766–774 (2011)
Z.P. Bažant, Q. Yu, G.-H. Li, Excessive long-term deflections of prestressed box girder: part I record-span bridge in Palau and other paradigms. J. Struct. Eng. **138**(6), 676–686 (2012a)
Z.P. Bažant, Q. Yu, G.-H. Li, Excessive long-term deflections of prestressed box girder: part II numerical analysis and lessons learned. J. Struct. Eng. **138**(6), 687–696 (2012b)
Z.P. Bažant, M.H. Hubler, M. Jirasek, Improved estimation of long-term relaxation function from compliance function of aging concrete. ASCE J. Eng. Mech. **139**(2), 146–152 (2012c)
Z.P. Bažant, M.H. Hubler, R. Wendner, Model B4 for creep, drying shrinkage and autogenous shrinkage of normal and high-strength concretes with multi-decade applicability. TC-242-MDC multi–decade creep and shrinkage of concrete: material model and structural analysis. RILEM Mater. Struct. (2014), in review
Berger/ABAM Engineers Inc, *Koror-Babeldaob Bridge Modifications and Repairs*, Oct 1995
J.J. Brooks, Accuracy of estimating long-term strains in concrete. Mag. Concr. Res. **36**(128), 131–145 (1984)
J.J. Brooks, 30-year creep and shrinkage of concrete. Mag. Concr. Res. **57**(9), 545–556 (2005)
R.D. Browne, P.P. Bamforth, The long term creep of Wylfa P. V. concrete for loading ages up to 12 1/2 years, in *3rd International Conference on Structural Mechanics in Reactor Technology* (*SMiRT-3*), paper H1/8, London, 1975
J.D. Buckler, C.F. Scribner, *Relaxation Characteristics of Prestressing Strand*. Engineering Studies, Report No. UILU-ENG-85-2011, University of Illinois, Urbana, 1985
O. Burdet, A. Muttoni, Evaluation des systèmes existants pour le suivi à long terme des déformations des ponts, Swiss Federal Roads Office, N° 607, p. 57, Bern, Switzerland (2006)
R.G. Burg, B.W. Orst, *Engineering Properties of Commercially Available High- Strength Concretes (Including Three Year Data)*. PCA Research and Development Bulletin RD104T (Portland Cement Association, Skokie, 1994), p. 58
I. Carol, Z.P. Bažant, Viscoelasticity with aging caused by solidification of nonaging constituent. J. Eng. Mech. ASCE **119**(11), 2252–2269 (1993)
Comité Euro-International du Béton (CEB), *Recommandations internationales pour le calcul et l'exécution des ouvrages en betón: Principes et recommandations* (CEB, Paris, 1972)
Comité Euro-International du Béton–Fédération International de la Précontrainte (CEB-FIP), *CEB-FIP Model Code for Concrete Structures* (Thomas Telford, London, 1990)
A.H. Cottrell, *The Mechanical Properties of Matter* (Wiley, New York, 1964)
DRC Consultants, Inc. (DRC). Koror-Babelthuap bridge: force distribution in bar tendons (1996)
J. Eliáš, J.L. Le, Modeling of mode-I fatigue crack growth in quasibrittle structures under cyclic compression. Eng. Fract. Mech. **96**, 26–36 (2012)
C. Fairhurst, F. Comet, Rock fracture and fragmentation, in *Rock Mechanics: from Research to Application. Proceedings of the 22nd U.S. Symposium on Rock Mechanics*, ed. by H.H. Einstein (MIT Press, Cambridge, MA, 1981), pp. 21–46.
G.N. Fernie, J.A. Leslie, Vertical and longitudinal deflections of major prestressed concrete bridges, in *Institution of Engineers*, Australia, No. 7516, *Symposium of Serviceability of Concrete*, Melbourne, 19 Aug 1975
FIB, *Structural Concrete: Textbook on Behaviour, Design and Performance, Updated Knowledge of the CEB/FIP Model Code 1990. Bulletin No. 2*, vol. 1 (Fédération internationale du béton (FIB), Lausanne, 1999), pp. 35–52
FIB, *Model Code 2010* (Fédération internationale de béton, Lausanne, 2010)
N.J. Gardner, M.J. Lockman, Design provisions of shrinkage and creep of normal strength concrete. ACI Mater. J. **98**(2), 159–167 (2001)
N.J. Gardner, Design provisions of shrinkage and creep of concrete, in Adam Neville Symposium: Creep and Shrinkage – Structural Design Effect, ed. by A. Al-Manaseer (American Concrete Institute, Farmington Hills, 2000), pp. 101–104
J.A. Hanson, *A Ten-Year Study of Creep Properties of Concrete*. Concrete Laboratory Report No. SP-38, US Department of the Interior, Bureau of Reclamation, Denver, 1953

E.M. Harboe et al., *A Comparison of the Instantaneous and the Sustained Modulus of Elasticity of Concrete,* Concrete Laboratory Report No.C-354, Division of Engineering Laboratories, US Department of the Interior, Bureau of Reclamation, Denver, 1958

W.K. Hatt, Notes on the effect of time element in loading reinforced concrete beams. Proc. Am. Soc. Test. Mater. **7**, 421–433 (1907)

J. Hellesland, R. Green, Sustained and cyclic loading of concrete columns. Proc. ASCE **97**(ST4), 1113–1128 (1971)

G.A. Hirst, A.M. Neville, Activation energy of concrete under short-term static and cyclic stresses. Mag. Concr. Res. **29**(98), 13–18 (1977)

M.H. Hubler, R. Wendner, Z.P. Bažant, Comprehensive database for concrete creep and shrinkage: analysis and recommendations for testing and recording. ACI (2014a), submitted

M.H. Hubler, R. Wendner, Z.P. Bažant, Statistical justification of model B4 for drying and autogenous shrinkage of concrete and comparisons to other models. RILEM Mater. Struct. (2014b), in review

Japan International Cooperation Agency, *Present Condition Survey of the Koror Babelthuap Bridge*, Feb 1990

Japan Road Association, *Specifications for Highway Bridges* (Japan Road Association, Tokyo, 2002)

M. Jirásek, Z.P. Bažant, *Inelastic Analysis of Structures* (Wiley, New York, 2002)

JSCE, Standard specification for design and construction of concrete structure. Jpn. Soc. Civ. Eng. (JSCE) (in Japanese) (1991)

I.M. Kemeny, N.G.W. Cook, Crack models for the failure of rock under compression, in *Proceedings of the 2nd International Conference on Constitutive Laws for Engineering Materials*, eds. by C.S. Desai et al., vol. 2 (Elsevier Science, New York, 1987), pp. 879–887

E. Kern, A. Mehmel, *Elastische und plastische Stauchungen von Beton infolge Druckschwell- und Standbelastung*. Deutscher Ausschuss fur Stahlbeton, Heft, vol. 153 (W. Ernst &Sohn, Berli, 1962)

R. L'Hermite, M. Mamillan, Retrait et fluage des bétons. Annales de L'Inst. Techn. Du Bâtiment et des Travaux Publics **21**(249), 1334 (1968)

R.G. L'Hermite, M. Mamillan, C. Lefèvre, Nouveax résultats de recherchessur la déformation et la rupture du béton. Ann. Inst. Techn. Bâtiment Trav. Publ. **23**, 5–6 (1965)

R.G. L'Hermite, M. Mamillan, A. Bouineau, Nouveaux résultats et récentes éu dessur le fluage du béton. Mater. Struct. (RILEM) **2**, 35–41 (1969)

H. Le Chatelier (translated by J.L. Mack), *Experimental Researches on the Constitution of Hydraulic Mortars* (McGraw Publishing, New York, 1905, original thesis published in 1887)

T.Y. Lin International (TYLI). Collapse of the Koror-Babelthuap Bridge. Technical Report, T.Y. Lin International, San Francisco, USA (1996)

P.Y. Manjure, Rehabilitation/Strengthening of Zuari Bridge on NH-15 in Goa, Paper No.490, Indian Roads Congress, p. 471, 2001–2002

B. McDonald, V. Saraf, B. Ross, A spectacular collapse: the Koror-Babeldaob (Palau) balanced cantilever prestressed, post-tensioned bridge. Ind. Concr. J. **77**(3), 955–962 (2003)

A. Patron-Solares, B. Godart, R. Eymard, Ětude des déformations différées du pont de Savines (Hautes-Alpes). Bulletin des laboratoires des Pontset Chaussées, May–June 1996

W. Pfeil, Twelve years monitoring of long span prestressed concrete bridge. Concr. Int. **3**(8), 79–84 (1981)

M. Pilz, The collapse of the KB bridge in 1996, Dissertation, Imperial College London, 1997

D. Pirtz, Creep characteristics of mass concrete for Dworshak Dam. Report No. 65–2, Structural Engineering Laboratory, University of California, Berkeley, 1968

RILEM Committee TC-69, State of the art in mathematical modeling of creep and shrinkage of concrete, in *Mathematical Modeling of Creep and Shrinkage of Concrete*, ed. by Z.P. Bažant (Wiley, Chichester/New York, 1988), pp. 57–215

F.S. Rostásy, K.-C. Thienel, On prediction of relaxation of cold drawn prestressing wire under constant and variable elevated temperature. Nucl. Eng. Des. **130**(1991), 221–227 (1991)

H.G. Russel, S.C. Larson, Thirteen years of deformations in water tower place. ACI Struct. J. **86** (2), 182–191 (1989)
C.G. Sammis, M.F. Ashby, The failure of brittle porous solids under compressive stress state. Acta Metall. Mat. **34**(3), 511–526 (1986)
S. Suresh, *Fatigue of Materials* (Cambridge University Press, Cambridge, UK, 1998). 193
S. Suresh, E.K. Tschegg, J.R. Brockenbrough, Crack growth in cementitious composites under cyclic compressive loads. Cem. Concr. Res. **19**, 827–833 (1989)
H. Trost, Auswirkungen des Superpositionsprinzip auf Kriech- und Relaxations Probleme bei Beton und Spannbeton. Beton- und Stahlbetonbau **62**, 230–238, 261–269 (1967)
G.E. Troxell, J.E. Raphael, R.W. Davis, Long-time creep and shrinkage tests of plain and reinforced concrete. Proc. ASTM **58**, 1101–1120 (1958)
R. Wendner, M. H. Hubler, and Z. P. Bazant, Optimization method, choice of form and uncertainty quantification of model B4 using laboratory and multi-decade bridge databases. RILEM Mater. Struct. in review, (2014a)
R. Wendner, M. H. Hubler, and Z. P. Bazant, Statistical justification of model B4 for multi-decade concrete creep and comparisons to other models using laboratory and bridge databases, RILEM Mater. Struct. in review, (2014b)
C.P. Whaley, A.M. Neville, Non-elastic deformation of concrete under cyclic compression. Mag. Concr. Res. **25**(84), 145–154 (1973)
A.A. Yee, Record span box girder bridge connects Pacific islands. Concr. Int. **1**, 22–25 (1979)
Q. Yu, Z.P. Bažant, R. Wendner, Improved algorithm for efficient and realistic creep analysis of large creep-sensitive concrete structures. ACI Struct. J. **109**(5), 665–676 (2012)

Damage to Ship Structures Under Uncertainty: Evaluation and Prediction

18

Dan M. Frangopol and Mohamed Soliman

Contents

Abstract

Ship structures are subjected to various deteriorating mechanisms throughout their service life. This deterioration is highly uncertain and can adversely affect the performance and safety of the vessel, and if not addressed properly, catastrophic failures may occur. In this chapter, deteriorating mechanisms affecting ship structures and their prediction models under uncertainty are discussed. In addition, the integration of these models into a general evaluation and management framework is introduced. This integration can support the optimal decision-making process regarding future structural interventions and, eventually, may

D.M. Frangopol (✉) • M. Soliman
Department of Civil and Environmental Engineering, Engineering Research Center for Advanced Technology for Large Structural Systems (ATLSS Center), Lehigh University, Bethlehem PA, USA
e-mail: dan.frangopol@lehigh.edu; mos209@lehigh.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_34

lead to safe and efficient service life extension. The role of structural health monitoring and nondestructive evaluation techniques in damage identification, assessment, and prediction is also discussed.

Introduction

Ships are often subjected to sudden and/or gradual (i.e., time-dependent) damage mechanisms throughout their service life. Sudden structural failures due to extreme events include collision, grounding, fire, and explosions, while time-dependent deterioration mechanisms include fatigue and corrosion. Each damage mechanism requires its own assessment methods suitable to support intervention decisions related to this damage type. Sudden structural damage requires fast damage quantification and assessment of the structural residual strength in order to make effective decisions regarding the future use of the ship. Although the occurrence of such events may be unpredictable, their effects can be correctly managed through the proper emergency response protocols imposed by the ship owners, in addition to various active and passive safety measures. For sudden structural damage, multiple methods can be used to assess the degree of damage and determine the residual structural strength ranging from complex nonlinear finite element analysis (FEA) to simplified formulae (Wang et al. 2002). Reliability and risk of failure due to insufficient residual longitudinal strength of ships damaged by collision or grounding have been also topics of active research (e.g., Fang and Das 2005; Hussein and Guedes Soares 2009; Saydam and Frangopol 2013).

Damage due to time-dependent deterioration, on the other hand, can be predicted through the appropriate modeling of the deterioration phenomena. This prediction process involves multiple sources of uncertainties; thus, it has to be performed probabilistically (Frangopol 2011; Frangopol et al. 2012; Soliman and Frangopol 2013b). Part of these uncertainties is associated with the natural randomness (i.e., aleatory uncertainties) and the other part is associated with inaccuracies in the adopted prediction models (i.e., epistemic uncertainties) (Ang and Tang 2007). The proper modeling of such uncertainties is a key factor affecting the effectiveness and accuracy of the prediction process.

Inspection actions provide valuable information on the actual damage level found at the time of inspection. This assists in the damage evaluation and enables updating the damage propagation model in order to achieve a better damage prognosis process (Soliman and Frangopol 2013a). Various nondestructive testing (NDT) methods can be employed to assess the time-dependent damage of ships. Some of these methods, such as the acoustic emission technique, received noticeable attention within the last decades. The acoustic emission technique was found to provide useful results regarding the damage identification and localization in ship structures.

Structural health monitoring (SHM) systems used to record the ship response are employed to study the structural performance of the ship under normal operational conditions and to validate the assumptions placed during the design phase.

SHM techniques have the potential for the detection and localization of structural damage occurring under severe operational conditions or slam impacts (Salvino and Collette 2009). Multiple approaches have been recently proposed to fulfill this task. However, most of these approaches, which have demonstrated their feasibility in laboratories and controlled environments, still require additional research before they are widely implemented on large and complex structures such as ships (Salvino and Brady 2008).

This chapter presents a brief overview of the damage evaluation and prediction techniques for ship structures, with emphasis on time-dependent damage prediction models. The probabilistic performance evaluation methods suitable to consider the uncertainties associated with these models are emphasized. The role of SHM and NDT in the damage identification process and the recent developments in the service life prediction and extension methodologies for ship structures are also presented.

Time-Based Structural Deterioration Under Fatigue and Corrosion

Time-based damage deterioration mechanisms, such as fatigue and corrosion, are among the major threats affecting ship performance and safety. Due to this deterioration, ship structures require frequent inspections and repairs. Ship deterioration occurs progressively as a result of normal ship operation in the surrounding environment (ISSC 2009). Corrosion can lead to thickness reduction in the affected areas which can ultimately reduce the hull bending capacity. Fatigue, on the other hand, results in cracks that may cause sudden fracture and drastically reduce the structural reliability. These aging effects, when combined with rough sea conditions, may lead to catastrophic ship failures. Generally, time-dependent deterioration initiation and propagation processes are highly uncertain. This adds challenges to the performance assessment and service life estimation. The time-dependent damage level with the effect of uncertainties is shown schematically in Fig. 1. As shown, at any point in time, the damage level can be described by its probability density function (PDF). Additionally, the time required to reach certain damage level carries significant uncertainty.

Maintenance actions applied throughout the service life of ships can either reduce the damage level (e.g., by replacing the damaged component), or prevent further damage propagation for a certain period of time (e.g., by applying corrosion coatings) (Kim et al. 2013). Both maintenance types, denoted as M_1 and M_2, respectively, result in an extension in the service life. The effect of both maintenance types on the time-dependent damage level is shown in Fig. 2.

Predicting fatigue and corrosion damage initiation and propagation has been an active research topic for decades. As a result, several analytical models have been proposed for predicting the structural capacity and service life. The next sections present the commonly used fatigue and corrosion damage prediction models. Later in this chapter, probabilistic performance assessment considering uncertainties associated with these models will be discussed.

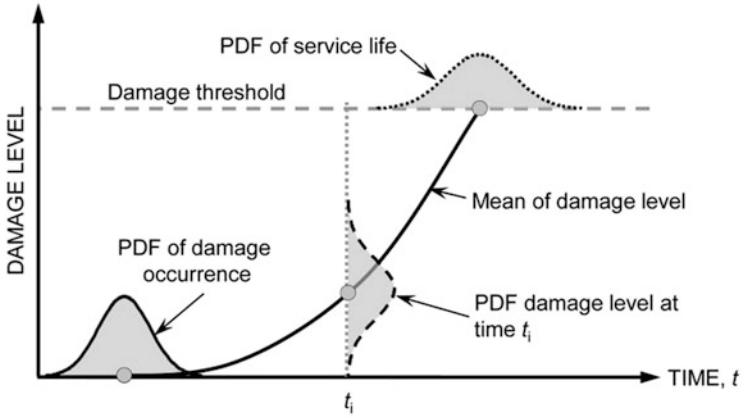


Fig. 1 Damage initiation and propagation under uncertainty

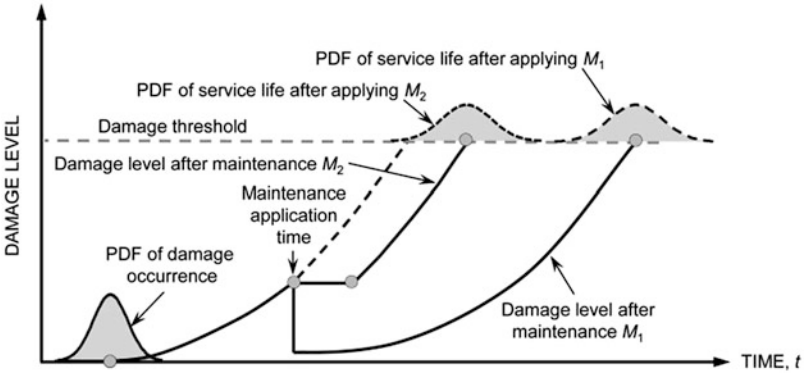


Fig. 2 Effect of maintenance on the damage level and service life

Corrosion in Ships

Several types of corrosion wastage in mild and low alloy steels in marine environments exist, such as uniform (general), pitting, stress, and galvanic corrosion. For corrosion management and control, both localized and general corrosion must be considered. The former can cause oil or gas leaks, while the latter, which spreads over the surface of the affected area, is more likely to lead to structural strength problems. Stress corrosion occurs in some alloys when exposed to corrosive environments while mechanically stressed. Furthermore, when two different metals are physically connected, galvanic accelerated corrosion occurs in the less noble metal (ISSC 2009). Factors affecting marine immersion corrosion include the type of structural material, corrosion protection method (e.g., coating, cathodic

protection), type of cargo or stored material, cycles of loading/unloading of cargo or stored material, humidity, and temperature (ISSC 2006).

In recent years, extensive work has been performed to investigate different parameters affecting general corrosion wastage and to formulate corrosion wastage prediction models (Paik et al. 2003a, b; Melchers 2002, 2003a, b, 2004c, 2006; Guedes Soares and Garbatov 1999; Guedes Soares et al. 2005). For example, Guedes Soares et al. (2005) investigated the influence of salt content, water temperature, dissolved oxygen, PH value, and water velocity on the general corrosion rate and included these effects in the nonlinear corrosion wastage model proposed in Guedes Soares and Garbatov (1999). Their model consists of three corrosion loss stages. The first is penetration of the water particles through the corrosion coating, the second is the formation of the two-dimensional monolayer oxide film, and the third is the start and growth of the three-dimensional oxide nuclei. In this model, the first two stages represent the coating effectiveness period where the corrosion depth at any time t can be found as (Guedes Soares et al. 2005)

$$d(t) = d_\infty \left(1 - e^{\frac{-(t-\tau_c)}{\tau_t}}\right) \quad \text{for} \quad t > \tau_c \tag{1a}$$

$$d(t) = 0 \quad \text{for} \quad t \le \tau_c \tag{1b}$$

where $d(t)$ is the time-dependent corrosion depth and d_∞, τ_c, and τ_t are model parameters depending on the coating type and operational and environmental conditions.

Melchers (2003a, b, 2006) developed a corrosion wastage prediction model consisting of the following phases of average corrosion loss: (a) short-term initial phase in which the corrosion is governed by the chemical kinetics, (b) approximated linear function dependent on the oxygen diffusion from surrounding water, (c) nonlinear function governed by oxygen diffusion through corrosion product layer, (d) anaerobic bacterial corrosion phase, and (e) linearly approximated long-term anaerobic bacterial corrosion phase.

Research work has also been performed to model pitting corrosion. However, the scarcity of corrosion depth measurements for this type of corrosion compared to the general corrosion poses additional challenges. In this context, Melchers (2004a, b) proposed a multiphase model for pitting corrosion loss as a function of exposure time.

Due to the importance of the corrosion assessment and repair topic, multiple classification societies issued recommendations and regulations for corrosion coating, prevention, inspection, and repair of corroded steel ships (e.g., DNV 1998, 1999; IACS 2003). Corrosion wastage prediction is a process covered by various uncertainties; thus, it has to be conducted probabilistically. Although many corrosion models are available, these models are based on statistical data collected from different vessels; as new construction techniques and materials emerge, these models should be updated and refined.

Time-dependent corrosion losses have an effect on the structural resistance of a ship and should be considered in its life-cycle performance assessment (Kwon and Frangopol 2012a). Corrosion losses may cause reduction in the hull structural

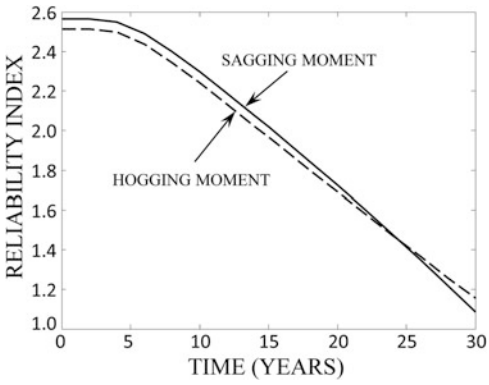


Fig. 3 Effect of corrosion on the time-variant performance (Adapted from Frangopol and Okasha (2010))

resistance, reduction in the local strength, and increase in the fatigue crack propagation within the affected areas. Considering general corrosion, multiple studies have been performed to predict the time-variant hull structural resistance by estimating the loss in the hull girder section modulus due to corrosion (e.g., Ayyub et al. 2000; Paik and Wang 2003; Okasha et al. 2010; Decò et al. 2011, 2012). Figure 3 shows the time-variant reliability index of a steel ship studied in Frangopol and Okasha (2010). As shown, the performance of the ship drops significantly due to corrosion. It is observed that most of the analytical studies tend to overestimate the effect of corrosion on the hull girder strength. In an attempt to address this point, Wang et al. (2008) presented a statistical study showing the loss in the hull girder section modulus in a database of 222 steel ships. This type of analysis can support the verification and calibration of the hull resistance prediction models.

Aluminum alloys used in ship construction, mainly 5xxx-series alloys, have excellent corrosion resistance in marine environments. Part of the corrosion resistance of aluminum is attributed to the formation of a thin oxide layer which prevents the core metal from any further corrosion. This layer is hard and renews itself almost instantly in case of any mechanical abrasion. It is very stable under most conditions except for extreme PH values where it may lose its stability; additionally, the self-renewal may not be fast enough to prevent further corrosion. However, since aluminum is a very active metal, it is highly prone to galvanic corrosion if not properly isolated. Galvanic action, especially at areas where both steel and aluminum are connected, makes the aluminum vulnerable to corrosion. The corrosion damage in this case may be very fast (ISSC 2009). An example of this type of problem was observed in the USS Independence LCS-2, a 127.4 m, high-speed trimaran capable of speeds up to 44 knots, in which corrosion initiated at the locations where the aluminum hull was in contact with the steel propulsion system (O'Rourke 2012). However, this mode of corrosion can be easily prevented by the use of appropriate isolations or cathodic protection systems.

Another mode of deterioration of aluminum ships is sensitization, which is a degradation mode that occurs in high-magnesium aluminum alloys (e.g., 5083,

5086, 5456, and 5383) when exposed to elevated temperatures (Sielski 2007). Under certain conditions, these alloys may suffer intergranular corrosion due to the precipitation of the beta-phase (Mg_2Al_3) on the grain boundaries. This precipitate is electromechanically more active than the aluminum matrix and can cause further intergranular corrosion with the continued grain boundary migration. Furthermore, this process increases the material susceptibility to stress corrosion cracking, exfoliation, and decreased ductility. Recent studies were carried out to find the time required to sensitize the material based on the thermal profile of the ship. However, this is directly related to the location of the plate within the ship as it is heavily dependent on the stress profile acting on the studied location (Sielski et al. 2012).

Fatigue in Steel and Aluminum Ships

Fatigue is one of the major stressors affecting ship structures. Although many classification societies issued codes and regulations for the proper fatigue design and assessment, ship structures still suffer fatigue cracking. Fatigue is the process of damage accumulation caused by repeated fluctuating loads. Fatigue damage can exist in mild environments as well as aggressive ones (i.e., corrosion-induced fatigue). For a component subjected to elastic stress fluctuations, fatigue damage may accumulate at regions of stress concentration, where the local stress exceeds the yield limit of the material (Barsom and Rolfe 1999). Stress concentrations can occur in a component due to the presence of initial flaws in the material, welding process, or fabrication. Initiation and propagation of cracks in the plastic localized region occur due to the cumulative damage acting over a certain number of stress fluctuations. These cracks can eventually cause the fracture of the component. This process can be minimized by adopting better details, avoiding stress concentrations, and decreasing the number of welded attachments, among others. Currently, design specifications give the guidelines for maximizing the fatigue life and offer the means for selecting details associated with high fatigue resistance (Fisher et al. 1998).

Fatigue for ship structures can generally be assessed by the *S-N* (i.e., stress-life) approach and the fracture mechanics approach (also known as the crack growth approach). The former gives the relationship between the stress acting on the detail and the predicted number of stress cycles to failure, while the latter provides a theoretical model to calculate the crack size in relation to the number of cycles acting on the detail. A brief discussion on both approaches is provided in the next subsections.

The *S-N* Approach

In the *S-N* approach, the fatigue life of a certain detail is determined in a laboratory test by applying constant or variable amplitude stress cycles to the detail until a crack with predefined size grows through the detail. The test is repeated for several specimens and for different stress amplitudes. Next, the stress-range amplitude is

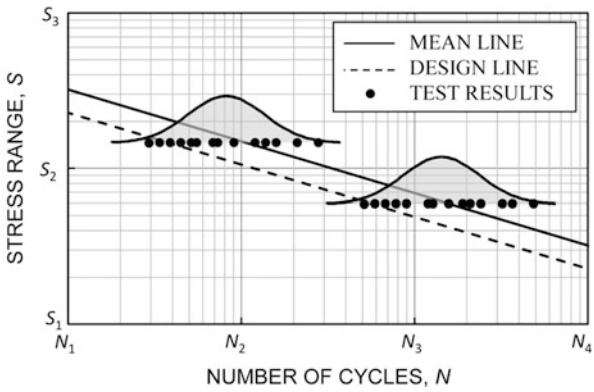


Fig. 4 *S-N* mean and design lines

plotted versus the number of cycles to failure in a logarithmic scale plot, as shown in Fig. 4, and a linear or multi-linear fitting of the data is performed yielding the mean *S-N* lines. Due to the variability in test results, a design line is usually defined by codes in which the mean line is shifted to the left by a certain amount sufficient to achieve a satisfactory probability of survival for designed structures. For example, the AASHTO LRFD design specifications (AASHTO 2010) shift the mean line to the left by two standard deviations indicating that approximately 95 % of the specimens would survive the associated number of cycles (Fisher et al. 1998). The resulting *S-N* relationship of a detail can be expressed, for a single-slope *S-N* relation, as

$$S = \left(\frac{A}{N}\right)^{\frac{1}{m}} \tag{2}$$

in which S is the stress range (i.e., fatigue resistance), A is a fatigue detail coefficient for each category, N is the number of cycles, and m is a material constant defining the value of the slope of the *S-N* line.

Ship details are normally subjected to variable amplitude stress-range cycles; hence, an equivalent constant amplitude stress range is needed for fatigue assessment. Miner's rule (Miner 1945) is widely used for ship structures to quantify the fatigue damage accumulation at details subjected to variable amplitude loading with a known stress-range histogram. By assuming a linear damage accumulation, Miner's damage accumulation index D is

$$D = \sum_{i=1}^{n_{ss}} \frac{n_i}{N_i} \tag{3}$$

where n_{ss} is the number of stress-range bins in a stress-range histogram, n_i is the number of stress cycles in the ith bin with stress range S_i, and N_i is the number of cycles to failure under the stress range S_i. According to Miner's damage

accumulation rule, the failure of the detail occurs when $D = 1.0$. However, research showed that this value is subjected to significant variability, and, up to date, no value is widely accepted by all research communities.

Based on Miner's damage accumulation rule, an equivalent constant amplitude stress range can be defined as

$$S_{re} = \left[\sum_{i=1}^{n_{ss}} \frac{n_i}{N_T} \cdot S_i^m \right]^{\frac{1}{m}} \tag{4}$$

where $N_T = \sum_{i=1}^{n_{ss}} n_i \cdot S_{re}$ can be alternatively calculated using the PDF $f_S(s)$ of the stress range S as

$$S_{re} = \left[\int_0^{\infty} s^{m_1} \cdot f_S(s) \cdot ds \right]^{\frac{1}{m_1}} \tag{5}$$

For ship details, the stress range can follow lognormal, Rayleigh, or Weibull distributions. The three-parameter PDFs of these distributions, including the cutoff threshold s_c, are expressed, respectively, as

$$f_S(s) = \frac{1}{(s - s_c) \cdot \zeta \cdot \sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2} \cdot \left(\frac{\ln(s - s_c) - \lambda}{\zeta} \right)^2 \right] \tag{6}$$

$$f_S(s) = \left(\frac{s - s_c}{S_{ro}^2} \right) \cdot \exp\left[-\frac{1}{2} \left(\frac{s - s_c}{S_{ro}} \right)^2 \right] \tag{7}$$

$$f_S(s) = \frac{\kappa}{\alpha} \cdot \left(\frac{s - s_c}{\alpha} \right)^{\kappa - 1} \cdot \exp\left[-\left(\frac{s - s_c}{\alpha} \right)^{\kappa} \right] \tag{8}$$

where $s > s_c$, α, and κ are the scale and shape parameters of the Weibull distribution, respectively; λ and ζ are the location parameter and scale parameters for the lognormal distribution, respectively; and S_{ro} is the mode of the Rayleigh distribution. Needless to say, depending on the stress-range bin histogram, a two-parameter PDF can also be used considering $s_c = 0$.

Using the equivalent constant amplitude stress range, fatigue life, measured as the number of cycles to failure, is calculated as

$$N = \frac{A}{S_{re}^m} \tag{9}$$

This number of cycles can be used in conjunction with the average annual number of cycles N_{avg} to estimate the fatigue life in years using the following equation:

$$t(\text{years}) = \frac{N}{N_{avg}} \tag{10}$$

The *S-N* approach has been widely used for fatigue assessment of steel and aluminum ship details. Multiple design specifications and research reports are available for fatigue design and assessment of ship details (e.g., BS 5400 1980; ABS 2010; DNV 1997, 2010; Eurocode 3 2010; Eurocode 9 2009). Since the estimation of the resistance and demand terms in the *S-N* approach is straightforward, this approach has been successfully used for the reliability-based fatigue assessment of ships. In this context, Ayyub et al. (2002) proposed reliability-based design guidelines for fatigue of ship details. They briefly discussed the available fatigue assessment methods for ship structures and their associated parameters. Kwon et al. (2013) conducted fatigue reliability assessment, based on SHM data, by estimating the probabilistic lifetime sea loads for high-speed ship structures. The British Standards *S-N* relationships (BS 5400 1980) were used in their approach.

The Fracture Mechanics Approach

Although the *S-N* approach is widely used for the fatigue assessment of ships, it cannot be used to study the crack condition at a given detail since it does not provide a direct relation between the crack size and the number of cycles affecting the detail. The approach based on fracture mechanics, on the other hand, can be used to study the crack conditions and stability in a damaged detail. In this method, the stresses near the crack tip, which are responsible for the crack propagation, are related to the stress intensity factor K. Linear elastic fracture mechanics (LEFM) can be applied through Paris' equation (Paris and Erdogan 1963) for assessing fatigue behavior of steel details. This equation relates the crack growth rate to the range of the stress intensity factor as follows:

$$\frac{da}{dN} = C \cdot (\Delta K)^m \tag{11}$$

where a is the crack size, N is the number of cycles, and ΔK is the range of the stress intensity factor. C and m are material parameters. The values for C and m can be found through experimental reports or code specifications. For example, the British Standards BS 7910 (2005) provides the values for C and m of 2.3×10^{-12} and 3.0, respectively, using the units of mm/cycle for da/dN and N/mm$^{3/2}$ for ΔK, for simplified assessment of steel details operating in marine environment. The range of the stress intensity factor can be expressed as

$$\Delta K = Y(a) \cdot S \cdot \sqrt{\pi a} \tag{12}$$

where S is the stress range and $Y(a)$ is a correction factor which depends on the crack orientation and shape. This correction factor takes into account the effects of the elliptical crack shape, free surface, finite width (or thickness), and nonuniform stress acting on the crack. More detailed empirical and exact solutions for these correction factors can be found in Tada et al. (2000).

Using Eqs. 11 and 12, the number of cycles associated with a growth in the crack size from an initial size of a_o to a size of a_t can be calculated as

$$N = \frac{1}{C \cdot S^m} \cdot \int_{a_o}^{a_t} \frac{1}{(Y(a) \cdot \sqrt{\pi a})^m} da \tag{13}$$

By setting a_t in Eq. 13 to be equal to the critical crack size a_f, the number of cycles to failure of the detail is obtained. This approach can also be implemented in the probabilistic fatigue life assessment and inspection and monitoring planning of ships. For instance, Kim and Frangopol (2011c) used this approach to find the optimum inspection times which minimize the damage detection delay in steel ship details.

Probabilistic Performance Assessment and Prediction

Probabilistic performance assessment methods are suitable for ships due to the presence of various uncertainties associated with sea loading, ship operation, damage initiation and propagation, and their impact on the structural resistance. Several probabilistic approaches are available to assess the structural performance (e.g., Ayyub et al. 2000; Okasha and Frangopol 2010b; Okasha et al. 2011; Kim and Frangopol 2011a, b, c; Kwon and Frangopol 2012b; Decò and Frangopol 2013). Some of them use solely the time-variant damage level, quantified by simulation techniques, to assess the performance, while others use probabilistic performance indicators such as reliability, redundancy, and risk. Each of these probabilistic performance indicators represents a distinctive structural feature that can be useful for performance assessment and life-cycle management under uncertainty. In the next example, probabilistic fatigue life estimation for a steel ship detail is performed using Monte Carlo simulation. Later in this section, structural reliability analysis is briefly discussed and an example of the reliability assessment and maintenance scheduling is presented.

Example 1 Fatigue cracking is a major safety concern for ship structures. Probabilistic simulation methods can be used to predict the fatigue damage propagation and provide an indication about the expected service life of the investigated location. As an example, a welded joint between the bottom plate and longitudinal stiffener in the hull structure of a steel ship, shown in Fig. 5, is considered. During the routine inspection, a crack with a mean size of 2.0 mm was found to initiate from the stiffener to bottom plate weld and was propagating transversally as shown in Fig. 5.

Crack propagation for such detail can be studied using Eq. 13 after determining the parameters C, m, a_o, and S. Moreover, if the average annual number of cycles N_{avg} is known, the crack length over time can be found. For this example, the fatigue crack growth parameters C, m, and a_o are assumed to follow lognormal distributions, whereas the stress range is treated as a random variable following a

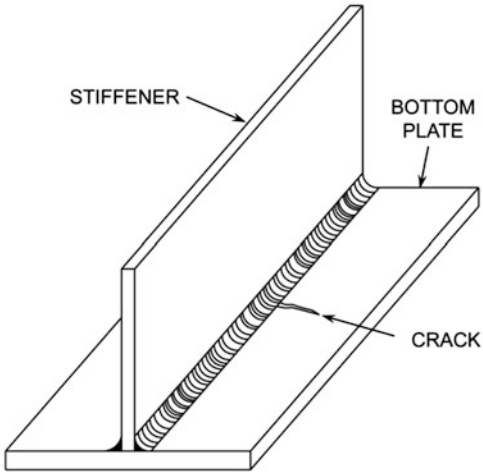


Fig. 5 Critical fatigue detail

Table 1 Random variables and deterministic parameters associated with the crack growth model

Variable	Notation (units)	Mean value	Coefficient of variation	Distribution type
Material crack growth parameter	C	3.54×10^{-11}	0.3	Lognormal
Material crack growth exponent	m	2.54	–	Deterministic
Initial crack size	a_o (mm)	2.0	0.2	Lognormal
Daily number of cycles	N_{avg} (cycles/year)	1.0×10^6	0.3	Lognormal
Stress range	S (MPa)	30	0.1	Weibull

Weibull distribution. The mean value of the parameter C is assumed 3.54×10^{-11}, using units of MPa for stress range and mm for crack size (this translates to 1.77×10^{-9} using units of ksi for stress and in for crack size), and m is assumed 2.54 (Dobson et al. 1983). The descriptors of the variables associated with the crack growth are given in Table 1. In this example, the geometric function $Y(a)$ is assumed to be one (Akpan et al. 2002).

For this detail, knowing the average number of cycles enables calculating the time associated with crack growth from the initial size a_o to a given size a_t as

$$t(\text{years}) = \frac{1}{N_{avg} \cdot C \cdot S^m} \cdot \int_{a_o}^{a_t} \frac{1}{(Y(a) \cdot \sqrt{\pi a})^m} da \tag{14}$$

Considering the final crack size to be 50 mm, the time associated with the growth from 2.0 to 50 mm can be found using Monte Carlo simulation in which the random variables are represented by their respective PDFs. For this example, a Monte Carlo simulation with 100,000 samples yields the histogram shown in Fig. 6 for the time

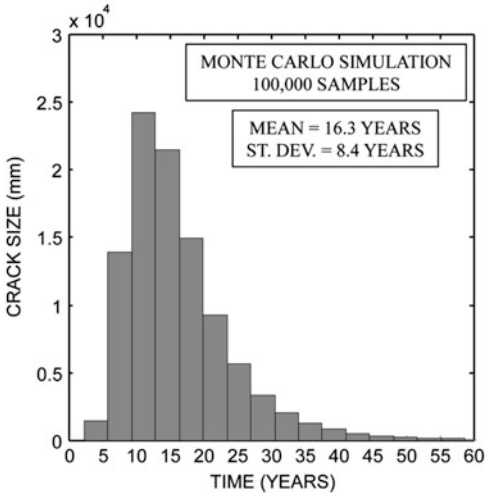


Fig. 6 Histogram of time required to reach a crack size of 50 mm

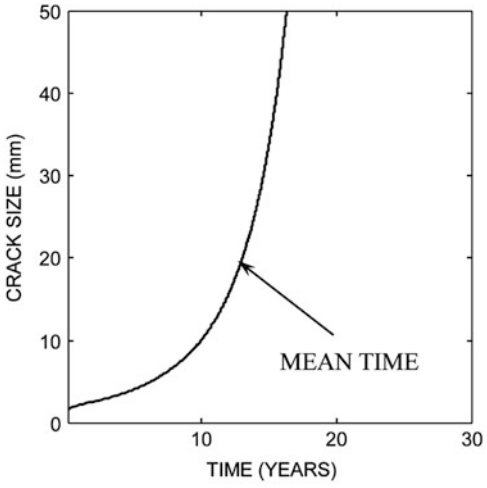


Fig. 7 Mean time to reach various crack sizes

to reach the final crack size. Additionally, as shown in Fig. 7, the simulation can be used to find the mean time for the crack to grow from the initial size to various crack sizes. Inspection and repair actions can be subsequently planned based on the required target safety levels.

Probabilistic performance indicators, such as the reliability index, provide measures for the structural reliability while considering the aforementioned uncertainties. Thus, they can be used to predict the service life and plan for future inspection, maintenance, and monitoring actions (Frangopol and Messervey 2009a, b; Frangopol and Kim 2011). Figure 8 shows schematically the probabilistic performance profile of a structure including effects of aging, sudden damage, and repair actions. Two maintenance types can be defined based on their application time and the performance

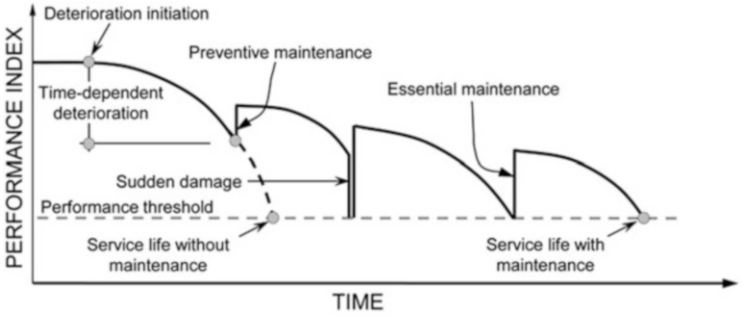


Fig. 8 Probabilistic performance index profile including effects aging, maintenance, and sudden damage

level at this time, namely, essential maintenance (EM) and preventive maintenance (PM). EM is performance based, in which the maintenance is performed when the performance indicator reaches its allowable threshold. In contrast, PM is usually time based in that it is typically applied at prescribed instants over the life cycle of the structure. PM can be performed either to delay the damage propagation for an effective period of time or to slightly improve the performance of the structure. EM, on the other hand, should significantly improve the performance of the structure in order to produce a substantial extension in the service life.

Structural Reliability Analysis

In general, the reliability of a structural component can be related to the probability of failure, defined as the probability of violating a certain limit state $g(\boldsymbol{X}) = 0$. The performance function $g(\boldsymbol{X})$ may be defined as the safety margin

$$g(X) = R - S \tag{15}$$

where R and S are the random capacity and demand of the structure, respectively, and $\boldsymbol{X}$ is the random variable vector. Based on the considered limit state, the probability of failure P_f can be defined as

$$P_f = P(g(X) \leq 0) \tag{16}$$

The PDFs of R, S, and safety margin (i.e., $R - S$) as well as the probability of failure P_f are represented in Fig. 9. Thus, the reliability index β can be defined as

$$\beta = \Phi^{-1}(1 - P_f) \tag{17}$$

where $\Phi^{-1}(\cdot)$ denotes the inverse standard normal cumulative distribution function (CDF).

For cases where R and S are statistically independent normally or lognormally distributed random variables, exact expressions for calculating the probability of failure can be formulated (see, e.g., Ang and Tang 1984). For more complex

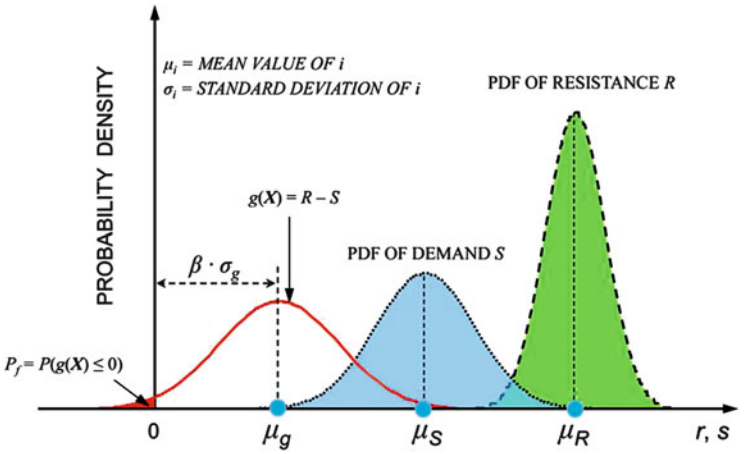


Fig. 9 PDFs of resistance, demand, and safety margin

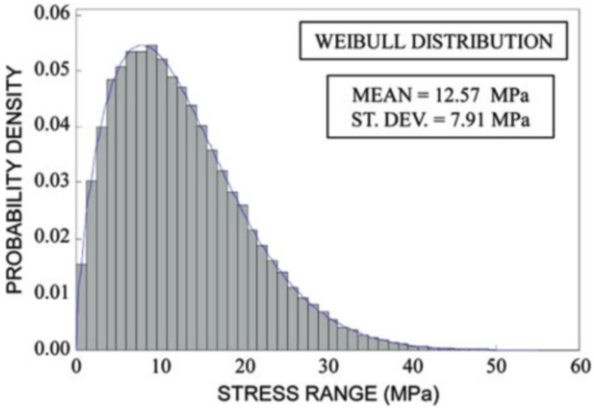


Fig. 10 Probability density of the stress range affecting the ship detail

problems, where R and/or S follow a PDF other than normal or lognormal, efficient reliability techniques can be used to evaluate the component reliability, such as the first-order reliability method (FORM), second-order reliability method (SORM), and Monte Carlo simulation. FORM and SORM have been widely employed in many structural reliability problems and various software packages, such as RELSYS (Estes and Frangopol 1998), to calculate the reliability indices of structural components and systems.

Example 2 To illustrate the reliability concepts for fatigue assessment of steel ship details, consider a steel ship detail subjected to the stress-range distribution shown in Fig. 10 with an average annual number of cycles of 1.5×10^6. Based on the S-N approach of the BS 5400 (1980) specifications, the detail is classified under fatigue category F of this code.

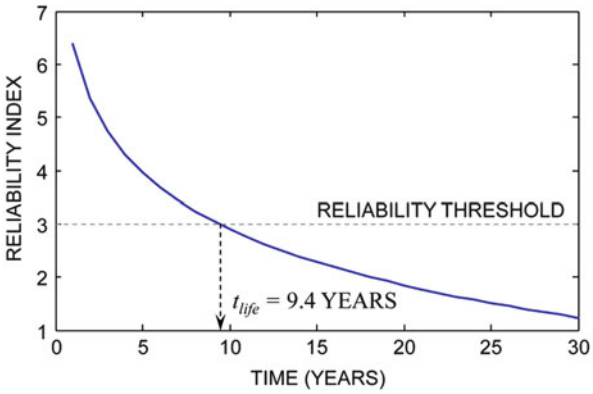


Fig. 11 Reliability index without maintenance

The material constant m for this detail is 3.0, while the constant A (see Eq. 2) is assumed to follow a lognormal distribution with mean of 6.29×10^{11} MPa^3 and a coefficient of variation of 0.54 (Kwon et al. 2013). Based on Eqs. 5 and 8, the equivalent constant amplitude stress range S_{re} is 17.64 MPa. To account for uncertainty in this value, S_{re} is assumed to follow a lognormal distribution with mean 17.64 MPa and coefficient of variation 0.1.

To study the fatigue reliability of the detail, a performance function can be defined as the safety margin

$$g(t) = \Delta - D(t) \tag{18}$$

where Δ = Miner's critical damage accumulation index, indicating the allowable accumulated damage and assumed lognormal distributed with mean 1.0 and coefficient of variation (COV) 0.3 (Wirsching 1984); $D(t)$ = Miner's damage accumulation index, which can be expressed as

$$D(t) = \frac{N(t)}{A} \cdot S_{re}^{m} \tag{19}$$

Based on Eqs. 18 and 19 and assuming that the random variables S_{re}, A, and Δ are also lognormally distributed, the fatigue reliability index β can be derived as (Kwon and Frangopol 2010)

$$\beta(t) = \frac{\lambda_\Delta + \lambda_A - m \cdot \lambda_{S_{re}} - \ln N(t)}{\sqrt{\zeta_\Delta^2 + \zeta_A^2 + \left(m \cdot \zeta_{S_{re}}\right)^2}} \tag{20}$$

where λ and ζ are the parameters associated with different random variables. Using Eq. 20, the reliability profile of the detail can be found as shown in Fig. 11. The fatigue life of the detail can be calculated by setting a threshold for the reliability index. For ship details subjected to fatigue, a reliability index threshold ranging

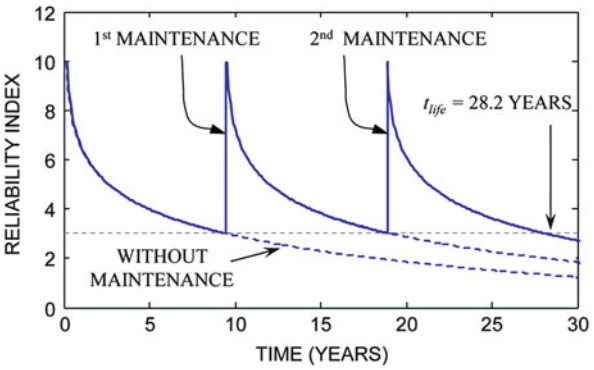


Fig. 12 Service life with EM maintenance

from 2.0 to 4.0 is appropriate (Mansour et al. 1996). For this example, this threshold is set to be 3.0 yielding a fatigue life without maintenance of 9.4 years.

Threshold-based EM, in which the performance is restored to the initial level, can be applied to extend the service life. As shown in Fig. 12, essential maintenance can be performed at 9.4 and 18.8 years yielding a total service life of 28.2 years (i.e., life extension of 18.8 years).

Although the maintenance planning provided in this example is straightforward, other cases of maintenance optimization are not as simple. This is especially true if multiple maintenance actions of varying types are applied to the structure specifically when each of them yields its own service life extension. In this case, probabilistic optimization techniques can be used efficiently to solve such problems. The topic of maintenance optimization is discussed in Okasha and Frangopol (2010a) and Kim et al. (2013).

Damage Evaluation Using NDT and SHM

NDT-based inspection plays a great role in the damage identification and assessment of ship structures. Up to date, the most widely adopted damage evaluation method is visual inspection. This is mainly due to the cost-effectiveness and the ease of application. However, successful visual inspection is challenged by multiple factors including the level of inspector's experience and accessibility problems due to fire protection and corrosion coating. On the other hand, NDT methods, such as ultrasonic inspection, face more challenges arising from the large scale of the structure and number of locations requiring inspection. In addition, the exact location of damage is generally required to apply these inspection methods, which is generally not the case. Research in the field of NDT methods that can identify the location and damage level is very active. These methods mostly rely on installing sensors that continuously record the structural response or emissions and attempt to identify and localize the damage based on the recorded data. These systems include regular strain gages, accelerometers, and acoustic emission sensors. Information from such systems can also be used to update and calibrate

performance prediction and damage propagation models to achieve more reliable and accurate performance assessment process (Zhu and Frangopol 2013a, b). In the next subsection, the recent developments in damage identification using the acoustic emissions and SHM are briefly discussed.

Damage Identification Using Acoustic Emission Technique

Within the last decade, acoustic emission technique has received considerable attention for its use in the fatigue and corrosion damage detection and localization in ships. In this approach, stress waves emitted by the material during sudden changes in the internal structure are recorded using special sensors and used to detect structural damage such as crack initiation and growth, fracture, plastic deformation, corrosion, and stress corrosion cracking, among others (Anastasopoulos et al. 2009). In general, a uniform steel specimen with no stress raisers will start emitting acoustic emissions when stressed to a level of 60 % of its yield stress (Anastasopoulos et al. 2009). During normal operation of the ship, these emissions can be continuously detected and recorded such that structural damage can be monitored. This approach has been successfully applied to different types of structures such as bridges, pressure vessels, and pipelines. Recently, research programs in Europe (see, e.g., Baran et al. 2012; Tscheliesnig 2006) and the United States (see, e.g., Wang et al. 2010) have shown the feasibility of such an approach in detecting corrosion and crack damage in ship structures. In these research programs, the results of controlled laboratory testing of specimens subjected to fatigue and accelerated corrosion, as well as oil tankers showed the feasibility of the approach. Since acoustic emission signals can be very weak, especially for corrosion detection, damage detection may be significantly affected by the noise arising from the normal ship operation. The research in this area also aimed to evaluate and isolate the noise under real operation conditions. Special pattern recognition techniques can be used to filter the noise (Baran et al. 2012). Multiple damage detection approaches have been developed, along with their necessary hardware. Some approaches use immersed sensors to detect the acoustic waves travelling through liquids in tankers, while others use sensors attached directly to the structure. The results of such research programs show that using acoustic emissions for the continuous real-time monitoring of damage due to fatigue or corrosion is a promising approach.

Application of SHM for Damage Identification in Ships

A parallel effort has been running to develop approaches which can support the use of SHM systems for damage detection in ships. SHM systems employ various types of sensors, accelerometers, and strain gages to record the structural response during normal ship operation. These systems can be used on multiple fronts such as the validation of design assumptions, monitoring the structural response under normal

operation, damage detection and diagnosis, prognosis, and useful life estimation (Salvino and Brady 2008). Validation of the design assumptions is performed typically after the ship is constructed. In this process, the ship is operated through predesigned seakeeping trial runs subjecting the ship to various combinations of operational conditions in terms of speed, sea state, and heading angles to ensure that the actual structural responses are within the design and allowable limits. Information from seakeeping trials may also be used to adjust the safe operational profile by removing some restrictive operational condition if the monitoring shows an acceptable response under these conditions (Salvino and Collette 2009). On the other hand, seakeeping trial data can be used to reduce the likelihood of the ship damage under conditions that were not shown to cause damage during design phase. Additionally, SHM systems can be also used to evaluate the integrity and vibration levels of the propulsion systems of ships (Brady 2004).

After the initial seakeeping trials, the monitoring system can be used for the continuous health assessment of the ship systems. Various high-speed ships are equipped with accelerometers that are always online and can warn the crew when the acceleration levels exceed the allowable threshold. Exceedance can occur due to slamming events and the crew can reduce the speed accordingly (Salvino and Collette 2009). The current research in this area aims to develop monitoring systems, acquisition systems, and the supporting software that is capable of providing real-time information to the ship crew regarding the structural system integrity and response under ship operation (Hess 2007; Salvino and Brady 2008; Swartz et al. 2012). Moreover, such a system should be able to enhance the ability for damage diagnosis and prognosis. These systems provide the possibility to identify damage at its early stages and support the scheduling of inspection and maintenance activities. SHM information can be used to aid in the detection of damage in the areas difficult to access. Moreover, it can be performed while the ship is in service; this minimizes the disturbance of ship operation and extends the operational time of the vessel. Up to date, the most common damage prognosis based on SHM data is applied in order to quantify fatigue damage in ship structures. This is performed by recording the strains at the monitored locations and converting those strains to stresses, and by using the appropriate classification guidelines, the fatigue life can be found using Miner's damage accumulation rule (Hess 2007). In this approach, the stresses and the cycle count are used to find the percentage of the consumed life under the vessel operational profile and find the remaining fatigue life. However, these prognosis methods cannot be directly used to study the crack conditions at a damaged location. Additionally, they cannot be used to assess other damage mechanisms such as corrosion or damage due to slamming.

Damage detection techniques based on SHM such as the vibration-based methods are under continuous development for use in ship structures. Vibration-based methods use advanced signal processing techniques such as the empirical mode decomposition and Hilbert-Huang transform to detect the damage by determining the change in the dynamic properties of the structure. This is based on the fact that a change in the mode shapes or frequencies would suggest that a change

has occurred to the physical properties of the structure (Salvino and Brady 2008). Due to the inherent randomness associated with the monitoring outcomes, it is necessary to integrate those uncertainties in the damage detection technique (Okasha et al. 2011). Methods such as vector autoregressive modeling can be used for the detection and localization of damage in high-speed naval vessels. In this method, the vibration signal obtained from the structure as a reference signal is modeled and this model is fitted to the measured structural response. The parameters of this model are the damage-sensitive features (Okasha et al. 2011). The model is assumed to provide an accurate prediction of the structural response; thus, an increase in the difference between the model data and the data measured in the future is interpreted as an indication of structural damage. Mattson and Pandit (2006) proposed a vector-based model which allows a signal to be described in terms of its own past values as well as the past values of other sensors.

A measure of the goodness of fit can be used to select the order of the autoregressive model which is a function of the predicted signal and the measured one. An application of such method was conducted by Mattson and Pandit (2006) on an experimental setup. Additionally, the feasibility of applying this model to ships has been tested in Okasha et al. (2011). Although the damage detection using vibration-based statistical methods is found to be a promising approach, more research is still required for verification, validation, and statistical quantification of such models in order to be reliably applied to SHM of ship structures.

Conclusions

This chapter briefly discussed the damage mechanisms affecting steel and aluminum ships with emphasis on time-dependent effects such as fatigue and corrosion. Different damage prediction models for fatigue and corrosion were briefly presented in addition to various sources of uncertainty associated with these deteriorating mechanisms. Additionally, damage identification through NDE methods and SHM was discussed.

The presence of uncertainties associated with the ship loading, operational conditions, and damage prediction models calls for the use of probabilistic performance indicators in the damage prediction process. Such indicators provide rational quantification of the ship performance and safety while considering various sources of randomness. Moreover, those performance indicators can be effectively integrated within the life-cycle management framework to support decision making regarding future inspection and maintenance activities.

Many of the damage evaluation and prediction techniques for marine structures, as well as life-cycle estimation, prediction, and extension, are also used for civil structures such as bridges and buildings (Frangopol and Liu 2007; Frangopol et al. 2008a, b; Frangopol and Okasha 2009; Kwon and Frangopol 2011; Strauss et al. 2008; Soliman et al. 2013; Okasha and Frangopol 2012).

Acknowledgments The authors greatly acknowledge the support of the US Office of Naval Research through the awards N00014-08-1-0188 and N00014-12-1-0023 (Structural Reliability Program, Director Dr. Paul E. Hess III, ONR, Code 331). Additionally, the constructive comments and suggestions offered by Alysson Mondoro (Ph.D. student) of Lehigh University are greatly acknowledged. The opinions and conclusions presented in this chapter are those of the authors and do not necessarily reflect the views of the sponsoring organization.

References

AASHTO, *AASHTO LRFD Bridge Design Specifications*, 5th edn. with Interims (American Association of State Highway and Transportation Officials, Washington, DC, 2010)

ABS, *Spectral-Based Fatigue Analysis for Floating Production, Storage and Offloading (FSPO) Installations* (American Bureau of Shipping, Houston, 2010)

U.O. Akpan, T.S. Koko, B. Ayyub, T.E. Dunbar, Risk assessment of aging ship hull structures in the presence of corrosion and fatigue. Mar. Struct. **15**(3), 211–231 (2002)

A. Anastasopoulos, D. Kourousis, S. Botten, G. Wang, Acoustic emission monitoring for detecting structural defects in vessels and offshore structures. Ships Offshore Struct. **4**(4), 363–372 (2009)

A.H.-S. Ang, W. Tang, *Probability Concepts in Engineering Planning and Design* (Wiley, New York, 1984)

A.H.-S. Ang, W. Tang, *Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering*, 2nd edn. (Wiley, New Jersey, 2007)

B. Ayyub, U.O. Akpan, G.F. DeSouza, T.S. Koto, X. Luo, *Risk-Bbased Life Cycle Management of Ship Structures, SSC-416* (Ship Structure Committee, Washington, DC, 2000)

B.M. Ayyub, I.A. Assakkaf, D.P. Kihl, M.W. Siev, Reliability-based design guidelines for fatigue of ship structures. Nav. Eng. J. **114**(2), 113–138 (2002)

I. Baran, M. Nowak, A. Jagenbrein, H. Buglacki, Acoustic emission monitoring of structural elements of a ship for detection of fatigue and corrosion damages, in *Proceedings of 30th European Conference on Acoustic Emission Testing & 7th International Conference on Acoustic Emission* (University of Granada, Granada, Spain, 2012)

J.M. Barsom, S.T. Rolfe, *Fracture and Fatigue Control in Structures: Applications of Fracture Mechanics* (ASTM, West Conshohocken, 1999)

T.F. Brady, HSV-2 Swift instrumentation and technical trials, Technical report NSWCCD-65-TR-2004/18 (Naval Surface Warfare Center, Carderock Division, West Bethesda, 2004)

BS 5400-Part 10, *Steel, Concrete, and Composite Bridges: Code of Practice for Fatigue* (British Standards Institute, London, 1980)

BS 7910, *Guide to Methods for Assessing the Acceptability of Flaws in Metallic Structures* (British Standards Institute, London, 2005)

A. Decò, D.M. Frangopol, Risk-informed optimal routing of ships considering different damage scenarios and operational conditions. Reliab. Eng. Syst. Saf. **19**, 126–140 (2013)

A. Decò, D.M. Frangopol, N.M. Okasha, Time-variant redundancy of ship structures. J. Ship Res. **55**(3), 208–219 (2011)

A. Decò, D.M. Frangopol, B. Zhu, Reliability and redundancy of ships under different operational conditions. Eng. Struct. **42**, 457–471 (2012)

DNV, *Fatigue Analysis of High Speed and Light Craft* (Det Norske Veritas Classification, Høvik, 1997)

DNV, *Type Approval Programme for Protective Coating Systems* (No. 1–602.1, Det Norske Veritas, Oslo, 1998)

DNV, *Corrosion Prevention of Tanks and Holds* (Classification Notes No. 33.1, Det Norske Veritas, Oslo, 1999)

DNV, *Fatigue Methodology of Offshore Ships* (Det Norske Veritas Classification, Høvik, 2010)

W.G. Dobson, R.F. Brodrick, J.W. Wheaton, J. Giannotti, K.A. Stambaugh, *Fatigue Considerations in View of Measured Load Spectra, SSC-315* (Ship Structure Committee, Washington, DC, 1983)

A.C. Estes, D.M. Frangopol, RELSYS, A computer program for structural system reliability analysis. Struct. Eng. Mech. **6**(8), 901–919 (1998)

Eurocode 3, *Design of Steel Structures Part 1–9. Fatigue Strength* (CEN – European Committee for Standardisation, Brussels, 2010)

Eurocode 9, *Design of Aluminium Structures Part 1–3, Additional Rules for Structures Susceptible to Fatigue* (CEN – European Committee for Standardisation, Brussels, 2009)

C. Fang, P.K. Das, Survivability and reliability of damaged ships after collision and grounding. Ocean Eng. **32**, 293–307 (2005)

J.W. Fisher, G.L. Kulak, I.F. Smith, *A Fatigue Primer for Structural Engineers* (National Steel Bridge Alliance, Chicago, 1998)

D.M. Frangopol, Life-cycle performance, management, and optimization of structural systems under uncertainty: accomplishments and challenges. Struct. Infrastruct. Eng. **7**(6), 389–413 (2011)

D.M. Frangopol, S. Kim, Service life, reliability and maintenance of civil structures, chapter 5, in *Service Life Estimation and Extension of Civil Engineering Structures*, ed. by L.S. Lee, V. Karbari (Woodhead Publishing, Cambridge, UK, 2011), pp. 145–178

D.M. Frangopol, M. Liu, Maintenance and management of civil infrastructure based on condition, safety, optimization, and life-cycle cost. Struct. Infrastruct. Eng. **3**(1), 29–41 (2007)

D.M. Frangopol, T.B, Messervey, Life-cycle cost and performance prediction: role of structural health monitoring, chapter 16, in *Frontier Technologies for Infrastructures Engineering*, *Structures and Infrastructures Book Series, vol. 4, DM Frangopol, Book Series Editor*, eds. by S.S. Chen, A.H.S. Ang (CRC Press/Balkema, Boca Raton/New York/Leiden, 2009a), pp. 361–381

D.M. Frangopol, T.B. Messervey, Maintenance principles for civil structures, chapter 89, in *Encyclopedia of Structural Health Monitoring*, ed. by C. Boller, F.-K. Chang, Y. Fujino, vol. 4 (Wiley, Chicester, 2009), pp. 1533–1562

D.M. Frangopol, N.M. Okasha, Multi-criteria optimization of life-cycle maintenance programs using advanced modeling and computational tools, chapter 1, in *Trends in Civil and Structural Computing*, ed. by B.H.V. Topping, L.F. Costa Neves, C. Barros (Saxe-Coburg Publications, Stirlingshire, 2009), pp. 1–26

D.M. Frangopol, N.M. Okasha, Life-cycle framework for maintenance, monitoring, and reliability of naval ship structures, in *Reliability and Optimization of Structural Systems*, ed. by D. Straub (CRC Press/Taylor & Francis Group, Leiden/London, 2010), pp. 69–76

D.M. Frangopol, A. Strauss, S. Kim, Use of monitoring extreme data for the performance prediction of structures: general approach. Eng. Struct. **30**(12), 3644–3653 (2008a)

D.M. Frangopol, A. Strauss, S. Kim, Bridge reliability assessment based on monitoring. J. Bridge Eng. **13**(3), 258–270 (2008b)

D.M. Frangopol, P. Bocchini, A. Decò, S. Kim, K. Kwon, N.M. Okahsa, D. Saydam, Integrated life-cycle framework for maintenance, monitoring, and reliability of naval ship structures. Nav. Eng. J. **124**(1), 89–99 (2012)

C. Guedes Soares, Y. Garbatov, Reliability of maintained, corrosion protected plate subjected to non-linear corrosion and compressive loads. Mar. Struct. **12**, 425–445 (1999)

C. Guedes Soares, Y. Garbatov, A. Zayed, G. Wang, Nonlinear corrosion model for immersed steel plates accounting for environmental factors, in *SNAME Maritime Technology Conference & Expo*, Houston, 2005

P.E. Hess, Structural health monitoring for high-speed naval ships, in *Proceedings of the 7th International Workshop on Structural Health Monitoring*, Stanford, 2007

A.W. Hussein, C. Guedes Soares, Reliability and residual strength of double hull tankers designed according to the new IACS common structural rules. Ocean Eng. **39**, 1446–1459 (2009)

IACS, Renewal criteria for side shell frames and brackets in single side skin bulk carriers not build in accordance with UR S12 Rev.1 or subsequent revisions (International Association of Classification Societies, London, 2003)

ISSC, Condition assessment of aged ships, in *International Ship and Offshore Structures Congress, Committee* V.6, vol. 2, 2006

ISSC, Condition assessment of aged ships and offshore structures. *Committee V.6, 17th International Ship and Offshore Structures Congress*, Seoul, 16–21 Aug 2009
S. Kim, D.M. Frangopol, Cost-based optimum scheduling of inspection and monitoring for fatigue-sensitive structures under uncertainty. J. Struct. Eng. **137**(11), 1319–1331 (2011a)
S. Kim, D.M. Frangopol, Cost-effective lifetime structural health monitoring based on availability. J. Struct. Eng. **137**(1), 22–33 (2011b)
S. Kim, D.M. Frangopol, Optimum inspection planning for minimizing fatigue damage detection delay of ship hull structures. Int. J.Fatigue **33**(3), 448–459 (2011c)
S. Kim, D.M. Frangopol, M. Soliman, Generalized probabilistic framework for optimum inspection and maintenance planning. J. Struct. Eng. **139**(3), 435–447 (2013)
K. Kwon, D.M. Frangopol, Bridge fatigue reliability assessment using probability density functions of equivalent stress range based on field monitoring data. Int. J. Fatigue **32**, 1221–1232 (2010)
K. Kwon, D.M. Frangopol, Bridge fatigue assessment and management using reliability-based crack growth and probability of detection models. Probabilist. Eng. Mech. **26**(3), 471–480 (2011)
K. Kwon, D.M. Frangopol, System reliability of ship hull structures under corrosion and fatigue. J. Ship Res. **95**(4), 234–251 (2012a)
K. Kwon, D.M. Frangopol, Fatigue life assessment and lifetime management of aluminum ships using life-cycle optimization. J. Ship Res. **56**(2), 91–105 (2012b)
K. Kwon, D.M. Frangopol, S. Kim, Fatigue performance assessment and service life prediction of high-speed ship structures based on probabilistic lifetime sea loads. Struct. Infrastruct. Eng. **9**(2), 102–115 (2013)
A.E. Mansour, P.H. Wirsching, G.J. White, B.M. Ayyub, Probability-based ship design: implementation of design guidelines. SSC 392, in *Ship Structures Committee*, Washington, DC, 1996
S.G. Mattson, S.M. Pandit, Statistical moments of autoregressive model residuals for damage localization. Mech. Syst. Signal Process. **20**(3), 627–645 (2006)
R.E. Melchers, Effect of temperature on the marine immersion corrosion of carbon steels. Corrosion (NACE) **58**(9), 768–782 (2002)
R.E. Melchers, Modeling of marine immersion corrosion for mild and low alloy steels – Part 1: phenomenological model. Corrosion (NACE) **59**(4), 319–334 (2003a)
R.E. Melchers, Probabilistic models for corrosion in structural reliability assessment – Part 2: models based on mechanics. J. Offshore Mech. Arctic Eng. **125**(4), 272–280 (2003b)
R.E. Melchers, Pitting corrosion of mild steel in marine immersion environment – part 1: maximum pit depth. Corrosion (NACE) **60**(9), 824–836 (2004a)
R.E. Melchers, Pitting corrosion of mild steel in marine immersion environment – part 2: variability of maximum pit depth. Corrosion (NACE) **60**(10), 937–944 (2004b)
R.E. Melchers, Mathematical modeling of the effect of water velocity on the marine immersion corrosion of mild steel coupons. Corrosion (NACE) **60**(5), 471–478 (2004c)
R.E. Melchers, Recent progress in the modeling of corrosion of structural steel immersed in seawaters. J. Infrastruct. Syst. **12**(3), 154–162 (2006)
M.A. Miner, Cumulative damage in fatigue. J. Appl. Mech **12**(3), 159–164 (1945)
R. O'Rourke, *Naval Littoral Combat Ship (LCS) Program: Background and Issues for Congress* (CRS Report for Congress, Congressional Research Service, Washington, DC, 2012)
N.M. Okasha, D.M. Frangopol, Novel approach for multi-criteria optimization of life-cycle preventive and essential maintenance of deteriorating structures. J. Struct. Eng. **136**(80), 1009–1022 (2010a)
N.M. Okasha, D.M. Frangopol, Efficient method based on optimization and simulation for the probabilistic strength computation of the ship hull. J. Ship Res. **54**(4), 244–256 (2010b)
N.M. Okasha, D.M. Frangopol, Integration of structural health monitoring in a system performance based life-cycle bridge management framework. Struct. Infrastruct. Eng. **8**(11), 999–1016 (2012)
N.M. Okasha, D.M. Frangopol, A. Decò, Integration of structural health monitoring in life-cycle performance assessment of ship structures under uncertainty. Mar. Struct. **23**(3), 303–321 (2010)
N.M. Okasha, D.M. Frangopol, D. Saydam, L.W. Salvino, Reliability analysis and damage detection in high speed naval crafts based on structural health monitoring data. Struct. Health Monit. **10**(4), 361–379 (2011)

J. Paik, G. Wang, Time-dependent risk assessment of ageing ships accounting for general/pit corrosion, fatigue cracking and local dent damage, in *World Maritime Technology Conference*, San Francisco, 2003

J.K. Paik, J.M. Lee, J.S. Hwang, Y.I. Park, A time-dependent corrosion wastage model for the structures of single and double hull tankers and FSOs and FPSOs. Mar. Technol. **40**(3), 201–217 (2003b)

J.K. Paik, J.M. Lee, Y.I. Park, J.S. Hwang, C.W. Kim, Time-variant ultimate longitudinal strength of corroded bulk carriers. Mar. Struct. **16**, 567–600 (2003c)

P.C. Paris, F.A. Erdogan, Critical analysis of crack propagation laws. J. Basic Eng. (Trans. ASME) **85**(Series D), 528–534 (1963)

L.W. Salvino, M.D. Collette, Monitoring marine structures, in *Encyclopedia of Structural Health Monitoring*, ed. by C. Boller, F.-K. Chang, Y. Fujino (Wiley, Chichester, 2009), pp. 2357–2372

L.W. Salvino, T.F. Brady, Hull monitoring system development using hierarchical framework for data and information management, in *Proceedings of the 7th International Conference on Computer and IT Applications in the Marine Industries (COMPIT'08)*, Liège, 2008

D. Saydam, D.M. Frangopol, Performance assessment of damaged ship hulls. Ocean Eng. **68**, 65–76 (2013)

R.A. Sielski, Research needs in aluminum structure, in *Proceedings of the 10th International Symposium on Practical Design of Ships and Other Floating Structures* (American Bureau of Shipping, Washington, DC, 2007)

R.A. Sielski, K. Nahshon, L.W. Salvino, K.Anderson, R. Dow, The ONR ship structural reliability program, in *Proceedings of the 2012 ASNE Day*, Arlington, 2012

M. Soliman, D.M. Frangopol, S. Kim, Probabilistic optimum inspection planning of steel bridges based on multiple fatigue sensitive details. Eng. Struct. **49**, 996–1006 (2013)

M. Soliman, D.M. Frangopol, Life-cycle management of fatigue sensitive structures integrating inspection information. J. Infrastruct. Syst. (2013a). doi:10.1061/(ASCE)IS.1943-555X.0000169, 04014001 (in press)

M. Soliman, D.M. Frangopol. Reliability and remaining life assessment of fatigue critical structures: Integrating inspection and monitoring information, *Proceedings of the ASCE Structures Congress*, Pittsburgh, PA, May 2–4, 2013; in *Proceedings of the 2013 Structures Congress* "Bridge Your Passion with Your Profession," B.J. Leshko and J. McHugh, eds. ASCE, CD-ROM, 709–720 (2013b)

A. Strauss, D.M. Frangopol, S. Kim, Use of monitoring extreme data for the performance prediction of structures: Bayesian updating. Eng. Struct. **30**(12), 3654–3666 (2008)

R.A. Swartz, A.T. Zimmerman, J.P. Lynch, J. Rosario, T.F. Brady, L.W. Salvino, K.H. Law, Hybrid wireless hull monitoring system for naval combat vessels. Struct. Infrastruct. Eng. **8**(7), 621–638 (2012)

H. Tada, P.C. Paris, G.R. Irwin, *The Stress Analysis of Cracks Handbook*, 3rd edn. (The American Society of Mechanical Engineers, New York, 2000)

P. Tscheliesnig, Detection of corrosion attack on oil tankers by means of acoustic emission (AE), in *Proceedings of the 12th A-PCNDT 2006 – Asia-Pacific Conference on NDT*, Auckland, 2006

G. Wang, J. Spencer, Y. Chen, Assessment of a ship's performance in accidents. Mar. Struct. **15**, 313–333 (2002)

G. Wang, A. Lee, L. Ivanov, T. Lynch, C. Serratella, R. Basu, A statistical investigation of time-variant hull girder strength of aging ships and coating life. Mar. Struct. **21**(2–3), 240–256 (2008)

G. Wang, L. Michael, C. Serratella, S. Botten, S. Ternowchek, D. Ozevin, J. Thibault, R. Scott, Testing of acoustic emission technology to detect cracks and corrosion in the marine environment. J. Ship Res. **26**(2), 106–110 (2010)

P.H. Wirsching, Fatigue reliability for offshore structures. J. Struct. Eng. **110**(10), 2340–2356 (1984)

B. Zhu, D.M. Frangopol, Incorporation of SHM data on load effects in the reliability and redundancy assessment of ship cross-sections using Bayesian updating. Struct. Health Monit. **12**(4), 377–392 (2013a)

B. Zhu, D.M. Frangopol, Reliability assessment of ship structures using Bayesian updating. Eng. Struct. **56**, 1836–1847 (2013b)

Elasto-dynamic Damage Evaluation of Bridges

19

Zhihai Xiang and Qiuhai Lu

Contents

Abstract

This chapter introduces two elasto-dynamic damage evaluation methods for bridge structures without traffic blocking. The first method identifies damage parameters from natural frequencies and mode shapes. Since this is an ill-posed problem, the identified damage parameters from all measurements could be

Z. Xiang (✉)
Department of Engineering Mechanics, School of Aerospace Engineering, Tsinghua University, Beijing, China
e-mail: xiangzhihai@tsinghua.edu.cn

Q. Lu
Department of Aeronautical & Astronautical Engineering, School of Aerospace Engineering, Tsinghua University, Beijing, China
e-mail: luqh@tsinghua.edu.cn

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_35

unreliable due to measurement noises. A feasible solution to this problem is using only part of natural frequencies and mode shape points in damage identification. For this purpose, an algorithm based on well-posedness analysis is proposed to select the optimal combination of natural frequencies and mode shape points. Another method is called the Tap-scan damage detection method, which is featured by extracting bridge damage information from the acceleration of a passing vehicle mounted with a well-controlled tapping device. The theoretical basis and hardware setup are presented in detail. In addition, several on-site experiments are given to illustrate its application potential to real bridge structures.

The main advantage of these two methods is not requiring traffic blocking during bridge evaluation. Besides, since they extract damage information from the self-comparison of bridge properties among different sections, they do not require the reference state of the intact bridge. These two features are crucial to their practical implementations.

Introduction

The health condition of bridge structures is the basis of their maintenance and prognosis (Okasha and Frangopol 2012). Besides visual inspection, the bridge condition is usually obtained from the structural dynamic properties, which are easy to acquire in practice (Farrar et al. 2001). However, traditional bridge evaluation methods, such as forced vibration tests by truck loadings or shakers and free vibration tests (Cunha et al. 2013), etc., require traffic blocking during bridge testing. This greatly limits their application to bridges that require real-time monitoring or are located at busy transportation links. For example, among 1,855 urban bridges in Beijing City (EPS Net 2010), only very few of them are filtered out for detailed testing and rehabilitation each year. The filtering work is mainly conducted by visual inspection because it is not permitted to stop the traffic flow in such a city suffered from severe traffic jam. Under this circumstance, efficient bridge evaluation methods free of traffic blocking are highly desired to reduce the labor cost and improve the inspection quality.

How to deal with the traffic loading is a key point to the success of such a method. The general answer could be very simple: either takes advantage of it or just avoids it. But the detailed strategy needs more consideration. To address this point, this chapter will introduce two relevant methods.

The first method detects the damage from structural dynamic properties, such as natural frequencies and mode shapes, which can be obtained by the output-only modal identification methods based on ambient excitations, such as wind and traffic flow (Deraemaeker et al. 2008; Gentile and Saisi 2011; Cunha et al. 2013). This method assumes the excitation input is a zero-mean Gaussian white noise and the bridge structure acts like a signal filter. From the output of this filter, one can

extract the structural dynamic properties that are related with damage, such as the change of stiffness, mass, or damping. Then, the occurrence of damage can be inferred from the difference between the dynamic properties of current state and those of the intact state (Salawu 1997; Farrar et al. 2001; Carden and Fanning 2004; Fan and Qiao 2011). However, since natural frequencies and mode shapes are global properties of a structure, while damage is a local phenomenon, lower natural frequencies and mode shapes could not be sensitive to damage (Farrar and Jauregui 1998; Chang et al. 2003; Carden and Fanning 2004; Xiang and Zhang 2009; Fan and Qiao 2011), so that the damage information obtained from these properties could be unreliable when measurement noises were presented. To ensure the success of damage identification, it was recommended to introduce some local information (Chang et al. 2003; Carden and Fanning 2004; Xiang and Zhang 2009). For example, the method introduced in section "Damage Detection from Selected Frequencies and Mode Shape Points" of this chapter just takes the effective stiffness of a bridge subdivision as the damage parameter, which is identified by selected natural frequencies and part of mode shape points (Xiang et al. 2012). The criterion of how to select the proper natural frequencies and mode shape points is established from the well-posedness analysis of the parameter identification procedure.

Another method is inspired by the hunting behavior of woodpeckers who detect worms under the bark by tapping trunks with their beaks. This sounds like the very traditional tap test, which can detect near-surface voids, debounding, and significant cracks even in noisy environment. However, tap test results based on sound difference could be very subjective and imprecise (Chang et al. 2003). This is probably because every strike of the hammer is different and the feeling of sound is operator dependent. Therefore, people tried to improve this method based on theoretical investigations. The first theory was given by Cawley and Adams (1988), who concluded that the local impedance of structures determines the impact force between the hammer and the structure surface, so that the change of impact force is the reason for the different tapping sounds when good and flawed regions are tested. Based on this theory, damage can be identified by comparing either the duration of the impact force or the area information of the force spectrum with intact ones. Recently, Xiang et al. (2010) pointed out that the acceleration of the passing vehicle with a tapping device contains the information of the structural local impedance of a bridge. Thus, the damage can be represented by its spectrum pattern. From the sensitivity analysis, it is recommended that the most favorable frequency of tapping force should be close to bridge natural frequencies. If this frequency was higher than the frequency band of environment noises, the vehicle acceleration spectrum at the interested frequency band could be free of environmental interference. In this way, the bridge condition can be efficiently evaluated without traffic blocking. This method is called the Tap-scan damage detection method, which will be introduced in section "The Tap-Scan Damage Detection Method" of this chapter.

Damage Detection from Selected Frequencies and Mode Shape Points

The Damage Detection Method

Although any change of stiffness, mass, damping, strength, etc., can be regarded as damage, here concerns only the first two factors, which cover most damage scenarios and are easy to evaluate. Since the stiffness and the mass cannot be simultaneously identified through natural frequencies and mode shapes (Baruch 1997), one can represent the damage as the change of the effective stiffness while keeping the density and geometry unchanged.

To increase the sensitivity of dynamic properties to local damage, the bridge is subdivided into several regions and the effective stiffness of each region is identified from a few selected frequencies and a portion of mode shape points. This approach can be formulated as a least-squares estimation problem:

$$\begin{array}{ll} \text{Minimize} & \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} \\ \text{Subject to:} & \left[\boldsymbol{K}(\boldsymbol{p}) - \omega_i^2(\boldsymbol{p})\boldsymbol{M}\right]\boldsymbol{X}_i(\boldsymbol{p}) = \boldsymbol{0} \quad i = 1, 2, \ldots \end{array} \tag{1}$$

where $\boldsymbol{p}$ is the effective stiffness parameter, which determines the ith natural angular frequency ω_i and the corresponding mode shape $\boldsymbol{X}_i$; $\boldsymbol{K}$ and $\boldsymbol{M}$ are stiffness and mass matrices, respectively; the residual vector is defined as $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{S}_\omega & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_X \end{bmatrix} \begin{Bmatrix} \overline{\boldsymbol{\omega}} - \boldsymbol{\omega} \\ \overline{\boldsymbol{X}} - \boldsymbol{X} \end{Bmatrix}$, with $\boldsymbol{\omega}$ and $\overline{\boldsymbol{\omega}}$ be the calculated and measured natural frequency vectors, $\boldsymbol{X} = (\boldsymbol{X}_1^{\mathrm{T}}, \boldsymbol{X}_2^{\mathrm{T}}, \ldots, \boldsymbol{X}_N^{\mathrm{T}})^{\mathrm{T}}$ and $\overline{\boldsymbol{X}}$ be the calculated and measured vectors of all mode shapes, respectively; $\boldsymbol{S}_\omega$ and $\boldsymbol{S}_X$ are selective matrices that determine the interested natural frequencies and mode shape points, respectively.

Equation 1 can be efficiently solved by Gauss-Newton method (Haber et al. 2000) iteratively:

$$\boldsymbol{p}^k = \boldsymbol{G}\left(\boldsymbol{p}^{k-1}\right) \equiv \boldsymbol{p}^{k-1} - \left[\left(\boldsymbol{J}^{k-1}\right)^T \boldsymbol{J}^{k-1}\right]^{-1} \left(\boldsymbol{J}^{k-1}\right)^T \boldsymbol{R}^{k-1} \quad k = 1, 2, \cdots \tag{2}$$

where the Jacobian matrix is defined as

$$\boldsymbol{J}(\boldsymbol{p}) = \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{p}} = -\begin{bmatrix} \boldsymbol{S}_\omega & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_X \end{bmatrix} \begin{Bmatrix} \dfrac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{p}} \\ \dfrac{\partial \boldsymbol{X}}{\partial \boldsymbol{p}} \end{Bmatrix} \tag{3}$$

Usually, the convergence criterion of Eq. 2 can be taken as

$$\mathrm{Max}\left\{\left|\frac{p_i^k - p_i^{k-1}}{p_i^k}\right|\right\} < 1 \times 10^{-3} \quad i = 1, 2, \cdots, N_p$$

$$\frac{\left\|\boldsymbol{S}_\omega\left(\boldsymbol{\omega}^k - \boldsymbol{\omega}^{k-1}\right)\right\|_2}{\left\|\boldsymbol{S}_\omega \boldsymbol{\omega}^k\right\|_2} < 5 \times 10^{-2} \quad \text{and} \quad \frac{\left\|\mathbf{S}_X\left(\boldsymbol{X}^k - \boldsymbol{X}^{k-1}\right)\right\|_2}{\left\|\mathbf{S}_X \boldsymbol{X}^k\right\|_2} < 5 \times 10^{-2} \tag{4}$$

where $\|\cdot\|_2$ is the Euclidian norm and N_p is the total number of parameters.

The Criterion of Optimal Measurements

The selected natural frequencies and mode shape points should ensure a unique and stable solution of Eq. 2. That means the parameter identification procedure should be well posed (Engl et al. 1996).

Since $\boldsymbol{p}$ is usually within a continuous interval $\boldsymbol{D}_p$, according to *the Brouwer's fixed-point theorem* (Griffel 2002), if the mapping function $\boldsymbol{G}$ in Eq. 2 is continuous, there certainly exists at least one solution. This can be satisfied when the number of measurements is not less than N_p and the interested natural frequencies and mode shape points are very sensitive to $\boldsymbol{p}$, so that the Jacobian matrix $\boldsymbol{J}$ is continuous and the Fisher information matrix $\boldsymbol{J}^T\boldsymbol{J}$ is nonsingular.

Assuming $\boldsymbol{x}$ and $\boldsymbol{y}$ be arbitrary two parameters, one can use the mean value theorem to obtain

$$\boldsymbol{G}(\boldsymbol{x}) - \boldsymbol{G}(\boldsymbol{y}) = \frac{\partial \boldsymbol{G}(\boldsymbol{\zeta})}{\partial \boldsymbol{p}}(\boldsymbol{x} - \boldsymbol{y}) \quad \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{D}_p \tag{5}$$

where $\boldsymbol{\zeta} = \boldsymbol{y} + \eta(\boldsymbol{x} - \boldsymbol{y})$, $0 < \eta < 1$ and

$$\frac{\partial \boldsymbol{G}(\boldsymbol{p})}{\partial \boldsymbol{p}} = \boldsymbol{\Omega}(\boldsymbol{p}) \equiv \left(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}\right)^{-1} \frac{\partial\left(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}\right)}{\partial \boldsymbol{p}} \left(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}\right)^{-1} \boldsymbol{J}^{\mathrm{T}}\boldsymbol{R} - \left(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}\right)^{-1} \frac{\partial \boldsymbol{J}^{\mathrm{T}}}{\partial \boldsymbol{p}} \boldsymbol{R} \tag{6}$$

Applying the norm of infinity $\|\cdot\|_\infty$ to both sides of Eq. 5 yields

$$\|\boldsymbol{G}(\boldsymbol{x}) - \boldsymbol{G}(\boldsymbol{y})\|_\infty \le \|\boldsymbol{\Omega}(\boldsymbol{\zeta})\|_\infty \|\boldsymbol{x} - \boldsymbol{y}\|_\infty \le L_\Omega \|\boldsymbol{x} - \boldsymbol{y}\|_\infty \tag{7}$$

where

$$L_\Omega \equiv \mathrm{Max}\,\|\boldsymbol{\Omega}(\boldsymbol{p})\|_\infty \tag{8}$$

Since $\boldsymbol{G}(\boldsymbol{p}) \in \boldsymbol{D}_p$, if $L_\Omega < 1$ the iteration procedure (Eq. 2) will converge to a unique solution, according to *the contraction mapping theorem* (Griffel 2002).

To study the stability of the solution, one can suppose the measurement contains absolute noises $\boldsymbol{\varepsilon}_\omega$ and $\boldsymbol{\varepsilon}_X$ to true values $\boldsymbol{\omega}^*$ and $\boldsymbol{X}^*$, respectively,

$$\left\{\begin{array}{c}\overline{\boldsymbol{\omega}}\\ \overline{\boldsymbol{X}}\end{array}\right\}=\left\{\begin{array}{c}\boldsymbol{\omega}^*\\ \boldsymbol{X}^*\end{array}\right\}+\left\{\begin{array}{c}\boldsymbol{\varepsilon}_\omega\\ \boldsymbol{\varepsilon}_X\end{array}\right\} \tag{9}$$

These small measurement noises may result in large identification errors because Eq. 2 is a nonlinear procedure. To trace the propagation of measurement noises through the damage identification procedure, one can examine the difference between the identified parameters and the true parameters $\boldsymbol{p}^*$ at the kth iteration step:

$$\boldsymbol{h}^k=\boldsymbol{p}^k-\boldsymbol{p}^* \tag{10}$$

Substituting Eqs. 2 and 6 into Eq. 10 yields

$$\boldsymbol{h}^k-\boldsymbol{h}^{k-1}=\left[\boldsymbol{\Omega}\left(\boldsymbol{p}^{\xi_{k-1}},\boldsymbol{\varepsilon}^{\xi_{k-1}}\right)-\boldsymbol{I}\right]\boldsymbol{h}^{k-1}+\boldsymbol{A}\left(\boldsymbol{p}^{\xi_{k-1}},\boldsymbol{\varepsilon}^{\xi_{k-1}}\right) \tag{11}$$

where $\boldsymbol{I}$ is the identity matrix and

$$\boldsymbol{A}(\boldsymbol{p},\boldsymbol{\varepsilon})=\frac{\partial \boldsymbol{G}(\boldsymbol{p},\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}}\boldsymbol{\varepsilon}=-\left(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}\right)^{-1}\boldsymbol{J}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{\varepsilon} \tag{12}$$

$\boldsymbol{p}^{\xi_{k-1}}=\boldsymbol{p}^*+\gamma\left(\boldsymbol{p}^{k-1}-\boldsymbol{p}^*\right),0<\gamma<1,\ \boldsymbol{S}=\begin{bmatrix}\boldsymbol{S}_\omega & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{S}_X\end{bmatrix},\boldsymbol{\varepsilon}=(\boldsymbol{\varepsilon}_\omega^{\mathrm{T}},\ \boldsymbol{\varepsilon}_X^{\mathrm{T}})^{\mathrm{T}}$, and $\boldsymbol{\varepsilon}^{\xi_{k-1}}=\psi\boldsymbol{\varepsilon},0<\psi<1$.

Applying the norm of infinity $\|\cdot\|_\infty$ to both sides of Eq. 11 yields

$$\left\|\boldsymbol{h}^k\right\|_\infty\le L_\Omega\left\|\boldsymbol{h}^{k-1}\right\|_\infty+L_A\le\cdots\le L_\Omega^k\|\delta\boldsymbol{p}\|_\infty+\left(1+L_\Omega+\ldots L_\Omega^{k-1}\right)L_A \tag{13}$$

where

$$L_A\equiv\mathrm{Max}\|\boldsymbol{A}(\boldsymbol{p},\boldsymbol{\varepsilon})\|_\infty \tag{14}$$

If $L_\Omega<1$, one can get the estimation of damage identification error:

$$B\equiv\lim_{k\to\infty}\left\|\boldsymbol{h}^k\right\|_\infty\le\frac{1}{1-L_\Omega}L_A\approx(1+L_\Omega)L_A \tag{15}$$

The above analyses reveal that the selected natural frequencies and mode shape points should minimize the estimation of identification error B while ensure $L_\Omega<1$.

The Calculation of the Jacobian Matrix and Its Derivative

In above sections, the Jacobian matrix $\boldsymbol{J}$ and its derivative $\partial\boldsymbol{J}/\partial\boldsymbol{p}$ are crucial to identifying damage parameters and estimating identification errors.

These values can certainly be obtained by finite difference method. From Eq. (3), it is easy to calculate the ith items in $\boldsymbol{J}$ as

$$\boldsymbol{J}_i(\boldsymbol{p}) \approx -\begin{bmatrix} \boldsymbol{S}_\omega & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_X \end{bmatrix} \begin{Bmatrix} \dfrac{\Delta \boldsymbol{\omega}}{\Delta p_i} \\ \dfrac{\Delta \boldsymbol{X}}{\Delta p_i} \end{Bmatrix} \quad i = 1, 2, \cdots, N_p \tag{16}$$

where $\Delta\boldsymbol{\omega}$ and $\Delta\boldsymbol{X}$ are the difference of natural frequencies and mode shapes when the ith parameter p_i changes to $p_i + \Delta p_i$, while fixing other parameters. Generally, $\boldsymbol{J}_i$ would be more accurate when Δp_i was smaller. However, owing to numerical truncation errors, etc., extremely small Δp_i could introduce large errors into $\boldsymbol{J}_i$ (Engl et al. 1996). Since it is very difficult to find the optimal Δp_i, it is recommended to try several times to ensure the best accuracy of $\boldsymbol{J}_i$. In addition, this method requires $N_p + 1$ times direct modal analyses to obtain $\boldsymbol{J}$ and $N_p^2 + 1$ times direct modal analyses to obtain $\partial\boldsymbol{J}/\partial\boldsymbol{p}$. This could be very time-consuming for large problems.

Besides using the aforementioned finite difference method, $\boldsymbol{J}$ and $\partial\boldsymbol{J}/\partial\boldsymbol{p}$ can be directly obtained if following a modified Nelson's method (Nelson 1976) to calculate $\partial\boldsymbol{\omega}/\partial\boldsymbol{p}$, $\partial\boldsymbol{X}/\partial\boldsymbol{p}$, $\partial^2\boldsymbol{\omega}/\partial\boldsymbol{p}^2$, and $\partial^2\boldsymbol{X}/\partial\boldsymbol{p}^2$ analytically. This is much more time efficient and accurate than the finite deference method with the expense of additional programming work.

In this method, each mode shape is firstly normalized into unity:

$$\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{X}_i = 1 \quad i = 1, 2, \ldots \tag{17}$$

From this relation, it is easy to obtain

$$\boldsymbol{X}_i^{\mathrm{T}} \frac{\partial \boldsymbol{X}_i}{\partial p_j} = 0 \tag{18}$$

Then, assuming $\partial\boldsymbol{X}_i/\partial p_j$ is a linear combination of an unknown vector $\boldsymbol{V}_i$ and $\boldsymbol{X}_i$,

$$\frac{\partial \boldsymbol{X}_i}{\partial p_j} = \boldsymbol{V}_i + c_j \boldsymbol{X}_i \tag{19}$$

where c_j is an unknown coefficient.

Substituting Eq. 19 into Eq. 18 and noticing Eq. 17, it is easy to obtain the relation between $\boldsymbol{V}_i$ and c_j as

$$c_j = -\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{V}_i \tag{20}$$

The partial derivative of the equation $(\boldsymbol{K} - \omega_i^2 \boldsymbol{M})\boldsymbol{X}_i = \boldsymbol{0}$ in Eq. 1 over p_j is

$$\left(\boldsymbol{K} - \omega_i^2 \boldsymbol{M}\right) \frac{\partial \boldsymbol{X}_i}{\partial p_j} = -\left(\frac{\partial \boldsymbol{K}}{\partial p_j} - 2\omega_i \frac{\partial \omega_i}{\partial p_j} \boldsymbol{M} - \omega_i^2 \frac{\partial \boldsymbol{M}}{\partial p_j}\right) \boldsymbol{X}_i \tag{21}$$

Substituting Eq. 19 into Eq. 21 and noticing $(\boldsymbol{K} - \omega_i^2 \boldsymbol{M})\boldsymbol{X}_i = \boldsymbol{0}$ yield

$$\left(\boldsymbol{K} - \omega_i^2 \boldsymbol{M}\right) \boldsymbol{V}_i = -\left(\frac{\partial \boldsymbol{K}}{\partial p_j} - 2\omega_i \frac{\partial \omega_i}{\partial p_j}\boldsymbol{M} - \omega_i^2 \frac{\partial \boldsymbol{M}}{\partial p_j}\right)\boldsymbol{X}_i \tag{22}$$

If left multiplying $\boldsymbol{X}_i^{\mathrm{T}}$ on both sides of Eq. 22 and using the symmetric properties of $\boldsymbol{K}$ and $\boldsymbol{M}$, one can obtain

$$\frac{\partial \omega_i}{\partial p_j} = \frac{\boldsymbol{X}_i^{\mathrm{T}}\left(\frac{\partial \boldsymbol{K}}{\partial p_j} - \omega_i^2 \frac{\partial \boldsymbol{M}}{\partial p_j}\right)\boldsymbol{X}_i}{2\omega_i \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{M} \boldsymbol{X}_i} \tag{23}$$

Then, $\boldsymbol{V}_i$ can be solved by substituting Eq. 23 into Eq. 22. Together with Eq. 20, one can obtain $\partial \boldsymbol{X}_i / \partial p_j$ from Eq. 19.

Similarly, to calculate $\partial^2 \boldsymbol{\omega}/\partial \boldsymbol{p}^2$ and $\partial^2 \boldsymbol{X}/\partial \boldsymbol{p}^2$, one can differentiate Eq. 18 with respect to the kth parameter p_k to obtain

$$\boldsymbol{X}_i^{\mathrm{T}} \frac{\partial^2 \boldsymbol{X}_i}{\partial p_j \partial p_k} + \frac{\partial \boldsymbol{X}_i^{\mathrm{T}}}{\partial p_k}\frac{\partial \boldsymbol{X}_i}{\partial p_j} = \boldsymbol{0} \tag{24}$$

And assume

$$\frac{\partial^2 \boldsymbol{X}_i}{\partial p_j \partial p_k} = \overline{\boldsymbol{V}}_i + c_{jk}\boldsymbol{X}_i \tag{25}$$

where $\overline{\boldsymbol{V}}_i$ is an unknown vector and c_{jk} is an unknown coefficient, which can be obtained by substituting Eq. 25 into Eq. 24 as

$$c_{jk} = -\boldsymbol{X}_i^{\mathrm{T}} \overline{\boldsymbol{V}}_i - \frac{\partial \boldsymbol{X}_i^{\mathrm{T}}}{\partial p_k}\frac{\partial \boldsymbol{X}_i}{\partial p_j} \tag{26}$$

Differentiating Eq. 21 with respect to p_k obtains

$$\begin{aligned}\left(\boldsymbol{K} - \omega_i^2 \boldsymbol{M}\right)\frac{\partial^2 \boldsymbol{X}_i}{\partial p_j \partial p_k} = -\Bigg[&\frac{\partial^2 \boldsymbol{K}}{\partial p_j \partial p_k} - 2\left(\frac{\partial \omega_i}{\partial p_j}\frac{\partial \omega_i}{\partial p_k} + \omega_i \frac{\partial^2 \omega_i}{\partial p_j \partial p_k}\right)\boldsymbol{M} \\ &-2\omega_i\left(\frac{\partial \omega_i}{\partial p_j}\frac{\partial \boldsymbol{M}}{\partial p_k} + \frac{\partial \omega_i}{\partial p_k}\frac{\partial \boldsymbol{M}}{\partial p_j}\right) - \omega_i^2 \frac{\partial^2 \boldsymbol{M}}{\partial p_j \partial p_k}\Bigg]\boldsymbol{X}_i \\ &-\left(\frac{\partial \boldsymbol{K}}{\partial p_j} - 2\omega_i \frac{\partial \omega_i}{\partial p_j}\boldsymbol{M} - \omega_i^2 \frac{\partial \boldsymbol{M}}{\partial p_j}\right)\frac{\partial \boldsymbol{X}_i}{\partial p_k} \\ &-\left(\frac{\partial \boldsymbol{K}}{\partial p_k} - 2\omega_i \frac{\partial \omega_i}{\partial p_k}\boldsymbol{M} - \omega_i^2 \frac{\partial \boldsymbol{M}}{\partial p_k}\right)\frac{\partial \boldsymbol{X}_i}{\partial p_j}\end{aligned} \tag{27}$$

Substituting Eq. 25 into Eq. 27 and noticing $(\boldsymbol{K} - \omega_i^2 \boldsymbol{M})\boldsymbol{X}_i = \boldsymbol{0}$ obtain

$$\begin{aligned}
&\left(\boldsymbol{K} - \omega_i^2 \boldsymbol{M}\right)\overline{\boldsymbol{V}}_i = \\
&-\left[\frac{\partial^2 \boldsymbol{K}}{\partial p_j \partial p_k} - 2\left(\frac{\partial \omega_i}{\partial p_j}\frac{\partial \omega_i}{\partial p_k} + \omega_i \frac{\partial^2 \omega_i}{\partial p_j \partial p_k}\right)\boldsymbol{M} - 2\omega_i\left(\frac{\partial \omega_i}{\partial p_j}\frac{\partial \boldsymbol{M}}{\partial p_k} + \frac{\partial \omega_i}{\partial p_k}\frac{\partial \boldsymbol{M}}{\partial p_j}\right) - \omega_i^2 \frac{\partial^2 \boldsymbol{M}}{\partial p_j \partial p_k}\right]\boldsymbol{X}_i \\
&-\left(\frac{\partial \boldsymbol{K}}{\partial p_j} - 2\omega_i \frac{\partial \omega_i}{\partial p_j}\boldsymbol{M} - \omega_i^2 \frac{\partial \boldsymbol{M}}{\partial p_j}\right)\frac{\partial \boldsymbol{X}_i}{\partial p_k} - \left(\frac{\partial \boldsymbol{K}}{\partial p_k} - 2\omega_i \frac{\partial \omega_i}{\partial p_k}\boldsymbol{M} - \omega_i^2 \frac{\partial \boldsymbol{M}}{\partial p_k}\right)\frac{\partial \boldsymbol{X}_i}{\partial p_j}
\end{aligned} \tag{28}$$

If left multiplying $\boldsymbol{X}_i^{\mathrm{T}}$ on both sides of Eq. 28 and using the symmetric properties of $\boldsymbol{K}$ and $\boldsymbol{M}$, one can obtain

$$\begin{aligned}
&\frac{\partial^2 \omega}{\partial p_j \partial p_k} = \\
&\boldsymbol{X}_i^{\mathrm{T}}\left\{\left[\frac{\partial^2 \boldsymbol{K}}{\partial p_j \partial p_k} - 2\frac{\partial \omega_i}{\partial p_j}\frac{\partial \omega_i}{\partial p_k}\boldsymbol{M} - 2\omega_i\left(\frac{\partial \omega_i}{\partial p_j}\frac{\partial \boldsymbol{M}}{\partial p_k} + \frac{\partial \omega_i}{\partial p_k}\frac{\partial \boldsymbol{M}}{\partial p_j}\right) - \omega_i^2 \frac{\partial^2 \boldsymbol{M}}{\partial p_j \partial p_k}\right]\boldsymbol{X}_i\right. \\
&\left.+\left(\frac{\partial \boldsymbol{K}}{\partial p_j} - 2\omega_i \frac{\partial \omega_i}{\partial p_j}\boldsymbol{M} - \omega_i^2 \frac{\partial \boldsymbol{M}}{\partial p_j}\right)\frac{\partial \boldsymbol{X}_i}{\partial p_k} + \left(\frac{\partial \boldsymbol{K}}{\partial p_k} - 2\omega_i \frac{\partial \omega_i}{\partial p_k}\boldsymbol{M} - \omega_i^2 \frac{\partial \boldsymbol{M}}{\partial p_k}\right)\frac{\partial \boldsymbol{X}_i}{\partial p_j}\right\} / 2\omega_i \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{M} \boldsymbol{X}_i
\end{aligned} \tag{29}$$

Then, $\overline{\boldsymbol{V}}_i$ can be solved by substituting Eq. 29 into Eq. 28. Together with Eq. 26, one can obtain $\frac{\partial^2 \boldsymbol{X}_i}{\partial p_j \partial p_k}$ from Eq. 25.

The Integrated Algorithm

With the criterion presented in section "The Criterion of Optimal Measurements," it is possible to find optimal measurement set $\boldsymbol{S}$ by certain combinatorial optimization method, if $\boldsymbol{\varepsilon}$, L_Ω, and L_A are known. However, the values of L_Ω and L_A depend on parameter $\boldsymbol{p}$, which is unknown before parameter identification. This sounds like a *chicken or the egg* causality dilemma. In addition, it is also difficult to get a precise estimation of measurement noises. To solve this problem, an iterative algorithm (see Fig. 1) that integrates measurement set selection and parameter identification is needed. In this way, all parameters are adaptively updated with the minimum a priori information, i.e., only the guess of the initial parameter $\boldsymbol{p}^0$ and measurement noise $\boldsymbol{\varepsilon}^0$.

Many combinatorial optimization methods, such as the genetic algorithm, etc., can be used to find the optimal measurement set $\boldsymbol{S}$ from possible candidates (Padula and Kincaid 1999). Since this time-consuming process is iteratively conducted in the integrated algorithm (see Fig. 1), a trade-off should be made between the solution quality and the computational expense. A possible balanced heuristic is presented in Fig. 2 that consists of:

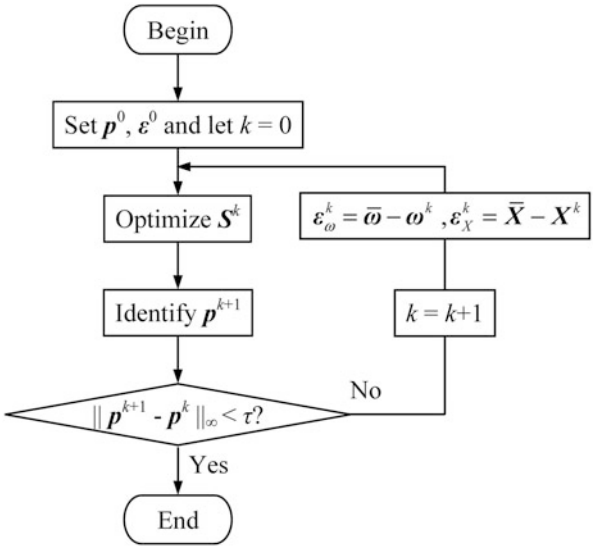


Fig. 1 The flow chart of the integrated algorithm

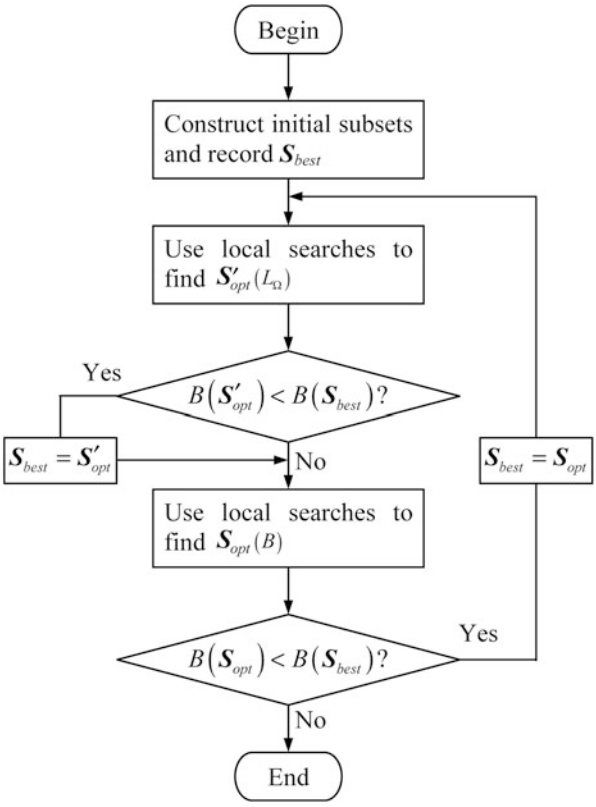


Fig. 2 The flow chart of the heuristic to optimize $\boldsymbol{S}$

- A simple process to construct the initial solution that tries to evenly distribute all possible natural frequencies and mode shape points into several feasible subsets that contain at least N_p measurements.
- An *intensification* mechanism that tries to find the optimal measurement set $\boldsymbol{S}_{opt}(B)$, which meets the primary objective of minimizing the identification error B while keeping $L_\Omega < 1$ in a local solution region.
- A *diversification* mechanism that tries to find the optimal measurement set $\boldsymbol{S}'_{\text{opt}}(L_\Omega)$, which meets the secondary objective of minimizing L_Ω. Since this objective is slightly different from the primary objective, the optimization process can temporarily accept worse solutions. However, this gives a chance to jump out from local optimum to find better solutions in adjacent new local solution regions.

Both *intensification* and *diversification* operations iteratively use *Exchange*, *Move1*, *Exchange*, and *Move2* elemental local searches to explore solution regions (Xiang et al. 2012) until there is no improvement of solutions. These elemental local searches are defined as:

- *Exchange* tries to exchange a pair of points between measurement sets $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. If either $B(\boldsymbol{S}_1)$ or $B(\boldsymbol{S}_2)$ (or $L_\Omega(\boldsymbol{S}_1)$ or $L_\Omega(\boldsymbol{S}_2)$) is reduced and the corresponding $L_\Omega < 1$, keep this move and stop; otherwise, recover $\boldsymbol{S}_1$, $\boldsymbol{S}_2$ and continue this process until each pair of points between $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ has been tested. This operation should apply to each pair of measurement sets $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ from all candidate measurement sets. During this process, update the best measurement set $\boldsymbol{S}_{\text{opt}}(B)$ (or $\boldsymbol{S}'_{\text{opt}}(L_\Omega)$).
- *Move1* tries to move one point from a measurement set $\boldsymbol{S}_1$ into another measurement set $\boldsymbol{S}_2$. If either $B(\boldsymbol{S}_1)$ or $B(\boldsymbol{S}_2)$ (or $L_\Omega(\boldsymbol{S}_1)$ or $L_\Omega(\boldsymbol{S}_2)$) is reduced and the corresponding $L_\Omega < 1$, keep this move and stop; otherwise, recover $\boldsymbol{S}_1$, $\boldsymbol{S}_2$ and continue this process until every point in $\boldsymbol{S}_1$ has been tested. This operation should apply to each pair of measurement sets $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ from all candidate measurement sets. During this process, update the best measurement set $\boldsymbol{S}_{\text{opt}}(B)$ (or $\boldsymbol{S}'_{\text{opt}}(L_\Omega)$).
- *Move2* tries to move two points from a measurement set $\boldsymbol{S}_1$ into another measurement set $\boldsymbol{S}_2$. If either $B(\boldsymbol{S}_1)$ or $B(\boldsymbol{S}_2)$ (or $L_\Omega(\boldsymbol{S}_1)$ or $L_\Omega(\boldsymbol{S}_2)$) is reduced and the corresponding $L_\Omega < 1$, keep this move and stop; otherwise, recover $\boldsymbol{S}_1$, $\boldsymbol{S}_2$ and continue this process until every two points in $\boldsymbol{S}_1$ has been tested. This operation should apply to each pair of measurement sets $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ from all candidate measurement sets. During this process, update the best measurement set $\boldsymbol{S}_{\text{opt}}(B)$ (or $\boldsymbol{S}'_{\text{opt}}(L_\Omega)$).

Examples

Simply Supported Beam

The main idea of using selected frequencies and mode shape points to identify damage parameters can be well demonstrated by a simple example depicted in Fig. 3. This is a simply supported beam with a uniform rectangular cross section of

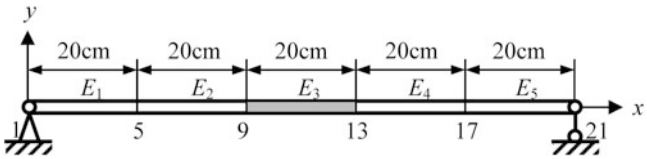


Fig. 3 A simply supported beam with damage

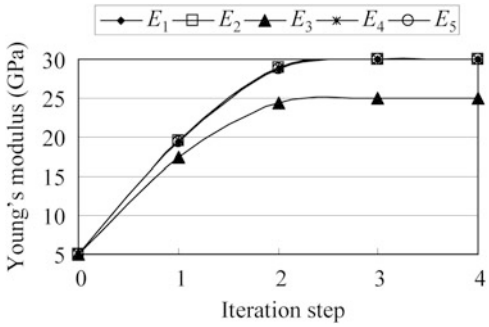


Fig. 4 Parameter identification without noise

20×10 cm and the density of 2,400 kg/m^3. To introduce an artificial damage, the Young's modulus of the center region is set as $E_3 = 25$GPa, which is smaller than those in other regions with $E_1 = E_2 = E_4 = E_5 = 30$GPa. In the following simulation, the Finite Element Method (FEM) is used to calculate all dynamic properties, which are regarded as the "true" values of measurements. For this purpose, 20 plane beam elements are uniformly distributed along the beam length, so that there are 21 nodes in total.

Firstly, use the noise free first four natural frequencies and the first mode shape "measured" at all 21 points to identify the Young's modulus of each region. As Fig. 4 shows, using the Gauss-Newton iteration method given in Eq. 2, the Young's moduli of all five regions can be successfully identified in a few steps even the initial values are far away from the true values. Since it is not easy to place dense sensors on structures in practice, another parameter identification is conducted by using only the noise free first four natural frequencies and the first mode shape "measured" at points of 1, 5, 9, 13, 17, and 21 (see Fig. 3). Not surprising, all parameters can be accurately identified through almost the same iteration procedure as that depicted in Fig. 4.

However, if add only 1 % noise into the first four natural frequencies and 10 % noise, which is evaluated as the mean square root error between the true and contaminated mode shapes, into the first mode shape at all 21 points, the parameter identification cannot converge, even if the initial parameters are taken as true values (see Fig. 5). This demonstrates that natural frequencies and mode shapes just have weak sensitivity to local damage. Based on these measurements, the identified damage parameters could be very unreliable due to the unavoidable measurement noises in practice.

If still use the measurements of the first four natural frequencies with 1 % noise and the first mode shape at all 21 points with 10 % noise, and set the initial

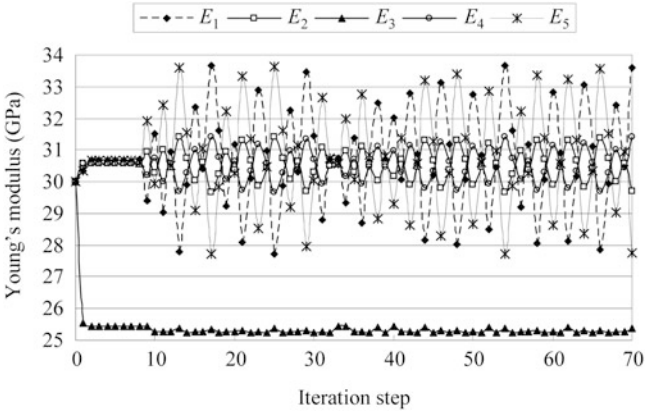


Fig. 5 Parameter identification with noise

parameters as true values, the algorithm proposed in Fig. 2 (without measurement noise updating since noises are known) cannot find a good measurement set to ensure a converged parameter identification procedure. The failure continues even if the first four mode shapes at 21 points with 10 % noises are used or even the noises in natural frequencies are reduced to 0.5 %. This is probably because natural frequencies are very insensitive to local damage. Unfortunately, it is difficult to introduce local information into natural frequencies to increase the sensitivity. However, natural frequencies are still required to work together with mode shapes to ensure the unique identification of stiffness. This can be easily understood from the equation $(\boldsymbol{K} - \omega_i^2 \boldsymbol{M})\boldsymbol{X}_i = \boldsymbol{0}$ in Eq. 1. Therefore, since natural frequencies can be measured very precisely in practice, it is recommended to ignore their measurement noises in damage parameter identification.

Table 1 compares the damage parameter identification results from different measurements. Besides the first four clean natural frequencies, the candidate measurements of case 1 through case 4 have additional 21 polluted mode shape points with 10 % noise, and case 5 contains all the first four polluted mode shapes. To check the validity of the criterion given in section "The Criterion of Optimal Measurements," the initial values of all parameters are all set as true values. Two scenarios are tested for each case in Table 1. One scenario just uses all candidate measurements and another scenario uses selected measurements obtained by the algorithm given in Fig. 2, which are listed in the 2nd through the 6th rows in Table 1. From the comparison of the estimated identification error B, the index L_Ω and identified parameters of each scenario, one can observe that:

- If feasible measurements with $L_\Omega < 1$ could be found, the identified parameters would have errors very close to the estimation B (see case 3, case 4, and case 5). However, feasible measurements with $L_\Omega < 1$ could not always be found (see case 1 and case 2).

Table 1 Damage identification with polluted measurements and known noises

		Case 1	Case 2	Case 3	Case 4	Case 5
Selected frequency		/	/	1, 2, 4	1, 2, 3	1, 2, 4
Selected mode 1 points		/	/	/	/	3, 7, 13
Selected mode 2 points		/	/	/	/	3, 13, 17, 19
Selected mode 3 points		/	/	3, 5, 7, 11, 15, 17, 19	/	9, 11, 15, 17, 19
Selected mode 4 points		/	/	/	3, 5, 9, 13, 15, 19	3, 5, 7, 9, 13, 17, 19
B (GPa)	S	/	/	0.00480	0.85833	0.02514
	A	1.71250	50.69158	0.02565	2.97000	3.02505
L_Ω	S	/	/	0.098	0.168	0.053
	A	33.634	2.272	0.470	0.343	0.317
E_1 (GPa)	S	/	/	30.00136	29.33297	30.02349[a]
	A	/	34.33958	29.99318	28.14913[a]	30.60870
E_2 (GPa)	S	/	/	29.99585	30.71625[a]	29.97784
	A	/	16.99820	29.99682	30.08291	27.81328
E_3 (GPa)	S	/	/	25.00436[a]	24.99584	25.00284
	A	/	30.76546	25.01743[a]	25.10786	25.01764
E_4 (GPa)	S	/	/	29.99585	29.31186	30.00484
	A	/	59.78787[a]	29.99682	30.00000	32.35985[a]
E_5 (GPa)	S	/	/	30.00136	30.71081	29.98789
	A	/	23.57690	29.99318	31.81542	29.46059

S use selected measurements, A use all measurements
[a]With the maximum error

- If $L_\Omega < 1$ is satisfied, the parameter identification procedure will converge. However, as indicated in section "The Criterion of Optimal Measurements," this is just a sufficient but not necessary condition of convergence. For example, case 2 still converges with $L_\Omega = 2.272$ when using all candidate measurements. But in this case, the identified parameters have very large errors.
- The results of case 5 are worse than those of case 4. This implies the algorithm can only find better solutions instead of the best one and using more measurements could result in worse solution when the measurements contain noises. The second point is further confirmed with the observation that if use all candidate measurements, the identification procedure either diverges or converges to the results with larger errors than those from selected measurements. This is mainly because more contaminated measurements would bring more errors into the final results.

Table 2 Damage identification with polluted measurements and unknown noises

	All measurements	No noise updating	Update $\boldsymbol{\varepsilon}_\omega$ and $\boldsymbol{\varepsilon}_X$	Update $\boldsymbol{\varepsilon}_\omega$
Selected frequency	All	/	/	1, 2
Selected mode 1 points	All	/	/	3, 5, 13, 15
Selected mode 2 points	All	/	/	5, 7, 15, 17, 19
Selected mode 3 points	All	/	/	All
Selected mode 4 points	All	/	/	All
E_1 (GPa)	31.22372	/	/	30.01390
E_2 (GPa)	28.37269	/	/	32.73441[a]
E_3 (GPa)	25.52086	/	/	25.59989
E_4 (GPa)	33.01066[a]	/	/	28.67603
E_5 (GPa)	30.05202	/	/	31.07655

[a]With the maximum error

- As case 5 shows, the selected measurements could contain points from different mode shapes. In this manner, the local information is introduced to increase the sensitivity.

In practice, it is very difficult to have accurate estimations of measurement noises. Therefore, it is necessary to update them with the integrated algorithm proposed in Fig. 2.

In the following tests, the initial values of all Young's moduli are set to be 20GPa. The candidate measurements are the first four natural frequencies with 1 % positive noises and the first four mode shapes at 21 points with 10 % noise. Since natural frequencies are very insensitive to local damage, it just assumes that they are noise free at the first iteration step of parameter identification procedure, i.e., $\boldsymbol{\varepsilon}_\omega^0 = \mathbf{0}$, so that good measurement sets can be found. In addition, the initial estimations of the measurement noises in mode shapes can be obtained from $\boldsymbol{\varepsilon}_X^0 = \overline{\boldsymbol{X}} - \tilde{\boldsymbol{X}}^*$, where $\overline{\boldsymbol{X}}$ is the "measured" mode shapes and $\tilde{\boldsymbol{X}}^*$ is the guess of true mode shapes. Since there are 10 % noises in "measured" mode shapes, $\tilde{\boldsymbol{X}}^*$ can be estimated by solving $\overline{\boldsymbol{X}} = \frac{\tilde{\boldsymbol{X}}^*+0.1}{\left\|\tilde{\boldsymbol{X}}^*+0.1\right\|_2}$ with the Gauss-Newton method similar to Eq. 2.

As Table 2 shows, if use all candidate measurements, the resultant parameters will have very large identification errors compared with the results in Table 1. The parameter identification procedure does not converge when there is no measurement noises updating or updates the measurement noises both in natural frequencies and mode shapes. The best identification results are obtained when updates only the measurement noises in mode shapes. All these tests illustrate that natural frequencies are very insensitive to local damage so that the damage parameter identification result is very sensitive to the measurement noise in natural frequencies.

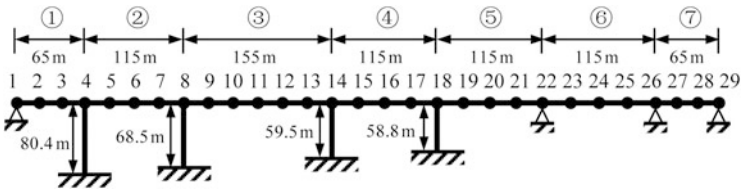


Fig. 6 The Shi Kong Shan Bridge

Shi Kong Shan Bridge

Besides the above academic example, the integrated algorithm was implemented to the right line of the Shi Kong Shan Bridge (see Fig. 6), a prestressed continuum bridge located in Fujian province, China. Because the mode shapes of spans ① through ⑤ and spans ⑥ and ⑦ were separately measured, only the effective Young's moduli of spans ① through ⑤ (E_1 to E_5) were identified based on the natural frequencies and mode shapes of the first two vertical modes. The FEM model of 226 beam elements was used in the identification. At the first step, all Young's moduli were set as 32GPa and it assumed that the natural frequencies were noise free and the mode shapes had 10 % noises, since exact measurement noises were not clear. These uncertain data were updated during the iterations of the integrated algorithm (see Fig. 1).

The identified parameters are $E_1 = 31.9$GPa, $E_2 = 25.5$GPa, $E_3 = 30.0$GPa, $E_4 = 26.0$GPa, and $E_5 = 35.9$GPa with the selected first two frequencies, 2, 3, 5, 10, 12, 19 points of mode 1 and 3, 9, 12, 13, 15, 20 points of mode 2. The apparent small stiffness coincides with the on-site visual observation, where some cracks were found in span ② and span ④. In addition, the identification could not converge if used all measurements.

The Tap-Scan Damage Detection Method

Theory

Although the aforementioned damage detection method tries hard to remove the influence of measurement noises, its success is still heavily dependent on the quality of modal testing. If traffic blocking was prohibited, frequencies and mode shapes could only be obtained by output-only methods. These methods assume that input environmental excitations are stationary white signals. However, this basic assumption is not easy to be satisfied in practice, which brings much trouble in preconditioning signals to remove the environmental influences (Deraemaeker et al. 2008; Gentile and Saisi 2011; Cunha et al. 2013). In addition, modal testing methods usually require the dense placement of transducers on the bridge. This could be time-consuming and expensive.

If 24 h' monitoring is not necessary, the Tap-scan damage detection method (Xiang et al. 2010) can be used to efficiently filter out damaged bridges

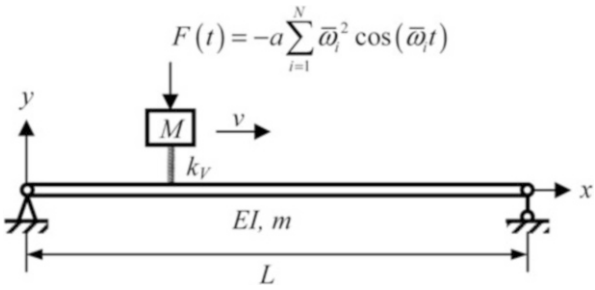


Fig. 7 The passing vehicle subject to tapping force

from large amount of candidates. The theoretic basis of this method can be illustrated by a vehicle-bridge interaction model shown in Fig. 7. In this model, a vehicle of mass M passes through the bridge at speed v with a tapping force F. The bridge is modeled as a simply supported beam with bending stiffness EI and mass per unit length m. The kinematic equations of the vehicle and the bridge are

$$M\ddot{y}_V + k\left(y_V - y_B|_{x=vt}\right) = F(t) \tag{30}$$

$$m\ddot{y}_B + EI\frac{\partial^4 y_B}{\partial x^4} = f(t)\delta(x - vt) \tag{31}$$

where head dot denotes the partial derivative over time t; y_V and y_B are the vehicle and bridge displacements measured from the static equilibrium position, respectively. The stiffness k is an effective value contributed by the series connection of vehicle suspension system of stiffness k_V and the bridge local stiffness of k_B at a position x. It could simply take $k = k_V k_B/(k_V + k_B)$, which varies along the bridge length because $k_B(x)$ can be evaluated as the inverse of bridge deflection subjected to a unit load at position x.

The function f in Eq. 31 is the vehicle-bridge interaction force:

$$f(t)\delta(x - vt) = k\left(y_V - y_B|_{x=vt}\right) - Mg \tag{32}$$

where δ is the Dirac symbol that defines the contact position $x = vt$. Since the local impedance $Z(x) = f(t)\delta(x - vt)/y_B(t)$ contains the information of stiffness, mass, and damping, the abrupt change of Z can indicate the occurrence of damage. From Eq. 30 through Eq. 32, one can easily find out that the vehicle acceleration (VA) contains the damage information:

$$\ddot{y}_V(x) = -\frac{y_B(x/v)}{M}Z(x) + \frac{F(x/v) - Mg}{M} \tag{33}$$

In addition, in the spectrum of VA the contributions of the last term of Eq. 33 are constant static values that have nothing to do with damage. This means the change of the VA spectrum is only related with Z. This is good news for damage identification.

The analytical solution of VA can be obtained by applying the modal superposition method over Eqs. 30 and 31 (Xiang et al. 2010):

$$\begin{aligned}\ddot{y}_V = \omega_V \sum_{j=1}^{\infty}\left[\sum_{i=1}^{N}\Delta_W\left(R_j^i\right)^2\ddot{P}_j^i(t) + \frac{\Delta_{Bj}}{1-S_j^2}\ddot{Q}_j(t)\right] \\ + \frac{2a}{M}\sum_{i=1}^{N}\overline{\omega}_i^2\left[\omega_V^2\cos\left(\omega_V t\right) - \frac{\overline{\omega}_i^2\cos\left(\overline{\omega}_i t\right)}{\overline{\omega}_i^2-\omega_V^2}\right]\end{aligned} \tag{34}$$

where $\omega_V = \sqrt{\frac{k}{M}}$, $\Delta_W = \frac{a}{mL}$, $\Delta_{Bj} = -\frac{2MgL^3}{(j\pi)^4 EI}$, $S_j = \frac{j\pi v}{L\omega_{Bj}}$, $R_j^i = \frac{\overline{\omega}_i}{\omega_{Bj}}$,

$$\begin{aligned}\ddot{Q}_j(t) = \frac{2\omega_V T_j^2\left[\cos\left(\omega_V t\right) - \cos\left(2T_j t\right)\right]}{4T_j^2-\omega_V^2} \\ -S_j\omega_V\left\{\left(T_j-\omega_{Bj}\right)^2\frac{\cos\left[\left(T_j-\omega_{Bj}\right)t\right]}{2\left[\left(T_j-\omega_{Bj}\right)^2-\omega_V^2\right]} - \left(T_j+\omega_{Bj}\right)^2\frac{\cos\left[\left(T_j+\omega_{Bj}\right)t\right]}{2\left[\left(T_j+\omega_{Bj}\right)^2-\omega_V^2\right]}\right. \\ \left. - \frac{2T_j\omega_{Bj}{\omega_V}^2\cos\left(\omega_V t\right)}{\left[\left(T_j+\omega_{Bj}\right)^2-\omega_V^2\right]\left[\left(T_j-\omega_{Bj}\right)^2-\omega_V^2\right]}\right\}\end{aligned} \tag{35}$$

$$\begin{aligned}\ddot{P}_j^i(t) = \frac{1}{2\left[\left(R_j^i+S_j\right)^2-1\right]}\left\{\frac{T_j\omega_V^2\cos\left(\omega_V t\right)}{T_j^2-\left(\overline{\omega}_i+T_j-\omega_V\right)^2} - \frac{T_j\omega_V^2\cos\left(\omega_V t\right)}{T_j^2-\left(\overline{\omega}_i+T_j+\omega_V\right)^2}\right. \\ \left. - \omega_V\left(2T_j+\overline{\omega}_i\right)^2\frac{\cos\left[\left(2T_j+\overline{\omega}_i\right)t\right]}{\left(2T_j+\overline{\omega}_i\right)^2-\omega_V^2} + \omega_V\overline{\omega}_i^2\frac{\cos\left(\overline{\omega}_i t\right)}{\overline{\omega}_i^2-\omega_V^2}\right\} \\ -\frac{1}{2\left[\left(R_j^i-S_j\right)^2-1\right]}\left\{\frac{T_j\omega_V^2\cos\left(\omega_V t\right)}{T_j^2-\left(\overline{\omega}_i-T_j-\omega_V\right)^2} - \frac{T_j\omega_V^2\cos\left(\omega_V t\right)}{T_j^2-\left(\overline{\omega}_i-T_j+\omega_V\right)^2}\right. \\ \left. + \omega_V\left(2T_j-\overline{\omega}_i\right)^2\frac{\cos\left[\left(2T_j-\overline{\omega}_i\right)t\right]}{\left(T_j-\overline{\omega}_i\right)^2-\omega_V^2} - \omega_V\overline{\omega}_i^2\frac{\cos\left(\overline{\omega}_i t\right)}{\overline{\omega}_i^2-\omega_V^2}\right\} \\ +\frac{S_j}{2\left[\left(R_j^i-1\right)^2-S_j^2\right]}\left\{\frac{T_j\omega_V^2\cos\left(\omega_V t\right)}{T_j^2-\left(\omega_{Bj}-\omega_V\right)^2} - \frac{T_j\omega_V^2\cos\left(\omega_V t\right)}{T_j^2-\left(\omega_{Bj}+\omega_V\right)^2}\right. \\ \left. -\omega_V\left(T_j+\omega_{Bj}\right)^2\frac{\cos\left[\left(T_j+\omega_{Bj}\right)t\right]}{\left(T_j+\omega_{Bj}\right)^2-\omega_V^2} + \omega_V\left(-T_j+\omega_{Bj}\right)^2\frac{\cos\left[\left(-T_j+\omega_{Bj}\right)t\right]}{\left(T_j-\omega_{Bj}\right)^2-\omega_V^2}\right\} \\ -\frac{S_j}{2\left[\left(R_j^i+1\right)^2-S_j^2\right]}\left\{\frac{T_j\omega_V^2\cos\left(\omega_V t\right)}{T_j^2-\left(\omega_{Bj}-\omega_V\right)^2} - \frac{T_j\omega_V^2\cos\left(\omega_V t\right)}{T_j^2-\left(\omega_{Bj}+\omega_V\right)^2}\right. \\ \left. -\omega_V\left(T_j+\omega_{Bj}\right)^2\frac{\cos\left[\left(T_j+\omega_{Bj}\right)t\right]}{\left(T_j+\omega_{Bj}\right)^2-\omega_V^2} + \omega_V\left(-T_j+\omega_{Bj}\right)^2\frac{\cos\left[\left(-T_j+\omega_{Bj}\right)t\right]}{\left(T_j-\omega_{Bj}\right)^2-\omega_V^2}\right\}\end{aligned} \tag{36}$$

and $T_j = j\pi v/L$.

If transform Eq. 34 into frequency domain, one can find seven peaks at ω_V, $\overline{\omega}_i$, $\overline{\omega}_i + 2T_j$, $\overline{\omega}_i - 2T_j$, $\omega_{Bj} + T_j$, $\omega_{Bj} - T_j$, and $2T_j$, respectively, in the spectrum. Usually, the vehicle speed v is not very high during bridge inspection. Therefore, T_j could be neglected compared with other frequencies and consequently only the peaks at ω_V, $\overline{\omega}_i$ and ω_{Bj} are dominant in the spectrum. It notices that the Frequency of Tapping Force (FTF) $\overline{\omega}_i$ dose not contain damage information and the bridge lower natural frequency ω_{Bj} is not sensitive to damage (Farrar and Jauregui 1998; Chang et al. 2003; Carden and Fanning 2004; Xiang and Zhang 2009; Fan and Qiao 2011), but ω_V has direct connection with the bridge stiffness k_B, which could have a great change with the presence of damage. In addition, from Eq. 34 through Eq. 36 one can observe that the change of ω_V has great impact on the amplitudes of peaks at ω_V, $\overline{\omega}_i$ and ω_{Bj} in the spectrum of VA. This means that the spectrum contour of the VA can be used as a damage feature and accordingly, one can define the Tap-scan damage detection method that consists of following steps:

1. Uniformly subdivide the VA in time domain into n segments.
2. Transform each segment into frequency domain to obtain its spectrum, which is expected to contain the damage information.
3. Record the contour of each spectrum at the interested frequency interval into a damage vector $\boldsymbol{Y}_i$, $i = 1, 2, \ldots n$.
4. Plot the damage index that is the difference between each pair of damage vectors $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$. For example, the MAC value (Allemang and Brown 1982) defined as $\boldsymbol{Y}_i \cdot \boldsymbol{Y}_j / (\|\boldsymbol{Y}_i\|_2 \times \|\boldsymbol{Y}_j\|_2)$ is commonly used for comparison. Since the local impedance of a healthy bridge should smoothly distributed, the abrupt change in the MAC value indicates the location of damage or structural discontinuities.

To fine-tune the controlling parameters in the Tap-scan damage detection method, such as the FTF $\overline{\omega}_i$, the vehicle speed v, the vehicle mass M, and the stiffness of vehicle suspension system k_V, one can conduct the sensitivity analysis of VA to damage. Since only ω_V is sensitive to damage, the sensitivity of VA to damage can be evaluated as

$$\frac{\partial \ddot{y}_V}{\partial \omega_V} = \sum_{j=1}^{\infty}\left[\sum_{i=1}^{N}\Delta_W\left(R_j^i\right)^2 \ddot{P}_j^i + \frac{\Delta_{Bj}}{1-S_j^2}\ddot{Q}_j\right] + \omega_V \sum_{j=1}^{\infty}\left[\sum_{i=1}^{N}\Delta_W\left(R_j^i\right)^2 \partial\frac{\ddot{P}_j^i}{\partial\omega_V} + \frac{\Delta_{Bj}}{1-S_j^2}\frac{\partial\ddot{Q}_j}{\partial\omega_V}\right] + \frac{2a\omega_V}{M}\sum_{i=1}^{N}\overline{\omega}_i^2\left\{\frac{2\overline{\omega}_i{}^2[\cos(\omega_V t) - \cos(\overline{\omega}_i t)] - \omega_V\left(\overline{\omega}_i{}^2 - \omega_V^2\right)\sin(\omega_V t)t}{\left(\overline{\omega}_i^2 - \omega_V^2\right)^2}\right\} \tag{37}$$

where $\frac{\partial \ddot{P}_j^i}{\partial \omega_V}$ and $\frac{\partial \ddot{Q}_j}{\partial \omega_V}$ can be easily obtained from Eqs. 35 and 36.

In practice, the vehicle speed usually satisfies $S_j << 1$; ω_V is different from interested bridge frequencies; and the tapping force is prohibited to cause the vehicle resonance. Therefore, Eqs. 35 through 37 reveal that if $R_j^i \approx 1$, i.e., the

FTF is close to the bridge natural frequencies, the VA will have high sensitivity to damage. In addition, if the FTF was higher than the frequency band of environmental noises, the VA at the frequency around the FTF could be free of environmental interference.

If remove the tapping force, the model in Fig. 7 becomes the classical passing vehicle model, which has been extensively discussed by Yang et al. (Yang et al. 2004; Yang and Chang 2009a, b; Yang et al. 2013). Based on the passing vehicle model, it can easily extract bridge natural frequencies from the spectrum of the vehicle acceleration signal (Lin and Yang 2005). However, this method could not be very effective for damage identification in noisy environment, because in this case the VA could not be very sensitive to damage due to the small $\frac{\partial \ddot{y}_V}{\partial \omega_V}$ when $R_j^i = 0$ and $\overline{\omega}_i = 0$ (see Eq. 37).

If the vehicle dose not move, $v = 0$, $S_j = 0$, $\ddot{P}_j^i = 0$ and $\ddot{Q}_j = 0$, which could be regarded as a traditional tap test. In this case, one can easily find from Eq. 34 that the peaks in the spectrum of VA could be sharply located at ω_V and $\overline{\omega}_i$. Therefore, it is possible to extract the damage information from the change of spectrum contour. In addition, Eq. 37 implies that the vehicle acceleration is not very sensitive to damage if $\overline{\omega}_i << \omega_V$. Therefore, it requires $\overline{\omega}_i > \omega_V$ to getting a better performance, which could be easily satisfied in practice. Owing to the controlled tapping force with the constant amplitude and the sensitive tapping frequency, this method is more reliable than the traditional tap test (Cawley and Adams 1988). This static inspection could be used complementarily to the Tap-scan method, if one wants to recheck the detected damage.

Numerical Examples

Numerical simulations on that vehicle-bridge interaction model shown in Fig. 7 are conducted to illustrate the validity of the theoretical basis of the Tap-scan method presented in section "Theory." All simulations are solved by ABAQUS implicit finite element package with time step of 0.001 s. And 5 % damping and smooth contact between the vehicle and the bridge are adopted.

The properties of the simply supported beam are length $L = 25$ m, cross-sectional area $A = 2.0$ m^2, moment of inertia $I = 0.12$ m^4, Young's modulus $E = 2.75 \times 10^{10}$N/m^2, and mass per unit length $m = 4{,}800$ kg/m. In the following simulations, this beam is discretized into 2-node plane beam elements in length of 0.01 m. And the damage is introduced as 10 % reduction of Young's modules in the region between 12.5 and 13 m. The vehicle of mass $M = 1{,}200$ kg and the suspension stiffness $k_V = 5{,}000$ kN/m passes through the beam at a constant speed $v = 1$ m/s. And the amplitude of tapping force is set as $a = 0.5$ kgm.

Following the Tap-scan method presented in section "Theory," the time series of VA is subdivided into 25 segments, i.e., 1 s or 1 m per segment. Then, the power spectrum of each segment is obtained by applying the short time Fourier transformation method. The contour of each power spectrum is compared with each other to

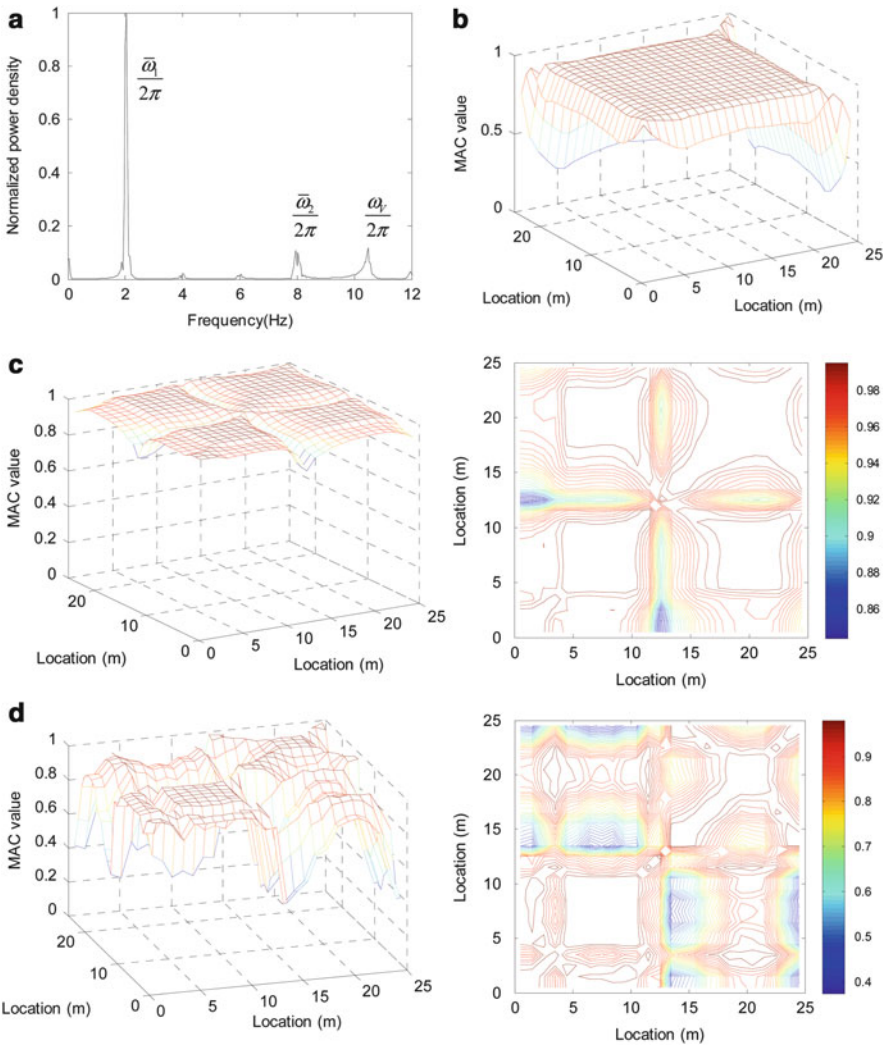


Fig. 8 Damage identification results of a simple beam (FTFs = 2.0, 8.0 Hz). (**a**) the spectrum of VA (**b**) SFs = [1, 4] Hz (**c**) SFs = [6, 9] Hz (**d**) SFs = [9, 12] Hz

get the MAC matrix based on Eq. 28, from which one can justify the damage information.

Since the natural frequency of this beam is $i^2 \times 2.08$ Hz ($i = 1, 2, \ldots$), according to the sensitivity analysis in section "Theory," the FTFs are firstly set as 2.0Hz and 8.0Hz, which are very close to the first two natural frequencies of this beam (2.08 and 8.34 Hz). As Fig. 8a shows, besides these two frequencies one can clearly identify another peak close to ω_V (about 10.27 Hz) in the spectrum. Then, MAC values are calculated from the damage feature vectors $\boldsymbol{Y}_i$, which are extracted

from the spectrum contour at different Sampling Frequencies (SFs). Since the MAC values are evaluated from each pair of damage vectors, one can obtain a symmetric MAC matrix. This matrix can be plotted as a surface (e.g., the left part of Fig. 8c) or a contour (e.g., the right part of Fig. 8c). Both abscissa and ordinates of these plots are the same traveling times along the beam, which can be easily converted into distances, since the vehicle speed is constant. Deep grooves on the MAC surface or dark strips on the contour indicate the abrupt change of MAC values, which correspond to the locations of damage.

As Fig. 8b shows, if SFs locate at the interval of 1–4 Hz, which covers the fundamental natural frequency of the beam, it is impossible to find the damage location from the MAC matrix. However, if the SFs locate at the interval of 6–9 Hz (covers the second natural frequency) or the interval of 9–12 Hz (covers the ω_V), the correct damage location can be clearly identified (see Fig. 8c, d). It seems that the damage feature vector around ω_V is more sensitive to damage than that around the beam natural frequencies, because compared with those in Fig. 8d the MAC surface in Fig. 8c is smoother and the MAC value at damage location is larger. This can be explained by Eq. 36, in which the coefficients at ω_V are more dominant to the coefficients at ω_{Bj}. In addition, the larger ω_{Bj}, the larger coefficients at ω_{Bj}, when $\omega_{Bj} < \omega_V$. Therefore, it is not difficult to understand why it fails to identify the damage when SFs locate around the fundamental natural frequency.

In practice, ω_V is usually larger than the fundamental natural frequency of the bridge but still in the range of environmental noises. In this case, the damage feature vectors usually do not cover ω_V, because the FTFs should be larger than the frequencies of environmental noises. This implies the resultant MAC surface could be smooth but less sensitive to damage. Sure, ω_V could be increased beyond the frequencies of environmental noises, if one can reduce the mass of the vehicle. So that one can still include ω_V into the damage feature vectors to increase the sensitivity. However, the designer should get the caution that the vehicle should be heavy enough to ensure the firm contact on the bridge under large tapping forces.

To illustrate the importance of the FTF, another simulation is conducted by setting the FTFs as 1.0 Hz, 2.5 Hz, 4.0 Hz, and 5.5 Hz. In this case, R_j^i are far from 1 and S_j, T_j, Δ_{Bj}, and Δ_W are very small. Therefore, the contributions of $\ddot{P}_j^i$ and $\ddot{Q}_j$ are very small in VA (see Eq. 34 through Eq. 36). Consequently, only the peaks at the given FTFs (1.0Hz peak can hardly be identified) and ω_V can be identified in the corresponding spectrum shown in Fig. 9a. In addition, the amplitudes of the peaks at FTFs are approximately proportional to the square of $\overline{\omega}_i$, which coincides with the F(t) shown in Fig. 7. This is different from the observation from Fig. 8a, where certain resonances occur due to the small gaps between the FTFs and the natural frequencies.

The MAC matrices for different SFs are shown in Fig. 9b through Fig. 9d. Not surprising, none of them can correctly identify the damage. This illustrates the importance of the FTFs to damage identification.

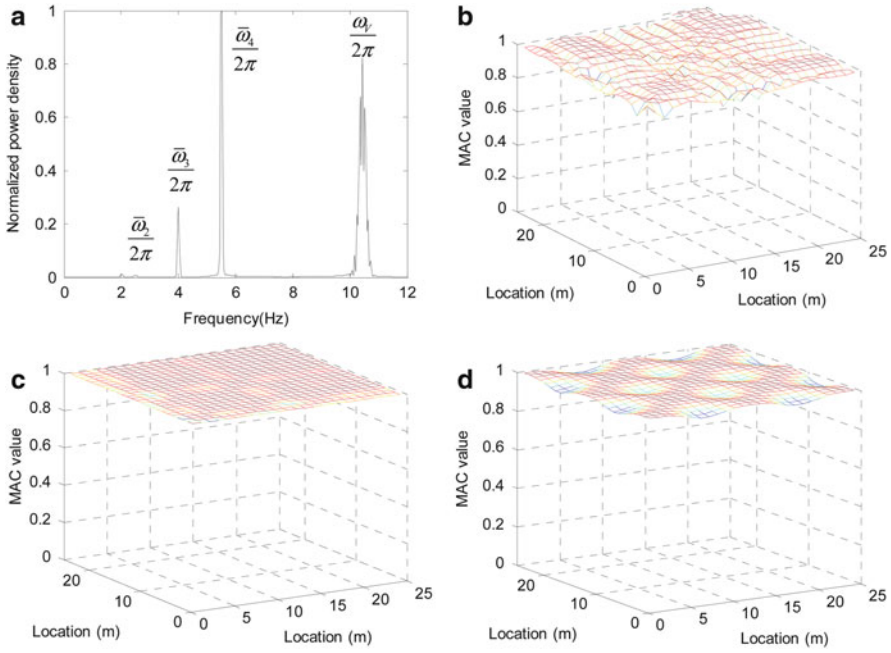


Fig. 9 Damage identification results of a simple beam (FTFs = 1.0, 2.5, 4.0, 5.5 Hz) (**a**) the spectrum of VA (**b**) SFs = [1, 4] Hz (**c**) SFs = [6, 9] Hz (**d**) SFs = [9, 12] Hz

The Setup of the Tap-Scan Damage Detection Vehicle

The physical realization of the Tap-scan damage detection method could be a tractor-trailer system shown in Fig. 10. The trailer acts as that moving sprung mass shown in Fig. 7. And the tapping force is generated by an electromagnetic shaker mounted on the trailer. The trailer is made from pure steel so that it can completely transfer the tapping force to the bridge surface and receives its responses by an accelerator fixed on the axle shaft. A laptop computer runs an integrated program that controls the tapping force, receives acceleration signals, and calculates damage indices. This computer together with a power amplifier, a charge conditioner as well as a power supply system are all loaded in a minivan that tows the trailer to scan each lane of a bridge.

This Tap-scan vehicle system takes the advantage of the passing vehicle (Yang et al. 2004; Lin and Yang 2005; Yang and Chang 2009a, b; Yang et al. 2013) and the controlled tapping force that is essential to damage detection (Xiang et al. 2010). Since the bridge responses are indirectly collected from only one accelerator mounted on the trailer, it is much more efficient and economic than traditional bridge inspection methods that install many sensors on the bridge.

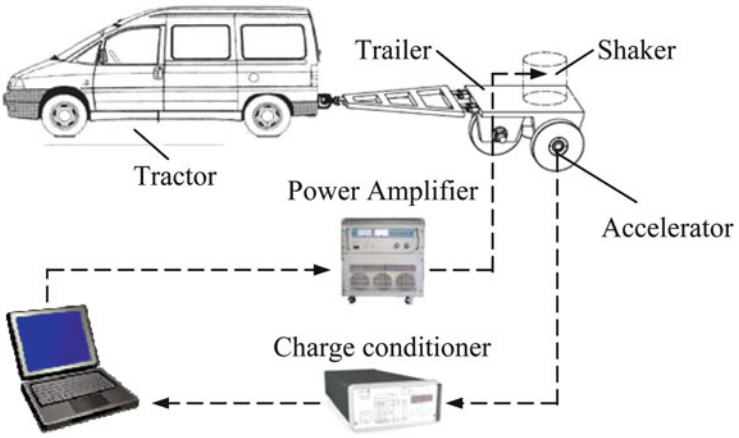


Fig. 10 The Tap-scan damage detection system

On-Site Experiments

According to the sensitivity analysis in section "Theory" and the numerical examples in section "Numerical Examples," the information of natural frequencies of a bridge is crucial to the success of the Tap-scan damage detection method. In on-site experiments, this information can be easily obtained from the spectrum of the acceleration signals when applying swept band exciting force. Then, the FTFs can be determined around certain bridge natural frequencies that are away from the frequency band of environmental noises. As discussed in section "Numerical Examples," these frequencies usually do not include the vehicle frequencies, which are mixed into the frequencies of environmental noises. Empirically, most traffic noises are lower than 70 Hz, so that the frequency band from 70 Hz through 110 Hz is a good choice for bridges with short spans about 20–50 m. This frequency band was adopted in the following bridge inspection experiments conducted by China Road & Bridge Corporation and Tsinghua University.

The Hong Shi Kan Bridge

The Hong Shi Kan Bridge is a 10 years old arch bridge located in Ping Gu district, Beijing, China (see Fig. 11a). The Tap-scan inspection test on this bridge was conducted on September 4th, 2012. For this short bridge of 50 m long, it took about one minute to scan a single lane with the inspection vehicle, without stopping the traffic flow (see Fig. 11b).

The inspection result is plotted in Fig. 12, which is the sharpened contour of the MAC matrix. Since the MAC value is calculated from each pair of spectrum vectors $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$, the resultant MAC matrix is symmetric. So that in Fig. 12, both abscissa and ordinates are traveling times along the bridge length and the dark strips are outliers of MAC values. Comparing with the photos shown in Fig. 13, it is clear that

Fig. 11 The Hong Shi Kan Bridge (**a**) elevation (**b**) on-site inspection

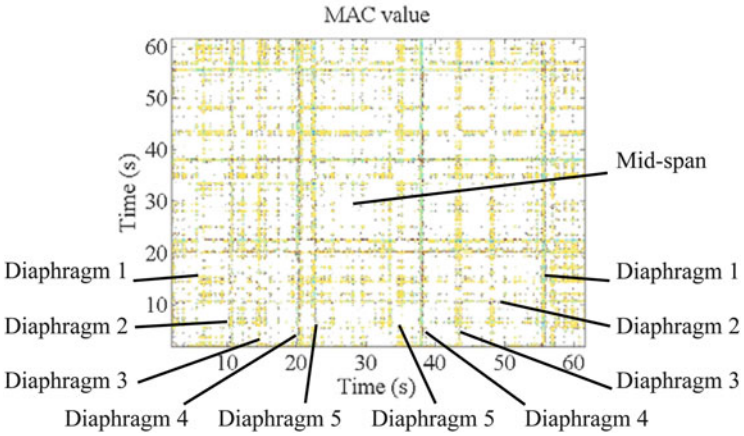


Fig. 12 Damage identification results of the Hong Shi Kan Bridge

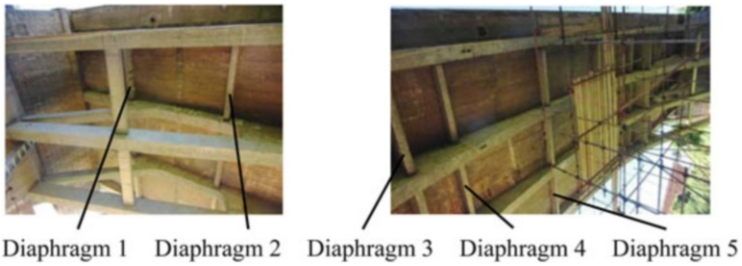


Fig. 13 The distribution of diaphragms of the Hong Shi Kan Bridge

the regularly distributed dark strips in Fig. 12 correspond to the transverse diaphragms under the deck of this bridge. Because it was difficult to keep a precisely constant driving speed, the time spacing is not completely proportional to distance. In addition, since there is no significant abnormal sign in Fig. 12 except for these dark strips, it can conclude that this young bridge is in good condition.

Fig. 14 The Chao He Bridge (**a**) elevation (**b**) on-site inspection

This example demonstrates the ability of the Tap-scan method to find slight structural discontinuities of a bridge under the interference of other passing vehicles.

The Chao He Bridge

The Chao He Bridge is a 45 years old simply supported bridge located in Mi Yun district, Beijing, China (see Fig. 14). This bridge has eight spans. Each span is about 20 m long. The Tap-scan inspection test on this bridge was conducted on September 19th, 2012. It took about 4 min to scan a single lane without stopping the traffic flow. In addition, a heavy bridge inspection crane of 45 t also worked on the bridge during the inspection.

The inspection result is plotted in Fig. 15, in which seven piers can be clearly identified. In addition, one can also find two abnormal strips. One strip is between pier 2 and pier 3, which is just the location of that heavy inspection crane. Another strip is near pier 5. With the help of the inspection crane, examiners found two clear repair marks on the girder close to pier 5 (see Fig.16). Although the detailed record of these repairs was missing, one can imagine the severity of the damage from the photos and Fig. 15.

This example demonstrates that the Tap-scan method can not only identify pier discontinuities but also small damage even the bridge surface is very rugged (see Fig. 14b). At the same time, it warns us that some static heavy trucks with running engine could also have some impact on inspection results. In addition, this also shows the power of using the Tap-scan vehicle to detect hidden damages on such bridges, in which people may not have easy access to conduct visual inspections.

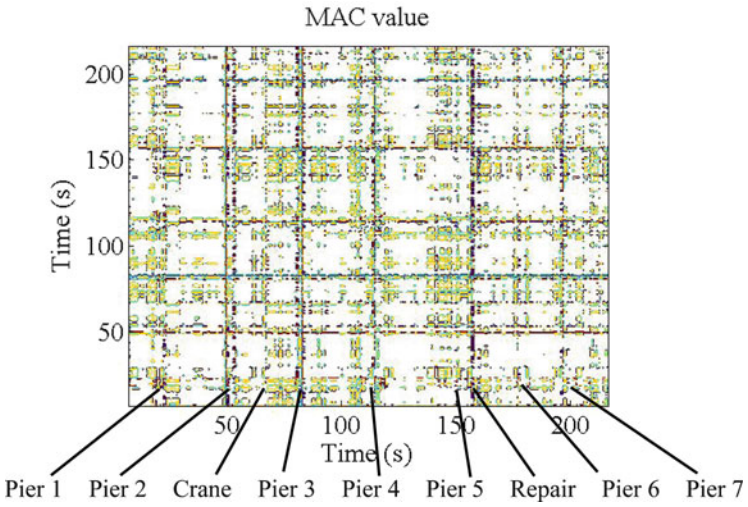


Fig. 15 Damage identification result of the Chao He Bridge

Fig. 16 The damages on the Chao He Bridge

The Yu Lin Bridge

The Yu Lin Bridge is located on S227 Road between Xinxiang City and Yanjin County, Henan Province, China (see Fig. 17). The Tap-scan inspection test on this bridge was conducted on March 16, 2013. For such a short bridge of 20 m long, only half a minute was needed to scan a single lane with the Tap-scan inspection vehicle free of traffic blocking.

The inspection result is plotted in Fig. 18, which clearly shows the left and right expansion joints as well as two damaged regions. Visual inspection verified that there were two severe seepage damages along with a few very small longitudinal cracks (see Fig. 17). In addition, from the darkness of the strips shown in Fig. 18, one can guess that Damage 2 is more severe than Damage 1, although this can hardly be justified only from the visual inspection result.

Fig. 17 The damage inspection of Yu Lin Bridge

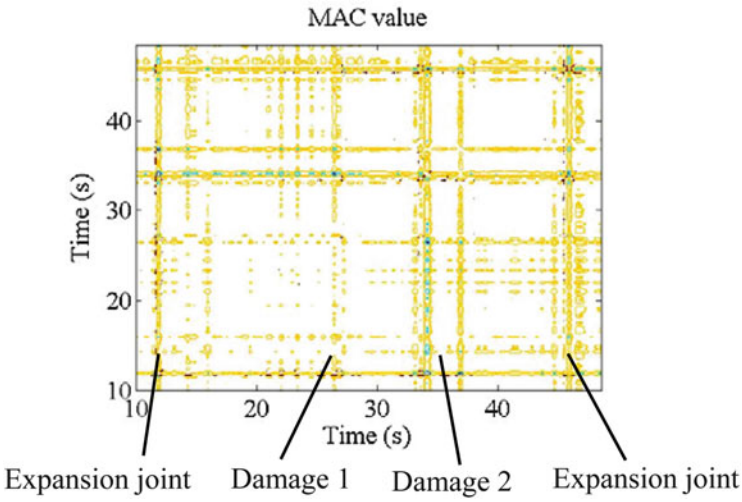


Fig. 18 Damage identification result of the Yu Lin Bridge

Actually, the seepage damage and longitudinal cracks have little contribution to the reduction of local stiffness. The success of identifying these kinds of damage from the Yu Lin Bridge demonstrate that the Tap-scan method actually uses the change of local impedance to detect damage, although the theoretical model in section "Theory" just discusses the change of local stiffness.

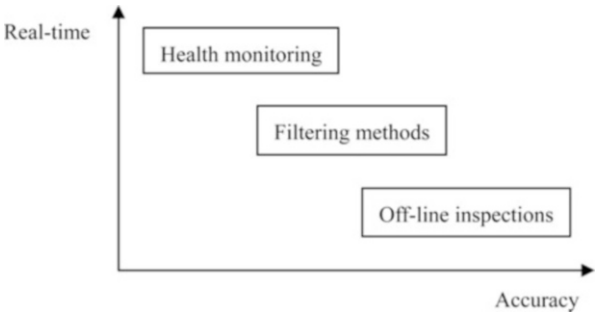


Fig. 19 The hierarchy model of bridge evaluation methods

Summaries

Dynamic signals are good choices for damage detection, because they can be easily acquired from bridges in normal service. If the quality of modal testing can be guaranteed, reliable damage detection can be conducted from the selected frequencies and mode shape points. On the other hand, the Tap-scan method gives a possibility of very robust and efficient bridge inspection without traffic blocking. Both of these two methods do not require the reference state of intact bridges, because they assume that the bridge properties, such as stiffness, mass, and damping, should be smoothly distributed over a healthy bridge and the abrupt change of these properties indicates the structural discontinuity or the occurrence of local damage. This greatly facilitates their practical implementations.

As Fig. 19 illustrates, bridge evaluation methods could be classified into three levels according to their accuracies and applying intervals. The method that uses the selected frequencies and mode shape points is promising for real-time health monitoring on some very important bridges. While for large amount of common bridges, one can use the Tap-scan inspection vehicle to quickly filter out a few suspects for further detailed examinations. This could help big cities to efficiently manage their bridge networks. Since these two kinds of methods just regard the damage as the change of some effective properties, off-line local inspections are still needed to give detailed reports of damage.

References

R.J. Allemang, D.L. Brown, Correlation coefficient for modal vector analysis, in *Proceedings of the 1st International Modal Analysis Conference* (Society for Experimental Mechanics, Orlando, 1982), pp. 110–116

M. Baruch, Modal data are insufficient for identification of both mass and stiffness matrices. AIAA J. **35**, 1797–1798 (1997)

E.P. Carden, P. Fanning, Vibration based condition monitoring: a review. Struct. Health Monit. **3**, 355–377 (2004)

P. Cawley, R.D. Adams, The mechanics of the coin-tap method of non-destructive testing. J. Sound Vib. **122**, 299–316 (1988)
P.C. Chang, A. Flatau, S.C. Liu, Review paper: health monitoring of civil infrastructure. Struct. Health Monit. **2**, 257–267 (2003)
A. Cunha, E. Caetano, F. Magalhães, C. Moutinho, Recent perspectives in dynamic testing and monitoring of bridges. Struct. Control Health Monit. **20**, 853–877 (2013)
A. Deraemaeker, E. Reynders, G. Roeck, J. De Kullaa, Vibration based structural health monitoring using output-only measurements under changing environment. Mech. Syst. Signal **22**, 34–56 (2008)
H.W. Engl, M. Hanke, A. Neubauer, *Regularization of inverse problems* (Kluwer, Dordrecht, 1996)
EPS Net (2010) China Macro Economy Database. http://www.epsnet.com.cn/
W. Fan, P. Qiao, Vibration-based damage identification methods: a review and comparative study. Struct. Health Monit. **10**, 83–111 (2011)
C.R. Farrar, D.A. Jauregui, Comparative study of damage identification algorithms applied to a bridge: I. experiment. Smart Mater. Struct. **7**, 704–719 (1998)
C.R. Farrar, S.W. Doebling, D.A. Nix, Vibration-based structural damage identification. Philos. Trans. R. Soc. A **359**, 131–149 (2001)
C. Gentile, A. Saisi, Ambient vibration testing and condition assessment of the Paderno iron arch bridge (1889). Construct. Build Mater. **25**, 3709–3720 (2011)
D.H. Griffel, *Applied Functional Analysis* (Dover, Mineola, 2002)
E. Haber, U.M. Ascher, D. Oldenburg, On optimization techniques for solving nonlinear inverse problems. Inverse Probl. **16**, 1263–1280 (2000)
C.W. Lin, Y.B. Yang, Use of a passing vehicle to scan the fundamental bridge frequencies. Eng. Struct. **27**, 1865–1878 (2005)
R. Nelson, Simplified calculation of eigenvector derivatives. AIAA J. **14**, 1201–1205 (1976)
N.M. Okasha, D.M. Frangopol, Integration of structural health monitoring in a system performance based life-cycle bridge management framework. Struct. Infrastruct. Eng. **8**, 999–1016 (2012)
S.L. Padula, R.K. Kincaid, Optimization strategies for sensor and actuator placement. NASA Technique Report: TM-1999-209126, 1999
O.S. Salawu, Detection of structural damage through changes in frequency: a review. Eng. Struct. **19**, 718–723 (1997)
Z.H. Xiang, Y. Zhang, Changes of modal properties of simply-supported plane beams due to damage. Interact. Multiscale Mech. **2**, 171–193 (2009)
Z.H. Xiang, X.W. Dai, Y. Zhang, Q.H. Lu, The tap-scan method for damage detection of bridge structures. Interact. Multiscale Mech. **3**, 173–191 (2010)
Z.H. Xiang, L.Q. Wang, M.S. Zhou, Suppressing damage identification errors from selected natural frequencies and mode shape points. Inverse Probl. Sci. Eng. **20**, 871–890 (2012)
Y.B. Yang, K.C. Chang, Extracting the bridge frequencies indirectly from a passing vehicle: parametric study. Eng. Struct. **31**, 2448–2459 (2009a)
Y.B. Yang, K.C. Chang, Extraction of bridge frequencies from the dynamic response of a passing vehicle enhanced by the EMD technique. J. Sound Vib. **322**, 718–739 (2009b)
Y.B. Yang, C.W. Lin, J.D. Yau, Extracting bridge frequencies from the dynamic response of a passing vehicle. J. Sound Vib. **272**, 471–493 (2004)
Y.B. Yang, K.C. Chang, Y.C. Li, Filtering techniques for extracting bridge frequencies from a test vehicle moving over the bridge. Eng. Struct. **48**, 353–362 (2013)

Materials Mechanical Characterizations and Structural Diagnoses by Inverse Analyses

20

Vladimir Buljak, Giuseppe Cocchetti, Aram Cornaggia, Tomasz Garbowski, Giulio Maier, and Giorgio Novati

Contents

V. Buljak
Faculty of Mechanical Engineering, Department of Strength of Materials, University of Belgrade, Belgrade, Serbia
e-mail: vladimir.buljak@polimi.it

G. Cocchetti • A. Cornaggia • G. Maier (✉) • G. Novati
Department of Civil and Environmental Engineering, Politecnico di Milano, Milano, Italy
e-mail: giuseppe.cocchetti@polimi.it; aram.cornaggia@polimi.it; giulio.maier@polimi.it; giorgio.novati@polimi.it

T. Garbowski
Institute of Structural Engineering, Poznan University of Technology, Poznań, Poland
e-mail: tomasz.garbowski@put.poznan.pl

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_33

Abstract

Mechanical damages in structures, in structural components of plants and in industrial products usually imply changes of parameters which have central roles in computational modelling apt to assess safety margins with respect to service loading. Such parameters may depend also on the production processes in industrial environments.

In this chapter, the parameter identification methodology by inverse analyses is dealt with under the following limitations: experiments at macroscale level, deterministic approaches, statical external actions and time independence in material behaviours. Semiempirical approaches frequently adopted in codes of practice are not dealt with here.

The inverse analysis methods outlined here are centered on computational simulations of tests (namely, direct analyses), sensitivity analyses for the optimal design of experiments, model reduction procedures and other provisions apt to make fast and economical the parameter estimation in engineering practice. The applications summarized here as examples concern structural diagnoses based on indentation tests, in situ diagnostic experiments on concrete dams and laboratory mechanical characterization of membranes and laminates.

Introductory Remarks

The "inverse analysis" methodology is an area of applied sciences which at present is still growing as for improvements of procedures and as for variety of engineering applications. Inverse analysis is based on information concerning the response of a "system" to external actions and leads to the identification of some features of the system, usually parameters included in its modelling and, hence, in the computer simulation of the system response to those actions.

In the present context of applied mechanics, the features to assess are usually either parameters contained in material constitutive models or stresses present in the system and included in the set of "parameters" to estimate or "identify." The system may be a laboratory specimen or a structural component as industrial product; however, frequently, it consists of a structure possibly affected by damages due to deterioration in service. Therefore, inverse analysis is becoming central to "structural diagnosis" intended to provide a reliable basis for the subsequent "direct analyses" apt to assess "margins of safety" with respect to collapses or to substantial further structural damages ("admissible stress" criteria being superseded now in more and more industrial codes).

Inverse analysis with the above purposes clearly requires synergistic convergence of experimental, structural and computational mechanics and applied mathematics. References to such synergy can be attributed to Richard Feynman's symbolic warning to ensure reliability of computer inputs ("garbage in, garbage out") after the 1986 disaster of the Space Shuttle Challenger (Gribbin and Gribbin (1997)). Several stimuli to structural diagnosis improvements arose from other

tragic events in engineering history (see, e.g., Levy and Salvadori (1992), such as Vajont Dam (1963), Alexander Kielland platform (1980) and Gulf of Mexico disaster (2010)).

Classifications of the inverse analysis procedures now presented in a broad literature (e.g., Bui 1994, 2006; Tarantola 2005; Mróz and Stavroulakis 2005) can be based on diverse standpoints. According to a meaningful methodological criterion, the following categories can be considered.

(a) Deterministic approaches which lead to estimates of the parameters to identify without quantification of their uncertainties. Randomness, such as experimental noise, can be possibly quantified by the covariance matrix of the experimental data. The inverse of this matrix is employed in the formulation of the "discrepancy function" to be minimized with respect to the sought parameters. Such formulation is motivated by advantages reasonably expected in estimation accuracy when "more weight" is conferred to more reliable experimental measurements. Clearly, repeated inverse analyses starting from measurable quantities obtained by random perturbations of experimental (or "pseudoexperimental") data may lead to parameter sets which can provide a quantification of the consequences on the estimates (e.g., standard deviations) of those perturbations in input. However, such approach based on deterministic inverse analysis may be clearly uneconomical since it may require many repeated inverse analyses.
(b) Stochastic approaches which provide assessments both of estimates and of estimation uncertainties due to random errors of experimental data. Among stochastic methods (which include popular techniques like Monte Carlo, Bayes, et alia), particular versatility and usefulness can at present be attributed to Kalman filters. Such "filtering" starts from the experimental data and their covariance matrix which quantifies random noise and, after a sequence of steps sequentially exploiting experimental data for estimation updates (by sequential "sensitivity" and "gain" matrices), provides both final parameter estimates and their covariance matrix.

In this chapter only deterministic approaches (a) are considered.

Other limitations, adopted in view of the contents of other chapters in this handbook, are pointed out in what follows.

Even if the inverse analysis methodology is clearly applicable at both macro- and microscales, only industrial applications at macroscale are here briefly discussed as illustrative examples. Particularly worth pointing out is that attention will be paid elsewhere to dynamical external actions (e.g., for structural diagnosis of possibly damaged bridges) and to time-dependent material models (e.g., visco–elastic–plastic modelling of prestressed concrete bridges damaged by creep, high-strain rates in structures exposed to impacts or explosions).

The purposes pursued in this chapter and its contents can be outlined as follows. Section "A Survey of a Practical Methodology for Damage Assessment in Structures" is devoted to a brief description, stage by stage, of the, partially innovative,

methodology of inverse analysis for mechanical characterizations of materials or for diagnoses of structures or structural components, in the industrial engineering context. Detailed information on mathematical developments, computational tools and engineering praxis can be found in the publications specified in the reference list. Section "Parameter Identifications by Indentation Tests" presents primarily recent research results on inverse analysis procedures based on "quasi-nondestructive" tests centered on indentations (originally "hardness tests") and devised primarily for damage assessments in metallic industrial products. The applications of inverse analysis dealt with as informative examples in sections "Damage Assessment in Dam Engineering" and "Inverse Analysis Procedures for Mechanical Characterizations of Foil Products" concern a variety of engineering areas, such as concrete dams, food-container productions and tension structure design. The remarks presented in section "Conclusion" aim at underlining the potential practical usefulness of recent results and current research developments for assessments of possibly deteriorated material properties and of mechanical damages in structures.

A Survey of a Practical Methodology for Damage Assessment in Structures

With the limitations specified in the preceding section, a rather general inverse analysis method is here outlined in operative terms, namely, stage by stage, with generality in view of its practical applications to be exemplified with some peculiar features in the subsequent sections.

Stage 1. Preliminary examinations of the structure to be investigated "in situ" or of the specimen to be tested in a laboratory are obviously required for the selection of the parameters to identify by inverse analysis, within constitutive models of materials (e.g., Lubliner 1990; Jirásek and Bažant 2001) and/or within a stress state. Clearly, related to such examinations are the experiment to be performed and its design. The design of experiments has to be performed in terms of external actions ("loads") and of quantities to measure in the system response to them, taking account (if possible through a covariance matrix) of expected experimental errors with randomness.

Stage 2. A conjecture to be exploited later has to be formulated by an "expert" on the following subject: a lower and upper bound should be specified for each one of the sought parameters, in order to define an interval over which the estimate search will later be confined. Thus, a "search domain" is defined in the parameter space. If "correlation" is predicted among parameters (namely, if one parameter is expected to be large if another will turn out to be so), some guess on correlation ratio might be requested from the "expert."

Stage 3. Computer modelling of the preselected tests is a crucial step. Even if a popular finite element (FE) commercial code is adopted to the present purposes, attention should obviously be paid to computational features: mesh generation, boundary conditions and interfaces, possible large-strain requirements, implementation of the

selected constitutive model, etc. A compromise has to be reached between accuracy of results and computational time for multiple test simulations.

Stage 4. Sensitivity analysis is a need which arises from the obvious remark that it is particularly worthwhile to measure, among test response quantities, those which are most influenced by the parameters to estimate through test simulations (see, e.g., Kleiber et al. 1997). Consequently, sensitivities are assessed by computing derivatives (of course usually approximated by finite differences) of the relationships by means of which measurable quantities depend on sought parameters through the system model and test simulations.

However, another criterion turns out to be useful in the present inverse analysis context, namely, larger (say by two orders of magnitude) than relevant experimental error should be the difference between the values of measurable quantities, which are computed by attributing the conjectured upper and lower bound to each parameter to identify. Both the above sensitivity assessments require merely direct analyses and may be meaningful for the design of the experiment (see the next stage and section "Parameter Identifications by Indentation Tests").

Stage 5. The design of the experiments, in order to accurately estimate parameters, obviously within basic constraints on the available resources, has to be carried out frequently, on the basis of sensitivity analysis (Stage 4), to the following purposes:

(a) selection of the quantities to be measured, in the response of the examined system to the test loading;
(b) "a priori" optimization of some features of the experiment (e.g., of the indenter shape in indentation tests, as pointed out in section "Parameter Identifications by Indentation Tests").

Purpose (a) is usually reached by the former approach mentioned above for sensitivity analyses (Stage 4) and purpose (b) by the latter approach as well.

Stage 6. Computer simulations of the test which has been designed in what precedes can lead from any vector $\mathbf{p}$ of P sought parameters to the M-vector $\mathbf{u}$ (usually called "snapshot") of measurable quantities; these quantities are called here "pseudoexperimental" data when they are merely computed as functions of $\mathbf{p}$.

Inverse analysis is mathematically formulated as minimization, with respect to the variable P-vector $\mathbf{p}$, of a "discrepancy function" $\omega(\mathbf{p})$ defined as a norm of the differences between experimental or pseudoexperimental data gathered in M-vector $\overline{\mathbf{u}}$ and their corresponding counterparts gathered in "snapshot" M-vector $\mathbf{u}$, which is a function of the parameters contained in $\mathbf{p}$. Formally,

$$\min_{\mathbf{p}} \omega(\mathbf{p}) = \omega_{\min} \tag{1}$$

$$\omega(\mathbf{p}) = [\overline{\mathbf{u}} - \mathbf{u}(\mathbf{p})]^{\mathrm{T}} \mathbf{C}^{-1} [\overline{\mathbf{u}} - \mathbf{u}(\mathbf{p})] \tag{2}$$

where $\mathbf{C}$ represents the covariance matrix of the data. When such quantification of the random noise affecting measurements is not available, the identity matrix replaces $\mathbf{C}$ without any qualitative change in the subsequent computations.

As mentioned in section "Introductory Remarks," in deterministic approaches to inverse analyses, covariance matrix in Eq. 2 merely implies useful attribution of "more weight" to more reliable data.

Stage 7. The above mathematical minimization problem, Eqs. 1 and 2, may exhibit lack of convexity in the objective function ω (hence, possible local minima), "non-smoothness" in the constraints and "ill-posedness." In order to reach the absolute minimum, the following kinds of algorithms can at present be adopted in practical applications: mathematical programming and particularly Trust Region Algorithms (TRA), e.g., Conn et al. (2000) and Coleman and Li (1996); Genetic Algorithms (GA), e.g., Koh and Perry (2009); Artificial Neural Networks (ANN), e.g., Haykin (1998), Hagan et al. (1996) and Waszczyszyn (1999). Such computational methods are implemented in software available on the market and described in the literature, but herein, they are only mentioned and later employed in problems outlined in the subsequent sections.

The following computational–mathematical circumstances have meaningful consequences to the present purposes of inverse analyses in engineering environments:

(i) TRA often requires diverse initializations in order to possibly avoid ending up at a local minimum and at each step, first-order derivatives are needed in order to approximate the Hessian (by gradient of Jacobian) in the step problem of quadratic programming in two variables;
(ii) in GA applications, each "member" $\mathbf{p}$ of each sequential "population" over the "search domain" defined in Stage 2 requires a test simulation;
(iii) in order to exhibit efficiency and stability, ANNs require a balance between the number of inputs and the number of outputs.

The above and other computational circumstances make highly desirable, for routine practical applications of inverse analyses, to reduce computational efforts by recourse to some procedure of "model reduction" like the one applied here as subsequent stages.

Stage 8. Over the "search domain," defined at Stage 2, within the space of parameters to identify (by means of the bounds suggested by the "expert"), a "grid of N nodes" ($\mathbf{p}_i$, $i = 1 \ldots N$) is now to be constructed. The simplest way of grid generation rests obviously on a subdivision of the search interval conjectured for each sought parameter in an equal number of equal subintervals and on the assumption as "nodes" $\mathbf{p}_i$ of the vertices of all cells thus generated. Clearly, with such grid generation (adopted in the inverse analysis problems presented in this chapter), the number of nodes grows exponentially with the number of the parameters to identify. When an "a priori" choice of the node number N becomes useful, the node distribution can be performed by alternative methods, not described here for brevity (see, e.g., Viana et al. 2010; Bates et al. 2004).

Stage 9. The "model reduction" procedure called "Proper Orthogonal Decomposition" (POD) is adopted here below and in the innovative inverse analysis procedures applied in the subsequent sections. It has remote origins in mathematics oriented to

economics and can be outlined, as follows, to the present purposes. Details can be found in Ruckelynck et al. (2006), Ostrowski et al. (2008), Nouy (2010), and Buljak (2012). By assuming the parameters $\mathbf{p}_i$ at each node of the grid generated in the preceding stage, the test simulation is carried out leading to the vector $\mathbf{u}_i$ containing the pseudoexperimental data as "snapshot" corresponding to $\mathbf{p}_i$, through direct analysis by the (say FE) discretization model elaborated at Stage 3. Let a $M \times N$ matrix $\mathbf{U}$ gather all such "snapshots." The number M of quantities measurable in the test is usually smaller than N, which grows exponentially with the parameter number.

In most real-life problems, the responses $\mathbf{u}_i$ of the tested system to the same given external actions, but with diverse parameters $\mathbf{p}_i$ internal to the "search domain," turn out to be "correlated," namely, "almost parallel" in their space. Such correlation, often physically motivated, can be easily checked on matrix $\mathbf{U}$ and computationally exploited by the procedure at Stage 10. Attention should be paid to "correlation" of test responses also because it may not hold in some situations (e.g., a small change in Young modulus may cause the transit from stability to instability in a structural component under dominant compression).

Stage 10. In the M-dimensional space of the snapshots $\mathbf{u}_i$ ($i = 1 \ldots N$), a new reference axis is singled out by maximizing, with respect to all directions, a Euclidean norm of the projections on it of all N snapshots $\mathbf{u}_i$. Then, another axis is found by similar maximization over the set of all directions orthogonal to the one above singled out. A sequence of such optimizations leads to a new reference system, or "new basis," analytically described by an orthonormal matrix $\mathbf{\Phi}$ of order M such that

$$\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi} = \mathbf{I}, \qquad \mathbf{U} = \mathbf{\Phi}\mathbf{A}, \qquad \mathbf{A} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{U} \tag{3}$$

where the $M \times N$ matrix $\mathbf{A}$ gathers, as columns, the vectors (called "amplitudes" $\mathbf{a}_i$ in the POD jargon) which describe the snapshots $\mathbf{u}_i$ in the "new basis." The abovementioned (Stage 9) "correlation" among test responses with various parameters varying within their "search domain" naturally motivates large differences among amplitude components and suggests simplifications based on removal of axes with negligible components in the new basis. Such "truncation," intuitively very natural and clear, had been investigated decades ago in economics-oriented mathematics. To the present purposes, some operative details can be found in Chatterjee (2000). Here, only the main features are mentioned, namely, the above "truncation" is based on computation of the eigenvalues λ_i ($i = 1 \ldots N$) of matrix $\mathbf{D} = \mathbf{U}^{\mathrm{T}}\mathbf{U}$ (of order N, symmetric positive definite or semidefinite) and on preservation of the axes corresponding to the λ_i larger by orders of magnitude than the smallest eigenvalues.

Thus, a "truncated basis" (matrix $\hat{\mathbf{\Phi}}$ of order $M \times K$ with $K \ll M$) is generated for approximations of the test responses $\mathbf{u}_i$ through their dependence on "reduced amplitudes" $\hat{\mathbf{a}}_i$, namely:

$$\mathbf{u}_i \cong \hat{\mathbf{\Phi}}\,\hat{\mathbf{a}}_i \quad (i = 1 \ldots N) \quad \text{or} \quad \mathbf{U} \cong \hat{\mathbf{\Phi}}\,\hat{\mathbf{A}} \tag{4}$$

An assessment of errors implied by the recourse to the truncated POD basis can be easily evaluated by comparing the sum of the eigenvalues λ_i ($i = 1 \ldots K$)

related to the K preserved directions (or "modes") to the sum of all the original ones λ_i $(i = 1 \ldots N)$.

The truncated basis $\hat{\mathbf{\Phi}}$ exhibits the mathematical features of the original basis $\mathbf{\Phi}$ expressed by Eq. 3; therefore, the reduced "amplitude" $\hat{\mathbf{a}}_i$ of any snapshot $\mathbf{u}_i$ can be computed with the approximation as in Eq. 4, namely:

$$\hat{\mathbf{a}}_i \cong \hat{\mathbf{\Phi}}^{\mathrm{T}} \mathbf{u}_i \quad (i = 1 \ldots N) \quad \text{or} \quad \hat{\mathbf{A}} \cong \hat{\mathbf{\Phi}}^{\mathrm{T}} \mathbf{U} \tag{5}$$

The model reduction procedure outlined in what precedes concerns the set of the N parameter vectors $\mathbf{p}_i$ preselected as grid nodes in the "search domain" and can be done once for all in view of repeated practical applications of inverse analyses, which can be made fast by the computational provisions outlined in the next stage.

Stage 11. The minimization of the discrepancy function $\omega(\mathbf{p})$, Eqs. 1 and 2, requires a high number of test simulations, as underlined earlier, both if a GA is employed and if an algorithm of mathematical programming as TRA is adopted. Such practical difficulty can be overcome by means of the computational provisions summarized in what follows (see, e.g., Buhmann 2003; Kansa 2001).

For each parameter grid node $\mathbf{p}_i$ $(i = 1 \ldots N)$, a Radial Basis Function (RBF) is considered, namely:

$$g_i(\mathbf{p}) = \left[(\mathbf{p} - \mathbf{p}_i)^{\mathrm{T}} (\mathbf{p} - \mathbf{p}_i) + r^2 \right]^{-\frac{1}{2}} \quad (i = 1 \ldots N) \tag{6}$$

with the "smoothing coefficient" r to be calibrated, once for all, for each kind of application. Each component $\hat{a}_j^k$ $(k = 1 \ldots K, j = 1 \ldots N)$ of the "reduced amplitude" vector $\hat{\mathbf{a}}_j$ corresponding to node parameters $\mathbf{p}_j$ defined in Stage 10 and by Eq. 5 is expressed as a linear combination of the values acquired there by the RBFs Eq. 6:

$$\hat{a}_j^k = \sum_{i=1}^{N} b_i^k g_i\left(\mathbf{p}_j\right) \quad (k = 1 \ldots K, \ j = 1 \ldots N) \quad \text{or} \quad \hat{\mathbf{A}} = \mathbf{B}\mathbf{G} \tag{7}$$

This Eq. 7 consists of $K \times N$ linear equations in $K \times N$ unknowns b_i^k, gathered in matrix $\mathbf{B}$, whereas matrix $\mathbf{G}$ contains the known values $g_i(\mathbf{p}_j)$ of all functions RBF in all N parameter nodes $\mathbf{p}_i$ of the grid over the search domain.

The simple solution of Eq. 7 provides the coefficients b_i^k for the linear combination which leads from any "new" parameter vector $\mathbf{p}$ out of the grid nodes to the reduced amplitudes $\hat{\mathbf{a}}$ of the "snapshot" $\mathbf{u}$; this vector $\mathbf{u}$ quantifies the pseudoexperimental data resulting from the test simulation based on the parameters contained in that vector $\mathbf{p}$. In matrix formulae,

$$\mathbf{u}(\mathbf{p}) = \hat{\mathbf{\Phi}} \hat{\mathbf{a}} = \hat{\mathbf{\Phi}} \mathbf{B}\, \mathbf{g}(\mathbf{p}) \tag{8}$$

where vector $\mathbf{g}$ gathers the N values of $g_i(\mathbf{p})$ $(i = 1 \ldots N)$ acquired by the RBF centered on $\mathbf{p}$ in all grid nodes (Eq. 6). When matrix $\mathbf{B}$ is available since it was

provided by the solution once for all of Eq. 7, any "direct analysis," i.e. any test simulation leading to the measurable quantities, can be carried out by Eq. 8, instead of by FEM or by other methods, with computing times by various orders of magnitude shorter, at comparable accuracy. Clearly, the practical benefits for parameter identifications by means of either TRA or GA are significant. This circumstance implies substantial computational advantages also when an ANN is adopted for fast inverse analyses, since the ANN input may consist of amplitude vector $\hat{\mathbf{a}}$ which represents the snapshot $\mathbf{u}$ with much lesser number of components.

Stage 12. A question which arises in engineering applications of inverse analysis concerns the consequences of experimental errors on the parameters estimates. Within the deterministic methodology considered in this chapter, by exploiting the computational efficiency acquired by POD, an answer to the above question can be achieved as follows: many parameter identifications are performed starting from randomly perturbed experimental data (such randomness quantifies the measurement accuracy); the consequent "perturbations" in the estimates is quantified, say, by mean value and standard deviation of their probability density distribution.

The inverse analysis procedure outlined stage by stage in what precedes can be regarded as a methodological platform for the contents of the sections which follow. The sequels in this chapter are aimed to illustrate potentialities and limitations of such methodology by means of research results achieved by the authors and oriented to improvements, in the present state of the art, for the assessment of structural damages and of mechanical properties of materials in engineering and industrial environments. Clearly, diverse technological areas and other approaches and research results far from those dealt with in this chapter may contribute to the above purposes, also through other chapters of this handbook.

A possible future methodological development based on what was outlined in this section, but with substantial diversity as well, is briefly presented here below. Several prestressed concrete structures, primarily bridges, are at present damaged by creep occurred along decades since their construction. Two circumstances can be supposed to affect the relevant engineering situation: experimental data have been accumulated on such structure along decades of its long life because of alarms in view of evident damaging processes due to concrete creep; diverse material models, say A and B, have been proposed in order to reliably predict the residual time of service of that bridge before collapse, if no interventions occur, and/or in order to design repairing provisions. Let the structural behaviour in time under the known loads be modelled, first by using time-dependent constitutive model A for concrete; subsequently, the same modelling work is done (same FEM code, same FE mesh and boundary conditions, same external actions, etc.), but using material model B. A discrepancy function $\omega^{\mathrm{A}}(\mathbf{p}^{\mathrm{A}})$ is defined and minimized with respect to the parameters which are not accurately known a priori in model A, by an inverse analysis according to the computational technique described in this chapter (but not necessarily by carrying out the POD and subsequent use of it, since repeated routine applications are not required by the present engineering problem). A second inverse analysis is done with a single change which concerns the constitutive model B of

concrete mechanical behaviour with emphasis on time dependence. Clearly, meaningful is a comparison between the minimized discrepancies $\omega_{\min}^{A}$ and $\omega_{\min}^{B}$: the model leading to a lower minimum turns out objectively to be preferable for reliable predictions on the future of the structure considered (and of other structures similarly affected by phenomena long lasting in time, hence seldom susceptible of new repeated reliable tests).

The above hints to unusual comparative applications of the inverse analysis method dealt with in this chapter may have conceptual connections and practical links with the contents of the chapter concerning prestressed bridges in this handbook.

Parameter Identifications by Indentation Tests

Calibration of Tensorial and Fracture Models on the Basis of Indentation Curves Alone

Penetration of a stiff tip into a structural component, possibly "in situ" (without extracting any specimen), superficially (say less than one millimeter) and, hence, as "non-destructive" test, has been since decades a popular experiment in engineering mechanics and a subject of an abundant literature; see, e.g., Oliver and Pharr (1992), Dao et al. (2001), Kucharski and Mróz (2004), and Bolzon et al. (2004).

Originally, "hardness" was the material parameter to be achieved in view of its metallurgical meaning. Later parameters in elastic–plastic constitutive models became the targets of estimation by means of indentation tests and empirical or semi-empirical formulae.

Recent developments are centered on inverse analysis methods. In what follows, "indentation curves" alone are sources of measured data; both indentation curves and imprint profiles will be considered in the subsequent subsections.

Diagnostic analyses of possibly deteriorated structures (e.g., pipelines, offshore platforms, ships, power plant components, tall steel buildings) frequently require a high number of tests which implies high costs if performed on specimens extracted, with obvious consequent damages, and taken to laboratories. Instrumented indenters can be easily employed "in situ," but the following problem arises in engineering practice: can damage parameters be identified fast and repeatedly "in situ"? Clearly, such a problem becomes more challenging if the sought parameters govern direction-dependent (tensorial) properties.

The following practical situations with the above difficulties for diagnosis have been subjects of recent research:

(A) residual stresses, particularly due to defective welding processes;
(B) anisotropy in elastic and/or plastic material behavior, e.g., sometimes due to lamination or other production treatments.

The above parameter estimation problems can be tackled by the following novel operative procedures:

(i) the geometrical shape of the indenter is preliminarily optimized by sensitivity analyses for each kind of practical situation, on the basis of the "search domain" suggested by an expert (see Stages 2, 4 and 5 in section "A Survey of a Practical Methodology for Damage Assessment in Structures");
(ii) the indenter with optimized cross section no longer axisymmetric (but, e.g., elliptical) is employed for two or three indentations in near locations, but rotated by 90 or twice by 45 around the axis.

In both practical problems (A) and (B), the stage sequence of inverse analysis turned out to be productive of reliable estimates and accelerated by POD-RBF-TRA (Stages 7, 8, 9, 10, 11) procedures, susceptible to be performed routinely "in situ" by a small computer.

A numerical exercise apt to illustrate residual stress assessment, namely, method (A), is concisely presented in what follows. Details are available in Buljak et al. (2012). A traditional spherical Brinell-type indenter with diameter D_s is considered in what follows. If the coordinate x along the indenter axis has its origin on the indenter tip, the two diameters of the ellipse generated as new cross section by shape modification (Fig. 1a) can be described by the following formulae:

$$D_0(x) = 2\left(D_s x - x^2\right)^{1/2}, \qquad D_{\max} = \beta D_0(x), \qquad D_{\min} = (1/\beta) D_0(x) \tag{9}$$

where $\beta > 1$ governs the sharpness of the new shape.

For preliminary computational exercises, the classical isotropic elastic–perfectly–plastic "associative" Huber–Hencky–Mises model has been adopted in Buljak et al. (2012). The reference values which characterize, in terms of principal stresses and directions, the residual stress state to identify are $\sigma_I = 500$ MPa and $\sigma_{II} = -500$ MPa, $\varphi = 20^\circ$ being the angle between axis 1 and direction of σ_I. The indenter which turned out to be preferable in comparative computations is ellipsoidal with shape defined by $D_s = 0.5$ mm and $\beta = 2$, Eq. 9; its geometry is specified in Fig. 1a. Figure 1b visualizes the influence on indentation curves of this indenter when indentations are performed three times by rotating it 45° each time. The adopted FE model for test simulations (a) and the indentation sequence (b) are shown in Fig. 2. The POD procedure was started by adopting in the parameter space a regular grid with 125 nodes by varying the three parameters within the following ranges: $-600 < \sigma_I < 600$ MPa, $-600 < \sigma_{II} < 600$ MPa; $0 < \varphi < 90^\circ$.

Figure 3 visualizes the TRA step sequences up to convergence of discrepancy function minimization. The same optimization procedures were repeated three additional times starting from different initialization vectors. The averages resulting from all the inverse analyses exhibit the following values: $\sigma_I = 503$ MPa, $\sigma_{II} = -499$ MPa and $\varphi = 21.9^\circ$. These values turn out to be satisfactory if compared to their counterparts earlier assumed as "targets" (500 MPa, −500 MPa

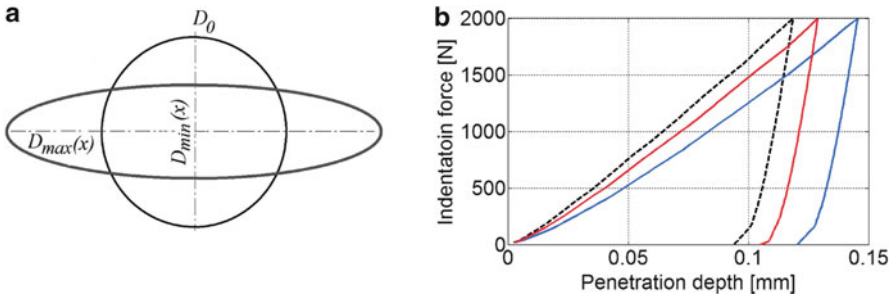


Fig. 1 (**a**) Generation of ellipsoidal indenter from a spherical one. (**b**) Indentation curves provided by three tests, rotated by 45° one from the other, in the presence of residual stresses

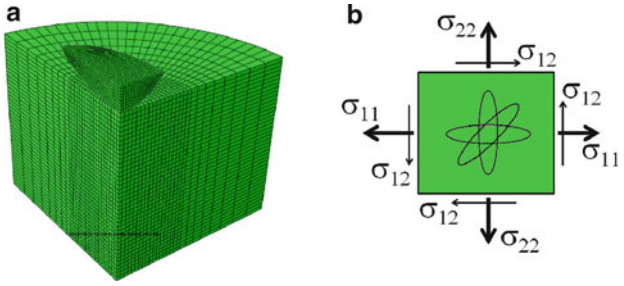


Fig. 2 (**a**) Finite element mesh for test simulation of an indentation test. (**b**) Three indentation tests rotated by 45° for the estimation of a residual stress state

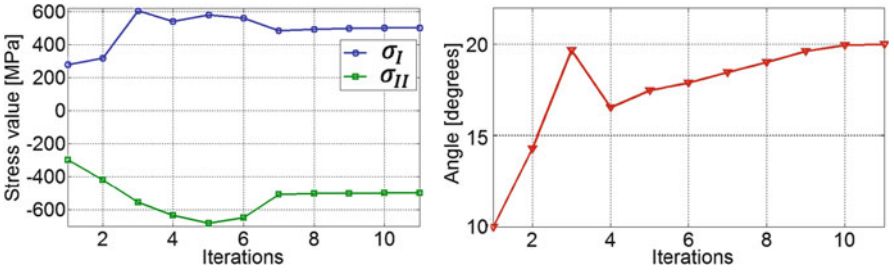


Fig. 3 Sequence of steps of a TRA procedure applied to the estimation of the three parameters which define a residual stress state ($\sigma_I = 500$ MPa, $\sigma_{II} = -500$ MPa, $\varphi = 20^\circ$)

and 20°, respectively). When residual stresses to be estimated exhibit principal directions known a priori, a third indentation (at 45°) is useless in view of the small sensitivities of its experimental data.

Details on a novel method (B) for calibration of anisotropic plasticity models, again by exploiting indentation curve only, can be found in Buljak et al. (2013a).

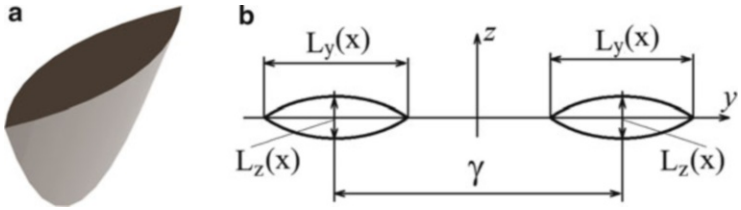


Fig. 4 (**a**) Shape of double sharp indenter. (**b**) The three geometrical parameters to be optimized in the fracture test design

Common features with method (A) are repeated indentations by means of elliptical or bicircular indenters.

A third kind of structural diagnosis can suitably be mentioned here as (C): identification of brittle fracture parameters.

Whereas residual stress states can be estimated also by non-destructive radiation or acoustic (supersonic) tests, identifications of fracture parameters so far are usually performed in laboratories on specimens.

Recently, it has been shown in Buljak et al. (2013b) that the two parameters governing a simple mode-I model of brittle fracture can be estimated economically "in situ" by means of "non-destructive" indentation curves alone, provided that, like in methods (A) and (B), the shape of the instrumented indenter has been optimized again by sensitivity analyses (Stage 4).

The purpose here pursued in the indenter design is to make a crack generation "dominant" among the consequences (elastic and inelastic strains) of the loading process. The novel geometry, which turned out to be suitable for the calibration of a brittle fracture model in glass, is visualized in Fig. 4: the "double bicircular blade" has been designed starting from the conical shape of a classical indenter and modifying the geometry governed by three parameters (one is the distance γ between the two axes, Fig. 4) in order to maximize the difference between pseudoexperimental indentation curves computed on the basis of parameter values at lower and upper bounds suggested by "experts" on the material (Stage 2 in section "A Survey of a Practical Methodology for Damage Assessment in Structures").

For preliminary computational validation of this innovative diagnostic procedure, the piecewise-linear relationship of Fig. 5a (tensile stress t versus crack opening displacement w) has been assumed, namely, the sought parameters are tensile strength t_Y and softening parameter H (or fracture energy G_f).

The discrepancy function $\omega(t_Y, H)$ defined according to Eqs. 1 and 2, Stage 6 in section "A Survey of a Practical Methodology for Damage Assessment in Structures," is graphically shown in Fig. 5b, which visualizes also the itineraries of step-by-step TRA solutions from two initializations to the absolute minimum within the "search domain" (Stages 8–10 in section "A Survey of a Practical Methodology for Damage Assessment in Structures").

It is worth noting that the piecewise-linear (PWL) approximation of Fig. 5a adopted in Buljak et al. (2013b) implies its mathematical representation as "linear

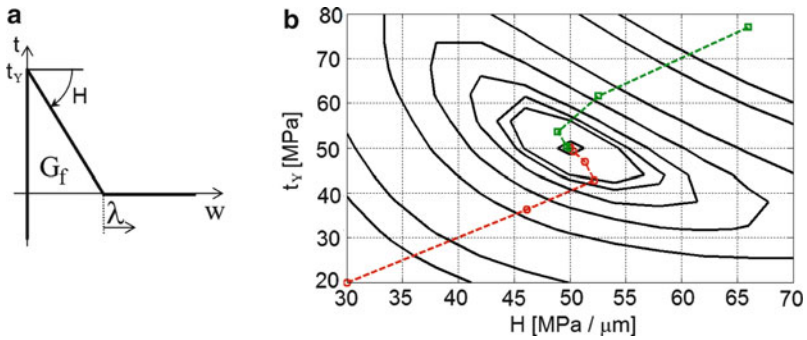


Fig. 5 (**a**) Mode-I cohesive crack model with two parameters to estimate. (**b**) Itineraries of two identification procedures by TRA

complementarity problem" (LCP), and therefore, the discrepancy minimization becomes "mathematical programming under equilibrium constraints" (MPEC), namely, a classical mathematical problem of minimization with possible lack of convexity due to its general formulation.

However, the following advantages emerge from PWL approximation in practical applications: algorithm software available on the market; same mathematical features of direct and inverse analyses when more general and accurate PWL brittle mechanics (or "cohesive crack") models are employed (e.g., more than one straight segment for the approximation of the softening branch; mixed-mode fracture), either with holonomic ("path independent") or non-holonomic assumptions.

Structure Model Calibrations Based on Measurements by both Instrumented Indenter and Imprint Profilometer

A recent development, started in Bolzon et al. (2004), in the structural diagnostic methodology centered on indentation tests is based on the employment of an instrument additional to the instrumented indenter and apt to provide digitalized data on the geometry of the generated residual imprint. Such instrument is a "laser profilometer" for macroscale experiment or an "atomic force microscope" at micro- or nanoscale.

The consequent availability of additional experimental data may provide the following possible advantages to parameter identification by inverse analysis: well-posedness of the minimization problem and convexity of the discrepancy function to minimize, faster computer solutions of the inverse problem and larger numbers of identifiable parameters.

The imprint visualized in Fig. 6a was generated by indentation on copper. Its average profile is compared in Fig. 6b to profiles computed by FE simulations on

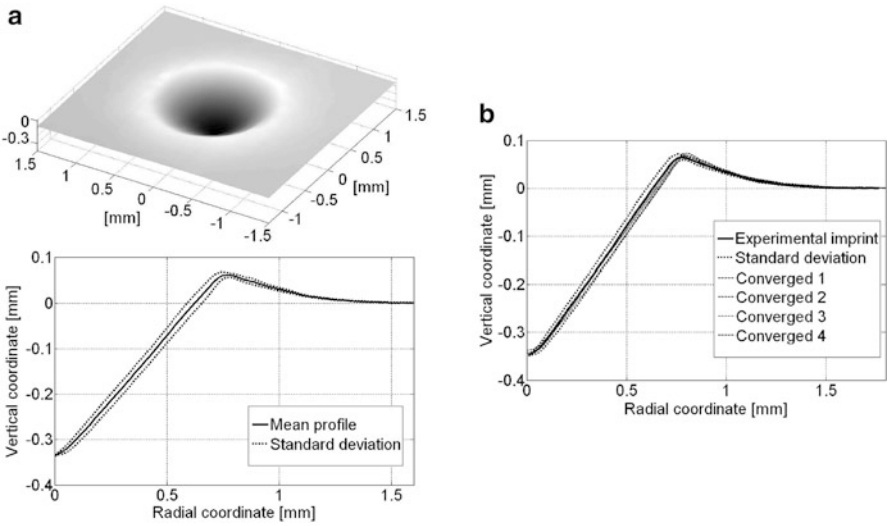


Fig. 6 (**a**) Imprint due to an indentation on copper and relevant radial profile. (**b**) Averaged imprint of profiles achieved by laser compared to those computed by parameter estimates resulting from diverse TRA inverse analyses

the experiment based on parameter estimates achieved by TRA with four different initializations. The inverse analyses concern here Young modulus E, yield stress σ_y and hardening exponent n contained in the classical Ramberg–Osgood elastoplastic isotropic model (e.g., Lubliner (1990)).

Figure 7 (from Bolzon et al. (2011)) provides some comparisons among three diagnostic procedures centered on indentation and on Ramberg–Osgood model, but based on (A) curves only, (B) curves and imprint profiles, and (C) imprint only.

Comparisons with the procedures, outlined in the preceding subsection "Calibration of Tensorial and Fracture Models on the Basis of Indentation Curves Alone," lead, e.g., to the following remarks: residual stresses and anisotropy are reflected by the geometry of the imprint which generally does not exhibit axial symmetry after extraction of popular usual axisymmetric indenters (e.g., Brinell or Rockwell); therefore, the parameter identification may occur on the basis of a single indentation, without repeated rotated tests as specified here earlier. Two instruments obviously mean more costs, and, as for laser profilometers, confinement to laboratories is expected (no longer "in situ"). However, the following prospect turns out to be promising and might be realized in the near future: indentions are performed "in situ" and the imprint shape is carried by a mold to the profilometer in a laboratory. Anyway, the employment of laser profilometers after indentations is potentially useful and practically advantageous in many situations of materials and structural engineering.

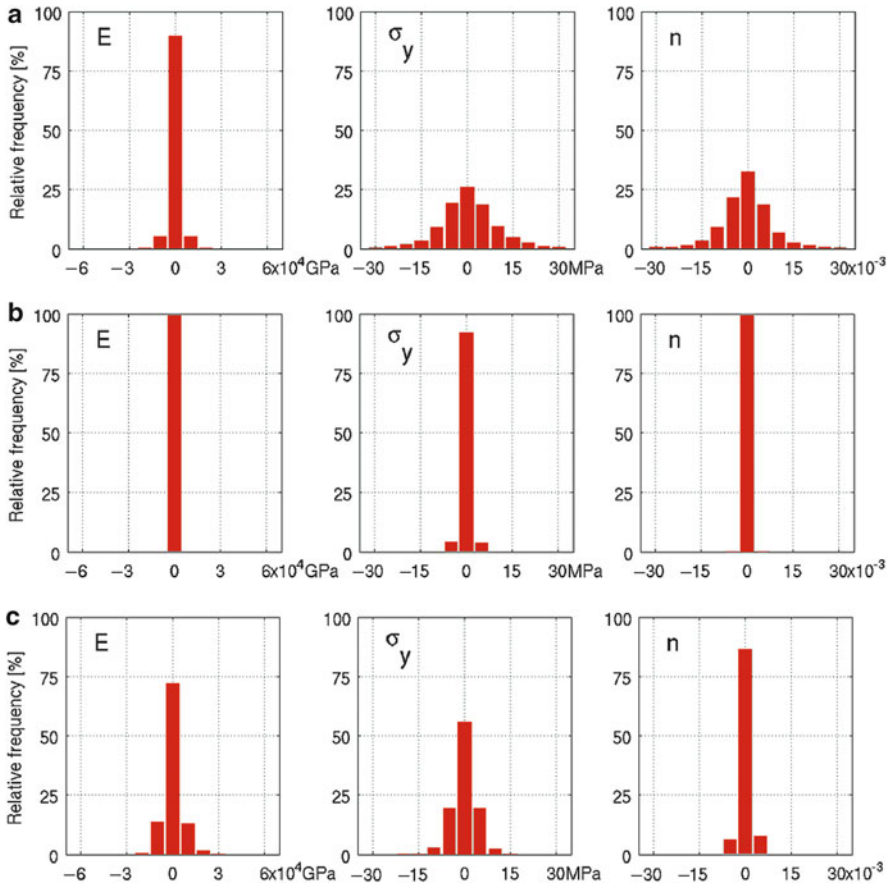


Fig. 7 Error distributions in estimates of Young modulus E, yield limit σ_y and hardening exponent n, achieved by the ANN-centered procedure with approaches A, B, and C: figures (**a**), (**b**), and (**c**), respectively

Damage Assessment in Dam Engineering

Diagnostic Analyses in Elasticity Ranges

At present there are about 45,000 "large" dams in the world (conventionally, "large" means higher than 15 m and/or with reservoir larger than 10^6 m^3). Most of them have been built up in the last century, particularly in decades after the Second World War. Besides possible damages, like in other categories of civil engineering structures (particularly, e.g., due to exceptional external actions like earthquakes), two sources of damages may occur specifically in dams as a slow but irremediable statical deterioration (see, e.g., Ahmed et al. (2003) and Comi et al. (2009)):

(i) "Alkali–Silica Reaction" (ASR) in concrete, namely, a physical–chemical phenomenon due to original wrong choices of ingredients and implying, after a "dormant stage" of years, drastic strength reduction and residual stresses due to local expansions; (ii) slow motions (say half centimeter per year) and possibly eventual instabilities (e.g., like in Vajont disaster) in the surrounding geological formation.

A frequently adopted criterion for diagnosis of concrete dams is based on overall estimations of elastic stiffness: the distribution of Young modulus over the whole dam is reasonably regarded as representative of possible deterioration (including those consisting of diffused cracks). Dynamical excitation by vibrodynes and response measurements by accelerometers provide a context suitable to inverse analyses apt to assess zonewise distribution of present average Young modulus.

An alternative diagnostic method leading to the same estimates (i.e. average local elastic stiffness), but based on statical (instead of dynamical) excitation, has been proposed in Ardito et al. (2008) and is outlined here below.

In most hydroelectrical systems, a substantial change of water level occurs between summer and winter. Such annual fluctuation in the reservoir can be exploited as economical, accurately quantifiable statical loading. The structural response of the dam, obviously in the elastic range, can at present be assessed by a radar instrumentation placed downstream. Many measurements are taken in order to provide displacements at preselected points of the dam surface and are exploited as input of inverse analysis. The parameters to identify are the values of the elastic Young modulus in preselected regions of the dam volume. Clearly, contributions to these displacements arise also from changes of the temperature in the environment and must be assessed by thermal analyses, which can hardly be performed very accurately (Ardito et al. 2008). Simplification and increase of accuracy, but obviously with higher costs, can be achieved by a fast (in a few days) "ad hoc"-designed change of reservoir level (see, e.g., Ardito and Cocchetti (2006)).

Superficial Structural Diagnoses Based on Flat-Jack Tests and "Full-Field" Displacement Measurements

The limitations to the elasticity range of the overall diagnostic methods on concrete dams can be overcome by the local parameter identification procedures outlined in what follows.

"Flat-jack" tests are employed since several decades for estimation of stress states and of stiffness in concrete dams superficially and in masonry structures. Besides the parameter identification methodology presented in this chapter, the main novelties with respect to the current flat-jack practice consist of "full-field" measurements of displacements by "Digital Image Correlation" (DIC), see, e.g., Avril et al. (2008) and Hild and Roux (2006). This modern experimental technique is described in chapters of this handbook.

Recent developments for more effective and fruitful "routine" applications of flat-jack experiments are presented in and outlined here by a list of operative stages on a concrete dam, with reference to Fig. 8.

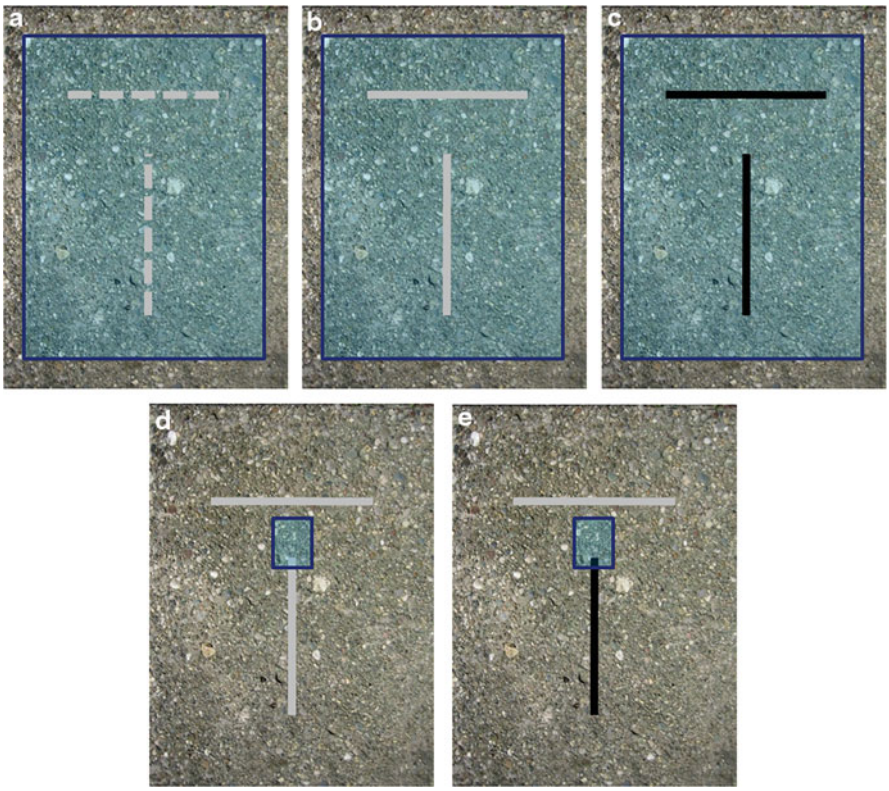


Fig. 8 Sequence of steps in the novel flat-jack tests combined with parameter identification by inverse analysis

(α) On the dam surface, the geometry of two future orthogonal slots is marked and a first photograph is taken over the "Region of Interest" (ROI), shown by a gray rectangle in Fig. 8.

(β) The two slots are generated (Fig. 8b) and a second photo is taken by means of the DIC equipment: thus, the displacements due to the release of the stresses in the cuts are measured at all grid nodes.

(γ) Two flat-jacks are inserted and pressurized (Fig. 8c) and DIC photos are taken again in order to measure the "full field" of new displacements.

(δ) The horizontal jack is removed and the vertical one is depressurized. A reference photo is taken of a (ROI) zone located near the upper end of the vertical slot, in which inelastic deformations are expected to develop in the subsequent step.

(ε) The vertical flat-jack is pressurized in order to generate inelastic strains first (plastic primarily) and later a quasi-brittle-fracturing process near the tips of the loaded slot. A sequence of DIC photos is taken to capture the nonlinear evolution of the displacements due to such loading by further jack pressurization.

The estimation of parameters is performed according to the following sequence: elastic moduli, on the basis of experimental data concerning transition from stage (β) to (γ); stresses, on the basis of the elastic moduli estimates and of data acquired at stages (α) and (β); inelastic parameters, on the basis of data concerning the transition from stage (δ) to the various deformation stages represented in the sequence of DIC photos taken in phase (ε).

The above novel procedure resting on inverse analyses exhibits the following advantages with respect to the present flat-jack practice: more information is provided, including inelastic parameters; more accurate are the estimates; less "destructivity" occurs due to simpler slot geometry. Operatively, the main difficulty consists in the design and realization of supports apt to carry the cameras by avoiding disturbances due to the slot drilling.

Since orthotropy with "transversal" isotropy in the horizontal plane is frequently generated in roller-compacted concrete, the identification procedure presented in Garbowski et al. (2011) concerns the three parameters which play the main role in the system response, namely horizontal E_H and vertical E_V Young modulus and the shear stiffness G_V. The preexisting stress state is assumed plane on the free surface of the dam and uniform over the volume involved in the test; therefore, parameters to identify are three stress components σ_H, σ_V, and τ_{HV}. As for inelastic behaviour, the classical Drucker–Prager model has been adopted (perfect plasticity with nonassociated flow rule) governed by three parameters.

Structural Diagnoses In Depth by Drilling and Dilatometric Tests

In order to compute safety factors of possibly deteriorated concrete dams, with respect to various kinds of failures, assessments of present concrete properties and of self-equilibrated stress states (due to ASR expansion or to slow motions of surrounding geological formations) are necessary also in depth, not only near the free surfaces. The in-depth material characterization is a well-developed topic in rock mechanics (see, e.g., Wittke (1990)). A diagnostic method, proposed in Zirpoli et al. (2008), can be outlined as follows.

(α) A perforation is performed by a traditional drilling equipment up to a preselected depth.
(β) A device called "dilatometer" is inserted; it consists of two sleeves equipped with radial displacement gauges and, between them, two movable steel "arches."
(γ) The drilling goes ahead, while the gauges measure the displacements due to the consequent stress relief.
(δ) Hydraulic jacks inside the equipment apply two growing radial forces to the two rigid steel "arches" and the gauges measure their displacements as linear-elastic response of the region surrounding the hole.
(ε) The elastic limit is overcome by increasing the jack pressure and plastic deformation occurs while monitoring goes on.

(ζ) Another instrumental equipment is inserted in the hole with blades instead of "arches": again small jacks push these sharp indenters against the cylindrical wall in order to promote dominant quasi-brittle fractures to be exploited in a procedure similar to that mentioned in section "Parameter Identifications by Indentation Tests" with reference to glass.

The sought parameters can be estimated in the following sequence: Young modulus and Poisson ratio, using the experimental data collected during phase (δ); the initial stresses, two normal and one tangential, in the plane orthogonal to the hole axis, by employing the estimated elasticity parameters and data coming from phase (γ); parameters governing a plastic constitutive model and/or a quasi-brittle fracture model, using the previous estimates and measurements in phase (ε) and/or (ζ).

With respect to the state-of-the-art diagnosis of concrete dams, the above diagnostic procedure is motivated by the following advantages:

(i) no specimen is extracted from the borehole to be tested in laboratory;
(ii) displacements are measured, not strains which are sensitive to local material properties (quite different from mortar to aggregate);
(iii) inelastic properties can be assessed "in situ" and in depth;
(iv) a portable computer, containing software elaborated "a priori" (by POD outlined in section "A Survey of a Practical Methodology for Damage Assessment in Structures") through FE simulations of the mechanical test, performs inverse analyses on the basis of the displacement data coming from the gauges.

The above outlined mechanical characterization procedures of material behaviours in depth provide obvious practical advantages if compared to the present techniques of "coring" and "overcoring," traditional and still fashionable in geotechnical engineering. Reliable quantitative assessment provided by inverse analyses on geological formations might contribute in the near future to predict earthquake in some areas, to avoid damage events like Vajont disaster (1963) and Macondo borehole collapse in Gulf of Mexico (2010). The latter historical event has led to a gigantic 10-year research project (starting in 2013, coordinated by the two National Academies in the USA). In this project, one of many purposes might be the realization of experimental equipment apt to mechanically characterize the crossed geological layers, in addition to the multiple instrumentation employed at present by the offshore industries for deep perforations (which at present sometimes exceed 12 km depth in some areas).

Inverse Analysis Procedures for Mechanical Characterizations of Foil Products

Industrial products consisting of thin layers in a plane, like textile membranes and laminates, are at present often mechanically characterized by means of laboratory experiments with the following features: cruciform specimen with branches on

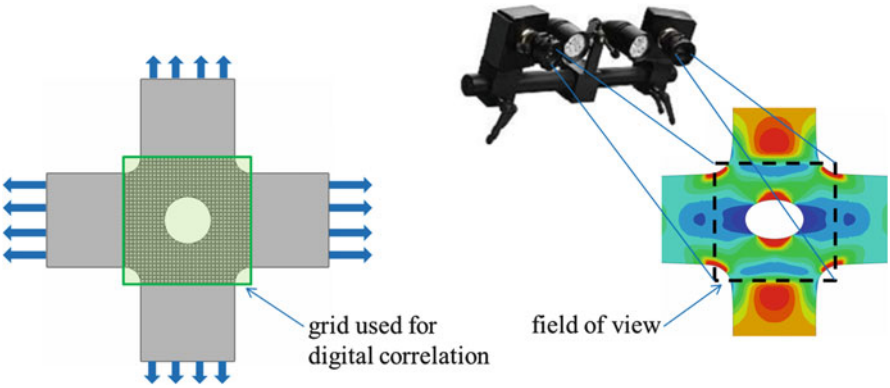


Fig. 9 Cruciform test on a membrane with full-field measurements of inhomogeneous displacements

which axial elongations are imposed by clamps and consequent "loads" are measured; measurements of two-dimensional displacements in the nodes of a grid by digital image correlation (DIC). "Full-field" assessment of displacements may confer well-posedness, particularly if combined with their inhomogeneity generated by a central hole shown in Fig. 9.

An elastic–plastic hardening associative constitutive model has been elaborated for paper and paperboard foils at MIT (Xia et al. 2002). A simplification of this material model proposed in Garbowski et al. (2012) reduces from 27 to 17 the number of parameters. In view of routine applications in an industrial environment, "feed-forward" artificial neural networks (ANNs) have been considered in Garbowski et al. (2012). Generally, for the design and the computational behavior of ANN, a balance is desirable between the dimensionalities of parameter vector **p** and of measurable quantities vector **u**. In the present context, the number of experimental data, i.e., the dimensionality of vector **u** containing displacement measurements by DIC, turns out to be by orders of magnitude larger than the dimension of the parameter vector **p**. Therefore, the role of ANN input is attributed here again, like in section "Parameter Identifications by Indentation Tests," to "amplitude" vector **a** which approximates the information contained in the "snapshot" **u** by compressing it through the POD procedure employed in the preceding section. The role of vector **a** is twofold: the preliminary generation of the ANN by means of the "patterns" ($\mathbf{p}_i$, $\mathbf{a}_i$) and the input of the ANN for the estimation of the parameters **p** on the basis of a test on cruciform specimen with DIC measurements.

An obvious difficulty intrinsic to the POD + ANN method is that the growth of the parameter space dimensionality implies exponential growth of the number of grid nodes over the search domain. With reasonable snapshot number in the preliminary POD computations, the density of nodes over the search domain may become low and this can jeopardize the accuracy of the estimates provided by the trained ANN. However, as already underlined in the preceding sections, the

increase of snapshot number N concerns only the preparatory computations to be done once for all and, hence, is quite possible in real-life applications.

The mechanical characterization of free foils by experiments with compression in their planes is required by the following circumstances: walls of food containers are often significantly compressed during transportation; anisotropic constitutive models like the ones considered in what precedes involve parameters which cannot be identified by tests with dominant tension. A novel experimental equipment ("sandwich system") and inverse analysis procedure have been proposed in Cocchetti et al. (2012). A rectangular specimen of the foil is inserted between two stabilizing elastic "blocks." The external actions consist of rotations imposed by two rigid clamps, apt to generate a chosen combination of compression and bending. At each step of the planned rotations, many displacements are measured by the DIC technique, both on the blocks and on the emerging specimen. FE computer simulations are performed, first for sensitivity analyses apt to design and optimize the procedure; subsequently, the POD + RBF + TRA procedure is applied to the estimation of parameters. A "fictitious homogeneous material" is attributed to the foil (even if it is a layered laminate) and its behaviour is described by an anisotropic elastic–plastic model. Homogeneity assumption is suggested by unpredictable changes in local properties of layers and interfaces due to the production processes.

Conclusion

The damages considered in this chapter primarily consist of unexpected changes in mechanical properties of materials and structures, changes which may reduce the margins of safety in service of industrial products, of structural components in power plants and civil engineering buildings and of structures in a variety of other engineering areas.

Such changes concern parameters which are included in the input of each structural "direct analysis" apt to assess the above safety margins. Clearly, back-analyses are needed in many other technical contexts without damages like assessment of conductivity in thermal problems.

Identifications of these damage parameters, usually contained in material models, by means of suitable experiments and "inverse analyses" based on them, can be at present considered as a growing area of applied mechanics. Within this area, the procedures outlined in the present chapter, by numerical examples as well, concern the damage assessments (or "structural diagnosis") by experiments listed here as conclusions: nondestructive indentation tests, with or without profilometric measurements as additional sources of data; flat-jack tests and dilatometric tests with drilling, superficially and in depth, respectively, for mechanical characterizations of concrete dams and of geomaterial or geological formation crossed by perforation instruments; calibration of anisotropic constitutive models for multilayer laminates by cruciform tests.

The inverse analysis methodology and the above listed applications have been confined to the following features in view of contents of other chapters in this handbook: deterministic approaches, macroscale modelling, and statical loads.

However, the methodological features outlined in section "A Survey of a Practical Methodology for Damage Assessment in Structures" (sensitivity analyses for test design, "full-field" measurements, model reduction by "Proper Orthogonal Decomposition," repeated interpolations instead of finite element simulations) and exemplified by the few abovementioned applications in the other sections can be regarded as innovative and potentially advantageous in a much broader engineering field.

References

T. Ahmed, E. Burley, S. Rigden, A.I. Abu-Tair, The effect of alkali reactivity on the mechanical properties of concrete. Construct. Build Mater. **17**(2), 123–144 (2003)

R. Ardito, G. Cocchetti, Statistical approach to damage diagnosis of concrete dams by radar monitoring: formulation and a pseudo-experimental test. Eng. Struct. **28**(14), 2036–2045 (2006)

R. Ardito, G. Maier, G. Massalongo, Diagnostic analysis of concrete dams based on seasonal hydrostatic loading. Eng. Struct. **30**, 3176–3185 (2008)

S. Avril, M. Bonnet, A.-S. Bretelle, M. Grédiac, F. Hild, P. Ienny, F. Latourte, D. Lemosse, S. Pagano, E. Pagnacco, F. Pierron, Overview of identification methods of mechanical parameters based on full-field measurements. Exp. Mech. **48**(4), 381–402 (2008)

S.J. Bates, J. Sienz, V.V. Toropov, Formulation of the optimal Latin hypercube design of experiments using a permutation genetic algorithm, in *Collection of Technical Papers – AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, vol. 7 (2004), pp. 5217–5223. Curran Associates, Palm Springs, California

G. Bolzon, G. Maier, M. Panico, Material model calibration by indentation, imprint mapping and inverse analysis. Int. J. Solids Struct. **41**(11–12), 2957–2975 (2004)

G. Bolzon, V. Buljak, G. Maier, B. Miller, Assessment of elastic-plastic material parameters comparatively by three procedures based on indentation test and inverse analysis. Inverse Probl. Sci. Eng. **19**(6), 815–837 (2011)

M.D. Buhmann, *Radial basis functions* (Cambridge University Press, Cambridge, 2003)

H.D. Bui, *Inverse Problems in the Mechanics of Materials: An Introduction* (CRC Press, Boca Raton, 1994)

H.D. Bui, *Fracture mechanics: inverse problems and solutions* (Springe, Dordrecht, 2006)

V. Buljak, *Inverse Analyses with Model Reduction: Proper Orthogonal Decomposition in Structural Mechanics* (Springer, New York, 2012)

V. Buljak, G. Maier, Identification of residual stresses by instrumented elliptical indentation and inverse analysis. Mech. Res. Commun. **41,** 21–29 (2012). Erratum, **46**, 90 (2012)

V. Buljak, M. Bocciarelli, G. Maier, Mechanical characterization of anisotropic elasto-plastic materials by indentation curves only. Meccanica (2014). doi: 10.1007/s11012-014-9940-y. Article in Press

V. Buljak, G. Cocchetti, G. Maier, Calibration of brittle fracture models by sharp indenters and inverse analysis. Int. J. Fract. **184**(1-2), 123–136 (2013).

A. Chatterjee, An introduction to the proper orthogonal decomposition. Curr. Sci. **78**(7), 808–817 (2000)

G. Cocchetti, M.R. Mahini, G. Maier, Mechanical characterization of foils with compression in their plane. Mech. Adv. Mater. Struct. (2013). doi: 10.1080/15376494.2012.726398. Article in Press

T.F. Coleman, Y. Li, An interior trust region approach for nonlinear minimization subject to bounds. SIAM J. Optim **6**(2), 418–445 (1996)
C. Comi, R. Fedele, U. Perego, A chemo-thermo-damage model for the analysis of concrete dams affected by alkali-silica reaction. Mech. Mater. **41**(3), 210–230 (2009)
A.R. Conn, N.I.M. Gould, P.L. Toint, *Trust-Region Methods* (Society for Industrial and Applied Mathematics, Philadelphia, 2000)
M. Dao, N. Chollacoop, K.J. Van Vliet, T.A. Venkatesh, S. Suresh, Computational modeling of the forward and reverse problems in instrumented sharp indentation. Acta Mater. **49**(19), 3899–3918 (2001)
T. Garbowski, G. Maier, G. Novati, Diagnosis of concrete dams by flat-jack tests and inverse analyses based on proper orthogonal decomposition. J. Mech. Mater. Struct. **6**(1–4), 181–202 (2011)
T. Garbowski, G. Maier, G. Novati, On calibration of orthotropic elastic-plastic constitutive models for paper foils by biaxial tests and inverse analyses. Struct. Multidiscip. Optim. **46**, 111–128 (2012)
J. Gribbin, M. Gribbin, *Richard Feynman, a life in science* (Dutton, New York, 1997)
M.T. Hagan, H.B. Demuth, M.H. Beale, *Neural network design* (PWS Publishing, Boston, 1996)
S. Haykin, *Neural networks: a comprehensive foundation*, 2nd edn. (Prentice Hall, Upper Saddle River, 1998)
F. Hild, S. Roux, Digital image correlation: from displacement measurement to identification of elastic properties – a review. Strain **42**(2), 69–80 (2006)
M. Jirásek, Z.P. Bažant, *Inelastic analysis of structures* (Wiley, New York, 2001)
E.J. Kansa, Motivations for using radial basis functions to solve PDEs (2001), http://rbf-pde.uah.edu/kansaweb.pdf
M. Kleiber, H. Antúnez, T.D. Hien, P. Kowalczyk, *Parameter Sensitivity in Nonlinear Mechanics. Theory and Finite Element Computations* (Wiley, New York, 1997)
C.G. Koh, M.J. Perry, *Structural Identification and Damage Detection using Genetic Algorithms* (CRC Press, New York, 2009)
S. Kucharski, Z. Mróz, Identification of material parameters by means of compliance moduli in spherical indentation test. Mater. Sci. Eng. A **379**(1–2), 448–456 (2004)
M. Levy, M. Salvadori, *Why buildings fall down* (W. W. Norton, New York, 1992)
J. Lubliner, *Plasticity theory* (Macmillan, New York, 1990)
Z. Mróz, G.E. Stavroulakis (eds.), *Parameter identification of materials and structures* (Springer, New York, 2005)
A. Nouy, A priori model reduction through Proper Generalized Decomposition for solving time-dependent partial differential equations. Comput. Methods Appl. Mech. Eng. **199**(23–24), 1603–1626 (2010)
W.C. Oliver, G.M. Pharr, An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. J. Mater. Res. **7**(6), 1564–1583 (1992)
Z. Ostrowski, R.A. Białecki, A.J. Kassab, Solving inverse heat conduction problems using trained POD-RBF network inverse method. Inverse Probl. Sci. Eng. **16**(1), 35–54 (2008)
D. Ruckelynck, F. Chinesta, E. Cueto, A. Ammar, On the a priori model reduction: overview and recent developments. Arch. Comput. Methods Eng. **13**(1), 91–128 (2006)
A. Tarantola, *Inverse problem theory and methods for model parameter estimation* (Society for Industrial and Applied Mathematics, Philadelphia, 2005)
F.A.C. Viana, G. Venter, V. Balabanov, An algorithm for fast optimal Latin hypercube design of experiment. Int. J. Numer. Methods Eng. **82**(2), 135–156 (2010)
Z. Waszczyszyn, *Neural Networks in the Analysis and Design of Structures* (Springer, New York, 1999)
W. Wittke, *Rock mechanics* (Springer, New York, 1990)
Q.S. Xia, M.C. Boyce, D.M. Parks, A constitutive model for the anisotropic elastic-plastic deformation of paper and paperboard. Int. J. Solids Struct. **39**(15), 4053–4071 (2002)
A. Zirpoli, G. Maier, G. Novati, T. Garbowski, Dilatometric tests combined with computer simulations and parameter identification for in-depth diagnostic analysis of concrete dams, in *Proceedings of the* 1st*International Symposium on Life-Cycle Civil Engineering (IALCCEE-08)*, eds. by F. Biondini, D.M. Frangopol (2008), pp. 259–264, Taylor & Francis Group, London

Section V

Damage Mechanics in Electronic Packaging

Adhesion and Failure of Polymer-Metal Interfaces in Microelectronic Packaging

21

Jianmin Qu

Contents

Abstract

This chapter focuses on polymer/metal interfaces. Attention is given to two common aspects of tremendous practical interest, namely, surface roughness and moisture. Debonding of polymer/metal interfaces often involves both interfacial and cohesive failure. Since the cohesive strength of polymers is usually much greater than the polymer/metal interfacial strength, cohesive failure near the interface is usually desired to enhance the interfacial adhesion. Roughened surfaces generally produce more cohesive failure and, therefore, are used commonly in practice to obtain better adhesion. In this chapter, a fracture mechanics model is developed that can be used to quantitatively predict the amount of cohesive failure once the surface roughness data are given. Moisture, on the other hand, tends to

J. Qu (✉)
Department of Civil and Environmental Engineering, Department of Mechanical Engineering, Northwestern University, Evanston, IL, USA
e-mail: j-qu@Northwestern.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_27

degrade the interfacial strength. To quantify such degradation, a systematic and multidisciplinary study is conducted to better understand the fundamental science of moisture-induced degradation of interfacial adhesion. The approach is comprised of both experimental and modeling components of analysis and addresses some of the key issues needed to advance the understanding of the effect of moisture on interfacial adhesion.

Introduction

Polymer/metal interfaces are ubiquitous. They can be found in numerous engineering applications. For example, polymers are used as adhesives and coatings for metals in aircrafts, automobiles, microelectronics, MEMS, etc., and polymer composites are routinely used for retrofitting damaged structures. The need for the miniaturization of microelectronic devices has led to the emerging technology of 3D packaging, in which numerous metal layers are bonded together by polymer dielectric adhesives. The reliability of microelectronics devices (cell phone, computers, etc.), to a large extent, depends on the integrity of these material interfaces, because they are often the "weakest links" and their failure often leads to the malfunction of the entire electronic device. Therefore, understanding the structure–performance relationship of polymer/metal interfaces is critical in designing, building, and operating various structures, components, and devices.

Bonding of a polymer to a metal may invoke several mechanisms. Some are chemical, others are physical and mechanical. At intimate contact between the polymer molecules and metal atoms, chemical bonding is generated through charge transfer processes, such as the hydrogen bonding, the acid–base interactions, and the electron pair donor–acceptor interactions which may involve both covalent and ionic forces. Chemical bonding can be significantly improved by coupling agents at the interface.

The van der Waals faces originated from the molecular dipole interactions give rise to physical adsorption in which wetting is necessary and essential. When integrated over the two mating surfaces, van der Waals faces turn out to be a weak but long range interaction, decreasing as r^{-2} or r^{-3} beyond 10 nm. Since adsorption is believed to be one of the most important mechanisms in achieving adhesion, diffusion and wetting are critical to attain good bonding.

Dissipation is the major mechanical adhesion mechanism at the microscopic scale. The two main dissipation mechanisms are the viscoelastic energy loss due to the translational motion of the polymer chains, and the plastic deformation near the delamination crack tip. The latter plays a major role in interfacial fracture mechanics, which is a subject studied extensively in many engineering applications. Chemical and physical adhesion provides the molecular scale bonding forces between the polymer and the metal. The macroscopic strength of the interface, however, depends on not only the molecular scale bonding forces but also flaws and defects on the interface. Such flaws and defects cause stress concentration which greatly reduces the macroscopic interfacial strength. Fracture mechanics is then needed to assess the integrity of the interface.

For many polymer/metal interfaces of practical interest, their fracture behavior can be captured by the linear elastic fracture mechanics or small-scale yielding elastic-plastic fracture mechanics. In both cases, the crack-tip stress field is scaled by a mixed mode stress intensity factor (SIF). The critical value of the SIF at which the crack starts to propagate defines the interfacial fracture toughness, which may also depend on the load mixity. In engineering applications today, the only way to obtain such mixed mode fracture toughness is through experimental measurements. In other words, there is no methodology that is capable of predicting the interfacial fracture toughness from the material's atomistic structure.

At the mesoscopic scale, mechanical interlocking may significantly improve interfacial strength, which can be achieved by various surface treatments of the metals to provide desired surface roughness and topologies (e.g., Lee and Qu 2003). It is to be noted that mechanical interlocking, strictly speaking, is not one of the adhesion mechanisms, at least not at the molecular level. It is only a technological means in achieving adhesive bonding.

This chapter focuses on two issues that are most critical to the strength of polymer/metal interfaces, namely, surface roughness and moisture.

Effects of Surface Roughness

Metal surfaces are microscopically rough. When a liquid state or gelatinous adhesive is applied to a rough surface, it conforms to the rough surface and tends to fill up the irregularities of the substrate surface such as microgroves, holes, or dips. Consequently, mechanical interlock forms after the adhesive is cured. For example, the anodization of aluminum alloys produces a deeply porous topography with many open porous structures, the adhesive typically penetrates to virtually the bottom of the pores, and so a "composite" interfacial region is created (Kinloch 1987). This composite region will have a modulus and strength intermediate between those of the polymeric adhesive and the aluminum oxide, and this would be expected to be beneficial from the viewpoint of joint strength and toughness. Such adhesion enhancement due to surface roughness has been demonstrated by many studies in the literature.

For polymer/metal interfaces, the mechanism of adhesion enhancement due to roughness is associated with the fact that separation of the interface does not occur completely along the interface line. Due to surface roughness, cohesive failure of the polymer adhesive occurs near the interface as the adhesive is peeled away from the substrate, resulting in the crack path deviating from the interface. Such deviation of crack path away from the interface usually requires additional energy associated with the crack propagation within the polymer adhesive. Therefore, in order to quantify the adhesion enhancement, this additional energy needs to be evaluated.

To this end, consider a polymer/metal interface with an idealized interface profile as shown in Fig. 1. The relevant dimensions are the height $2R$ and the half-wavelength λ. Furthermore, it is necessary to introduce a nondimensional parameter η so that

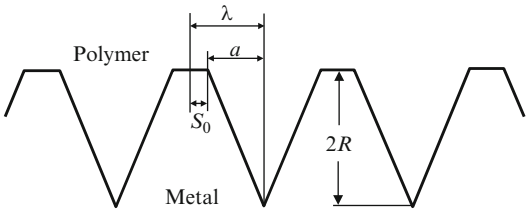


Fig. 1 An idealized interface profile

$$S_0 = \eta\lambda. \tag{1}$$

Obviously, η is limited by $0 < \eta < 1$. These three parameters, R, λ, and η, are sufficient to describe the interface profile uniquely.

In reality, rough surfaces are random and can only be characterized by statistical averages. In this case, the parameters R, λ can be chosen as the mean values of the random variable. Techniques to measure and evaluate these means are discussed in Yao and Qu.

Using the idealized interface profile described above, one may assume that interfacial failure first occurs in the flat area due to relatively weak interfacial adhesion. As the crack propagates along the interface (Fig. 2a), the driving force at the crack tip changes due to different mode mixity. Consequently, the interfacial crack may be deflected into the polymer material (Fig. 2b). This crack deflection results in the cohesive failure of the polymer material (Fig. 2c). Therefore, the objective here is to identify the conditions for crack deflection.

Let the fracture toughness of the interface be G_{ic} and that of the polymer be G_{pc}. Then, based on the linear elastic fracture mechanics, the crack is likely to continue along the interface if the condition $G_i \geq G_{ic}$ is met, where G_i is the crack-tip energy release rate for a crack along the interface (Fig. 2a). On the other hand, the crack is likely to be deflected into the polymer material if the condition $G_p \geq G_{pc}$ is met, where G_p is the crack-tip energy release rate for the crack shown in Fig. 2b. Since the fracture toughness of the metal is much higher than that of the polymers and the interface, the possibility of the crack kinking into the metal may be ruled out.

To formulate the above discussion into a mathematical form convenient for quantitative analysis, let us define the energy release rate ratio

$$G_R = \frac{G_i}{G_p}. \tag{2}$$

Then, the conditions for the crack to select its paths can be formulated as

$$G_R = \frac{G_i}{G_p} > \frac{G_{ic}}{G_{pc}} \Rightarrow \text{cracking along the interface,} \tag{3}$$

$$G_R = \frac{G_i}{G_p} < \frac{G_{ic}}{G_{pc}} \Rightarrow \text{cracking into the polymer.} \tag{4}$$

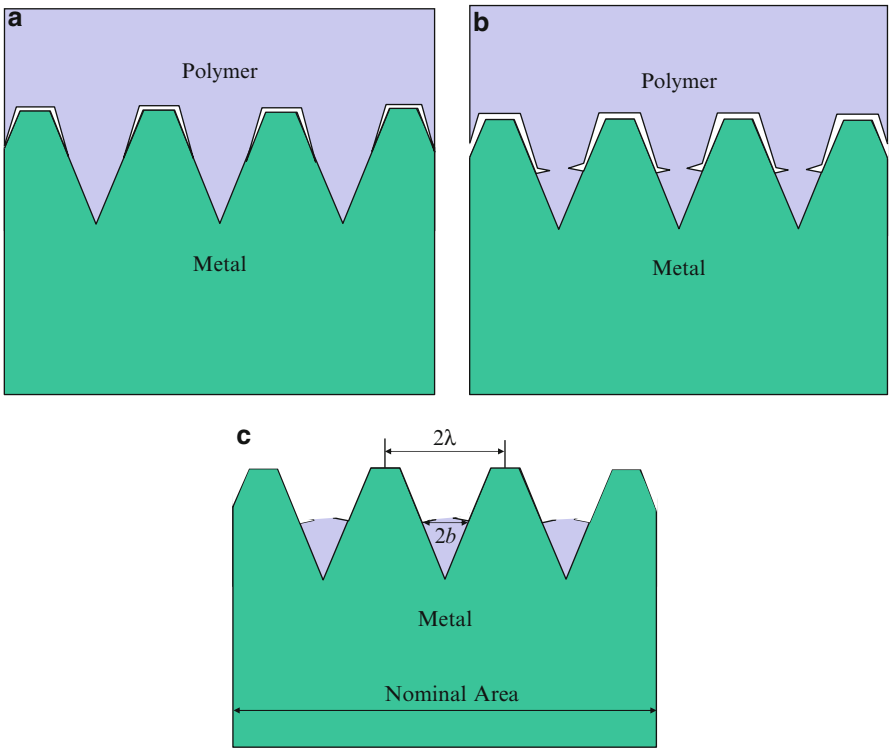


Fig. 2 Interfacial and cohesive failure

So, the condition for impending crack branching is

$$G_R = \frac{G_{ic}}{G_{pc}}. \tag{5}$$

The point for crack kinking into the adhesive can then be determined from Eq. 5 once the quantities on both sides of Eq. 5 are determined.

Note that G_{ic} and G_{pc} are material properties. Although G_{ic} is usually a function of loading phase angle, it is nevertheless given for a specified interface. In other words, the right-hand sides of Eq. 5 are known for a given interface. The energy release rate ratio, G_R, on the other hand, is a function of loads. In linear elastic fracture mechanics, the energy release rate represents the amount of energy available at the crack tip for crack growth. In general, for a given geometry and material pair, the energy release rate ratio G_R can be computed for any given loading. Yao and Qu developed an approximate formula for G_R

$$G_R = \frac{\lambda^2}{2\gamma R(\lambda - b - S_0)} \frac{g_i}{g_p}, \tag{6}$$

where γ is a nondimensional fitting parameter,

$$g_i = \left(\frac{1-\nu_1}{\mu_1} + \frac{1-\nu_2}{\mu_2}\right) \frac{1}{4\cosh^2(\pi\varepsilon)}, \quad g_p = 1.25\pi \frac{1-\nu_2}{2\mu_2}, \tag{7}$$

where ν_n and μ_n are the Poisson's ratio and shear moduli of the polymer ($n = 1$) and the metal ($n = 2$), respectively. The nondimensional parameter ε is given by

$$\varepsilon = \frac{1}{2\pi} \ln\left(\frac{1-\beta}{1+\beta}\right) \tag{8}$$

with β being the second Dundur's bimaterial constant:

$$\beta = \frac{1}{2} \frac{(1-2\nu_2)/\mu_2 - (1-2\nu_1)/\mu_1}{(1-\nu_2)/\mu_2 + (1-\nu_1)/\mu_1}. \tag{9}$$

Substituting Eq. 6 into Eq. 5 and rearranging the terms yield

$$1 - \frac{b}{\lambda} = \frac{F}{2\gamma} \frac{\lambda}{R} + \eta, \tag{10}$$

with

$$F = \frac{g_i G_{pc}}{g_p G_{ic}}. \tag{11}$$

It is seen from Fig. 2c that b/λ is the percent of area covered by the residual polymer as a result of local cohesive failure relative to the total nominal area (horizontal projection of the actual surface area). Consequently, $1 - b/\lambda$ represents interfacial failure as a percent of the nominal area.

Furthermore, it is easy to see that Eq. 10 has a solution for b in the range of $0 \leq b < \lambda$ only if

$$R \geq \frac{F\lambda}{2\gamma(1-\eta)}. \tag{12}$$

When Eq. 12 is not satisfied,

$$G_R = \frac{\lambda^2}{2\gamma R(\lambda - b - S_0)} \frac{g_i}{g_p} > \frac{G_{ic}}{G_{pc}},$$

which means that the crack does not kink into the polymer material. Instead, it will continue to grow along the interface leading to pure interfacial failure. Thus, the area of interfacial failure as a percentage of the nominal interface area is given by

$$S(R,\lambda,F,\eta,\gamma) = \begin{cases} \left(1 - \frac{b}{\lambda}\right) = \frac{F}{2\gamma}\frac{\lambda}{R} + \eta, R \geq \frac{F\lambda}{2\gamma(1-\eta)} \\ 1 \qquad\qquad\qquad\qquad\quad , R \geq \frac{F\lambda}{2\gamma(1-\eta)} \end{cases} \tag{13}$$

Since the roughness is random, the profile height $2R$ and the half-wavelength λ are random variables. Assume the surface is Gaussian and the probability distributions for R and λ are given, respectively, by

$$p_R(R) = \frac{1}{\sigma_R\sqrt{2\pi}} \exp\left[-\frac{\left(R-\overline{R}\right)^2}{2{\sigma_R}^2}\right], \tag{14}$$

$$p_\lambda(R) = \frac{1}{\sigma_\lambda\sqrt{2\pi}} \exp\left[-\frac{\left(\lambda-\overline{\lambda}\right)^2}{2{\sigma_\lambda}^2}\right], \tag{15}$$

where the means $\overline{R}$, $\overline{\lambda}$ are discussed in the Appendix and σ_R, σ_λ are the standard deviations from the means.

The statistical average of the percent of interfacial failure is thus given by

$$\overline{S} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(R,\lambda,F,\eta,\lambda)p_R(R)p_\lambda(\lambda)d\lambda dR. \tag{16}$$

Substituting Eq. 13 into Eq. 16 and carrying out the integration with respect to λ yields

$$\overline{S}\left(\overline{R},\overline{\lambda},F,\eta,\gamma\right) = \int_{-\infty}^{\infty} S\left(R,\overline{\lambda},F,\eta,\gamma\right)p_R(R)dR. \tag{17}$$

This is the mean (average) percent of interfacial failure on a rough polymer/metal interface.

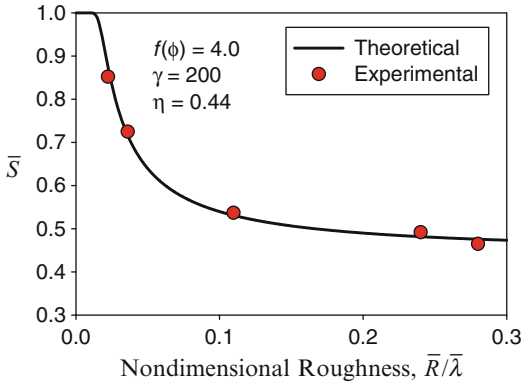


Fig. 3 Comparison between experimental and theoretical results

It is seen that the $\overline{S}$ given by Eq. 17 depends on several parameters. First, it is a function of the material parameter F defined in Eq. 11. This parameter is a combination of elastic stiffness and fracture toughness of both materials and the fracture toughness of their interface. In principle, all these properties are measurable for given materials and their interfaces. Since G_{ic} is a function of the phase angle, F varies with the crack-tip phase loading angle as well. For example, for epoxy and aluminum, $g_i \approx 0.015$/GPa and $g_p \approx 1.064$/GPa, respectively. The cohesive fracture toughness of the epoxy is $G_{pc} \approx 7.55\text{kJ/m}^2$ and G_{ic} ranges between 20 ~ 40 J/m^2 for the epoxy–aluminum interface used in this study. Thus, F ranges approximately from 2.7 to 5.3 for aluminum–epoxy interfaces depending upon the loading phase angle.

Second, $\overline{S}$ depends on the interface roughness profile described by $\overline{R}$, $\overline{\lambda}$, and η. Among them, $\overline{R}$ and $\overline{\lambda}$ can be easily measured using conventional profilometers. The nondimensional parameter η is a consequence of the idealization of the random surface profile. Generally, it needs to be determined indirectly.

Finally, $\overline{S}$ is also a function of γ, an ad hoc parameter introduced in the approximate solution of the energy release rate for the interface crack problem. It should be pointed out that if an "exact" numerical solution were used for the interface crack problem, it would be unnecessary to introduce γ.

Based on the above discussion, it may be concluded that once the material and interface properties (including the interface profile) are given, the percent of nominal area having purely interfacial failure can be predicted by Eq. 33. In this simple model, two parameters (γ and η) need to be determined empirically.

Experimental work was also carried out for various degrees of surface roughness. Figure 3 shows the experimental data together with the theoretically predicted values using Eq. 17. It is seen that the agreement is excellent. The parameters used in the theoretical analysis are also indicated in the figure.

It is interesting to notice the two extreme values of the roughness. For very roughness surface ($\overline{R} \gg 1$), it follows from Eq. 13 that the curve approaches its asymptote of $\overline{S} \to \eta = S_0/\lambda$, which is the ratio of the plateau area and the entire

nominal area. This means that interfacial failure occurs only in the plateau area. On the other hand, for extremely smooth surfaces ($\overline{R} \ll 1$), it follows from Eq. 13 that $\overline{S} = 1$. This means that for very smooth surfaces, interfacial failure occurs over the entire surface. In this case, there is no adhesion enhancement. The roughness threshold for any adhesion enhancement is $R_{th} = F\lambda/2\lambda(1 - \eta)$, according to Eq. 13. As a design guideline, it is interesting to note that the adhesion enhancement increases very rapidly once the roughness exceeds the threshold value. The increase tapers off quickly as the roughness further increases. The maximum adhesion enhancement from roughness is limited by the parameter η. The smaller the η, the better. Physically, this means shape asperities on the roughness surface may yield greater adhesion enhancement.

Effects of Moisture

A significant problem in the microelectronic packaging industry is the presence of moisture-induced failure mechanisms. Moisture is a multidimensional concern in packaging, having an adverse effect on package reliability by introducing corrosion, development of hygro-stresses, and degradation of polymers present in the package. Moisture can also accelerate delamination by deteriorating the polymer interfaces within the package. As the interfacial adhesion between the chip, underfill, and substrate decreases, the likelihood of delamination at each encapsulant interface increases. Once the package delaminates, the solder joints in the delaminated area are exposed to high stress concentrations, resulting in a reduction of overall package life.

Moisture can affect interfacial adhesion through two primary mechanisms. The first mechanism is the direct presence of moisture at the interface altering the interfacial integrity of the adhesive joint. The second mechanism is the absorbed moisture in either the adhesive or substrate altering the mechanical properties of those materials, which changes the response of the adhesive structure in the presence of an externally applied load. Inevitably, the effect of moisture on the adhesion and fracture of interfaces entails a multidisciplinary study, and several aspects should be considered. From a global perspective, the primary aspects include moisture transport behavior, changes in bulk material properties from moisture absorption, effect of moisture on interfacial adhesion, and recovery from moisture upon fully drying, although several subsections within each major group occur due to the complexity of the problem.

In this chapter, a systematic and multidisciplinary study is presented to address the fundamental science of moisture-induced degradation of interfacial adhesion. First, the moisture transport behavior within underfill adhesives is experimentally characterized. The results are incorporated into a finite element model to depict the moisture ingress and interfacial moisture concentration after moisture preconditioning. Second, the effect of moisture on the variation of the adhesive elastic modulus is demonstrated and the physical mechanisms for the change identified. Third, the aggregate effect of moisture on the interfacial fracture toughness is determined. This includes the primary

effect of moisture being physically present at the interface and the secondary effect of moisture changing the elastic modulus of the adhesive when absorbed. Both reversible and irreversible components of the interfacial moisture degradation are evaluated. Using adsorption theory in conjunction with fracture mechanics, an analytical model is developed that predicts the loss in interfacial fracture toughness as a function of moisture content. The model incorporates key parameters relevant to the problem of moisture in epoxy joints identified from the experimental portion of this research, including the interfacial hydrophobicity, active nanopore density, saturation concentration, and density of water.

The effect of moisture on interfacial adhesion is governed by two fundamental mechanisms. The first is the rate at which moisture is delivered to the interface, and the second is the change in adhesion performance as a consequence of moisture being present in the adhesive structure. This includes not only the primary effect of moisture being directly present at the interface itself but also the secondary effect of moisture altering the mechanical performance of the two materials that constitute the bimaterial interface. Having previously quantified both the rate at which moisture is delivered to the interface and the degrading effect of moisture on the elastic modulus of the materials that constitute the bimaterial interface, a model depicting the intrinsic change in interfacial adhesion as a function of moisture concentration is developed. Interfacial fracture mechanics is used to characterize this change to develop relationships that are independent of test specimen geometry.

Interfacial Fracture Testing

Interfacial fracture toughness is defined as the critical value of the energy release rate, G_c, at which a bimaterial interface will begin to delaminate. It is a property that characterizes the adhesion of a bimaterial interface, independent of the size and geometry of the cracked body. For a bimaterial interface loaded in four-point bending under plane strain conditions, it can be shown that the critical value of the energy release rate, G_c, can be determined using the following equation (Hutchinson and Suo 1992):

$$G = \frac{1}{2\overline{E_1}}\left(\frac{12M^2}{h^3}\right) - \frac{1}{2\overline{E_2}}\left(\frac{M^2}{Ih^3}\right) \tag{18}$$

where

$$\overline{E_i} \equiv \frac{E_i}{1-\nu_i^2} \tag{19}$$

M is the moment, ν is Poisson's ratio, E is the elastic modulus, subscript 1 refers to material 1, subscript 2 refers to material 2, h is the height of material 1, and I is the dimensionless moment of inertia.

Since the interfacial fracture toughness only specifies the magnitude of the crack-tip singularity, the mode mixity, ψ, must be determined from the complex stress intensity factor K. For a two-dimensional system, the complex stress intensity factor, K, is given by

$$K = K_1 + iK_2 \tag{20}$$

For four-point loading conditions, it can be shown that (Hutchinson and Suo 1992)

$$K = h^{-i\varepsilon}\sqrt{\frac{1-\alpha}{1-\beta^2}}\left(\frac{P}{\sqrt{2hU}} - ie^{i\gamma}\frac{M}{\sqrt{2h^3V}}\right)e^{i\omega} \tag{21}$$

with the mode mixity given by

$$(K_1 + iK_2)L^{i\varepsilon} = |(K_1 + iK_2)|e^{i\psi} \tag{22}$$

$$\psi = \tan^{-1}\left(\frac{\mathrm{Im}\left(KL^{i\varepsilon}\right)}{\mathrm{Re}\left(KL^{i\varepsilon}\right)}\right) \tag{23}$$

where L is the characteristic length and ε is a dimensionless quantity given by Hutchinson and Suo (1992). As shown in Eq. 23, the mode mixity for a test specimen requires the specification of some length quantity, L. The choice for L is arbitrary, but it should be selected as a fixed length and reported with the calculated values for the mode mixity.

The flexural beam test for interfacial fracture testing has three primary benefits. First, it yields intermediate values for mode mixity, which is representative of the values experienced by electronic devices during actual application. Second, it provides a means for successful interfacial fracture test specimen construction utilizing substrates and adhesives common to microelectronic packaging. Last, the flexural beam test configuration yields an open-faced test specimen design, which allows saturated, steady-state conditions to be reached in the test specimens in a relatively short amount of time. This is due to the large surface area for moisture uptake relative to the short diffusion path to the interface.

Effect of Moisture Preconditioning on Adhesion

Interfacial fracture mechanics was used to characterize the intrinsic effect of moisture on adhesion. The adhesive used was an epoxy-based underfill developed for no-flow assembly, designated as UR-B in this research. This particular underfill was determined to be ideal for studying the fundamental effect of moisture on interfacial adhesion due to its moisture diffusion kinetics and saturation behavior established from the moisture absorption portion of this research. The substrate

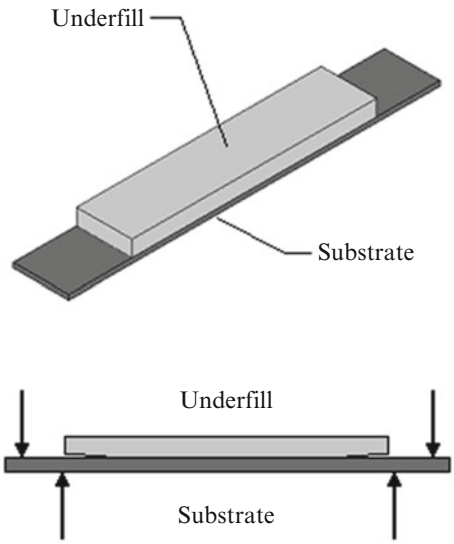


Fig. 4 Interfacial fracture toughness test specimen

used was oxygen-free electronic grade copper, alloy 101. The copper substrates were polished to a mirror finish and cleaned using the routine procedure given by Shi and Wong (1998) prior to bonding. This was done to isolate the intrinsic effect of moisture on adhesion without mechanical interlocking and/or surface contamination from influencing the results. Symmetric interface cracks were introduced into the underfill/copper bilayer test specimens by using a molding compound release agent (Ferguson and Qu 2004).

Based on the results from the moisture absorption analysis, a waterproof perimeter was applied to the interfacial fracture test specimens during moisture preconditioning and removed before fracture testing. This perimeter served two purposes. First, the application of the perimeter forced 1D diffusion through the top, open surface of the underfill, yielding uniform concentrations of moisture spatially across the entire interface for the full duration of exposure to the humid preconditioning environment. Second, the waterproof perimeter prevented moisture wicking at the interface, which allowed identification of the test specimen moisture concentration by utilizing the inherent moisture absorption characteristics of the adhesive. Completed specimens were tested in a four-point bend test at room temperature to measure the critical load of fracture for the interface. A completed representative interfacial fracture toughness test specimen is shown in Fig. 4.

Test specimens were divided into five test groups and subjected to four different levels of moisture preconditioning to ascertain the effect of moisture on interfacial fracture toughness. The test groups included fully dry, 85 °C only, 85 °C/50 %RH, 85 °C/65 %RH, and 85 °C/85 %RH, with the latter four test groups being environmentally preconditioned for 168 h. All test specimens were baked at 115 °C for at least 12 h to remove any moisture that may have been introduced during sample preparation prior to environmental aging, which was performed in a humidity

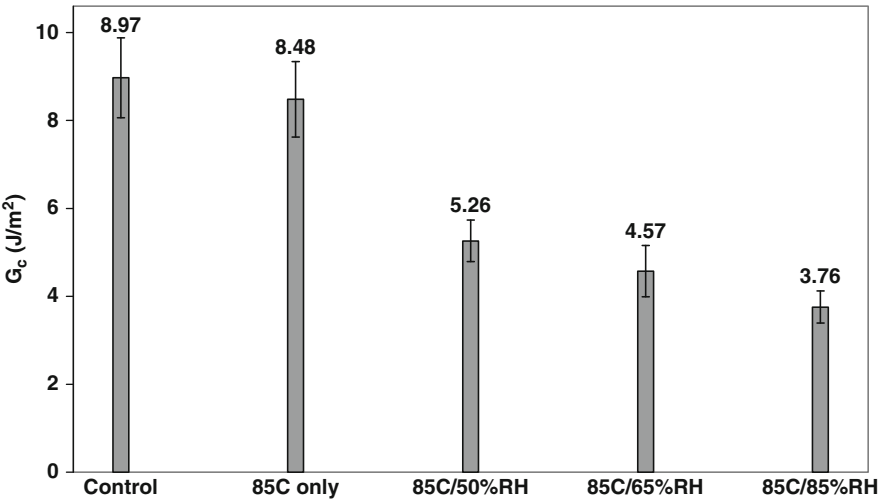


Fig. 5 Effect of environmental preconditioning on the interfacial fracture toughness of the underfill/copper interface

chamber in an atmosphere maintained at a constant temperature (±1 °C), humidity (±1 °C), and pressure (P_{atm}). All interface fracture tests were performed with both the surrounding environment and test specimens being at room temperature after environmental preconditioning. No measurable loss in moisture uptake occurred in the test specimens from the time they were removed from the environmental chamber, allowed to cool to room temperature, and experimentally tested.

Using the experimentally measured value for the critical load of fracture in conjunction with previously identified elastic modulus results, the interfacial fracture toughness of the underfill/copper test specimens was determined using Eq. (18) for each particular level of moisture preconditioning. Figure 5 provides a graphical depiction of the results depicting the effect of environmental preconditioning on the underfill/copper interfacial fracture toughness.

The entire range of mode mixity for all interfacial test specimens fell between -37.41^{o} and -37.64^{o}. The substrate height was used to define the characteristic length for all reported toughness values when evaluating the mode mixity. Since the variation in mode mixity was negligible, the effect of this variation affecting interfacial fracture toughness results between different test groups is insignificant. Consequently, interfacial fracture toughness results for different moisture preconditioned test groups can be compared to one another to ascertain the effect of increasing moisture content on toughness values. In addition, saturation was reached in each moisture preconditioning environment prior to fracture testing. As a result, a gradient of moisture concentration did not exist in the interfacial fracture toughness test specimens during testing. As shown in Fig. 5, it is clear that the contribution of thermal aging at 85 °C did not significantly affect the interfacial fracture toughness of the underfill/copper interface. It is important to remember that

Table 1 Change in the underfill/copper test specimen interfacial fracture toughness from moisture uptake

T (C)	RH (%)	C_{sat} (wt%)	C_{sat} (mg H_2O/mm^3)	G_c (J/m^2)	Toughness change (%)
Control	–	0	0.0000	8.97 ± 0.91	–
85	50	0.65	0.0075	5.26 ± 0.47	41.4
85	65	0.77	0.0089	4.57 ± 0.58	49.1
85	85	1.02	0.0118	3.76 ± 0.36	58.1

all tests were performed at room temperature; hence, only the effects of thermal aging were evaluated rather than the effect of testing at higher temperatures. Since all environmental preconditioned test groups were exposed to the same temperature component of 85 °C and duration of 168 h, any observed changes in the fracture toughness after moisture preconditioning can be attributed to the contribution of moisture. Moisture preconditioning at 85 °C/50 %RH, 85 °C/65 %RH, and 85 °C/85 %RH had a substantial effect on the interfacial fracture toughness and yielded decreases of 41.4 %, 49.1 %, and 58.1 %, respectively. A summary of the effect of moisture preconditioning on the interfacial fracture toughness is provided in Table 1, where C_{sat} represents the saturation concentration of moisture for each respective level of moisture preconditioning and given as a percent weight change (wt%).

Figures 6 and 7 depict the inherent change in the underfill/copper interfacial fracture toughness as a function of moisture concentration. It is seen that the change in the interfacial fracture toughness is sensitive to small amounts of moisture. A significant reduction in interfacial adhesion was observed for concentrations as low as 0.65 wt%. Since the moisture did not significantly alter the elastic modulus of the underfill adhesive for the moisture conditions evaluated for the interfacial fracture toughness, plasticization of the underfill from moisture contributed little to the change in the interfacial fracture toughness. As a result, the reduction in toughness is primarily attributed to the weakening of the underfill/copper interface due to the direct presence of moisture at the interface. The moisture at the interface could decrease the adhesion through displacement of the underfill reducing van der Waals forces as well as possible chemical degradation of adhesive bonds. Further investigations into the exact failure mechanism from moisture at the interface are provided in detail in subsequent sections of this chapter.

Moisture-Induced Swelling

In addition to the mechanical load applied to test specimens during interfacial fracture testing, the interface is also subjected to hygro-swelling and thermal contraction mismatch effects between the adhesive and substrate. These two effects

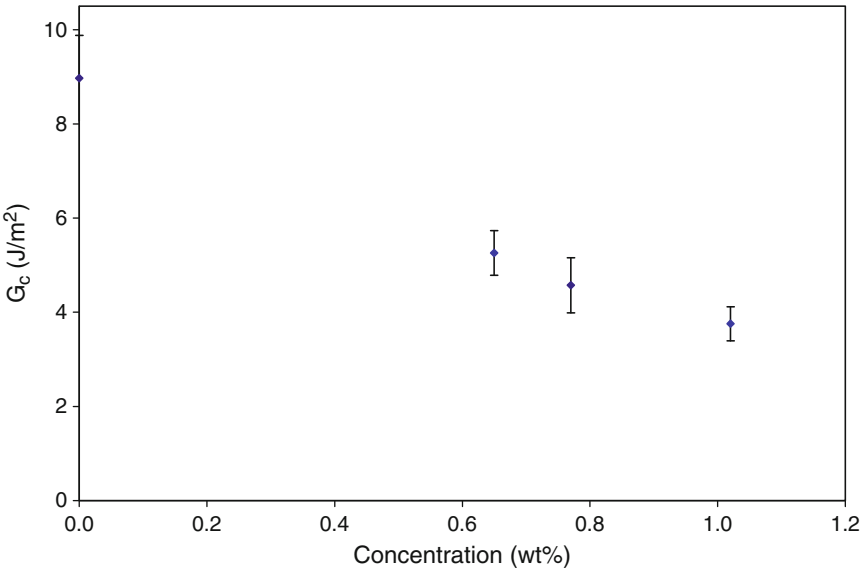


Fig. 6 Underfill/copper interfacial fracture toughness variation as a function of moisture concentration (wt%)

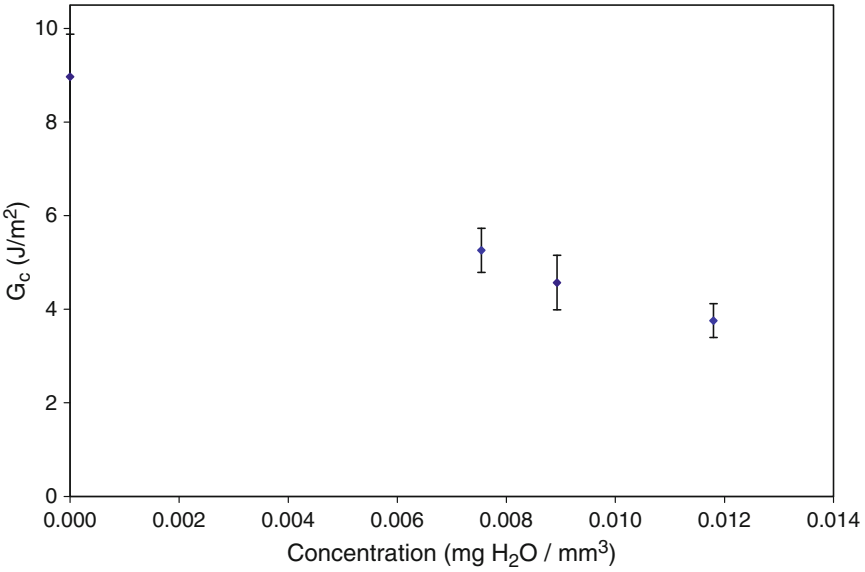


Fig. 7 Underfill/copper interfacial fracture toughness variation as a function of moisture concentration (mg H_2O/mm^3)

have opposite outcomes on the interface, as the contribution from the hygro-swelling mismatch will cause the underfill to be in compression, while the contribution from the thermal contraction mismatch will cause the underfill to be in tension. This is attributed to the different stress-free environments for each case. For the case of the hygro-swelling mismatch, fully dry conditions represent a stress-free state for the interface. As moisture is absorbed in the underfill, it will cause the underfill to expand, while the moisture impermeable substrate will retain its original dimensions. Since the moisture expansion in the underfill will be constrained by the substrate, the expansion in the underfill will yield compressive stresses within the underfill. For the case of the thermal contraction mismatch, the curing temperature of the underfill represents a stress-free state for the interface. Once test specimens are removed from the oven and allowed to cool to room temperature, the thermal mismatch between the copper and the underfill will cause the underfill to be in tension due to it wanting to shrink more than the copper substrate (CTE of experimental materials: underfill = 75 ppm/°C, copper = 17 ppm/°C). Whether the interface is dominated by the hygro-swelling mismatch, thermal contraction mismatch, or possibly neither due to the effects of one another canceling each other out for a particular moisture saturation level will depend on the characteristics of the materials that constitute each bimaterial interface relative to their moisture preconditioning environment.

To investigate the effect of hygro-swelling on interfacial fracture test results, the moisture swelling coefficient, β, of the underfill was experimentally determined for each moisture preconditioning environment. The moisture swelling coefficient is defined as

$$\beta = \frac{\Delta \ell / \ell_o}{C_{sat}} \tag{24}$$

where $\Delta\ell$ is the change in length of the specimen due to moisture absorption, ℓ_o is the initial dry length of the specimen, and C_{sat} is the saturation moisture concentration. Using Eq. 24 with experimental test data, the moisture swelling coefficient was determined for conditions of 85 °C/50 %RH (β = 1,987 ppm/wt %), 85 °C/65 %RH (β = 1,907 ppm/wt %), and 85 °C/85 %RH (β = 1,808 ppm/wt %). Having identified the moisture swelling coefficient for each moisture preconditioning environment, a comparison can be made between the hygro-swelling and thermal mismatch strains for the underfill/copper interface. The hygro-swelling mismatch strain, ε_h, and thermal mismatch strain, ε_t, are defined as follows:

$$\varepsilon_h = \beta_1 C_{sat,1} - \beta_2 C_{sat,2} \tag{25}$$

$$\varepsilon_t = (\alpha_1 - \alpha_2)(T_f - T_i) \tag{26}$$

where β is the moisture swelling coefficient, C_{sat} is the equilibrium moisture saturation concentration, α is the coefficient of thermal expansion, T is the temperature, and subscripts 1 and 2 refer to the two materials that constitute the bimaterial

Table 2 Comparison of hygro-swelling and thermal mismatch strains for the underfill/copper interfacial fracture test specimens

Environment	β (ppm/wt%)	C_{sat} (wt%)	ε_h	α_{uf} (ppm/C)	α_{Cu} (ppm/C)	T_i (C)	T_f (C)	ε_t
85C/50%RH	1,987	0.65	**0.0013**	75	17	190	25	**0.0096**
85C/65%RH	1,907	0.77	**0.0015**	75	17	190	25	**0.0096**
85C/85%RH	1,808	1.02	**0.0018**	75	17	190	25	**0.0096**

interface. The hygro-swelling mismatch strain and thermal expansion mismatch strain were calculated using Eqs. 25 and 26 respectively for each moisture preconditioning environment. Since the cooling of the interfacial fracture test specimens from the cure temperature to room temperature will result in a thermal contraction, while the uptake of moisture will result in an expansion from swelling, it should be noted that the hygro-swelling and thermal expansion mismatch strains act in opposite directions. The results are given in Table 2.

As shown in Table 2, the thermal mismatch strains were significantly greater than the hygro-swelling mismatch strains for all moisture preconditioning environments by roughly an order of magnitude. It is clear that the thermal mismatch strain dominated the interaction at the interface and was only slightly offset by a small contribution from the hygro-swelling mismatch strain for this particular bimaterial interface. As a result, the underfill will be in tension during interfacial fracture testing, effectively preloading the interface and requiring a lower critical load of fracture, P_c, from mechanical testing to advance the interface crack. Consequently, interfacial fracture toughness values will represent a conservative estimate of the interfacial fracture toughness of the interface. In addition, it is clear that increasing the saturation concentration did not significantly increase the hygro-swelling mismatch strain. All interfaces for all environments experienced similar hygro-swelling mismatch strains for the materials and moisture preconditioning environments tested in this study. Consequently, the trends exhibited in the interfacial fracture toughness as moisture concentration increases are essentially independent of the hygro-swelling mismatch relative to one another, and the observed changes between the different moisture preconditioning environments can be predominately attributed to more moisture being present at the interface resulting in a greater loss of adhesion.

Interfacial Hydrophobicity

The polarity of the water molecule will affect its behavior at the interface, which can influence the extent of environmental degradation of an adhesive joint due to the presence of moisture (Luo 2003). The polar behavior of water arises from its structure, which is composed of a single oxygen atom bonded to two hydrogen atoms. The hydrogen atoms are covalently bonded to the oxygen atom through shared

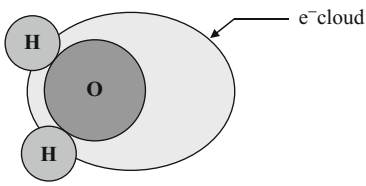


Fig. 8 Electron cloud distribution on a water molecule

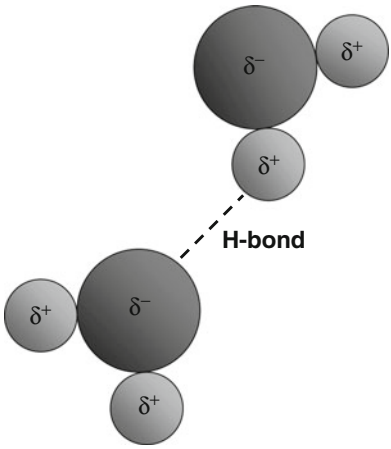


Fig. 9 Hydrogen bonding between water molecules

electrons. Two pairs of electrons surrounding the oxygen atom are involved in covalent bonds with hydrogen; however, there are also two unshared pairs of electrons (lone pair) on the other side of the oxygen atom, which shift the electron cloud of the water molecule over to the oxygen atom as shown in Fig. 8.

This uneven distribution of electron density in the water molecule yields a partial negative charge (δ^-) on the oxygen atom and a partial positive charge (δ^+) on the hydrogen atoms, giving rise to the polarity of the water molecule. Polarity allows water molecules to bond with each other, and hydrogen bonds will form between two oppositely charged ends of a water molecule as shown in Fig. 9.

The hydrogen bonds have about a tenth of the strength of an average covalent bond and are being constantly broken and reformed in liquid water. The polarity will also allow water to molecules to bond with other polar molecules, which will affect how the water will wet on different surfaces. Surfaces that contain polar molecules are hydrophilic. They interact with the water molecules to enhance wetting, causing the water to smear flat. If a surface contains alcohols, O, or N, it will probably be hydrophilic. Conversely, surfaces that contain nonpolar substances are hydrophobic. They cannot interact with the water molecules, causing it to form a bubble on the surface. In general, if a surface contains C, H, or F, it will probably be hydrophobic.

Most materials will not be purely hydrophobic or hydrophilic, but will have varying degrees to which they are considered one or the other. This is addressed in hydrophobicity, which is the study of the wetting characteristics of water on

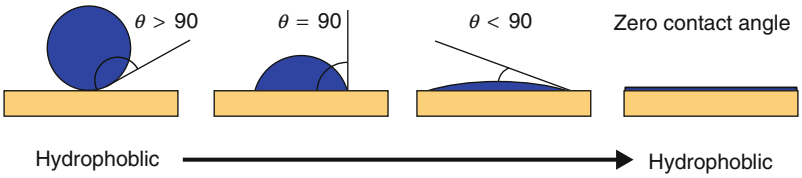


Fig. 10 Hydrophobic and hydrophilic water contact angle behavior

surfaces. One method used to test the hydrophobicity of a surface is through measurement of the contact angle, θ, using water as the probe liquid. The contact angle represents a balance between the adhesive forces between the liquid and solid and cohesive forces in the liquid. The adhesive forces cause the liquid drop to spread, while the cohesive forces cause the liquid drop to retain the shape of a sphere. The contact angle is a direct measure of wettability and provides an effective means to evaluate many surface properties such as surface contamination, surface hydrophobicity, surface energetics, and surface heterogeneity. When $\theta > 0$, the liquid is nonspreading and reaches an equilibrium position between the liquid–fluid and solid–liquid interfaces. When $\theta = 0$ the liquid wets without limit and spontaneously spreads freely over the surface. Hydrophobic surfaces repel water and produce high contact angles. Hydrophilic surfaces attract water and produce low contact angles. Figure 10 illustrates the contact angle behavior of water on both hydrophobic and hydrophilic surfaces.

By utilizing water as the probe liquid, the interfacial hydrophobicity can be ascertained by measuring the water contact angle of both the adhesive and substrate. To determine the hydrophobicity of interfacial fracture test specimens, contact angle measurements were made for the adhesive and substrate evaluated in this study. Both the clean copper substrate and underfill adhesive exhibited fairly hydrophobic behavior with contact angles of 74^{o} and 83^{o} respectively. Having established the hydrophobicity of the substrate and adhesive, the interfacial hydrophobicity of the underfill/copper interfacial fracture test specimens can be evaluated. When addressing the relative hydrophobicity of the substrate and adhesive to moisture behavior at the interface, the interaction can become complex. The surface with the most dominant degree of hydrophobicity will govern the shape and response of the water at the interface. For example, if a hydrophobic substrate is bonded with a hydrophilic adhesive, then the water at the interface will want to minimize contact with the substrate and maximize contact with the adhesive. Depending on imperfections in the bonding, surface roughness, and the relative degree of hydrophobicity of the substrate to the adhesive, water at the interface will more or less form a somewhat hemispherical shape at the interface, with the spherical end minimizing contact on the substrate and the open end maximizing contact on the adhesive. Naturally, the shape of the water at the interface can have various permutations of the aforementioned shape depending on the degree of hydrophobic behavior of the substrate relative to the hydrophilic behavior of the adhesive, but the general idea remains the same. For other systems with varying

degrees of hydrophobicity, the shape of the water at the interface relative to the hydrophobicity of the substrate and adhesive can be extremely difficult to characterize; however, qualitative conclusions can be made. For the case of the underfill/copper interfacial fracture test specimens, the relative hydrophobicity of the adhesive to the substrate was similar; consequently, the wetting behavior of the moisture at the interface would not be significantly dominated by either the adhesive or substrate.

An additional consideration unique to environmental preconditioning is the growth of oxides affecting the interfacial hydrophobicity. Copper has a strong affinity to oxygen, and the development of an oxidation layer between the substrate and adhesive after bonding is inevitable. Initially, cuprous oxide, Cu_2O, will form followed by the formation of a layer of cupric oxide, CuO (Cho and Cho 2000). The oxidation of copper substrates can be significant, and previous studies have shown that the water contact angle on copper is affected by oxidation (Cho and Cho 2000; Yi et al. 1999; Hong et al. 1994; Kim 1991). Due to oxidation growth on the copper substrates, contact angle measurements were made for each preconditioning environment to monitor any change in the hydrophobicity of the copper surface.

Since the copper bonding surface of the interfacial fracture test specimen will be shielded by the underfill adhesive, the oxidation growth rate will be different than for bare copper environmentally aged for a similar duration of time. Therefore, water contact angles for each environmental test group were measured using special test specimens that mimicked the exposure of the copper bonding surface to similar amounts of oxygen and moisture as the interfacial fracture test specimens. These specimens used the same geometry as the interfacial fracture test specimens, but the underfill adhesive was cured separately in an individual mold. After curing the adhesive, the underfill was placed on top of the copper substrate and held in place by c-clamps. Similar to the interfacial fracture test specimens, a waterproof sealant was applied around the perimeter of the test specimen to eliminate wicking of moisture at the interface and force 1D diffusion through the top surface of the underfill. After environmental preconditioning, the waterproof perimeter, c-clamps, and underfill were removed from the test specimen for contact angle measurement of the copper surface.

Experimentally measured water contact angle results were as follows: 76^o for 85 °C thermal aging, 76^o for 85 °C/50 %RH moisture preconditioning, 77^o for 85 °C/65 %RH moisture preconditioning, and 77^o for 85 °C/85 %RH moisture preconditioning. All test groups were preconditioning for the same duration of 168 h, which was the same criteria used in the evaluation of the effect of moisture on interfacial adhesion. Based on these results, it is evident that all levels of environmental preconditioning did not significantly alter the water contact angle and associated hydrophobicity of the interface. As a result, similar interfacial wetting characteristics of moisture at the interface will occur for all preconditioning environments.

Although the contact angle did not significantly change, there did appear to be a slight increase in the water contact angle with moisture preconditioning.

Previous studies have shown both an increase (Yi et al. 1999; Kim 1991) and decrease (Cho and Cho 2000; Hong et al. 1994) in the water contact angle of copper with oxidation. The oxidation–reduction chemistry occurring at the interface relative to environmental preconditioning is complex, and the differences in trends could be attributed to the degree of oxidation altering the surface chemistry (Cho and Cho 2000), change in surface roughness of the substrate from oxidation growth (Hong et al. 1994), and contamination of the surface by hydrocarbons from the environment (Luo 2003). In addition, Yi et al. (1999) have provided data correlating the oxide layer thickness on copper leadframes to water contact angles. These data shows a slow, gradual increase in oxide thickness from water contact angles ranging from 72° to 78° but depicts a sharp increase in oxide layer thickness for contact angles exceeding 80°. Based on results for the water contact angle on copper in this study, all measurements yielded average contact angles less than 78° with very little variation with each other. This indicates a similar level of interfacial hydrophobicity and oxide layer thickness for all environmentally preconditioned test groups. Both Mino et al. (1998) and Chong et al. (1995) have shown that the development of the copper oxide layer thickness is significantly slower and minimal for temperatures below 100 °C and 120 °C. Since the test specimens in this study had a temperature component of only 85 °C, it is anticipated that the oxide layer thickness that developed on test specimens would have a minimal effect on toughness results. This is also supported by X-ray photoelectron spectroscopy (XPS) results. XPS showed the presence of cupric oxide not only in the 85 °C/50 %RH, 85 °C/65 %RH, and 85 °C/85 %RH test groups but also in the 85 °C thermal aging test group. As a result, identical oxide chemical formations existed at the interface for all environmentally preconditioned test groups. In addition, similar atomic percentages of cupric oxide were obtained when comparing thermal aging at 85 °C to the moisture preconditioning environments of 85 °C/50 %RH, 85 °C/65 %RH, and 85 °C/85 %RH, indicating that the moisture component had a minimal contribution to oxidation growth rates on the copper compared to the available oxygen in the air common to all environmental preconditioned environments. Consequently, a similar level of oxidation thickness existed on all environmentally preconditioned test specimens, which supports the results from the water contact angle measurements.

Since oxides were removed from the copper surface before adhesive bonding and the flux present in the no-flow underfill would have removed any oxides that developed during adhesive curing, it is possible that the oxidation growth from environmental preconditioning would have an effect on the interfacial fracture toughness results. This oxide growth could displace the underfill from the copper substrate after bonding to contribute to the observed loss in adhesion after moisture preconditioning shown in Fig. 5. Since both water contact angle measurements and XPS results demonstrate a similar oxidation thickness existed on all environmentally preconditioned test specimens, the 85 °C thermal aging results can be compared to the control test results to ascertain the effect of oxidation growth on the loss in adhesion without the contribution from moisture. As shown in Fig. 5, thermal aging at 85 °C produced little to no effect on interfacial fracture toughness results,

thus oxidation growth displacing the underfill after adhesive bonding had an insignificant effect on the adhesion loss compared to the effect of moisture from moisture preconditioning.

Interfacial Fracture Toughness Recovery from Moisture Uptake

The underfill/copper interface was found to be very sensitive to moisture, with large decreases in interfacial fracture toughness occurring for moisture preconditioning environments of 85 °C/50 %RH, 85 °C/65 %RH, and 85 °C/85 %RH (Fig. 5). To further investigate the reversible and irreversible nature of moisture on the interfacial adhesion of the underfill/copper interface, additional test specimens were moisture preconditioned for each condition for 168 h followed by baking at 95 °C until fully dry. A fully dried state was established when there was no measurable change in the weight of a specimen for a period of 24 h. Upon reaching a dry state, specimens were fracture tested to ascertain the interfacial fracture toughness. The entire range of mode mixity for all interfacial test specimens fell between -37.43^{o} and -37.48^{o}. The substrate height was used to define the characteristic length for all reported toughness values when evaluating the mode mixity. Since the variation in mode mixity was negligible, the effect of this variation influencing interfacial fracture toughness results between different test groups is insignificant. Consequently, toughness recovery results for different moisture preconditioned test groups can be compared to one another to ascertain the effect of increasing moisture content on toughness values. Figure 11 provides a graphical depiction of the effect of environmental preconditioning and recovery of the underfill/copper interfacial fracture toughness.

As shown in Fig. 11, most of the loss in interfacial fracture toughness from moisture was not recovered upon fully drying. Since the small change in the underfill elastic modulus from moisture was recoverable upon fully drying, the permanent reduction in the toughness of the underfill/copper interface is attributed to the direct presence of moisture at the interface debonding the underfill adhesive to the copper substrate. Similar in form to the recoverability of the elastic modulus given by Eq. 26, the recoverability for the interfacial fracture toughness will be defined as follows:

$$\text{Recoverability } (\%) = \frac{G_{c,\,\text{recovery}} - G_{c,\,sat}}{G_{c,\,\text{dry}} - G_{c,\,sat}} \cdot 100 \tag{27}$$

where $G_{c,\text{recovery}}$ is value of the interfacial fracture toughness upon fully drying from the moisture saturated state, $G_{c,sat}$ is the saturated value of the interfacial fracture toughness after moisture absorption, and $G_{c,\text{dry}}$ is the unaged, control value of the interfacial fracture toughness. Equation 27 only applies when the mode mixity of the interfacial fracture toughness before and after moisture preconditioning remains relatively unchanged, otherwise changes in the toughness due to a contribution from a change in the mode mixity will introduce error in the

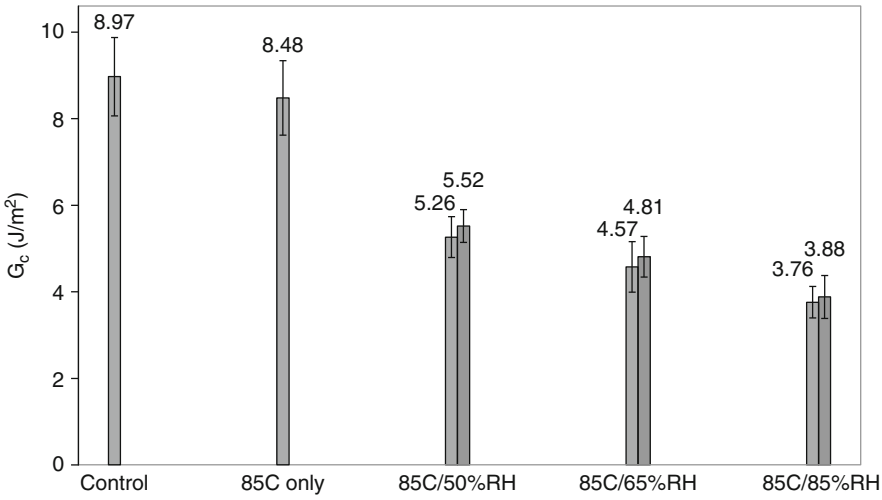


Fig. 11 Recovery of the underfill/copper interfacial fracture toughness on removal of moisture

Table 3 Recoverability of the underfill/copper interfacial fracture toughness from moisture uptake after subsequent drying

T (°C)	RH (%)	C_{sat} (wt%)	$G_{c,sat}$ (J/m^2)	$G_{c,recovery}$ (J/m^2)	Recoverability (%)
Control	–	0.00	8.97 ± 0.91	–	–
85	50	0.65	5.26 ± 0.47	5.52 ± 0.38	7.0
85	65	0.77	4.57 ± 0.58	4.81 ± 0.47	5.5
85	85	1.02	3.76 ± 0.36	3.88 ± 0.50	2.3

recoverability results. The recoverability of the underfill/copper interfacial fracture toughness is given in Table 3.

As shown by Table 3, the irreversible damage on interfacial fracture toughness from exposure to moisture was substantial for the underfill/copper interface. Very little of the underfill/copper interfacial fracture toughness was recoverable after fully drying, with recoverability values for all moisture preconditioning environments less than 7 %. It is also evident that a relatively small amount of moisture reaching the interface causes the structural integrity of the adhesive bond to be noticeably, permanently compromised.

Interfacial Fracture Toughness Moisture Degradation Model

Having implemented an extensive experimental program to ascertain the role of moisture in adhesion degradation and the physical mechanisms responsible for the change in interfacial adhesion, the focus now shifts to developing a model depicting the intrinsic loss in interfacial fracture toughness as a function of the critical

parameters relevant to moisture. At the root of this model is characterizing the dominant mechanism for adhesion between the adhesive and substrate. There are four primary mechanisms for adhesion which have been proposed. They include mechanical interlocking, diffusion theory, electronic theory, and adsorption theory (Kinloch 1987). For the underfill/copper interface, the contributions of interfacial diffusion and electrostatic forces between the adhesive and substrate causing adhesion are far lower than the effects of mechanical interlocking and adsorption. Since the copper substrates in this study were polished to a mirror finish, the effects from mechanical interlocking of the adhesive into irregularities present on the substrate surface will be small compared to the effects from intermolecular secondary forces (i.e., van der Waals) between the atoms and molecules in the surfaces of the adhesive and substrate. Consequently, adsorption theory will dominate the adhesive bonding at the underfill/copper interface.

Provided adsorption theory governs adhesion and only secondary forces are acting across an interface; the stability of an adhesive/substrate interface in the presence of moisture can be ascertained from thermodynamic arguments. The thermodynamic work of adhesion, W_A, in an inert medium is given by (Kinloch 1987)

$$W_A = \gamma_a + \gamma_s - \gamma_{as} \tag{28}$$

where γ_a is the surface-free energy of the adhesive, γ_s is the surface-free energy of the substrate, and γ_{as} is the interfacial free energy. In the presence of a liquid, the thermodynamic work of adhesion, W_{Al}, is given by

$$W_{Al} = \gamma_{al} + \gamma_{sl} - \gamma_{as} \tag{29}$$

where γ_{al} and γ_{sl} are the interfacial free energies between the adhesive/liquid and substrate/liquid interfaces, respectively. Typically the thermodynamic work of adhesion of an adhesive/substrate interface in an inert medium, W_A, is positive, which indicates the amount of energy required to separate a unit area of the interface. However, the thermodynamic work of adhesion in the presence of a liquid, W_{Al}, can be negative, which indicates the interface is unstable and will separate when it comes in contact with the liquid. Thus, the calculation of W_A and W_{Al} can indicate the environmental stability of the adhesive/substrate interface. Kinloch (1987) has shown that W_A and W_{Al} may be calculated from the following expressions:

$$W_A = 2\sqrt{\gamma_a^D \gamma_s^D} + 2\sqrt{\gamma_a^P \gamma_s^P} \tag{30}$$

$$W_{Al} = 2\left(\gamma_{lv} - \sqrt{\gamma_a^D \gamma_{lv}^D} - \sqrt{\gamma_a^P \gamma_{lv}^P} - \sqrt{\gamma_s^D \gamma_{lv}^D} - \sqrt{\gamma_s^P \gamma_{lv}^P} + \sqrt{\gamma_a^D \gamma_s^D} + \sqrt{\gamma_a^P \gamma_s^P}\right) \tag{31}$$

where γ^D is the dispersion component of surface-free energy, γ^P is the polar component of surface-free energy, and γ_{lv} is the surface-free energy of the

Table 4 Polar and dispersion surface-free energies of epoxy, copper, and water (Kinloch 1987)

Substance	γ (mJ/m^2)	γ^D (mJ/m^2)	γ^P (mJ/m^2)
Epoxy	46.2	41.2	5.0
Copper	1,360	60	1,300
Water	72.2	22.0	50.2

liquid. Table 4 gives the polar and dispersion surface-free energies of epoxy, copper, and water.

Using the values given in Table 4 and substituting into Eq. 30, the thermodynamic work of adhesion of the epoxy/copper interface is 260.7 mJ/m^2. If water is present at the epoxy/copper interface, the thermodynamic work of adhesion given by Eq. 31 is -270.4 mJ/m^2. Therefore, since the work of adhesion is positive before exposure to moisture and negative after exposure, all adhesion of the epoxy/copper interface is lost if water comes in contact with the interface. This is supported by the fact that virtually none of the observed loss in adhesion from moisture exposure was recovered upon fully drying.

Using adsorption theory as the physical basis for the loss in adhesion from moisture, expressions are now developed depicting the amount of moisture delivered to the underfill/copper interface. Since the interfacial fracture test specimens were designed to prevent wicking of moisture at the interface and the copper substrate provides a barrier for moisture transport, the moisture transport to the interface is governed by the epoxy network of the underfill. Soles and Yee (2000) have shown that water traverses within the epoxy through the network of nanopores inherent in the epoxy structure. A typical nanopore ranges from 5.0 to 6.1 Å in diameter. Figure 12 illustrates the transport of moisture through the bulk epoxy of an interfacial fracture test specimen.

Assuming that the nanopore channels are the only mechanism by which moisture can be delivered to the interface, the saturation concentration in the epoxy expressed in mg H_2O/mm^3 is given by

$$C_{sat} = \frac{\rho(NV)}{V_{tot}} \tag{32}$$

where ρ is the density of water measured in milligram per cubic millimeter (mg/mm^3), N is the number of nanopores actively participating within the epoxy network, V is the volume occupied by a single nanopore in the epoxy network, and V_{tot} is the total volume of the epoxy. After rearrangement of Eq. 32, the number of nanopores actively participating within an epoxy system for a given saturation concentration is as follows:

$$N = \frac{4AC_{sat}}{\pi\rho D^2} \tag{33}$$

where A is the total area of the interface and D is the nanopore diameter. Assuming adsorption theory holds, the adhesive bond area, A_{bond}, that remains intact after

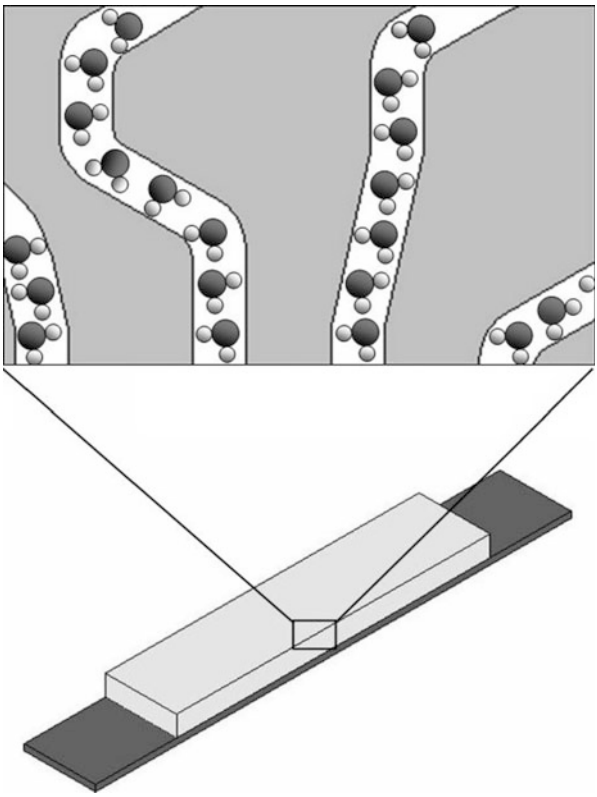


Fig. 12 Moisture transport through the bulk epoxy of an interfacial fracture test specimen

exposure to moisture will depend on the area occupied by the moisture at the interface, A_{H2O}:

$$A_{\text{bond}} = A - A_{H_2O} \tag{34}$$

Relating this adhesive bond area to the number of nanopores actively participating in transport yields

$$A_{\text{bond}} = A - \pi N r_{\text{debond}}^2 \tag{35}$$

where r_{debond} represents the debond radius of moisture at the interface that occurs at each nanopore. The debond radius must be greater or equal to the nanopore radius and is governed by the interfacial hydrophobicity of the adhesive/substrate interface. Figure 13 provides a graphical depiction of the parameter, r_{debond}, at the interface.

Substituting Eq. 32 into Eq. 35 provides an expression for the adhesive bond area that remains intact after exposure to a particular moisture saturation concentration:

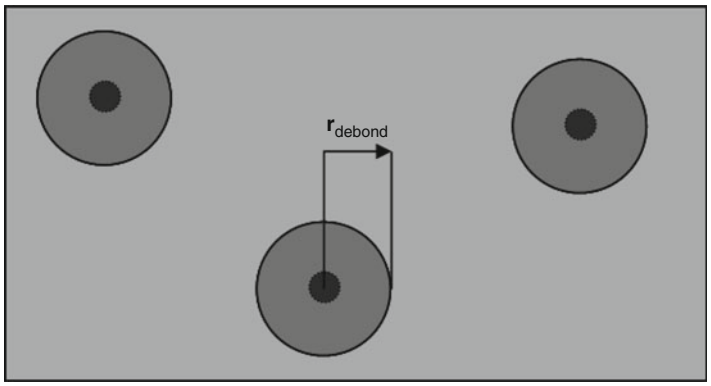


Fig. 13 Graphical illustration of the parameter, r_{debond}, at the interface

$$A_{\text{bond}} = A - \frac{4AC_{sat}r_{\text{debond}}^2}{\rho D^2} \tag{36}$$

A fracture mechanics development may be employed to relate the change in bond area due to the presence of moisture at the interface to the interfacial moisture concentration. Recall from fracture mechanics the general form of the stress intensity factor:

$$K = S\sigma\sqrt{\pi a} \tag{37}$$

where S is a dimensionless constant that depends on the geometry and mode of loading, σ is the remotely applied stress, and a is the crack length. The stress intensity factor is related to the fracture toughness, G_c, by the following expression:

$$G_c = Z\sigma^2 \tag{38}$$

where

$$Z = \frac{\pi a S^2(1-\upsilon^2)}{E}$$

Based on the thermodynamic work of adhesion for the epoxy/copper interface, the interface will become unstable and debond in the presence of moisture; however, since interfacial fracture toughness is a material property that characterizes the adhesion of the interface, the toughness must be the same in all areas that remain bonded after exposure to moisture. Using mode I loading and making the following three assumptions – (1) adsorption theory dominates the interfacial bonding, (2) the change in the mechanical properties of both the adhesive and substrate from moisture is small relative to the change in bond area from moisture, and (3) the relative change in fracture toughness from moisture remains constant irrespective to the

means of measuring the toughness for a given moisture saturation concentration – an expression is obtained relating the change in bond area due to the presence of moisture to the change in the critical load of fracture:

$$\frac{P_{\text{wet}}}{A - \pi N r_{\text{debond}}^2} = \frac{P_{\text{dry}}}{A} \tag{39}$$

Rearranging Eq. 39 to obtain an expression for P_{wet} and substituting that value into Eq. 38 for the wet, saturated case yields the following expression:

$$G_{c,\,\text{wet}} = \left(1 - \frac{\pi N r_{\text{debond}}^2}{A}\right)^2 G_{c,\,\text{dry}} \tag{40}$$

As the saturation moisture concentration increases, so will the number of active nanopores participating. The incremental change in fracture toughness due to the participation of a single additional nanopore, $N + 1$, is given by

$$G_{c,\,\text{wet}} = \left(1 - \frac{\pi (N+1) r_{\text{debond}}^2}{A}\right)^2 G_{c,\,\text{dry}} \tag{41}$$

For convenience, define f such that for N nanopores participating

$$f_N = \frac{\pi N r_{\text{debond}}^2}{A} \tag{42}$$

For $N + 1$ nanopores participating,

$$f_{N+1} = \frac{\pi r_{\text{debond}}^2}{A}(N+1) \tag{43}$$

Restating Eqs. 40 and 41 in terms of f,

$$G_{c,\,\text{wet}}(f_N) = \left(1 - \frac{\pi N r_{\text{debond}}^2}{A}\right)^2 G_{c,\,\text{dry}} \tag{44}$$

$$G_{c,\,\text{wet}}(f_{N+1}) = \left(1 - \frac{\pi r_{\text{debond}}^2}{A}\right)^2 G_{c,\,\text{wet}}(f_{N_N}) \tag{45}$$

Subtracting Eq. 44 from Eq. 45 and dividing by $f_{N+1} - f_N$ give

$$\frac{G_{c,\,\text{wet}}(f_{N+1}) - G_{c,\,\text{wet}}(f_N)}{f_{N+1} - f_N} = \frac{\left[1 - \frac{\pi r_{\text{debond}}^2}{A}\right]^2 G_{c,\,\text{wet}}(f_N) - G_{c,\,\text{wet}}(f_N)}{f_{N+1} - f_N} \tag{46}$$

Utilizing a Taylor series expansion of f_N with first-order accuracy and substituting Eqs. 42 and 35 into Eq. 46 yield

$$\frac{dG_{c,\,\mathrm{wet}}(f_N)}{df_N} = \frac{\left[1 - \frac{\pi r_{\mathrm{debond}}^2}{A}\right]^2 G_{c,\,\mathrm{wet}}(f_N) - G_{c,\,\mathrm{wet}}(f_N)}{\frac{\pi r_{\mathrm{debond}}^2}{A}} \tag{47}$$

Simplification and elimination of higher-order terms gives the following differential equation characterizing the loss in interfacial fracture toughness due to moisture:

$$\frac{dG_{c,\,\mathrm{wet}}(f_N)}{df_N} = -2G_{c,\,\mathrm{wet}}(f_N) \tag{48}$$

subject to the boundary condition

$$G_{c,\,\mathrm{wet}}(f_N = 0) = G_{c,\,\mathrm{dry}} \tag{49}$$

Solution of Eq. 48 gives

$$G_{c,\,\mathrm{wet}} = G_{c,\,\mathrm{dry}} \exp\left[\frac{-8C_{sat} r_{\mathrm{debond}}^2}{\rho D^2}\right] \tag{50}$$

Equation 50 characterizes the loss in interfacial fracture toughness from moisture in terms of key parameters relevant to moisture. Using the value for the density of water at room temperature (0.998 mg/mm^3), an average nanopore diameter of 5.5 Å, and the saturation concentration determined from the experimental portion of this study in conjunction with Eqs. 32 and 50, the number of active nanopores participating, N, and value of r_{debond} can be determined by the intrinsic response of the material system to each level of moisture preconditioning. The results are shown in Table 5.

As shown in Table 5, the number of nanopores participating increases with saturation concentration. This is expected since an increase in saturation concentration would increase the available moisture for transport through the nanopores. In addition, the values for r_{debond} were similar for each moisture preconditioning environment for both respective interfaces, which is also expected since X-ray photoelectron spectroscopy and water contact angle results did not indicate a change in the interfacial hydrophobicity of the copper surface from moisture preconditioning. The slight variation in the values for r_{debond} could in part be attributed to experimental scatter. Since the results were similar, they were averaged to obtain a representative value for r_{debond} in the presence of moisture for each interface.

Using the moisture parameters identified for each interfacial material system, Eq. 50 was used to predict the interfacial fracture toughness for the underfill/copper interface as a function of increasing saturation concentration.

As shown in Fig. 14, Eq. 50 accurately predicted the loss in interfacial fracture toughness as a function of increasing moisture concentration. Since Eq. 50 was based on the physics of adsorption theory, it will yield a loss in interfacial fracture toughness provided there is moisture at the interface, no matter how small the

Table 5 Key parameters relevant to moisture for the underfill/copper interface

Environment	Substrate	Adhesive	C_{sat} (mg H_2O/mm^3)	N	r_{debond} (mm)
85C/50 %RH	Copper	Underfill	0.0075	1.006E+13	1.640E-06
85C/65 %RH	Copper	Underfill	0.0089	1.194E+13	1.692E-06
85C/85 %RH	Copper	Underfill	0.0118	1.583E+13	1.669E-06

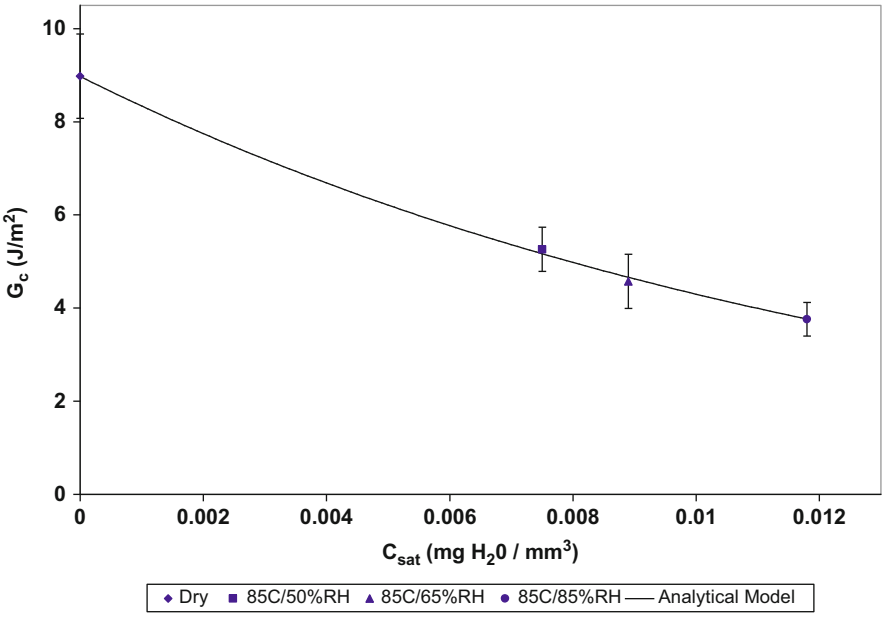


Fig. 14 Analytical prediction of the loss in interfacial fracture toughness from moisture for the underfill/copper interface

concentration. This contradicts the results of previous studies, who have reported that a critical concentration of water may exist below which there is no measurable loss in adhesion (Comyn et al. 1994; Gledhill et al. 1980; Kinloch 1979). Based on the results of adsorption theory, it does not appear possible that a critical concentration of water could exist in theory. It is possible in those studies that other mechanisms for adhesion in addition to adsorption theory governed the adhesion at the interface, which could explain why a critical concentration of water was observed. An additional consideration is the method of testing used to obtain adhesion results. The aforementioned studies used lap shear test specimens to determine the interfacial strength after moisture preconditioning. Due to lacking a precrack at the interface and the applied load being distributed over the entire bonding area, these test specimens are not as sensitive to interfacial failure, consequently, possibly also explaining why in part a critical concentration of water appeared to exist for low concentrations of moisture. Conversely, interfacial fracture toughness test specimens are designed for interfacial failure through the

use of a precrack at the interface, making them more sensitive to subtle changes in adhesion at the interface. The work of Wylde and Spelt (1998) supports this observation. Using interfacial fracture toughness test specimens with a similar material system previously reported to exhibit a critical concentration of water from lap shear results, they found a decrease in the interfacial toughness from moisture for all concentrations of moisture, including those lower than the previously reported critical concentration of water. Consequently, provided adsorption theory dominates the adhesive bonding at the adhesive/substrate interface and the assumptions in the development of the model as satisfied, Eq. 50 should accurately predict the loss in interfacial fracture toughness for a given moisture concentration.

Conclusion

This chapter consists of two major contributions. First, a mechanics- and physics-based model has been developed to determine the amount of interfacial failure versus cohesive failure of polymer–metal interfaces. Based on this model, the amount of cohesive failure near the adhesion interface can be predicted, and therefore, the adhesion enhancement can be quantitatively determined. Because the cohesive failure strength of the adhesive is usually much greater than the polymer–metal interfacial strength, the cohesive failure near the interface due to surface roughness can yield significant improvement of the apparent interface fracture toughness.

It should be mentioned that the model development here assumes perfect linear elastic crack-tip fields. When significant plasticity takes place near the crack tip, some modifications are needed to include the plastic deformation. Furthermore, it is clear that the other adhesion mechanisms including physical and chemical adhesion play significant roles in the adhesion enhancement.

Second, effects of moisture on the interfacial fracture toughness are investigated in great details. Two epoxy-based, no-flow underfills, designated UR-A and UR-B in this research, were examined for moisture transport behavior. Based on the results of the diffusion analysis, it was clear that very different behavior was exhibited by each underfill. Although UR-A absorbed more aggregate moisture than UR-B, the moisture diffused more easily through UR-B than UR-A. This behavior is attributed to the different chemistry in each underfill, where the presence of amine functional groups in UR-A retarded moisture transport, while the absence of amine function groups in UR-B yielded enhanced diffusion rates. A finite element model was developed to analytically and visually depict the moisture transport characteristics of UR-A and UR-B. The model shows that moisture will initially reach the interface for microelectronic assemblies that use UR-B before comparably sized assemblies utilizing UR-A; however, due to the higher saturation concentration of UR-A, more moisture will arrive at the interface for UR-A assemblies if exposed to the moist environment for longer durations. This presents an interesting scenario for microelectronic applications when considered with interfacial fracture toughness results. Based on interfacial fracture toughness

results, it was found that the critical aspect in the loss in interfacial adhesion is not the degradation of the adhesive from moisture uptake, but the amount of moisture that arrives at the interface for the adhesives and substrates evaluated in this study. With that in mind, depending on the service environment and duration of exposure to that environment, one underfill may yield significantly better interfacial adhesion than the other. For instance, if the microelectronic package is exposed to a moist environment for a long duration of time and assuming similar adhesion characteristics for both underfills, the non-amine containing resin UR-B would be a better choice in terms of reliability. This is a result of the lower saturation concentration of UR-B to UR-A; consequently, the total amount of moisture that arrives to the interface is limited by the inherent moisture saturation behavior of the underfill. Conversely, if the microelectronic package is going to be exposed to a moist environment for a short period of time and again assuming similar adhesion characteristics of both underfills, the amine containing resin UR-A would be a better choice for reliability. This is a result of the amine functional groups present in UR-A retarding moisture transport through the resin; consequently, it will take longer for moisture to reach the interface for assemblies using UR-A than comparably sized assemblies encapsulated with UR-B. Naturally both of these scenarios assume that the only means for moisture transport to the interface is by bulk diffusion through the underfill and caution should be implemented to insure moisture does not wick at the interface in addition to the bulk diffusion.

Another consideration when evaluating the problem of moisture is the effect of moisture on the bulk properties of the adhesive and substrate. Absorbed moisture can alter the mechanical characteristics of the adhesive and substrate, which can indirectly affect interfacial adhesion. A change in the elastic modulus can alter interfacial fracture toughness results considerably. Since the substrates evaluated in this research were metallic and impermeable to moisture, only the variation in the underfill elastic modulus due to moisture uptake was considered. The elastic modulus was measured for several different moisture preconditioning environments and subsequent saturation concentrations. It is important to note that specimens were fully saturated with moisture at the time of testing; thus, a gradient of moisture did not exist within the specimens at the time of testing and the inherent, wet modulus was identified for each condition. In addition, thermal aging test results showed no change in the elastic modulus from the temperature component of the moisture preconditioning environment; consequently, all observed losses can be attributed to the presence of moisture. Results show a gradual decrease in the elastic modulus for concentrations $<$1.02 wt% (0.0118 mg H2O/mm3) with a more noticeable decrease (17 %) occurring at concentrations of 1.19 wt% (0.0138 mg H2O/mm3). Since the inherent wet modulus was identified for several different saturation concentrations, results depict the inherent change in elastic modulus of the underfill as a function of increasing moisture concentration, which can be used to model the transient change in the underfill elastic modulus as moisture is absorbed. To evaluate the recovery of the underfill elastic modulus from moisture uptake, specimens were allowed to reach saturation followed by baking in a convection oven until fully dry. The recovery results indicate that the majority of

the loss in underfill elastic modulus was recovered upon fully drying, although some permanent loss did occur. Since plasticization from moisture is the only known reversible mechanism for the change in mechanical characteristics due to moisture uptake, the recovery results demonstrate that plasticization was the dominant mechanism responsible for the loss in the elastic modulus. The slight irreversible effect from moisture uptake can be attributed in part to hydrolysis, which was supported by a slight, net permanent weight gain in the underfill after fully drying. It should be noted that DSC results demonstrated that the underfill was fully cured before moisture preconditioning, so the contribution of incomplete curing in the underfill reacting with moisture is unlikely. Since moisture did not significantly alter the elastic modulus of the underfill and bearing in mind that the majority of change in the elastic modulus was recovered upon drying, the long-term reliability of the underfill in microelectronic applications is not a primary concern when considering the effect of moisture. Since plasticization from moisture was found to be the dominant mechanism responsible for the change in the underfill modulus, variations in the underfill chemistry can be addressed to yield products that are intentionally more resistant to plasticization from moisture if known to be exposed to moist environments.

Having established the moisture absorption kinetics and change in properties of the adhesive and substrate from moisture, the final aspect considered is the effect of moisture on interfacial adhesion. Underfill/copper and undefill/FR-4 board interfacial fracture toughness test specimens were made to evaluate this effect. By implementing several different moisture preconditioning environments and by using the critical load of fracture determined from the test specimens in conjunction with the moisture concentration and elastic modulus variation for each environment, the interfacial fracture toughness was determined as function of increasing moisture concentration. Failures occurred at the underfill/copper interface for the underfill/copper interfacial test specimens and at the solder mask/copper interface for the underfill/FR-4 board test specimens for all environments. Moisture preconditioning results demonstrate that both interfaces were very sensitive to moisture, with significant changes in interfacial toughness for concentrations as lows as 0.65 wt% (0.0089 mg H2O/mm3). Since there is both a temperature and moisture component to moisture preconditioning, thermal aging tests were performed to delineate the contributions from both on interfacial fracture results. The thermal aging test results showed no significant change in the toughness from the temperature component of the moisture preconditioning environment; consequently, all observed losses can be attributed to the presence of moisture. In addition, since the moisture did not significantly change the elastic modulus of the underfill adhesive for the moisture conditions evaluated for the interfacial fracture toughness, plasticization of the underfill from moisture contributed little to the change in the interfacial fracture toughness. As a result, the reduction in the toughness is primarily attributed to the weakening of the interface due to the direct presence of moisture at the interface. This has a very significant implication for practical application, demonstrating that the critical aspect to consider when minimizing the loss in interfacial adhesion from moisture is preventing moisture from

physically reaching the interface. Using adsorption theory, the stability of an adhesive/substrate interface in the presence of moisture can be ascertained from thermodynamic arguments. The work of adhesion was determined to be positive before exposure to moisture and negative after exposure, indicating all adhesion of the epoxy/copper interface is lost if water comes in contact with the interface. This is supported by recovery results, which showed very little if any of the interfacial fracture toughness is recovered upon fully drying. Consequently, the results indicate that the adsorption theory of bonding is the dominant bonding mechanism for the epoxy/metal interfaces studied in this research. Using adsorption theory in conjunction with fracture mechanics, an analytical model was developed that predicts the loss in interfacial fracture toughness as a function of moisture content. The model incorporates key parameters relevant to the problem of moisture in epoxy joints identified in this research, including the interfacial hydrophobicity, active nanopore density, saturation concentration, and the density of water. The model correlated well with experimental results, suggesting that if adsorption theory dominates the adhesive bonding at the adhesive/substrate interface, the model should accurately predict the loss in interfacial fracture toughness for a given moisture concentration. The predictive model provides a useful tool for developing new adhesives, innovative surface treatment methods, and effective protection methodologies for enhancing interfacial adhesion.

Acknowledgments The work reported here is based on the PhD theses of two former students, Matthew Yao and Timothy Ferguson.

References

L. Butkus, Environmental durability of adhesively bonded joints, Doctoral thesis, Georgia Institute of Technology, Woodruff School of Mechanical Engineering, Atlanta, 1997

Y. Cengel, M. Boles, *Thermodynamics: An Engineering Approach* (McGraw-Hill, New York, 1994)

K. Cho, E. Cho, Effect of the microstructure of copper oxide on the adhesion behavior or epoxy/copper leadframe joints. J. Adhes. Sci. Technol. **14**(11), 1333–1353 (2000)

C. Chong, A. Leslie, L. Beng, C. Lee, Investigation on the effect of copper leadframe oxidation on package delamination, in *Proceedings of the 45th Electronic Components and Technology Conference*, (1995), pp. New York, N.Y. 463–469

P. Chung, M. Yuen, P. Chan, N. Ho, D. Lam, Effect of copper oxide on the adhesion behavior of epoxy molding compound-copper interface, in *Proceedings of the 52nd Electronic Components and Technology Conference*, 2002, pp. 1665–1670

J. Comyn, C. Groves, R. Saville, Durability in high humidity of glass-to-lead alloy joints bonded with and epoxide adhesive. Int. J. Adhes. Adhes. **14**, 15–20 (1994)

T. Ferguson, J. Qu, Moisture and temperature effects on the reliability of interfacial adhesion of a polymer/metal interface, in *Proceedings of the 54th Electronic Components and Technology Conference*, 2004

R. Gledhill, A. Kinloch, J. Shaw, A model for predicting joint durability. J. Ahdes. **11**, 3–15 (1980)

K. Hong, H. Imadojemu, R. Webb, Effects of oxidation and surface roughness on contact angle. Exp. Therm. Fluid Sci. **8**, 279–285 (1994)

J. Hutchinson, Z. Suo, Mixed mode cracking in layered materials, in *Advances in Applied Mechanics*, vol. 29 (Academic, New York, 1992)

S. Kim, The role of plastic package adhesion in IC performance, in *Proceedings of the 41st Electronic Components and Technology Conference*, 1991, pp. 750–758

A. Kinloch, Interfacial fracture mechanical aspects of adhesive bonded joints – a review. J. Adhes. **10**, 193–219 (1979)

A.J. Kinloch, *Adhesion and Adhesives Science and Technology* (Chapman and Hall, London, 1987)

H. Lee, J. Qu, Microstructure, adhesion strength and failure path at a polymer/roughened metal interface. J. Adhes. Sci. Technol. **17**(2), 195–215 (2003)

S. Luo, Study on adhesion of underfill materials for flip chip packaging, Doctoral thesis, Georgia Institute of Technology, School of Textile and Fibers Engineering, Atlanta, 2003

T. Mino, K. Sawada, A. Kurosu, M. Otsuka, N. Kawamura, H. Yoo, Development of moisture-proof thin and large QFP with copper lead frame, in *Proceedings of the 48th Electronic Components and Technology Conference*, 1998, pp. 1125–1131

S. Shi, C.P. Wong, Study of the fluxing agent effects on the properties of no-flow underfill materials for flip-chip applications, in *Proceedings of the 48th Electronic Components and Technology Conference*, 1998, pp. 117–124

C. Soles, A. Yee, A discussion of the molecular mechanisms of moisture transport in epoxy resins. J. Polym. Sci. B **38**, 792–802 (2000)

C. Soles, F. Chang, D. Gidley, A. Yee, Contributions of the nanovoid structure to the kinetics of moisture transport in epoxy resins. J. Polym. Sci. B **38**, 776–791 (2000)

J. Wylde, J. Spelt, Measurement of adhesive joint fracture properties as a function of environmental degradation. Int. J. Adhes. Adhes. **18**, 237–246 (1998)

S. Yi, C. Yue, J. Hsieh, L. Fong, S. Lahiri, Effects of oxidation and plasma cleaning on the adhesion strength of molding compounds to copper leadframes. J. Adhes. Sci. Technol. **13**, 789–804 (1999)

Damage Mechanics Unified Constitutive Modeling for Polymers

22

Cemal Basaran and Eray Gunel

Contents

Abstract

Adequacy of a material model that is based on isothermal properties becomes questionable for non-isothermal cases such as polymer processing which includes continuous change in temperature and rate. For true prediction of thermomechanical response of amorphous polymers under non-isothermal conditions, it is necessary to formulate temperature-dependent material properties and flow rules to provide a smooth transition around glass transition temperature.

An improved version of dual-mechanism viscoplastic constitutive model is presented that is used to describe thermomechanical response of amorphous polymers below and above glass transition temperature. Material property definitions, evolution of internal state variables, and plastic flow rules were revisited to provide a smooth and continuous transition in material response

C. Basaran (✉) • E. Gunel
Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA
e-mail: cjb@buffalo.edu; eraygunel@gmail.com

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_28

around glass transition temperature, θ_g. The elastic-viscoplastic constitutive model is developed based on a thermodynamics framework. For damage evolution in complex thermomechanical problems such as polymer processing, irreversible entropy production rate is used as a damage metric [a.k.a Basaran Damage Evolution Model].

Introduction

Particle-filled acrylics are widely used in domestic and industrial applications due to their aesthetic quality, high heat and weather resistance. Applications of acrylic particulate composites focus on solid surface industry in place of traditional materials such as natural stone, wood, laminate products. In addition, a new trend for designers and architects is to use acrylic sheets for artistic works such as wall panels, furniture, and sculptures. Thermoforming is the typical manufacturing method of acrylic products during which desired structural design is achieved by stretching heated panels over or into molds by manual or pneumatic methods. Large deformation of particle reinforced acrylics at elevated temperatures followed by cooling of material under fixed configuration may create several problems in parts both during and after thermoforming procedure. Stress whitening is one of the major problems observed in thermoformed particle acrylics.

Computational mechanic models can be used for simulation of the manufacturing processes. However, this requires sophisticated constitutive models that can simulate the material behavior accurately. The applicability of material models developed for large deformation behavior of amorphous polymers over a wide range of temperatures and strain rates has only been verified so far under completely controlled cases (constant strain rate, constant temperature). However, in actual polymer processing operations, temperature drop from temperatures above to temperatures below causes a transition from rubbery state to solid state and a significant change in material response. In order to resolve problems associated with material response predictions around glass transition temperature, an improved version of dual-mechanism elastic-viscoplastic material model is presented in this chapter. Evolution of state variables and flow rules that control viscoplastic characteristics in the model and material property definitions is carefully formulated to ensure smooth transition in response around glass transition temperature. Since material state continuously changes with temperature during non-isothermal cases, experimental measures cannot be used for continuous monitoring of damage evolution. A new method for damage quantification and evolution in polymer processing technique has been proposed based on irreversible entropy production during the process. An entropy based damage model is used to predict the material degradation and failure in polymer processing operations.

Large Deformation Concept

For studying large deformation behavior of polymeric materials, it is necessary to describe kinematics of constitutive model based on finite deformation tensors. Consider a body with a volume of V_o in undeformed (original or initial) configuration ($\boldsymbol{\Sigma}_o$) at time t_o which deforms into a volume V in current (deformed) configuration ($\boldsymbol{\Sigma}$) at time t, as depicted in Fig. 1. Motion of a point in the body can be uniquely defined with a continuous one-to-one mapping ($\boldsymbol{\chi}$) of position vectors $\mathbf{X}$ and $\mathbf{x}$ in original configuration and current configuration, respectively (Fig. 1).

Gradient of deformation ($\mathbf{F}$) describes transformation of a line element ($\mathbf{dX}$) at position of $\mathbf{X}$ in original configuration to a deformed line element ($\mathbf{dx}$) at position of $\mathbf{x}$ in current configuration (Eq. 1). Unique transformation ensures a non-singular, nonnegative determinant of deformation gradient or Jacobian (J) (Eq. 3). Velocity gradient ($\mathbf{L}$) is the gradient of velocity field ($\mathbf{v}$) and relates deformation gradient to its material time derivative $\left(\dot{\mathbf{F}}\right)$ (Eqs. 4–6):

$$\mathbf{x} = \boldsymbol{\chi}(\mathbf{X}, t) \tag{1}$$

$$\mathbf{F} = \nabla_{\mathbf{X}}(\chi) = \frac{\partial \boldsymbol{\chi}}{\partial \mathbf{X}} \tag{2}$$

$$J = \det(\mathbf{F}) > 0 \tag{3}$$

$$\mathbf{v}(\mathbf{x}, t) = \dot{\mathbf{x}}(\mathbf{x}, t) \tag{4}$$

$$\mathbf{L} = \operatorname{grad}(\mathbf{v}) = \frac{\partial \dot{\mathbf{x}}}{\partial \mathbf{X}} \tag{5}$$

$$\dot{\mathbf{F}} = \mathbf{LF} \tag{6}$$

Deformation gradient ($\mathbf{F}$) is assumed to multiplicatively decompose into elastic part ($\mathbf{F}^e$) and plastic part ($\mathbf{F}^p$) (Ames et al. 2009; Anand 1986), and information related to thermal deformation is included in the elastic part of deformation gradient (Eq. 7). Plastic deformation gradient ($\mathbf{F}^p$) defines a deformation from initial configuration ($\boldsymbol{\Sigma}_o$) to an intermediate, "relaxed," or "natural" configuration ($\boldsymbol{\Sigma}'$) which is followed by an elastic transformation ($\mathbf{F}^e$) to the final configuration ($\boldsymbol{\Sigma}$). Symmetric part of velocity gradient ($\mathbf{L}$) is defined as stretching rate tensor ($\mathbf{D}$), and asymmetric part of velocity gradient is defined as spin tensor ($\mathbf{W}$) (Eq. 10):

$$\mathbf{F} = \mathbf{F}^e\mathbf{F}^p; J = J^e J^p \tag{7}$$

$$J^e = \det(\mathbf{F}^e) > 0 \tag{8}$$

$$J^p = \det(\mathbf{F}^p) > 0 \tag{9}$$

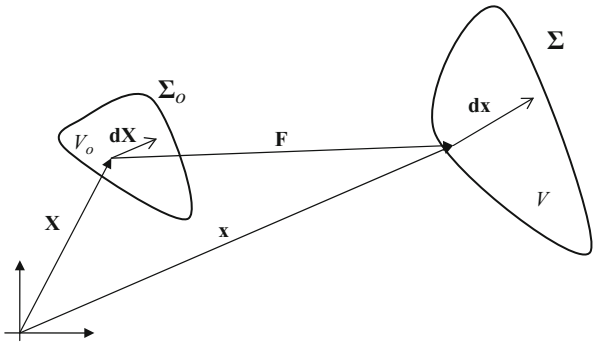


Fig. 1 Deformation from original configuration to current configuration

$$\text{sym}(\mathbf{L}) = \mathbf{D}; \text{asym}(\mathbf{L}) = \mathbf{W}; \mathbf{L} = \mathbf{D} + \mathbf{W} \tag{10}$$

Using Eq. 7, elastic and plastic part of velocity gradient can be found as

$$\mathbf{L} = \mathbf{L}^e + \mathbf{F}^e \mathbf{L}^p \mathbf{F}^{e-1} \tag{11}$$

where

$$\mathbf{L}^e = \dot{\mathbf{F}}^e \mathbf{F}^{e-1} \tag{12}$$

$$\mathbf{L}^p = \dot{\mathbf{F}}^p \mathbf{F}^{p-1} \tag{13}$$

Stretching rate tensor (**D**) and spin tensor (**W**) for elastic and plastic parts can be derived similarly from Eq. 10 where superscripts "e" and "p" represent elastic and plastic parts of corresponding quantity, respectively (Eqs. 14 and 15):

$$\text{sym}(\mathbf{L}^e) = \mathbf{D}^e; \text{asym}(\mathbf{L}^e) = \mathbf{W}^e; \mathbf{L}^e = \mathbf{D}^e + \mathbf{W}^e \tag{14}$$

$$\text{sym}(\mathbf{L}^p) = \mathbf{D}^p; \text{asym}(\mathbf{L}^p) = \mathbf{W}^p; \mathbf{L}^p = \mathbf{D}^p + \mathbf{W}^p \tag{15}$$

Right and left stretch tensors (**U**, **V**) and rotation tensor (**R**) can be found from right and left polar decompositions of deformation gradient as follows:

$$\mathbf{F} = \mathbf{RU} \tag{16}$$

$$\mathbf{F} = \mathbf{VR} \tag{17}$$

U and **V** stretch tensors are positive-definite symmetric tensors, and **R** is a proper orthogonal tensor. Cauchy (**C**) and Almansi (**B**) tensors can be formulated as follows:

$$\mathbf{C} = \mathbf{F}^T \mathbf{F} = \mathbf{UU} \tag{18}$$

$$\mathbf{B}^{-1} = \mathbf{F}^{-T} \mathbf{F}^{-1} = \mathbf{V}^{-1} \mathbf{V}^{-1} \tag{19}$$

Similarly, from Eqs. 16–19, the following relations can be written for elastic and plastic part of deformation gradient:

$$\mathbf{F}^e = \mathbf{R}^e\mathbf{U}^e \tag{20}$$

$$\mathbf{F}^e = \mathbf{V}^e\mathbf{R}^e \tag{21}$$

$$\mathbf{C}^e = \mathbf{F}^{eT}\mathbf{F}^e = \mathbf{U}^e\mathbf{U}^e \tag{22}$$

$$\mathbf{B}^{e-1} = \mathbf{F}^{e-T}\mathbf{F}^{e-1} = \mathbf{V}^{e-1}\mathbf{V}^{e-1} \tag{23}$$

$$\mathbf{F}^p = \mathbf{R}^p\mathbf{U}^p \tag{24}$$

$$\mathbf{F}^p = \mathbf{V}^p\mathbf{R}^p \tag{25}$$

$$\mathbf{C}^p = \mathbf{F}^{pT}\mathbf{F}^p = \mathbf{U}^p\mathbf{U}^p \tag{26}$$

$$\mathbf{B}^{p-1} = \mathbf{F}^{p-T}\mathbf{F}^{p-1} = \mathbf{V}^{p-1}\mathbf{V}^{p-1} \tag{27}$$

Multiplicative split of deformation gradient produces nonunique, locally defined intermediate (relaxed) configurations (Anand and On 1979). One convenient way to solve this problem is to introduce corotational rate definitions of elastic and plastic deformation gradients based on (director) orthonormal vectors which was proposed by Mandel (Anand et al. 2009). Resulting elastic and plastic parts of rate deformation tensors are invariant upon superposed rigid body rotations (Aravas 1994; Arruda and Boyce 1993; Arruda et al. 1995). However, arbitrariness of intermediate configuration can be also removed by setting plastic spin tensor equal to zero (Eq. 28; Basaran and Lin 2007; Basaran and Nie 2007; Basaran and Yan 1998; Bauwens-Crowet et al. 1969). In this case, elastic and plastic deformation gradients will include rotations which can be handled with proper selection of stress and rate measures establishing a frame-indifferent model:

$$\mathbf{W}^p = 0 \tag{28}$$

$$J^p = 1 \tag{29}$$

Assuming that plastic flow is irrotational (Eq. 28) and incompressible (Eq. 29), Jacobian of total deformation gradient (J) and material time derivative of plastic deformation gradient $\left(\dot{\mathbf{F}}^p\right)$ will become

$$J = J^e \tag{30}$$

$$\dot{\mathbf{F}}^p = \mathbf{D}^p\mathbf{F}^p \tag{31}$$

As a result of incompressible plastic flow and irrotational plastic flow assumptions,

$$\mathbf{L}^p = \mathbf{D}^p \tag{32}$$

$$\mathrm{tr}(\mathbf{L}^p) = \mathrm{tr}(\mathbf{D}^p) = 0 \tag{33}$$

$$\mathbf{L} = \mathbf{L}^e + \mathbf{F}^e\mathbf{D}^p\mathbf{F}^{e-1} \tag{34}$$

System is assumed to be at rest originally ($t = 0$) which provides the following initial conditions for elastic and plastic deformation gradient:

$$\mathbf{F}^p(\mathbf{X}, 0) = \mathbf{I} \tag{35}$$

$$\mathbf{F}^e(\mathbf{X}, 0) = \mathbf{I} \tag{36}$$

Frame Indifference of Kinematic Descriptors

According to Eringen (1967), material constitutive relations must be invariant to changes of current reference frame (Bergström and Boyce 1998). Therefore, a quantity or an equation is frame indifferent or objective if it is invariant to changes of reference frame. Let rigid body motion of any material point $\boldsymbol{\chi}(\mathbf{X}, t)$ be defined by a proper orthogonal rotation tensor $\boldsymbol{\Omega}(t)$ and a vector $\mathbf{X_o}(t)$ for any time frame t as

$$\overline{\boldsymbol{\chi}}(\mathbf{X}, t) = \boldsymbol{\Omega}(t)[\boldsymbol{\chi}(\mathbf{X}, t) - \mathbf{O}] + \mathbf{X_o}(t) \tag{37}$$

where orthogonal characteristic of rotation tensor is defined as

$$\boldsymbol{\Omega}^T\boldsymbol{\Omega} = \mathbf{I} \tag{38}$$

In Eq. 37, $\boldsymbol{\Omega}(t)$ represents rigid body rotation and $\mathbf{X_o}(t)$ represents rigid body translation. In this context, frame indifference of a vector ($\mathbf{a}$) and a 2nd-order tensor ($\mathbf{A}$) are defined in terms of transformation rules (Bergström and Boyce 1998; Boltzmann 1995) as follows:

$$\overline{\mathbf{a}} = \boldsymbol{\Omega}\mathbf{a} \tag{39}$$

$$\overline{\mathbf{A}} = \boldsymbol{\Omega}\mathbf{A}\boldsymbol{\Omega}^T \tag{40}$$

Transformation of deformation gradient with respect to a change in reference frame can be obtained from Eq. 1 as

$$\overline{\mathbf{F}} = \boldsymbol{\Omega}\mathbf{F} \tag{41}$$

Therefore, deformation gradient is not objective (frame indifferent) according to Eq. 40. Using Eq. 41, Cauchy tensor ($\mathbf{C}$) in transformed configuration can be obtained as

$$\overline{\mathbf{C}} = \overline{\mathbf{F}}^T \overline{\mathbf{F}} = \mathbf{F}^T \boldsymbol{\Omega}^T \boldsymbol{\Omega} \mathbf{F} = \mathbf{F}^T \mathbf{F} = \mathbf{C} \tag{42}$$

Since Cauchy tensor is a Lagrangian tensor referring to original reference frame, Cauchy tensor should not be expected to change with current reference frame according to Eq. 40. Therefore, Cauchy tensor (**C**) is not objective, but still invariant to changes in current reference frame in deformed configuration. Using Eq. 41 and right polar decomposition rule in Eq. 16,

$$\overline{\mathbf{F}} = \overline{\mathbf{R}}\,\overline{\mathbf{U}} = \boldsymbol{\Omega} \mathbf{F} = \boldsymbol{\Omega} \mathbf{R} \mathbf{U} \tag{43}$$

Since right polar decomposition is unique, Eq. 43 implies that

$$\overline{\mathbf{R}} = \boldsymbol{\Omega} \mathbf{R} \tag{44}$$

$$\overline{\mathbf{U}} = \mathbf{U} \tag{45}$$

Therefore, right stretch tensor (**U**) and rotation tensor (**R**) are not objective. However, similar to Cauchy tensor, right stretch tensor refers to original reference frame, and hence right stretch tensor is invariant to changes in current reference frame. Similarly, using Eq. 19, Almansi tensor (**B**) in the transformed configuration can be obtained as

$$\overline{\mathbf{B}} = \overline{\mathbf{F}}\,\overline{\mathbf{F}}^T = \boldsymbol{\Omega} \mathbf{F} \mathbf{F}^T \boldsymbol{\Omega}^T = \boldsymbol{\Omega} \mathbf{B} \boldsymbol{\Omega}^T \tag{46}$$

Therefore, Almansi tensor (**B**) is objective. Using Eq. 41 and left polar decomposition rule in Eq. 17,

$$\overline{\mathbf{F}} = \overline{\mathbf{V}}\,\overline{\mathbf{R}} = \boldsymbol{\Omega} \mathbf{F} = \boldsymbol{\Omega} \mathbf{V} \mathbf{R} \tag{47}$$

According to transformation of rotation tensor in Eqs. 44, and 47 can be rewritten as

$$\overline{\mathbf{F}} = \overline{\mathbf{V}}\,\overline{\mathbf{R}} = \boldsymbol{\Omega} \mathbf{V} \boldsymbol{\Omega}^T \boldsymbol{\Omega} \mathbf{R} \tag{48}$$

Since left polar decomposition is also unique, Eq. 48 implies that

$$\overline{\mathbf{V}} = \boldsymbol{\Omega} \mathbf{V} \boldsymbol{\Omega}^T \tag{49}$$

Therefore, left stretch tensor (**V**) is objective. Taking time derivative of Eq. 41, it can be shown that material time derivative of deformation gradient is not objective (Eq. 50):

$$\overline{\mathbf{F}}. = \dot{\boldsymbol{\Omega}} \mathbf{F} + \boldsymbol{\Omega} \dot{\mathbf{F}} \tag{50}$$

Using Eq. 50 and the definition of velocity gradient in Eq. 6,

$$\overline{\mathbf{F}}. = \overline{\mathbf{L}}\,\overline{\mathbf{F}} = \dot{\mathbf{\Omega}}\mathbf{F} + \mathbf{\Omega}\dot{\mathbf{F}} = \left(\dot{\mathbf{\Omega}}\mathbf{F}\mathbf{F}^{-1}\mathbf{\Omega}^{T} + \mathbf{\Omega}\dot{\mathbf{F}}\mathbf{F}^{-1}\mathbf{\Omega}^{T}\right)\mathbf{\Omega}\mathbf{F} \tag{51}$$

Equation 51 implies that velocity gradient is not objective as shown in Eq. 52:

$$\overline{\mathbf{L}} = \mathbf{\Omega}\mathbf{L}\mathbf{\Omega}^{T} + \dot{\mathbf{\Omega}}\mathbf{\Omega}^{T} \tag{52}$$

Using definitions of stretch rate tensor and spin tensor in Eq. 10,

$$\overline{\mathbf{L}} = \overline{\mathbf{D}} + \overline{\mathbf{W}} = \mathbf{\Omega}(\mathbf{D} + \mathbf{W})\mathbf{\Omega}^{T} + \dot{\mathbf{\Omega}}\mathbf{\Omega}^{T} \tag{53}$$

which implies that

$$\overline{\mathbf{D}} = \mathbf{\Omega}\mathbf{D}\mathbf{\Omega}^{T} \tag{54}$$

$$\overline{\mathbf{W}} = \mathbf{\Omega}\mathbf{W}\mathbf{\Omega}^{T} + \dot{\mathbf{\Omega}}\mathbf{\Omega}^{T} \tag{55}$$

Therefore, stretch rate tensor ($\mathbf{D}$) is objective, while spin tensor ($\mathbf{W}$) is not objective. Using multiplicative decomposition definition of deformation gradient in Eq. 7,

$$\overline{\mathbf{F}} = \overline{\mathbf{F}^{e}}\,\overline{\mathbf{F}^{p}} = \mathbf{\Omega}\mathbf{F}^{e}\mathbf{F}^{p} \tag{56}$$

Equation 56 implies that

$$\overline{\mathbf{F}^{e}} = \mathbf{\Omega}\mathbf{F}^{e} \tag{57}$$

$$\overline{\mathbf{F}^{p}} = \mathbf{F}^{p} \tag{58}$$

Therefore, elastic and plastic deformation gradients are not objective. Since plastic deformation gradient refers to original and intermediate reference frames, plastic deformation gradient is not objective but invariant to changes in current reference frame. Using definitions for elastic and plastic rotation tensors, left stretch tensor, right stretch tensor, Cauchy tensor, and Almansi tensor in Eqs. 20–27, the following relations can be derived:

$$\overline{\mathbf{F}^{e}} = \overline{\mathbf{R}^{e}}\,\overline{\mathbf{U}^{e}} = \mathbf{\Omega}\mathbf{R}^{e}\mathbf{U}^{e} \tag{59}$$

$$\overline{\mathbf{R}^{e}} = \mathbf{\Omega}\mathbf{R}^{e} \tag{60}$$

$$\overline{\mathbf{U}^{e}} = \mathbf{U}^{e} \tag{61}$$

$$\overline{\mathbf{F}^{e}} = \overline{\mathbf{V}^{e}}\,\overline{\mathbf{R}^{e}} = \mathbf{\Omega}\mathbf{V}^{e}\mathbf{\Omega}^{T}\mathbf{\Omega}\mathbf{R}^{e} \tag{62}$$

$$\overline{\mathbf{V}^{e}} = \mathbf{\Omega}\mathbf{V}^{e}\mathbf{\Omega}^{T} \tag{63}$$

$$\overline{\mathbf{C}^e} = \overline{\mathbf{U}^e}\,\overline{\mathbf{U}^e} = \mathbf{U}^e\mathbf{U}^e = \mathbf{C}^e \tag{64}$$

$$\overline{\mathbf{B}^e} = \overline{\mathbf{V}^e}\,\overline{\mathbf{V}^e} = \boldsymbol{\Omega}\mathbf{V}^e\boldsymbol{\Omega}^T\boldsymbol{\Omega}\mathbf{V}^e\boldsymbol{\Omega}^T = \boldsymbol{\Omega}\mathbf{B}^e\boldsymbol{\Omega}^T \tag{65}$$

$$\overline{\mathbf{F}^p} = \overline{\mathbf{R}^p}\,\overline{\mathbf{U}^p} = \boldsymbol{\Omega}\mathbf{R}^p\mathbf{U}^p \tag{66}$$

$$\overline{\mathbf{R}^p} = \boldsymbol{\Omega}\mathbf{R}^p \tag{67}$$

$$\overline{\mathbf{U}^p} = \mathbf{U}^p \tag{68}$$

$$\overline{\mathbf{F}^p} = \overline{\mathbf{V}^p}\,\overline{\mathbf{R}^p} = \boldsymbol{\Omega}\mathbf{V}^p\boldsymbol{\Omega}^T\boldsymbol{\Omega}\mathbf{R}^p \tag{69}$$

$$\overline{\mathbf{V}^p} = \boldsymbol{\Omega}\mathbf{V}^p\boldsymbol{\Omega}^T \tag{70}$$

$$\overline{\mathbf{C}^p} = \overline{\mathbf{U}^p}\,\overline{\mathbf{U}^p} = \mathbf{U}^p\mathbf{U}^p = \mathbf{C}^p \tag{71}$$

$$\overline{\mathbf{B}^p} = \overline{\mathbf{V}^p}\,\overline{\mathbf{V}^p} = \boldsymbol{\Omega}\mathbf{V}^p\boldsymbol{\Omega}^T\boldsymbol{\Omega}\mathbf{V}^p\boldsymbol{\Omega}^T = \boldsymbol{\Omega}\mathbf{B}^p\boldsymbol{\Omega}^T \tag{72}$$

Therefore, elastic and plastic rotation tensors ($\mathbf{R}^e$, $\mathbf{R}^p$) are not objective. Elastic and plastic right stretch tensors and Cauchy tensors ($\mathbf{C}^e$, $\mathbf{C}^p$, $\mathbf{U}^e$, $\mathbf{U}^p$) are not objective, but invariant under changes in current reference frame. Elastic and plastic left stretch tensors and Almansi tensors ($\mathbf{B}^e$, $\mathbf{B}^p$, $\mathbf{V}^e$, $\mathbf{V}^p$) are objective. Using elastic and plastic velocity gradient definitions in Eqs. 12 and 13 with transformation rules for elastic and plastic deformation gradient in Eqs. 57 and 58,

$$\overline{\mathbf{L}^e} = \dot{\overline{\mathbf{F}^e}}\,\overline{\mathbf{F}^e}^{-1} = \left(\boldsymbol{\Omega}\mathbf{F}^e + \dot{\boldsymbol{\Omega}}\mathbf{F}^e\right)\left(\mathbf{F}^{e-1}\boldsymbol{\Omega}^T\right) = \boldsymbol{\Omega}\mathbf{L}^e\boldsymbol{\Omega}^T + \dot{\boldsymbol{\Omega}}\boldsymbol{\Omega}^T \tag{73}$$

$$\overline{\mathbf{L}^p} = \dot{\overline{\mathbf{F}^p}}\,\overline{\mathbf{F}^p}^{-1} = \dot{\mathbf{F}}^p\mathbf{F}^{p-1} = \mathbf{L}^p \tag{74}$$

Equation 73 shows that elastic velocity gradient is not objective. According to Eq. 74, plastic velocity gradient is not objective but invariant to changes in current reference frame, since plastic velocity gradient refers to original and intermediate reference frames, but does not refer to current reference frame. Using definitions of elastic stretch rate tensor and spin tensor in Eq. 14,

$$\overline{\mathbf{L}^e} = \overline{\mathbf{D}^e} + \overline{\mathbf{W}^e} = \boldsymbol{\Omega}(\mathbf{D}^e + \mathbf{W}^e)\boldsymbol{\Omega}^T + \dot{\boldsymbol{\Omega}}\boldsymbol{\Omega}^T \tag{75}$$

which implies that

$$\overline{\mathbf{D}^e} = \boldsymbol{\Omega}\mathbf{D}^e\boldsymbol{\Omega}^T \tag{76}$$

$$\overline{\mathbf{W}^e} = \boldsymbol{\Omega}\mathbf{W}^e\boldsymbol{\Omega}^T + \dot{\boldsymbol{\Omega}}\boldsymbol{\Omega}^T \tag{77}$$

Therefore, elastic stretch rate tensor ($\mathbf{D}^e$) is objective, while elastic spin tensor ($\mathbf{W}^e$) is not objective. Similarly, using plastic stretch rate definition in Eq. 15 and applying irrotational plastic flow assumption in Eq. 28,

$$\overline{\mathbf{L}^p} = \overline{\mathbf{D}^p} = \mathbf{D}^p \tag{78}$$

which implies that

$$\overline{\mathbf{D}^p} = \mathbf{D}^p \tag{79}$$

Therefore, plastic stretch rate tensor ($\mathbf{D}^p$) is not objective but invariant to changes in current reference frame. To summarize, it has been shown that $\mathbf{F}, \mathbf{R}, \dot{\mathbf{F}}$, $\mathbf{L}, \mathbf{W}, \mathbf{F}^e, \mathbf{R}^e, \mathbf{R}^p, \mathbf{L}^e, \mathbf{W}^e$ are not objective; $\mathbf{C}$, $\mathbf{U}$, $\mathbf{F}^p$, $\mathbf{U}^e$, $\mathbf{C}^e$, $\mathbf{U}^p$, $\mathbf{C}^p$, $\mathbf{L}^p$, $\mathbf{D}^p$ are not objective but invariant to changes in current reference frame and $\mathbf{B}$, $\mathbf{V}$, $\mathbf{D}$, $\mathbf{V}^e$, $\mathbf{B}^e$, $\mathbf{V}^p$, $\mathbf{B}^p$, $\mathbf{D}^e$ are objective.

Thermodynamic Framework

Definition of a thermodynamic potential which is concave with respect to temperature (θ) and convex with respect to other internal state variables will provide a basis for satisfying conditions of thermodynamic stability imposed by Clausius-Duhem inequality. Specific Helmholtz free energy (ψ) forms such a basis which is defined as difference between specific internal energy (u) and product of absolute temperature (θ) and specific entropy (s) (Eq. 80):

$$\psi = u - \theta s \tag{80}$$

The first law of thermodynamics states that energy can be transported, or converted from one form to another, but cannot be destroyed or created. Accordingly, internal energy of a system can be increased by heat flow into the system (δq), heat generated within the system by external agencies (e.g., inductive heating) (r), mechanical work done on the system by external pressure (δw), or all other kinds of work done on the system ($\delta w'$) during any process. Considering only thermomechanical effects, rate of change in internal energy can be stated as

$$\dot{u} = Q + l + r \tag{81}$$

In Eq. 81, Q is rate of net heat flow into the system and l is rate of net work done on the system. Second law of thermodynamics states that there exists a state function, entropy, which increases in universe for all types of processes due to entropy production. Unlike energy, entropy (s) is not only transferred across boundaries of system, but it may be created in the system which is called as entropy production (η). Clausius-Duhem inequality describes second law of thermodynamics in terms of a nonnegative entropy production rate (γ) per unit volume for any kind of irreversible process which is defined as

$$\gamma = \rho \dot{s} + \operatorname{div}(\mathbf{J}_s) > 0 \tag{82}$$

$\mathbf{J}_s$ in Eq. 82 is net entropy flux into system. Heat flow into body can be defined in terms of heat flux as shown in Eq. 83.

$$\rho Q = -\mathrm{div}\left(\mathbf{J}_q\right) \tag{83}$$

Substituting time derivative of Eq. 80, internal energy definition in Eq. 81, and heat flow equation in Eq. 83 into Eq. 82, internal entropy production density rate can be rewritten as

$$\gamma = \mathrm{div}(\mathbf{J}_s) - \frac{\mathrm{div}\left(\mathbf{J}_q\right)}{\theta} + \frac{\rho r}{\theta} + \frac{\rho}{\theta}\left(l - \dot{\psi} - \dot{\theta} s\right) > 0 \tag{84}$$

r in Eq. 84 is internal heat source strength.

Thermodynamic Restrictions

In dual-mechanism viscoplastic constitutive model, material response is resolved into two components which necessitate multi-mechanism generalization of multiplicative decomposition in Eq. 7 and description of different Helmholtz free energy functions and associated entropy functions assuming that linear addition is applicable for Eq. 80. Accordingly, Eqs. 12–15, 20–27, and 31–36 hold true for each component of resistance. Subscripts "I" and "M" will be used henceforth to designate the component of a quantity in intermolecular mechanism and molecular network mechanism, respectively. For description of dissipation inequality, total Helmholtz free energy density in (original) reference configuration is written as summation of defect energy (Ψ_D) and elastic energy stored in intermolecular structure (Ψ_I) and molecular network structure (Ψ_M):

$$\Psi\left(\mathbf{C}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right) = \Psi_I\left(\mathbf{E}_I^e, \theta\right) + \Psi_M\left(\mathbf{C}_M^e, \theta\right) + \Psi_D(\mathbf{A}, \theta) \tag{85}$$

In Eq. 85, defect energy (Ψ_D) is assumed to depend on a stretch-like tensor ($\mathbf{A}$) and temperature (θ), elastic energy in intermolecular structure (Ψ_I) is assumed to depend on logarithmic elastic strain in intermolecular structure ($\mathbf{E}_I^e$) and temperature (θ), and elastic energy in molecular network structure (Ψ_M) is assumed to depend on elastic Cauchy tensor in molecular network structure ($\mathbf{C}_M^e$) and temperature (θ). Assuming that a similar decomposition also holds for specific entropy and specific Helmholtz free energy,

$$s\left(\mathbf{C}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right) = s_I\left(\mathbf{E}_I^e, \theta\right) + s_M\left(\mathbf{C}_M^e, \theta\right) + s_D(\mathbf{A}, \theta) \tag{86}$$

$$\psi\left(\mathbf{C}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right) = \psi_I\left(\mathbf{E}_I^e, \theta\right) + \psi_M\left(\mathbf{C}_M^e, \theta\right) + \psi_D(\mathbf{A}, \theta) \tag{87}$$

Helmholtz free energy density in reference configuration (Ψ) can be simply related to specific Helmholtz free energy function (ψ) through Eq. 88 and in rate form through Eq. 89:

$$\rho_o \psi\left(\mathbf{C}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right) = \Psi\left(\mathbf{C}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right) \tag{88}$$

$$\rho \dot{\psi}\left(\mathbf{C}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right) = J^{-1} \dot{\Psi}\left(\mathbf{C}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right) \tag{89}$$

In Eqs. 88 and 89, ρ_o and ρ are densities in reference configuration and deformed configuration, respectively. Note that since eigenvalues of elastic Cauchy tensor ($\mathbf{C}_I^e$) and logarithmic elastic strain tensor ($\mathbf{E}_I^e$) corresponding to intermolecular structure are related through Eq. 90 and eigenvectors of these tensors are identical, it is possible to consider Helmholtz free energy density (Ψ_I) and specific entropy (s_I) associated with intermolecular structure as functions of temperature (θ) and elastic Cauchy tensor ($\mathbf{C}_I^e$) (Eqs. 91 and 92):

$$\text{eigenval}\left(\mathbf{E}_I^e\right) = \frac{1}{2} \ln\left(\text{eigenval}\left(\mathbf{C}_I^e\right)\right) \tag{90}$$

$$\Psi_I\left(\mathbf{E}_I^e, \theta\right) \equiv \Psi_I\left(\mathbf{C}_I^e, \theta\right) \tag{91}$$

$$s_I\left(\mathbf{E}_I^e, \theta\right) \equiv s_I\left(\mathbf{C}_I^e, \theta\right) \tag{92}$$

Time derivative of Helmholtz free energy can be formulated as

$$\begin{aligned} \dot{\Psi}\left(\mathbf{C}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right) = \Bigg[& \frac{\partial \Psi_I\left(\mathbf{E}_I^e, \theta\right)}{\partial \mathbf{C}_I^e} : \dot{\mathbf{C}}_I^e + \frac{\partial \Psi_M\left(\mathbf{C}_M^e, \theta\right)}{\partial \mathbf{C}_M^e} : \dot{\mathbf{C}}_M^e + \frac{\partial \Psi_D(\mathbf{A}, \theta)}{\partial \mathbf{A}} : \dot{\mathbf{A}} \\ & + \frac{\partial \Psi_I\left(\mathbf{E}_I^e, \theta\right)}{\partial \theta} : \dot{\theta} + \frac{\partial \Psi_M\left(\mathbf{C}_M^e, \theta\right)}{\partial \theta} : \dot{\theta} + \frac{\partial \Psi_D(\mathbf{A}, \theta)}{\partial \theta} : \dot{\theta} \Bigg] \end{aligned} \tag{93}$$

According to principle of virtual work (power), rate of work done per unit volume of deformed body (external work power) is balanced with internal work power (Eq. 94), while total work done on the system is stored as elastic strain energy (represented with the first two terms in Eq. 95) and plastically dissipated (represented with the last two terms in Eq. 95):

$$\rho l = \dot{w}_{\text{int}} \tag{94}$$

$$\dot{w}_{\text{int}} = \mathbf{\Gamma}_I^e : \mathbf{L}_I^e + \mathbf{\Gamma}_M^e : \mathbf{L}_M^e + J^{-1} \mathbf{\Gamma}_I^p : \mathbf{L}_I^p + J^{-1} \mathbf{\Gamma}_M^p : \mathbf{L}_M^p \tag{95}$$

In Eq. 95, $\mathbf{\Gamma}_I^e$, $\mathbf{\Gamma}_M^e$, $\mathbf{\Gamma}_I^p$, and $\mathbf{\Gamma}_M^p$ are stress measures conjugate to rate of deformation measures $\mathbf{L}_I^e$, $\mathbf{L}_M^e$, $\mathbf{L}_I^p$, and $\mathbf{L}_M^p$ which were defined in Eqs. 12 and 13. Noting that $J^{-1} = J^{e^{-1}}$ (Eq. 30), the J^{-1} multiplier in front of the last two terms of Eq. 95 recovers work power definitions from intermediate configuration to deformed configuration. Requirement of frame indifference of internal work power definition can be described as

$$\dot{w}_{\text{int}} = \dot{\overline{w}}_{\text{int}} \tag{96}$$

which implies that

$$\begin{pmatrix} \boldsymbol{\Gamma}_I^e : \mathbf{L}_I^e + \boldsymbol{\Gamma}_M^e : \mathbf{L}_M^e \\ + J^{-1}\boldsymbol{\Gamma}_I^p : \mathbf{L}_I^p + J^{-1}\boldsymbol{\Gamma}_M^p : \mathbf{L}_M^p \end{pmatrix} = \left(\overline{\boldsymbol{\Gamma}}_I^e : \overline{\mathbf{L}}_I^e + \overline{\boldsymbol{\Gamma}}_M^e : \overline{\mathbf{L}}_M^e + J^{-1}\overline{\boldsymbol{\Gamma}}_I^p : \overline{\mathbf{L}}_I^p + J^{-1}\overline{\boldsymbol{\Gamma}}_M^p : \overline{\mathbf{L}}_M^p \right) \tag{97}$$

Using transformation rules for elastic and plastic velocity gradients in Eqs. 73 and 74,

$$\begin{pmatrix} \boldsymbol{\Gamma}_I^e : \mathbf{L}_I^e + J^{-1}\boldsymbol{\Gamma}_I^p : \mathbf{L}_I^p \\ + \boldsymbol{\Gamma}_M^e : \mathbf{L}_M^e + J^{-1}\boldsymbol{\Gamma}_M^p : \mathbf{L}_M^p \end{pmatrix} = \begin{bmatrix} \overline{\boldsymbol{\Gamma}}_I^e : \left(\boldsymbol{\Omega}\mathbf{L}_I^e\boldsymbol{\Omega}^T + \dot{\boldsymbol{\Omega}}\boldsymbol{\Omega}^T \right) + J^{-1}\overline{\boldsymbol{\Gamma}}_I^p : \mathbf{L}_I^p + \overline{\boldsymbol{\Gamma}}_M^e : \\ \left(\boldsymbol{\Omega}\mathbf{L}_M^e\boldsymbol{\Omega}^T + \dot{\boldsymbol{\Omega}}\boldsymbol{\Omega}^T \right) + J^{-1}\overline{\boldsymbol{\Gamma}}_M^p : \mathbf{L}_M^p \end{bmatrix} \tag{98}$$

or

$$\begin{pmatrix} \boldsymbol{\Gamma}_I^e : \mathbf{L}_I^e + J^{-1}\boldsymbol{\Gamma}_I^p : \mathbf{L}_I^p \\ + \boldsymbol{\Gamma}_M^e : \mathbf{L}_M^e + J^{-1}\boldsymbol{\Gamma}_M^p : \mathbf{L}_M^p \end{pmatrix} = \begin{bmatrix} \left(\boldsymbol{\Omega}^T\overline{\boldsymbol{\Gamma}}_I^e\boldsymbol{\Omega} \right) : \mathbf{L}_I^e + \left(\boldsymbol{\Omega}^T\overline{\boldsymbol{\Gamma}}_M^e\boldsymbol{\Omega} \right) : \mathbf{L}_M^e + J^{-1}\overline{\boldsymbol{\Gamma}}_I^p : \mathbf{L}_I^p \\ + J^{-1}\overline{\boldsymbol{\Gamma}}_M^p : \mathbf{L}_M^p + \overline{\boldsymbol{\Gamma}}_I^e : \dot{\boldsymbol{\Omega}}\boldsymbol{\Omega}^T + + \overline{\boldsymbol{\Gamma}}_M^e : \dot{\boldsymbol{\Omega}}\boldsymbol{\Omega}^T \end{bmatrix} \tag{99}$$

The first two terms on right-hand side of Eq. 99 indicate that stress measures corresponding elastic work ($\boldsymbol{\Gamma}_I^e$, $\boldsymbol{\Gamma}_M^e$) are objective as shown in Eqs. 100 and 101. Since $\dot{\boldsymbol{\Omega}}\boldsymbol{\Omega}^T$ is a skew symmetric tensor (Eqs. 55 and 77), the last two terms on right-hand side of Eq. 99 imply that stress measures corresponding elastic work ($\boldsymbol{\Gamma}_I^e$, $\boldsymbol{\Gamma}_M^e$) are symmetric (Eqs. 102 and 103). The third and fourth terms on right-hand side of Eq. 99 show that stress measures corresponding plastic work ($\boldsymbol{\Gamma}_I^p$, $\boldsymbol{\Gamma}_M^p$) are not objective but invariant to changes in current reference frame as shown in Eqs. 104 and 105:

$$\overline{\boldsymbol{\Gamma}}_I^e = \boldsymbol{\Omega}\boldsymbol{\Gamma}_I^e\boldsymbol{\Omega}^T \tag{100}$$

$$\overline{\boldsymbol{\Gamma}}_M^e = \boldsymbol{\Omega}\boldsymbol{\Gamma}_M^e\boldsymbol{\Omega}^T \tag{101}$$

$$\overline{\boldsymbol{\Gamma}}_I^e = \overline{\boldsymbol{\Gamma}}_I^{eT}; \boldsymbol{\Gamma}_I^e = \boldsymbol{\Gamma}_I^{eT} \tag{102}$$

$$\overline{\boldsymbol{\Gamma}}_M^e = \overline{\boldsymbol{\Gamma}}_M^{eT}; \boldsymbol{\Gamma}_M^e = \boldsymbol{\Gamma}_M^{eT} \tag{103}$$

$$\overline{\boldsymbol{\Gamma}}_I^p = \boldsymbol{\Gamma}_I^p \tag{104}$$

$$\overline{\boldsymbol{\Gamma}}_M^p = \boldsymbol{\Gamma}_M^p \tag{105}$$

Using symmetry property of stress measures in Eqs. 102 and 103, irrotational plastic flow definition in Eq. 28, total internal work power over whole volume of system can be written as

$$\dot{W}_{\text{int}} = \int_V \dot{w}_{\text{int}} dV = \int_V \left[\boldsymbol{\Gamma}_I^e : \mathbf{D}_I^e + \boldsymbol{\Gamma}_M^e : \mathbf{D}_M^e + J^{-1}\boldsymbol{\Gamma}_I^p : \mathbf{D}_I^p + J^{-1}\boldsymbol{\Gamma}_M^p : \mathbf{D}_M^p \right] dV \tag{106}$$

Total external work power on the system can be described in terms of surface tractions on boundaries of the system and body force acting on the system as

$$\dot{W}_{\text{ext}} = \int_V \rho l dV = \int_S \underline{t} \bullet \dot{\chi} dS + \int_V \underline{b} \bullet \dot{\chi} dV \tag{107}$$

Consider principal of virtual power for a special case defined from Eq. 34 as

$$\mathbf{L} = \text{grad}(\dot{\chi}) = \mathbf{L}_I^e = \mathbf{L}_M^e \tag{108}$$

where it is assumed that

$$\mathbf{D}_I^p = \mathbf{D}_M^p = 0 \tag{109}$$

Principal of virtual power can be rewritten for this special case from Eqs. 106 and 107 as

$$\int_S \underline{t} \bullet \dot{\chi} dS + \int_V \underline{b} \bullet \dot{\chi} dV = \int_V \left[\boldsymbol{\Gamma}_I^e : \mathbf{D}_I^e + \boldsymbol{\Gamma}_M^e : \mathbf{D}_M^e \right] dV \tag{110}$$

$$\int_S \underline{t} \bullet \dot{\chi} dS + \int_V \underline{b} \bullet \dot{\chi} dV = \int_V \left[\left(\boldsymbol{\Gamma}_I^e + \boldsymbol{\Gamma}_M^e \right) : \text{grad}(\dot{\chi}) \right] dV \tag{111}$$

$$\int_S \underline{t} \bullet \dot{\chi} dS + \int_V \underline{b} \bullet \dot{\chi} dV = \int_V \text{div}\left[\dot{\chi} \bullet \left(\boldsymbol{\Gamma}_I^e + \boldsymbol{\Gamma}_M^e \right) \right] dV - \int_V \text{div}\left(\boldsymbol{\Gamma}_I^e + \boldsymbol{\Gamma}_M^e \right) \bullet \dot{\chi} dV \tag{112}$$

Since Eq. 112 is true for any choice of V and $\text{grad}(\dot{\chi})$, from the first terms on left- and right-hand side of Eq. 112,

$$\underline{t} = \left(\boldsymbol{\Gamma}_I^e + \boldsymbol{\Gamma}_M^e \right) \underline{n} \tag{113}$$

which is essentially Cauchy's stress theorem describing the relation between stress tensor and surface tractions. From the second terms on left- and right-hand side of Eq. 112,

$$\text{div}\left(\boldsymbol{\Gamma}_I^e + \boldsymbol{\Gamma}_M^e \right) + \underline{b} = 0 \tag{114}$$

which represents Cauchy's equation of motion for stationary systems. Therefore, stress measures in Eqs. 113 and 114 identically correspond to Cauchy stress ($\mathbf{T}$) components in intermolecular mechanism ($\mathbf{T}_I$) and molecular network mechanism ($\mathbf{T}_M$):

$$\mathbf{T} = \boldsymbol{\Gamma}_I^e + \boldsymbol{\Gamma}_M^e \tag{115}$$

$$\mathbf{\Gamma}_I^e = \mathbf{T}_I = \mathbf{T}_I{}^T \tag{116}$$

$$\mathbf{\Gamma}_M^e = \mathbf{T}_M = \mathbf{T}_M{}^T \tag{117}$$

Consider principal of virtual power for a second special case defined from Eq. 34 such that

$$\mathbf{L} = \text{grad}(\dot{\chi}) = \mathbf{L}_I^e + \mathbf{F}_I^e \mathbf{D}_I^p \mathbf{F}_I^{e^{-1}} = \mathbf{L}_M^e + \mathbf{F}_M^e \mathbf{D}_M^p \mathbf{F}_M^{e^{-1}} = 0 \tag{118}$$

or

$$\mathbf{L}_I^e = -\mathbf{F}_I^e \mathbf{D}_I^p \mathbf{F}_I^{e^{-1}} \tag{119}$$

$$\mathbf{L}_M^e = -\mathbf{F}_M^e \mathbf{D}_M^p \mathbf{F}_M^{e^{-1}} \tag{120}$$

Accordingly, principal of virtual power can be rewritten for this special case from Eqs. 106 and 107 as

$$\dot{W}_{\text{int}} = \int_V \begin{bmatrix} \mathbf{\Gamma}_I^e : \left(-\mathbf{F}_I^e \mathbf{D}_I^p \mathbf{F}_I^{e^{-1}}\right) + J^{-1}\mathbf{\Gamma}_I^p : \mathbf{D}_I^p \\ +\mathbf{\Gamma}_M^e : \left(-\mathbf{F}_M^e \mathbf{D}_M^p \mathbf{F}_M^{e^{-1}}\right) + J^{-1}\mathbf{\Gamma}_M^p : \mathbf{D}_M^p \end{bmatrix} dV \tag{121}$$

or

$$\dot{W}_{\text{int}} = \int_V \left[\left(J^{-1}\mathbf{\Gamma}_I^p - \mathbf{F}_I^{e^T}\mathbf{\Gamma}_I^e \mathbf{F}_I^{e^{-T}}\right) : \mathbf{D}_I^p + \left(J^{-1}\mathbf{\Gamma}_M^p - \mathbf{F}_M^{e^T}\mathbf{\Gamma}_M^e \mathbf{F}_M^{e^{-T}}\right) : \mathbf{D}_M^p \right] dV \tag{122}$$

Since velocity gradient is assumed as zero for this special case, velocity field will be also equal to zero. Therefore, external power work will be equal to zero $\left(\dot{W}_{\text{ext}} = 0\right)$. As a result, individual terms inside parenthesis in Eq. 122 should be equal to zero for arbitrary selection of V, $\mathbf{D}_I^p$, and $\mathbf{D}_M^p$:

$$J^{-1}\mathbf{\Gamma}_I^p - \mathbf{F}_I^{e^T}\mathbf{\Gamma}_I^e \mathbf{F}_I^{e^{-T}} = 0 \tag{123}$$

$$J^{-1}\mathbf{\Gamma}_M^p - \mathbf{F}_M^{e^T}\mathbf{\Gamma}_M^e \mathbf{F}_M^{e^{-T}} = 0 \tag{124}$$

Using definitions of stress measures in Eqs. 116 and 117, it can be shown that

$$\mathbf{\Gamma}_I^p = J\mathbf{F}_I^{e^T}\mathbf{T}_I\mathbf{F}_I^{e^{-T}} \tag{125}$$

$$\mathbf{\Gamma}_M^p = J\mathbf{F}_M^{e^T}\mathbf{T}_M\mathbf{F}_M^{e^{-T}} \tag{126}$$

which form definition of elastic symmetric Mandel stress in intermolecular structure and molecular network structure as

$$\mathbf{M}_I^e = J\mathbf{F}_I^{e^T}\mathbf{T}_I\mathbf{F}_I^{e^{-T}} \tag{127}$$

$$\mathbf{M}_M^e = J\mathbf{F}_M^{e^T}\mathbf{T}_M\mathbf{F}_M^{e^{-T}} \tag{128}$$

whereas since trace of plastic stretch rate is equal to zero due to incompressible plastic flow assumption (Eq. 33), stress conjugate to plastic stretch rate should be a deviatoric tensor. Therefore,

$$\mathbf{\Gamma}_I^p = \mathrm{dev}\left(\mathbf{M}_I^e\right) \tag{129}$$

$$\mathbf{\Gamma}_M^p = \mathrm{dev}\left(\mathbf{M}_M^e\right) \tag{130}$$

Finally, elastic 2nd Piola-Kirchhoff stress tensor in intermolecular structure and molecular network structure can be defined as

$$\mathbf{S}_I^e = J\mathbf{F}_I^{e^{-1}}\mathbf{T}_I\mathbf{F}_I^{e^{-T}} \tag{131}$$

$$\mathbf{S}_M^e = J\mathbf{F}_M^{e^{-1}}\mathbf{T}_M\mathbf{F}_M^{e^{-T}} \tag{132}$$

Using definitions of elastic Mandel stress (Eqs. 127 and 128) and symmetric 2nd Piola-Kirchhoff stress (Eqs. 131 and 132),

$$\mathbf{M}_I^e = \mathbf{C}_I^e\mathbf{S}_I^e \tag{133}$$

$$\mathbf{M}_M^e = \mathbf{C}_M^e\mathbf{S}_M^e \tag{134}$$

Therefore, it has been shown that stress measures ($\mathbf{\Gamma}_I^e$, $\mathbf{\Gamma}_M^e$, $\mathbf{\Gamma}_I^p$, $\mathbf{\Gamma}_M^p$) or (T_I, T_M, $\mathbf{M}_I^e$, $\mathbf{M}_M^e$) with deformation rate conjugates of ($\mathbf{D}_I^e$, $\mathbf{D}_M^e$, $\mathbf{D}_I^p$, $\mathbf{D}_M^p$) form a frame-indifferent framework for dual-mechanism elastic-viscoplastic constitutive model. The thermodynamic restrictions on constitutive relations can be obtained by substituting Eqs. 86, 89, and 93 into Eq. 84, as follows:

$$\mathrm{div}(\mathbf{J}_s) - \mathrm{div}\left(\frac{\mathbf{J}_q}{\theta}\right) = 0 \tag{135}$$

$$\mathbf{J}_s = \frac{\mathbf{J}_q}{\theta} \tag{136}$$

$$\gamma_{\mathrm{ther}} = -\frac{1}{\theta^2}\mathrm{div}\left(\mathbf{J}_q\right) \bullet \nabla_x(\theta) + \frac{\rho r}{\theta} > 0 \tag{137}$$

$$\left[J^{-1}\left(\frac{\partial \Psi_I\left(\mathbf{E}_I^e,\theta\right)}{\partial\theta} + \frac{\partial \Psi_M\left(\mathbf{C}_M^e,\theta\right)}{\partial\theta} + \frac{\partial \Psi_D(\mathbf{A},\theta)}{\partial\theta}\right) + \rho s\right]\dot{\theta} = 0 \tag{138}$$

$$\rho s\left(\mathbf{E}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right) = -J^{-1}\frac{\partial \Psi\left(\mathbf{E}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right)}{\partial\theta} \tag{139}$$

$$s\left(\mathbf{E}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right) = -\frac{\partial \psi\left(\mathbf{E}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right)}{\partial \theta} \tag{140}$$

$$\frac{1}{\theta}\left(\boldsymbol{\Gamma}_I^e : \mathbf{L}_I^e - J^{-1}\frac{\partial \Psi_I\left(\mathbf{E}_I^e, \theta\right)}{\partial \mathbf{C}_I^e} : \dot{\mathbf{C}}_I^e\right) = \tag{141}$$

$$\boldsymbol{\Gamma}_I^e : \mathbf{L}_I^e = \mathbf{T}_I : \mathbf{L}_I^e = \mathbf{T}_I : \mathbf{D}_I^e = J^{-1}\frac{\partial \Psi_I\left(\mathbf{E}_I^e, \theta\right)}{\partial \mathbf{C}_I^e} : \dot{\mathbf{C}}_I^e \tag{142}$$

$$\mathbf{T}_I : \mathbf{D}_I^e = \frac{1}{2}\left(\mathbf{F}_I^{e^{-1}}\mathbf{T}_I\mathbf{F}_I^{e^{-T}}\right) : \dot{\mathbf{C}}_I^e = J^{-1}\frac{\partial \Psi_I\left(\mathbf{E}_I^e, \theta\right)}{\partial \mathbf{C}_I^e} : \dot{\mathbf{C}}_I^e \tag{143}$$

$$J\left(\mathbf{F}_I^{e^{-1}}\mathbf{T}_I\mathbf{F}_I^{e^{-T}}\right) = \mathbf{S}_I^e = 2\frac{\partial \Psi_I\left(\mathbf{E}_I^e, \theta\right)}{\partial \mathbf{C}_I^e} \tag{144}$$

$$\frac{1}{\theta}\left(\boldsymbol{\Gamma}_M^e : \mathbf{L}_M^e - J^{-1}\frac{\partial \Psi_M\left(\mathbf{C}_M^e, \theta\right)}{\partial \mathbf{C}_M^e} : \dot{\mathbf{C}}_M^e\right) = \tag{145}$$

$$\boldsymbol{\Gamma}_M^e : \mathbf{L}_M^e = \mathbf{T}_M : \mathbf{L}_M^e = \mathbf{T}_M : \mathbf{D}_M^e = J^{-1}\frac{\partial \Psi_M\left(\mathbf{C}_M^e, \theta\right)}{\partial \mathbf{C}_M^e} : \dot{\mathbf{C}}_M^e \tag{146}$$

$$\mathbf{T}_M : \mathbf{D}_M^e = \frac{1}{2}\left(\mathbf{F}_M^{e^{-1}}\mathbf{T}_M\mathbf{F}_M^{e^{-T}}\right) : \dot{\mathbf{C}}_M^e = J^{-1}\frac{\partial \Psi_M\left(\mathbf{C}_M^e, \theta\right)}{\partial \mathbf{C}_M^e} : \dot{\mathbf{C}}_M^e \tag{147}$$

$$J\left(\mathbf{F}_M^{e^{-1}}\mathbf{T}_M\mathbf{F}_M^{e^{-T}}\right) = \mathbf{S}_M^e = 2\frac{\partial \Psi_M\left(\mathbf{C}_M^e, \theta\right)}{\partial \mathbf{C}_M^e} \tag{148}$$

$$\gamma_{\text{mech}} = \frac{1}{\theta}\left(J^{-1}\boldsymbol{\Gamma}_I^p : \mathbf{L}_I^p + J^{-1}\boldsymbol{\Gamma}_M^p : \mathbf{L}_M^p - J^{-1}\frac{\partial \Psi_D(\mathbf{A}, \theta)}{\partial \mathbf{A}} : \dot{\mathbf{A}}\right) > 0 \tag{149}$$

Equation 136 relates entropy flux to heat flux, and Eq. 137 defines irreversible entropy production per unit volume in deformed configuration associated with heat conduction which is always positive according to Fourier's law (Eq. 150). Equation 140 provides the relation between specific entropy and specific Helmholtz energy, while Eqs. 144 and 148 describe stress measures derived from Helmholtz free energy functions. Equation 149 defines irreversible entropy production due to mechanical dissipation per unit volume in deformed configuration. The reason for appearance of J^{-1} term in front of Eq. 149 is that all stress measures and their conjugate rate measures refer to intermediate (relaxed configuration), while irreversible entropy production rate density refers to deformed configuration:

$$\mathbf{J}_q = -k\nabla_x(\theta) \tag{150}$$

In Eq. 150, k is temperature-dependent thermal conductivity of material. Specific heat can be expressed in terms of specific entropy (Eq. 151) and in terms of Helmholtz

free energy (Eq. 153) by exploiting the relation in Eq. 139. Using linear decomposition assumption in Eqs. 85 and 86, specific heat can be rewritten in terms of separate component of specific entropy (Eq. 152) and Helmholtz free energy (Eq. 154):

$$c = \theta \frac{\partial s\left(\mathbf{E}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right)}{\partial \theta} \tag{151}$$

$$c = \theta \left(\frac{\partial s_I\left(\mathbf{E}_I^e, \theta\right)}{\partial \theta} + \frac{\partial s_M\left(\mathbf{C}_M^e, \theta\right)}{\partial \theta} + \frac{\partial s_D\left(\mathbf{A}, \theta\right)}{\partial \theta} \right) \tag{152}$$

$$c = -\theta \frac{\partial}{\partial \theta} \left[\frac{J^{-1}}{\rho} \frac{\partial \Psi\left(\mathbf{E}_I^e, \mathbf{C}_M^e, \mathbf{A}, \theta\right)}{\partial \theta} \right] \tag{153}$$

$$c = -\theta \frac{\partial}{\partial \theta} \left[\frac{J^{-1}}{\rho} \left(\frac{\partial \Psi_I\left(\mathbf{E}_I^e, \theta\right)}{\partial \theta} + \frac{\partial \Psi_M\left(\mathbf{C}_M^e, \theta\right)}{\partial \theta} + \frac{\partial \Psi_D\left(\mathbf{A}, \theta\right)}{\partial \theta} \right) \right] \tag{154}$$

Constitutive Relations and Flow Rules

In viscoplastic constitutive modeling of amorphous polymers at large deformations, dual decomposition of material response into two parallel working mechanisms of intermolecular structure and molecular network structure is widely used (Basaran and Lin 2007; Basaran and Yan 1998; Bauwens-Crowet et al. 1969; Boyce et al. 2000; Chudnovsky et al. 1973). More recently, a trial-mechanism has been also proposed to include a secondary mechanism for molecular network structure (Eringen 1967). Both dual- and trial-mechanism models are proven to be successful in describing large deformation behavior of amorphous polymers at different isothermal test conditions. In order to extend applicability of such models to non-isothermal conditions, several refinements on material property definitions and viscoplastic flow rule definitions are necessary.

In dual-mechanism constitutive models, material response is assumed to be controlled by states of two parallel working mechanisms (intermolecular structure and molecular network structure), as depicted in Fig. 2. Since intermolecular mechanism (I) and molecular network mechanism (M) are working in parallel, deformation in both mechanisms is equal to each other and equal to total deformation (Eq. 155), while total stress is the summation of stresses due to intermolecular interactions (I) and molecular network interactions (M) (Eq. 156). Subscripts "I" and "M" represent intermolecular and molecular network components of associated quantity, respectively:

$$\mathbf{F} = \mathbf{F}_I = \mathbf{F}_M \tag{155}$$

$$\mathbf{T} = \mathbf{T}_I + \mathbf{T}_M \tag{156}$$

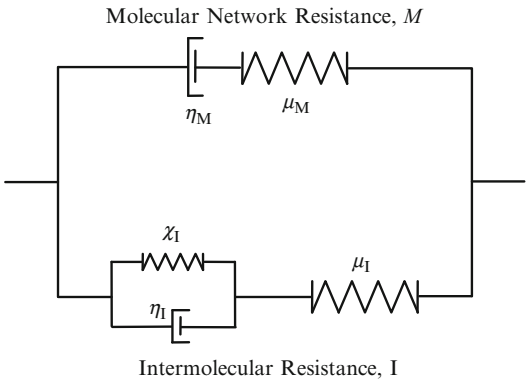


Fig. 2 Schematic representation of material model

Intermolecular Resistance (I)

Initial elastic response to deformation due to intermolecular resistance is governed by van der Waals interactions with surrounding molecules. A Helmholtz free energy per unit volume in reference configuration is considered for constitutive relation describing intermolecular resistance which was developed by Anand (Fotheringham and Cherry 1978; Fotheringham et al. 1976):

$$\Psi_I\left(\mathbf{E}_I^e,\theta\right)=\left\{\begin{array}{l} G\left|\mathrm{dev}\left(\mathbf{E}_I^e\right)\right|^2+\frac{1}{2}\left(K-\frac{2}{3}G\left[\mathrm{tr}\left(\mathbf{E}_I^e\right)\right]^2\right) \\ -3K\alpha(\theta-\theta_o)\mathrm{tr}\left(\mathbf{E}_I^e\right) \end{array}\right\} \tag{157}$$

In Eq. 157, θ_o is initial temperature, $\theta_o=\theta(\mathbf{X},t_o)$; G, K, α are temperature-dependent shear modulus, bulk modulus, and coefficient of thermal expansion, respectively. Elastic logarithmic strain ($\mathbf{E}_I^e$) in Eq. 157 is related to right elastic stretch tensor ($\mathbf{C}_I^e$) and elastic deformation gradient ($\mathbf{F}_I^e$) through Eqs. 158 and 159:

$$\mathbf{E}_I^e=\frac{1}{2}\ln\mathbf{C}_I^e \tag{158}$$

$$\mathbf{E}_I^e=\frac{1}{2}\ln\left(\mathbf{F}_I^{e^T}\mathbf{F}_I^e\right) \tag{159}$$

According to state laws (Eq. 144), symmetric second Piola-Kirchhoff stress ($\mathbf{S}_I^e$) and Cauchy stress ($\mathbf{T}_I$) can be obtained from Helmholtz free energy density function corresponding to intermolecular resistance as

$$\mathbf{S}_I^e=2\frac{\partial\Psi_I\left(\mathbf{E}_I^e,\theta\right)}{\partial\mathbf{C}_I^e} \tag{160}$$

$$\mathbf{T}_I=J^{-1}\mathbf{F}_I^e\mathbf{S}_I^e\mathbf{F}_I^{e^T} \tag{161}$$

Since Helmholtz free energy density corresponding to macroscopic elastic energy stored (Ψ_I) is an isotropic function of elastic right Cauchy tensor ($\mathbf{C}_I^e$), $\mathbf{C}_I^e$ and $\partial\Psi_I/\partial\mathbf{C}_I^e$ are coaxial, and their product is a symmetric tensor, elastic Mandel stress ($\mathbf{M}_I^e$) (Eq. 162) is defined as

$$\mathbf{M}_I^e = \mathbf{C}_I^e\mathbf{S}_I^e \tag{162}$$

Relation between elastic Mandel stress ($\mathbf{M}_I^e$) and elastic logarithmic strain ($\mathbf{E}_I^e$) can be obtained from Helmholtz free energy function (Eq. 157) and second Piola-Kirchhoff stress ($\mathbf{S}_I^e$) definition (Eq. 160) as

$$\mathbf{M}_I^e = 2G\mathrm{dev}\left(\mathbf{E}_I^e\right) + K\left[\mathrm{tr}\left(\mathbf{E}_I^e\right) - 3\alpha(\theta-\theta_o)\right]\mathbf{I} \tag{163}$$

Kinematic hardening characteristics in intermolecular structure are modeled by a defect energy function per unit volume in intermediate (relaxed) configuration which was developed by Anand (Boyce et al. 2000; Eringen 1967; Francisco et al. 1996):

$$\Psi_D(\mathbf{A},\theta) = \frac{1}{4}B\left[\ln(a_1)^2 + \ln(a_2)^2 + \ln(a_3)^2\right] \tag{164}$$

a_i in Eq. 164 represents eigenvalues of a stretch-like internal variable ($\mathbf{A}$) which is a symmetric unimodular tensor, $\det(\mathbf{A}(\mathbf{x},t)) = 1$. Since defect energy ($\Psi_D$) is an isotropic function of symmetric unimodular stretch-like tensor ($\mathbf{A}$), $\mathbf{A}$ and $\partial\Psi_D/\partial\mathbf{A}$ are coaxial and their product is symmetric deviatoric back stress tensor ($\mathbf{M}_{\mathrm{back}}$). Back stress (Eq. 165) and evolution equation for ($\mathbf{A}$) (Eq. 166) with initial condition in Eq. 167 are defined as

$$\mathbf{M}_{\mathrm{back}} = 2\mathrm{dev}\left(\frac{\partial\Psi(\mathbf{A},\theta)}{\partial\mathbf{A}}\mathbf{A}\right) = B\ln(\mathbf{A}) \tag{165}$$

$$\dot{\mathbf{A}} = \mathbf{D}_I^p\mathbf{A} + \mathbf{A}\mathbf{D}_I^p - \gamma\mathbf{A}\ln(\mathbf{A})\nu_I^p \tag{166}$$

$$\mathbf{A}(\mathbf{X},0) = \mathbf{I} \tag{167}$$

In Eq. 166, γ represent the dynamic recovery, B is temperature-dependent back stress modulus, and ν_I^p is equivalent plastic stretch rate in intermolecular structure. Driving stress for plastic flow in intermolecular structure is defined as

$$\mathbf{M}_{\mathrm{eff}} = \mathrm{dev}\left(\mathbf{M}_I^e - \mathbf{M}_{\mathrm{back}}\right) \tag{168}$$

Equivalent plastic stretch rate (Eq. 169), effective equivalent shear stress (Eq. 170), and mean normal pressure (Eq. 171) are defined in terms of tensorial variables as follows:

$$\nu_I^p = \sqrt{2}\left|\mathbf{D}_I^p\right| \tag{169}$$

$$\tau_I = \frac{1}{\sqrt{2}} \left| \mathbf{M}_{eff} \right| \tag{170}$$

$$p_I = -\frac{1}{3} \mathrm{tr}\left(\mathbf{M}_I^e \right) \tag{171}$$

Evolution of plastic deformation gradient in intermolecular mechanism can be rewritten from Eq. 31 as

$$\dot{\mathbf{F}}_I^p = \mathbf{D}_I^p \mathbf{F}_I^p \tag{172}$$

$$\mathbf{F}_I^p(\mathbf{X}, 0) = \mathbf{I} \tag{173}$$

Effective equivalent shear stress is the part of stress that drives plastic flow, and it is the source of plastic dissipation, while some significant amount of plastic work is stored as energy during large deformation of amorphous polymers associated with back stress (Bauwens-Crowet et al. 1969; Gent 1996). Once effective shear stress level reaches a critical level so that energy barrier to molecular chain segment rotation is exceeded, plastic flow takes place. According to cooperative model, viscous flow in a solid amorphous polymer may take place only when a number of polymer segments move cooperatively which also account for the significance of activation volume during yield process. The flow rule for amorphous polymers is essentially based on the energy distribution statistics of individual segments (Gomez and Basaran 2006). In simple terms, cooperative model flow rule is based on average probability of simultaneous occurrence of n thermally activated transitions across an energy barrier (activation energy, Q) inducing a macroscopic strain increment of ν_o (Gomez and Basaran 2006; Kachanov 1986). Yield characteristics of amorphous polymers are strongly temperature and rate dependent. According to strain rate-temperature superposition principle, an increase in temperature will have the same effect on the yield stress as a decrease in strain rate (Kontou and Spathis 2006; Kröner 1959; Lee 1969). Equivalence of time and temperature basically describes that yielding of amorphous polymers at low temperatures is comparable to that at high strain rates. Therefore, Eyring plots (yield stress-temperature ratio versus plastic strain rate curves) for various temperatures can be shifted vertically and horizontally with respect to a reference temperature (θ_{ref}) in order to obtain a master curve describing yield stress behavior over a wide range of temperatures and strain rates.

Recently, Richeton (Chudnovsky et al. 1973) has proposed that both horizontal shift (ΔH_h) and vertical shift (ΔH_v) should follow Arrhenius-type temperature dependence. Resulting yield stress definition relates yield behavior of polymer with β mechanical loss peak at temperatures below θ_g through introducing activation energy at β-transition temperature, i.e., yield behavior is controlled by segmental motions of polymer chains and reference state for yielding is chosen as β-transition. Increase in yield stress due to an increase in strain rate is attributed to decrease in molecular mobility of polymer chains, while a slow deformation rate allows polymer chains to slip past each other, resulting in a lower resistance to flow.

At low temperatures near secondary transition temperature (θ_β), secondary molecular motions are restricted and chains become stiffer which also increase yield stress, while increase in temperature provides more energy to polymer chains facilitating relative motion between polymer chains. For temperatures above θ_g, characteristic plastic strain rate equation was modified by Williams-Landel-Ferry (WLF) parameters (c_1, c_2) (Bauwens-Crowet et al. 1969; Mandel 1972). Although characteristic plastic strain rate definitions at temperatures below and above θ_g are continuous functions of temperature in separate domains, piecewise definition with respect to glass transition (θ_g) results in unrealistic change in plastic flow behavior around glass transition, i.e., derivative of plastic strain rate equation is discontinuous at θ_g. Recently, Anand (Eringen 1967) proposed a modified version of flow rule in intermolecular structure which incorporates different values of activation energy for glassy region and rubbery region, yet abrupt change in activation energies at θ_g still creates problem in material response. In order to provide a smoother transition in flow characteristics around θ_g, characteristic plastic strain rate (Eq. 174) and equivalent shear plastic stretch rate (Eq. 175) are proposed in following forms:

$$\nu^* = \nu_I^o \exp\left(-\frac{Q_I}{k_B\theta}\right)\left[1+\exp\left(\frac{\ln(10)c_1\left(\theta-\theta_g\right)}{c_2+\left(\theta-\theta_g\right)}\right)\right] \tag{174}$$

$$\nu_I^p = \nu^*\left[\sinh\left(\frac{\overline{\tau}_I V}{2k_B\theta}\right)\right]^{n_I} \tag{175}$$

In Eqs. 174 and 175, ν_I^o is pre-exponential factor, Q_I is activation energy for plastic flow in intermolecular structure, k_B is the Boltzmann's constant, c_1 and c_2 are WLF parameters, n_I number of thermally activated transitions necessary for plastic flow, V activation volume, and $\overline{\tau}_I$ net effective stress which is defined in Eq. 176:

$$\overline{\tau}_I = \tau_I - S_I - \alpha_P p_I \tag{176}$$

In Eq. 176, α_p is pressure sensitivity parameter and S_I is plastic flow resistance in intermolecular structure. Evolution of intermolecular resistance to plastic flow is defined in Eq. 177 with initial condition in Eq. 178:

$$\dot{S}_I = h_I\left(S_I^* - S_I\right)\nu_I^p \tag{177}$$

$$S_I^o = S_I(\mathbf{X}, 0) \tag{178}$$

In Eq. 177, h_I is a parameter characterizing hardening-softening, and S_I^* is saturation value for plastic flow resistance in intermolecular structure which is defined in Eq. 179:

$$S_I^* = b(\varphi^* - \varphi) \tag{179}$$

b in Eq. 179 is a temperature- and rate-dependent parameter which relates saturation value of plastic flow resistance to an order function ($\varphi^* - \varphi$). Resistance

to plastic flow (S_I) increases with disorder in the material and becomes constant (S_I^*) when order parameter (φ) reaches a critical value (φ^*) which is also a temperature- and rate-dependent variable. When intermolecular resistance reaches saturation value, steady-state plastic flow occurs and plastic flow rate becomes equal to applied strain rate. Evolution equation for order parameter is defined as

$$\dot{\varphi} = g(\varphi^* - \varphi)\nu_I^p \tag{180}$$

$$\varphi_o = \varphi(\mathbf{X}, 0) \tag{181}$$

g in Eq. 180 is a temperature-dependent parameter. Evolution equation for plastic deformation gradient in Eqs. 172 and 173 completes definition of material behavior in intermolecular structure. Strain hardening becomes insignificant as temperatures approach θ_g and completely vanishes above θ_g (Chudnovsky et al. 1973). Definition of a vanishing internal resistance right at θ_g causes also discontinuity in yield behavior of polymer. Since annealing at high temperatures well-above θ_g clears past thermomechanical history of materials by providing an alternative stationary molecular configuration at a higher energy level, internal resistance is bound to vanish above or around θ_g. Therefore, underlying problem is essentially on the assumption that glass transition takes place at a single temperature and internal resistance becomes zero abruptly at θ_g (Chudnovsky et al. 1973). Similarly, variables (b, g, S_I^*, φ^*, h_I) that characterize hardening-softening behavior in post-yield region shall also provide a smooth transition from temperatures below θ_g to temperatures above θ_g.

It should be noted that viscoplastic models currently available in literature are all phenomenological and serve as a mathematical tool to fit experimentally observed behavior into a curve. Measure of applicability of such formulations is the accuracy of representing physical reality. These models can provide reasonably accurate predictions for yield characteristics of amorphous polymers for only isothermal cases. In the case of non-isothermal tests which include temperature change in the material concurrently with loading, most material models in literature by Anand (Boyce et al. 2000; Eringen 1967; Francisco et al. 1996) or Richeton (Bauwens-Crowet et al. 1969; Mandel 1972) would predict unrealistic results. A comparison of viscoplastic models for amorphous polymers from literature and improved version of dual-mechanism model in this research is presented in Fig. 3.

Temperature variations of characteristic viscoplastic shear strain rates in different models are presented by normalizing with respect to characteristic viscoplastic strain rate at reference glass transition temperature for PMMA (387 K). Material properties were assumed to be the same and taken from Eringen (1967), while WLF parameters in this research and Richeton's work (Bauwens-Crowet et al. 1969) were taken as their original values. Viscoplastic models presented in Eringen (1967) and Bauwens-Crowet et al. (1969) and this research are applicable for temperatures both above and below glass transition, while viscoplastic model in Boyce et al. (2000) is only applicable for temperatures below glass transition and

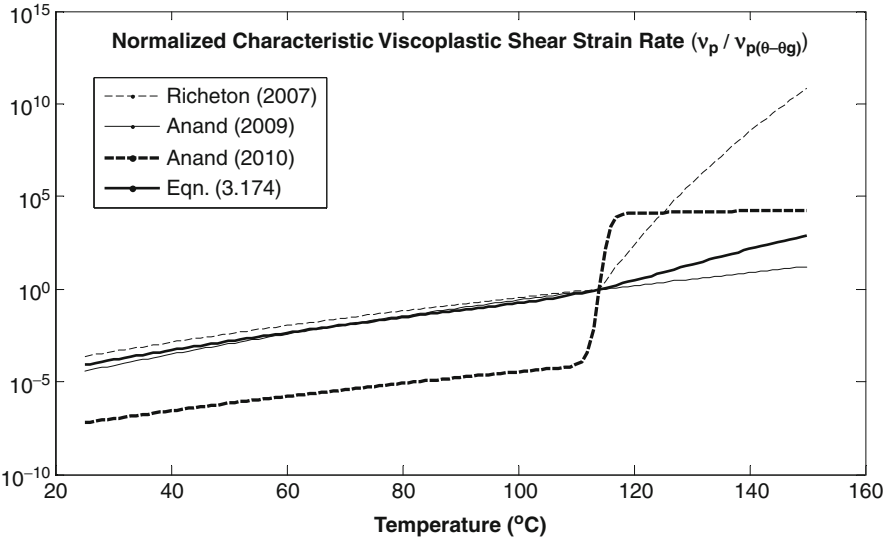


Fig. 3 Comparison of different viscoplastic models in literature in terms of temperature dependence of normalized characteristic viscoplastic shear strain rate $\left(\nu_p/\nu_{p(\theta=\theta_g)}\right)$

presented only for comparison. In Anand's model (Eringen 1967), temperature dependence is conveyed through temperature dependence of activation energy term included in equivalent viscoplastic stretch rate equation. In Richeton's model (Bauwens-Crowet et al. 1969), a piecewise definition of viscoplastic stretch rate is employed to describe temperature and rate dependence of yield strength. On the other hand, in this research, temperature dependence of viscoplastic stretch rate is directly employed utilizing physically motivated Williams-Landel-Ferry parameters in a completely new form expression as presented in Eq. 174. It is clear that temperature-dependent activation energy approach in Anand's model (Eringen 1967) predicts not a gradual change in behavior in intermolecular mechanism but an abrupt increase in viscoplastic strain rate over a relatively short temperature range. Therefore, according to this model, there is actually no rubbery region, but material response is liquid-like as viscoplastic strain rate increases by six orders in a narrow temperature interval (2 °C) around glass transition temperature. On the other hand, in Richeton's model (Bauwens-Crowet et al. 1969), there is also a remarkable change in viscoplastic strain rate due to the piecewise definition with respect to glass transition temperature, but it also causes discontinuous derivative of viscoplastic rate at glass transition. Accordingly material response predicted by Anand's or Richeton's viscoplastic model will present a significant (abrupt) change in stress. Improved version of dual-mechanism model in this research predicts definitely a more gradual transition in material response with respect to temperature around glass transition. Though Anand's model has a continuous definition of activation energy in temperature domain, remarkable difference between activation

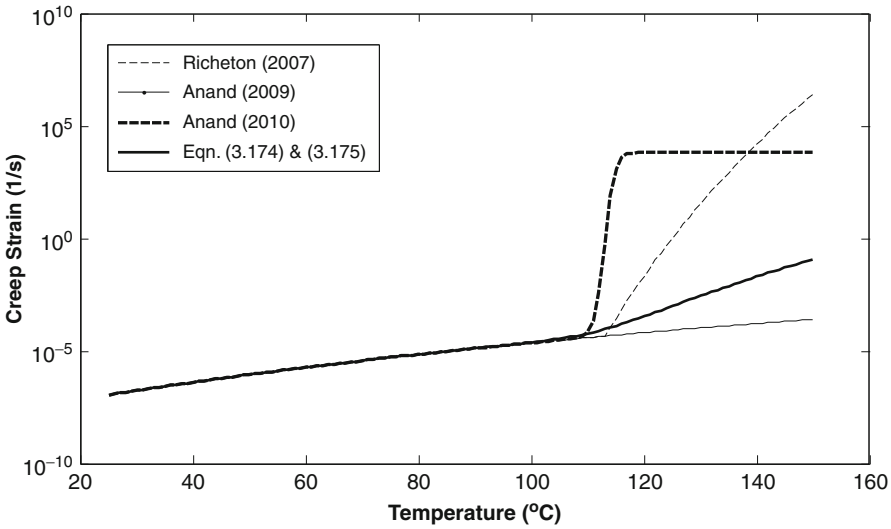


Fig. 4 Comparison of different viscoplastic models in literature in terms of creep strain rates at different temperatures in response to an applied stress of 0.6 MPa

energies in glassy and rubbery region still produces an abrupt change in response. According to Fig. 3, Anand's viscoplastic model is invariant of temperature above glass transition temperature, while Richeton's model provides relatively gentler transition in response. As a result of this rapid change in viscoplastic response in Anand's model, it cannot predict material behavior accurately under non-isothermal conditions. These models were further studied in terms of predictions for creep strain rate in a creep test with a stress level of 0.6 MPa which is conducted at different temperatures as depicted in Fig. 4.

Figure 4 definitely supports previous observations on different models. All models yield same predictions for creep strain rate at temperatures below glass transition. However, according to Anand's model (Eringen 1967), influence of glassy-to-rubbery transition on creep strain rate will take place over only "3C" temperature range with a five-order increase in creep strain rate. This means that if a sample is tested at 112C, creep strain rate will be predicted as $3.7 \times 10^{-3}\ s^{-1}$, while for another sample tested at 115 °C, creep strain rate from Anand's model (Eringen 1967) will be predicted as $1.3 \times 10^{3}\ s^{-1}$. A 5-order increase in creep strain rate for a 3C difference in temperature is completely unrealistic. Moreover, according to Anand's model, creep strain rate is invariant of temperature right above glass transition temperature which is completely inaccurate and contradicts experimental studies in the literature (Chudnovsky et al. 1973; Mandel 1972). On the other hand, Richeton model (Bauwens-Crowet et al. 1969) predicts a more gradual change in creep strain rate over temperature, yet piecewise definition at glass transition temperature is still problematic for transient thermal conditions. Variation of creep strain rate with temperature predicted by Eq. 174 in this research provides a

smoother transition around glass transition and a more gradual change in creep strain rate at temperatures above glass transition temperature.

For assurance of accurate and realistic modeling of material response around glass transition, every aspect of material property definitions that describe hardening-softening behavior and flow characteristics should be continuous in temperature domain and should have continuous (at least) first derivatives with respect to temperature. It is clear that Anand's model and other available viscoplastic models do not fulfill this crucial and fundamental necessity, while temperature dependence of all material characteristics presented in this research conforms to the abovementioned criteria.

Molecular Network Resistance (M)

Resistance in molecular network to deformation is based on molecular orientation and relaxation process. If there is enough stretch in polymer chains, network resists relaxation and resistance increases with increasing stretch. In literature, there is a general consensus on modeling of plastic flow behavior in intermolecular structure based on Eyring cooperative model, but there is still some debate on modeling of molecular network structure. Arruda and Boyce (Palm et al. 2006) modeled molecular network resistance with rubber elasticity model based on eight-chain network of non-Gaussian chains similar to transient response of elastomers which represents nonlinear rate-dependent deviation from equilibrium state (Basaran and Nie 2007). However, network resistance described by 8-chain model is not accurate enough to be a physically consistent model describing orientational hardening behavior in amorphous polymers, since temperature dependence of rubbery modulus and number of rigid links between polymer chain segments which essentially control response does not match with experimental observations. Therefore, molecular network description based on 8-chain model (Basaran and Lin 2007; Bauwens-Crowet et al. 1969; Gent 1996; Palm et al. 2006) becomes merely a numerical tool to match experimentally observed stress-strain response. Instead, a simpler two-constant constitutive relation for rubber networks developed by Gent (Povolo and Élida 1995) was shown to describe strain hardening due to polymer chain stretching better than statistical-mechanical entropic rubber elasticity models (8-chain model) while resulting a similar stress-strain response (Boyce et al. 2000; Eringen 1967). Gent free energy per unit volume in reference configuration (Povolo and Élida 1995) describes elastic energy stored in molecular network structure in terms of first invariant of stretch in polymer chains as

$$\Psi_M\left(\mathbf{C}_M^e, \theta\right) = -\frac{1}{2}\mu_M I_M \ln\left(1 - \frac{I_1 - 3}{I_m}\right) \tag{182}$$

μ_M and I_M in Eq. 182 are temperature-dependent rubbery shear modulus and limit on extensibility of polymer chains, respectively. Since volume change in material is considered in elastic deformation gradient associated with intermolecular structure, it is essential to define Gent free energy in terms of distortional elastic deformation gradient in network structure $(\mathbf{F}_M^e)_d$ (Eq. 183) which produces no change in volume

(Eq. 184). I_1 is first invariant (Eq. 186) of elastic distortional Cauchy tensor in network structure, $(\mathbf{C}_M^e)_d$ (Eq. 185):

$$\left(\mathbf{F}_M^e\right)_d = J^{-1/3}\mathbf{F}_M^e \tag{183}$$

$$\det\left(\left(\mathbf{F}_M^e\right)_d\right) = 1 \tag{184}$$

$$\left(\mathbf{C}_M^e\right)_d = \left(\mathbf{F}_M^e\right)_d^T\left(\mathbf{F}_M^e\right)_d \tag{185}$$

$$I_1 = \mathrm{tr}\left[\left(\mathbf{C}_M^e\right)_d\right] \tag{186}$$

Second Piola-Kirchhoff stress ($\mathbf{S}_M^e$) and Cauchy stress ($\mathbf{T}_M$) can be derived from Gent free energy as follows:

$$\mathbf{S}_M^e = 2\frac{\partial\Psi_M\left(\mathbf{C}_M^e,\theta\right)}{\partial\mathbf{C}_M^e} \tag{187}$$

$$\mathbf{T}_M = J^{-1}\mathbf{F}_M^e\mathbf{S}_M^e\mathbf{F}_M^{e^T} \tag{188}$$

Using Gent free energy definition in Eq. 182 and stress definitions in Eqs. 187 and 188,

$$\mathbf{S}_M^e = J^{-2/3}\mu_M\left(1-\frac{I_1-3}{I_M}\right)^{-1}\left[\mathbf{I}-\frac{1}{3}\mathrm{tr}\left(\left(\mathbf{C}_M^e\right)_d\right)\left(\mathbf{C}_M^e\right)_d{}^{-1}\right] \tag{189}$$

$$\mathbf{T}_M = J^{-1}\mu_M\left(1-\frac{I_1-3}{I_M}\right)^{-1}\mathrm{dev}\left(\left(\mathbf{B}_M^e\right)_d\right) \tag{190}$$

Elastic distortional Almansi tensor $(\mathbf{B}_M^e)_d$ in Eq. 191 is defined in terms of distortional elastic deformation gradient ($(\mathbf{F}_M^e)_d$) as

$$\left(\mathbf{B}_M^e\right)_d = \left(\mathbf{F}_M^e\right)_d\left(\mathbf{F}_M^e\right)_d^T \tag{191}$$

Elastic Mandel stress (Eqs. 192 and 193) and equivalent shear stress (Eq. 194) and equivalent plastic shear strain rate (Eq. 195) for molecular relaxation in network structure can be found as

$$\mathbf{M}_M^e = \mathbf{C}_M^e\mathbf{S}_M^e \tag{192}$$

$$\mathbf{M}_M^e = \mu_M\left(1-\frac{I_1-3}{I_M}\right)^{-1}\mathrm{dev}\left(\left(\mathbf{C}_M^e\right)_d\right) \tag{193}$$

$$\tau_M = \frac{1}{\sqrt{2}}\left|\mathbf{M}_M^e\right| \tag{194}$$

$$\nu_M^p = \sqrt{2}\left|\mathbf{D}_M^p\right| \tag{195}$$

Evolution of plastic deformation gradient in molecular network mechanism can be rewritten from Eq. 31 as

$$\dot{\mathbf{F}}_M^p = \mathbf{D}_M^p \mathbf{F}_M^p \tag{196}$$

$$\mathbf{F}_M^p(\mathbf{X}, 0) = \mathbf{I} \tag{197}$$

Molecular network is responsible for resistance to chain alignment which resists relaxation as stretch in network increases (Basaran and Lin 2007; Gent 1996). A similar observation for elastomers also found that plastic chain stretch is inversely proportional to effective creep rate (Basaran and Nie 2007). In experimental and numerical studies, it was observed that in post-yield region (at large deformations), controlling mechanism is molecular network mechanism. Molecular network mechanism had a dominant contribution to stress change in post-yield region, while amount elastic recovery upon unloading is associated with plastic strain in molecular network mechanism. Temperature dependence of molecular relaxation in network structure is characterized with a classical Arrhenius term, and flow rule is completed with a simple power law as follows:

$$\nu_M^p = \nu_M^o \exp\left(-\frac{Q_M}{k_B\theta}\right)\left(\frac{\tau_M}{S_M}\right)^{n_M} \tag{198}$$

In Eq. 198, ν_M^o is pre-exponential factor, Q_M is activation energy for molecular relaxation in network structure, n_M is a strain-rate sensitivity parameter, and S_M is a stress measure describing resistance of network structure to relaxation which increases with increasing plastic stretch rate as defined in Eq. 199 with initial condition in Eq. 200:

$$\dot{S}_M = h_M(\lambda_M^p - 1)\left(S_M^* - S_M\right)\nu_M^p \tag{199}$$

$$S_M^o = S_M(\mathbf{X}, 0) \tag{200}$$

In Eq. 199, h_M is a parameter characterizing molecular relaxation in material, S_M^* is temperature-rate-dependent saturation value for network resistance, and λ_M^p is plastic stretch which is related to plastic Almansi tensor in network structure ($\mathbf{B}_M^p$) as follows:

$$\lambda_M = \sqrt{\frac{\mathrm{tr}(\mathbf{B}_M^p)}{3}} \tag{201}$$

$$\mathbf{B}_M^p = \mathbf{F}_M^p \mathbf{F}_M^{p^T} \tag{202}$$

Evolution equation for plastic deformation gradient in Eqs. 196 and 197 completes definition of material behavior in molecular network structure. According to Eq. 199, network resistance will increase continuously as plastic stretch (λ_M^p) in polymer chains increases and reaches a constant value (S_M^*) which depends on temperature and stretch rate. Plastic stretch-dependent evolution of resistance to plastic flow also ensures correct prediction of elastic recovery in unloading path (Eringen 1967). In Anand's model (Boyce et al. 2000; Eringen 1967), net driving stress for plastic flow in intermolecular mechanism includes an additional resistance term accounting for dissipative resistance to plastic flow (S_b or S_2) taking place at large deformations. According to Anand's model, this dissipative resistance evolves with plastic stretch in intermolecular mechanism (Boyce et al. 2000) or total stretch in intermolecular mechanism (Eringen 1967). In this research, dissipative resistance (S_M) at large deformations in molecular network mechanism evolves with plastic stretch in molecular network branch which actually controls material response at large deformations (post-yield region). Anand et al. (Eringen 1967) attribute necessity of introducing a third mechanism for true prediction of elastic recovery during unloading and cooling of a preheated sample. It was argued that a second mechanism for molecular network structure was introduced due to a necessity driven by an experimentally observed complex response, yet source of this observation was neither referred from any other work nor included in their work. However, experimental framework in this research involves non-isothermal condition mentioned in Anand's work. It is clear that Anand's models (Boyce et al. 2000; Eringen 1967) do not properly employ a locking mechanism that will truly predict elastic recovery and necessitate an additional mechanism. In Anand's recent model (Eringen 1967), resistance to plastic flow associated with first molecular network mechanism (activated at low temperatures) assumed to be constant, and resistance to plastic flow associated with second molecular network mechanism (activated at high temperatures) evolves with increasing plastic stretch in intermolecular structure. It is their argument that this new additional mechanism depending on plastic stretch controls elastic recovery. In improved version of dual-mechanism viscoplastic model, molecular network mechanism dominates large deformation behavior (post-yield region) and stress level, while resistance to plastic flow in this regime is controlled by plastic stretch in molecular network. As plastic stretch in molecular network at large deformations is retained upon unloading, there was no problem observed in elastic recovery as mentioned. Therefore, elastic recovery can be truly predicted without artificially introducing an additional third mechanism.

Finally, from relation in Eq. 80 and specific Helmholtz free energy definitions in Eqs. 157, 164, and 182, governing equation for temperature can derived as

$$\begin{aligned}\rho c\dot{\theta} = \nabla_x(k\nabla_x(\theta)) + r + J^{-1}\left(\left(\tau_I + \frac{1}{2}\gamma B|\ln(\mathbf{A})|^2\right)\nu_I^p + \tau_M\nu_M^p\right)\\ + J^{-1}\theta\left(\frac{1}{2}\frac{\partial \mathbf{S}_I^e}{\partial\theta} : \dot{\mathbf{C}}_I^e + \frac{1}{2}\frac{\partial \mathbf{S}_M^e}{\partial\theta} : \dot{\mathbf{C}}_M^e + \frac{1}{2}\frac{\partial\left(\mathbf{M}_{back}\mathbf{A}^{-1}\right)}{\partial\theta} : \dot{\mathbf{A}}\right)\end{aligned} \tag{203}$$

The first two terms in Eq. 203 are heat conduction representing heat transfer within material during transient state and heat source due to passive heating (external heating) or active heating (internally generated heat). The last two terms in Eq. 203 represent heat induced due to intrinsic dissipation and thermoelastic effect representing conversion between mechanical and thermal energy in elastic range. In the case of coupled thermomechanical loading, temperature increase due to mechanical work and temperature change due to heat transfer between material and surroundings are mixed together. Based on descriptions of stress components and strain rate measures, irreversible entropy production due to mechanical dissipation per unit volume in deformed configuration can be rewritten from Eq. 149 as

$$\gamma_{\text{mech}} = \frac{J^{-1}}{\theta}\left[\begin{array}{l}\left(\text{dev}\left(\mathbf{M}_I^e\right) - 2\text{dev}\left(\dfrac{\partial \Psi_D(\mathbf{A},\theta)}{\partial \mathbf{A}}\mathbf{A}\right)\right) : \mathbf{D}_I^p \\ +\gamma\left(\dfrac{\partial \Psi_D(\mathbf{A},\theta)}{\partial \mathbf{A}}\mathbf{A}\right) : \ln(\mathbf{A})\nu_I^p + \text{dev}\left(\mathbf{M}_M^e\right) : \mathbf{D}_M^p\end{array}\right] > 0 \tag{204}$$

$$\gamma_{\text{mech}} = \frac{J^{-1}}{\theta}\left[\left(\tau_I + \frac{1}{2}\gamma B|\ln(\mathbf{A})|^2\right)\nu_I^p + \tau_M \nu_M^p\right] > 0 \tag{205}$$

Due to associated plastic flow assumption described in Eqs. 206 and 207, irreversible mechanical entropy production (γ_{mech}) is always positive:

$$\mathbf{N}_I^p = \frac{\mathbf{D}_I^p}{\left|\mathbf{D}_I^p\right|} = \frac{\mathbf{M}_{eff}}{\left|\mathbf{M}_{eff}\right|} \tag{206}$$

$$\mathbf{N}_M^p = \frac{\mathbf{D}_M^p}{\left|\mathbf{D}_M^p\right|} = \frac{\mathbf{M}_M^e}{\left|\mathbf{M}_M^e\right|} \tag{207}$$

Damage Evolution

Damage characterization of non-isothermal stretching of PMMA necessitates alternative methods due to complex form of testing. Traditional experimental damage quantification methods such as elastic modulus degradation cannot be directly used for prediction of damage evolution under transient thermal conditions. Damage evolution during non-isothermal stretching of amorphous polymers is modeled with isotropic damage evolution function which was originally proposed by Ye et al. (2003) (Richeton et al. 2005a, b, 2006, 2007; Srivastava et al. 2010).

From thermodynamics point of view, degradation of a material is the result of irreversible thermodynamic processes that create disorder in matter. Since entropy is a thermodynamic indicator of disorder, it is a natural measure of material degradation. In local form of entropy change at a material point, entropy production is associated with irreversibility of a process (Eq. 84) and entropy flow describes reversible component of the process (Eq. 136). Entropy production can be further

decomposed into thermal dissipation (Eq. 137) due to heat exchange between system and surroundings and mechanical dissipation (Eq. 205) as a result of permanent changes in material structure. Critical entropy production (η_{cr}) is a characteristic value for a material which is independent of loading conditions and geometry of specimen and only depends on mode of failure. Link between irreversible material degradation (damage) and amount of heat generated due to some nonconservative forces (plastic dissipation and friction forces) or entropy production due to mechanical dissipation was observed in several studies. Heat conduction within a metal will be very fast leading to negligible thermal gradients and thermal dissipation due to high thermal diffusivity of metals. In the case of materials with low thermal diffusivity such as polymers, heat transfer may take place over a prolonged period of time with significant thermal gradients. However, thermal gradients cannot be responsible for failure of chemical bonds since material degradation or damage is a consequence of formation of small voids or cracks at microscale by breakage of chemical bonds between molecules. Thermal dissipation may result in deterioration of material properties (e.g., thermal fatigue) which is insignificant compared to degradation by mechanical dissipation. Damage induced by temperature reversals will take much longer time to produce an equivalent damage effect created by mechanical work in a finite period of time.

According to Boltzmann (Sweeney et al. 1997), entropy of a micro-system (S) can be related to probability of existence of the system at a microstate with respect to all other possible microstates (Ω) as

$$S = k_B \ln(\Omega) \tag{208}$$

where k_B is Boltzmann's constant. Let that probability of a material being in a completely ordered ground state be equal to Ω_o. In an alternative configuration, material deviates from this perfectly ordered reference state under action of some external effects (mechanical, thermal, chemical, environmental, or a combination of these effects) to another disordered state with a probability of Ω. These changes in material with respect to reference (ground) state can be recoverable, and material may return to its original state with removal of external effects for a (hypothetical) reversible process. In the case of an irreversible process, external effects will create permanent changes in structure of material described as a positive entropy production or an overall net increase in entropy of material. In terms of statistical interpretation of entropy, such irreversible changes are described as an increase in probability of the microstate in new configuration which also indicates tendency of a system to attain a less ordered state (Sweeney et al. 1997). Since a disordered state is formed from an ordered state through introduction of damage (change) in system, damage (the measure of changes) and entropy (the measure of disorder) are naturally related. In final stage, material reaches a critical state such that disorder in the system is maximized (Ω_{max}) and material integrity no longer preserved ($D = 1$). At this stage, entropy will also reach a critical level (η_{cr}) which is a characteristic of material (Truesdell 2004). Since material in ground (reference) state is free of any possible defects, imperfections, i.e., damage, it can be assumed that damage

in material is equal to zero ($D = 0$). Kachanov similarly relates continuity of a material as an alternative definition to damage (Wallin and Ristinmaa 2005). In this case, continuity in material (ζ) at ground state is considered to be equal to "1," while continuity function (ζ) in a completely damaged material approaches to "0" since material integrity (continuity) is vanished when failure takes place in the material. Therefore, relation between continuity function and damage function can be established as

$$D = 1 - \xi \tag{209}$$

In order to relate entropy and damage, consider a system in ground state ($D = 0$) with a total entropy of S_o and an associated probability of Ω_o. In an alternative disordered (damaged) state, S is total entropy of the same system (matter) with an associated probability of Ω and a damage level of D. At failure, entropy reaches a critical level (S_{cr}) and damage approaches to 1, while probability of this particular state will be maximum ($\Omega_{\max}$) with respect to all other possible states. It is possible to develop a relation between entropy and damage in terms of probabilities of damaged state and ground state as defined in Eq. 210:

$$D = f(\Omega, \Omega_o) \tag{210}$$

Furthermore, it is assumed that damage can be related to the difference of damaged state probability from ground state probability (Eq. 211):

$$D = f(\Omega - \Omega_o) \tag{211}$$

Equation 211 interprets damage associated with different states in terms of deviation from reference state which is also the principal approach in all damage mechanic models and experimental damage measures. In Eq. 211, however, probabilities of states are taken as basis for comparison. Damage evolution does not only depend on deviation from original state, but it also depends on damage in current state. Therefore, Eq. 211 is revised as

$$D = f\left(\frac{\Omega - \Omega_o}{\Omega}\right) \tag{212}$$

Finally, damage at any possibility is defined in terms of probabilities of damaged state and ground state as

$$D = D_{\text{cr}}\left(\frac{\Omega - \Omega_o}{\Omega}\right) \tag{213}$$

where D_{cr} is critical damage parameter which essentially controls damage evolution in system. Damage evolution depends on several factors such as frequency of

external disturbances, temperature of system, and other possible factors that directly influence state of the system. Therefore, damage evolution under different irreversible process paths will not be identical. Using Eq. 208, Eq. 213 can be rewritten as

$$D = D_{\text{cr}}\left(\frac{\exp\left({}^{S}/_{k_B}\right) - \exp\left({}^{S_o}/_{k_B}\right)}{\exp\left({}^{S}/_{k_B}\right)}\right) \tag{214}$$

which can be simplified as

$$D = D_{\text{cr}}\left[1 - \exp\left(\frac{S_o - S}{k_B}\right)\right] \tag{215}$$

It is clear from Eq. 215 that when material is ground state ($S = S_o$), damage is equal to zero ($D = 0$). For the case when entropy in system reaches a critical entropy level (S_{cr}) at failure ($D = 1$), critical damage parameter (D_{cr}) can be formulated as shown in Eq. 216. Since critical damage parameter (D_{cr}) depends on critical entropy level (S_{cr}), it is also a characteristic of material and should be estimated for different types of materials separately:

$$D_{\text{cr}} = \left[1 - \exp\left(\frac{S_o - S_{\text{cr}}}{k_B}\right)\right]^{-1} \tag{216}$$

Damage evolution function in Eq. 217 is a refinement of Eq. 215 for a single material point for which specific entropy production definition is used instead of total entropy definition in whole volume of system. In derivation of damage evolution function in Eq. 215, source of entropy was assumed to be both entropy flow into system and entropy generation in the system. Equation 217 describes the damage evolution in a solid material in terms of only irreversible processes which create permanent changes in material:

$$D = D_{\text{cr}}\left[1 - \exp\left(-\frac{m_s}{R}\eta\right)\right] \tag{217}$$

where η is internal entropy production due to mechanical and thermal dissipation (Eqs. 137 and/or 205), m_s is specific mass, R is gas constant, and D_{cr} is temperature-dependent critical damage parameter. During any irreversible process inducing degradation in microstructure, internal entropy production increases according to second law of thermodynamics. Total entropy production due to mechanical and thermal dissipation can be calculated at any time step using Eqs. 218–222:

$$\eta_{mech} = \eta_{mech}|_{t=t_o} + \int_{t_o}^{t} \gamma_{mech} dt \tag{218}$$

$$\eta_{mech} = \eta_{mech}|_{t=t_o} + \int_{t_o}^{t} \left\{ \frac{J^{-1}}{\theta} \left[\left(\tau_I + \frac{1}{2}\gamma B |\ln(\mathbf{A})|^2 \right) \nu_I^p + \tau_M \nu_M^p \right] \right\} dt \tag{219}$$

$$\eta_{ther} = \eta_{ther}|_{t=t_o} + \int_{t_o}^{t} \gamma_{ther} dt \tag{220}$$

$$\eta_{ther} = \eta_{ther}|_{t=t_o} + \int_{t_o}^{t} \left\{ -\frac{1}{\theta^2} \mathrm{div}(\mathbf{J}_q) \bullet \nabla_x(\theta) + \frac{\rho r}{\theta} \right\} dt \tag{221}$$

$$\eta_{mech}|_{t=t_o} = \eta_{ther}|_{t=t_o} = 0 \tag{222}$$

Since a virgin material is at ground (undamaged) state ($t = t_o$), both mechanical and thermal entropy production can be taken as zero (Eq. 222). It is proposed that at failure internal entropy production reaches a critical value (η_{cr}) which depends on only temperature. Unlike metals, temperature dependence of critical entropy production is essential for the case of stretching of polymers due to change in failure mode of amorphous polymers. Amorphous polymers display a brittle failure at low temperatures ($\theta << \theta_g$) without any significant plastic dissipation, while at high temperatures ($\theta > \theta_g$) ductile failure occurs after significant amount of plastic work. Critical damage parameter (D_{cr}) is defined in such a way that at constant temperatures as $\eta \to \eta_{\mathrm{cr}}(\theta)$, $D \to 1$. Nonnegative entropy production assures that $D \geq 0$, while for an undamaged material ($\eta = 0$), damage is equal to zero ($D = 0$):

$$D_{\mathrm{cr}}(\theta) = \left[1 - \exp\left(-\frac{m_s}{R} \eta_{\mathrm{cr}}(\theta) \right) \right]^{-1} \tag{223}$$

According to incremental form of damage evolution in Eq. 224, damage will increase at a much faster rate at low temperatures (sudden brittle failure), while increase in damage will be relatively smaller at high temperatures (prolonged ductile failure). The power of damage evolution function (Eq. 217) lies in definition of critical damage parameter which exponentially decreases with increasing temperature. Damage evolution merely depends on evolution of a state function (entropy) which incorporates all aspects of material response to thermomechanical loading, and it can represent damage accumulation prior to different types of failure:

$$\Delta D = D_{\mathrm{cr}} \frac{m_s}{R} \exp\left(-\frac{m_s}{R} \eta_{mech} \right) \Delta\eta \tag{224}$$

Material Property Definitions

Material property definitions are keystones for complete and accurate material constitutive models. A proper definition for material property over a large temperature range should be continuous and smooth over transition region

(Bauwens-Crowet et al. 1969) and should include rate dependency of the property. It is important to note that selection of material property formulation is completely based on researchers' decision and can be substituted by other forms. Expressions for material properties presented herein are indeed mathematical tools to describe influence of temperature and rate on material behavior in a continuous form. Formulations for these material properties representing temperature dependence are completely in a form that provides continuity in temperature domain and has continuous first derivative with respect to temperature. A significant part of material properties can be obtained by conducting isothermal tests at different temperatures (both above and below glass transition temperature) and at different rates such as $\{E, \nu, I_m, \nu_I^o, Q_I, n_I, B_g, X_B, V, \alpha_p, \gamma\}$. Though the number of parameters is quite large, actual number of material properties is quite small. There is no direct way of observing some material parameters such as $\{h_I, b, g, \nu_M^o, Q_M, h_M, n_M, \mu_M, \phi^*, S_M^*\}$, yet these properties/parameters can be easily obtained by trial-error method in a simple version of numerical algorithm of material model developed in MATLAB®.

According to free-volume theory of William et al. (1955), plastic flow rule can be constructed for equivalent plastic shear strain rate at temperatures above θ_g using Williams-Landel-Ferry (WLF) equations. Similarly, rate dependence of glass transition temperature can be considered in terms of temperature-time equivalence of glass transition (Eq. 225):

$$\theta_g = \begin{cases} \theta_g^{ref} + \dfrac{c_2^g \log\left({}^{\nu}/_{\nu^{ref}}\right)}{c_1^g - \log\left({}^{\nu}/_{\nu^{ref}}\right)} & \nu > \nu^{ref} \\ \theta_g^{ref} & \nu \le \nu^{ref} \end{cases} \tag{225}$$

c_1^g and c_2^g in Eq. 225 are WLF parameters associated with θ_g, ν^{ref} is reference stretch rate, and ν is equivalent stretch rate which is defined in Eq. 226:

$$\nu = \sqrt{2}|\mathbf{D}| \tag{226}$$

Temperature and rate dependence of elastic modulus (E) is considered as

$$E = \left[\begin{array}{l} \dfrac{1}{2}\left(E_g + E_r\right) - \dfrac{1}{2}\left(E_g - E_r\right)\tanh\left(\dfrac{\theta - \left(\theta_g + \theta_E\right)}{\Delta_E}\right) \\ + X_E\left[\theta - \left(\theta_g + \theta_E\right)\right] \end{array}\right]\left[1 + s_E \log\left(\frac{\nu}{\nu^{ref}}\right)\right] \tag{227}$$

$$X_E = \begin{cases} X_g^E & \theta \le \theta_g + \theta_E \\ X_r^E & \theta > \theta_g + \theta_E \end{cases} \tag{228}$$

In Eq. 227, E_g and E_r are ground glassy and rubbery elastic modulus corresponding to temperatures confining glass-rubber transition region. X_E^g and X_E^r in Eq. 228 represent rate of change of elastic modulus with respect to temperature in glassy and rubbery domains, respectively. s_E is rate sensitivity of elastic modulus, while θ_E and Δ_E define origin temperature and width of glass-rubber transition, respectively. Experimental studies on temperature and rate dependence of storage

modulus and elastic modulus of PMMA indicate that PMMA is highly sensitive to rate and temperature. Modulus of PMMA continuously decreases with increasing temperature with a remarkable drop around θ_g over a 10–20 °C temperature domain depending on frequency of loading (Ye et al. 2003). Poisson's ration ($\overline{\nu}$) is assumed to be only temperature dependent as defined in Eq. 229:

$$\overline{\nu} = \frac{1}{2}\left(\overline{\nu}_g + \overline{\nu}_r\right) - \frac{1}{2}\left(\overline{\nu}_g - \overline{\nu}_r\right)\tanh\left(\frac{\theta - \left(\theta_g + \theta_E\right)}{\Delta_E}\right) \tag{229}$$

$\overline{\nu}_g$ and $\overline{\nu}_r$ in Eq. 229 are ground glassy and rubbery Poisson's ratio, respectively. Shear modulus (G) and bulk modulus (K) are defined in Eqs. 230 and 231:

$$G = \frac{E}{2(1+\overline{\nu})} \tag{230}$$

$$K = \frac{E}{3(1-2\overline{\nu})} \tag{231}$$

Temperature and rate dependence of rubbery modulus (μ_M) is modeled similar to elastic modulus (Eq. 233):

$$\mu_M = \left[\begin{array}{l}\frac{1}{2}\left(\mu_M^g + \mu_M^r\right) - \frac{1}{2}\left(\mu_M^g - \mu_M^r\right)\tanh\left(\frac{\theta - \left(\theta_g + \theta_\mu\right)}{\Delta_\mu}\right) \\ + X_\mu\left[\theta - \left(\theta_g + \theta_\mu\right)\right]\end{array}\right]\left[1 + s_\mu \log\left(\frac{\nu}{\nu^{ref}}\right)\right] \tag{232}$$

$$X_\mu = \begin{cases} X_\mu^g & \theta \le \theta_g + \theta_\mu \\ X_\mu^r & \theta > \theta_g + \theta_\mu \end{cases} \tag{233}$$

Definitions of rubbery shear modulus parameters in Eq. 233 are identical to those for elastic modulus. Temperature dependence of critical value of order parameter (φ^*), limit of polymer chain extensibility (I_M), and saturation value of plastic flow resistance of molecular network (S_M^*) are defined in Eqs. 234–239:

$$\varphi^* = \frac{1}{2}\left(\varphi_g^* + \varphi_r^*\right) - \frac{1}{2}\left(\varphi_g^* - \varphi_r^*\right)\tanh\left(\frac{\theta - \left(\theta_g + \theta_\varphi\right)}{\Delta_\varphi}\right) + X_\varphi\left[\theta - \left(\theta_g + \theta_\varphi\right)\right] \tag{234}$$

$$X_\varphi = \begin{cases} X_\varphi^g & \theta \le \theta_g + \theta_\varphi \\ X_\varphi^r & \theta > \theta_g + \theta_\varphi \end{cases} \tag{235}$$

$$I_M = \frac{1}{2}\left(I_M^g + I_M^r\right) - \frac{1}{2}\left(I_M^g - I_M^r\right)\tanh\left(\frac{\theta - \left(\theta_g + \theta_M\right)}{\Delta_M}\right) + X_M\left[\theta - \left(\theta_g + \theta_M\right)\right] \tag{236}$$

$$X_M = \begin{cases} X_M^g & \theta \le \theta_g + \theta_\mu \\ X_M^r & \theta > \theta_g + \theta_\mu \end{cases} \tag{237}$$

$$S_M^* = \frac{1}{2}\left(S_M^g + S_M^r\right) - \frac{1}{2}\left(S_M^g - S_M^r\right)\tanh\left(\frac{\theta - \left(\theta_g + \theta_S\right)}{\Delta_S}\right) + X_S\left[\theta - \left(\theta_g + \theta_S\right)\right] \quad (238)$$

$$X_S = \begin{cases} X_S^g & \theta \le \theta_g + \theta_S \\ X_S^r & \theta > \theta_g + \theta_S \end{cases} \quad (239)$$

Definitions of parameters in Eqs. 234–239 are identical to those for elastic modulus. Saturation value for network resistance (S_M^*) and critical value of disorder parameter (φ^*) were assumed to decrease with increasing temperature, while limited chain extensibility (I_M) increases with increasing temperature based on observations in experiments. Similar to models of Richeton (Bauwens-Crowet et al. 1969) and Anand (Boyce et al. 2000), back stress is assumed to vanish above θ_g, but decrease in back stress modulus (B) with increasing temperature is finalized with asymptotical approach to zero at temperature around θ_g, as defined in Eqs. 240–241:

$$B = B_g\left(1 - \tanh\left(\frac{\theta - \theta_g}{\Delta_B}\right)\right) + X_B\left(\theta_g - \theta\right) \quad (240)$$

$$X_B = \begin{cases} X_B^g & \theta \le \theta_g \\ 0 & \theta > \theta_g \end{cases} \quad (241)$$

Parameters b and g characterizing hardening-softening behavior in intermolecular structure are defined in Eqs. 242–244, respectively:

$$g = \frac{1}{2}\left(g_g + g_r\right) - \frac{1}{2}\left(g_g - g_r\right)\tanh\left(\frac{\theta - \left(\theta_g + \theta^g\right)}{\Delta^g}\right) + X_g\left[\theta - \left(\theta_g + \theta^g\right)\right] \quad (242)$$

$$X_g = \begin{cases} X_g^g & \theta \le \theta_g + \theta^g \\ 0 & \theta > \theta_g + \theta^g \end{cases} \quad (243)$$

$$b = b_1 \exp(b_2\theta)\left(\frac{\nu_I^p}{\nu_{ref}^p}\right)^{b_3} \quad (244)$$

Other parameters involved in material model $\{\nu_I^o, Q_I, V, \alpha_p, n_I, h_I, \gamma, \nu_M^o, Q_M, h_M, n_M\}$ are assumed to be constant.

Conclusions

An improved version of dual-mechanism viscoplastic material model is presented to predict thermomechanical behavior of PMMA under non-isothermal conditions. In order to resolve problems associated with material response predictions around glass transition temperature in non-isothermal conditions similar to actual polymer processing operations which includes a temperature drop from temperatures above to temperatures below inducing a transition from rubbery state to solid

state, evolution of state variables and flow rules that control viscoplastic characteristics in the model and material property definitions is carefully formulated to ensure smooth transition in response around.

In order to successfully simulate material response under isothermal and non-isothermal conditions, material parameters for the model were determined from isothermal tests on PMMA. Since PMMA response is highly sensitive to time and temperature, tests at different loading rates and different temperatures were conducted. In numerical simulations of isothermal and non-isothermal stretching of PMMA, dual-mechanism viscoplastic model is implemented in ABAQUS® through user-defined material subroutine. Simulation and experimental results are reasonably in good agreement for both isothermal and non-isothermal cases, and simulations can correctly predict rate and temperature dependence of material response. The accuracy of numerical simulation of non-isothermal simulations depends on the description of heat transfer model which requires extensive preliminary studies.

Non-isothermal simulation results corresponding to forming step were especially accurate describing all possible salient features of PMMA under all test conditions (non-isothermal H series, near-isothermal M series, and isothermal L series). Unified formulation of plastic flow rule for intermolecular structure and appropriate material property definitions ensure smooth transition of response around which is bound to take place at some stage of all non-isothermal simulations. Simulation results are reasonably good in predicting some important features observed in dwell step such as molecular relaxation and increase in axial force with thermal shrinkage at fixed deformation. Softening/hardening behavior of PMMA was also predicted accurately to certain extent in most of the cases.

Since material state continuously changes with temperature during non-isothermal cases, experimental measures cannot be used for continuous monitoring of damage evolution. Entropy based damage model that is used for damage quantification in polymer processing technique was used to predict the material degradation and failure in polymer processing operations. Thermal part of irreversible entropy production is shown to be unrelated to mechanical damage, while damage evolution based on irreversible mechanical entropy production was able to predict the failure in non-isothermal tests for low forming temperatures. It is shown that in damage induced by polymer processing operations, a sample increases with decreasing forming rate and forming temperature.

References

N.M. Ames et al., A thermo-mechanically coupled theory for large deformations of amorphous polymers. Part II: applications. Int. J. Plast. **25**(8), 1495–1539 (2009)

L. Anand, Moderate deformations in extension-torsion of incompressible isotropic elastic materials. J. Mech. Phys. Solids **34**, 293–304 (1986)

L. Anand, H. On, Hencky's approximate strain-energy function for moderate deformations. J. Appl. Mech. **46**, 78–82 (1979)

L. Anand et al., A thermo-mechanically coupled theory for large deformations of amorphous polymers. Part I: formulation. Int. J. Plast. **25**(8), 1474–1494 (2009)

N. Aravas, Finite-strain anisotropic plasticity and the plastic spin. Model. Simul. Mater. Sci. Eng. **2**(3A), 483–504 (1994)
E.M. Arruda, M.C. Boyce, Evolution of plastic anisotropy in amorphous polymers during finite straining. Int. J. Plast. **9**(6), 697–720 (1993)
E.M. Arruda, M.C. Boyce, R. Jayachandran, Effects of strain rate, temperature and thermomechanical coupling on the finite strain deformation of glassy polymers. Mech. Mater. **19**(2–3), 193–212 (1995)
C. Basaran, M. Lin, Damage mechanics of electromigration induced failure. Mech. Mater. **40** (1–2), 66–79 (2007)
C. Basaran, S. Nie, A thermodynamics based damage mechanics model for particulate composites. Int. J. Solids Struct. **44**(3–4), 1099–1114 (2007)
C. Basaran, C.Y. Yan, A thermodynamic framework for damage mechanics of solder joints. J. Electron. Packag. **120**(4), 379–384 (1998)
C. Bauwens-Crowet, J.C. Bauwens, G. Homès, Tensile yield-stress behavior of glassy polymers. J. Polym. Sci. Part A-2 Polym. Phys. **7**(4), 735–742 (1969)
J.S. Bergström, M.C. Boyce, Constitutive modeling of the large strain time-dependent behavior of elastomers. J. Mech. Phys. Solids **46**(5), 931–954 (1998)
L. Boltzmann, *Lectures on Gas Theory* (Dover, New York, 1995)
M.C. Boyce, S. Socrate, P.G. Llana, Constitutive model for the finite deformation stress–strain behavior of poly(ethylene terephthalate) above the glass transition. Polymer **41** (6), 2183–2201 (2000)
A. Chudnovsky et al., in *On Fracture of Solids in Studies on Elasticity and Plasticity*, ed. by L. Kachanov (Leningrad University Press, Leningrad, 1973), pp. 3–41
A.C. Eringen, *Mechanics of Continua* (Wiley, New York, 1967)
D.G. Fotheringham, B.W. Cherry, The role of recovery forces in the deformation of linear polyethylene. J. Mater. Sci. **13**(5), 951–964 (1978)
D. Fotheringham, B.W. Cherry, C. Bauwens-Crowet, Comment on "the compression yield behaviour of polymethyl methacrylate over a wide range of temperatures and strain-rates". J. Mater. Sci. **11**(7), 1368–1371 (1976)
P. Francisco, S. Gustavo, B.H. Élida, Temperature and strain rate dependence of the tensile yield stress of PVC. J. Appl. Polym. Sci. **61**(1), 109–117 (1996)
A.N. Gent, A new constitutive relation for rubber. Rubber Chem. Technol. **69**(1), 59–61 (1996)
J. Gomez, C. Basaran, Damage mechanics constitutive model for Pb/Sn solder joints incorporating nonlinear kinematic hardening and rate dependent effects using a return mapping integration algorithm. Mech. Mater. **38**(7), 585–598 (2006)
L.M. Kachanov, Introduction to continuum damage mechanics (M. Nijhoff, Dordrecht/Boston, 1986)
E. Kontou, G. Spathis, Application of finite strain viscoplasticity to polymeric fiber composites. Int. J. Plast. **22**(7), 1287–1303 (2006)
E. Kröner, Allgemeine Kontinuumstheorie der Versetzungen und Eigenspannungen. Arch. Ration. Mech. Anal. **4**(1), 273–334 (1959)
E.H. Lee, Elastic–plastic deformation at finite strains. J. Appl. Mech. **36**, 1–6 (1969)
J. Mandel, *Plasticite Classique et Viscoplasticite (Lecture Notes)* (International Center for Mechanical Sciences, Udine, 1972)
G. Palm, R.B. Dupaix, J. Castro, Large strain mechanical behavior of poly(methyl methacrylate) (PMMA) near the glass transition temperature. J. Eng. Mater. Technol. **128**(4), 559–563 (2006)
F. Povolo, B.H. Élida, Phenomenological description of strain rate and temperature-dependent yield stress of PMMA. J. Appl. Polym. Sci. **58**(1), 55–68 (1995)
J. Richeton et al., A formulation of the cooperative model for the yield stress of amorphous polymers for a wide range of strain rates and temperatures. Polymer **46**(16), 6035–6043 (2005a)
J. Richeton et al., A unified model for stiffness modulus of amorphous polymers across transition temperatures and strain rates. Polymer **46**(19), 8194–8201 (2005b)

J. Richeton et al., Influence of temperature and strain rate on the mechanical behavior of three amorphous polymers: characterization and modeling of the compressive yield stress. Int. J. Solids Struct. **43**(7–8), 2318–2335 (2006)
J. Richeton et al., Modeling and validation of the large deformation inelastic response of amorphous polymers over a wide range of temperatures and strain rates. Int. J. Solids Struct. **44**(24), 7938–7954 (2007)
V. Srivastava et al., A thermo-mechanically-coupled large-deformation theory for amorphous polymers in a temperature range which spans their glass transition. Int. J. Plast. **26**(8), 1138–1182 (2010)
J. Sweeney et al., Application of an elastic model to the large deformation, high temperature stretching of polypropylene. Polymer **38**(24), 5991–5999 (1997)
C. Truesdell, *The Non-linear Field Theories of Mechanics*, 3rd edn. (Springer, New York, 2004)
M. Wallin, M. Ristinmaa, Deformation gradient based kinematic hardening model. Int. J. Plast. **21** (10), 2025–2050 (2005)
M.L. Williams, R.F. Landel, J.D. Ferry, The temperature dependence of relaxation mechanisms in amorphous polymers and other glass-forming liquids. J. Am. Chem. Soc. **77**(14), 3701–3707 (1955)
H. Ye, C. Basaran, D.C. Hopkins, Damage mechanics of microelectronics solder joints under high current densities. Int. J. Solids Struct. **40**(15), 4021–4032 (2003)

Thermodynamics Theory for Damage Evolution in Solids

23

Cemal Basaran, Shihua Nie, Juan Gomez, Eray Gunel, Shidong Li, Minghui Lin, Hong Tang, Chengyong Yan, Wei Yao, and Hua Ye

Contents

Abstract

In this chapter the thermodynamic theory behind damage mechanics is presented. The presented damage evolution model is purely physical, rather than empirical. Entropy production rate is used as a damage metric. It is

C. Basaran (✉) • S. Nie • J. Gomez • E. Gunel • S. Li • M. Lin • H. Tang • C. Yan • W. Yao • H. Ye
Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA
e-mail: cjb@buffalo.edu; shihwa.nie@gmail.com; jgomezc1@eafit.edu.co; eraygunel@gmail.com; shidongli@us.ibm.com; minghuil@hotmail.com; hongtang2000@yahoo.com; weiyao@buffalo.edu; weiyaoepl@gmail.com; huaye@microsoft.com

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_31

shown that when entropy production rate is used as a damage metric, damage due to numerous related and unrelated external and internal sources can be combined into a single universal damage term, which is not possible with any phenomenological damage evolution model.

Introduction

The science of thermodynamics, which began by treating the relations among heat, work, and the intrinsic properties of the systems in equilibrium, has developed into a very general science of energetics for all types of systems: mechanical, chemical, and electrical, whether in equilibrium or not (De Groot and Mazur 1962; Yourgrau et al. 1966; Haase 1969; Germain et al. 1983; Ericksen 1998). Irreversible thermodynamics provides a general framework for the macroscopic description of irreversible processes. In irreversible thermodynamics, the so-called balance equation for the entropy plays a central role. This equation expresses the fact that the entropy of a volume element changes with time for two reasons. First, it changes because entropy flows into the controlled volume, and second, it changes because there is an entropy source due to irreversible phenomena inside the volume element. The entropy source is always a nonnegative quantity, since entropy can only be created, never destroyed. For reversible transformations the entropy source vanishes. Entropy is a measure of how much energy is unavailable for work.

The entropy of the universe increases or remains constant in all natural processes. It is possible to find a system for which entropy decreases but only due to a net increase in a related system. For example, the originally hot objects and cooler objects reaching thermal equilibrium in an isolated system may be separated and some of them put in a refrigerator. The objects would again have different temperatures after a period of time, but now the system of the refrigerator would have to be included in the analysis of the complete system. No net decrease in entropy of all the related system occurs. This is yet another way of stating the second law of thermodynamics (DeHoff 1993).

The concept of entropy has far-reaching implications that tie the order of our universe to probability and statistics. Imagine a new deck of cards in order by suits, with each unit in numerical order. As the deck is shuffled, no one would expect the original order to return. There is a probability that the randomized order of the shuffled deck would return to the original format, but it is exceedingly small. An ice cube melts, and the molecules in the liquid form have less order than in the frozen form. An infinitesimally small probability exists that all of the slower moving molecules will aggregate in one space so that the ice cube will reform in the exactly same lattice formation. The entropy, or disorder, of the universe increases as hot bodies cool and cold bodies warm. Eventually, the entire universe will be at the same temperature so the energy will be no longer usable (DeHoff 1993).

To relate the entropy source explicitly to the various irreversible processes that occur in a system, one needs the macroscopic conservation laws of mass, momentum, and energy in local, i.e., differential form. These conservation laws contain a number

of quantities such as the diffusion flows, the heat flow, and the stress tensor, which are related to the transport of mass, energy, and momentum. The entropy source may then be calculated by using the thermodynamic Gibbs relation, which connects the rate of the change in entropy in the medium to the rate of the change in energy and work. "It turns out that the entropy source has a very simple appearance: it is a sum of terms each being a product of a flux characterizing an irreversible process, and a quantity called thermodynamic force which is related to the non-uniformity of the system"(Groot and Mazur 1962). The entropy source strength can thus serve as a basis for the systematic description of the irreversible processes occurring in a system.

"As yet the set of conservation laws, together with the entropy balance equation and the equations of state are to a certain extent empty, since this set of equations contain the irreversible fluxes as unknown parameters and can therefore not be solved with the given initial and boundary conditions for the state of the system." At this point we must therefore supplement the equations by an additional set of relationships, which relate the irreversible fluxes and the thermodynamic forces appearing in the entropy source strength. Irreversible thermodynamics, in its present form, is mainly restricted to the study of the linear relationship between the fluxes and the thermodynamic forces as well as possible cross-effects between various phenomena. This is not a very serious restriction however, since even rather extreme physical situations are still described by linear laws" (Groot and Mazur 1962).

Conservation Laws

Thermodynamics is based on two fundamental laws: the first law of thermodynamics (law of conservation of energy) and the second law of thermodynamics (entropy law). A systematic macroscopic scheme for the description of irreversible processes must also be built upon these two laws. However, it is necessary to formulate these laws in a suitable way. Since we wish to develop a theory applicable to systems of which the properties are continuous functions of space coordinates and time, we shall give a local formulation of the law of conservation of energy. As the local momentum and mass densities may change in time, we will also need local formulations of the laws of conservation of momentum and mass. For general purposes in solid mechanics, the thermodynamic system will usually be chosen as a given collection of continuous matter.

Conservation of Mass

Consider an arbitrary volume V fixed in space, bounded by surface Ω. The rate of change of the mass within the volume V is (Malvern 1969)

$$\frac{d}{dt}\int^{V} \rho dV = \int^{V} \frac{\partial \rho}{\partial t} dV \tag{1}$$

where ρ is the density (mass per unit volume). If no mass is created or destroyed inside V, this quantity must be equal to the rate of the material flow into the volume V through its surface Ω (Malvern 1969):

$$\int_V \frac{\partial \rho}{\partial t} dV = -\int_\Omega \rho \boldsymbol{\upsilon} \cdot d\boldsymbol{\Omega} \tag{2}$$

where $\boldsymbol{\upsilon}$ is the velocity and $d\boldsymbol{\Omega}$ is a vector with magnitude $d\Omega$ normal to the surface and counted positive from the inside to the outside. The quantities ρ and $\boldsymbol{\upsilon}$ are all functions of time and of space coordinates. Applying Gauss's theorem to the surface integral in Eq. 2, we obtain

$$\frac{\partial \rho}{\partial t} = -\mathrm{div} \rho \boldsymbol{\upsilon} \tag{3}$$

Equation 3 is valid for an arbitrary volume V, which expresses the fact that the total mass is conserved, i.e., that the total mass in any volume element of the system can only change if matter flows into (or out of) the volume element. This equation has the form of a so-called balanced equation: the local change of the density is equal to the negative divergence of the flow of mass. The continuity equation in the vector form of Eq. 3 is independent of any choice of coordinate system.

The conservation of mass equation can also be written in an alternative form by introducing the substantial time derivative (Groot and Mazur 1962):

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \boldsymbol{\upsilon} \cdot \mathrm{grad} \tag{4}$$

With the help of Eqs. 4 and 3, it becomes (Groot and Mazur 1962)

$$\frac{d\rho}{dt} = -\rho \mathrm{div} \boldsymbol{\upsilon} \tag{5}$$

With the specific volume $v = \rho^{-1}$, the formula (5) may also be written as (Groot and Mazur 1962)

$$\rho \frac{dv}{dt} = \mathrm{div} \boldsymbol{\upsilon} \tag{6}$$

Finally the following relation is valid for an arbitrary local property a that may be a scalar or a component of a vector or tensor (Groot and Mazur 1962):

$$\rho \frac{da}{dt} = \frac{\partial a \rho}{\partial t} + \mathrm{div} a \rho \boldsymbol{\upsilon} \tag{7}$$

which is a consequence of Eqs. 3 and 4. We can verify Eq. 7 directly. According to Eq. 4, the left-hand side of Eq. 7 is

$$\rho \frac{da}{dt} = \rho \frac{\partial a}{\partial t} + \rho \boldsymbol{\upsilon} \cdot \text{grad}\, a$$

According to Eq. 3, the right side of Eq. 7 is

$$\begin{aligned} \frac{\partial a\rho}{\partial t} + \text{div} a\rho \boldsymbol{\upsilon} &= a \frac{\partial \rho}{\partial t} + \rho \frac{\partial a}{\partial t} + a\text{div}\rho \boldsymbol{\upsilon} + \rho \boldsymbol{\upsilon} \cdot \text{grad}\, a \\ &= a(-\text{div}\rho \boldsymbol{\upsilon}) + \rho \frac{\partial a}{\partial t} + a\text{div}\rho \boldsymbol{\upsilon} + \rho \boldsymbol{\upsilon} \cdot \text{grad}\, a \\ &= \rho \frac{\partial a}{\partial t} + \rho \boldsymbol{\upsilon} \cdot \text{grad}\, a \end{aligned}$$

So Eq. 7 is true.

Momentum Principle

The momentum principle for a collection of particles states that the time rate of the change in the total momentum for a given set of particles equals to the vector sum of all the external forces acting on the particles of the set, provided Newton's third law of action and reaction governs the initial forces (Malvern 1969). Consider a given mass of the medium, instantaneously occupying a volume V bounded by surface Ω and acted upon by external surface $\mathbf{t}$ and body force $\mathbf{b}$. Then the momentum principle can be expressed as (Malvern 1969)

$$\int_{\Omega} \mathbf{t} d\Omega + \int_{V} \rho \mathbf{b} dV = \frac{d}{dt} \int_{V} \rho \boldsymbol{\upsilon} dV \tag{8}$$

or in rectangular coordinates

$$\int_{\Omega} t_i d\Omega + \int_{V} \rho b_i dV = \frac{d}{dt} \int_{V} \rho \upsilon_i dV \tag{9}$$

Substituting $t_i = \sigma_{ji} n_j$ and transforming the surface integral by using the divergence theorem, we obtain (Malvern 1969)

$$\int_{V} \left(\frac{\partial \sigma_{ji}}{\partial x_j} + \rho b_i - \rho \frac{d\upsilon_i}{dt} \right) dV = 0 \tag{10}$$

for an arbitrary volume V. Whence at each point we have (Malvern 1969)

$$\rho \frac{d\upsilon_i}{dt} = \frac{\partial \sigma_{ji}}{\partial x_j} + \rho b_i \tag{11}$$

where n_j is the component of the normal unit vector $\mathbf{n}$, $v_i (i = 1, 2, 3)$ is a Cartesian component of $\boldsymbol{\upsilon}$, and $x_j (j = 1, 2, 3)$ is the Cartesian coordinates. The quantities $\sigma_{ji} (i, j = 1, 2, 3)$ and $b_i (i = 1, 2, 3)$ are the Cartesian components of the stress tensor $\boldsymbol{\sigma}$ and body force $\mathbf{b}$, respectively. For a nonpolar case the stress tensor $\boldsymbol{\sigma}$ is symmetric, namely,

$$\sigma_{ij} = \sigma_{ji} \quad (i, j = 1, 2, 3) \tag{12}$$

In tensor notation Eq. 11 is written as (Groot and Mazur 1962)

$$\rho \frac{d\boldsymbol{\upsilon}}{dt} = \mathrm{Div}\boldsymbol{\sigma} + \rho \mathbf{b} \tag{13}$$

From a microscopic point of view, the stress tensor $\boldsymbol{\sigma}$ results from the short-range interactions between the particles of the system, whereas $\mathbf{b}$ contains the external forces as well as a possible contribution from long-range interactions in the system.

Using relation (7), the equation of motion (13) can also be written as

$$\frac{\partial \rho \boldsymbol{\upsilon}}{\partial t} = -\mathrm{Div}(\rho \boldsymbol{\upsilon}\boldsymbol{\upsilon} - \boldsymbol{\sigma}) + \rho \mathbf{b} \tag{14}$$

where $\boldsymbol{\upsilon}\boldsymbol{\upsilon} = \boldsymbol{\upsilon} \otimes \boldsymbol{\upsilon}$ is an ordered (dyadic) product. This equation also has the form of a balance equation for the momentum density $\rho\boldsymbol{\upsilon}$. In fact one can interpret the quantity $(\rho\boldsymbol{\upsilon}\boldsymbol{\upsilon} - \boldsymbol{\sigma})$ as a momentum flow with a convective part $\rho\boldsymbol{\upsilon}\boldsymbol{\upsilon}$ and the quantity $\rho\mathbf{b}$ as a source of momentum.

It is also possible to derive from Eq. 11 a balance equation for the kinetic energy of the center of gravity motion by multiplying both members with the component v_i of $\boldsymbol{\upsilon}$ and summing over i

$$\rho \frac{d\frac{1}{2}\boldsymbol{\upsilon}^2}{dt} = \sum_{i,j=1}^{3} \frac{\partial}{\partial x_j}\left(\sigma_{ji} v_i\right) - \sum_{i,j=1}^{3} \sigma_{ji} \frac{\partial}{\partial x_j} v_i + \rho b_i v_i \qquad (i = 1, 2, 3) \tag{15}$$

or in tensor notation

$$\rho \frac{d\frac{1}{2}\boldsymbol{\upsilon}^2}{dt} = \mathrm{div}(\boldsymbol{\sigma} \cdot \boldsymbol{\upsilon}) - \boldsymbol{\sigma} : \mathbf{L} + \rho \mathbf{b} \cdot \boldsymbol{\upsilon} \tag{16}$$

where $\mathbf{L} = \mathrm{Grad}\,\boldsymbol{\upsilon}$ is the spatial gradient of the velocity. $\mathbf{L}$ can be written as the sum of a symmetric tensor $\mathbf{D}$ called the rate of deformation tensor or the stretch tensor and a skew symmetric tensor $\mathbf{W}$ called the spin tensor or the vorticity tensor as follows (Malvern 1969):

$$\mathbf{L} = \mathbf{D} + \mathbf{W} \tag{17}$$

where $\mathbf{D} = \frac{1}{2}\left(\mathbf{L} + \mathbf{L}^T\right)$, $\mathbf{W} = \frac{1}{2}\left(\mathbf{L} - \mathbf{L}^T\right)$.

Since $\mathbf{W}$ is skew symmetric, while $\boldsymbol{\sigma}$ is symmetric, it follows that

$$\boldsymbol{\sigma} : \text{Grad}\ \boldsymbol{\upsilon} = \sigma_{ij} L_{ij} = \sigma_{ij} D_{ij} = \boldsymbol{\sigma} : \mathbf{D} \tag{18}$$

We can also establish the relationship between the strain rate $d\boldsymbol{\varepsilon}/dt$ and the rate of the deformation tensor $\mathbf{D}$ (Malvern 1969):

$$\frac{d\boldsymbol{\varepsilon}}{dt} = \mathbf{F}^T \cdot \mathbf{D} \cdot \mathbf{F} \tag{19}$$

where $\mathbf{F}$ is the deformation gradient referred to the undeformed configuration. When the displacement gradient components are small compared to unity, Eq. 19 is reduced to (Malvern 1969)

$$\frac{d\boldsymbol{\varepsilon}}{dt} \approx \mathbf{D} \tag{20}$$

With the help of Eq. 7, Eq. 16 becomes

$$\frac{\partial \frac{1}{2}\rho \boldsymbol{\upsilon}^2}{\partial t} = -\text{div}\left(\frac{1}{2}\rho \boldsymbol{\upsilon}^2 \cdot \boldsymbol{\upsilon} - \boldsymbol{\sigma} \cdot \boldsymbol{\upsilon}\right) - \boldsymbol{\sigma} : \mathbf{D} + \rho \mathbf{b} \cdot \boldsymbol{\upsilon} \tag{21}$$

For the conservative body forces which can be derived from a potential ψ independent of time (Groot and Mazur 1962),

$$\mathbf{b} = -\text{grad}\ \psi, \quad \frac{\partial \psi}{\partial t} = 0 \tag{22}$$

We can now establish an equation for the rate of change of the potential energy density $\rho\psi$. In fact it follows from Eqs. 3 and 22 that

$$\begin{aligned} \frac{\partial \rho\psi}{\partial t} &= \psi \frac{\partial \rho}{\partial t} + \rho \frac{\partial \psi}{\partial t} = \psi(-\text{div}\rho\boldsymbol{\upsilon}) \\ &= -\text{div}\rho\psi\boldsymbol{\upsilon} + \rho\boldsymbol{\upsilon} \cdot \text{grad}\ \psi = -\text{div}\rho\psi\boldsymbol{\upsilon} - \rho\mathbf{b} \cdot \boldsymbol{\upsilon} \end{aligned} \tag{23}$$

Adding Eqs. 22 and 23 for the rate of change of the kinetic energy $\frac{1}{2}\rho\boldsymbol{\upsilon}^2$ and the potential energy $\rho\psi$,

$$\frac{\partial \rho\left(\frac{1}{2}\boldsymbol{\upsilon}^2 + \psi\right)}{\partial t} = -\text{div}\left\{\rho\left(\frac{1}{2}\boldsymbol{\upsilon}^2 + \psi\right)\boldsymbol{\upsilon} - \boldsymbol{\sigma} \cdot \boldsymbol{\upsilon}\right\} - \boldsymbol{\sigma} : \mathbf{D} \tag{24}$$

This equation shows that the sum of kinetic and potential energy is not conserved, since a sink term appears at the right-hand side.

Conservation of Energy

The first law of thermodynamics relates the work done on the system and the heat transfer into the system to the change in energy of the system. Suppose that the only energy transferred to the system is by mechanical work done on the system by surface tractions and body forces, by heat exchange through the boundary, and by the heat generated within the system by external agencies (inductive heating). According to the principle of conservation of energy, the total energy content within an arbitrary volume V in the system can only change if energy flows into (out of) the volume considered through its boundary Ω, which can be expressed as (Malvern 1969)

$$\frac{d}{dt}\overset{V}{\int}\rho e dV = \overset{V}{\int}\frac{\partial \rho e}{\partial t}dV = -\overset{S}{\int}\mathbf{J_e}\cdot d\mathbf{\Omega} + \overset{V}{\int}\rho r dV \tag{25}$$

where e is the energy per unit mass, $\mathbf{J_e}$ is the energy flux per unit surface and unit time, and r is the strength of the distributed internal heat source per unit mass. We shall refer to e as the total specific energy, because it includes all forms of energy in the system. Similarly we shall call $\mathbf{J_e}$ the total energy flux. With the help of Gauss's theorem, we obtain the differential or local form of the law of conservation of energy:

$$\frac{\partial \rho e}{\partial t} = -\mathrm{div}\mathbf{J_e} + \rho r \tag{26}$$

In order to relate this equation to the previously obtained Eq. 24 for the kinetic and potential energy, we must specify what are the various contributions to the energy e and the flux $\mathbf{J_e}$. The total specific energy e includes the specific kinetic energy $\frac{1}{2}\boldsymbol{\upsilon}^2$, the specific potential energy ψ, and the specific internal energy u (Groot and Mazur 1962):

$$e = \frac{1}{2}\boldsymbol{\upsilon}^2 + \psi + u \tag{27}$$

From a macroscopic point of view, this relation can be considered as the definition of internal energy, u. From a microscopic point of view, u represents the energy of thermal agitation as well as the energy due to the short-range molecular interactions.

Similarly, the total energy flux includes a convective term $\rho e\boldsymbol{\upsilon}$, an energy flux $\boldsymbol{\sigma}\cdot\boldsymbol{\upsilon}$ due to the mechanical work performed on the system, and finally a heat flux $\mathrm{J_q}$ (Groot and Mazur 1962):

$$\mathbf{J_e} = \rho e\boldsymbol{\upsilon} - \boldsymbol{\sigma}\cdot\boldsymbol{\upsilon} + \mathrm{J_q} \tag{28}$$

This equation may be also considered as defining the heat flux $\mathrm{J_q}$. Then the heat flowing rate per unit mass is

$$\rho \frac{dq}{dt} = -\mathrm{div}\mathrm{J_q} \tag{29}$$

where q is the heat flowing into the system per unit mass. If we subtract Eq. 24 from Eq. 26, we obtain, using also Eqs. 27 and 28, the balance equation for the internal energy u:

$$\frac{\partial \rho u}{\partial t} = -\mathrm{div}\left\{\rho u \boldsymbol{\upsilon} + \mathrm{J_q}\right\} + \boldsymbol{\sigma} : \mathbf{D} + \rho r \tag{30}$$

It is apparent from Eq. 30 that the internal energy u is not conserved. In fact a source term appears which is equal but of opposite sign to the source term of the balance Eq. 24 for kinetic and potential energy.

With the help of Eqs. 7 and 30, it may be written in an alternative form:

$$\rho \frac{du}{dt} = -\mathrm{div}\mathrm{J_q} + \boldsymbol{\sigma} : \mathbf{D} + \rho r \tag{31}$$

The total stress tensor $\boldsymbol{\sigma}$ can be split into a scalar hydrostatic pressure part p and a deviatoric stress tensor $\boldsymbol{\sigma}'$ (Malvern 1969):

$$\boldsymbol{\sigma} = -p\mathbf{I} + \boldsymbol{\sigma}' \tag{32}$$

where $\mathbf{I}$ is the unit matrix with element $\delta_{ij} (\delta_{ij} = 1, \text{if } i = j; \delta_{ij} = 0, \text{if } i \neq j), p = -\frac{1}{3}\sigma_{kk}$. With the help of Eqs. 32 and 31, it becomes

$$\rho \frac{du}{dt} = -\mathrm{div}\mathrm{J_q} - p\mathrm{div}\boldsymbol{\upsilon} + \boldsymbol{\sigma}' : \mathbf{D} + \rho r \tag{33}$$

where use has been made of the equality

$$\mathbf{I} : \mathbf{D} = \mathbf{I} : \mathrm{Grad}\ \boldsymbol{\upsilon} = \sum_{i,j=1}^{3} \delta_{ij} \frac{\partial}{\partial x_j} \upsilon_i = \sum_{i=1}^{3} \frac{\partial}{\partial x_i} \upsilon_i = \mathrm{div}\ \boldsymbol{\upsilon}$$

With Eq. 6, the first law of thermodynamics can finally be written in the form

$$\frac{du}{dt} = -v\ \mathrm{div}\mathrm{J_q} - p\frac{dv}{dt} + v\boldsymbol{\sigma}' : \mathbf{D} + r \tag{34}$$

where $v \equiv \rho^{-1}$ is the specific volume.

Entropy Production and Entropy Balance

Thermodynamics in the traditional sense is concerned with the study of reversible process. For an irreversible process in which the thermodynamic state of a solid changes from some initial state to a current state, it can be assumed that such a

process can occur along an imaginary reversible isothermal path. The processes defined in this way will be thermodynamically admissible if the Clausius-Duhem inequality is satisfied.

According to the principles of thermodynamics, two more new variables – temperature T and entropy S – are introduced for any macroscopic system. The entropy of the universe, taken as a system plus whatever surroundings are involved in producing the change within the system, can only increase. Entropy changes in the solids are always irreversible processes because of friction, which results in the production of entropy and thus a permanent change in the universe (Dehoff 1993).

The variation of the entropy dS may be written as the sum of two and only two terms for a closed system (Groot and Mazur 1962):

$$dS = dS_e + dS_i \tag{35}$$

where dS_e is the entropy derived from the transfer of heat from external sources across the boundary of the system and dS_i is the entropy produced inside the system. The second law of thermodynamics states that dS_i must be zero for any reversible (or equilibrium) process and positive for irreversible process of the system, namely (Groot and Mazur 1962),

$$dS_i \geq 0 \tag{36}$$

The entropy supplied, dS_e, on the other hand may be positive, zero, or negative, depending on the interaction of the system with its surroundings.

As we know, thermodynamics in the traditional sense is concerned with the study of the reversible transformation for which the equality in Eq. 36 holds. For an irreversible process in which the thermodynamic state of a solid changes from some initial state to a current state, it is assumed that such a process can occur along an imaginary reversible isothermal path which consists of a two-step sequence (Krajcinovic 1996). This is the so-called local equilibrium assumption, which postulates that the thermodynamic state of a material medium at a given point and instant is completely defined by the knowledge of the values of a certain number of variables at that instant. The method of local state implies that the laws which are valid for the macroscopic system remain valid for infinitesimally small parts of it. This method also implies, on a microscopic model, that the local macroscopic measurements performed on the system are really measurements of the properties of small parts of the system, which still contain a large number of the constituting particles. This hypothesis of "local equilibrium" can, from a macroscopic point of view, only be justified by virtue of the validity of the conclusions derived from it. Ultrarapid phenomena for which the timescale of the evolutions is at the same order as the relaxation time for a return to thermodynamic equilibrium are excluded from this theory's field of application (Lemaitre and Chaboche 1990). The physical phenomena that can be described with a precision depend on the choice of the number of state variables if the Clausius-Duhem inequality is satisfied.

In irreversible thermodynamics, one of the important objectives is to relate the dS_i, the internal entropy production, to the various irreversible phenomena which

may occur inside the system. Before calculating the entropy production in terms of quantities which characterize the irreversible phenomena, we shall rewrite Eqs. 35 and 36 in a form which is more suitable for the description of the systems in which the densities of the extensive properties (such as mass and energy considered in conservation laws) are continuous functions of spatial coordinates (Groot and Mazur 1962):

$$S = \int^{V} \rho s dV \tag{37}$$

$$\frac{dS_e}{dt} = -\int^{\Omega} \mathrm{J}_{\mathrm{S},tot} \cdot d\boldsymbol{\Omega} \tag{38}$$

$$\frac{dS_i}{dt} = \int^{V} \gamma dV \tag{39}$$

where s is the entropy per unit mass, $\mathrm{J}_{\mathrm{S},tot}$ is the total entropy flux which is a vector that coincides with the direction of entropy flow and has a magnitude equal to the entropy crossing unit area perpendicular to the direction of flow per unit time, and γ is the entropy source strength or entropy production per unit volume and unit time.

With Eqs. 37, 38, and 39, the formula (35) may be written, using Gauss's theorem, in the form (Groot and Mazur 1962)

$$\int^{V} \left(\frac{\partial \rho s}{\partial t} + \mathrm{div}\ \mathrm{J}_{\mathrm{S},tot} - \gamma \right) dV = 0 \tag{40}$$

where the divergence of $\mathrm{J}_{\mathrm{S},tot}$ simply represents the net entropy leaving unit volume per unit time. From this relation, it follows, since Eq. 40 must be hold for an arbitrary volume V, that

$$\frac{\partial \rho s}{\partial t} = -\mathrm{div}\ \mathrm{J}_{\mathrm{S},tot} + \gamma \tag{41}$$

$$\gamma \geq 0 \tag{42}$$

These two formulations are the local forms of Eqs. 35 and 36, i.e., the local mathematical expressions for the second law of thermodynamics. Equation 41 is formally a balance equation for the entropy density ρs with a source γ which satisfies the important inequality (42). With the help of Eqs. 7 and 41, it can be rewritten in a slightly different form:

$$\rho \frac{ds}{dt} = -\mathrm{div}\ \mathrm{J}_{\mathrm{S}} + \gamma \tag{43}$$

where the entropy flux J_S is the difference between the total entropy flux $J_{S,tot}$ and a convective term $\rho s\boldsymbol{\upsilon}$:

$$J_S = J_{S,tot} - \rho s\boldsymbol{\upsilon} \tag{44}$$

For application in continuum mechanics, we must relate the changes in the properties of the system to the rate of change in entropy, which will enable us to obtain more explicit expressions for the entropy flux J_S and the entropy source strength γ that appears in Eq. 43.

We postulate the existence of a thermodynamic potential from which the state laws can be derived. Without entering the details, let us say that the specification of a function with a scalar value, concave with respect to T and convex with respect to other variables, allows us to satisfy a priori the conditions of thermodynamic stability imposed by Clausius-Duhem inequality. Here we choose the specific Helmholtz free energy φ, which is defined as the difference between the specific internal energy density u and the product between the absolute temperature T and specific entropy s:

$$\varphi = u - Ts \tag{45}$$

Differentiating this and with the help of the law of conservation of energy, we have

$$\begin{aligned} d\varphi &= du - Tds - sdT \\ &= \delta q + \delta w - Tds - sdT \\ &= \delta q + \left(\delta w^d + \delta w^e\right) - Tds - sdT \\ &= \left(\delta q + \delta w^d - Tds\right) + \left(\delta w^e - sdT\right) \end{aligned}$$

where q is the total heat flowing into the system per unit mass, including the conduction through the surface and the distributed internal heat source; w is the total work done on the system per unit mass by external pressure and body force; w^d is the lost energy associated with the total work, which is generally dissipated in the form of heat; and w^e is the elastic energy associated with the total work. For the quantitative treatment of entropy for irreversible processes, let us introduce the definition of entropy for irreversible processes:

$$ds = \frac{\delta q + \delta w^d}{T} \tag{46}$$

With the help of Eq. 46, we have

$$Tds = du - dw^e \tag{47}$$

This is the Gibbs relation which combines the first and second laws. From the definition of the entropy, we also have

$$dw^e = d\varphi + sdT \tag{48}$$

The Helmholtz free energy is the isothermal recoverable elastic energy. It should be pointed out that the specific elastic energy w^e, namely, the work stored in the system per unit mass during a process, is path independent. The elastic energy is the maximum amount of work that could be produced by a device between any given two states. If the device is work absorbing, the elastic energy work of the process is the minimum amount of work that must be supplied (Li 1989).

In order to find the explicit form of the entropy balance Eq. 43, we insert the expressions (34) for du/dt into Eq. 47 with the time derivatives given by Eq. 4:

$$\rho\frac{ds}{dt} = -\frac{\text{div J}_\text{q}}{T} + \frac{1}{T}\boldsymbol{\sigma} : \mathbf{D} - \frac{\rho}{T}\frac{dw^e}{dt} + \frac{\rho r}{T} \tag{49}$$

Noting that

$$\frac{\text{div J}_\text{q}}{T} = \text{div}\frac{\text{J}_\text{q}}{T} + \frac{1}{T^2}\text{J}_\text{q} \cdot \text{grad } T$$

it is easy to cast Eq. 49 into the form of a balance Eq. 43:

$$\rho\frac{ds}{dt} = -\text{div}\frac{\text{J}_\text{q}}{T} - \frac{1}{T^2}\text{J}_\text{q} \cdot \text{grad } T + \frac{1}{T}\boldsymbol{\sigma} : \mathbf{D} - \frac{\rho}{T}\frac{dw^e}{dt} + \frac{\rho r}{T} \tag{50}$$

From comparison with Eq. 43, it follows that the expressions for the entropy flux and the entropy production rate are given by

$$\text{J}_\text{S} = \frac{\text{J}_\text{q}}{T} \tag{51}$$

$$\gamma = \frac{1}{T}\boldsymbol{\sigma} : \mathbf{D} - \frac{\rho}{T}\frac{dw^e}{dt} - \frac{1}{T^2}\text{J}_\text{q} \cdot \text{grad } T + \frac{\rho r}{T} \tag{52}$$

Equation 51 shows that for the closed systems, the entropy flow consists of only one part: the "reduced" heat flow J_q/T. Equation 52 represents the entropy production by the internal dissipations. The sum of the first two terms is called the intrinsic dissipation or mechanical dissipation. It consists of plastic dissipation plus the dissipation associated with the evolution of other internal variables; it is generally dissipated by the volume element in the form of heat. The last two terms are the thermal dissipation due to the conduction of heat and the internal heat source. The structure of the expression for γ is that of a bilinear form: it consists of a sum of products of two factors. One of these factors in each term is a flow quantity (heat flow J_q, momentum flow or pressure tensor $\boldsymbol{\sigma}$) already introduced in the conservation of laws. The other factor in each term is related to a gradient of an intensive state variable (gradients of temperature and velocity). These quantities which multiply the fluxes in the expression for the entropy production are called thermodynamic forces.

The way in which the separation of the right-hand side of Eq. 49 into the divergence of a flux and a source term has been achieved may at first sight seem to be to some extent arbitrary. The two parts of Eq. 50 must, however, satisfy a number of requirements which determine this separation uniquely (Groot and Mazur 1962). The entropy source strength γ must be zero if the thermodynamic equilibrium conditions are satisfied within the system. Another requirement which Eq. 52 must satisfy is that it be invariant under the transformation of different reference frames, since the notions of reversible and irreversible behavior must be invariant under such a transformation. It can be seen that Eq. 52 satisfies these requirements. Finally, it may be noted that Eq. 50 also satisfies the Clausius-Duhem inequality:

$$\boldsymbol{\sigma} : \mathbf{D} - \rho\left(\frac{d\varphi}{dt} + s\frac{dT}{dt}\right) - \mathrm{J_q} \cdot \frac{\mathrm{grad}\, T}{T} \geq 0 \tag{53}$$

Between two particles of a solid body which are at different temperatures, heat is transferred only by conduction, a process which takes place at the molecular and atomic levels. The law of heat conduction for isotropic bodies may be stated as follows (Boley and Weiner 1988):

$$\mathrm{J_q} = -k\mathrm{grad}\, T \tag{54}$$

where k, with typical units of $Btu/ft \cdot hr \cdot {}^{\circ}F$, is termed the thermal conductivity of the solid and where $\mathrm{J_q}$ is the heat flux.

This law of heat conduction was stated first by Fourier who based it on experimental observation. Fourier's law expresses a linear relation between the heat flux vector $\mathrm{J_q}$ and its dual variable grad T. Since solid, opaque bodies are of primary interest here, heat is transferred from point to point within this body solely by conduction. The field equation of the boundary-value problem will, therefore, always be some form of the Fourier heat conduction equation. Of course, heat may be transferred to the surface of the body by other modes of heat transfer which correspond to various thermal boundary conditions.

Then the expression for the internal entropy production can be simplified as

$$\gamma = \frac{1}{T}\boldsymbol{\sigma} : \mathbf{D} - \frac{\rho}{T}\frac{dw_{ava}}{dt} + \frac{k}{T^2}|\mathrm{grad}\, T|^2 + \frac{\rho r}{T} \tag{55}$$

Fully Coupled Thermomechanical Equations

The formalism of continuum mechanics and thermodynamics requires the existence of a certain number of state variables. We limit ourselves to two observable variables – the temperature T and the total strain $\boldsymbol{\varepsilon}$ – as they are the only ones which occur in elasticity. For dissipative phenomena the current state also depends on the past history and path followed. Plasticity and viscoplasticity require the

introduction of the plastic (or viscoplastic) strain $\boldsymbol{\varepsilon}^p$ as a variable. Other phenomena, such as hardening, damage, and fracture, require the introduction of other internal variables of less obvious in nature. These variables represent the internal state of matter (density of dislocations, crystalline of microstructures, configuration of microcracks and cavities, etc.) (Lemaitre and Chaboche 1990). There is no objective way to choose the nature of the internal variables best suited to the study of a phenomenon. For general study, here these variables will be denoted by $V_k (k = 1, 2 \ldots)$ representing either a scalar or a tensorial variable.

For small strains, the plastic strain is the permanent strain associated with the relaxed configuration which is obtained by elastic unloading, leading to the additive strain decomposition:

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^e + \boldsymbol{\varepsilon}^p \tag{56}$$

The relations existing between the energy, stress tensor, and strain tensor can be obtained using the formalism of thermodynamics with internal variables. Here we choose the specific Helmholtz free energy φ, which depends on observable variables and internal variables:

$$\varphi = (\boldsymbol{\varepsilon}, T, \boldsymbol{\varepsilon}^e, \boldsymbol{\varepsilon}^p, V_k) \tag{57}$$

For small strains, the strains appear only in the form of their additive decomposition, so that

$$\varphi((\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^p), T, V_k) = \varphi(\boldsymbol{\varepsilon}^e, T, V_k) \tag{58}$$

which shows that (Lemaitre and Chaboche 1990)

$$\frac{\partial \varphi}{\partial \boldsymbol{\varepsilon}^e} = \frac{\partial \varphi}{\partial \boldsymbol{\varepsilon}} = -\frac{\partial \varphi}{\partial \boldsymbol{\varepsilon}^p} \tag{59}$$

and the following expressions define the thermodynamic laws (Lemaitre and Chaboche 1990):

$$\boldsymbol{\sigma} = \rho \frac{\partial \varphi}{\partial \boldsymbol{\varepsilon}^e} \tag{60}$$

$$s = -\frac{\partial \varphi}{\partial T} \tag{61}$$

$$A_k = \rho \frac{\partial \varphi}{\partial V_k} \tag{62}$$

where A_k is the thermodynamic force associated with the internal variables V_k, s, $\boldsymbol{\sigma}$ and A_k constitute the associated variables. The vector formed by the variables is the gradient of the function φ in the space of the variables T, $\boldsymbol{\varepsilon}^e$, and V_k. This vector is normal to the surface $\varphi =$ constant.

The equation of the conservation of energy for small strains (from Eqs. 20 and 31) can be written as

$$\rho\dot{u} = -\text{div } \mathrm{J_q} + \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} + \rho r \tag{63}$$

and replace $\rho\dot{u}$ by the expression derived from Eq. 45

$$\rho\dot{u} = \rho\dot{\varphi} + \rho\dot{s}T + \rho s\dot{T} \tag{64}$$

and $\dot{\varphi}$ and $\dot{s}$ by its expression as a function of the state variables with the help of Eqs. 60, 61, and 62:

$$\dot{\varphi} = \frac{\partial\varphi}{\partial\boldsymbol{\varepsilon}^{\mathbf{e}}} : \dot{\boldsymbol{\varepsilon}}^e + \frac{\partial\varphi}{\partial T}\dot{T} + \frac{\partial\varphi}{\partial V_k}\dot{V}_k = \frac{1}{\rho}\boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^e - s\dot{T} + A_k\dot{V}_k \tag{65}$$

$$\dot{s} = -\frac{\partial^2\varphi}{\partial\boldsymbol{\varepsilon}^{\mathbf{e}}\partial T} : \dot{\boldsymbol{\varepsilon}}^e - \frac{\partial^2\varphi}{\partial T^2}\dot{T} - \frac{\partial^2\varphi}{\partial V_k\partial T}\dot{V}_k = -\frac{1}{\rho}\frac{\partial\boldsymbol{\sigma}}{\partial T} : \dot{\boldsymbol{\varepsilon}}^e + \frac{\partial s}{\partial T}\dot{T} - \frac{1}{\rho}\frac{\partial A_k}{\partial T}\dot{V}_k \tag{66}$$

We obtain

$$-\text{div } \mathrm{J_q} = \rho T\frac{\partial s}{\partial T}\dot{T} - \boldsymbol{\sigma} : (\dot{\boldsymbol{\varepsilon}} - \dot{\boldsymbol{\varepsilon}}^e) + A_k V_k - \rho r - T\left(\frac{\partial\boldsymbol{\sigma}}{\partial T} : \dot{\boldsymbol{\varepsilon}}^e + \frac{\partial A_k}{\partial T}\dot{V}_k\right) \tag{67}$$

by introducing the specific heat defined by

$$C = T\frac{\partial s}{\partial T} \tag{68}$$

and taking into account Fourier's law for isotropic materials

$$\text{div } \mathrm{J_q} = -k\,\text{div}(\text{grad } T) = -k\nabla^2 T \tag{69}$$

where ∇^2 denotes the Laplacian operator.

We obtain, using $\dot{\boldsymbol{\varepsilon}}^p = \dot{\boldsymbol{\varepsilon}} - \dot{\boldsymbol{\varepsilon}}^e$,

$$k\nabla^2 T = \rho C\dot{T} - \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^p + A_k V_k - \rho r - T\left(\frac{\partial\boldsymbol{\sigma}}{\partial T} : \dot{\boldsymbol{\varepsilon}}^e + \frac{\partial A_k}{\partial T}\dot{V}_k\right) \tag{70}$$

This is the fully coupled thermomechanical equation, which can simulate the evolution of temperature influenced by the mechanical work with properly imposed boundary conditions. $A_k\dot{V}_k$ represents the non-recoverable energy stored in the materials corresponding to other dissipated phenomena such as hardening, damage, and fracture. It represents only 5–10 % of the term $\boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^p$ and is often negligible (Lemaitre 1992; Lemaitre and Chaboche 1990; Chaboche and Lesne 1988):

$$A_k\dot{V}_k \approx 0 \tag{71}$$

which results in the fully coupled elastoplastic thermomechanical equation

$$k\nabla^2 T = \rho C\dot{T} - \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^p - \rho r - T\frac{\partial \boldsymbol{\sigma}}{\partial T} : \dot{\boldsymbol{\varepsilon}}^e \tag{72}$$

which has been used to simulate the thermal effects on the material behavior by many researchers (Sluzalec et al. 1988; Hong 1999). Equation 72 also allows us to calculate heat flux J_q generated due to elastic and/or inelastic work in a solid body.

For the isotropic linear thermoelastic materials, the stress-strain relationship is

$$\sigma_{ij} = \lambda\delta_{ij}\varepsilon_{kk} + 2\mu\varepsilon_{ij} - (3\lambda + 2\mu)\delta_{ij}\alpha(T - T_0) \tag{73}$$

where T_0 is the reference temperature, α is the isotropic thermal expansion coefficient, and λ and μ are the Lame's coefficients:

$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)};\ \mu = \frac{E}{2(1+\nu)} \tag{74}$$

If the internal generation of heat created by external sources is neglected, Eq. 70 becomes for isotropic linear thermoelastic materials

$$k\nabla^2 T = \rho C\dot{T} + (3\lambda + 2\mu)\alpha T\dot{\varepsilon}_{kk} \tag{75}$$

The last term represents the interconvertibility of the thermal and mechanical energy.

Thermodynamic Damage Evolution Function

Changes are always irreversible processes because of friction, which results in the production of entropy and thus a permanent change in the universe (Dehoff 1993). Damage is the progressive deterioration which occurs in materials prior to failure. Cumulative damage analysis plays a key role in the life prediction of components and structures subjected to load histories. As a result, many damage models have been proposed in the literature, such as linear damage models, nonlinear damage models, linear elastic fracture mechanics models, continuum damage mechanics models, and energy-based damage models (Bazant 1991; Bonora and Newaz 1998; Chaboche 1981, 1988; Chow and Chen 1992; Ju 1989, 1990; Kachanov 1986, 1986; Krajcinovic 1989; Lemaitre 1992; Murakami 1988; Murakami and Kamiya 1997; Rabotnov 1969a; Shi and Voyiadjis 1997; Voyiadjis and Thiagarajan 1996). Damage evolution function based on thermodynamics and statistical mechanics was first introduced in the literature by Basaran and Yan (1998), who established a relationship between entropy and damage for solids undergoing plastic deformations. Yet their model is not general enough to account for elastic deformations and to relate entropy production with material stiffness degradation. Basaran et al. (2003); Gomez and Basaran (2005); Li and Basaran (2009); Gunel and Basaran (2011); Yao and Basaran (2013); and Basaran et al. (2004) developed a generalized version of this damage evolution model.

There are many metrics to measure degradation in materials, such as direct measurements of the total crack areas lying on a surface, degradation of the elastic modulus, degradation of ultrasonic wave propagation speed, degradation of the microhardness, change in density, increase in electrical resistance, variation in the cyclic plastic response, change in creep properties, change in acoustic emission properties, remaining life, and cumulative hysteresis dissipation. The damage process corresponding to the degradation of microstructure is, in general, irreversible. During the cumulative damage process, the internal entropy production, which is a measure of disorder in the system, must increase according to the second law of thermodynamics, so that internal entropy production can be used as a criterion for quantification of damage. Entropy in statistical physics and in the thermodynamic sense is really the same thing (Malvern 1969). The statistical physics interpretation in terms of probability and tendency toward disordered microstates furnishes a physical significance for the otherwise rather abstract thermodynamic concept of entropy. Boltzmann (1898) first used statistical mechanics to give a precise meaning to disorder and established a connection between disorder and entropy for the whole system:

$$S = k_0 \ln W \tag{76}$$

where k_0 is Boltzmann constant and W is the disorder parameter which is the probability that the system will exist in the state relative to all the possible states it could be in. Statistical mechanics assigns an exact meaning to the probability of a state and supplies a general expression for W that employs the idea of the distribution function of a system. This function measures the probability for the coordinates and velocities of the molecules of the system to have specified values at a given time. The relation between the entropy per unit mass and the disorder parameter is given by Basaran and Yan (1998):

$$s = \frac{R}{m_s} \ln W \tag{77}$$

where s is the entropy per unit mass, m_s is the specific mass, and R is the gas constant.

According to Eq. 77 we have the disorder function as follows:

$$W = e^{\frac{sm_s}{R}} \tag{78}$$

Select an initial reference state of the continuous medium with disorder W_0; then the change in disorder at any arbitrary time with respect to the initial reference state is given by

$$\Delta W = W - W_0 = e^{\frac{sm_s}{R}} - e^{\frac{s_0 m_s}{R}} \tag{79}$$

where s_0 is the entropy for the initial reference state. According to Basaran and Nie (2004), the isotropic damage variable D is defined as the ratio of the change in

disorder parameter to the current state disorder parameter with a proportional critical disorder coefficient D_{cr}:

$$D = D_{cr} \frac{\Delta W}{W} \tag{80}$$

where D_{cr} allows us to correlate the value of entropy production-based damage D with other material coordinates, such as degradation of material stiffness. D_{cr} is easy to determine from experimental data, but it can vary for different loading profiles such as monotonic and cyclic loading.

With the help of Eqs. 78 and 79, the relation between the damage parameter D and the change in entropy can be written as

$$D = D_{cr}\left[1 - e^{-\frac{m_s}{R}(s - s_0)}\right] \tag{81}$$

With the help of Eqs. 49 and 70, the rate of total specific entropy change under the condition of small strains is given by

$$\frac{ds}{dt} = \frac{c}{T}\frac{\partial T}{\partial t} - \frac{1}{\rho}\left(\frac{\partial \boldsymbol{\sigma}}{\partial T} : \dot{\boldsymbol{\varepsilon}}^e + \frac{\partial A_k}{\partial T} : V_k\right) \tag{82}$$

where the identity $\boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} - \rho \dot{w}^e = \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^p - A_k \dot{V}_k$ is used, which represents the total mechanical dissipation rate. Equation 82 can also be obtained by its expression as a function of the state variables the same as Eq. 66.

With the help of Eq. 55, the specific entropy production rate for small strains becomes

$$\frac{ds_i}{dt} = \frac{\gamma}{\rho} = \frac{\boldsymbol{\sigma} : \dot{\varepsilon}^p}{T\rho} + \frac{k}{T^2\rho}|\mathrm{grad}T|^2 + \frac{r}{T} \tag{83}$$

where the identity $\boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} - \rho \dot{w}^e = \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^p - A_k \dot{V}_k$ is also used and $A_k \dot{V}_k$ is omitted.

The fundamental equations governing the temperature, stresses, deformation, and the entropy production rate in a continuum medium have been derived in the previous section. From a strict viewpoint, these quantities are all interrelated and must be determined simultaneously. However, for most practical problems, the effect of the stresses and deformations upon the temperature distribution is quite small and can be neglected. This procedure allows the determination of the temperature distribution in the solid resulting from prescribed thermal conditions to become the first and independent step of a thermal-stress analysis; the second step of such an analysis is then the determination of the stresses, deformations, and damage in the body due to this temperature distribution (Boley and Weiner 1988). Because the entropy change caused by the heat transfer between systems and surroundings has no influence on the degradation of the materials, only the entropy source strength, namely, the entropy created in the system, should be used as a basis for the systematic description of the irreversible processes.

So the damage evolution Eq. 73 can be implemented in a numerical analysis procedure, where

$$\Delta s = \Delta s_i = \int_{t_0}^{t} \frac{\boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^p}{T\rho} dt + \int_{t}^{t_0} \left(\frac{k}{T^2 \rho} |\text{grad} T|^2 \right) dt + \int_{t_0}^{t} \frac{r}{T} dt \tag{84}$$

According to Eq. 84, it is obvious that $D \geq 0$ is always satisfied because of the nonnegative entropy source strength. $D = 0$ when $\Delta s = 0$ and $D = D_{cr}$ when $\Delta s \to \infty$. Equation 84 shows that the damage is a function not only of the loading or straining process but also of the temperature. However, a uniform increase in temperature in a stress-free field does not cause any damage.

The advantage of the proposed procedure is quite obvious. Instead of separate formulations of constitutive and damage evolution equations in the continuum damage mechanics (CDM) theory, a unified description of CDM is possible only by establishing the constitutive relationship. So the whole problem of modeling the damage phenomenon lies in the determination of the analytical expressions for the constitutive relationship and its identification in characteristic experiments. The specific functional form of the constitutive relationship depends on the damage mechanism and the deformation itself. Moreover, this unified approach nullifies the need for a damage potential surface.

Damage Evolution and Entropy Production Under Electrical Current

When there is no mechanical load (in the form of a surface traction or a point load) acting on the system, according to Eq. 55 irreversible entropy rate production would be zero. However, if only electrical current is present and if the electrical current constitutes high current density, it will lead to significant irreversible entropy production. In the following section, this special case is formulated.

Modeling Electromigration Process in Thermodynamics

Electromigration is an electron flow-assisted diffusion process that takes place in conductor solids under high current density. The process can be assumed to be controlled by a vacancy diffusion mechanism, in which the diffusion takes place by vacancies switching lattice sites with adjacent atoms. In isothermal condition, the process is driven by electrical current-caused mass diffusion, stress gradient-induced diffusion, and diffusion due to atomic vacancy concentration. In the presence of electrical current, due to electrical resistance, there is always heat production (joule heating) that leads to thermomigration, which interacts with other diffusive forces. Under the presence of these four forces, the atomic vacancy flux equation can be given by combining Huntington [14] and Kirchheim [16] flux

definitions, adding the influence of temperature gradient and vacancy concentration, which yields

$$\frac{\partial C_v}{\partial t} = -\vec{\Delta} \cdot \vec{q} + G \tag{85}$$

$$\vec{q} = -D_v \left[\vec{\nabla} C_v + \frac{C_v Z^* e}{kT} \left(-\rho \vec{j} \right) - \frac{C_v}{kT} (-f\Omega) \vec{\nabla} \sigma + \frac{C_v Q^*}{kT} \frac{\vec{\nabla} T}{T} \right] \tag{86}$$

Combining these two equations yields

$$\frac{\partial C_v}{\partial t} = D_v \left[\nabla^2 C_v - \frac{Z^* e \rho}{kT} \vec{\nabla} \bullet \left(C_v \vec{j} \right) + \frac{f\Omega}{kT} \vec{\nabla} \bullet \left(C_v \vec{\nabla} \sigma \right) + \frac{Q^*}{kT^2} \vec{\nabla} \left(C_v \nabla T \right) \right] + G \tag{87}$$

where

C_v, vacancy concentration
D_v, vacancy diffusivity

$\vec{q}$, vacancy flux vector

Z^*, vacancy effective charge number
e, electron charge
ρ, metal resistivity

$\vec{j}$, current density vector

f, vacancy relaxation ratio, the ratio of the volume of an atom and the volume of a vacancy
Ω, atomic volume
Q^* heat of transport, the isothermal heat transmitted by moving the atom in the process of jumping a lattice site less the intrinsic enthalpy
k, Boltzmann's constant
T, absolute temperature
σ = trace(σ_{ij})/3, hydrostatic or spherical part of the stress tensor, Sarychev and Zhinikov [23]
G, vacancy generation rate,

$$G = -\frac{C_v - C_{ve}}{\tau_s}, \tag{88a}$$

C_{ev}, thermodynamic equilibrium vacancy concentration

$$C_{ve} = C_{v0} e^{\frac{(1-f)\Omega\sigma}{kT}} \tag{88b}$$

C_{v0}, equilibrium vacancy concentration in the absence of stress
τ_s, characteristic vacancy generation/annihilation time

If we define $C \equiv C_v/C_{v0}$ as the normalized concentration, then the vacancy diffusion equation could be rewritten as

$$\frac{\partial C_v}{\partial t} = D_v\left[\nabla^2 C - \frac{Z^* e\rho}{kT}\vec{\nabla}\bullet\left(C\vec{j}\right) + \frac{f\Omega}{kT}\vec{\nabla}\bullet\left(C\,\vec{\nabla}\,\sigma\right) + \frac{Q^*}{kT^2}\vec{\nabla}\left(C\vec{\nabla}\,T\right)\right] + \frac{G}{C_{vo}} \tag{89}$$

where, initially, $C = 1$(or $C_v = C_{v0}$).

Using the conservation of energy equation in the form derived in previous sections, we can write the rate of change of entropy density as follows:

$$\begin{aligned}\rho\frac{ds}{dt} = &\frac{1}{T} - \mathrm{div}\mathrm{J_q} + \sigma : \mathbf{Grad}(\mathbf{v}) \\ &+ \sum_k \mathrm{J_k}\bullet\mathrm{F_k} - \rho\left(\frac{\partial\Psi}{\partial\varepsilon^e} : \frac{d\varepsilon^e}{dt} + \frac{\partial\Psi}{\partial T} : \frac{dT}{dt} + \frac{\partial\Psi}{\partial V_k} : \frac{dV_k}{dt} - s\frac{dT}{dt}\right)\end{aligned} \tag{90}$$

where

$$\sigma : \mathbf{Grad}(\mathbf{v}) = \sigma : (\mathbf{D} + \mathbf{W}) \tag{91}$$

and **D** (symmetric) and **W** (skew symmetric) are the rate of deformation tensor and spin tensor, respectively.

Due to the symmetry of σ,

$$\sigma : (D + W) = \sigma : D \tag{92}$$

For small deformation, we can make the following assumption:

$$\sigma : \mathbf{D} = \sigma : \frac{d\varepsilon}{dt} = \sigma : \left(\frac{d\varepsilon^e}{dt} + \frac{d\varepsilon^p}{dt}\right) \tag{93}$$

Rearranging Eq. 90 and comparing with Eq. 27, we can get $\mathrm{J_s} = \frac{1}{T}\mathrm{J_q}$ and the following entropy production rate term:

$$\begin{aligned}\gamma = &-\frac{1}{T^2}\mathrm{J_q}\bullet\mathrm{Grad}(T) + \frac{1}{T}\sum_k \mathrm{J_k}\bullet\mathrm{F_k} + \frac{1}{T}\sigma : \frac{d\varepsilon^p}{dt} + \frac{1}{T}\left(\sigma : \frac{d\varepsilon^e}{dt}, -\rho\frac{\partial\Psi}{\partial\varepsilon^e}, : \frac{d\varepsilon^e}{dt}\right) \\ &+ \frac{\rho}{T}\left(s + \frac{\partial\Psi}{\partial T}\right)\frac{dT}{t} - \frac{\rho}{T}\frac{\partial\Psi}{\partial V_k} : \frac{dV_k}{dt}\end{aligned} \tag{94}$$

In solids with internal friction, all deformations cause positive entropy production rate $\gamma \geq 0$ (which is also referred to as the Clausius-Duhem inequality, Malvern [22]).

Using the following relations,

$$\sigma = \rho(\partial\Psi/\partial\varepsilon^e) \tag{95}$$

$$s = -\frac{\partial\Psi}{\partial T} \tag{96}$$

we can simplify Eq. 94 as follows:

$$\gamma = -\frac{1}{T^2}\mathrm{J_q}\bullet\mathrm{Grad}(T) + \frac{1}{T}\sum_k \mathrm{J_k}\bullet\mathrm{F_k} + \frac{1}{T}\sigma : \dot{\varepsilon}^p - \frac{\rho}{T}\frac{\partial\Psi}{\partial V_k} : \frac{dV_k}{dt} \tag{97}$$

or if the heat flux term $\mathrm{J_q}$ is replaced by

$$\mathrm{J_q} = \frac{1}{T^2}\mathrm{C}|\mathrm{Grad}(T)|^2 \tag{98}$$

where C is the thermal conductivity tensor, the entropy production rate can be given by

$$\gamma = -\frac{1}{T^2}\mathrm{C}|\mathrm{Grad}(T)|^2 + \frac{1}{T}\sum_k \mathrm{J_k}\bullet\mathrm{F_k} + \frac{1}{T}\sigma : \dot{\varepsilon}^p - \frac{\rho}{T}\frac{\partial\Psi}{\partial V_k} : \frac{dV_k}{dt} \tag{99}$$

We identify $\mathrm{J_k}$ in Eq. 99 as $\vec{q}$ in Eq. 86 and the effective driving force terms $\mathrm{F_k}$ as

$$\mathrm{F_k} = \left[Z^* ej\rho + (-f\Omega)\vec{\nabla}\,\sigma - \frac{Q}{T}\;\vec{\nabla}\,T - \frac{kT}{\mathrm{C}}\vec{\nabla}\,\mathrm{C}\right] \tag{100}$$

From Eq. 99, we can see that the irreversible dissipation includes two parts: the first term is called heat dissipation caused by conduction inside the system, while the second, third, and fourth terms account for other irreversible processes in the system; we will call it intrinsic dissipation.

According to second law of thermodynamics, all systems must fail when entropy is at maximum and the production rate is at minimum. Based on this law, entropy production rate given in the form shown in Eq. 99 allows accounting for electromigration-induced entropy production in the system. Because electromigration is an irreversible process and leads to failure of the system, we assume that it must fall within the laws of thermodynamics. Entropy production rate given in Eq. 99 could be written in many different forms, but this form of this equation is the most appropriate for our purpose.

With the help of Eqs. 99 and 100, we can write Eq. 16 as

$$\Delta s = \int_{t_0}^{t} \left(\frac{1}{T^2} \mathrm{C} |\mathrm{Grad}(T)|^2 + \frac{C_v D_{\text{effective}}}{kT^2} \left(Z_l^* e \rho j - f \Omega \nabla \sigma + \frac{Q \vec{\nabla} T}{T} + \frac{kT}{\mathrm{C}} \vec{\nabla} \mathrm{C} \right)^2 + \frac{1}{T} \sigma : \varepsilon^p - \frac{\rho}{T} \frac{\partial \Psi}{\partial V_k} : \frac{dV_k}{dt} \right) dt \tag{101}$$

If we only consider the damage caused by electrical driving forces and disregard other factors such as stress gradient, temperature gradient, and atomic vacancy concentration, the damage evolution formula can be given by

$$D = D_{cr} \left[1 - e^{\frac{\int_{t_0}^{t} -\frac{C_v D_{\text{effective}}}{kT^2} (Z^* e \rho j)^2 dt}{N_0 k}} \right] \tag{102}$$

If we assume that the solder joint has failed when the degradation reaches some critical value (e.g., it is 5 % drop in electrical resistance of a solder joint in the US microelectronics industry), defined as D_{cr}, the time required to reach the failure can be obtained from Eq. 102 as

$$t = \frac{N_0 T^2 k^2}{C_v D_{\text{effective}} (Z^* e \kappa j)^2} \ln \frac{1}{(1 - D_{cr})} \tag{103}$$

From Eq. 102, we observe that the dependence on temperature is to the power of two. One may wonder why the time to failure is larger when the temperature is larger. The fact is that the diffusivity is going to change with temperature too. The relationship between the temperature and the diffusivity is represented by Arrhenius function,

$$D_{\text{effective}} = D_0 e^{-\frac{Q}{kT}} \tag{104}$$

Inserting Eq. 104 into Eq. 102, we can get

$$t = \frac{N_0 T^2 k^2}{N_l D_0 \left(Z_l^* e \kappa j \right)^2} \left(\ln \frac{1}{(1 - D_{cr})} \right) e^{\frac{Q}{kT}} \tag{105}$$

The last term (exponent) dominates the temperature effect so the time to failure decreases with temperature as expected.

Damage-Coupled Viscoplasticity

The damage-coupled material constitutive model is ideally suited to characterize the mutual interaction between the macro-level mechanical properties and material damage due to microstructure degradation. The damage variable can be directly used as the fatigue damage criterion for the numerical model, from which the number of cycles to failure can be determined. Furthermore, the damage distribution and progressive damage evolution can be obtained.

Effective Stress and Strain Equivalence Principle

Lemaitre and Chaboche (1990) considered a certain section of the representative volume element (RVE) under uniaxial force **F** as shown in Fig. 1, where δS is the initial area of the undamaged section and δS_D denotes the lost area as result of damage. $\delta S - \delta S_D$ can be interpreted as the actual area of the section. The values δS and δS_D are to be understood in the sense of appropriate averaging.

The nominal stress can be defined as

$$\boldsymbol{\sigma} = \frac{\mathbf{F}}{\delta S} \tag{106}$$

Rabotnov (1969) introduced the concept of effective stress $\widetilde{\boldsymbol{\sigma}}$, which relates to the surface that effectively resists the load, namely, $(\delta S - \delta S_D)$,

$$\widetilde{\boldsymbol{\sigma}} = \frac{\mathbf{F}}{\delta S - \delta S_D} \tag{107}$$

Kachanov (1986) gave the definition of damage at microscale as the creation of microsurface discontinuities: breaking of atomic bonds and plastic enlargement of microcavities. This isotropic damage variable is defined as

$$D = \frac{\delta S_D}{\delta S} \tag{108}$$

So, we have

$$\widetilde{\boldsymbol{\sigma}} = \frac{\boldsymbol{\sigma}}{1 - D} \tag{109}$$

Lemaitre and Chaboche (1990) assumed that the strain response of the body is modified by damage only through the actual stress. Thus, the stress-strain behavior of the damaged material can be represented by the constitutive equation of the virgin material (without damage) with the stress in it replaced by the effective stress. And this is the strain equivalence principle: "Any strain constitutive equation for a damaged material may be derived in the same way as for a virgin material except that the usual stress is replaced by the effective stress."

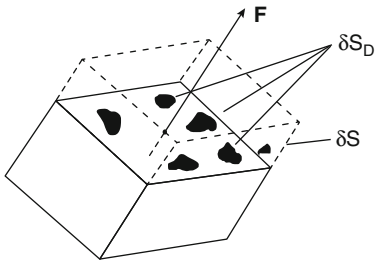


Fig. 1 Schematic illustration of definition of damage (Lemaitre and Chaboche 1990)

The strain equivalent principle demonstrates the major role of the concept of actual stress. According to the strain equivalence principle, the elastic strain of a damaged material is

$$\boldsymbol{\varepsilon}^e = \frac{\widetilde{\boldsymbol{\sigma}}}{\tilde{E}} = \frac{1}{E}\frac{\boldsymbol{\sigma}}{1-D} \tag{110}$$

So, the Hook's law here has its usual form with the Young's modulus E being replaced by $\tilde{E}$, where $\tilde{E}$ is the Young's modulus associated with the damaged status.

In the case of an elastic-plastic deformation when the damage is a result of a large strain, it is natural to assume that damage does not depend on the elastic strain; hence

$$\frac{dD}{d\boldsymbol{\varepsilon}^e} = 0 \tag{111}$$

This condition leads to the relation

$$D = 1 - \frac{\tilde{E}}{E} \tag{112}$$

Hence, the damage may be estimated by measuring the elastic response. Note that $\tilde{E}$ can be identified with the unloading modulus. So the change in modulus with strain was considered by the damage parameter D, which reflects the decrease in load carrying ability of particles in particulate composites as they crack or debond.

Damage-Coupled Isotropic Viscoplasticity

In order to simulate damage behavior of solid materials, there is a need for a progressive constitutive degradation model. Damage mechanics provides basic framework to develop damage evolution models at small strains.

Damage-Coupled Constitutive Equations

In accordance with the strain equivalence principle and Hook's law, the elasticity constitutive relationship may be written as

$$d\boldsymbol{\sigma} = (1 - D)\mathbf{C}^e d\boldsymbol{\varepsilon}^e \tag{113}$$

where $d\boldsymbol{\varepsilon}^e$ is the elastic strain increment vector, $\mathbf{C}^e$ is the elastic constitutive matrix, $d\boldsymbol{\sigma}$ is the total stress increment vector, and D is the isotropic damage parameter.

Assuming deformations of small strain, the total strain increment can be separated into three components,

$$\{d\boldsymbol{\varepsilon}\} = \left\{d\boldsymbol{\varepsilon}^{th}\right\} + \{d\boldsymbol{\varepsilon}^e\} + \{d\boldsymbol{\varepsilon}^{vp}\} \tag{114}$$

where $d\boldsymbol{\varepsilon}^{th}$, $d\boldsymbol{\varepsilon}^e$, and $d\boldsymbol{\varepsilon}^{vp}$ are the incremental thermal, elastic, and viscoplastic strains, respectively.

The thermal strain increment is

$$d\boldsymbol{\varepsilon}^{th} = \alpha_T dT\mathbf{I} \tag{115}$$

where α_T is the coefficient of thermal expansion, dT is the temperature increment, and $\mathbf{I}$ is the second-order unit tensor. $\{d\boldsymbol{\varepsilon}^{vp}\}$ is viscoplastic strain increment, which can be determined using the viscoplasticity theory.

So the total stress increment can be obtained by

$$d\boldsymbol{\sigma} = (1 - D)\mathbf{C}^e\left(d\boldsymbol{\varepsilon} - d\boldsymbol{\varepsilon}^{vp} - d\boldsymbol{\varepsilon}^{th}\right) \tag{116}$$

Damage-Coupled Yield Surface

Von Mises-type yield surface with isotropic and kinematic hardening is used in the constitutive model here. Von Mises criterion states that the viscoplastic strain is governed by the elastic shear (or deviatoric) energy density, the so-called J_2 theory:

$$F = q - \overline{\sigma} \tag{117}$$

where the Mises equivalent stress q is

$$q = \sqrt{\frac{3}{2}(\mathbf{S} - \boldsymbol{\alpha}) : (\mathbf{S} - \boldsymbol{\alpha})} \tag{118}$$

$\mathbf{S}$ is the deviatoric stress tensor defined by

$$\mathbf{S} = \boldsymbol{\sigma} - \frac{1}{3}\sigma_{kk}\mathbf{I} \tag{119}$$

$\boldsymbol{\alpha}$ is the deviatoric stress component of back stress tensor corresponding to kinematic hardening.

$\overline{\sigma}$ is the equivalent yield stress:

$$\overline{\sigma} = R + \sigma_y \tag{120}$$

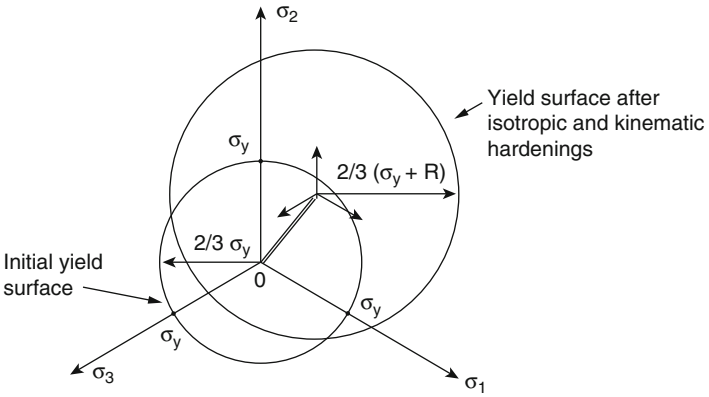


Fig. 2 Yield locus in the principal stress space (Lemaitre and Chaboche 1990)

R is the evolution of the size of the yield surface corresponding to isotropic hardening; σ_y is the initial size of the yield surface. Figure 2 shows the yield surface in the principal stress space.

The kinematic hardening represents an approximation of accounting for the Bauschinger effect. The corresponding deviatoric back stress represents the translation of the center of the yield surface in the deviatoric space. Isotropic hardening stress R measures the increase in radius of the yield cylinder in the stress space.

Damage has significant effects on the yield surface. In order to model the behavior of a damaged material, the strain equivalence principle is needed. According to the strain equivalence principle, the normal stress is replaced by the effective stress in the yield function, and all other variables remain unchanged. So the yield function has the same form with the modified Mises equivalent stress:

$$q^* = \sqrt{\frac{\mathbf{3}}{\mathbf{2}}\left(\frac{\mathbf{S}}{1-D} - \boldsymbol{\alpha}\right) : \left(\frac{\mathbf{S}}{1-D} - \boldsymbol{\alpha}\right)} \tag{121}$$

If the nominal stress σ is used, instead of the effective stress $\widetilde{\sigma}$, the yield surface can be written as

$$F = q^d - \overline{\sigma}^d \tag{122}$$

with the damage-coupled back stress, Mises equivalent stress, and yield stress as follows:

$$\boldsymbol{\alpha}^d = (1-D)\boldsymbol{\alpha} \tag{123}$$

$$\overline{\sigma}^d = (1 - D)\overline{\sigma} \tag{124}$$

$$q^d = \sqrt{\frac{3}{2}(\mathbf{S} - \boldsymbol{\alpha}^d) : (\mathbf{S} - \boldsymbol{\alpha}^d)} \tag{125}$$

It shows that the damage equally decreases the yield stress, the isotropic strain hardening stress, and the back stress.

It is emphasized that the fully coupled elastoplastic thermomechanical Eq. 70, the damage evolution function (81), the entropy production Eq. 84, and the constitutive Eq. 116 completely characterize the progressive damage behavior of any material.

Examples

Damage-coupled plasticity is ideal for the damage-coupled plasticity calculations, either as a rate-dependent or as a rate-independent model. Damage-coupled linear kinematic hardening plasticity and damage-coupled isotropic hardening plasticity have particularly simple forms. Because of this simplicity, the algebraic equations associated with integrating this model are easily developed in terms of a single variable, and the material stiffness matrix can be written explicitly. For simplicity of notation, all quantities not explicitly associated with a time point are assumed to be evaluated at the end of the increment.

Damage-Coupled Plasticity with Isotropic Hardening

The Mises yield function with associated flow means that there is no volumetric plastic strain, and since the elastic bulk modulus is quite large, the volume change will be small, so that we can define the volume strain as

$$\varepsilon_{vol} = \mathbf{I} : \boldsymbol{\varepsilon} = \text{trace}(\boldsymbol{\varepsilon}) \tag{126}$$

where $\mathbf{I}$ is the second-order unit tensor.

The deviatoric strain is

$$\mathbf{e} = \boldsymbol{\varepsilon} - \frac{1}{3}\varepsilon_{vol}\mathbf{I} = \boldsymbol{\varepsilon} - \frac{1}{3}\mathbf{II} : \boldsymbol{\varepsilon} = \left(\mathbf{\Pi} - \frac{1}{3}\mathbf{I} \otimes \mathbf{I}\right) : \boldsymbol{\varepsilon} \tag{127}$$

where $\mathbf{\Pi}$ is the fourth-order unit tensor.

The equivalent pressure stress

$$p = -\frac{1}{3}\text{trace}(\boldsymbol{\sigma}) = -\frac{1}{3}\mathbf{I} : \boldsymbol{\sigma} \tag{128}$$

The deviatoric stress is defined as

$$\mathbf{S} = \boldsymbol{\sigma} + p\mathbf{I} \tag{129}$$

The Mises equivalent stress

$$q = \sqrt{\frac{3}{2}(\mathbf{S} : \mathbf{S})} \tag{130}$$

where

$$\frac{3}{2}(\mathbf{S} : \mathbf{S}) = \frac{1}{2}\left[(\sigma_{11} - \sigma_{22})^2 + (\sigma_{22} - \sigma_{33})^2 + (\sigma_{33} - \sigma_{11})^2 + 6\left(\sigma_{12}^2 + \sigma_{23}^2 + \sigma_{13}^2\right)\right] \tag{131}$$

Strain-rate decomposition is

$$d\boldsymbol{\varepsilon} = d\boldsymbol{\varepsilon}^{el} + d\boldsymbol{\varepsilon}^{pl} \tag{132}$$

Using the standard definition of corotational measures, this can be written in the integrated form as

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{el} + \boldsymbol{\varepsilon}^{pl} \tag{133}$$

The elasticity is linear and isotropic and may therefore be written in terms of two temperature-dependent material parameters bulk modulus K and shear modulus G, which are readily computed from Young's modulus E and Poisson ratio v as

$$K = \frac{E}{3(1 - 2v)} \tag{134}$$

and

$$G = \frac{E}{2(1 + v)} \tag{135}$$

The elasticity coupled with isotropic damage may then be written in volumetric and deviatoric components as follows:

$$p = -(1 - D)K\varepsilon_{vol} = -(1 - D)K\mathbf{I} : \boldsymbol{\varepsilon} \tag{136}$$

$$\mathbf{S} = 2(1 - D)G\mathbf{e}^{el} \tag{137}$$

The flow rule is

$$d\mathbf{e}^{pl} = d\overline{e}^{pl}\mathbf{n} \tag{138}$$

where

$$\mathbf{n} = \frac{3}{2}\frac{\mathbf{S}}{q} \tag{139}$$

$d\overline{e}^{pl}$ is the scalar equivalent plastic-strain rate. The evolution of the equivalent plastic strain is obtained from the following equivalent plastic work expression:

$$\sigma_y \dot{\overline{\varepsilon}}^{pl} = \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^{pl} \tag{140}$$

which yields

$$\dot{\overline{\varepsilon}}^{pl} = \sqrt{\frac{2}{3}\dot{\varepsilon}^{pl} : \dot{\varepsilon}^{pl}} \tag{141}$$

for isotropic Mises plasticity. Using Eqs. 138 and 139, we can prove this result as follows:

$$\sigma_y \dot{\overline{\varepsilon}}^{pl} = \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^{pl} = (\mathbf{S} - p\mathbf{I}) : \dot{\boldsymbol{\varepsilon}}^{pl} = \mathbf{S} : \dot{\boldsymbol{\varepsilon}}^{pl}$$

Then,

$$\dot{\overline{\varepsilon}}^{pl} = \frac{\mathbf{S}}{\sigma_y} : \dot{\boldsymbol{\varepsilon}}^{pl} = \frac{2}{3}\mathbf{n} : \dot{\boldsymbol{\varepsilon}}^{pl} = \frac{2}{3}\frac{\dot{\boldsymbol{\varepsilon}}^{pl}}{\dot{\overline{e}}^{pl}} : \dot{\boldsymbol{\varepsilon}}^{pl} = \frac{2}{3}\frac{\dot{\boldsymbol{\varepsilon}}^{pl}}{\dot{\overline{\varepsilon}}^{pl}} : \dot{\boldsymbol{\varepsilon}}^{pl}$$

Therefore, we have

$$\dot{\overline{\varepsilon}}^{pl} = \sqrt{\frac{2}{3}\dot{\boldsymbol{\varepsilon}}^{pl} : \dot{\boldsymbol{\varepsilon}}^{pl}}$$

The plasticity requires that the material satisfy a uniaxial-stress, plastic-strain, and strain-rate relationship. If the material is rate independent, this is the yield condition:

$$q = (1 - D)\sigma^0\left(\overline{e}^{pl}, T\right) \tag{142}$$

If the material is rate dependent, the relationship is the uniaxial flow rate definition:

$$\dot{\overline{e}}^{pl} = h\left(\frac{q}{1 - D}, \overline{e}^{pl}, T\right) \tag{143}$$

where h is a known function.

Integrating this relationship by the backward Euler method gives

$$\Delta\overline{e}^{pl} = h\left(\frac{q}{1 - D}, \overline{e}^{pl}, T\right)\Delta t \tag{144}$$

The backward Euler method is an unconditional stable and second-order accurate algorithm. This equation can be inverted (numerically, if necessary) to give q as a function of $\overline{e}^{pl}$ at the end of the increment. Thus both the rate-independent model and the integrated rate-dependent model give the general uniaxial form

$$q = (1 - D)\overline{\sigma}\left(\overline{e}^{pl}, T\right) \tag{145}$$

where $\overline{\sigma} = \sigma^0$ for the rate-independent model and $\overline{\sigma}$ is obtained by the inversion of Eq. 144 for the rate-dependent model.

In any increment when plastic flow is occurring (which is determined by evaluating q based on purely elastic response and finding that its value exceeds $(1 - D)\overline{\sigma}$), these equations must be integrated and solved for the state at the end of the increment. The integration is done by applying the backward Euler method to the flow rule (138), giving

$$\Delta \mathbf{e}^{pl} = \Delta \overline{e}^{pl} \mathbf{n} \tag{146}$$

Combining this with the deviatoric elasticity (137) and the integrated strain-rate decomposition (133) gives

$$\mathbf{S} = 2(1 - D)G\left(\mathbf{e}^{el}\big|_t + \Delta \mathbf{e} - \Delta \overline{e}^{pl} \mathbf{n}\right) \tag{147}$$

Then using the integrated flow rule (146), together with the Mises definition of the flow direction $\mathbf{n}$, this becomes

$$\left(1 + 3(1 - D)\frac{G}{q}\Delta \overline{e}^{pl}\right)\mathbf{S} = \mathbf{S}^{pr} \tag{148}$$

where

$$\mathbf{S}^{pr} = 2(1 - D)G\hat{\mathbf{e}} \tag{149}$$

$$\hat{\mathbf{e}} = \mathbf{e}^{el}\big|_t + \Delta \mathbf{e} \tag{150}$$

With the help of Eq. 130, taking the inner product of Eq. 148 with itself gives

$$q + 3(1 - D)G\Delta \overline{e}^{pl} = q^{pr} \tag{151}$$

where

$$q^{pr} = \sqrt{\frac{3}{2}\left(\mathbf{S}^{pr} : \mathbf{S}^{pr}\right)} \tag{152}$$

where q^{pr} is the elastic predictor based on purely elastic behavior.

The Mises equivalent stress q must satisfy the uniaxial form defined in Eq. 145, so that, from Eq. 151,

$$q^{pr} - 3(1 - D)G\Delta \overline{e}^{pl} - (1 - D)\overline{\sigma} = 0 \tag{153}$$

This is a nonlinear equation for $\Delta \overline{e}^{pl}$ in the general case, which can be solved iteratively by Newton's method ($x_{n+1} = x_n + f(x_n)/f'(x_n)$):

$$c^{pl}|_k = \frac{q^{pr} - 3(1-D)G\Delta\overline{e}^{pl}|_k - (1-D)\overline{\sigma}|_k}{(1+D)(3G+H)} \tag{154}$$

$$\Delta\overline{e}^{pl}\big|_{k+1} = \Delta\overline{e}^{pl}\big|_k + c^{pl}|_k \tag{155}$$

where

$$H = (1-D)\frac{\partial\overline{\sigma}}{\partial\overline{e}^{pl}}|_k \tag{156}$$

and we iterate until convergence is achieved. Once $\Delta\overline{e}^{pl}$ is known, the solution is fully defined.

Using Eq. 145,

$$q = (1-D)\overline{\sigma}\left(\overline{e}^{pl}, T\right) \tag{157}$$

Using Eqs. 148 and 151,

$$\mathbf{S} = \frac{\mathbf{S}^{pr}}{1 + 3(1-D)\dfrac{G}{q}\Delta\overline{e}^{pl}} = \frac{q}{q^{pr}}\mathbf{S}^{pr} \tag{158}$$

Using Eq. 139,

$$\mathbf{n} = \frac{3}{2}\frac{\mathbf{S}}{q} \tag{159}$$

Using Eq. 146,

$$\Delta\mathbf{e}^{pl} = \Delta\overline{e}^{pl}\mathbf{n} \tag{160}$$

For cases where three direct strain components are provided by the kinematic solution, Eq. 136 defines

$$p = -(1-D)K\varepsilon_{vol} = -\frac{1}{3}\sigma_{kk}^{pr} \tag{161}$$

so that the solution is then fully defined, and the material stiffness matrix can be derived without the need for matrix inversion as follows. Taking the variation of Eq. 158 with respect to all quantities at the end of the increment gives

$$\partial\mathbf{S} = q\partial\left(\frac{\mathbf{S}^{pr}}{q^{pr}}\right) + \frac{\mathbf{S}}{q}\partial q \tag{162}$$

From Eq. 149,

$$\partial \mathbf{S}^{pr} = 2(1-D)G\partial\hat{\mathbf{e}} \tag{163}$$

And from Eq. 152,

$$\partial q^{pr} = 3(1-D)G\frac{\mathbf{S}^{pr}}{q^{pr}} : \partial\hat{\mathbf{e}} = \frac{3(1-D)G}{q}\mathbf{S} : \partial\hat{\mathbf{e}} \tag{164}$$

where Eqs. 158 and 163 have been used.

So,

$$\partial\left(\frac{\mathbf{S}^{pr}}{q^{pr}}\right) = \frac{q^{pr}\partial\mathbf{S}^{pr} - \mathbf{S}^{pr}\partial q^{pr}}{(q^{pr})^2} = \frac{2(1-D)G}{q^{pr}}\partial\hat{\mathbf{e}} - \frac{3(1-D)G}{q^2 q^{pr}}\mathbf{S}\otimes\mathbf{S} : \partial\hat{\mathbf{e}} \tag{165}$$

where Eqs. 158 and 164 have been used.

From Eq. 145,

$$\partial q = (1-D)H\partial\overline{e}^{pl} \tag{166}$$

and from Eq. 151,

$$\partial q + 3(1-D)G\partial\overline{e}^{pl} = \partial q^{pr} \tag{167}$$

Combining Eqs. 166, 167, and 164,

$$\partial\overline{e}^{pl} = \frac{\partial q^{pr}}{(3G+H)} = \frac{3G}{(3G+H)q}\mathbf{S} : \partial\hat{\mathbf{e}} \tag{168}$$

So,

$$\partial q = \frac{3(1-D)GH}{(3G+H)q}\mathbf{S} : \partial\hat{\mathbf{e}} \tag{169}$$

Combining these results with Eq. 162 gives

$$\partial\mathbf{S} = 2(1-D)G\frac{q}{q^{pr}}\partial\hat{\mathbf{e}} - \frac{3(1-D)G}{q^2}\left[\frac{q}{q^{pr}} - \frac{H}{(3G+H)}\right]\mathbf{S}\otimes\mathbf{S} : \partial\hat{\mathbf{e}} \tag{170}$$

So we have

$$\partial\mathbf{S} = [Q\mathbf{\Pi} - R\mathbf{S}\otimes\mathbf{S}] : \partial\hat{\mathbf{e}} \tag{171}$$

where

$$Q = 2(1-D)G\frac{q}{q^{pr}} \tag{172}$$

$$R = \frac{3(1-D)G}{q^2}\left[\frac{q}{q^{pr}} - \frac{H}{3G+H}\right] = \frac{9(1-D)G^2\left(q-(1-D)H\Delta\overline{e}^{pl}\right)}{q^2 q^{pr}(3G+H)} \tag{173}$$

For all cases where three direct strains are defined by the kinematic solution, the material stiffness is completed by (from Eq. 136)

$$\partial p = -(1-D)K\mathbf{I} : \partial\boldsymbol{\varepsilon} \tag{174}$$

so that (from Eq. 129)

$$\boldsymbol{\sigma} = \mathbf{S} - p\mathbf{I} \tag{175}$$

and (from Eq. 127)

$$\partial\hat{\mathbf{e}} = \partial\mathbf{e} = \left(\boldsymbol{\Pi} - \frac{1}{3}\mathbf{I}\otimes\mathbf{I}\right) : \partial\boldsymbol{\varepsilon} \tag{176}$$

Because

$$\mathbf{e} = \hat{\mathbf{e}} + \mathbf{e}^{pl}\big|_t \tag{177}$$

we have

$$\partial\boldsymbol{\sigma} = \left[\left((1-D)K - \frac{1}{3}Q\right)\mathbf{I}\otimes\mathbf{I} + Q\boldsymbol{\Pi} - R\mathbf{S}\otimes\mathbf{S}\right] : \partial\boldsymbol{\varepsilon} \tag{178}$$

Equation 178 can also be written in an alternate form:

$$\Delta\dot{\sigma}_{ij} = \lambda^*\delta_{ij}\Delta\dot{\varepsilon}_{kk} + 2\mu^*\Delta\dot{\varepsilon}_{ij} + \left(\frac{(1-D)H}{1+H/3G} - 3\mu^*\right)\eta_{ij}\eta_{kl}\Delta\dot{\varepsilon}_{kl} \tag{179}$$

where

$$\mu^* = (1-D)Gq/q^{pr} \tag{180}$$

$$\lambda^* = (1-D)K - \frac{2}{3}\mu^* \tag{181}$$

$$\eta_{ij} = S_{ij}^{pr}/q^{pr} \tag{182}$$

Using Eqs. 158, 161, and 175, we have

$$\sigma_{ij} = \eta_{ij}q + \frac{1}{3}\delta_{ij}\sigma_{kk}^{pr} \tag{183}$$

Damage-Coupled Plasticity with Linear Kinematic Hardening

The Mises equivalent stress is defined as

$$q = \sqrt{\frac{3}{2}(\mathbf{S} - \boldsymbol{\alpha}) : (\mathbf{S} - \boldsymbol{\alpha})} \tag{184}$$

where $\boldsymbol{\alpha}$ is the back stress, which gives the center of the yield surface in deviatoric stress space.

The flow rule is

$$d\mathbf{e}^{pl} = d\overline{e}^{pl}\mathbf{n} \tag{185}$$

where

$$\mathbf{n} = \frac{3}{2}\frac{(\mathbf{S} - \boldsymbol{\alpha})}{q} \tag{186}$$

and $d\overline{e}^{pl}$ is the scalar equivalent plastic-strain rate.

Prager-Ziegler (linear) kinematic hardening model is

$$d\boldsymbol{\alpha} = \frac{2}{3}(1 - D)Cd\mathbf{e}^{pl} = \frac{2}{3}(1 - D)Cd\overline{e}^{pl}\mathbf{n} \tag{187}$$

The plasticity requires that the material satisfy a uniaxial-stress, plastic-strain, and strain-rate relationship:

$$q = (1 - D)\sigma_y \tag{188}$$

In any increment when plastic flow is occurring (which is determined by evaluating q based on the purely elastic response and finding that its value exceeds $(1 - D)\sigma_y$), these equations must be integrated and solved for the state at the end of the increment. The integration is done by applying the backward Euler method to the flow rule (185) and hardening law (187), giving

$$\Delta\mathbf{e}^{pl} = \Delta\overline{e}^{pl}\mathbf{n} \tag{189}$$

$$\Delta\boldsymbol{\alpha} = \frac{2}{3}(1 - D)C\Delta\mathbf{e}^{pl} = \frac{2}{3}(1 - D)C\Delta\overline{e}^{pl}\mathbf{n} \tag{190}$$

The backward Euler method is an unconditional stable and second-order accurate algorithm. Combining Eq. 189 with the deviatoric elasticity (137) and the integrated strain-rate decomposition (133) gives

$$\mathbf{S} = 2(1 - D)G\left(\mathbf{e}^{el}\big|_t + \Delta\mathbf{e} - \Delta\overline{e}^{pl}\mathbf{n}\right) \tag{191}$$

Rearranging Eq. 191, together with subtracting $\boldsymbol{\alpha}|_t$ from both sides, this becomes

$$(\mathbf{S}-\boldsymbol{\alpha})+\Delta\boldsymbol{\alpha}+2(1-D)G\Delta\bar{e}^{pl}\mathbf{n}=2(1-D)G\hat{\mathbf{e}}-\boldsymbol{\alpha}|_t \tag{192}$$

where

$$\hat{\mathbf{e}}=\mathbf{e}^{el}|_t+\Delta\mathbf{e} \tag{193}$$

$$\boldsymbol{\alpha}|_t=\boldsymbol{\alpha}-\Delta\boldsymbol{\alpha} \tag{194}$$

Then using the integrated hardening law together with the Mises definition of the flow direction $\mathbf{n}$, this becomes

$$\left(1+(1-D)\frac{3G+C}{q}\Delta\bar{e}^{pl}\right)(\mathbf{S}-\boldsymbol{\alpha})=\mathbf{S}^{pr}-\boldsymbol{\alpha}|_t \tag{195}$$

where

$$\mathbf{S}^{pr}=2(1-D)G\hat{\mathbf{e}} \tag{196}$$

With the help of Eq. 184, taking the inner product of Eq. 195 with itself gives

$$q+(1-D)(3G+C)\Delta\bar{e}^{pl}=q^{pr} \tag{197}$$

where q^{pr} is the elastic predictor based on purely elastic behavior:

$$q^{pr}=\sqrt{\frac{3}{2}(\mathbf{S}^{pr}-\boldsymbol{\alpha}|_t):(\mathbf{S}^{pr}-\boldsymbol{\alpha}|_t)} \tag{198}$$

The Mises equivalent stress q must satisfy the uniaxial form defined in Eq. 188, so that, from Eq. 197,

$$q^{pr}-(1-D)(3G+C)\Delta\bar{e}^{pl}-(1-D)\sigma_y=0 \tag{199}$$

where $\Delta\bar{e}^{pl}$ can be solved by

$$\Delta\bar{e}^{pl}=\frac{q^{pr}-(1-D)\sigma_y}{(1-D)(3G+C)} \tag{200}$$

Once $\Delta\bar{e}^{pl}$ is known, the solution is fully defined.
Using Eq. 188,

$$q=(1-D)\sigma_y \tag{201}$$

Using Eqs. 195, 197, and 201,

$$\mathbf{S}-\boldsymbol{\alpha}=\frac{\mathbf{S}^{pr}-\boldsymbol{\alpha}|_t}{1+\dfrac{3G+C}{\sigma_y}\Delta\bar{e}^{pl}}=(1-D)\frac{\sigma_y}{q^{pr}}(\mathbf{S}^{pr}-\boldsymbol{\alpha}|_t) \tag{202}$$

Using Eqs. 186 and 202,

$$\mathbf{n} = \frac{3}{2}\frac{\mathbf{S}-\boldsymbol{\alpha}}{(1-D)\sigma_y} = \frac{3}{2}\frac{\mathbf{S}^{pr}-\boldsymbol{\alpha}|_t}{q^{pr}} \tag{203}$$

Using Eq. 189,

$$\Delta\mathbf{e}^{pl} = \Delta\bar{e}^{pl}\mathbf{n} \tag{204}$$

Using Eq. 190,

$$\Delta\boldsymbol{\alpha} = \frac{2}{3}(1-D)C\Delta\mathbf{e}^{pl} \tag{205}$$

So, the back stress and deviatoric stress tensor can be determined as

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}|_t + \Delta\boldsymbol{\alpha} \tag{206}$$

$$\mathbf{S} = (1-D)\frac{\sigma_y}{q^{pr}}(\mathbf{S}^{pr}-\boldsymbol{\alpha}|_t) + \boldsymbol{\alpha} \tag{206}$$

For cases where three direct strain components are provided by the kinematic solution, Eq. 136 defines

$$p = -(1-D)K\varepsilon_{vol} = -(1-D)K\mathbf{I}:\boldsymbol{\varepsilon} \tag{207}$$

so that the solution is then fully defined and the material stiffness matrix can be derived without the need for matrix inversion as follows. Taking the variation of Eq. 197 with respect to all quantities at the end of the increment gives

$$\partial\mathbf{S} = (1-D)\sigma_y\partial\left(\frac{\mathbf{S}^{pr}-\boldsymbol{\alpha}|_t}{q^{pr}}\right) + \partial\boldsymbol{\alpha} \tag{208}$$

From Eq. 196,

$$\partial\mathbf{S}^{pr} = 2(1-D)G\partial\hat{\mathbf{e}} \tag{209}$$

and from Eq. 198,

$$\partial q^{pr} = 3(1-D)G\frac{(\mathbf{S}^{pr}-\boldsymbol{\alpha}|_t)}{q^{pr}}:\partial\hat{\mathbf{e}} = \frac{3G}{\sigma_y}(\mathbf{S}-\boldsymbol{\alpha}):\partial\hat{\mathbf{e}} \tag{210}$$

where Eq. 206 has been used.

Then

$$\partial\left(\frac{\mathbf{S}^{pr}-\boldsymbol{\alpha}|_t}{q^{pr}}\right) = \frac{2(1-D)G}{q^{pr}}\partial\hat{\mathbf{e}} - \frac{3G}{(1-D)\sigma_y^2 q^{pr}}(\mathbf{S}-\boldsymbol{\alpha})\otimes(\mathbf{S}-\boldsymbol{\alpha}):\partial\hat{\mathbf{e}} \tag{211}$$

where Eq. 206 has been used.

From Eq. 200,

$$\partial\overline{e}^{pl} = \frac{\partial q^{pr}}{(1-D)(3G+C)} = \frac{3G}{(1-D)(3G+C)\sigma_y}(\mathbf{S}-\boldsymbol{\alpha}) : \partial\hat{\mathbf{e}} \tag{212}$$

From Eqs. 203 and 204,

$$\partial\mathbf{e}^{pl} = \frac{3}{2}\left[\frac{3G}{(1-D)^2(3G+C)\sigma_y^2}(\mathbf{S}-\boldsymbol{\alpha})\otimes(\mathbf{S}-\boldsymbol{\alpha}) : \partial\hat{\mathbf{e}} + \Delta\overline{e}^{pl}\partial\left(\frac{\mathbf{S}^{pr}-\boldsymbol{\alpha}|_t}{q^{pr}}\right)\right] \tag{213}$$

where Eqs. 206 and 212 have been used.

From Eqs. 205 and 213,

$$\begin{aligned}\partial\boldsymbol{\alpha} = &\frac{3GC}{(1-D)(3G+C)\sigma_y^2}(\mathbf{S}-\boldsymbol{\alpha})\otimes(\mathbf{S}-\boldsymbol{\alpha}) : \partial\hat{\mathbf{e}} \\ &+ (1-D)C\Delta\overline{e}^{pl}\partial\left(\frac{\mathbf{S}^{pr}-\boldsymbol{\alpha}|_t}{q^{pr}}\right)\end{aligned} \tag{214}$$

Combining these results with Eq. 208 gives

$$\begin{aligned}\partial\mathbf{S} = &\frac{2(1-D)^2G\left(\sigma_y + C\Delta\overline{e}^{pl}\right)}{q^{pr}}\partial\hat{\mathbf{e}} \\ &-\frac{3G}{\sigma_y^2}\left[\frac{\left(\sigma_y + C\Delta\overline{e}^{pl}\right)}{q^{pr}} - \frac{C}{(1-D)(3G+C)}\right](\mathbf{S}-\boldsymbol{\alpha})\otimes(\mathbf{S}-\boldsymbol{\alpha}) : \partial\hat{\mathbf{e}}\end{aligned} \tag{216}$$

So, we have

$$\partial\mathbf{S} = [Q\boldsymbol{\Pi} - R(\mathbf{S}-\boldsymbol{\alpha})\otimes(\mathbf{S}-\boldsymbol{\alpha})] : \partial\hat{\mathbf{e}} \tag{217}$$

where

$$Q = (1-D)^2\frac{2G\left(\sigma_y + C\Delta\overline{e}^{pl}\right)}{q^{pr}} \tag{218}$$

$$R = \frac{3G}{\sigma_y^2}\left[\frac{\sigma_y + C\Delta\overline{e}^{pl}}{q^{pr}} - \frac{C}{(1-D)(3G+C)}\right] = \frac{9G^2}{(3G+C)q^{pr}\sigma_y} \tag{219}$$

For all cases where three direct strains are defined by the kinematic solution, the material stiffness is completed by (from Eq. 136)

$$\partial p = -(1-D)K\mathbf{I} : \partial\boldsymbol{\varepsilon} \tag{220}$$

so that (from Eq. 129)

$$\boldsymbol{\sigma} = \mathbf{S} - p\mathbf{I} \tag{221}$$

and (from Eq. 127)

$$\partial\hat{\mathbf{e}} = \partial\mathbf{e} = \left(\mathbf{\Pi} - \frac{1}{3}\mathbf{I}\otimes\mathbf{I}\right) : \partial\boldsymbol{\varepsilon} \tag{222}$$

Because

$$\mathbf{e} = \hat{\mathbf{e}} + \mathbf{e}^{pl}\big|_t \tag{223}$$

we have

$$\partial\boldsymbol{\sigma} = \left[\left((1-D)K - \frac{1}{3}Q\right)\mathbf{I}\otimes\mathbf{I} + Q\mathbf{\Pi} - R(\mathbf{S}-\boldsymbol{\alpha})\otimes(\mathbf{S}-\boldsymbol{\alpha})\right] : \partial\boldsymbol{\varepsilon} \tag{224}$$

Equation 224 can also be written in alternate form:

$$\Delta\dot{\sigma}_{ij} = \lambda^*\delta_{ij}\Delta\dot{\varepsilon}_{kk} + 2\mu^*\Delta\dot{\varepsilon}_{ij} + \left(\frac{(1-D)C}{1+C/3G} - 3\mu^*\right)\eta_{ij}\eta_{kl}\Delta\dot{\varepsilon}_{kl} \tag{225}$$

where

$$\mu^* = (1-D)^2\frac{G\left(\sigma_y + C\Delta\bar{e}^{pl}\right)}{q^{pr}} \tag{226}$$

$$\lambda^* = (1-D)K - \frac{2}{3}\mu^* \tag{227}$$

$$\eta_{ij} = \left(S_{ij}^{pr} - \alpha_{ij}\big|_t\right)/q^{pr} \tag{228}$$

From Eqs. 196, 206, 207, and 221, we have

$$\sigma_{ij} = \alpha_{ij} + \eta_{ij}\sigma_y + \frac{1}{3}\delta_{ij}\sigma_{kk}^{pr} \tag{229}$$

Conclusions

In this chapter a thermodynamic theory of damage mechanics has been presented. This physics-based damage evolution theory avoids using phenomenological damage surfaces. Damage evolution is modeled by irreversible entropy production rate in the system. It is shown that only the irreversible portion of the entropy production contributes to damage in solids.

References

M.E. arychev, Zhinikov, General model for mechanical stress evolution during electromigration. Journal of Applied Physics, **86**(6), p.3068–3075 (1999)
C. Basaran, S. Nie, An irreversible thermodynamics theory for damage mechanics of solids. Int. J. Damage Mech. **13**(3), 205–223 (2004)
C. Basaran, C.Y. Yan, A thermodynamic framework for damage mechanics of solder joints. ASME J. Electron. Pack. **120**, 379–384 (1998)
C. Basaran, M. Lin, H. Ye, A thermodynamic model for electrical current induced damage. Int. J. Solids Struct. **40**(26), 7315–7327 (2003)
C. Basaran, H. Tang, S. Nie, Experimental damage mechanics of microelectronics solder joints under fatigue loading. Mech. Mater. **36**, 1111–1121 (2004)
Z.P. Bazant, Why continuum damage is nonlocal: micromechanics arguments. J. Eng. Mech. ASCE **117**(5), 1070–1087 (1991)
B.A. Boley, J.H. Weiner, *Theory of Thermal Stress* (Dover Publications, New York, 1988)
I. Boltzmann, *Lectures on Gas Theory* (University of California Press, Berkeley, 1898) (Translation by S. Brush, 1964)
N. Bonora, G.M. Newaz, Low cycle fatigue life estimation for ductile metals using a nonlinear continuum damage mechanics model. Int. J. Solids Struct. **35**(16), 1881–1894 (1998)
J.L. Chaboche, Continuum damage mechanics – a tool to describe phenomena before crack initiation. Nucl. Eng. Design **64**, 233–267 (1981)
J.L. Chaboche, Continuum damage mechanics: parts I & II. ASME J. Appl. Mech. **55**, 59–72 (1988)
J.L. Chaboche, P.M. Lesne, A non-linear continuous fatigue damage model. Fatigue Fract. Eng. Mater. Struct. **II**, 1–17 (1988)
S.R. De Groot, P. Mazur, *Non-equilibrium Thermodynamics* (North-Holland, Amsterdam, 1962)
R.T. DeHoff, *McGraw-Hill Series in Materials Science and Engineering: Thermodynamics in Materials Science* (McGraw-Hill, Boston, 1993)
J.L. Ericksen, *Introduction to the Thermodynamics of Solids*. Applied Mathematical Sciences, vol. 131 (Springer, New York, 1998)
P. Germain, Q.S. Nguyen, P. Suquet, Continuum thermodynamics. J. Appl. Mech. Trans. ASME **50**(4b), 1010–1020 (1983)
J. Gomez, C. Basaran, A thermodynamics based damage mechanics constitutive model for low cycle fatigue analysis of microelectronics solder joints incorporating size effect. Int. J. Solids Struct. **42**(13), 3744–3772 (2005)
E.M. Gunel, C. Basaran, Damage characterization in non-isothermal stretching of acrylics: part I theory. Mech. Mater. **43**(12), 979–991 (2011)
R. Haase, *Thermodynamics of Irreversible Processes* (Addison-Wesley, Reading, 1969)
B.Z. Hong, Analysis of thermomechanical interactions in a miniature solder system under cyclic fatigue loading. J. Electron. Mater. **28**(9), 1071–1077 (1999)
H.B. Huntington, A.R. Grone, Current-induced marker motion in gold wires. Journal of Physics and Chemistry of Solids, 1961. **20**(1–2)76–87
J.W. Ju, On energy-based coupled elastoplastic damage theories: constitutive modeling and computational aspects. Int. J. Solids Struct. **25**(7), 803–833 (1989)
J.W. Ju, Isotropic and anisotropic damage variables in continuum damage mechanics. J. Eng. Mech. **116**(12), 2764–2770 (1990)
L.M. Kachanov, *Introduction to Continuum Damage Mechanics* (Martinus Nijhoff, Boston, 1986)
R. Kircheim, Stress and Electromigration in Al-lines of Integrated Circuits" Acta Mettallurgica et Materilia. **40**(2), 309–323 (1992)
D. Krajcinovic, Damage mechanics. J. Mech. Mater. **8**, 117–197 (1989)
D. Krajcinovic, *North-Holland Series in Applied Mathematics and Mechanics*, Elsevier, Amsterdam (1996)
S. Li, C. Basaran, A computational damage mechanics model for thermomigration. Mech. Mater. **41**(3), 271–278 (2009)

J. Lemaitre, *A Course on Damage Mechanics* (Springer, Berlin, 1992)
J. Lemaitre, J.-L. Chaboche, *Mechanics of Solid Materials* (University Press, Cambridge, UK, 1990)
K.W. Li, *Applied Thermodynamics: Availability Method and Energy Conversion* (Taylor & Francis, New York, 1989)
L.E. Malvern, *Introduction to the Mechanics of a Continuous Medium* (Prentice-Hall, Englewood Cliffs, 1969)
S. Murakami, Mechanical modeling of material damage. ASME J. Appl. Mech. **55**, 280–286 (1988)
S. Murakami, K. Kamiya, Constitutive and damage evolution equations of elastic-brittle materials based on irreversible thermodynamics. Int. J. Solids Struct. **39**(4), 473–486 (1997)
Y.N. Rabotnov, *Creep Problems in Structural members* (North-Holland, Amsterdam, 1969a)
Y.N. Rabotnov, Fundamental problems in visco-plasticity, in *Recent Advances in Applied Mechanics* (Academic, New York, 1969b)
G.Y. Shi, G.Z. Voyiadjis, A new free energy for plastic damage analysis. Mech. Res. Commun. **24** (4), 377–383 (1997)
A. Sluzalec, An analysis of the thermal effects of coupled thermo-plasticity in metal forming processes. Commun. Appl. Numer. Methods **4**, 675–685 (1988)
G.Z. Voyiadjis, G. Thiagarajan, Cyclic anisotropic-plasticity model for metal matrix composites. Int. J. Plast. **12**(1), 69–91 (1996)
W. Yao, C. Basaran, Computational damage mechanics of electromigration and thermomigration. J. Appl. Phys. **114**, 103708 (2013)
W. Yourgrau, A.V.D. Merwe, G. Raw, *Treatise on Irreversible and Statistical Thermophysics: An Introduction to Nonclassical Thermodynamics* (Macmillan, New York, 1966)

Section VI

Damage Mechanics in Metal Forming

Damage Prediction in Metal Forming Process Modeling and Optimization: Simplified Approaches

24

Ying-Qiao Guo, Yuming Li, Boussad Abbès, Hakim Naceur, and Ali Halouani

Contents

Y.-Q. Guo (✉) • Y. Li • B. Abbès • A. Halouani
Université de Reims Champagne-Ardenne, GRESPI/MPSE, Faculté des Sciences Exactes et Naturelles, Reims Cedex 2, France
e-mail: yq.guo@univ-reims.fr; yuming.li@univ-reims.fr; boussad.abbes@univ-reims.fr; alihalouani@yahoo.fr

H. Naceur
Université Lille Nord de France, Laboratoire LAMIH UMR 8201 CNRS, Valenciennes Cedex 9, France
e-mail: hakim.naceur@univ-valenciennes.fr

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_41

Abstract

Some simplified numerical methods for damage predictions in metal forming process modeling and optimization are presented in this chapter. The incremental approaches including advanced damage models lead to accurate results, but the simulations are tedious and time-consuming. An efficient solving algorithm called *inverse approach* (IA) allows the fast modeling of forming processes in only one step between the known final part and the initial blank, avoiding the contact treatment and the incremental plastic integration. To improve the stress estimation in the IA, the so-called pseudo-inverse approach (PIA) has been developed. Some intermediate configurations are geometrically created and corrected by a free surface method to consider the deformation path, and the plastic integration based on the flow theory is carried out incrementally to consider the loading history. A simplified 3D strain-based damage model is coupled with the plasticity and implemented into a direct scalar integration algorithm of plasticity (without local iterations), which makes the plastic integration very fast and robust even for very large strain increments. These simplified approaches lead to very fast and useful numerical tools in the preliminary design and optimization.

Introduction

Nowadays, the forming industry needs to increase the product quality and to reduce the production costs and delay. The preliminary design of forming processes implies expensive trials-corrections on forming tools. The actual tendency is to use the numerical simulations in order to predict the forming feasibility (material flow, stresses, damage, etc.) and to optimize the process parameters and the tool geometry.

The numerical process modeling is a difficult task due to the involved complex phenomena: large strains, viscoplasticity, damage, contact-friction, thermal effects, etc. The incremental approaches with advanced damage models can give accurate results, but the simulation remains tedious and time-consuming; hence, it quickly becomes unfeasible in an optimization process. In this chapter, some simplified methods for damage modeling in metal forming process and optimization will be presented: (1) the fast forming algorithm IA which allows the calculation in only one step between the known final part and the initial blank, (2) the PIA which considers the loading path and improves significantly the stress estimation of the

IA, (3) a simplified 3D strain-based damage model and an efficient direct scalar integration algorithm of plasticity, and (4) the fast and robust forming process optimization using the above simplified approaches.

Two main approaches are widely used for the metal forming simulation: the incremental approach and the inverse approach (IA or one-step approach). The first one simulates the real multi-physic phenomena step by step, which makes it fairly accurate but very time-consuming. The second one exploits at maximum the knowledge of the final part shape; it performs the calculation from the known final configuration to the initial one in only one step to determine the strain and stress fields satisfying the equilibrium (Guo et al. 1990; Lee and Huh 1998, among others). The IA is based on two main assumptions: the tool actions are replaced by simplified nodal forces to avoid the contact-friction treatment, and the loading is supposed proportional to avoid the incremental plastic integration procedure. The strains and stresses are calculated by directly comparing the initial and final configurations. The IA is very fast and gives fairly good strain estimation. Nowadays, it is largely used as a valuable numerical tool in the preliminary design stages in various forming processes (stamping, hydroforming, etc.), in order to optimize the tool geometry and process parameters, such as the shape of initial metal sheet, the addendum surfaces, the drawbead sizes and positions, the holding forces, the springback compensation, etc. (Naceur et al. 2006; Dong et al. 2007; Azaouzi et al. 2008).

However, the IA cannot consider the loading history, leading to poor stress state estimation. The PIA has been developed to improve the stress estimation (Guo et al. 2004; Halouani et al. 2012a). Some intermediate configurations are geometrically created and mechanically corrected to take into account the deformation paths. The coupled damage–plasticity model is based on the flow theory of plasticity, and the plastic integration is carried out in an incremental form. The fast direct scalar algorithm of plasticity (DSAP) is used to speed up the procedure and avoid divergence problems in case of large strain increments. The PIA possesses not only the advantages of the IA (simple and fast), but also the advantages of the incremental approach (loading history, good stress estimation). Many research works based on the forward or backward methods have investigated the tool preform optimization (Kobayashi et al. 1989; Kim and Kobayashi 1990; Fourment et al. 1996), but too much computation time is required to carry out an optimization procedure. The PIA has been used for the automatic design and optimization of tool preforms. Genetic optimization algorithms and surrogate meta-models are adopted for the multi-objective optimization process in order to obtain the Pareto front (Halouani et al. 2012b).

Two main theories are extensively used to describe the ductile damage occurrence and its effect on the metal behavior. The first one was pioneered by Gurson (1977) and improved later by other researchers (Rousselier 1987, etc.). It is based on the micro-mechanisms of void nucleation, growth, and coalescence. It uses the void volume fraction as a "scalar" damage variable in the plastic potential in order to model the void effect on the plastic flow. The second one ignores the micro-defect mechanisms; it represents the damage effect on the overall elastoplastic

behavior of the material. The continuum damage mechanics (CDM) (Chaboche 1988; Lemaître and Chaboche 1990) uses a scalar or tensorial damage variables to represent the ductile defect evolution and their consequences on the other thermomechanical fields. This kind of CDM-based phenomenological approach has been widely applied to various metal forming and machining processes (Saanouni and Chaboche 2003).

There are two principal methods for the damage modeling: the uncoupled approach and the fully coupled one. The first one calculates the damage distribution with the stress and strain fields at the end of a FE analysis without taking into account its effect on other mechanical fields. It has been used by many authors to analyze damaged zones in the final workpiece (Hartley et al. 1989, among others). In some other works, the damage is used to find the forming limit strains in metal forming (Gelin et al. 1985; Cordebois and Ladevèze 1985). On the other hand, in the fully coupled approach, the damage effect is introduced directly into the constitutive equations, so it affects the other thermomechanical fields. The widely used models are based on the Gurson damage theory considering the void volume fraction evolution (Aravas 1986; Onate and Kleiber 1988; Picart et al. 1998). Other works are based on the continuum damage mechanics theory (CDM) (Lee et al. 1985; Zhu et al. 1992; Saanouni et al. 2000). A simplified approach based on Prandtl–Reuss plasticity model with nonlinear isotropic hardening is proposed in Mathur and Dawson (1987) and Brunet et al. (1996); the damage effect is taken into account by using the damage factor $(1 - D)$ on the stress vectors. Based on the thermodynamics of irreversible processes with state variables, an advanced approach aims to model the multi-physics coupling between the main thermomechanical phenomena including the isotropic and anisotropic damage (Mariage et al. 2002; Saanouni 2012, see also ▶ Chap. 25, "Ductile Damage in Metal Forming: Advanced Macroscopic Modeling and Numerical Simulation" of this volume).

In this chapter, a simplified model called 3D strain-based damage model (Lemaître and Chaboche 1990) and its applications in the IA and PIA will be presented. This ductile damage model is based on the dissipation potential and dedicated to the isotropic damage and hardening materials. The assumptions of the hardening saturation after the damage threshold and the constant triaxiality under proportional loading lead to a damage expression in terms of the equivalent plastic strain in a rate form, even in an integrated form. In the IA, an integrated constitutive equation is used to avoid the incremental plastic integration; the total damage expression is used to determine the damage distribution in the final workpiece (Cherouat et al. 2004). In the PIA, the damage effect is coupled with the plasticity by introducing the damage variable into the plasticity criterion; the damage rate expression is used to take account for their reciprocal effects (Guo et al. 2004).

In the PIA, the plastic strain increments are very large, so the classical iterative plastic integration based on the *return mapping algorithm* (Simo and Taylor 1986) requires much CPU time and may lead to divergence problems. A direct scalar algorithm of plasticity (DSAP) enables to directly performing the plastic integration without iterations (Li et al. 2007). The basic idea is to transform the unknown stress

vectors into the equivalent stresses which can be determined by using the tensile curve; then, the plastic multiplier λ can be directly calculated. The numerical results have shown very good agreement between the two algorithms, as well as the rapidity and robustness of the new DSAP.

Several examples will be presented to show the efficiency and limitations of the IA and PIA. The results are compared with those obtained by the classical incremental approach ABAQUS®/Explicit.

Inverse Approach for Sheet Forming Modeling

Basic Concept of the Inverse Approach

The IA has been proposed initially by Guo et al. (1990), for the estimation of the large elastoplastic strains encountered in the sheet forming process. The calculation is carried out from the known final workpiece to obtain the positions of material points in the initial blank (Fig. 1). Two main assumptions are adopted: the simplified tool action assumption to avoid the contact treatment and the proportional loading assumption to avoid the incremental plastic integration. This basic concept makes the IA very fast.

Since then, the IA has known numerous improvements, allowing the simulation of more complex 3D workpieces, including the friction, drawbeads, 3D anisotropy, rotation-free shell model, initial target solutions, spares solvers, etc. An important extension of the method has been done in order to deal with the tube hydroforming (Chebbah et al. 2011) and the cold forging (Halouani et al. 2010).

Formulation of the Inverse Approach

The IA formulation is presented below for the sheet forming and tube hydroforming by using a shell element. The FE formulation for axisymmetric cold forging can be found in Halouani et al. (2010).

The IA using the deformation theory of plasticity has different features comparing to the classical incremental approach: the known quantities are the final workpiece shape C, the FE mesh on C, and the thickness of the initial blank C^0, while the unknowns are the horizontal coordinates of the nodes in the initial

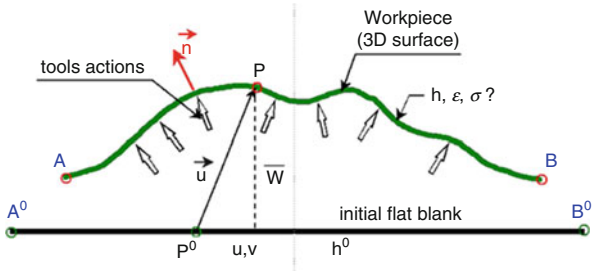


Fig. 1 Basic concept of the inverse approach

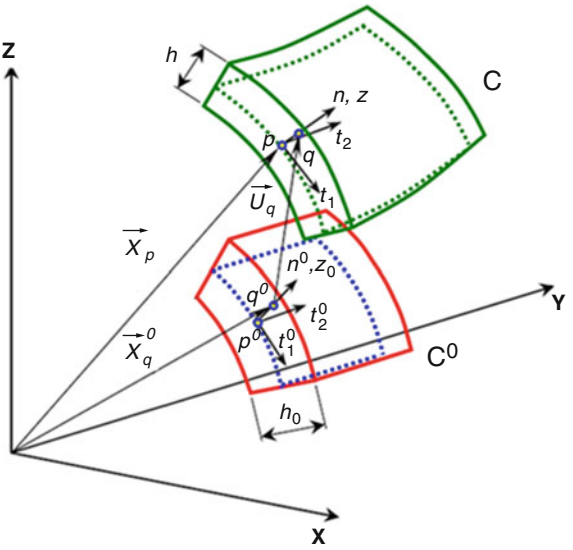


Fig. 2 Kinematics of 3D thin shell

flat blank and the thickness distribution in the final workpiece (Fig. 1). These unknowns are obtained by directly comparing C and C^0 in an iterative manner to satisfy the equilibrium conditions on C.

3D Shell Kinematics

Figure 2 illustrates the movement of a thin shell from its initial configuration C^0 to its final configuration C. p^0 is a material point on the shell mid-surface on C^0, q^0 is a point on the surface normal at p^0, and p and q are their final positions on C. Using the Kirchhoff assumption on the normal conservation, the following kinematic relations are obtained:

$$\boldsymbol{X}_{\boldsymbol{q}}^0 = \boldsymbol{X}_{\boldsymbol{p}}^0 + z^0\,\boldsymbol{n}^0 = \boldsymbol{X}_{\boldsymbol{p}} - \boldsymbol{U}_{\boldsymbol{p}} + \frac{z}{\lambda_3}\boldsymbol{n}^0; -\frac{h^0}{2} \le z^0 \le \frac{h^0}{2}; \quad z^0 = \frac{z}{\lambda_3} \tag{1}$$

$$\boldsymbol{X}_{\boldsymbol{q}} = \boldsymbol{X}_{\boldsymbol{p}} + z\,\boldsymbol{n}; -\frac{h}{2} \le z \le \frac{h}{2} \tag{2}$$

where $\boldsymbol{X}_p^0$ and $\boldsymbol{X}_p$ are the position vectors of p^0 and p, $\boldsymbol{U}_{\boldsymbol{p}}$ is the displacement vector between them, h^0 and h are the initial and final thicknesses, z^0 and z are the coordinates of q^0 and q through the thickness, λ_3 is the thickness stretch, and $\boldsymbol{n}^{\boldsymbol{0}}$ and $\boldsymbol{n}$ are the unit normal vectors at p^0 and p, respectively. It is noted that the final configuration is known and taken as reference.

Large Strain Measurement

Using two tangent vectors and the normal vector ($\boldsymbol{t_1}$, $\boldsymbol{t_2}$, $\boldsymbol{n}$) in p on the mid-surface of C, a local reference (x, y, z) is established:

$$\boldsymbol{t}_1 = \frac{\boldsymbol{X}_{p,x}}{\left\|\boldsymbol{X}_{p,x}\right\|} \quad ; \quad \boldsymbol{t}_2 = \frac{\boldsymbol{X}_{p,y}}{\left\|\boldsymbol{X}_{p,y}\right\|} \quad ; \quad \boldsymbol{n} = \boldsymbol{t}_1 \wedge \boldsymbol{t}_2 \tag{3}$$

The differentiation of Eqs. 1 and 2 gives the deformation gradient tensors with respect to the local reference (x, y, z):

$$d\boldsymbol{X}_q^0 = \left(\boldsymbol{F}_{\mathbf{x}}^0\right)^{-1} d\boldsymbol{x} \text{ with } \left(\boldsymbol{F}_{\mathbf{x}}^0\right)^{-1} = \left[\boldsymbol{X}_{p,x} - \boldsymbol{U}_{p,x} \vdots \boldsymbol{X}_{p,y} - \boldsymbol{U}_{p,y} \vdots \boldsymbol{n}^0/\lambda_3\right] \tag{4}$$

$$d\boldsymbol{X}_q = \boldsymbol{F}_{\mathbf{x}}\, d\boldsymbol{x} \text{ with } \boldsymbol{F}_{\mathbf{x}} = \left[\mathbf{X}_{p,x} + z\, \boldsymbol{n}_{,x} \vdots \boldsymbol{X}_{p,y} + z\, \boldsymbol{n}_{,y} \vdots \boldsymbol{n}\right] \tag{5}$$

The above two equations enable to obtain the inverse of deformation gradient tensor ($\boldsymbol{F}^{-1}$) between $d\boldsymbol{X}_q^0$ and $d\boldsymbol{X}_q$ and then the inverse of Cauchy–Green left tensor:

$$\boldsymbol{B}^{-1} = \boldsymbol{F}^{-T}\boldsymbol{F}^{-1} \tag{6}$$

The eigenvalues (λ_1^{-2}, λ_2^{-2}) of the tensor $\boldsymbol{B}^{-1}$ give the two in-plane principal stretches; the eigenvectors give their directions. The thickness elongation can be obtained by the incompressibility condition $\lambda_1\lambda_2\lambda_3 = 1$. The logarithmic principal strains in the final configuration C are given by

$$< \varepsilon_1\ \varepsilon_2\ \varepsilon_3 > = < \ln\lambda_1\ \ \ln\lambda_2\ \ \ln\lambda_3 > \tag{7}$$

Integrated Constitutive Law

In the IA, the proportional loading assumption is adopted to obtain an integrated constitutive law between the initial and final configurations. The loading history is ignored, leading to a total strain-stress law (see section "Simplified Plastic Ductile Damage Models and Direct Integration Algorithms" for more details):

$$\boldsymbol{\sigma} = \left[\boldsymbol{H}^{-1} + \left(\frac{1}{E_S} - \frac{1}{E}\right)\boldsymbol{P}\right]^{-1} \boldsymbol{\varepsilon} \tag{8}$$

where $\boldsymbol{\varepsilon}$ is the total strain vector, $\boldsymbol{H}$ is Hooke's elastic constitutive matrix, E_s is the secant modulus, E is Young's modulus, and $\boldsymbol{P}$ is the matrix defined by the von Mises isotropic criterion or Hill anisotropic criterion. The damage effect is uncoupled with the plasticity and evaluated as post-processing.

A non-quadratic anisotropic yield surface was proposed by Barlat et al. (2003) to deal with aluminum-based material parts:

$$f = \Phi' + \Phi'' - 2\overline{\sigma}_f^m = \left|X_1' - X_2'\right|^m + \left|2X_2'' - X_1''\right|^m + \left|2X_1'' - X_2''\right|^m - 2\overline{\sigma}_f^m = 0 \tag{9}$$

where $\overline{\sigma}_f^m$ is the updated effective yielding stress and the exponent m is mainly associated with the crystal structure of the material: a great m value corresponds to a small curvature radius at the rounded vertices of the yield surface. Typically, $m = 6$ and 8 are recommended, and X_1', X_2', X_1'', X_2'' are the principal values of the transformed deviatoric stresses $\boldsymbol{S}$:

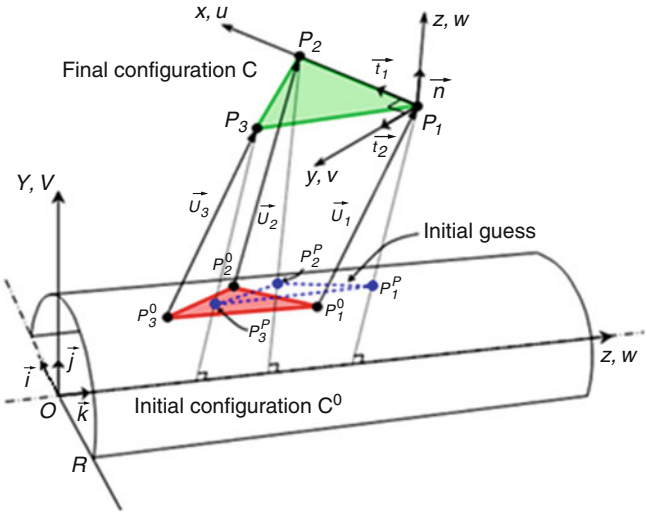


Fig. 3 3D mesh mapping onto the initial cylindrical tube surface

$$\begin{cases} \boldsymbol{X'} = \boldsymbol{C'}\,\boldsymbol{S} \\ \boldsymbol{X''} = \boldsymbol{C''}\,\boldsymbol{S} \end{cases} \tag{10}$$

where $\boldsymbol{C'}$ $\boldsymbol{C''}$ are the tensors of material parameters given in Barlat et al. (2003) and $\boldsymbol{\sigma}$ is the Cauchy stress tensor.

Using the normality law and the proportional loading assumption of the IA, the plastic strain rate can be integrated analytically, leading to a direct relationship between the total plastic strains and deviatoric stresses:

$$\boldsymbol{S} = \boldsymbol{H}^{p}\,\boldsymbol{\varepsilon}^{p} \tag{11}$$

Initial Solution Using Geometric Mapping Method

Several initial geometric guess techniques have been introduced by Naceur et al. (2002) to speed up the convergence of the static implicit solver in the IA. The authors proposed also other techniques for the treatment of vertical walls (balancing, opening, matrix condensing) which have been proven essential for complex 3D industrial parts.

To explain the basic concept of the initial solution, the geometric mapping method is described for the hydroforming of cylindrical tubes. The final mid-surface is discretized into triangular shell elements and mapped onto the initial cylindrical tube surface (Fig. 3).

Knowing the positions of the element nodes in the final configuration, the first guess can be achieved by radial projection of the nodes onto the initial cylindrical tube surface. These positions will be modified iteratively to meet the equilibrium in the final workpiece.

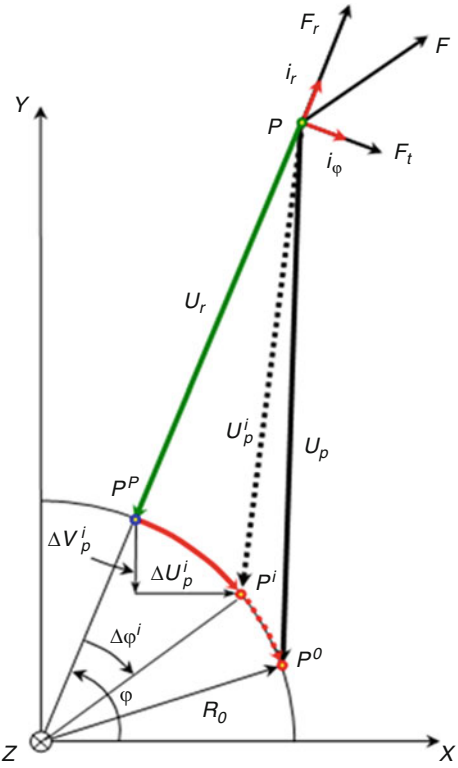


Fig. 4 Radial projection of point P and its movement on the tube surface

Since the mapping is made onto the known initial cylindrical surface, the radial displacement U_r of each node is known. The two other unknown displacements are axial and circumferential ones. The circumferential displacement is dependent on the unknown rotation angle $\Delta\varphi$. Its Cartesian components can be calculated as follows (Fig. 4):

$$\begin{array}{rcl} U_p^i & = & U_r \cos\varphi - \Delta U_p^i \\ V_p^i & = & U_r \sin\varphi + \Delta V_p^i \end{array} \text{ with } \begin{array}{rcl} \Delta U_p^i & = & R_0\left(\cos\left(\varphi - \Delta\varphi^i\right) - \cos\varphi\right) \\ \Delta V_p^i & = & R_0\left(\sin\left(\varphi - \Delta\varphi^i\right) - \sin\varphi\right) \end{array} \quad (12)$$

where R_0 is the radius of the tube mid-surface.

Variational Formulation

The discretization of the final shape of the desired workpiece is done by using a rotation-free shell element called DKTRF (Guo et al. 2002). This element is based on the membrane element CST and the plate element DKT6 (Batoz and Dhatt 1990). The DKTRF formulation involves the three neighboring elements in order to define the bending curvatures and build the element stiffness matrix without

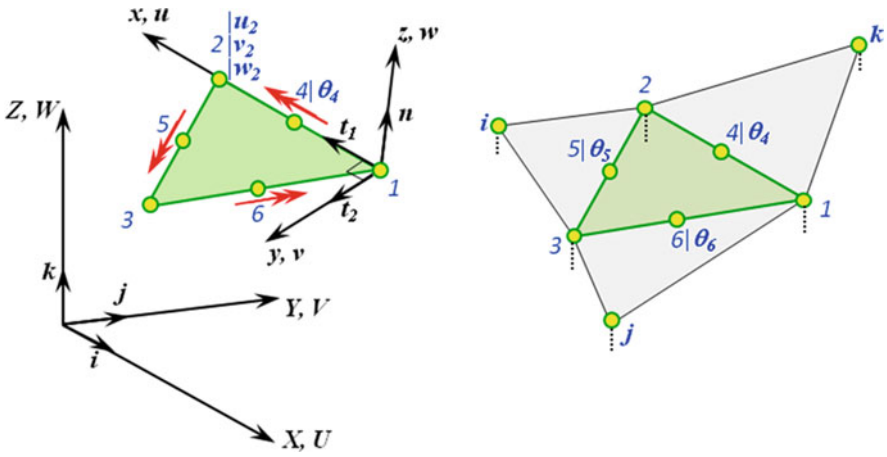


Fig. 5 DKT rotation-free triangular shell element

rotation degrees of freedom (Fig. 5). The resulting element has only three translations DOF per node.

Approximation of Strain Displacements

In the local reference of an element, the virtual membrane strains are expressed in terms of the two in-plane displacements along x and y:

$$\boldsymbol{e}^* = < \; u^*_{,x} \quad v^*_{,y} \quad u^*_{,y}+v^*_{,x} \; > \tag{13}$$

Linear approximations are used for u^* and v^* (constant strain triangle membrane element, CST) to obtain a constant membrane strain operator:

$$\boldsymbol{e}^* = \boldsymbol{B_m}\,\delta\left(\boldsymbol{u^*_n}\right)_m; \;\; (\boldsymbol{u_n})_m = \langle\, u_1 \quad v_1 \quad u_2 \quad v_2 \quad u_3 \quad v_3 \,\rangle \tag{14}$$

For the bending part, the rotations around the three element sides (θ_4, θ_5, θ_6) are expressed in terms of the transverse displacements of the six nodes (w_1, w_2, w_3, w_i, w_j, w_k, Fig. 5), the rotations normal to the three sides are expressed in terms of the transverse displacements of the three nodes (w_1, w_2, w_3, Fig. 5). Finally, a constant bending curvature operator is obtained, which is free from rotation degrees of freedom (Guo et al. 2002).

$$\boldsymbol{\chi}^* = \boldsymbol{B_f}\,\left(\boldsymbol{u^*_n}\right)_f; \;\; (\boldsymbol{u_n})_f = < \; w_1 \quad w_2 \quad w_3 \quad w_i \quad w_j \quad w_k \; > \tag{15}$$

Internal Force Vector

The principle of virtual work is used to establish the equilibrium on the final workpiece. The transverse shear effects are neglected for the thin sheet forming process. The virtual work of internal forces in an element is given by:

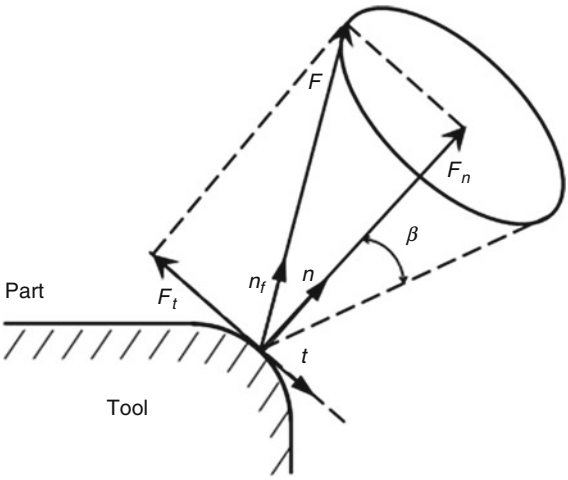


Fig. 6 Determination of the tool contact force by the friction cone

$$W_{\text{int}}^{e} = \int_{\text{V}^{\text{e}}} \boldsymbol{\varepsilon}^{*T}\, \boldsymbol{\sigma}\, dV = \boldsymbol{u}_{\boldsymbol{n}}^{*\text{T}} \boldsymbol{f}_{\text{int}}^{e} \tag{16}$$

By using the above FE approximation, the internal force vector in the global coordinate system is obtained:

$$W_{\text{int}}^{e} = \boldsymbol{U}_{\boldsymbol{n}}^{*\text{T}}\, \boldsymbol{F}_{\text{int}}^{e} \tag{17}$$

$$\boldsymbol{F}_{int}^{e} = \boldsymbol{T}^{\mathbf{T}} \int_{V^{e}} \left(\boldsymbol{B}_{\mathbf{m}}{}^{\text{T}} + z\, \boldsymbol{B}_{\mathbf{f}}{}^{\text{T}}\right) \quad \boldsymbol{\sigma}\, dV = A^{e}\, \boldsymbol{T}^{\text{T}} \left(\boldsymbol{B}_{\mathbf{m}}{}^{\text{T}} \boldsymbol{N} + \boldsymbol{B}_{\mathbf{f}}{}^{\text{T}} \boldsymbol{M}\right) \tag{18}$$

where $\boldsymbol{T}$ is the transformation matrix between the local and global references, $\boldsymbol{N}$ is the internal membrane force vector, and $\boldsymbol{M}$ is the internal bending–torsion moment vector.

External Force Vector

In the IA, the tool actions are simply represented by some external nodal forces to avoid the contact treatment. At a node, the resultant tool force $\boldsymbol{F}$ is composed of a normal pressure force $\boldsymbol{F}_n$ and a tangential friction force $\boldsymbol{F}_t$. $\boldsymbol{F}$ is situated on the friction cone surface defined by $\beta = arc\ \tan \mu$ (μ is friction coefficient, Fig. 6). Its direction $\boldsymbol{n}_f$ can be determined by the friction cone and the slide direction:

$$\boldsymbol{F} = F_n \boldsymbol{n} - F_t \boldsymbol{t} = \frac{F}{\sqrt{1+\mu^2}} (\boldsymbol{n} - \mu \boldsymbol{t}) = F \boldsymbol{n}_f \tag{19}$$

where $\boldsymbol{n}$ is the unit normal vector of the contour and $\boldsymbol{t}$ is the unit vector of the node displacement on the tangent direction.

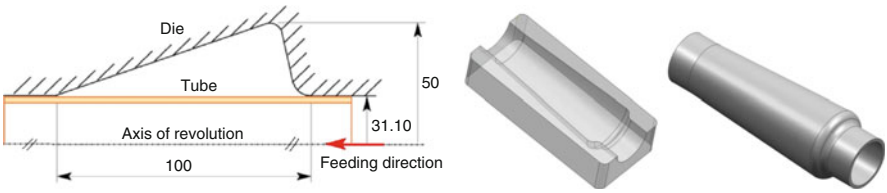


Fig. 7 Tube geometry and conical die

The value of $\boldsymbol{F}$ can be determined by the equilibrium condition at the contour. The following equilibrium equation can be established on a node k:

$$\boldsymbol{F}_{\text{ext}}^{k}\left(F^{k}\right)-\boldsymbol{F}_{\text{int}}^{i}=\left\{\begin{array}{c}F^{k}n_{X}^{k}\\F^{k}n_{Y}^{k}\\F^{k}n_{Z}^{k}\end{array}\right\}_{\text{ext}}-\left\{\begin{array}{c}F_{X}^{k}\\F_{Y}^{k}\\F_{Z}^{k}\end{array}\right\}_{\text{int}}=\left\{\begin{array}{c}0\\0\\0\end{array}\right\} \tag{20}$$

where $\left\langle\, n_{r}^{k} \quad n_{Z}^{k}\,\right\rangle^{T}=\boldsymbol{n}_{f}^{k}$ represents the resultant force direction at the node k. Then, the value of F^k and the nodal external force vector are obtained:

$$F^{k}=\left\langle\, n_{X}^{k} \quad n_{Y}^{k} \quad n_{Z}^{k}\,\right\rangle\left\{\begin{array}{c}F_{X}^{k}\\F_{Y}^{k}\\F_{Z}^{k}\end{array}\right\}_{\text{int}};\quad \boldsymbol{F}_{\text{ext}}^{k}=F^{k}\left\{\begin{array}{c}F_{X}^{k}\\F_{Y}^{k}\\F_{Z}^{k}\end{array}\right\}_{\text{ext}} \tag{21}$$

Application: Hydroforming of an Aluminum Alloy Conical Tube

The hydroforming of a conical tube has been studied by Jansson et al. (2008) (Fig. 7). The circumferential expansion attains 47.5 % from its original position. This is a difficult process because it requires a massive metal feeding into the expansion zone in order to avoid a burst fracture.

A 6-axis hydraulic press of 75 tons is used for the tube hydroforming. The axial feeding is provided on both sides of the tube. The whole device is instrumented with encoders and pressure sensors which allow the application of the axial feeding and pressure at any time during forming. Figure 8 shows the evolutions of the normalized pressure and feeding.

The tube is made of an aluminum alloy AA6063-T4. Both *Hill (1948)* and *Barlat (2000)* criteria are used. R_{00}, R_{45}, and R_{90} are the anisotropic coefficients; F, G, H, N, L are material parameters for *Hill* criterion; α_1–α_8 are the material parameters for *Barlat* criterion; they enable to consider the anisotropy in the evolution of the plastic flow surface for the case of uniaxial stress or biaxial stress state. The material data are as follows: $E = 68300$ MPa, $\nu = 0.3$, $\sigma_{00} = 78$ MPa, $\sigma_{45} = 76$ MPa, $\sigma_{90} = 74$ MPa, $\sigma_{11} = 23.4$ MPa, $\sigma_{22} = 85$ MPa, $R_{00} = 0.47$, $R_{45} = 0.12$, $R_{90} = 1.5$, $F = 0.43$, $G = 1.36$, $H = 0.64$, $N = 1.11$, $L = 0.43$, and α_1–$\alpha_8 = 0.72, 1.29, 0.99, 0.97, 1.03, 0.98, 0.16, 1.23$.

The curves of the pressure and feeding in function of time (Fig. 8) allow controlling the strain path, leading to a fully formed tube without bursting (Fig. 9). Firstly, the pressure is increased to 10 MPa with a moderate feeding in

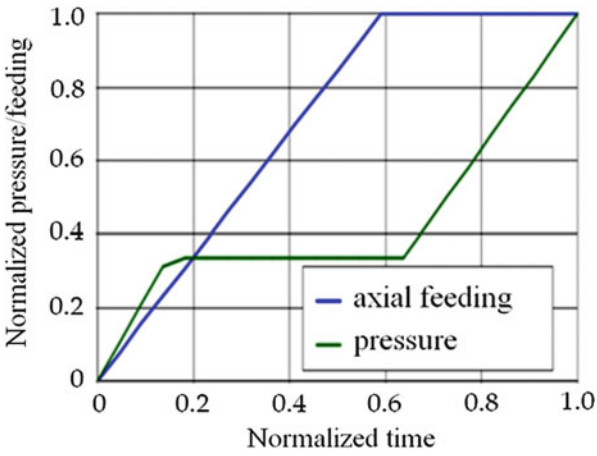


Fig. 8 Evolutions of the stroke (maximal value 33 mm) and pressure (maximal value 33 MPa) during forming

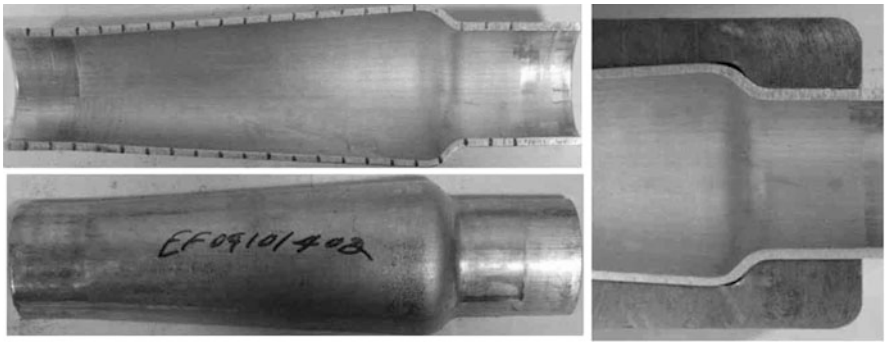


Fig. 9 Entirely formed conical tube

order to obtain a good compromise between the expansion and shortening; then, it is maintained constant while the axial feeding is increased until its maximum value of 33 mm. Secondly, the pressure is increased from 10 to 33 MPa to finish the expansion of the tube in the die.

In the numerical simulation, a quarter of the tube is discretized into 4536 DKTRF shell elements for the IA (Fig. 10). The same mesh is used for S3 shell elements in ABAQUS. The IA calculation uses only 38.7 s of CPU time, whereas the ABAQUS/Explicit calculation uses 275 s.

Figure 11 shows the thickness variation along the tube axis. The result obtained by the IA using the *Barlat (2000)* yielding criteria is in good agreement with that of ABAQUS using the same criterion.

It is noted that the thickness variation obtained by ABAQUS is closer to the experimental result than that by IA, especially at the die corner radius (at 80 and 210 mm). This can be explained by the fact that friction in the corner radii plays an important role, whereas the IA uses a simplified tooling action without loading history consideration.

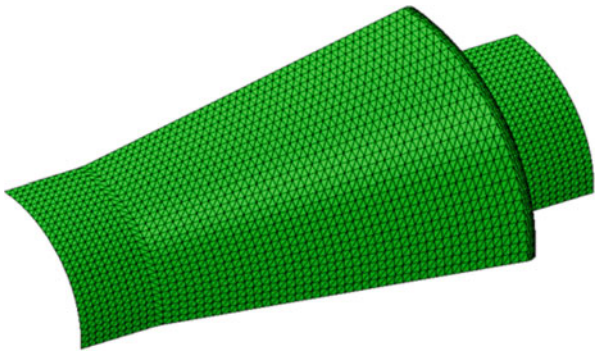


Fig. 10 Finite element mesh used for the IA calculation

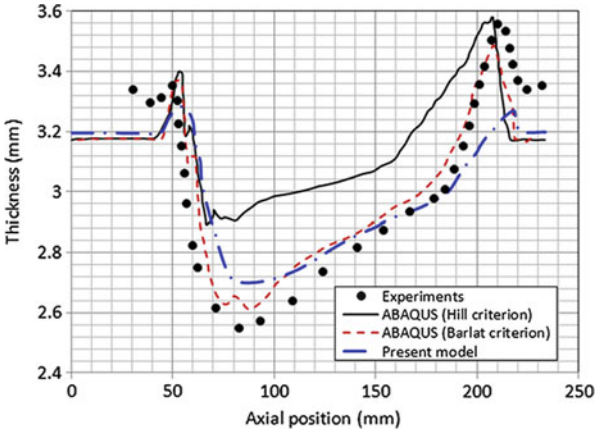


Fig. 11 Comparison of thickness variations along the tube axis

Pseudo-inverse Approach for Forming Process Modeling

In the IA, only the initial and final configurations are considered to calculate the total strains and then the total stresses by using the deformation theory of plasticity. This allows obtaining good strain estimation, but poor stress estimation. In the PIA, some *realistic* intermediate configurations are introduced to consider the deformation paths, and the incremental flow theory of plasticity with the damage is adopted to take into account the loading history.

Creation of Intermediate Configurations

Creation of Intermediate Configurations for Sheet Forming Modeling

In a 3D sheet forming process, since the final part shape and the tools are known, the intermediate configurations can be approximately generated by a geometrical

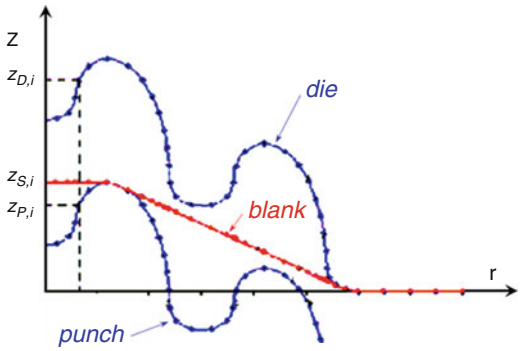


Fig. 12 Creation of an intermediate configuration by minimizing sheet surface

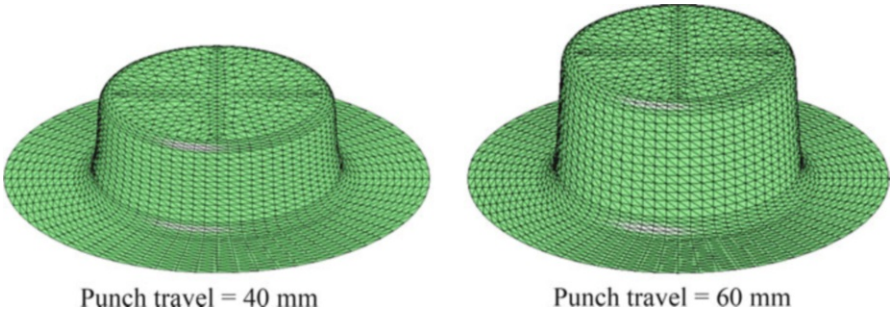


Fig. 13 Generation of the 3D mesh of an intermediate sheet configuration

method before the mechanical calculation (Guo et al. 2004). Considering a sheet as a stretched membrane in accordance with the tools (Fig. 12), its shape can be determined by minimizing its total surface:

$$J = Min \sum_{e} A_e; \quad z_{P,i} \leq z_{S,i} \leq z_{D,i} \tag{22}$$

where A_e is an element area and z_P, z_S, and z_D are the nodal vertical positions of a point on the punch, sheet, and die, respectively.

For a 2D or axisymmetric part, the minimization of the sheet surfaces can be done by using its profile. In a 3D case, the workpiece is divided into several sectors, and each of them is treated as a 2D case; all sectors are assembled together by linear interpolation (Fig. 13).

It is noted that the nodes having the same number in the intermediate and final meshes do not represent the same material point; a transfer of the strain and stress fields should be done between these two independent meshes.

Free Surface Method for Axisymmetric Cold Forging

In the axisymmetric cold forging process, some intermediate configurations are generated to consider the deformation history. An intermediate configuration is

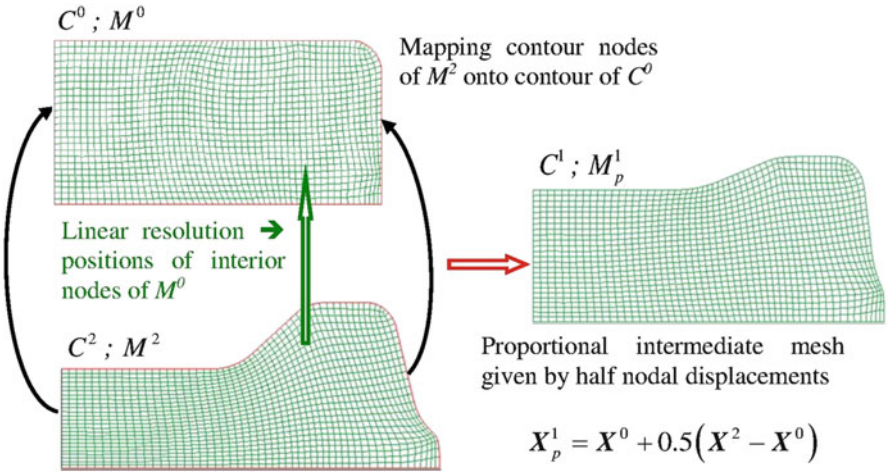


Fig. 14 Proportional intermediate mesh for PIA in a two-stage forging

created geometrically and then corrected by using a free surface method (Halouani et al. 2012), in order to obtain the free surface shape satisfying the equilibrium conditions.

Geometric Proportional Intermediate Configurations

In the PIA, the FE mesh M^2 is created on the known final part C^2. The contour of the initial billet C^0 is also known. For a two-step forging process, a geometric proportional configuration is generated as follows (Fig. 14):

1. The contour nodes of M^2 are mapped onto the contour of C^0, and the positions of the interior nodes on C^0 are determined by a linear solution on M^2 with the imposed contour node displacements from C^2 to C^0 as boundary conditions.
2. The intermediate mesh M^1_p is created by using a geometric proportional interpolation:

$$\boldsymbol{X}^1_p = \boldsymbol{X}^0 + 0.5\left(\boldsymbol{X}^2 - \boldsymbol{X}^0\right) \tag{23}$$

   where $\boldsymbol{X}^0$, $\boldsymbol{X}^1_p$, and $\boldsymbol{X}^2$ are the nodal position vectors of the meshes on the initial, intermediate, and final configurations.
3. The kinematic conditions of the mesh M^1_p are checked. If some nodes penetrate into the tool, they are mapped back on the tool contour (Fig. 15).

Determination of the Free Surface of an Intermediate Configuration

Once the kinematically admissible intermediate mesh is obtained, an inverse calculation is carried out between M^1_p and M^0. The mesh M^0 is fixed, the mesh M^1_p is taken as a reference to calculate the strains and stresses; M^1_p is modified iteratively in order to satisfy the equilibrium and contact conditions. Thus, the free surface

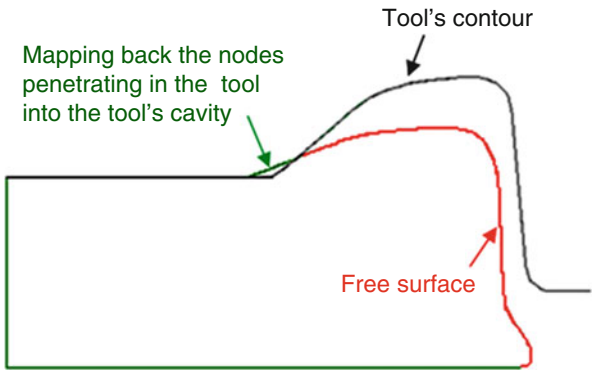


Fig. 15 Kinematically admissible free surface

shape is determined at the end of the equilibrium loop. The boundary conditions of M_p^1 are defined as follows (Fig. 15):

- If the nodal force points outward from the billet ($F_n > 0$, false), the node should be on a free surface having the boundary conditions: $U \neq 0$, $V \neq 0$, $\sigma_n = 0$, $\tau_n = 0$.
- If the nodal force points toward inside ($F_n < 0$), the contact conditions between the billet and tool give the tangent displacement $U_i' \neq 0$ and the normal one $V_i' = 0$ on the tool contour.
- Then, the IA calculation is performed and the nodal positions are updated in the intermediate configuration M_p^1.

The above operations are repeated in the equilibrium iteration loop until the convergence. The incompressibility condition, the contact condition, and the equilibrium at the free surface enable to obtain the mesh M_1^1 (superscript means step 1, subscript means configuration C^1) representing the real shape of the free surface.

Figure 16 shows the intermediate configurations obtained by the geometric proportional method, the PIA free surface method and ABAQUS® at the sixth step for a PIA calculation. It is noted that the proportional mesh is a good initial mesh but has a notable difference compared to the ABAQUS mesh (zones A and B). The realistic mesh obtained by the free surface method is very close to the ABAQUS® mesh.

Calculation of Large Strain Increments

The calculation of the large logarithmic strains in the IA is done in one step by directly comparing the initial billet and the final part (Halouani et al. 2010). A similar calculation is kept in the PIA but between two successive configurations.

For an axisymmetric problem, each material point moves in its meridian plane; the displacement field is therefore independent of the circumferential coordinate. The movement of a material point between two successive configurations C^{n-1} and

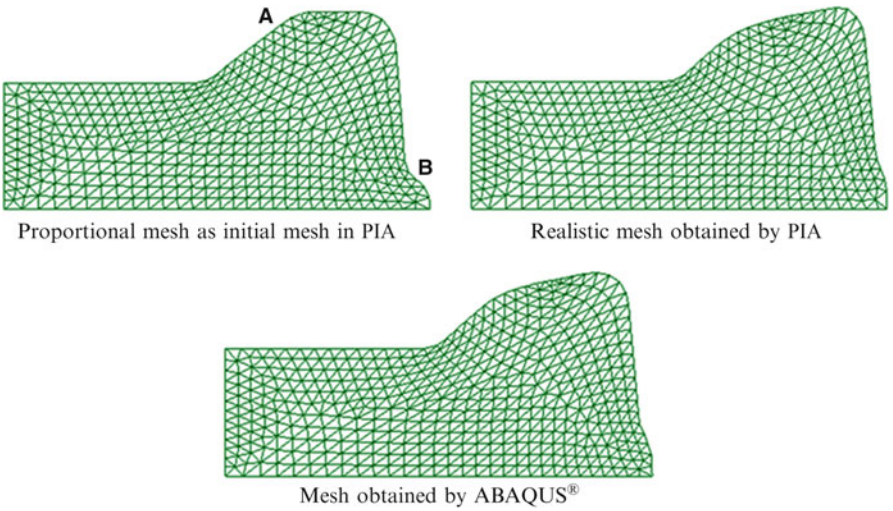


Fig. 16 Proportional mesh, PIA free surface mesh, and ABAQUS® mesh at step 6

C^n is expressed by $\boldsymbol{r}^{n-1} = \boldsymbol{r} - \Delta\boldsymbol{u}$ where $\boldsymbol{r}^{n-1}$ and $\boldsymbol{r}$ are the position vectors in C^{n-1} and C^n and $\Delta\boldsymbol{u}$ is the displacement increment vector in the radial plane.

Taking the known configurations C^n as reference, the inverse deformation gradient tensor is defined in the reference (r, z) as follows:

$$d\boldsymbol{r}^{n-1} = \frac{\partial \boldsymbol{r}^{n-1}}{\partial \boldsymbol{r}} d\boldsymbol{r} = \left(\boldsymbol{I} - \frac{\partial \Delta\boldsymbol{u}}{\partial \boldsymbol{r}}\right) d\boldsymbol{r} = \boldsymbol{F}_{\mathrm{L}}^{-1} d\boldsymbol{r} \tag{24}$$

with

$$\boldsymbol{F}_{\mathrm{L}}^{-1} = \begin{bmatrix} 1 - \Delta u_{,r} & 0 & -\Delta u_{,z} \\ 0 & 1 - \dfrac{\Delta u}{r} & 0 \\ -\Delta w_{,r} & 0 & 1 - \Delta w_{,z} \end{bmatrix} \tag{25}$$

The inverse of the Cauchy–Green left tensor is defined by:

$$(d\boldsymbol{r}^n)^{\mathrm{T}} d\boldsymbol{r}^n = d\boldsymbol{r}^{\mathrm{T}} \boldsymbol{F}_{\mathrm{L}}^{-\mathrm{T}} \boldsymbol{F}_{\mathrm{L}}^{-1} d\boldsymbol{r} = d\boldsymbol{r}^{\mathrm{T}} \boldsymbol{B}^{-1} d\boldsymbol{r} \tag{26}$$

$$\boldsymbol{B}^{-1} = \begin{bmatrix} (1 - \Delta u_{,r})^2 + (\Delta w_{,r})^2 & 0 & -\Delta u_{,z}(1 - \Delta u_{,r}) - \Delta w_{,r}(1 - \Delta w_{,z}) \\ 0 & (1 - \frac{\Delta u}{r})^2 & 0 \\ -\Delta u_{,z}(1 - \Delta u_{,r}) - \Delta w_{,r}(1 - \Delta w_{,z}) & 0 & (1 - \Delta w_{,z})^2 + (\Delta u_{,z})^2 \end{bmatrix} \tag{27}$$

The eigenvalues ($\Delta\lambda_1^{-2}$, $\Delta\lambda_2^{-2}$, $\Delta\lambda_3^{-2}$) of the tensor $\boldsymbol{B}^{-1}$ give the three principal elongations ($\Delta\lambda_1$, $\Delta\lambda_2$, $\Delta\lambda_3$), and the eigenvectors define the directions of these principal elongations:

$$\boldsymbol{B}^{-1} = \boldsymbol{M}\begin{bmatrix} \lambda_1^{-2} & 0 & 0 \\ 0 & \lambda_2^{-2} & 0 \\ 0 & 0 & \lambda_3^{-2} \end{bmatrix}\boldsymbol{M}^T \tag{28}$$

Then, the principal logarithmic strain increment is given by

$$\Delta\boldsymbol{\varepsilon} = \begin{Bmatrix} \Delta\varepsilon_1 \\ \Delta\varepsilon_2 \\ \Delta\varepsilon_3 \end{Bmatrix} = \begin{Bmatrix} \ln\Delta\lambda_1 \\ \ln\Delta\lambda_2 \\ \ln\Delta\lambda_3 \end{Bmatrix} \tag{29}$$

The large logarithmic strains in the reference (r, z) can be obtained by the following transformation (Batoz and Dhatt 1990):

$$\Delta\boldsymbol{\varepsilon} = \begin{Bmatrix} \Delta\varepsilon_r \\ \Delta\varepsilon_\theta \\ \Delta\varepsilon_z \\ \Delta\gamma_{rz} \end{Bmatrix} = \begin{bmatrix} \cos^2\varphi & 0 & \sin^2\varphi \\ 0 & 1 & 0 \\ \sin^2\varphi & 0 & \cos^2\varphi \\ 2\sin\varphi\cos\varphi & 0 & -2\sin\varphi\cos\varphi \end{bmatrix}\begin{Bmatrix} \Delta\varepsilon_1 \\ \Delta\varepsilon_2 \\ \Delta\varepsilon_3 \end{Bmatrix} \tag{30}$$

where φ is the angle from the r axis to the first principal strain axis.

In the PIA, the inverse calculation is carried out between two successive configurations using the strains and stresses obtained in the previous step. At the step n-1, the FE mesh is created on C^{n-1} and modified by the free surface method; at the step n, the mesh is created on C^n and mapped on C^{n-1} for the inverse calculation. These two meshes on C^{n-1} are completely independent, so a transfer of the strain and stress fields should be done between them.

Direct Integration of Plasticity and Damage for Large Strain Increments

The return mapping algorithm (RMA) (Simo and Taylor 1986) is the most widely used iterative scheme. It is considered as an efficient method for the plastic integration, but it consumes much computation time because of numerous integration points in whole structure and numerous iterations in the global equilibrium loop. Moreover, this iterative scheme may cause divergence problems for large strain increments. The new algorithm called direct scalar algorithm of plasticity (DSAP) proposed by Li et al. (2007) is very fast and robust without local iterative loop. The basic idea of the DSAP is to transform the constitutive equations with unknown stress vectors into a scalar equation in terms of equivalent stresses which can be obtained by using the tensile curve, leading to a direct solution to obtain the plastic multiplier $\Delta\lambda$ (see section "► Simplified Plastic Ductile Damage Models and Direct Integration Algorithms").

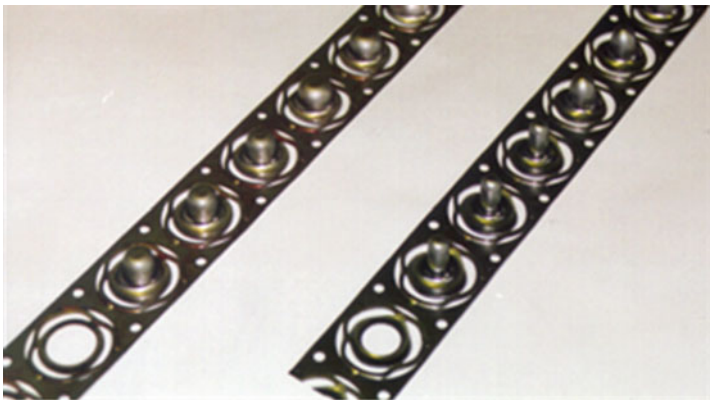


Fig. 17 Photo of an axisymmetric pot stamping in three stages

Numerical Results: Simulation of a Three-Stage Stamping Process

An axisymmetric pot is realized by three successive stamping stages (Fig. 17). The stamping is carried out continuously on a sheet band. The geometrical dimensions are given in Fig. 18. The first two stages are designed to obtain the biaxial strain states below the FLC curve.

The sheet is cut by six arc knives before the stamping to facilitate the sheet draw-in. In the third stage, the pot with a large diameter should be pushed by the punch into the die of smaller diameter; this implies the bending–unbending effect. The intermediate shape of the part is determined by the geometrical relations between the sheet and the tools (Fig. 19).

The material of the sheet is the steel DC04, its thickness is 1 mm. The part heights at the three stages are 13, 15 and 15 mm, respectively. The friction coefficient is 0.144 between the sheet and die.

In stage 1, a good agreement of the thickness distributions is obtained between the numerical results of the PIA and the experimental results of CETIM. It is found that the maximal thickness thinning is 13.2 % for CETIM and 10.6 % for PIA. The comparison of the FLD (forming limit diagram) shows also a good coherence between the numerical and experimental strain states.

In stage 2, the thickness distributions obtained by CETIM and PIA are also very similar. The maximal thinning is situated at the upper radius of the part. The maximal values are 17.4 % by CETIM and 15.3 % by PIA (Fig. 20). The experimental FLD points correspond well to those obtained by the PIA (Fig. 21). It is noted that only a small zone on the part has been measured in the experimental test.

In stage 3, the difference of the thickness distributions becomes notable between the numerical and experimental results: 30.3 % of thinning for PIA but 40.7 % for CETIM (Fig. 22). The reason of this error is due to the assumption on the

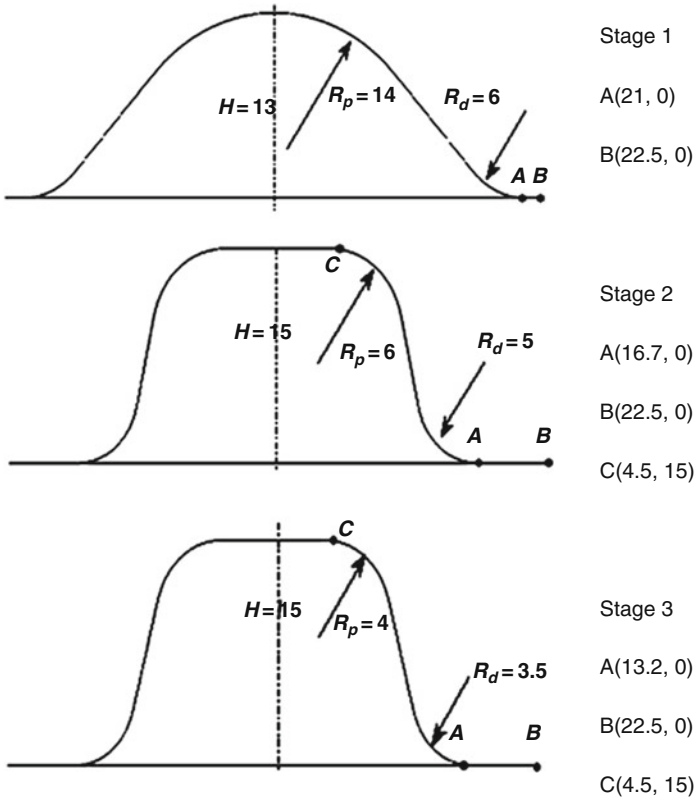


Fig. 18 Part geometry at the three stamping stages

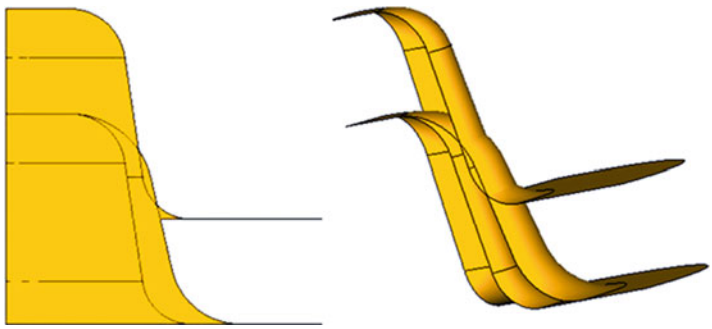


Fig. 19 Determination of the intermediate sheet shape in the third stage

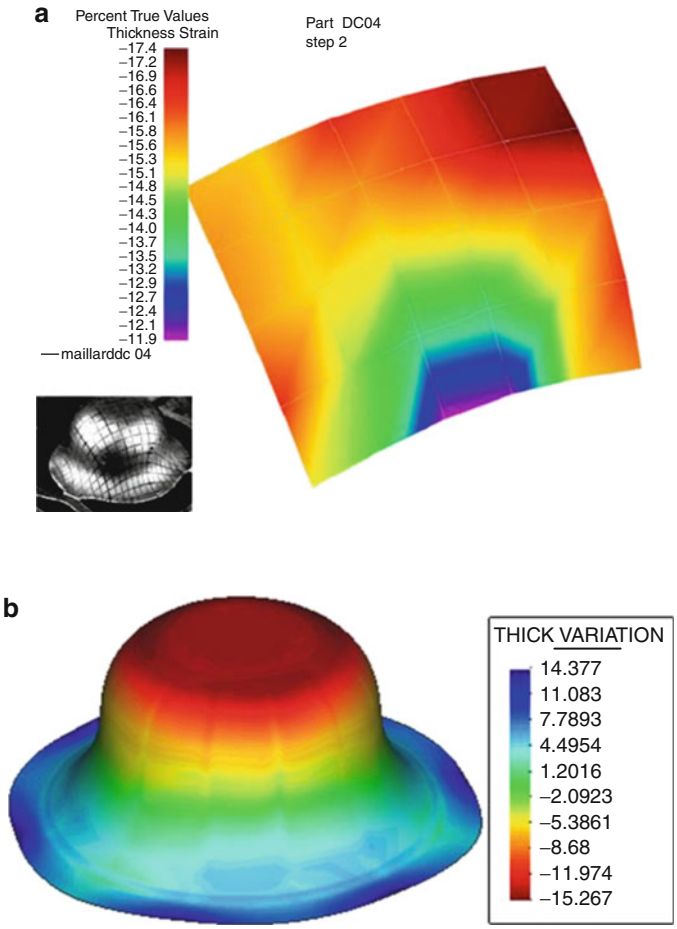


Fig. 20 Thickness variation (%) obtained by PIA and by test in stage 2. (**a**) Experimental test of CETIM. (**b**) Numerical simulation of PIA

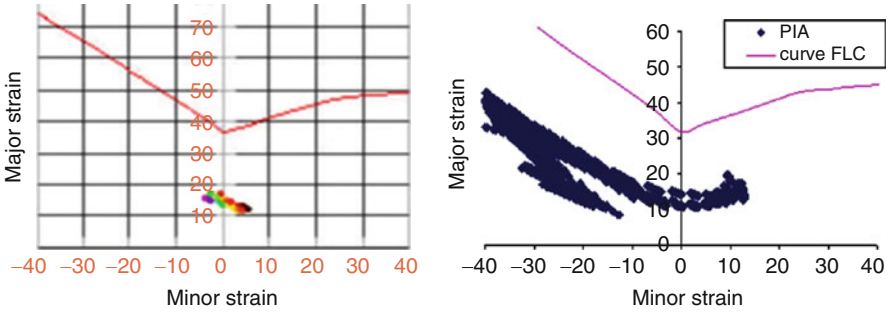


Fig. 21 FLC and FLD diagrams in stage 2

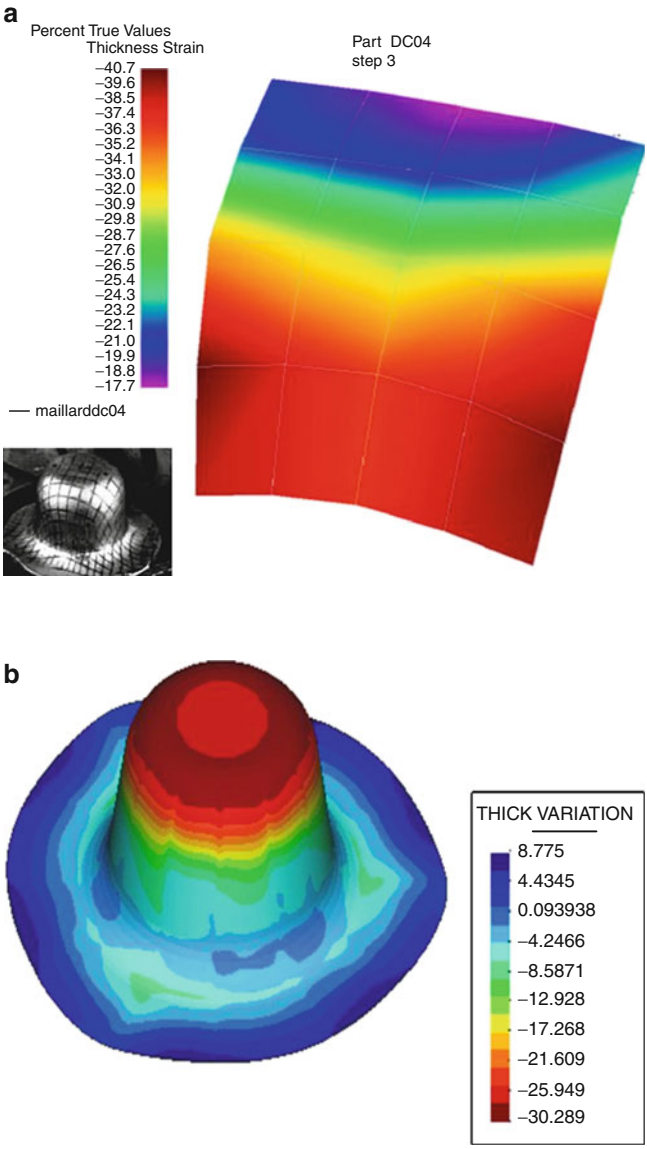


Fig. 22 Thickness variation (%) obtained by PIA and by test in stage 3. (**a**) Experimental test of CETIM. (**b**) Numerical simulation of PIA

contact-friction effects. Despite this difference, PIA is able to find correctly the maximal thinning zone.

In stage 3, the FLD numerical and experimental results are also fairly coherent (Fig. 23). In a "good part" just before the rupture, the points are very close to the FLC in both cases.

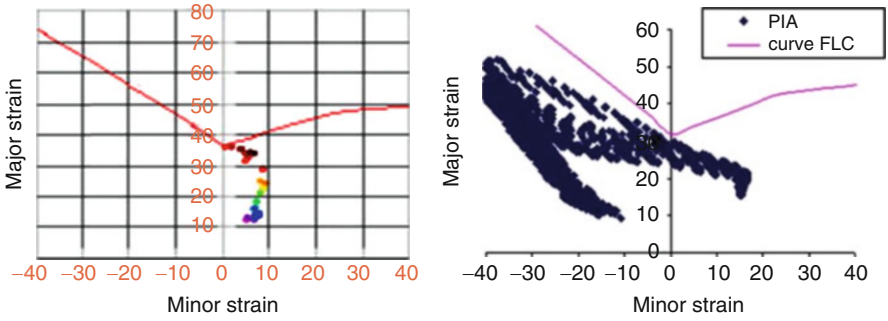


Fig. 23 FLC and FLD diagrams in stage 3

Simplified Plastic Ductile Damage Models and Direct Integration Algorithms

Strain-Based Damage Model

The damage models based on the continuum damage mechanics (CDM) were presented by (Chaboche 1988; Lemaître and Chaboche 1990). The damage effect is represented on the overall elastoplastic behavior of the material. This phenomenological model uses a scalar damage variable to describe the ductile defect evolution and thermomechanical behaviors. Advanced models on the CDM ductile damages and their strong couplings with elasto-viscoplastic behaviors are presented in ▶ Chap. 25, "Ductile Damage in Metal Forming: Advanced Macroscopic Modeling and Numerical Simulation". In the present chapter, a simplified damage model called "3D strain-based damage model" (Lemaître and Chaboche 1990) is used. This model partially ignores the loading history and is weakly coupled with the plastic behavior. It makes the formulation very simple and well adapted to the PIA, leading to an efficient numerical damage modeling.

The damage potential φ_D^* is chosen as the function of the strain energy density release rate $(-Y)$, so the damage rate $\dot{D}$ for a material with the isotropic hardening and damage can be written as follows:

$$\dot{D} = -\dot{\lambda}_D \frac{\partial \varphi_D^*}{\partial Y} = \left(\frac{-Y}{S_0}\right)^{s_0} \dot{\bar{\varepsilon}}^p \tag{31}$$

$$-Y = \frac{\sigma_{eq}^2}{2E(1-D)^2}\left[\frac{2}{3}(1+\nu) + 3(1-2\nu)\left(\frac{\sigma_{\mathrm{H}}}{\sigma_{eq}}\right)^2\right] \tag{32}$$

where $\dot{\lambda}_D$ is the damage multiplier rate, s_0 and S_0 are material coefficients in function of the temperature, σ_{eq} is the equivalent stress, σ_H is the hydrostatic stress, $\dot{\bar{\varepsilon}}^p$ is the equivalent plastic strain rate, E is Young's modulus, and ν is Poisson's coefficient.

Two assumptions are made to obtain a strain-based damage model: the assumption of hardening saturation after the damage threshold giving a perfect plasticity behavior and the assumption of proportional loading giving a constant triaxiality ratio σ_H/σ_{eq}.

Introducing the damage threshold ε_D, the equivalent plastic strain ε_R at the rupture, and the damage value at the rupture D_c (experimentally available, Zhu et al. 1992), a simplified strain-based damage model can be obtained in a rate form or in an integrated form:

$$\dot{D} = \frac{D_c}{\varepsilon_R - \varepsilon_D}\left[\frac{2}{3}(1+\nu) + 3(1-2\nu)\left(\frac{\sigma_H}{\sigma_{eq}}\right)^2\right]\dot{\bar{\varepsilon}}^p \quad ; \quad \text{if } (\bar{\varepsilon}^p > \varepsilon_D \text{ and } \sigma_H > 0) \tag{33}$$

$$D = \frac{D_c}{\varepsilon_R - \varepsilon_D}\left(\bar{\varepsilon}^p\left[\frac{2}{3}(1+\nu) + 3(1-2\nu)\left(\frac{\sigma_H}{\sigma_{eq}}\right)^2\right] - \varepsilon_D\right) \tag{34}$$

where $\dot{\bar{\varepsilon}}^p$ is the equivalent plastic strain rate, a compressive stress state ($\sigma_H < 0$) cannot induce the damage, giving $\dot{D} = 0$.

Constitutive Equations

In this chapter, the material is supposed to obey von Mises isotropic yield criterion (for cold forging) or Hill anisotropic yield criterion (for sheet forming). These criteria of plasticity with the damage consideration are given by

$$f = \frac{\sigma_{eq}}{1-D} - \bar{\sigma}(\bar{\varepsilon}^p) = 0 \tag{35}$$

with

$$\sigma_{eq} = \left(\boldsymbol{\sigma}^T \boldsymbol{P}\, \boldsymbol{\sigma}\right)^{1/2} \tag{36}$$

where $\bar{\sigma} = \bar{\sigma}(\bar{\varepsilon}^p)$ represents the uniaxial tensile curve, σ_{eq} is the equivalent stress, and $\boldsymbol{P}$ is isotropic or anisotropic matrix defined below.

The plastic normality rule is used as the flow rule to obtain the plastic strain rate:

$$\dot{\boldsymbol{\varepsilon}}^p = \dot{\lambda}\frac{\partial f}{\partial \boldsymbol{\sigma}} = \dot{\lambda}\frac{\partial f}{\partial \sigma_{eq}}\frac{\partial \sigma_{eq}}{\partial \boldsymbol{\sigma}} = \dot{\lambda}\frac{\boldsymbol{P}\,\boldsymbol{\sigma}}{(1-D)\sigma_{eq}} \tag{37}$$

Using the equivalent plastic work $\dot{\bar{\varepsilon}}^p \sigma_{eq} = (\dot{\boldsymbol{\varepsilon}}^p)^{\mathrm{T}}\boldsymbol{\sigma}$, the relation between the equivalent plastic strain rate and the plastic multiplier rate $\dot{\lambda}$ is obtained:

$$\dot{\bar{\varepsilon}}^p = \frac{1}{\sigma_{eq}}\frac{\dot{\lambda}\,\boldsymbol{\sigma}^T\boldsymbol{P}^T}{(1-D)\sigma_{eq}}\boldsymbol{\sigma} = \frac{\dot{\lambda}}{1-D} \tag{38}$$

The equivalent plastic strain is defined by

$$\dot{\bar{\varepsilon}}^{\,p} = (\dot{\boldsymbol{\varepsilon}}^{\boldsymbol{p}})^{\boldsymbol{T}} \boldsymbol{A}\, \dot{\boldsymbol{\varepsilon}}^{\boldsymbol{p}} \tag{39}$$

For the axisymmetrical cold forging, the isotropic material gives

$$\{\sigma\} = \begin{Bmatrix} \sigma_r \\ \sigma_\theta \\ \sigma_z \\ \sigma_{rz} \end{Bmatrix}; \quad \{\varepsilon\} = \begin{Bmatrix} \varepsilon_r \\ \varepsilon_\theta \\ \varepsilon_z \\ \varepsilon_{rz} \end{Bmatrix}; \quad \boldsymbol{P} = \begin{bmatrix} 1 & -0.5 & -0.5 & 0 \\ -0.5 & 1 & -0.5 & 0 \\ -0.5 & -0.5 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix};$$

$$\boldsymbol{A} = \frac{2}{3}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix}$$

For the thin sheet forming, the assumptions of plane stress, transverse anisotropy, and isotropic hardening are adopted. Using the equivalent plastic work $\dot{\bar{\varepsilon}}^{\,p}\sigma_{eq} = (\dot{\boldsymbol{\varepsilon}}^{p})^{T}\boldsymbol{\sigma}$ and Eqs. 33, 35, the following relations in a local reference are obtained:

$$\{\sigma\} = \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \sigma_{xy} \end{Bmatrix}; \quad \{\dot{\varepsilon}^{\mathrm{p}}\} = \begin{Bmatrix} \dot{\varepsilon}^{\mathrm{p}}_x \\ \dot{\varepsilon}^{\mathrm{p}}_y \\ \dot{\varepsilon}^{\mathrm{p}}_{xy} \end{Bmatrix}; \quad \boldsymbol{A} = \boldsymbol{P}^{-1} = \begin{bmatrix} 1 & \frac{-\bar{r}}{1+\bar{r}} & 0 \\ \frac{-\bar{r}}{1+\bar{r}} & 1 & 0 \\ 0 & 0 & \frac{2(1+2\bar{r})}{1+\bar{r}} \end{bmatrix}^{-1}$$

with the average transverse anisotropy coefficient $\bar{r} = \frac{1}{4}(r_0 + 2r_{45} + r_{90})$.

Integrated Constitutive Law

In the IA, the proportional loading assumption postulates that the stress tensor at a point is proportional to an initial tensor independent of the time:

$$\boldsymbol{\sigma}(\boldsymbol{x}, t) = \alpha(t)\boldsymbol{\sigma}(\boldsymbol{x}, t_0) \tag{40}$$

so the term $\boldsymbol{\sigma}/\sigma_{eq}$ is independent of the time and Eq. 37 can be analytically integrated:

$$\boldsymbol{\varepsilon}^{p} = \frac{\dot{\bar{\varepsilon}}^{p}}{\sigma_{eq}}\boldsymbol{P}\,\boldsymbol{\sigma} = \left(\frac{1}{E_S} - \frac{1}{E}\right)\boldsymbol{P}\,\boldsymbol{\sigma} \tag{41}$$

where the relation $\dot{\bar{\boldsymbol{\varepsilon}}}^{p} = \dot{\lambda}$ in Eq. 38 has been used (but without the damage consideration). Adding the elastic strain vector in Eq. 40, the total strain–stress relation is obtained:

$$\boldsymbol{\sigma} = \left[\boldsymbol{H}^{-1} + \left(\frac{1}{E_S} - \frac{1}{E}\right)\boldsymbol{P}\right]^{-1}\boldsymbol{\varepsilon} \tag{42}$$

where $\boldsymbol{H}$ is the elastic constitutive matrix. The damage effect is uncoupled with the plasticity and evaluated in the post-processing using Eq. 34.

Classical Return Mapping Algorithm for Plasticity–Damage

The elastic law coupled with damage can be written in a total or rate form:

$$\boldsymbol{\sigma} = (1 - D)\boldsymbol{H}\boldsymbol{\varepsilon}^{\mathrm{e}} \tag{43}$$

$$\dot{\boldsymbol{\sigma}} = (1 - D)\boldsymbol{H}(\dot{\boldsymbol{\varepsilon}} - \dot{\boldsymbol{\varepsilon}}^{\mathrm{p}}) - \frac{\dot{D}\boldsymbol{\sigma}}{1 - D} \tag{44}$$

While from Eqs. 33 and 38, the damage rate can be defined by

$$\dot{D} = \frac{\hat{Y}}{1 - D}\dot{\lambda} \tag{45}$$

with

$$\hat{Y} = \frac{D_c}{\varepsilon_R - \varepsilon_D}\left[\frac{2}{3}(1 + \nu) + 3(1 - 2\nu)\left(\frac{\sigma_H}{\sigma_{eq}}\right)^2\right] \tag{46}$$

Using Eqs. 37 and 44, 45, and 46, the stress rate can be expressed in function of the plastic multiplier rate $\dot{\lambda}$:

$$\dot{\boldsymbol{\sigma}} = (1 - D)\boldsymbol{H}\dot{\boldsymbol{\varepsilon}} - \dot{\lambda}\left(\frac{\boldsymbol{HP}}{\sigma_{eq}} + \frac{\hat{Y}}{(1 - D)^2}\boldsymbol{I}\right)\boldsymbol{\sigma} \tag{47}$$

Thus, the stress vector at the step n can be expressed in an incremental form:

$$\boldsymbol{\sigma}_n - \boldsymbol{\sigma}_{n-1} = (1 - D_n)\boldsymbol{H}\Delta\boldsymbol{\varepsilon} - \Delta\lambda\left(\frac{\boldsymbol{HP}}{\sigma_{eq,n}} + \frac{\hat{Y}}{(1 - D_n)^2}\boldsymbol{I}\right)\boldsymbol{\sigma}_n \tag{48}$$

where an implicit scheme is taken to ensure the numerical stability. Equation 44 can be rewritten as

$$\left(\boldsymbol{I} + \Delta\lambda\left(\frac{\boldsymbol{HP}}{\sigma_{eq,n}} + \frac{\hat{Y}}{(1 - D_n)^2}\boldsymbol{I}\right)\right)\boldsymbol{\sigma}_n = \boldsymbol{\sigma}_{n-1} + (1 - D_n)\boldsymbol{H}\Delta\boldsymbol{\varepsilon} \tag{49}$$

where the stress vector $\boldsymbol{\sigma}_n$ can be determined by using an elastic prediction then a plastic correction. The elastic prediction gives a trial stress state as follows:

$$\boldsymbol{\sigma}_n^{\mathrm{e}} = \boldsymbol{\sigma}_{n-1} + (1 - D_n)\boldsymbol{H}\Delta\boldsymbol{\varepsilon} \tag{50}$$

The above elastic stress vector is substituted into the plastic criterion (Eq. 35), noted as f^{e}. $f^{\mathrm{e}} < 0$ means the stress state inside the flow surface (elastic prediction is true), giving $\Delta\lambda = 0$; $f^{\mathrm{e}} > 0$ means the plasticity occurrence; a plastic correction

should be used to determine a new stress state on the flow surface ($f = 0$). In the Simo's return mapping method, one substitutes $\boldsymbol{\sigma}_n$ (in Eq. 49) into the plastic criterion (Eq. 35) and solves the nonlinear equation $f(\Delta\lambda) = 0$ by using the Newton–Raphson iterative method to obtain $\Delta\lambda$. It is noted that a weak coupling method between the damage and plasticity is often adopted by using the damage value D_n at the previous equilibrium iteration.

Direct Scalar Algorithm of Plasticity (DSAP) for Fast Plastic Integration

In this direct algorithm, using the equivalent stress operator, Eq. 45 with unknown stress vectors is transformed into a scalar equation in terms of the equivalent stresses which can be determined by using the tensile curve. Thus, a quadratic equation with an unique unknown $\Delta\lambda$ is obtained, leading to a direct solution.

Calculation of Approximate Ratio of the Elastic Strain Part in a Strain Increment

For a given strain increment, if the elastic and plastic parts can be separated (even approximately), the equivalent plastic strain $\overline{\varepsilon}_n^{\mathrm{p}} = \overline{\varepsilon}_{n-1}^{\mathrm{p}} + \Delta\overline{\varepsilon}^{\mathrm{p}}$ can be obtained, and then the equivalent stress can be calculated by using the tensile curve $\overline{\sigma}_n = \overline{\sigma}\left(\overline{\varepsilon}_n^{\mathrm{p}}\right)$.

In a loading increment, a material point may undergo an elastic unloading (*AD* in Fig. 24) and reloading (*DA*), and then an elastoplastic loading (*AC*) which is modeled numerically by an elastic prediction (*AB*) and plastic correction (*BC*). How can we determine the ratios of the elastic and plastic parts in a strain increment ($\gamma\Delta\boldsymbol{\varepsilon}$ and $(1-\gamma)\Delta\boldsymbol{\varepsilon}$)?

Supposing that the elastic part $\gamma\Delta\boldsymbol{\varepsilon}$ enables the stress state to reach the flow surface, the criterion of plasticity should be satisfied, leading to the following equations (Eq. 35):

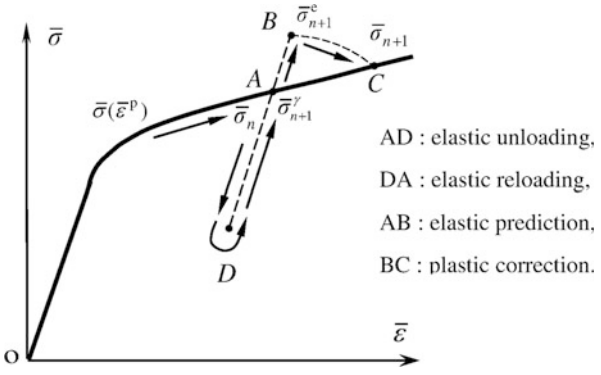


Fig. 24 Elastic unloading (*AD*), reloading (*DA*), and elastoplastic loading (*AC*)

$$\boldsymbol{\sigma}_n^{\gamma} = \boldsymbol{\sigma}_{n-1} + \gamma(1 - D_n)\boldsymbol{H}\Delta\boldsymbol{\varepsilon} \tag{51}$$

$$f\left(\boldsymbol{\sigma}_n^{\gamma}(\gamma)\right) = 0 \tag{52}$$

Equation 52 can be solved by the Newton–Raphson method. In order to avoid the iterative solution, the notion of the equivalent stress is adopted to transform Eq. 51 into a scalar equation by using the operation $\boldsymbol{\sigma}^T\boldsymbol{P}\boldsymbol{\sigma}$ at the two sides of Eq. 51:

$$\begin{aligned}\left(\sigma_{eq,n}^{\gamma}\right)^2 = \left(\sigma_{eq,n-1}\right)^2 &+ 2(1 - D_n)\gamma\boldsymbol{\sigma}_{n-1}^T\boldsymbol{P}\boldsymbol{H}\Delta\boldsymbol{\varepsilon} \\ &+ \gamma^2(1 - D_n)^2\Delta\boldsymbol{\varepsilon}^T\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}\Delta\boldsymbol{\varepsilon}\end{aligned} \tag{53}$$

In the above equation, $\sigma_{eq,n}^{\gamma}$ and $\sigma_{eq,n-1}$ are determined by using the tensile curve, so γ can be obtained directly without iterations. The elastic ratio γ should be between 0 and 1. $\gamma > 1$ means that the total strain increment is insufficient to bring the stresses state back to the flow surface, so $\gamma = 1$ should be taken.

Once the elastic percentage γ is obtained, the equivalent plastic strain and equivalent stress at the step n can be calculated by

$$\overline{\varepsilon}_n^{\mathrm{p}} = \overline{\varepsilon}_{n-1}^{\mathrm{p}} + (1 - \gamma)\Delta\overline{\varepsilon}_n \tag{54}$$

$$\sigma_{eq,n} = (1 - D_n)\overline{\sigma}\left(\overline{\varepsilon}_n^{p}\right) \tag{55}$$

The stress $\sigma_{eq,n}$ will be used in Eq. 57 to calculate the plastic multiplier $\Delta\lambda$.

Direct Calculation of the Plastic Multiplier $\Delta\lambda$

In the plastic correction phase, Eqs. 49 and 50 can be rewritten as follows:

$$\boldsymbol{\sigma}_n^e = \left(\boldsymbol{I} + \Delta\lambda\left(\frac{\boldsymbol{HP}}{\sigma_{eq,n}} + \frac{\hat{Y}}{(1 - D_n)^2}\boldsymbol{I}\right)\right)\boldsymbol{\sigma}_n \tag{56}$$

By using the equivalent stress notion, the operation $\boldsymbol{\sigma}^T\boldsymbol{P}\boldsymbol{\sigma}$ is done on the two sides of the above equation, leading to an equation of second degree in $\Delta\lambda$:

$$\begin{aligned}\left(\sigma_{eq,n}^e\right)^2 = (\sigma_{eq,n})^2 &+ 2\Delta\lambda\,\boldsymbol{\sigma}_n^T\left(\frac{\boldsymbol{HP}}{\sigma_{eq,n}} + \frac{\hat{Y}}{(1 - D_n)^2}\boldsymbol{I}\right)\boldsymbol{P}\boldsymbol{\sigma}_n \\ &+ \Delta\lambda^2\boldsymbol{\sigma}_n^T\left(\frac{\boldsymbol{HP}}{\sigma_{eq,n}} + \frac{\hat{Y}}{(1 - D_n)^2}\boldsymbol{I}\right)^T\boldsymbol{P}\left(\frac{\boldsymbol{HP}}{\sigma_{eq,n}} + \frac{\hat{Y}}{(1 - D_n)^2}\boldsymbol{I}\right)\boldsymbol{\sigma}_n\end{aligned} \tag{57}$$

Normally, this nonlinear equation requires an iterative solution. However, if one uses the equivalent stress obtained by Eq. 55, the damage value D_n at the previous equilibrium iteration, and an approximate stress normal direction, then $\Delta\lambda$ can be directly obtained by solving Eq. 57 without iterations. Using the criterion of plasticity (Eq. 35) and the plastic normality flow rule, the normal of the flow surface can be determined by

$$\boldsymbol{n} = \frac{\partial f}{\partial \boldsymbol{\sigma}} = \frac{\boldsymbol{P}\boldsymbol{\sigma}}{(1-D)\sigma_{\mathrm{eq}}} \quad \rightarrow \quad \boldsymbol{n} = \frac{\boldsymbol{P}\boldsymbol{\sigma}_n}{(1-D_n)\sigma_{\mathrm{eq},n}} = \frac{\boldsymbol{P}\boldsymbol{\sigma}_n}{(1-D_n)^2\overline{\sigma}_n} \tag{58}$$

where the last known stress normal direction can be used. Finally, Eq. 57 can be simplified to a second-order equation of $\Delta\lambda$:

$$\begin{aligned}&\Delta\lambda^2\left[(1-D_n)^2\boldsymbol{n}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}\boldsymbol{n} + \left(\frac{\hat{Y}}{(1-D_n)^2}\right)^2\left(\sigma_{\mathrm{eq},n}\right)^2 + 2\sigma_{\mathrm{eq},n}\hat{Y}\boldsymbol{n}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{n}\right]\\ &+2\Delta\lambda\left[(1-D_n)^2\sigma_{\mathrm{eq},n}\boldsymbol{n}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{n} + \frac{\hat{Y}}{(1-D_n)^2}\left(\sigma_{\mathrm{eq},n}\right)^2\right] + \left(\sigma_{\mathrm{eq},n}\right)^2 - \left(\sigma^{\mathrm{e}}_{\mathrm{eq},n}\right)^2 = 0\end{aligned} \tag{59}$$

It is noted that some quantities such as γ, $\boldsymbol{n}$, and D_n have been calculated approximately, an improvement can be done by replacing $(1-\gamma)\Delta\bar{\varepsilon}_n$ in Eq. 54 by $\Delta\lambda/(1-D)$ (Eq. 38) and repeat the operations of Eqs. 55 and 57 once again. But the numerical tests have shown a good agreement with the classical return mapping method even without this improvement.

This direct scalar algorithm to obtain $\Delta\lambda$ is very fast and robust; it enables to use large strain increments without divergence problems.

Numerical Results of Damage Prediction

Sheet Forming of a Square Box

An advanced fully coupled damage model is developed and implemented in ABAQUS/Explicit (Saanouni et al. 2000; Cherouat et al. 2004). This program is used to simulate the damage evolution in the sheet forming of a square box and to validate the PIA with the damage consideration.

The geometric data are as follows: initial blank 200 × 200 × 0.82 mm^3, punch section of 100 × 100 mm^2 with a round radius of 8 mm, and die cavity of 102.5 × 102.5 mm^2 with a round radius of 5 mm. The punch travel is 36 mm. The material properties are as follows: friction coefficient $\mu = 0.144$, Young's modulus $E = 210$ GPa, Poisson's coefficient $\nu = 0.3$, yield stress $\sigma_{\mathrm{y}} = 400$ MPa, and isotropic plasticity law $\overline{\sigma} = Q\left(1-e^{-b\bar{\varepsilon}}\right) = 1000\left(1-e^{-5\bar{\varepsilon}}\right)$ MPa. In the PIA, the used damage parameters ($D_c = 0.95$, $\varepsilon_R = 0.7$, $\varepsilon_D = 0$) give a similar damage behavior with that of Cherouat et al. (2004), but PIA damage model is not able to describe very large damage until the rupture. Figure 25 shows the damage distributions obtained by ABAQUS with coupled or uncoupled plasticity–damage models. It is found that the damage is always located in the same zone, but the damage value given by the coupled model is more concentrated and much bigger ($D_{\max} = 90.5$ %) in the coupled case than in the uncoupled case ($D_{\max} = 53.48$ %).

In Fig. 26, the damage distributions obtained by PIA are presented for the coupled and uncoupled cases. A similar phenomenon is observed, but the damage evolutions show some difference due to the different damage models: the rigidity in the ABAQUS simulation decreases more rapidly after the ultimate load.

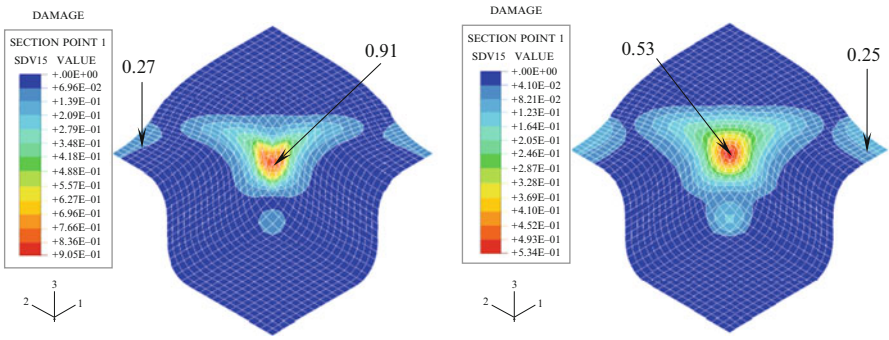


Fig. 25 Damage distributions by ABAQUS with advanced damage model. (**a**) Coupled damage–plasticity. (**b**) Non-coupled damage–plasticity

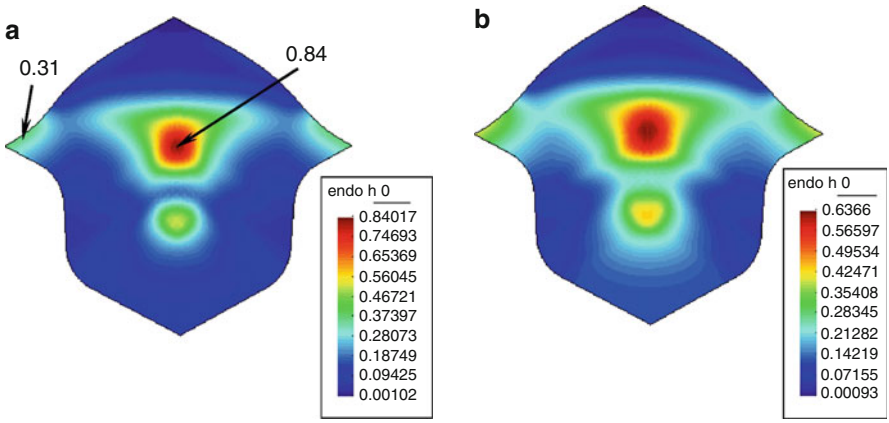


Fig. 26 Damage distributions by PIA with simplified strain-based damage model. (**a**) Coupled damage–plasticity. (**b**) Uncoupled damage–plasticity

SWIFT Stamping Simulation (Steel DC04)

This example is treated experimentally by CETIM and numerically by using the commercial code STAMPACK and the simplified PIA. Only eight steps are used for the PIA calculations. The geometry is presented in Fig. 27, and the material and process parameters are as follows: punch diameter $D = 33$ mm; die diameter $D_M = 35.2$ mm; friction coefficient $\mu = 0.144$; blank-holding force 500 daN; damage parameters for PIA $D_c = 0.4$, $\varepsilon_R = 0.7$, $\varepsilon_D = 0.2$; punch travel 14 mm; initial sheet diameter 74 mm; sheet thickness $t = 1$ mm; fillet radii $r_P = 5$ mm, $r_M = 4$ mm; Young's modulus $E = 82.377$ GPa; anisotropic coefficients $r_0 = 1.87$, $r_{45} = 1.12$, $r_{90} = 2.02$; and plastic behavior law $\overline{\sigma} = 559.66\left(\overline{\varepsilon}_p + 0.0057\right)^{0.226}$.

In Fig. 28, the FLC curves and FLD diagrams obtained by PIA and STAMPACK (without damage consideration) are presented. It is observed that these two codes give very similar FLD.

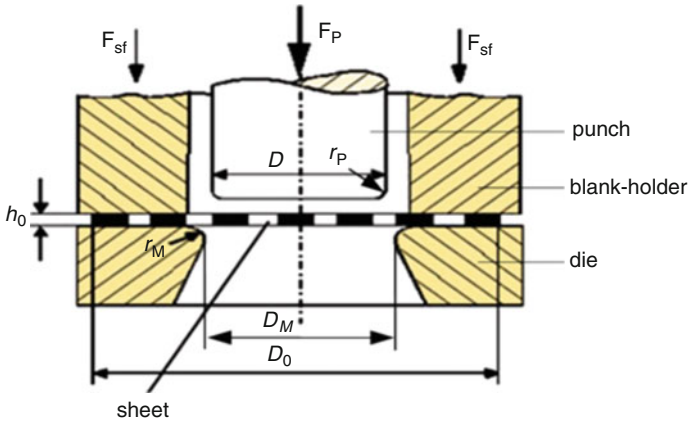


Fig. 27 Geometry of SWIFT stamping test

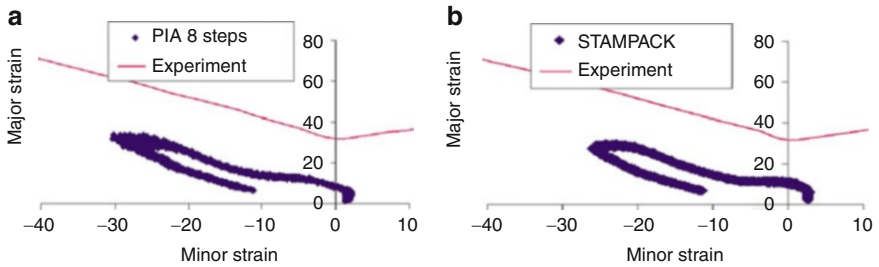


Fig. 28 FLD diagrams obtained by PIA and STAMPACK. (**a**) PIA eight steps. (**b**) STAMPACK

The damage distribution obtained by PIA is presented in Fig. 29. The strong damage zones are situated on the punch radius and on the die's entrance. The numerical results are in good agreement with the experimental results of CETIM. The rupture on the punch's radius and the necking on the die's entrance are clearly shown in the photo of CETIM.

Cold Forging of a Wheel

In this section, the cold forging modeling of a wheel is presented to show the efficiency and limitations of the PIA for the forging process. The results of the PIA including the strain-based damage model are compared to those obtained by the incremental approach ABAQUS/Explicit.

The geometry of the billet and punch is shown in Fig. 30. Due to the symmetry of the wheel, only a quarter of the part section is considered. The symmetric boundary conditions are imposed on the vertical axis and horizontal plane. To compare the two approaches, the final mesh obtained by ABAQUS® is taken for the PIA modeling, containing 1,402 nodes and 1,324 axisymmetric quadrangle elements. The tools are supposed rigid and modeled by analytic rigid wires.

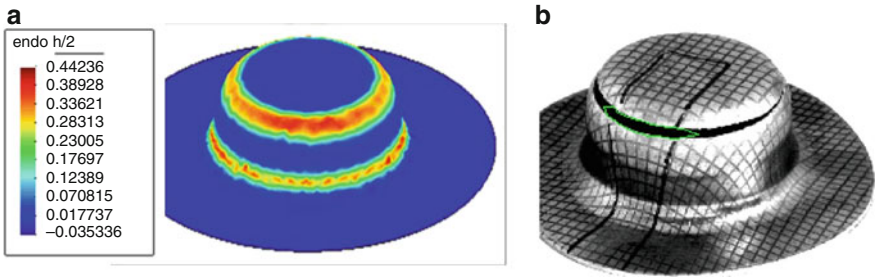


Fig. 29 Damage distribution obtained by PIA and rupture obtained by CETIM test. (**a**) Damage obtained by PIA. (**b**) Rupture in an experimental test

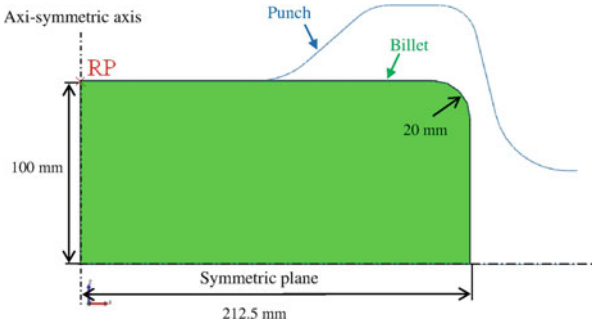


Fig. 30 Geometry of the billet and punch

The material properties of the billet in lead are Young's modulus $E = 17$ GPa, Poisson's ratio $\nu = 0.42$, friction coefficient $\mu = 0.05$, mass density $\rho = 11.35$ g/cm^3, and Hollomon strain–stress curve $\overline{\sigma} = 65.8(\overline{\varepsilon}^{p})^{0.27}$ MPa. The vertical punch travel is 38.8 mm. The damage parameters are $D_c = 0.5$, $\varepsilon_R = 0.315$ and $\varepsilon_D = 0.05$. Only 14 steps are needed for the PIA simulation, the results obtained by using more steps are almost unchanged.

Figure 31 shows the distributions of the equivalent plastic strain obtained by the PIA and ABAQUS®/Explicit. It is observed that the distributions are very similar and the maximal and minimal values are in good agreement between the two approaches.

The distributions of the equivalent stress obtained by the PIA and ABAQUS® are shown in Fig. 32. It is observed that the stress distributions are quantitatively very similar to each other. The maximum equivalent stresses are 57.59 MPa (PIA) and 57.49 MPa (ABAQUS®), respectively, giving an error of 0.2 %.

The damage distributions obtained by PIA and ABAQUS®/Explicit are presented in Fig. 33. It is found that the two approaches give very close damage values in the same zone: $D_{max} = 20.9$ % by the PIA and $D_{max} = 19.7$ % by ABAQUS®.

The PIA leads to a considerable gain of the CPU time compared to ABAQUS®. The ABAQUS®/Explicit uses 2,126 s, but the PIA uses only 460 s, saving 79 % of CPU time.

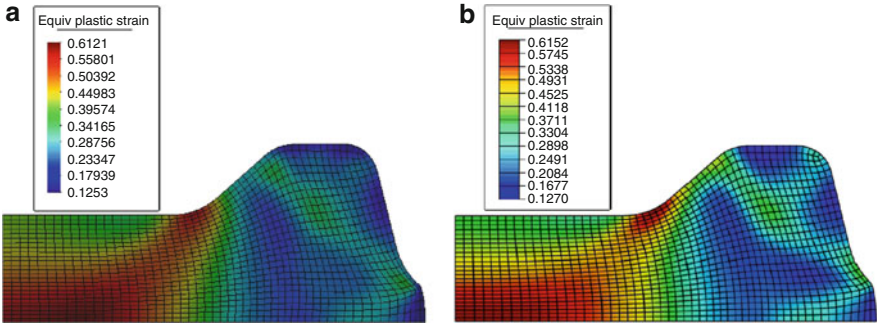


Fig. 31 Equivalent plastic strain distribution obtained by PIA and ABAQUS®. (**a**) Pseudo-inverse approach (14 steps). (**b**) ABAQUS® (339268 increments)

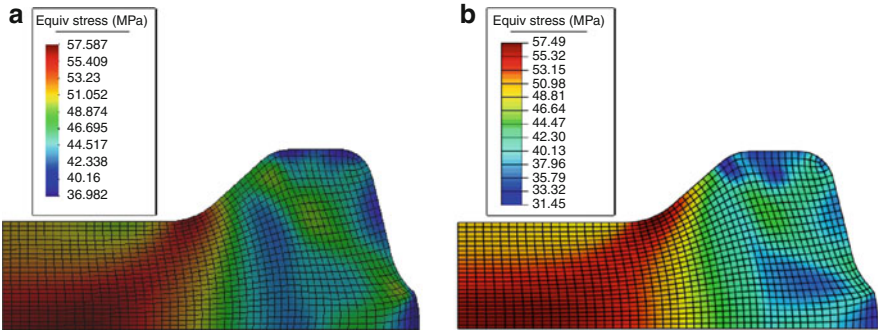


Fig. 32 Equivalent stress distribution obtained by PIA and ABAQUS®. (**a**) Pseudo-inverse approach (14 steps). (**b**) ABAQUS® (339268 increments)

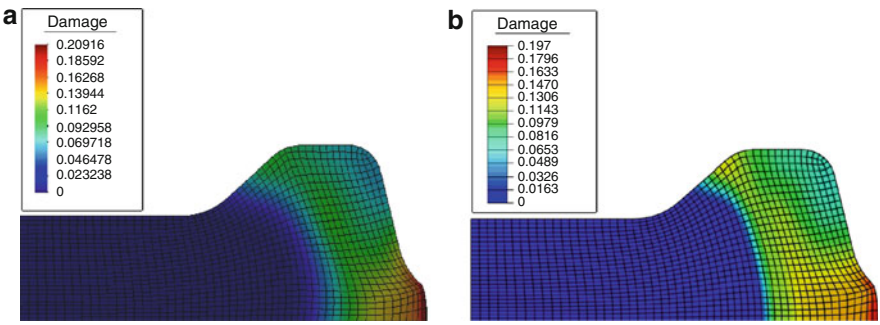


Fig. 33 Damage distributions obtained by PIA and ABAQUS®. (**a**) Pseudo-inverse approach (14 steps). (**b**) ABAQUS® (339268 increments)

Forming Process Optimization Using IA and PIA

General Aspects of the Forming Process Optimization

In a forming process, the original problem is usually a design or optimization problem. Firstly, its solution requires an accurate and efficient simulation of the process in multistages by an incremental or inverse method, taking into account the uncertainties, the affecting parameters, and the fine identification of material behaviors and interfaces. Secondly, the process and part geometry must be parameterized to reduce at maximum the computation time. Thirdly, global, robust, multi-objective, and parallel optimization algorithms should be used to find the optimal process and shape parameters.

The process optimization can largely improve the formability of the material and the robustness of the process. The combination of a numerical forming solver with an optimization algorithm allows an automatic design and control of process parameters, such as the material properties, the holding forces, the punch velocity and force, the geometry of the tools and initial billets, the addendum surfaces, the number and shapes of the forming tools, the friction aspects, the thermal effects, etc.

In the sheet forming field, many studies were presented on the optimization of forming process parameters such as the blank holding forces, the drawbead restraining forces, etc. (Jansson et al. 2005; Shim and Son 2000). Gelin et al. (2005) presented their works on the optimal design and control strategies for the sheet forming and tube hydroforming processes. Many works were done on the optimization of geometrical parameters such as the initial blank shape and the binder surface (Azaouzi et al. 2008, among others). Schenk and Hillmann (2004) proposed an approach for the design and optimization of addendum surfaces by changing the profile of the protection walls and the drawbead restraining forces. More recently, Dong et al. (2007) proposed an automatic procedure for the design and optimization of addendum surfaces by using the fast IA solver and the OpenCascade (2006) free library.

In the forging field, Kobayashi et al. (1989) firstly developed the backward tracing method for the preform design shape. Other groups worked on this method later and used it for the optimization procedure (Han et al. 1993). Zhao et al. (1997) presented an optimization method for the preform die shape design in metal forming processes. Fourment et al. (1996) and Vieilledent and Fourment (2001) made a great progress in this field. They developed shape sensitivity methods for the optimization of nonsteady-state forging processes; the preform shape was defined by B-spline curve taking the control points as design variables. Zhao et al. (2004) presented their studies on the multiple objective preform die shape optimal design by using the forward simulation and sensitivity analysis. Meng et al. (2010) worked on the multi-objective optimization of multistage forging by using advanced thermo-viscoplasticity-damage models and meta-models to optimize the tool shapes. Castro et al. (2010) worked on the optimization of shape and process parameters in metal forging using the genetic algorithms.

Halouani et al. (2012b) developed a fast forging solver called PIA for the multi-objective optimization of tool preform shapes.

For most of nongradient optimization algorithms (such as the response surface, the genetic method or simulated annealing), an important step is to carefully select a number of sampling points by a design of experiments (DOE). This selection has a great influence on the efficiency and accuracy of the optimization procedure. The DOE consists in selecting assessment of points in the design space. The difficulty is how to use the minimum points to obtain the best distribution of the sampling points. Several DOE can be found in the literature (Myers and Montgomery 2002). The well-known methods are the factorial design, central composite design, Latin hypercube, D-optimal, Box–Behnken, etc. A space-filling Latin hypercube design (LHD) is a good and popular DOE strategy for constructing meta-models from deterministic computer experiments (such as FE simulations) (McKay et al. 1979; Santner et al. 2003).

To limit the number of forming simulations, surrogate meta-models are often used to construct an approximate response surface based on the real simulation results for the optimal solution searching. In the literature, there are the moving least square method (Breitkopf et al. 2005; Naceur et al. 2010), the Kriging method (Emmerich et al. 2006), the diffuse approximation (Nayrolles et al. 1992), etc.

Since the metal forming processes involve very complicated phenomena, the multi-objective optimization with several constraints should be considered. The nongradient optimization algorithms are often adopted to avoid the gradient computation, to have a robust searching procedure, and to find global optimal solutions. Among the stochastic methods, the genetic algorithms and the simulated annealing algorithms are largely used (Fourment et al. 1996; Castro et al. 2010; Meng et al. 2010) to determine the Pareto front points and then to find the optimal solution according to other technological constraints. However, these algorithms are time-consuming; it is indispensible to reduce the number of design variables and to use a fast forming solver (Halouani et al. 2012b).

Optimization Procedure for Forming Processes

An optimization procedure comprises four steps: defining the objective functions, selecting the design variables, defining constraint functions, and finding the optimal design variables. The first three steps are denoted as the optimization "modeling." The fourth step is the optimization "solving" problem.

Design Variables

In metal forming processes, the design variables can be divided into geometrical, material and process-related variables. For the workpiece, the geometrical parameters are their shapes and dimensions; for the tools, the parameters are related to the die and punch geometries, including the holding part and drawbeads. The material parameters concern the Young's modulus, Poison coefficient, hardening behavior, anisotropy, damage, viscosity, etc. The process variables include the holding forces, punch travel and velocity, temperatures, friction, etc.

The shape optimization involves much more design variables than the process optimization. Since the computation time strongly depends on the number of design variables, so it is indispensible to parameterize the tools' geometries. This parameterization can be made by using segments and radii for the simple geometry such as the initial billet and final part (Meng et al. 2010). For more complicated geometry such as preforms, the B-spline curves and surfaces are adopted (Halouani et al. 2012).

Objective Functions

The optimization targets can be defined by multi-objective functions:

$$\begin{aligned} &\min[f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots] \\ &\boldsymbol{x} = < x_1, x_2, \ldots, x_n >^{\mathrm{T}} \quad ; \quad x_{iL} \le x_i \le x_{iU}; \quad i = 1, 2, \ldots, n \end{aligned} \tag{60}$$

where $f_i(x)$ are the objective functions, x_i the design variables, and x_{iL} and x_{iU} the lower and upper bounds of the design variables.

Metal forming processes are very complicated, so the optimization procedure often involves several objective functions which depend on the forming process:

- Deep drawing: the objectives can be to minimize the thickness variation, reduce the number of forming stages, improve the surface aspect, minimize the springback, prevent the wrinkling or necking, minimize the blank weight, control the punch force, etc. The following objective function was proposed by Naceur et al. (2001) to minimize the thickness variation and avoid the necking and wrinkling:

$$f = \min \frac{1}{N_{elt}} \sum_{e=1}^{N_{elt}} \left(\frac{h^e - h^0}{h^0} \right)^p \tag{61}$$

where h^0 is the initial sheet thickness, h^e is the final thickness of an element, and p is a positive pair integer ($p = 2, 4, \ldots$).

- Forging: the objectives can be to optimize the grain sizes, reduce the punch force or forging energy, minimize the strain variance, avoid the folding, etc. The following objective function was used to minimize the strain variance (Meng et al. 2010; Halouani et al. 2012b):

$$f = min \frac{1}{V_t} \sum_{i=1}^{N_{elt}} V_i \left(\bar{\varepsilon}_i^{\mathrm{p}} - \bar{\varepsilon}_{\mathrm{avg}}^{\mathrm{p}} \right)^2 \quad \text{with} \quad \bar{\varepsilon}_{\mathrm{avg}}^{\mathrm{p}} = \frac{1}{V_t} \sum_{i=1}^{N_{elt}} V_i \bar{\varepsilon}_i^{\mathrm{p}} \tag{62}$$

where $\bar{\varepsilon}_i^{\mathrm{p}}$ is the equivalent plastic strain of the element i, $\bar{\varepsilon}_{\mathrm{avg}}^{\mathrm{p}}$ is the average equivalent plastic strain, V_i is the volume of the element i, and V_t is the total volume.

Constraint Functions

The constraints and objective functions are related to each other in the sense that they are often exchangeable. In an optimization modeling, one should decide which

quantity is selected as objective, which as constraint. For example, to avoid excessive thickness uniformity, the objective function (60) can be replaced by a constraint on the thinning and thickening. The implicit constraint functions are defined as follows:

$$g_i(\boldsymbol{x}) \leq 0 \qquad i = 1, 2, \ldots, n \tag{63}$$

In the case of the deep drawing, it is not allowed that the strain states exceed the forming limit curve (FLC). The constraint can be that all FLD points are situated below the FLC curve. In the case of forging, the constraint functions can be the conditions on the damage, wrinkling, filling, and volume (Meng et al. 2012; Halouani et al. 2012b): the maximal damage should be inferior to the damage threshold, the contour of the forged part should not have sudden changes (folding), the volume should remain constant, etc.

Optimization Algorithms

Five types of algorithms are often used for the forming process optimization: iterative algorithms, evolutionary and genetic algorithms, approximate optimization algorithms, adaptive optimization algorithms, hybrid and combined optimization algorithms.

(a) Iterative Algorithms

Optimization of metal-forming processes can be performed by using classical iterative algorithms (SIMPLEX, conjugate gradient, SQP, BFGS, etc.). These algorithms usually require the sensitivities of the objective function and constraint functions with respect to the design variables. In the case of metal forming, the FEM calculations are very time-consuming and may give inaccurate sensitivities. Generally, the iterative algorithms are inadaptable to the multi-objective optimization and may get trapped in local optima.

(b) Evolutionary and Genetic Algorithms

Genetic and evolutionary algorithms are promising because of their tendency to find the global optimum and the possibility for parallel computing. Furthermore, they do not require the sensitivity computation. However, the large number of function evaluations is a serious drawback. The non-dominated sorting genetic algorithm NSGA-II (Deb 2000) is appealing to many authors for the metal forming optimization.

(c) Approximate Algorithms

The response surface method (RSM) is a well-known representative of approximate optimization algorithms. RSM is based on fitting a low-order polynomial meta-model through real response points, which are obtained by running FEM calculations for some chosen design variable settings. Next to the RSM, other meta-modeling techniques are Kriging and neural networks. Allowing the parallel computing and avoiding the sensitivity calculation, the approximate optimization is a preferred technique for many authors. A disadvantage of these methods is that the result is an approximate optimum rather than the real global optimum.

(d) Adaptive Algorithms

Adaptive algorithms are incorporated within FEM codes and generally optimize the time-dependent load paths of the metal-forming process during each increment of the FEM calculation. For example, to optimize the time-dependent pressure load path in hydroforming, one should keep a sufficient pressure to avoid the wrinkling. When such risk is detected during a loading increment, the pressure is increased in the next increment to avoid wrinkles in the final product. The advantage of these algorithms is that the optimum is obtained in only one FEM simulation. However, the access inside the FEM software is necessary, and only time-dependent design variables can be considered. These disadvantages seriously limit the general applicability of these algorithms.

(e) Hybrid and Combined Algorithms

Many researchers tried to combine the advantages of different optimization algorithms. Within the metal forming community, most authors used approximate algorithms to establish a meta-model and adopted an iterative algorithm to find the optimum. Some others constructed relatively noisy meta-models (i.e., many local optima) by using the Kriging and neural networks techniques and then used a global genetic algorithm to solve the optimization problem. For adaptive optimization algorithms, certain choose iterative algorithms, others genetic ones. It is also possible to enhance an evolutionary algorithm with the information provided by a meta-model-based approximate algorithm to make it more efficient and to overcome the difficulty of the large number of function evaluations.

(f) Simulated Annealing Method

This stochastic optimization method was developed by Kirkpatrick et al. 1983. This method is derived from an analogy with the slow cooling phenomenon of a molten body, which leads to a low-energy solid state. It should slowly lower the temperature, marking long plateaus so that the body reaches the thermodynamic equilibrium at each temperature plateau. For materials, this low energy manifests itself by obtaining a regular structure, such as crystals in the steel. The analogy used by the simulated annealing is to search a physic state p minimizing the energy function $\Phi(p)$. Simulated annealing usually exploits the criteria defined by the algorithm of Metropolis et al. (1953) for the acceptance of a solution obtained by perturbation of the current solution. Theoretical studies show that the simulated annealing algorithm converges to a global optimum under certain conditions. The main drawback is related to the choice of numerous annealing parameters such as the initial temperature, the decay rate of the temperature, the stopping criteria, or the lengths of the temperature plateaus. These parameters are often chosen empirically.

Preform Design and Optimization

The two-stage cold forging of an axisymmetric wheel is simulated and optimized. The forging process is composed of a preforming stage using a preform tool (Fig. 34) and forging stage using the final tool given by the desired part (Fig. 35).

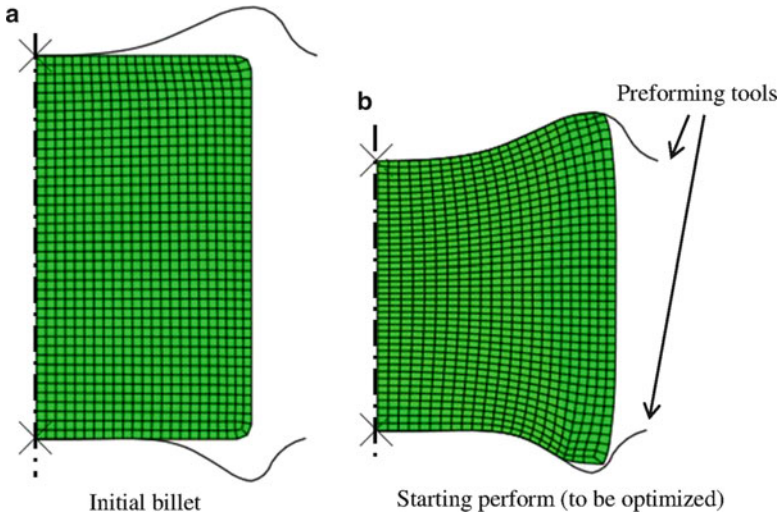


Fig. 34 Preforming stage using starting tool preform. (**a**) Initial billet. (**b**) Starting tool preform (to be optimized)

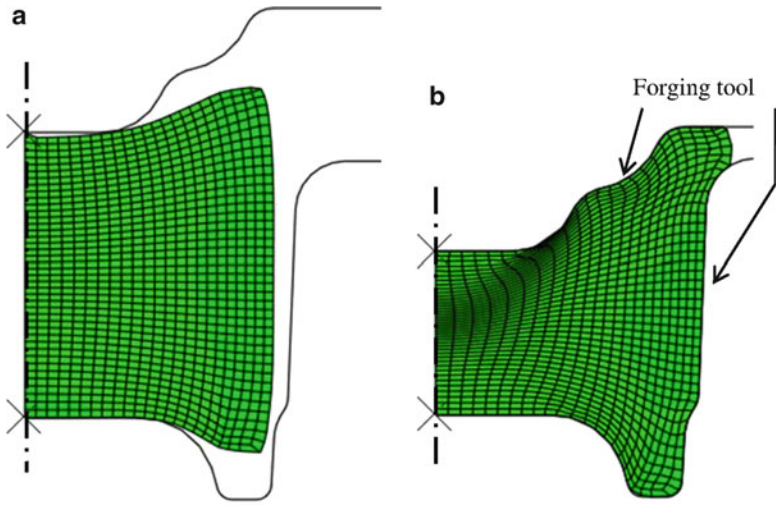


Fig. 35 Forging stage using preformed billet and final forging tools. (**a**) Billet obtained by performing stage. (**b**) Desired final forged part

The initial billet is a cylinder (height = 80 mm, radius = 45 mm). The geometry of the billet, the starting preform shape, and the final tools are shown in Figs. 34 and 35. The axisymmetric boundary conditions are imposed. The section is meshed with 830 nodes and 774 quadrilateral elements. The tools are supposed rigid and modeled by analytic rigid wires. The billet material is the lead: Young's modulus

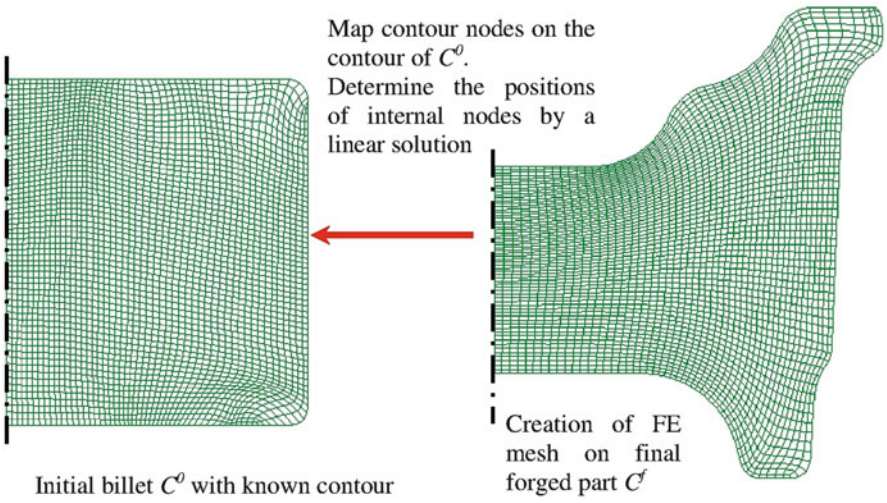


Fig. 36 FE mesh mapped from final part to initial billet

$E = 17$ GPa, Poisson's ratio $\nu = 0.42$, friction coefficient $\mu = 0.05$, and Hollomon tensile curve $\overline{\sigma} = 65.8(\overline{\varepsilon}^{\mathrm{p}})^{0.27}$ MPa.

In this work, the starting preform is created as follows:

1. Mesh mapping from the final forged part C^f to the initial billet C^0. A FE mesh is created on the known C^f, and the nodes at the contour of C^f are mapped on the contour of C^0; the positions of other nodes (interior nodes) in C^0 are determined by a linear solution with the imposed displacements on the contour (Fig. 36).
2. Creation of the geometrically proportional FE mesh between C^0 and C^f (Fig. 37): $\boldsymbol{X}_{\mathrm{p}}^1 = \boldsymbol{X}^f - (\boldsymbol{X}^f - \boldsymbol{X}^0)/2$.
3. Generation of a starting tool preform. The B-spline curve of this preform should fit well the proportional mesh contour except for the free surface part (Fig. 37). The right extremity (F) of the punch curve has the same height than that of the left extremity (E) and the same horizontal position than the maximal radial position (G) of the proportional preform. This choice gives a notable gap between the punch curve and preform shape on the zone B, but this gap has a little influence on the preform optimization. The B-spline curve of the lower die can be obtained by the same method.

A B-spline curve of the preform is defined by a polygonal contour having $n + 1$ ($n + 1 \geq 4$) control points $C_1 \ldots C_{n+1}$. These control points can be active or passive. Figure 38 shows the punch shape curve with seven control points; only their vertical displacements are taken as geometrical parameters to reduce the number of design variables. C_2 and C_6 are the passive points having the same vertical positions than C_1 and C_7 in order to keep the horizontal tangents at C_1 and C_7; other five are active points giving only five optimization design variables. The

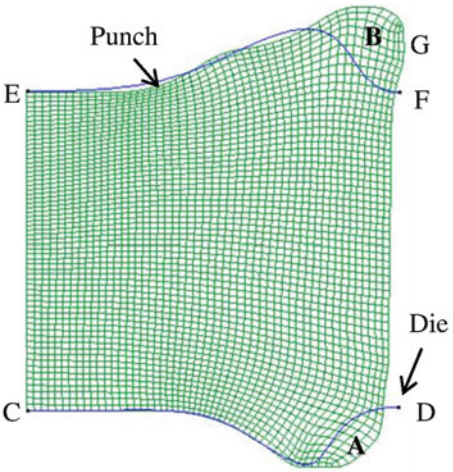


Fig. 37 Generation of proportional mesh and B-spline curves of the starting perform tools

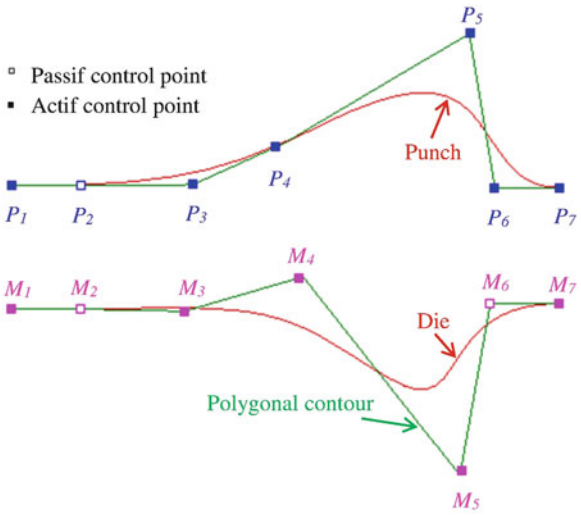


Fig. 38 Control points of the starting B-spline curves for preform tools

die shape curve is defined in same manner. Finally, there are only ten optimization design variables. These starting B-spline curves are then modified in the optimization loop to minimize the objective functions.

The validation of the PIA is done by using the software ABAQUS/Explicit. The punch travels are 23.7 mm in the preforming stage and 24.9 mm in the forming stage. The PIA is used firstly between the preform and initial billet, and then between the final part and preform. According to our numerical tests, the PIA results are no longer sensitive to the number of steps, from 11 steps for the preforming stage and from 12 for the forging stage.

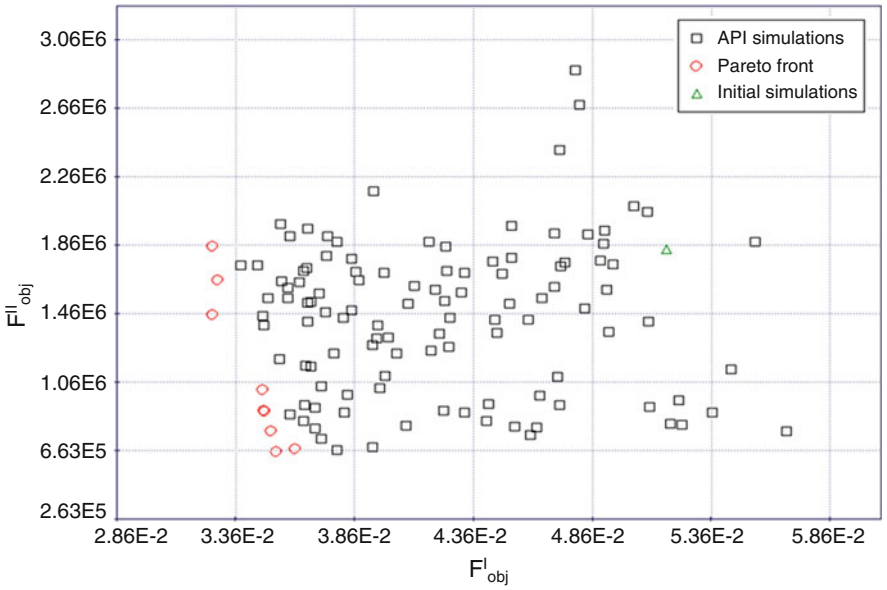


Fig. 39 Initial Pareto points obtained by using the real simulation results

In a multi-objective optimization, the concept of the best design is replaced by the concept of dominant design. This set of dominant designs is called Pareto frontier. The designer should find a good compromise for all objective functions according to other technical or economic constraints.

Two objective functions are adopted to minimize the plastic strain variation (Eq. 62) and the maximal punch force. A simulated annealing optimization algorithm called MOSA is used (ModeFRONTIER™ 4, User Manual). The initial Pareto points for this two-objective optimization problem are obtained by using 200 PIA simulations (or 200 iterations in the MOSA optimization loop). The distribution of these Pareto points (marked by red circles) is presented in the objective function plan ($F^{\mathrm{I}}_{\mathrm{obj}}$ and $F^{\mathrm{II}}_{\mathrm{obj}}$, Fig. 39).

The multi-objective optimization algorithms require a large amount of simulations, so it will be very expensive to minimize the objective functions entirely using real FE simulations. The Kriging method is adopted to build the surrogate meta-model for the two objective functions. Kriging method is a nonparametric interpolation model which interpolates the responses exactly at all sampling points. Figure 40 shows the surrogated metal-models of the two objective functions $F^{\mathrm{I}}_{\mathrm{obj}}$ and $F^{\mathrm{II}}_{\mathrm{obj}}$ using Gaussian Kriging method.

To get the optimal design values after building the meta-model, the genetic optimization algorithm called NSGA-II in the software modeFRONTIER™ is used. The distribution of the Pareto points obtained by using NSGA-II algorithm coupled with the Gaussian Kriging model is shown in Fig. 41. During the optimization, lots of new solutions are generated. This enables to have more optimal values on the

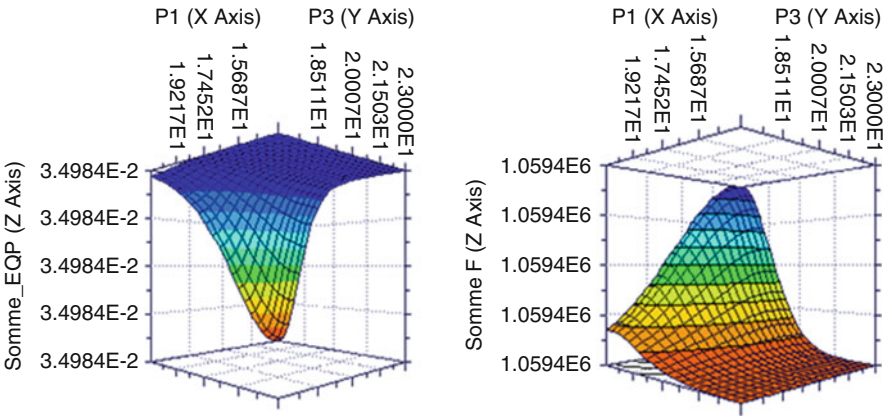


Fig. 40 Kriging surrogate meta-model of $F^{\mathrm{I}}_{\mathrm{obj}}$ and $F^{\mathrm{II}}_{\mathrm{obj}}$ in function of two design variables (vertical displacements of P1 and P3 in Fig. 38)

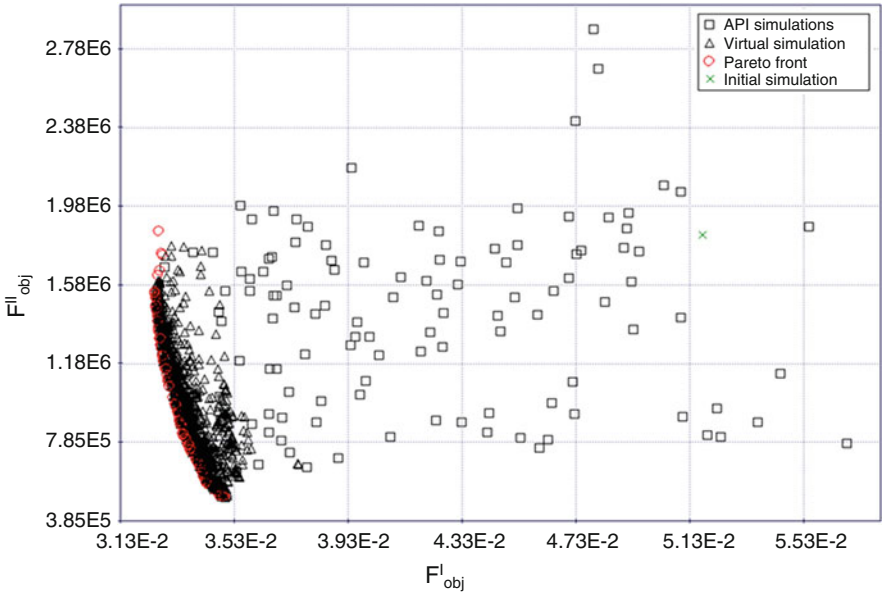


Fig. 41 Pareto front given by NSGA-II/Kriging

Pareto front, giving the final optimal solutions (marked by red circles in Fig. 41). It can be seen that the two objective functions largely decrease during the optimization ($F^{\mathrm{I}}_{\mathrm{obj}} = 0.052 \rightarrow 0.035$, $F^{\mathrm{II}}_{\mathrm{obj}} = 1842733.7 \rightarrow 504926.9$, Fig. 41), giving 33 % of reduction for the equivalent plastic strain and 72 % of reduction for the punch force.

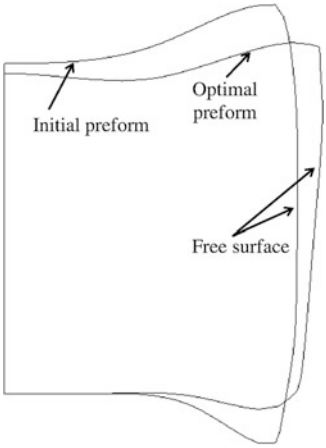


Fig. 42 Initial and optimal preform shapes

Figure 42 shows a comparison between the initial preform shape and the optimal one; a better result was achieved by using the proposed concept.

The CPU times for this two-stage forging simulation are also compared: the PIA uses only 285 s, but the ABAQUS/Explicit uses 1,453 s (5.1 times). Therefore, in the optimization procedure using 200 API simulations, the gain of the CPU time becomes enormous.

Conclusions

In this chapter, several simplified and practical approaches for damage prediction in metal-forming process modeling and optimization were presented.

- The IA exploits the knowledge of the final part shape and executes the calculation from the final part mesh to the initial blank or billet. The assumptions of proportional loading and simplified tool actions make the IA calculation very fast. This method gives fairly good strain estimation for the deep drawing, hydroforming, and cold forging processes, but poor stress evaluation. A simplified total strain-based damage model is implemented into the IA without considering the coupling effect between the plasticity and damage. The IA can be used as a numerical tool for the preliminary design and optimization in forming processes.
- The PIA is a good compromise between the IA and incremental approaches. The contact treatment is avoided by using some simplified tool actions as in the IA. In order to consider the strain path, some intermediate configurations are determined geometrically and corrected by using a surface minimization method or a free surface method, allowing very large strain increments. A 3D strain-based damage model in a rate form is adopted and coupled with the plasticity. An efficient direct scalar algorithm for damage–plasticity integration enables to take

into account the loading history so obtain good stress estimation. The PIA combines the advantages of the IA and incremental approaches: it gives better stress estimation then that of the IA and it is much faster than the incremental approach. The PIA is an efficient numerical tool for the damage prediction and forming processes optimization.

- The 3D strain damage model is based on the assumptions of the hardening saturation after the damage threshold and the proportional loading condition. It is implemented into the IA in a total form without considering its coupling with the plasticity. In the PIA, the model coupling the damage and plasticity is formulated and implemented in a rate form. The efficient direct scalar algorithm of plasticity (DSAP) for the integration of coupled damage–plasticity is developed to take into account the loading history. Using the notion of the equivalent stress, the constitutive equations in stress vectors are transformed into a scalar equation in which the equivalent stresses can be obtained by the tensile curve. Thus, the plastic multiplier can be obtained directly without iterations. This DSAP enables to largely reduce the CPU time and to avoid divergence problems even though for very large strain increments.
- Some optimization algorithms are combined with the IA using an integrated material law or with the PIA using the DSAP. These simplified methods make the optimization very efficient and robust, allowing to use the time-consuming optimization algorithms (such as genetic algorithm, simulated annealing method, etc.) in order to find globally optimal Pareto points for multi-objective functions.

Further research investigations will be devoted to continuously improve the original approaches IA and PIA. In the future, the authors will implement an adaptive meshing algorithm in the PIA in order to deal with complex parts undergoing very large plastic strains. In a forthcoming research, the authors will also include viscoplastic and thermomechanical material models in the PIA for the hot forging simulation. The fast PIA will be used to optimize tool preform shapes and other parameters of the forging process.

References

N. Aravas, The analysis of void growth that leads to central burst during extrusion. J. Mech. Phys. Solid **34**, 55–79 (1986)

M. Azaouzi, H. Naceur, A. Delameziere, J.L. Batoz, S. Belouettar, An heuristic optimization algorithm for the blank shape design of high precision metallic parts obtained by a particular stamping process. Finite Elem. Anal. Des. **44**, 842–850 (2008)

F. Barlat, J.C. Brem, J.W. Yoon, K. Chung, R.E. Dick, D.J. Lege, F. Pourboghrat, S.–.H. Choi, E. Chu, Plane stress yield function for aluminum alloy sheets–Part I: theory. Int. J. Plast. **19**, 1297–1319 (2003)

J.L. Batoz, G. Dhatt, *Modélisation des structures par éléments fini*, vol. 1, 3 (Edition HERMES, Paris, 1990)

P. Breitkopf, H. Naceur, A. Rassineux, P. Villon, Moving least squares response surface approximation: formulation and metal forming applications. Comput. Struct. **83**(17–18), 1411–1428 (2005)

M. Brunet, F. Sabourin, S. Mguil-Touchal, The prediction of necking and failure in 3d sheet forming analysis using damage variable. J. Phys. III **6**, 473–482 (1996)

F. Castro Catarina, C. António Carlos, C. Sousa Luisa, Pareto-based multi-objective hot forging optimization using a genetic algorithm, in *2nd International Conference on Engineering Optimization*, Lisbon, 2010

J.L. Chaboche, Continuum damage mechanics I-general concepts. II-damage growth, crack initiation, and crack growth. ASME Trans. J. Appl. Mech. **55**, 59–72 (1988)

M. Chebbah, H. Naceur, A. Gakwaya, A fast algorithm for strain prediction in tube hydroforming based on one-step inverse approach. J. Mater. Process. Technol. **211**(11), 1898–1906 (2011)

A. Cherouat, Y.Q. Guo, K. Saanouni, Y.M. Li, K. Debray, G. Loppin, Incremental versus inverse numerical approaches for ductile damage prediction in sheet metal forming. Int. J. Form. Process. **7**(1–2), 99–122 (2004)

J.P. Cordebois, P. Ladevèze, Necking criterion applied in sheet metal forming, in *Plastic Behavior of Anisotropic Solids*, ed. by J.P. Boehler (Editions CNRS, Paris, 1985)

K. Deb, An efficient constraint handling method for genetic algorithms. Comput. Method Appl. Mech. Eng. **186**(2–4), 311–338 (2000)

M. Dong, K. Debray, Y.Q. Guo, J.L. Shan, Design and optimization of addendum surfaces in sheet metal forming process. Int. J. Comput. Method. Eng. Sci. Mech. **8**(4), 211–222 (2007)

M. Emmerich, K. Giannakoglou, B. Naujoks, Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels. IEEE Trans. Evolut. Comput. **10**(4), 421–439 (2006)

L. Fourment, T. Balan, J.L. Chenot, Optimal design for nonsteady-state metal forming processes – I shape optimization method. Int. J. Numer. Method Eng. **39**(1), 33–65 (1996)

J.C. Gelin, J. Oudin, Y. Ravalard, An imposed finite element method for the analysis of damage and ductile fracture in cold metal forming processes. Ann CIRP **34**(1), 209–213 (1985)

J.C. Gelin, C. Labergère, S. Thibaud, Recent advances in process optimization and control for the design of sheet and tube hydroforming processes, in Numisheet, Detroit, edited by L.M. Smith et al., Vol. A, pp. 825–830, 2005

Y.Q. Guo, J.L. Batoz, J.M. Detraux, P. Duroux, Finite element procedures for strain estimations of sheet metal forming parts. Int. J. Numer. Method Eng. **30**, 1385–1401 (1990)

Y.Q. Guo, W. Gati, H. Naceur, J.L. Batoz, An efficient DKT rotation free shell element for springback simulation in sheet metal forming. Comput. Struct. **80**(27–30), 2299–2312 (2002)

Y.Q. Guo, Y.M. Li, F. Bogard, K. Debray, An efficient pseudo-inverse approach for damage modeling in the sheet forming process. J. Mater. Process. Technol. **151**(1–3), 88–97 (2004)

A.L. Gurson, Porous rigid-plastic materials containing rigid inclusions – yield function, plastic potential and void nucleation, in *Proceedings of the Conference on Fracture*, vol. 2, pp. 357–364, 1977

A. Halouani, Y.M. Li, B. Abbès, Y.Q. Guo, An axisymmetric inverse approach for cold forging modelling. Eng. Lett. **18**(4), 376–383 (2010)

A. Halouani, Y.M. Li, B. Abbès, Y.Q. Guo, Simulation of axi-symmetrical cold forging process by efficient pseudo inverse approach and direct algorithm of plasticity. Finite Elem. Anal. Des. **61**, 85–96 (2012a)

A. Halouani, Y.M. Li, B. Abbès, Y.Q. Guo, F.J. Meng, C. Labergere, P. Lafon, Optimization of forging preforms by using pseudo inverse approach. Key Eng. Mater. **504–506**, 613–618 (2012b)

C.S. Han, R.V. Grandhi, R. Srinivasan, Optimum design of forging die shapes using nonlinear finite element analysis. AIAA J. **31**(4), 774–781 (1993)

P. Hartley, S.E. Clift, J. Salimi, C.E.N. Sturgess, I. Pillinger, The prediction of ductile fracture initiation in metalforming using a finite element method and various fracture criteria. Res. Mech. **28**, 269–293 (1989)

T. Jansson, A. Anderson, L. Nilsson, Optimization of draw-in for an automotive sheet metal part: an evaluation using surrogate models and response surfaces. J. Mater. Process. Technol. **159**(3), 426–434 (2005)

M. Jansson, L. Nilsson, K. Simonsson, Tube hydroforming of aluminium extrusions using a conical die and extensive feeding. J. Mater. Process. Technol. **198**(1–3), 14–21 (2008)
N. Kim, S. Kobayashi, Preform design in H-shape cross section axisymmetric forging by finite element method. Int. J. Mach. Tool Manuf. **30**, 243–268 (1990)
S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by Simulated Annealing. Science, New Series. **220**(4598), 671–680 (1983)
S. Kobayashi, S.I. Oh, T. Altan, *Metal Forming and Finite Element Method* (Oxford University Press, Oxford, 1989)
C.H. Lee, H. Huh, Blank design and strain estimation for sheet metal forming processes by a finite element inverse approach with initial guess of linear deformation. J. Mater. Process. Technol. **82**, 145–155 (1998)
H. Lee, K.E. Peng, J. Wang, An anisotropic damage criterion for deformation instability and its application to forming limit analysis of metal plates. Eng. Fract. Mech. **21**(5), 1031–1054 (1985)
J. Lemaître, J.L. Chaboche, *Mechanics of Solid Materials* (Cambridge University Press, Cambridge, 1990)
Y.M. Li, B. Abbès, Y.Q. Guo, Two efficient algorithms of plastic integration for sheet forming modeling. ASME J. Manuf. Sci. Technol. **129**, 698–704 (2007)
J.F. Mariage, K. Saanouni, P. Lestriez, A. Cherouat, Numerical simulation of ductile damage in metal forming processes: a simple predictive model. Part I. Theoretical and numerical aspects. Int. J. Form. Process **5**(2–3–4), 363–376 (2002)
K. Mathur, P. Dawson, Damage evolution modeling in bulk forming processes, in *Computational Methods for Predicting Material Processing Defects* (Elsevier, Predeleanu, 1987)
M.D. McKay, W.J. Conover, R.J. Beckman, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics **21**, 239–245 (1979)
F.J. Meng, C. Labergere, P. Lafon, Methodology of the shape optimization of forging dies. Int. J. Mater. Form **3**(Suppl 1), 927–930 (2010)
F. Meng, Multi-objective optimization of several stages forging by using advanced numerical simulation and Meta-model, PhD thesis, Université de Technologie de Troyes. (2012)
N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087–1092 (1953)
R. Myers, D. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd edn. (Wiley, New York, 2002). ISBN 0-471-41255-4
H. Naceur, Optimisation de forme de structures minces en grandes transformations, Edition EUE, ISBN-13: 978-613-1-54700-3, p. 240, (2010)
H. Naceur, Y.Q. Guo, W. Gati, New enhancements in the inverse approach for the fast modeling of autobody stamping process. Int. J. Comput. Eng. Sci. **3**(4), 355–384 (2002)
H. Naceur, Y.Q. Guo, J.L. Batoz, C. Knopf-Lenoir, Optimization of drawbead restraining forces and drawbead design in sheet metal forming process. Int. J. Mech. Sci. **43**(10), 2407–2434 (2001)
H. Naceur, Y.Q. Guo, S. Ben-Elechi, Response surface methodology for design of sheet forming parameters to control springback effects. Comput. Struct. **84**, 1651–1663 (2006)
B. Nayrolles, G. Touzot, P. Villon, Generalizing the Finite Element Method: Diffuse approximation and diffuse elements. Comput. Mech. **10**, 307–318 (1992)
E. Onate, M. Kleiber, Plastic and viscoplastic flow of void containing metal – applications to axisymmetric sheet forming problem. Int. J. Numer. Meth. Eng. **25**, 237–251 (1988)
P. Picart, O. Ghouati, J.C. Gelin, Optimization of metal forming process parameters with damage minimization. J. Mater. Process. Technol. **80–81**, 597–601 (1998)
G. Rousselier, Ductile fracture models and their potential in local approach of fracture. Nucl. Eng. Des. **105**(1), 97–111 (1987)
K. Saanouni, *Damage Mechanics in Metal Forming. Advanced Modeling and Numerical Simulation* (ISTE/Wiley, London, 2012). ISBN 978-1-8482-1348-7

K. Saanouni, J.L. Chaboche, Computational damage mechanics, application to metal forming, in *Comprehensive Structural Integrity, Chapter 7*, ed. by R. de Borst, H.A. Mang. Numerical and Computational Methods, vol. 3 (Elsevier, Amsterdam, 2003)

K. Saanouni, K. Nesnas, Y. Hammi, Damage modeling in metal forming processes. Int. J. of Damage Mechanics. **9**(3), 196–240 (2000)

T. Santner, B. Williams, W. Notz, *The Design and Analysis of Computer Experiments* (Springer, New York, 2003)

O. Schenk, M. Hillmann, Optimal design of metal forming die surfaces with evolution strategies. Comp. Struct. **82**, 1695–1705 (2004)

H.B. Shim, K.C. Son, Optimal blank shape design by sensitivity method. J. Mater. Process. Technol. **104**, 191–199 (2000)

J.C. Simo, R.L. Taylor, A return mapping algorithm for plane stress elastoplasticity. Int. J. Numer. Method Eng. **22**, 649–670 (1986)

D. Vieilledent, L. Fourment, Shape optimization of axisymmetric preform tools in forging using a direct differentiation method. Int. J. Numer. Method Eng. **52**, 1301–1321 (2001)

G. Zhao, E. Wright, R.V. Grandhi, Preform die shape design in metal forming using an optimization method. Int. J. Numer. Method Eng. **40**(7), 1213–1230 (1997)

G. Zhao, X. Ma, X. Zhao, R.V. Grandhi, Studies on optimization of metal forming processes using sensitivity analysis methods. J. Mater. Process. Technol. **147**, 217–228 (2004)

Y.Y. Zhu, S. Cescotto, A.M. Habraken, A fully coupled elastoplastic damage modeling and fracture criteria in metal forming processes. J. Meter. Process. Technol. **32**, 197–204 (1992)

Ductile Damage in Metal Forming: Advanced Macroscopic Modeling and Numerical Simulation

25

Khemais Saanouni, Mohamed Hamed, Carl Labergère, and Houssem Badreddine

Contents

K. Saanouni (✉) • M. Hamed • C. Labergère • H. Badreddine
ICD/LASMIS, STMR UMR CNRS 6281, University of Technology of Troyes, Troyes, France
e-mail: khemais.saanouni@utt.fr; mohamed.hamed@utt.fr; carl.labergere@utt.fr; houssem.badreddine@utt.fr

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_37

Abstract

This chapter is dedicated to the presentation of an advanced fully adaptive numerical methodology for virtual sheet and/or bulk metal forming simulation to predict any defects occurrence. First, the detailed formulation of thermodynamically consistent fully coupled, nonlocal constitutive equations is given. Formulated in the framework of the generalized micromorphic continua, the proposed nonlocal constitutive equations account for the main material nonlinearities as the isotropic and kinematic hardening, the thermal exchanges, and the isotropic ductile damage under large inelastic strains. Second, the related numerical aspects required to solve the initial and boundary value problem (IBVP) are presented in the framework of a fully adaptive finite element method. First, the strong and weak forms for the multifunctional IBVP are posed. The well-known static implicit (SI) and dynamic explicit (DE) global resolution schemes are summarized, followed by the detailed presentation of the local integration scheme of the fully coupled constitutive equations. The numerical treatment of the contact and friction is reviewed in the framework of master/slave surfaces method. Finally, some typical examples of sheet and bulk metal forming processes are numerically simulated using the proposed fully adaptive methodology.

Introduction

The main objective of modern metal forming processes is to design robust and lightweight structural components which help reducing carbon dioxide emissions during both the manufacturing process and the future use of the final product, in connection with the new worldwide challenge related to the climate change. Accordingly, the rising demands of customers concern the lightweight design in order to reduce significantly energy consumption and cost efficiency while increasing the structures' service life by enhancing their stability and deformation resistance under various thermomechanical loading paths. These objectives cannot be reached without the help of an "efficient" and "robust" numerical or virtual design methodology based on:

- "Advanced" constitutive equations to describe, as accurately as possible, the main thermomechanical fields and their various interactions (full coupling effects) during their evolution
- "Advanced" numerical tools to predict robustly and accurately the evolution of the deformation processes and the possible defects occurrence during the manufacturing processes or during the use of the final component in any mechanical system

In fact, when formed or machined by large elasto–inelastic strains under room or high temperature, metallic materials undergo a strong localization of inelastic flow which is often at the origin of the initiation, growth, and coalescence of microcracks and/or microvoids usually called the ductile damage. Depending on the geometrical complexity of the forming tools, which define the shape of applied loading path,

this well-known mechanism of ductile fracture may cause the loss of the quality of the formed part by the formation of macroscopic cracks propagating inside this formed part. This ductile damage can be seen as a natural consequence of the large inelastic strain of the solid which itself is strongly dependent on the main thermomechanical phenomena such as the mixed isotropic and kinematic hardening, the heat flux, the various initial and induced anisotropies, and the initial microstructure of the material and its change (texture) under the applied load. Accordingly, the constitutive equations used to simulate and to optimize numerically these metal forming processes should account for these thermomechanical phenomena and their mutual interactions or strong coupling.

Many published works have been devoted to the optimization of bulk and sheet metal forming processes using various more or less simplified approaches. For example, in the forming of thick or thin metallic sheets, the goal is to enhance the capacity of the sheet to carry a large inelastic "homogeneous" strain without any strong localization, giving some through thickness necking prior to a macroscopic crack formation. In engineering practice, the material formability is usually assessed with strain-based forming limit diagrams (FLD) in the case of linear (or proportional) strain loading paths as pioneered by Mariciniak and coworkers (Mariciniak and Kunczynski 1967; Mariciniak et al. 1973). These forming limit diagrams or curves are determined from the experimental measurement of the necking or local fracture onset under linear strain paths using the minor and major principal strain diagram. However, it has been shown (Ghosh and Laukonis 1976; Arrieux et al. 1982; Arrieux and Boivin 1987; Graf and Hosford 1993; Arrieux 1995; Stoughton 2001; Matin et al. 2006; Assempour et al. 2009) that these strain-based forming limit criteria are not efficient when the applied strain path is not linear (or is nonproportional). Unfortunately, in major forming processes, the strain paths supported by deforming material points during the deformation process have been shown neither linear nor monotonic. This is mainly due to the complexity of the tools' (dies, punches) geometry which cause locally reversed strain path exhibiting non-negligible Bauschinger effect. This is clearly the case in sheet forming processes, for which the FLD prediction underestimates the failure strain as observed in many works (Chien et al. 2004; Yoshida et al. 2005; Yoshida and Kuwabara 2007; Hora and Tong 2009; Carbonnière et al. 2009; Le Maout et al. 2009 among others). To avoid these drawbacks, some authors proposed to construct the FLD (or FLC) in the stress space instead of the strain space leading to the so-called stress-based forming limit diagram, FLSD (Ghosh and Laukonis 1976; Arrieux et al. 1982; Arrieux and Boivin 1987; Graf and Hosford 1993; Arrieux 1995; Stoughton 2001; Matin et al. 2006; Assempour et al. 2009). However, this approach has been shown to be not efficient for complex combined stress paths (mainly nonproportional loading paths exhibiting additional hardening), where the material hardening is strongly dependent on the shape of the loading path (Yoshida et al. 2005; Yoshida and Kuwabara 2007; Hora and Tong 2009). On the other hand, when the necking takes place somewhere in the sheet, the plane stress assumption, on which is based the FLSD, becomes highly questionable, and the predicted local stress state is less accurate or simply wrong.

Another way, proposed in many works in order to enhance the predictivity of the forming limit curves, consists in completing the yield function, of von Mises or Hill types, by appropriate instability criteria based on the pioneering works by Swift (1952), Storen and Rice (1975), and Bressan and Williams (1983). Most instability theories assume the existence of an initial imperfection with a given geometrical definition, leading to a high sensitivity to the size of such an assumed initial imperfection. On the other hand, the prediction of plastic strain and its value at the final fracture is highly dependent on the used constitutive equations and whether or not they account for nonlinear mixed isotropic and kinematic hardening as well as the ductile damage effect on plastic flow and hardening evolution. In order to avoid this problem, many authors proposed to replace the initial imperfection by using an appropriate ductile damage theory, which allows catching naturally the instability conditions due to the damage initiation without assuming the existence of any initial imperfection (Needleman and Triantafyllidis 1980; Chu 1980; Chu and Needleman 1980; Brunet and Morestin 2001 among many others).

An alternative approach, proposed in the recent last two decades, to predict the localized neck prior to fracture in sheet or bulk metal forming, is the full coupling between the material behavior and the ductile damage using either macroscopic monoscale or micro-macro multiscale modeling approaches as can be found in the recent books dedicated to the damage prediction in metal forming (Dixit and Dixit 2008; Saanouni 2012a, b). Two different kinds of damage theories are used in metal forming problems: the Gurson-based damage theories (Gelin et al. 1985; Cordebois and Ladeveze 1985; Lee et al. 1985; Mathur and Dawson 1987; Onate and Kleber 1988; Hartley et al. 1989; Gelin 1990; Bontcheva and Iankov 1991; Zhu and Cescotto 1991; Brunet et al. 1996, 2005; Picart et al. 1998 among others) and the continuum damage mechanics (CDM)-based theories (Zhu et al. 1992; Zhu and Cescotto 1995; Saanouni et al. 2000, 2001; Villon et al. 2002; Cherouat et al. 2002a, b; Cherouat and Saanouni 2003; Saanouni and Chaboche 2003; Lestriez et al. 2004, 2005; Saanouni et al. 2004, 2008, 2010, 2011; Mariage et al. 2005; Saanouni 2006, 2008, 2012a, b; Chaboche et al. 2006; Cesar de Sa et al. 2006; Badreddine et al. 2007; Soyarslan et al. 2008; Boudifa et al. 2009; Saanouni and Lestriez 2009; Badreddine et al. 2010; Issa et al. 2011, 2012; Sornin and Saanouni 2011; Labergère et al. 2011). The equivalence between the CDM and the Gurson-type damage coupling has been investigated in Chaboche et al. (2006) where the potentialities of the CDM approach compared to the Gurson-type approach have been discussed mainly concerning the damage-induced anisotropy and its effect on the other fields (strong coupling). This kind of fully coupled approach accounts for the direct interactions (or strong coupling) between the inelastic flow, including different kinds of hardening, and the ductile damage initiation and growth. This full coupling allows the "natural" description of the strain localization modes inside the deformed part on the basis of the effect of the ductile damage evolution in the other mechanical fields under concern. Hence, it provides a simple and helpful way to predict where and when the inelastic flow localizes due to the earliest stage of ductile damage initiation without reference to any initial imperfection. The main advantages of this fully coupled approach are (Saanouni 2012a, b):

- It can be used, without any limitation, with advanced constitutive equations accounting for initial and induced anisotropies described by various quadratic or non-quadratic yield functions and plastic potentials. Many physical phenomena related to large inelastic strain coupled with ductile isotropic or anisotropic damage can be taken into account.
- The effect of the loading path shape in the stress or strain space (non-proportionality) and in time (cyclic loading) is considered including the reversibility of the load with or without compressive phase. This allows to accounting for:
  - The Bauschinger effect (kinematic hardening).
  - The closure of microcracks and/or microvoids (unilateral effect) under the compressive phase of the loading path and its effect on the recovery of some physical properties as the micro-defects close.
- Due to the localization modes giving rise to highly varying thickness, this approach will be used within full 3D or specific thick shell formulations in order to avoid the weakness related mainly to the plane stress assumption.

This kind of approach leads to thermodynamically consistent constitutive equations (for elasticity, (visco)plasticity, mixed hardening, damage, friction, thermal exchange, environment effect, etc.) with material parameters having a clear intrinsic character. In fact, since each main phenomenon is represented by a couple of state variables, which in turn are governed by appropriate ordinary differential equations, the identification procedure is decomposed in different steps, and the material parameters are determined using an inverse numerical approach for each phenomenon, while the other parameters are maintained fixed (Saanouni 2012a, b).

This chapter is dedicated to the presentation of an advanced fully adaptive numerical methodology for virtual sheet and/or bulk metal forming simulation. Section "Thermodynamically Consistent Modeling of the Behavior and Damage of Metallic Materials" gives the detailed formulation of thermodynamically consistent fully coupled multiphysical constitutive equations formulated in the framework of the generalized micromorphic continua and accounting for the main material nonlinearities as the isotropic and kinematic hardening, the thermal exchanges, and the nonlocal ductile damage under large inelastic strains. After the short summary of the main physical phenomena inherent to large inelastic strains, some modeling schemes usually used in continuum mechanics are briefly discussed. The mathematical representation of the ductile damage and its representation by scalar or tensorial variables is then presented. The kinematics of the homogeneous finite transformation together with the main conservation laws leading to the derivation of appropriate balance equations is presented. The formulation of the constitutive equations describing the thermomechanical behavior of the material is described in detail based on the appropriate choices of the state and dissipation potentials. Finally, the modeling of the constitutive equations for the contact interfaces is also presented.

The section "Numerical Aspects" is dedicated to the related numerical aspects. Firstly, the strong and weak forms for the multifunctional initial and boundary value problem (IBVP) are posed. Then the time and space discretization of the

IBVP in the framework of the finite difference method and finite element method respectively is discussed with the definition of some typical solid (3D) finite elements. The well-known static implicit (SI) and dynamic explicit (DE) global resolution schemes are summarized, followed by the detailed presentation of the local integration scheme of the fully coupled constitutive equations. The numerical treatment of the contact between deformable solids is reviewed in the framework of master/slave surfaces method taking into account of the friction constitutive equations. Finally, a brief description of the fully adaptive numerical methodology for virtual metal forming is presented.

Finally, section "Some Typical Examples of Virtual Metal Forming Processes" shows some application of the proposed virtual metal forming methodology to various sheet and bulk metal forming processes with damage occurrence. A short discussion of the identification procedure in order to determine the best values of the material parameters entering the fully coupled constitutive equations is first given. Then, some applications to various metal forming processes are presented and their results discussed.

Throughout this chapter, the following notations are used: $x, \vec{x}, \underline{x}, \underline{\vec{x}}, \underline{\underline{x}}$ indicate the zero (scalar), first (vector), second, third, and fourth rank tensors, respectively. The usual tensorial product is indicated by the symbol $\otimes$, while the contracted (or inner) product is indicated by $\cdot$, :, $\therefore$, and :: for the simple, double, triple, and quadruple contractions, respectively. In indicial form with respect to any orthonormal Cartesian frame of basis $\vec{e}_j$ ($\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$), this gives: $\vec{x}.\ \underline{y} = x_i y_{ij} = a_j \vec{e}_j = \vec{a}$, $\underline{x} : \underline{y} = x_{ijkl}\ y_{kl} = a_{ij} = \underline{a}$, $\underline{\vec{x}} \therefore \underline{\underline{y}} = x_{ijk}\ y_{ijkl} = a_l \vec{e}_l = \vec{a}$, or $\underline{\underline{x}} :: \underline{\underline{y}} = x_{ijkl}\ y_{ijkl} = a$.

Thermodynamically Consistent Modeling of the Behavior and Damage of Metallic Materials

About the Main Physical Phenomena Inherent to Large Inelastic Strains and Ductile Damage in Metallic Materials

Metallic components formed by large inelastic deformation under room or more or less high temperatures involve time-independent (plastic) or time-dependent (viscoplastic) large inelastic strains and very small elastic strains. When the large inelastic strains with hardening localize inside typical narrow zones, the ductile damage can take place according to the well-known mechanisms of nucleation, growth, and coalescence of microvoids around specific inclusions, second-phase particles, and other precipitates initially embedded inside the metallic matrix.

Briefly speaking, when a polycrystalline specimen containing inclusions is loaded (e.g., in tension), the following three mechanisms are successively observed:

- Inelastic flow with strain hardening and microcracks nucleation: Creation, multiplication, interaction, and arrangement of dislocations at the origin of the

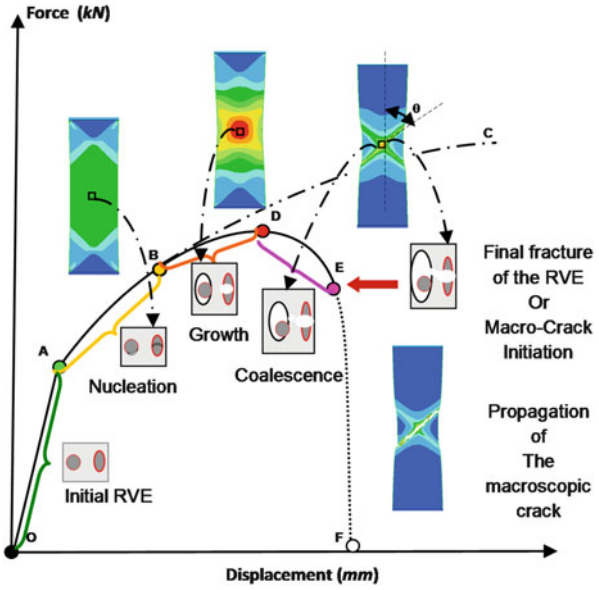


Fig. 1 Schematic representation of the ductile damage (nucleation, growth, and coalescence of microvoids) effects on the force–displacement curve under tensile tension until (Saanouni 2012a, b)

plastic strain by the well-known slip inside the slip planes of each crystal (or grain) favorably oriented with respect to the direction of the applied load. If the temperature is high enough, some thermally activated (diffusional) mechanisms can take place at the grain boundaries. Also for the large inelastic strain, grains can rotate (texture evolution) and new crystallographic slip systems become active. This leads to increasing the internal stresses giving rise to the strain hardening illustrated by the path ABC in Fig. 1. In fact during the path AB the inelastic strain with hardening takes place and the microcracks nucleation occurs without affecting the material behavior (neither elastic nor inelastic). Note that at this stage, the mechanical fields are homogeneous inside a large central part of the specimen.

- Microvoids growth and its effect on the inelastic flow with hardening: From point B (Fig. 1), the microvoids grow enough and their effect becomes sensitive introducing a non-negligible softening which decreases the hardening modulus to zero (point D). Note that at point D, the tangent modulus is zero and the maximum force is reached. During this stage, the elastic and inelastic behaviors of a typical RVE (Representative Volume Element defined as an aggregate of grains) are progressively influenced by the growing microcracks and microvoids inside the zone where the inelastic strain is maximum (path BD in Fig. 1). This zone is located at the central area of the specimen where a clear diffuse necking is observed.
- Micro-defects (microvoids and or microcracks) coalescence and fracture of the RVE: From point D, the force decreases quickly under the effect of the micro-defects coalescence inside the RVE located at the intersection of the two shear bands, where the inelastic dissipation is double, inducing a clear softening (path

DE, Fig. 1). In fact two shear bands form inside the central part of the specimen and intersect at the center of the specimen (localized necking). The RVE located at the intersection of the two shear bands contain the coalescing microvoids inside the typical RVE until its final fracture (path DE in Fig. 1). This defines the initiation of the first macroscopic crack inside the specimen.

- Macroscopic crack propagation (path EF, Fig. 1): Once the first RVE at the center of the specimen is fully fractured, other neighboring RVEs inside one shear band undergo the same scenario of the ductile fracture, creating the propagation of the macroscopic crack following one among the two shear bands previously formed. This phenomenon is very rapid (dynamic fracture) leading to the final fracture of the specimen as shown in Fig. 1 (path EF).

Clearly, these four phenomena, very sensitive to the temperature, are highly interrelated and interdependent. They are somewhat the consequences of each other and must be modeled as such without neglecting their various couplings and strong interactions. Consequently, any modeling work should account, as accurately as possible, for not only the kinetics of evolution of each phenomenon (elastic strain, thermal exchange, large inelastic strain, isotropic and anisotropic hardenings, damage, etc.) but also for their mutual "strong" coupling and interactions.

On Some Modeling Schemes

In solid and structural mechanics, the modeling job requires two main aspects: (i) the derivation of the main conservation laws (mass conservation, momentum conservation, energy conservation, positivity of the entropy production) and (ii) the derivation of the constitutive equations describing the evolution of the thermomechanical fields for each continuum under concern.

Generalized Conservation Laws

As depicted in the schematic representation of Fig. 2, the kinematics of the motion for any RVE in continuum mechanics is based in two assumptions as can be found in Truesdell and Noll (1965, 2004). The first one is the assumption of local action on which is based the classical (local) continua well known under the name of Cauchy continuum for which the knowledge of the first gradient of displacement (or the transformation gradient) is enough to determine the overall kinematics of the motion as well as all the variables needed to define its behavior. The second one is the assumption of the nonlocal action based on the integral form of the conservation laws as can be found in Eringen (1999, 2002). When transformed under local form of partial differential equations, these conservation laws give rise to a new term called the localization residuals together with appropriated jump conditions. This leads to a somewhat so complicated theory not very easy to use from both theoretical and numerical points of view.

However, the very helpful local action assumption can be used to define extended generalized continuum theories which may be grouped on two families:

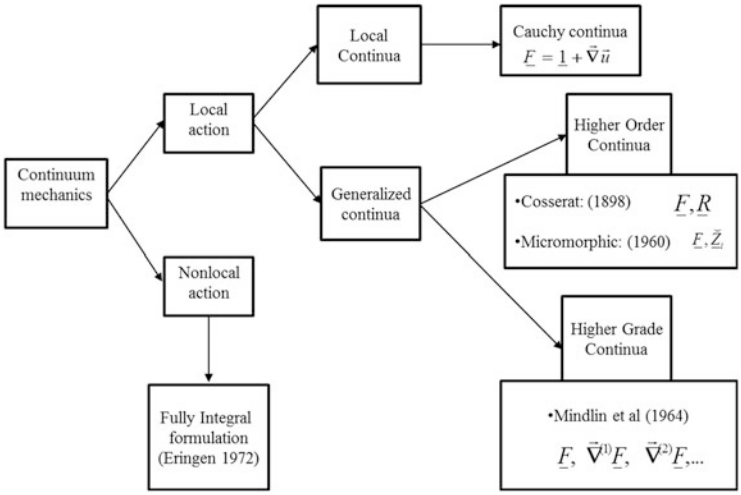


Fig. 2 Schematic representation of the different modeling schemes in continuum mechanics

(i) the higher-order continua (HOC) and (ii) the higher-grade continua (HGC) as can be found in Forest (2006, 2009), Forest and Sievert (2003), Saanouni (2012a, b), and Saanouni and Hamed (2013). Initially proposed by the Cosserat brothers (Cosserat and Cosserat 1909, 2009), the HOC theory consists in adding new degrees of freedom (dofs) in order to describe the kinematics of the transformation of the RVE. These additional dofs can be limited to the rigid body rotation tensor of the frame related to each RVE as proposed in the original Cosserat brothers' theory. They also can be a certain number ($i \in \{1, 2, 3 \ldots n\}$) of state variables here taken as scalar-valued variables formally noted $\left(\breve{z}_i, \breve{Z}_i\right)$ in which $\breve{z}_i$ are the strain-like variables and $\breve{Z}_i$ are their associated stress-like variables as proposed in the framework of the micromorphic theory (Eringen 1999, 2002; Forest and Aifantis 2010; Forest and Sievert 2003; Forest 2006, 2009; Saanouni 2012a, b; Saanouni and Hamed 2013). The HGC theory consists in introducing higher gradients of the displacement in addition to the transformation gradient $\underline{F}$ as the first gradient $\vec{\nabla}^{(1)}\underline{F}$, the second gradient $\vec{\nabla}^{(2)}\underline{F}$, the third $\vec{\nabla}^{(3)}\underline{F}$, and so on.

Applying the generalized principle of virtual power leads to the derivation of $(i + 1)$ balance equations: one classical balance of momentum (equilibrium equation) and i) additional micromorphic balance equations related to the i) introduced micromorphic dofs (see the next section).

Formulation of the Constitutive Equations

The standard thermodynamics of irreversible processes with state variables is still used in the same way as in classical local continuum mechanics (Germain 1973; Truesdell and Noll 1965, 2004; Besson et al. 2001; Lemaitre et al. 2009; Saanouni 2012a, b).

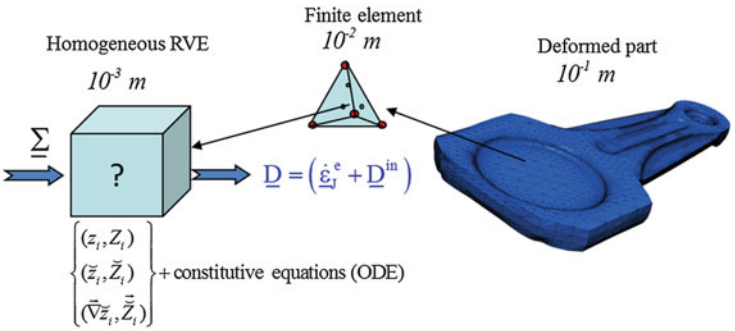


Fig. 3 Schematic representation of the macroscopic monoscale modeling approach (Saanouni 2012a, b)

However, the space of the state variables is enlarged by adding the new couple of micromorphic state variables $\left(\breve{z}_i, \breve{Z}_i\right)$ as well as their first gradients $\left(\vec{\nabla}\,\breve{z}_i, \vec{\breve{Z}}_i\right)$ if a first gradient theory is considered. A state potential is then constructed in the strain space (e.g., the Helmholtz free energy) as a closed and convex scalar-valued function of the strain-like variables from which all the classical and the micromorphic stress-like variables are deduced. In addition, appropriate yield functions and dissipation potentials have to be constructed in order to derive, using the generalized normality rule, the local and micromorphic flux variables defining the evolution of the strain-like variables associated to the local and micromorphic phenomena (Forest 2006; Saanouni 2012a, b; Saanouni and Hamed 2013) as can be developed in the next section.

These constitutive equations can be formulated using either the macroscopic monoscale approach or the micro–macro, multiscale approach. Schematized in Fig. 3, the macroscopic monoscale approach is phenomenological in nature and consists in attaching many couples of local state variables, namely, (z_i, Z_i), and the micromorphic state variables $\left(\breve{z}_i, \breve{Z}_i\right)$ and $\left(\vec{\nabla}\,\breve{z}_i, \vec{\breve{Z}}_i\right)$ to each RVE representing an integration or Gauss point (GP) of any finite element of the structure. Then knowing the macroscopic stress tensor (or total strain tensor) at each GP, compute its answer in terms of total strain tenor (or total stress tensor), through the numerical integration of the fully coupled ordinary differential equations which are nothing but the constitutive equations describing the evolution of the overall dissipative phenomena. In this approach each physical (local or micromorphic) phenomenon is represented by a couple of state variables (tensor of any rank) and each state variable is governed by an appropriate ordinary differential equation (ODE). Note that this kind of approach is of **deductive** type in such a manner that the material behavior is deduced from the behavior of the RVE defined for representative experimental conditions, and then **extrapolated** for more large conditions covering the future domain of use.

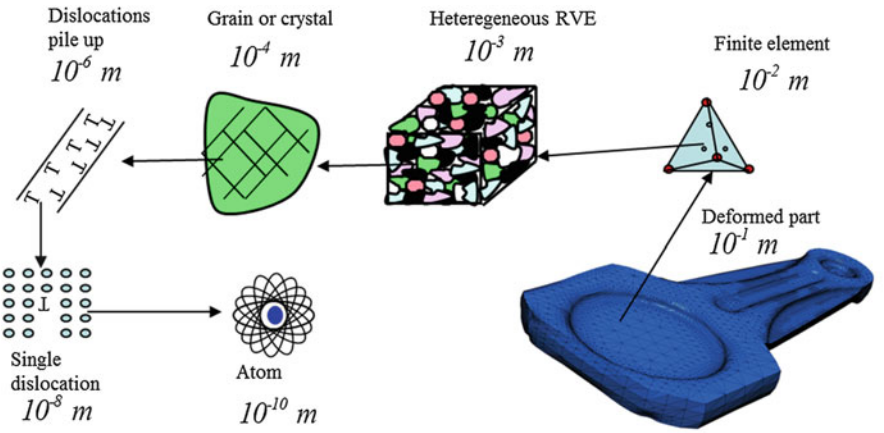


Fig. 4 Schematic representation of the micro–macro multiscale modeling approach (Saanouni 2012a, b)

The micro–macro multiscale approach is more physically based, looking for the fine description of each physical phenomenon (hardening, damage, etc.) at the adequate scale (Schmid and Boas 1968; Bunge 1982; Mura 1987; Havner 1992; Nemat-Nasser and Hori 1993; Yang and Lee 1993; Kocks et al. 1998; Raabe 1998; Bornert et al. 2001; Gambin 2001; Nemat-Nasser 2004; Asaro and Lubarda 2006). In this approach, each RVE is defined as a 3D aggregate of N_g grains, each one being a single crystal of a given phase defined by its phase nature, shape, size, and crystallographic orientations (Fig. 4). The idea is to use appropriate localization–homogenization methods in order to deduce the macroscopic behavior of the RVE from the knowledge of the local behaviors of each of the elementary constituents (grains, dislocations, atoms, etc.) at lower scales. This therefore requires recourse to appropriate localization and homogenization methods. Briefly speaking, two families of inductive methods are often used: (i) **full fields** methods, based on numerical treatment, by, for instance, the finite elements method, of all of the microstructure components (crystals), by assuming appropriate hypotheses of spatial and temporal periodicity, thanks to homogenization techniques, and (ii) **mean fields** methods, which rely on quasi-analytical methods to describe the various heterogeneities of matter by using, for example, self-consistent methods based on the basic solutions of the Eshelby inclusion problem. This approach is of inductive nature since it consists of describing accurately the physical mechanisms at the appropriate lower scales inducing naturally the macroscopic behavior thanks to the homogenization methods.

In this chapter dedicated to the damage prediction in metal forming using the macroscopic approach, only the monoscale modeling approach is used. The multiscale localization–homogenization-based approach will be used in ▶ Chap. 28 "Micromechanical Polycrystalline Damage-plasticity Modeling for Metal Forming Processes" of the present volume.

Balance Equations and Thermodynamics of Irreversible Processes for the Micromorphic Continua

Throughout this chapter, and unless direct specification, only isothermal processes are considered for the sake of shortness. The consideration of anisothermal processes, combining mechanical and thermal applied loading paths, can be found in Saanouni (2012a, b).

The Proposed Micromorphic Balance Equations

In this section, the conservation laws for the micromorphic continua will be derived. Consider an isothermal body with a volume $\Omega(\vec{x},t)$, surrounded by the boundary Γ and moving with the velocity field $\vec{u}(\vec{x},t)$ where $\vec{x}$ is the spatial position of any material point of the body at the current (deformed and damaged) configuration. To the classical dofs, represented by the displacement field $\vec{u}$, three new micromorphic dofs are added: the isotropic micromorphic damage $\breve{D}$ ($\breve{Y}$ its associated force), the micromorphic isotropic hardening $\breve{r}$ ($\breve{R}$ its associated force), and the macroscopic kinematic hardening $\breve{\underline{\alpha}}$ ($\breve{\underline{X}}$ its associated force).

The first conservation law concerns the conservation of mass. For the sake of simplicity, it is assumed that the micromorphic density is proportional to the classical material density $\rho(\vec{x},t)$. Accordingly, the conservation of mass keeps its classical form as in local continuum mechanics leading to the following integral and differential forms:

$$\frac{d}{dt}\int_{\Omega}\rho\left(\vec{x},t\right)d\Omega=0 \quad \text{or} \quad \frac{\partial\rho\left(\vec{x},t\right)}{\partial t}+div\left(\rho\left(\vec{x},t\right)\vec{u}\left(\vec{x},t\right)\right)=0 \tag{1}$$

where the notation $d(-)/dt$ stands for the material time derivatives of the scalar-valued function $(-)$.

The virtual power of internal forces δP_{int} of the classical local continuum is extended to the micromorphic continuum using the virtual dofs introduced above to have:

$$\begin{aligned}\delta P_{\text{int}}=&-\int_{\Omega}\left(\underline{\sigma}:\vec{\nabla}\,\dot{\delta\vec{u}}\right)d\Omega+\int_{\Omega}\left(\breve{R}\dot{\delta\breve{r}}+\vec{\breve{R}}\cdot\vec{\nabla}\,\dot{\delta\breve{r}}\right)d\Omega+\\&\int_{\Omega}\left(\breve{Y}\dot{\delta\breve{D}}+\vec{\breve{Y}}.\vec{\nabla}\,\dot{\delta\breve{D}}\right)d\Omega+\int_{\Omega}\left(\breve{\underline{X}}:\dot{\delta\breve{\underline{\alpha}}}+\vec{\breve{\underline{X}}}\therefore\vec{\nabla}\,\dot{\delta\breve{\underline{\alpha}}}\right)d\Omega\end{aligned} \tag{2}$$

where $\dot{\delta\vec{u}}$, $\dot{\delta\breve{r}}$, $\dot{\delta\breve{D}}$ and $\dot{\delta\breve{\underline{\alpha}}}$ are the kinematically admissible virtual velocity fields and $\vec{\nabla}$ stands for the spatial gradient operator. Note that the generalized forces $\underline{\sigma}$ (the Cauchy stress tensor), $\vec{\breve{R}}$, $\vec{\breve{Y}}$ (vectors), and $\vec{\breve{\underline{X}}}$ (third-rank tensor) are the thermodynamic forces associated with the total strain (or the symmetric part

of $\vec{\nabla}\vec{u}$), the first gradient of the micromorphic isotropic hardening $\vec{\nabla}\breve{r}$, the first gradient of the micromorphic damage $\vec{\nabla}\breve{D}$, and the first gradient of the micromorphic kinematic hardening $\vec{\nabla}\breve{\underline{\alpha}}$.

On the other hand, the virtual power of external forces δP_{ext} can be written under the additive form of the classical local term with three new terms related to the three micromorphic phenomena under concern:

$$\begin{aligned}\delta P_{\text{ext}} = &\int_{\Omega}\left(\vec{f}^{u}.\,\delta\dot{\vec{u}}\right)d\Omega + \int_{\Gamma}\left(\vec{F}^{u}.\,\delta\dot{\vec{u}}\right)ds + \int_{\Omega}\left(f^{\breve{r}}\delta\dot{\breve{r}} + \vec{f}^{g\breve{r}}.\,\vec{\nabla}\delta\dot{\breve{r}}\right)d\Omega \\ &+ \int_{\Gamma}\left(F^{\breve{r}}\delta\dot{\breve{r}}\right)ds + \int_{\Omega}\left(f^{\breve{D}}\delta\dot{\breve{D}} + \vec{f}^{g\breve{D}}.\,\vec{\nabla}\,\delta\dot{\breve{D}}\right)d\Omega + \int_{\Gamma}\left(F^{\breve{D}}\delta\dot{\breve{D}}\right)ds \\ &+ \int_{\Omega}\left(\underline{f}^{\breve{\alpha}} : \delta\dot{\underline{\alpha}} + \vec{\underline{f}}^{g\breve{\alpha}} \therefore \vec{\nabla}\delta\dot{\underline{\alpha}}\right)d\Omega + \int_{\Gamma}\left(\underline{F}^{\breve{\alpha}} : \delta\dot{\underline{\alpha}}\right)ds\end{aligned} \tag{3}$$

where $\vec{f}^{u}, f^{\breve{r}}, \vec{f}^{g\breve{r}}, f^{\breve{D}}, \vec{f}^{g\breve{D}}, \underline{f}^{\breve{\alpha}}$, and $\vec{\underline{f}}^{g\breve{\alpha}}$ are the simple and generalized body forces associated with the displacement, the micromorphic dofs, and their respective first gradients, while $\vec{F}^{u}$, $\underline{F}^{\breve{\alpha}}$, $F^{\breve{r}}$, and $F^{\breve{D}}$ are the simple and generalized micromorphic forces acting on the appropriate parts of the boundary Γ.

Similarly, the virtual power of the acceleration (or inertia) forces δP_a can be also written under the additive form of the classical local inertia force enhanced by the contribution of the three micromorphic phenomena:

$$\delta P_a = \int_{\Omega}\left(\rho\ddot{\vec{u}}.\,\delta\dot{\vec{u}}\right)d\Omega + \int_{\Omega}\rho\left(\zeta_{\breve{r}}\ddot{\breve{r}}\delta\dot{\breve{r}} + \zeta_{\breve{D}}\ddot{\breve{D}}\delta\dot{\breve{D}} + \zeta_{\breve{\alpha}}\ddot{\underline{\alpha}} : \delta\dot{\underline{\alpha}}\right)d\Omega \tag{4}$$

in which $\ddot{\vec{u}}$ is the classical local acceleration vector and the scalars $\ddot{\breve{r}}, \ddot{\breve{D}}$, and the second-rank tensor $\ddot{\underline{\alpha}}$ are the generalized micromorphic "accelerations." The micromorphic parameters $\zeta_{\breve{r}}$, $\zeta_{\breve{D}}$, and $\zeta_{\breve{\alpha}}$, homogenous to square lengths, are scale factors which map the local density $\rho\left(\vec{x},t\right)$ governed by Eq. 1 to the micromorphic one related to each micromorphic phenomenon.

The generalized principle of virtual power, with respect to the virtual velocity and micromorphic fields, takes the following form:

$$\delta P_{\text{int}} + \delta P_{\text{ext}} = \delta P_a \quad \forall\ \delta\dot{\vec{u}},\ \delta\dot{\breve{r}}\ ,\ \delta\dot{\breve{D}}\ ,\ \delta\dot{\underline{\alpha}}\ \ K.A. \tag{5}$$

where *K.A.* stands for kinematically admissible fields in the usual sense, i.e., fulfilling the Neumann-type boundary conditions (BC). By using Eqs. 2, 3, and 4, and with the help of the divergence theorem to transform the volume integrals,

Eq. 5 leads to four balance equations. The first one is nothing but the classical local balance of momentum expressed in terms of the Cauchy stress tensor:

$$\begin{cases} div(\underline{\sigma}) + \vec{f}^{u} = \rho \ddot{\vec{u}} & \text{in } \Omega \\ \underline{\sigma} \,.\, \vec{n} = \vec{F}^{u} & \text{on } \Gamma \end{cases} \tag{6}$$

The three additional generalized balance of micromorphic momentum equations are given, in terms of the micromorphic forces, by:

$$\begin{cases} div\left(\breve{\vec{\underline{X}}} - \vec{\underline{f}}^{g\breve{\alpha}}\right) - \left(\breve{\underline{X}} - \underline{f}^{\breve{\alpha}}\right) = \rho \zeta_{\breve{\alpha}} \ddot{\breve{\underline{\alpha}}} & \text{in } \Omega \\ \left(\breve{\vec{\underline{X}}} - \vec{\underline{f}}^{g\breve{\alpha}}\right).\vec{n} = \underline{F}^{\breve{\alpha}} \text{ on } \Gamma \end{cases} \tag{7}$$

$$\begin{cases} div\left(\breve{\vec{R}} - \vec{f}^{g\breve{r}}\right) - \left(\breve{R} - f^{\breve{r}}\right) = \rho \zeta_{\breve{r}} \ddot{\breve{r}} & \text{in } \Omega \\ \left(\breve{\vec{R}} - \vec{f}^{g\breve{r}}\right).\vec{n} = F^{\breve{r}} & \text{on } \Gamma \end{cases} \tag{8}$$

$$\begin{cases} div\left(\breve{\vec{Y}} - \vec{f}^{g\breve{D}}\right) - \left(\breve{Y} - f^{\breve{D}}\right) = \rho \zeta_{\breve{d}} \ddot{\breve{d}} & \text{in } \Omega \\ \left(\breve{\vec{Y}} - \vec{f}^{g\breve{D}}\right).\vec{n} = F^{\breve{D}} & \text{on } \Gamma \end{cases} \tag{9}$$

The four partial differential equations (6), (7), (8), and (9) are nothing but the strong forms associated with the micromorphic initial and boundary value problem (IBVP) in which the body forces $\vec{f}^{u}, f^{\breve{r}}, \vec{f}^{g\breve{r}}, f^{\breve{D}}, \vec{f}^{g\breve{D}}, \underline{f}^{\breve{\alpha}}$, and $\vec{\underline{f}}^{g\breve{\alpha}}$ as well as the "contact" or surface forces $\vec{F}^{u}, F^{\breve{r}}, F^{\breve{D}}$, and $\underline{F}^{\breve{\alpha}}$ are known (given) quantities, while the stress-like variables $\breve{\underline{\sigma}}, \breve{\vec{R}}, \vec{R}, \breve{\underline{X}}, \breve{\vec{\underline{X}}}, \breve{Y}$, and $\breve{\vec{Y}}$ have to be defined by the appropriated constitutive equations developed in the framework of the generalized thermodynamics of irreversible processes with state variables, as can described here after.

Let's examine now the principle of energy conservation known as the first principle of thermodynamics. By assuming that the micromorphic thermal phenomena are neglected, it is easy to obtain, using the definition of the internal power similar to Eq. 2, the generalized thermal equation governing the heat transfer under the following form:

$$\underline{\sigma} : \underline{D} + \breve{\underline{X}} : \dot{\breve{\underline{\alpha}}} + \breve{\vec{\underline{X}}} \therefore \vec{\nabla} \dot{\breve{\underline{\alpha}}} + \breve{R} \dot{\breve{r}} + \breve{\vec{R}} . \vec{\nabla} \dot{\breve{r}} + \breve{Y} \dot{D} + \breve{\vec{Y}} . \vec{\nabla}$$

$$\dot{\breve{D}} + \xi - \rho\left(\dot{\psi} + T\dot{s} + s\dot{T}\right) - div\left(\vec{q}\right) = 0 \tag{10}$$

in which $\underline{D} = \underline{\dot{\varepsilon}}^{eJ} + \underline{D}^{in}$ is the spatial total strain rate which decomposes additionally to a sum of the Jaumann small elastic strain rate and the inelastic (plastic or viscoplastic) strain rate; $\psi\left(\underline{\varepsilon}^e, \underline{\alpha}, r, D, T, \breve{\underline{\alpha}}, \breve{r}, \breve{D}, \vec{\nabla}\breve{\underline{\alpha}}, \vec{\nabla}\breve{r}, \vec{\nabla}\breve{D}\right)$ is the specific Helmholtz free energy, a convex and closed function of the strain-like state variables, taken as a state potential; $\underline{\alpha}$ is the kinematic hardening strain tensor conjugate to the stress tensor $\underline{X}$, and r is the isotropic hardening strain conjugate to the Orowan stress R, and finally D is the isotropic damage variable conjugate to the force Y. The variable s is the specific entropy of the material which is the force-like variable associated with the absolute temperature T. and $\vec{q}$ is the heat flux vector directly related to the first gradient of the temperature using the well-known Fourier linear heat transfer model. Note that if the derivatives of ψ and s with respect to time are calculated and substituted into Eq. 10, then the final heat equation for the micromorphic continuum is easily obtained (Saanouni 2012a, b). This heat equation, together with adequate thermal boundary conditions, may be added to Eqs. 6, 7, 8, and 9 in order to define completely the strong form of the IBVP (see later).

Before closing this section, let's introduce the second principle of thermodynamics, which expresses the positivity of the entropy production rate and the Clausius–Duhem inequality that can be derived from the combination of the second and first principles of the thermodynamics. If the micromorphic thermal phenomena are neglected, the entropy production rate can be expressed under the following classical form:

$$\rho\dot{s} + div\left(\frac{\vec{q}}{T}\right) - \frac{\xi}{T} \geq 0 \tag{11}$$

The combination of Eqs. 10 and 11 leads to the so-called Clausius–Duhem (CD) inequality for the micromorphic continua:

$$\begin{aligned} \underline{\sigma} : \underline{D} + \breve{\underline{X}} : \dot{\breve{\underline{\alpha}}} + \vec{\breve{\underline{X}}} \therefore \vec{\nabla}\dot{\breve{\underline{\alpha}}} + \breve{R}\dot{\breve{r}} + \vec{\breve{R}} \cdot \vec{\nabla}\dot{\breve{r}} + \breve{Y}\dot{\breve{D}} + \vec{\breve{Y}} \cdot \vec{\nabla}\dot{\breve{D}} \\ + \xi - \rho\left(\dot{\psi} + s\dot{T}\right) - \frac{\vec{q}}{T} \cdot \overrightarrow{grad}(T) \geq 0 \end{aligned} \tag{12}$$

This basic inequality will be used to derive both the state relations and the residual dissipation inequality as discussed hereafter.

Thermodynamics of Irreversible Processes for the Micromorphic Continua

The total time derivative of the state potential is given by

$$\begin{aligned} \dot{\psi} = \frac{\partial\psi}{\partial\underline{\varepsilon}^e} : \dot{\underline{\varepsilon}}^e + \frac{\partial\psi}{\partial\underline{\alpha}} : \dot{\underline{\alpha}} + \frac{\partial\psi}{\partial r}\dot{r} + \frac{\partial\psi}{\partial D}\dot{D} + \frac{\partial\psi}{\partial T}\dot{T} + \frac{\partial\psi}{\partial\breve{\underline{\alpha}}} : \dot{\breve{\underline{\alpha}}} + \frac{\partial\psi}{\partial\breve{r}}\dot{\breve{r}} + \frac{\partial\psi}{\partial\breve{D}}\dot{\breve{D}} \\ + \frac{\partial\psi}{\partial\vec{\nabla}\breve{\underline{\alpha}}} \therefore \vec{\nabla}\dot{\breve{\underline{\alpha}}} + \frac{\partial\psi}{\partial\vec{\nabla}\breve{r}} \cdot \vec{\nabla}\dot{\breve{r}} + \frac{\partial\psi}{\partial\vec{\nabla}\breve{D}} \cdot \vec{\nabla}\dot{\breve{D}} \end{aligned} \tag{13}$$

Using Eq. 13 in Eq. 12 together with $\underline{D} = \underline{\dot{\varepsilon}}_f^{re} + \underline{D}^{ir}$ (see later), the CD inequality in Eq. 12 transforms to

$$\left(\underline{\sigma} - \rho\frac{\partial\psi}{\partial\underline{\varepsilon}^e}\right) : \underline{\dot{\varepsilon}}_f^{re} + \rho\left(s + \frac{\partial\psi}{\partial T}\right)\dot{T} + \left(\breve{\underline{X}} - \rho\frac{\partial\psi}{\partial\breve{\underline{\alpha}}}\right) : \dot{\breve{\underline{\alpha}}} + \left(\breve{R} - \rho\frac{\partial\psi}{\partial\breve{r}}\right)\dot{\breve{r}} + \left(\breve{Y} - \rho\frac{\partial\psi}{\partial\breve{D}}\right)\dot{\breve{D}}$$
$$+ \left(\vec{\breve{\underline{X}}} - \rho\frac{\partial\psi}{\partial\vec{\nabla}\breve{\underline{\alpha}}}\right) .\vec{\nabla}\dot{\breve{\underline{\alpha}}} + \left(\vec{\breve{R}} - \rho\frac{\partial\psi}{\partial\vec{\nabla}\breve{r}}\right) .\vec{\nabla}\dot{\breve{r}} + \left(\vec{\breve{Y}} - \rho\frac{\partial\psi}{\partial\vec{\nabla}\breve{D}}\right) .\vec{\nabla}\dot{\breve{D}}$$
$$+ \underline{\sigma} : \underline{D}^{ir} - \rho\frac{\partial\psi}{\partial\underline{\alpha}} : \dot{\underline{\alpha}} - \rho\frac{\partial\psi}{\partial r}\dot{r} - \rho\frac{\partial\psi}{\partial D}\dot{D} - \frac{\vec{q}}{T} . \overrightarrow{grad}(T) \geq 0 \tag{14}$$

Following the standard arguments of the rational thermodynamics (Truesdell and Noll 1965, 2004), Eq. 14 leads to the definition of the:

- State relations: which defines the stress-like state variables associated with the classical strain-like variables and with the micromorphic ones classified as follows:
  - State relations associated with the five local strain-like variables:

$$\underline{\sigma} = \rho\frac{\partial\psi}{\partial\underline{\varepsilon}^e}, s = -\frac{\partial\psi}{\partial T}, \underline{X} = \rho\frac{\partial\psi}{\partial\underline{\alpha}}, R = \rho\frac{\partial\psi}{\partial r}, Y = -\rho\frac{\partial\psi}{\partial D} \tag{15}$$

  - State relations associated with the three micromorphic strain-like variables:

$$\breve{\underline{X}} = \rho\frac{\partial\psi}{\partial\breve{\underline{\alpha}}}, \breve{R} = \rho\frac{\partial\psi}{\partial\breve{r}}, \breve{Y} = \rho\frac{\partial\psi}{\partial\breve{D}} \tag{16}$$

  - State relations associated with the first gradient of each of the three micromorphic strain-like variables:

$$\vec{\breve{\underline{X}}} = \rho\frac{\partial\psi}{\partial\vec{\nabla}\breve{\underline{\alpha}}}, \vec{\breve{R}} = \rho\frac{\partial\psi}{\partial\vec{\nabla}\breve{r}}, \vec{\breve{Y}} = \rho\frac{\partial\psi}{\partial\vec{\nabla}\breve{D}} \tag{17}$$

- The volumetric dissipation inequality: which is nothing but the residual terms of the CD inequality in Eq. 14 rewritten here under the equivalent form:

$$\wp^v = \underline{\sigma} : \underline{D}^{ir} - \underline{X} : \dot{\alpha} \quad \underline{\alpha} - R : \dot{r} + Y\dot{D} - \frac{\vec{q}}{T} . \overrightarrow{grad}(T) \geq 0 \tag{18}$$

Clearly, Eq. 18 indicates that there is no dissipation from the micromorphic phenomena leading to a purely local volumetric dissipation. This assumed fact can be easily avoided by decomposing all the micromorphic state variables to reversible and irreversible parts similar to the total strain rate tensor as discussed in Saanouni (2012a, b). However, doing this requires the complete experimental knowledge of the micromorphic flow in order to choose the corresponding yield functions, which is clearly out of our purpose at this time.

Proposed Fully Coupled Micromorphic Constitutive Equations

Focusing our attention on the purely isothermal processes for the sake of the simplicity and the shortness of the paper, all the thermal aspects are neglected (see Saanouni (2012a, b), for more complete modeling including the strong thermomechanical coupling). Consequently, the time-independent finite plasticity flow is taken as the irreversible large strains ($\underline{\varepsilon}^{ir} = \underline{\varepsilon}^{p}$), while the small elastic strains are considered as the reversible strains ($\underline{\varepsilon}^{re} = \underline{\varepsilon}^{e}$). Accordingly, the couples of state variables associated with the phenomena taken into account in this formulation are $(\underline{\varepsilon}^{e}, \underline{\sigma})$, for the plastic flow, (r, R) for the isotropic hardening, $(\underline{\alpha}, \underline{X})$ for the kinematic hardening, and (D, Y) for the isotropic ductile damage in addition to the following pairs of micromorphic state variables: $\left(\breve{r}, \breve{R}\right)$, $\left(\vec{\nabla}\breve{r}, \vec{\breve{R}}\right)$, $\left(\breve{\underline{\alpha}}, \breve{\underline{X}}\right)$, $\left(\vec{\nabla}\breve{\underline{\alpha}}, \vec{\breve{\underline{X}}}\right)$, $\left(\breve{D}, \breve{Y}\right)$, and $\left(\vec{\nabla}\breve{D}, \vec{\breve{Y}}\right)$.

Micromorphic

State potential and state relationships: The Helmholtz free energy is taken as a state potential written, in the effective (damage free) actual locally rotated configuration, as a convex closed function of the strain-like variables under the following form (Hamed 2012; Saanouni 2012a, b; Saanouni and Hamed 2013):

$$\begin{aligned}\rho\psi = &\frac{1}{2}(1-D)\left[\underline{\varepsilon}^{e} : \underline{\underline{\Lambda}} : \underline{\varepsilon}^{e} + \frac{2}{3}C\underline{\alpha} : \underline{\alpha} + Qr^{2}\right] \\ &+ \frac{1}{2}\left[\frac{2}{3}\breve{C}\left(\sqrt{(1-D)}\underline{\alpha} - \breve{\underline{\alpha}}\right) : \left(\sqrt{(1-D)}\underline{\alpha} - \breve{\underline{\alpha}}\right) + \breve{Q}\left(\sqrt{(1-D)}r - \breve{r}\right)^{2} + \breve{H}\left(D - \breve{D}\right)^{2}\right] \\ &+ \frac{1}{2}\left[\frac{3}{2}\breve{C}^{g}\left(\vec{\nabla}\breve{\underline{\alpha}}\right) \therefore \left(\vec{\nabla}\breve{\underline{\alpha}}\right) + \breve{Q}^{g}\left(\vec{\nabla}\breve{r}\right).\left(\vec{\nabla}\breve{r}\right) + \breve{H}^{g}\left(\vec{\nabla}\breve{D}\right).\left(\vec{\nabla}\breve{D}\right)\right]\end{aligned} \tag{19}$$

where $\underline{\underline{\Lambda}} = \lambda_e \underline{1} \otimes \underline{1} + 2\mu_e \underline{\underline{1}}$ is the positive definite symmetric fourth-rank tensor of elastic moduli for the isotropic solid defined by the well-known Lame's parameters λ_e and μ_e. C and Q are the kinematic and isotropic hardening macroscopic moduli; $\breve{C}$, $\breve{Q}$, and $\breve{H}$ are the coupling moduli with respect to the kinematic hardening, isotropic hardening, and damage, respectively; and $\breve{C}^{g}, \breve{Q}^{g}, \breve{H}^{g}$ are the micromorphic moduli relative to the micromorphic kinematic hardening, isotropic hardening, and damage, respectively.

Using the state relations (15), (16), and (17), the local (macro) and nonlocal (micromorphic) thermodynamic forces associated with the macro- and micromorphic state variables are deduced from the state potential (Hamed 2012; Saanouni 2012a, b; Saanouni and Hamed 2013). After some, rather fastidious than difficult, algebraic calculations, these can be put under the following forms:

$$\underline{\sigma} = \rho\frac{\partial\psi}{\partial\underline{\varepsilon}^e} = (1-D)\underline{\underline{\Lambda}} : \underline{\varepsilon}^e = (1-D)[\lambda_e tr(\underline{\varepsilon}^e)\underline{1} + 2\mu_e\underline{\varepsilon}^e] = \underline{\sigma}_{loc} \tag{20}$$

$$\underline{X} = \rho\frac{\partial\psi}{\partial\underline{\alpha}} = \underline{X}_{loc} + \underline{X}_{nloc}, \quad \underline{X}_{loc} = \frac{2}{3}(1-D)C\underline{\alpha}, \quad \underline{X}_{nloc} = -\sqrt{(1-D)}\breve{\underline{X}} \tag{21}$$

$$R = \rho\frac{\partial\psi}{\partial r} = R_{loc} + R_{nloc}, \quad R_{loc} = (1-D)Qr, \quad R_{nloc} = -\sqrt{(1-D)}\breve{R} \tag{22}$$

$$\begin{aligned} &Y = -\rho\frac{\partial\psi}{\partial D} = Y_{loc} + Y_{nloc}, \quad Y_{loc} = \frac{1}{2}\underline{\varepsilon}^e : \underline{\underline{\Lambda}} : \underline{\varepsilon}^e + \frac{1}{2}Qr^2 + \frac{1}{3}C\underline{\alpha} : \underline{\alpha} \\ &Y_{nloc} = \breve{Y} - \frac{1}{2\sqrt{1-D}}\left[\breve{R}r + \breve{\underline{X}} : \underline{\alpha}\right] \end{aligned} \tag{23}$$

$$\breve{\underline{X}} = \rho\frac{\partial\psi}{\partial\breve{\underline{\alpha}}} = -\frac{2}{3}\breve{C}\left(\sqrt{(1-D)}\underline{\alpha} - \breve{\underline{\alpha}}\right), \quad \vec{\breve{\underline{X}}} = \rho\frac{\partial\psi}{\partial\vec{\nabla}\,\breve{\underline{\alpha}}} = \breve{C}^g\left(\vec{\nabla}\,\breve{\underline{\alpha}}\right) \tag{24}$$

$$\breve{R} = \rho\frac{\partial\psi}{\partial\breve{r}} = -\breve{Q}\left(\sqrt{(1-D)}r - \breve{r}\right), \quad \vec{\breve{R}} = \rho\frac{\partial\psi}{\partial\vec{\nabla}\,\breve{r}} = \breve{Q}^g\left(\vec{\nabla}\,\breve{r}\right) \tag{25}$$

$$\breve{Y} = \rho\frac{\partial\psi}{\partial\breve{D}} = -\breve{H}\left(D - \breve{D}\right), \qquad \vec{\breve{Y}} = \rho\frac{\partial\psi}{\partial\vec{\nabla}\,\breve{D}} = \breve{H}^g\left(\vec{\nabla}\,\breve{D}\right) \tag{26}$$

Excepting the Cauchy stress tensor which is still fully local quantity as indicated by Eq. 20, the three other stress-like variables associated respectively with the kinematic hardening, isotropic hardening, and damage have additive contributions of the classical local variables noted with the subscript (*loc*) and nonlocal contributions due to the micromorphic variables noted with the subscript (*nloc*) as indicated by Eqs. 21, 22, and 23. It is worth noting that if the micromorphic effects are neglected, i.e., if $\breve{C} = \breve{Q} = \breve{H} = 0$ and $\breve{C}^g = \breve{Q}^g = \breve{H}^g = 0$, then all the micromorphic contribution vanish and the classical fully coupled local state relations originally given in Saanouni et al. (1994) and Saanouni (2012a, b) are integrally recovered.

Dissipations Analysis and Evolution Equations

According to the assumption taken above neglecting the micromorphic sources of the dissipation, the classical nonassociative plasticity theory is assumed. To account for the isotropic damage effects, and limiting ourselves to the von Mises plastic

flow for the sake of simplicity, the yield function and dissipation potentials taken from (Saanouni 2012a, b) are considered:

$$f(\underline{\sigma},\underline{X},R,D)=\frac{\|\underline{\sigma}-\underline{X}\|-R}{\sqrt{(1-D)}}-\sigma_y\leq 0 \tag{27}$$

$$\begin{aligned}F(\underline{\sigma},\underline{X},R,D)=&f+\frac{1}{2}\frac{a\|\underline{X}\|}{C(1-D)}+\frac{1}{2}\frac{bR^2}{Q(1-D)}\\&+\frac{S}{s+1}\left\langle\frac{Y-Y_0}{S}\right\rangle^{s+1}\frac{1}{(1-D)^{\beta}}\end{aligned} \tag{28}$$

For any second-rank stress tensor $\underline{T}$ having $\underline{T}^{dev}$ as a deviator, the norm $\|\underline{T}\|$ is defined in the stress space by $\|\underline{T}\|=\|\underline{T}\|_M=\sqrt{(3/2)\underline{T}^{dev}:\underline{T}^{dev}}$ if the von Mises isotropic plastic flow is considered or by $\|\underline{T}\|=\|\underline{T}\|_H=\sqrt{\underline{T}:\underline{\underline{H}}:\underline{T}}$ if the Hill-type anisotropic plastic flow is taken with $\underline{\underline{H}}$ being the fourth-rank orthotropic operator. Note that the kinematic stress tensor $\underline{X}$ is a purely deviatoric stress tensor. The parameters a and b characterize the nonlinearity of the kinematic and isotropic hardening, respectively, while S, s, β and Y_0 characterize the nonlinear evolution of the ductile damage. Applying the generalized normality rule in the framework of the nonassociative plasticity theory yields to the following evolution equations for all the dissipative phenomena under concern:

$$\underline{D}^p=\dot{\lambda}^p\frac{\partial F}{\partial\underline{\sigma}}=\dot{\lambda}^p\frac{\partial f}{\partial\underline{\sigma}}=\dot{\lambda}^p\underline{n}=\dot{\lambda}^p\frac{\widetilde{\underline{n}}}{\sqrt{(1-D)}} \tag{29}$$

$$\begin{aligned}\dot{\alpha}\quad\underline{\alpha}&=-\dot{\lambda}^p\frac{\partial F}{\partial\underline{X}}=\left(\underline{D}^p-a\dot{\lambda}^p\left(\frac{\left(C+\breve{C}\right)}{C}\underline{\alpha}-\frac{\breve{C}}{C}\frac{\breve{\underline{\alpha}}}{\sqrt{(1-D)}}\right)\right)\\&=\frac{\dot{\lambda}^p}{\sqrt{(1-D)}}\left(\widetilde{\underline{n}}-a\left(\sqrt{(1-D)}\frac{\left(C+\breve{C}\right)}{C}\underline{\alpha}-\frac{\breve{C}}{C}\breve{\underline{\alpha}}\right)\right)\end{aligned} \tag{30}$$

$$\dot{r}=-\dot{\lambda}^p\frac{\partial F}{\partial R}=\frac{\dot{\lambda}^p}{\sqrt{(1-D)}}\left(1-b\left(\sqrt{(1-D)}\frac{\left(Q+\breve{Q}\right)}{Q}r-\frac{\breve{Q}}{Q}\breve{r}\right)\right) \tag{31}$$

$$\dot{D} = \dot{\lambda}^p \frac{\partial F}{\partial Y} = \frac{\dot{\lambda}^p}{(1-D)^\beta} \left(\frac{\langle (Y_{loc} + Y_{nloc}) - Y_0 \rangle}{S} \right)^s = \dot{\lambda}^p Y_D^* \tag{32}$$

where $\underline{n} = \partial f / \partial \underline{\sigma}$ is the outward normal to the yield surface in the rotated stress space, $\widetilde{\underline{n}} = \partial f / \partial \widetilde{\underline{\sigma}}$ is the same normal but expressed in the effective rotated stress space, and Y_D^* is introduced in Eq. 32 for the simple convenience. From Eq. 27, the normal to the yield surface in the rotated effective stress space is given by

$$\widetilde{\underline{n}} = \begin{cases} \dfrac{3}{2} \dfrac{(\underline{\sigma}^{dev} - \underline{X})}{\|\underline{\sigma} - \underline{X}\|_M} & \text{if isotropic von Mises plastic flow} \\ \dfrac{\underline{\underline{H}} : (\underline{\sigma} - \underline{X})}{\|\underline{\sigma} - \underline{X}\|_H} & \text{if anisotropic Hill plastic flow} \end{cases} \tag{33}$$

The plastic multiplier $\dot{\lambda}^p$ appearing in all the evolution equations given above is also deeply affected by the micromorphic variables and consequently is of a nonlocal nature. In this case of time-independent plasticity, $\dot{\lambda}^p$ is classically deduced from the consistency condition applied to the yield function (Eq. 30): $\dot{f}(\underline{\sigma}, R, \underline{X}, D) = 0$ if $f(\underline{\sigma}, R, \underline{X}, D) = 0$. By limiting ourselves to the corotationnel or Jaumann frame for the sake of shortness (see Badreddine et al. (2010) for the Green–Naghdi total or plastic frames), the plastic multiplier can be calculated. After some algebraic transformations through the derivatives of the stress-like variables and with the help of the evolution equations, one can obtain

$$\dot{\lambda}^p = \frac{1}{H_{pd}} \left\langle \frac{2}{3} \left(3\mu_e \sqrt{(1-D)} \underline{D} + \overset{\smile}{C} \overset{\smile}{\underline{\alpha}} \right) : \widetilde{\underline{n}} + \overset{\smile}{Q} \overset{\smile}{r} \right\rangle \tag{34}$$

in which $H_{pd} > 0$ is the generalized elastoplastic hardening modulus defined, in the rotated strain-like variables space, by the following additive contributions:

$$H_{pd} = H_{pd}^{loc} + H_{pd}^{nloc} \tag{35}$$

where the local H_{pd}^{loc} and nonlocal H_{pd}^{nloc} contributions are given, in terms of stress-like variables, by

$$H_{pd}^{loc} = 3\mu + C + Q - \frac{1}{\sqrt{(1-D)}} (a\widetilde{\underline{n}} : \underline{X}_{loc} + bR_{loc}) + \frac{Y_D^* \sigma_y}{2(1-D)} \tag{36}$$

$$H_{pd}^{nloc} = \breve{C} + \breve{Q} - a\frac{\breve{C}}{C}\left[\frac{\underline{X}_{loc}}{\sqrt{(1-D)}} - \frac{\left(C+\breve{C}\right)}{C}\breve{\underline{X}}\right] : \underline{\widetilde{n}} - b\frac{\breve{Q}}{Q}\left[\frac{R_{loc}}{\sqrt{(1-D)}} - \frac{\left(Q+\breve{Q}\right)}{Q}\breve{R}\right]$$
$$- \frac{Y_D^*}{2(1-D)}\left[\left(\frac{\frac{\breve{C}}{C}\underline{X}_{loc}}{\sqrt{(1-D)}} - \breve{\underline{X}}\right) : \underline{\widetilde{n}} + \frac{\breve{Q}}{Q}\frac{R_{loc}}{\sqrt{(1-D)}} - \breve{R}\right] \tag{37}$$

Finally, the equivalent (or accumulated) plastic strain rate is still related, as in a fully local case, to the plastic multiplier by

$$\dot{p} = \sqrt{\frac{2}{3}\underline{D}^p : \underline{D}^p} = \frac{\dot{\lambda}^p}{\sqrt{(1-D)}} \tag{38}$$

Clearly, the highly nonlocal character of both the plastic multiplier and the elastoplastic hardening modulus appears from Eqs. 34 to 38. It is worth noting that, if the micromorphic state variables are zero, the classical local constitutive equations are fully retrieved as a particular case (Saanouni 2012a, b). Also let's note that the analytical expression in Eq. 34 of the plastic multiplier is purely indicative, since $\dot{\lambda}^p$ is the main unknown to be determined numerically at each Gauss point of each finite element when the numerical solution is performed (see section "Local Integration Scheme: Computation of the State Variables at Each Integration Point").

To close this section, let's give the final expression of the elastoplastic tangent operators required for any implicit resolution scheme to solve the IBVP. These are obtained from the time derivatives of the stress tensor given by Eq. 20 using the constitutive equations and the plastic multiplier:

$$\begin{aligned}\dot{\underline{\sigma}} &= (1-D)\underline{\underline{\Lambda}} : \underline{D} - \dot{\lambda}^p\left[\sqrt{(1-D)}\underline{\underline{\Lambda}} : \underline{\widetilde{n}} + Y_D^*\frac{\underline{\sigma}}{(1-D)}\right] \\ &= \underline{\underline{L}}_{\sigma} : \underline{D} - \underline{\underline{L}}_{\alpha}^{\smile} : \dot{\breve{\alpha}} - \underline{L}_r^{\smile}\dot{\breve{r}}\end{aligned} \tag{39}$$

where $\underline{\underline{L}}_{\sigma}$ is the fourth-rank nonsymmetric elastoplastic-damage tangent operator defined by

$$\underline{\underline{L}}_{\sigma} = (1-D)\underline{\underline{\Lambda}} - \frac{\sqrt{(1-D)}}{H_{pd}}\left[\sqrt{(1-D)}\left(\underline{\widetilde{n}} : \underline{\underline{\Lambda}}\right)^T \otimes \left(\underline{\underline{\Lambda}} : \underline{\widetilde{n}}\right) + Y_D^*\left(\underline{\widetilde{n}} : \underline{\underline{\Lambda}}\right) \otimes \frac{\underline{\sigma}}{(1-D)}\right] \tag{40}$$

$\underline{\underline{L}}_{\breve{\alpha}}$ is a fourth-rank nonsymmetric tensor and $\underline{L}_r^{\smile}$ is a second-rank symmetric tensor representing the micromorphic tangent operators given by

$$\underline{\underline{L}}_{\breve{\alpha}} = \frac{2}{3}\frac{\breve{C}}{H_{pd}}\left[\underline{\underline{\tilde{n}}} \otimes \left(\underline{\underline{\Lambda}} : \underline{\underline{\tilde{n}}}\right) + Y_D^*\left(\underline{\underline{\tilde{n}}} \otimes \frac{\underline{\underline{\sigma}}}{(1-D)}\right)\right] \tag{41}$$

$$\underline{L}_{\breve{r}} = \frac{\breve{Q}}{H_{pd}}\left[\sqrt{(1-D)}\left(\underline{\underline{\Lambda}} : \underline{\underline{\tilde{n}}}\right) + Y_D^*\frac{\underline{\underline{\sigma}}}{(1-D)}\right] \tag{42}$$

Note that if these two micromorphic hardening contributions are zero (i.e., $\breve{C} = \breve{Q} = 0$), then $\underline{\underline{L}}_{\breve{\alpha}} = \underline{\underline{0}}$ and $\underline{L}_{\breve{r}} = \underline{0}$ and Eq. 39 shrinks to its well-known classical form found in Saanouni (2012a, b).

Extension to Large Irreversible Strains Framework

As indicated above, the small elastic strains and large inelastic (plastic or viscoplastic) strains as well as the multiplicative decomposition of the transformation gradient into reversible and irreversible parts (i.e., $\underline{F} = \underline{F}^{re}.\underline{F}^{ir}$) assumptions are assumed. This leads to the decomposition of the total strain rate $\underline{D} = \underline{L}^{Sy} = \left(\dot{\underline{F}}.\underline{F}^{-1}\right)^{Sy}$ according to $\underline{D} = \dot{\underline{\varepsilon}}^{re} + 2\left(\overline{\underline{\varepsilon}}^{re}.\overline{\underline{\Omega}}\right)^{Sy} + \overline{\underline{D}}^{ir} = \overset{\cdot}{\underline{\varepsilon}}_J^{re} + \underline{D}^{ir}$ in which the irreversible strain rate tensor is defined by $\underline{D}^{ir} = \left(\underline{L}^{ir}\right)^{Sy} = \left(\dot{\underline{F}}^{ir}.\left(\underline{F}^{ir}\right)^{-1}\right)^{Sy}$ and the reversible Jaumann strain rate $\overset{\cdot}{\underline{\varepsilon}}_J^{re}$ (Sy stands for the symmetric part). The extension of the constitutive equations developed above to the finite irreversible strains can be easily made in the framework of the so-called rotating frame formulation using various rotated frames as found, for example, in Sidoroff and Dogui (2001) or Nemat-Nasser (2004), Badreddine et al. (2010) among many others. The idea behind this rotating frame formulation is to write the constitutive equations on spatial deformed and damaged configuration locally rotated with the rotation orthogonal tensor $\underline{Q}$ to have the same fixed orientation as the initial configuration. This is the simplest way to fulfill the objectivity requirement. If the total rotation frames using either the Jaumann or the Green–Naghdi objective derivatives are used, the rotation tensor is governed by the following ordinary differential equation:

$$\begin{cases} \dot{\underline{Q}}.\underline{Q}^T = \underline{W}_Q = \underline{\Omega} - \overline{\underline{\Omega}} \\ \underline{Q}(t) = \underline{1} \quad for \quad t = t_0 \end{cases} \tag{43}$$

where $\underline{\Omega} = \underline{L}^{Sk} = \left(\dot{\underline{F}}.\underline{F}^{-1}\right)^{Sk}$ is the usual spatial total material spin (Sk stands for skew-symmetric part) in the actual frame and $\overline{\underline{\Omega}}$ is the same quantity defined in the rotated frame. First Eq. 43 is numerically integrated to compute the rotation tensor

$\underline{Q}$ which is used to rotate any second-rank tensor $\underline{T}$ and the fourth-rank tensor $\underline{\underline{T}}$ according to $\overline{\underline{T}} = \underline{Q}^T.\underline{T}.\underline{Q}$ and $\overline{\underline{\underline{T}}} = \left(\underline{Q} \otimes \underline{Q}^T\right) : \underline{\underline{T}} : \left(\underline{Q}^T \otimes \underline{Q}\right)$, respectively. Accordingly, all the tensorial variables have to be rotated in order to transport them to the rotated configuration where they will be numerically integrated (Badreddine et al. 2010; Saanouni 2012a, b). Note that since the additional degrees of freedom are independent from the deformation gradient $\underline{F}$, the rotating frame formulation to be used with the micromorphic formulation is still unchanged compared to the classical local formulation.

In what follows all the formulation and the numerical integration of the fully coupled micromorphic constitutive equations will be performed in the total rotated frame without any distinction in their notation for the sake of simplicity. Also the following definitions are used: $\underline{\varepsilon} = \int_t \underline{D} dt$ and $\underline{\varepsilon}^{ir} = \int_t \underline{D}^{ir} dt$. Throughout this chapter where plastic strains are involved, irreversible quantities $(-)^{ir}$ transform to plastic quantities $(-)^p$ as used in section "Proposed Fully Coupled Micromorphic Constitutive Equations" above.

Back to the Micromorphic Balance Equations

By using the micromorphic state relations given by Eqs. 24, 25, and 26, the micromorphic balance equations (7), (8), and (9) can be transformed to be rewritten in terms of the strain-like variables to get

$$\begin{cases} \breve{C}^g Lap\left(\breve{\underline{\alpha}}\right) + \breve{C}\left(\sqrt{(1-D)}\underline{\alpha} - \breve{\underline{\alpha}}\right) - \left(div\left(\vec{\underline{f}}^{g\breve{\underline{\alpha}}}\right) - \underline{f}^{\breve{\underline{\alpha}}}\right) = \rho\zeta_{\breve{\alpha}}\ddot{\breve{\underline{\alpha}}} \ \text{ in } \ \Omega \\ \left(\breve{C}^g\left(\vec{\nabla}\,\breve{\underline{\alpha}}\right) - \vec{\underline{f}}^{g\breve{\underline{\alpha}}}\right).\vec{n} = \underline{F}^{\breve{\alpha}} \qquad \text{on } \Gamma_{\underline{F}^{\breve{\alpha}}} \end{cases} \tag{44}$$

$$\begin{cases} \breve{Q}^g Lap\left(\breve{D}\right) + \breve{Q}\left(\sqrt{(1-D)}r - \breve{r}\right) - \left(div\left(\vec{f}^{g\breve{r}}\right) - f^{\breve{r}}\right) = \rho\zeta_{\breve{r}}\ddot{\breve{r}} \quad \text{in } \ \Omega \\ \left(\breve{Q}^g\left(\vec{\nabla}\ \breve{r}\right) - \vec{f}^{g\breve{r}}\right).\vec{n} = F^{\breve{r}} \qquad \text{on } \Gamma_{F^{\breve{r}}} \end{cases} \tag{45}$$

$$\begin{cases} \breve{H}^g Lap\left(\breve{D}\right) + \breve{H}\left(D - \breve{D}\right) - \left(div\left(\vec{f}^{g\breve{D}}\right) - f^{\breve{D}}\right) = \rho\zeta_{\breve{d}}\ddot{\breve{D}} \quad \text{in } \ \Omega \\ \left(\breve{H}^g\left(\vec{\nabla}\ \breve{D}\right) - \vec{f}^{g\breve{D}}\right).\vec{n} = F^{\breve{D}} \qquad \text{on } \Gamma_{F^{\breve{D}}} \end{cases} \tag{46}$$

where the notation $Lap(X)$ stands for the Laplacian of (X). Without any available experimental information concerning the overall micromorphic body and contact forces, it is very difficult to choose their values. Accordingly, and without limiting

the generalities, these forces are neglected by tacking $\overrightarrow{\underset{\sim}{f}}^{g\underset{\sim}{\alpha}} = \overrightarrow{0}$, $\underset{\sim}{f}^{\underset{\sim}{\alpha}} = \underline{0}$, $\overrightarrow{f}^{g\underset{\sim}{r}} = \overrightarrow{f}^{g\underset{\sim}{D}} = \overrightarrow{0}$, $f^{\underset{\sim}{r}} = f^{\underset{\sim}{D}} = 0$, $F^{\underset{\sim}{D}} = F^{\underset{\sim}{r}} = 0$, and $\underline{F}^{\underset{\sim}{\alpha}} = \underline{0}$; the above micromorphic balance equations (Eqs. 44, 45, and 46) are simplified and rearranged under the following form:

$$\begin{cases} \ell_{\underset{\sim}{\alpha}}^2 Lap\left(\underset{\sim}{\underline{\alpha}}\right) + \left(\sqrt{(1-D)}\underline{\alpha} - \underset{\sim}{\underline{\alpha}}\right) = \rho \dfrac{\zeta_{\underset{\sim}{\alpha}}}{\underset{\sim}{C}} \ddot{\underset{\sim}{\underline{\alpha}}} & \text{in } \Omega \\ \left(\underset{\sim}{C}^g \left(\vec{\nabla}\, \underset{\sim}{\underline{\alpha}}\right)\right).\vec{n} = \underline{0} & \text{on } \Gamma_{\underline{F}^{\underset{\sim}{\alpha}}} \end{cases} \tag{47}$$

$$\begin{cases} \ell_{\underset{\sim}{r}}^2 Lap\left(\underset{\sim}{r}\right) + \left(\sqrt{(1-D)}r - \underset{\sim}{r}\right) = \dfrac{\rho\zeta_{\underset{\sim}{r}}}{\underset{\sim}{Q}} \ddot{\underset{\sim}{r}} & \text{in } \Omega \\ \left(\underset{\sim}{Q}^g \left(\vec{\nabla}\, \underset{\sim}{r}\right)\right).\vec{n} = 0 & \text{on } \Gamma_{F^{\underset{\sim}{r}}} \end{cases} \tag{48}$$

$$\begin{cases} \ell_{\underset{\sim}{d}}^2 Lap\left(\underset{\sim}{D}\right) + \left(D - \underset{\sim}{D}\right) = \rho \dfrac{\zeta_{\underset{\sim}{d}}}{\underset{\sim}{H}} \ddot{\underset{\sim}{D}} & \text{in } \Omega \\ \left(\underset{\sim}{H}^g \left(\vec{\nabla}\, \underset{\sim}{D}\right)\right).\vec{n} = 0 & \text{on } \Gamma_{F^{\underset{\sim}{D}}} \end{cases} \tag{49}$$

where $\ell_{\underset{\sim}{\alpha}}$, $\ell_{\underset{\sim}{r}}$, and $\ell_{\underset{\sim}{d}}$ are the internal length scales relative to the micromorphic kinematic hardening, isotropic hardening, and damage, respectively, all defined as the ratio of the micromorphic moduli by

$$\ell_{\underset{\sim}{\alpha}} = \sqrt{\frac{\underset{\sim}{C}^g}{\underset{\sim}{C}}} \quad (a), \qquad \ell_{\underset{\sim}{r}} = \sqrt{\frac{\underset{\sim}{Q}^g}{\underset{\sim}{Q}}} \quad (b), \qquad \ell_{\underset{\sim}{d}} = \sqrt{\frac{\underset{\sim}{H}^g}{\underset{\sim}{H}}} \quad (c) \tag{50}$$

Let's mention that if the micromorphic damage inertia is neglected in Eq. 49 (i.e., $\zeta_{\underset{\sim}{d}} = 0$), it leads to the well-known Helmholtz equation proposed in the framework of the so-called implicit nonlocal damage formulations proposed in Engelin et al. (2003), Geers et al. (2003), Geers (2004) and widely used in the literature to regularize the IBVPs exhibiting damage-induced softening.

Modeling of the Contact with Friction

In sheet or bulk metal forming by large inelastic strains, the contact and friction between the deformed part and the rigid and/or deforming tools, as well as the auto-contact and friction between different parts of the same deforming solid, cannot be

ignored. In fact, the thermal and mechanical phenomena occurring at the tools–part interfaces, or even at the boundaries of the parts from the same deforming solid coming in auto-contact, have to be modeled accurately as a constitutive equations for the contact interfaces. On the other hand, these contact interfaces evolve during the deformation process and are therefore among the unknowns of the evolution problem. This introduces severe geometrical nonlinearities making the IBVP hard to solve, especially when the contact interfaces undergo large changes during the deformation history.

The first difficulty concerns the determination of contact interfaces during the deformation history. Without going into details, the system of solids in contact should fulfill additional equations, namely, (i) the impenetrability condition, (ii) the non-adhesion condition, and (iii) the contact unilaterality condition. These three conditions lead to the formulation of Kuhn–Tucker optimality conditions in terms of the local velocity and the contact force at any point of the contact interface:

$$F_N \geq 0 \ , \ \dot{\vartheta}_N \leq 0 \ et \ F_N \dot{\vartheta}_N = 0 \quad sur \ \Gamma_t^c \tag{51}$$

in which F_N is the normal component of the contact force and $\dot{\vartheta}_N$ is the normal interpenetration rate in the direction of the outward normal to the contact interface Γ_t^c. Note simply, that the contact forces at each point of the contact interfaces should be added to the external forces defining the virtual power of external forces defined by Eq. 3 under the contact constraint defined by Eq. 51 above. More details related to the formulation and numerical treatment of the contact conditions can be found in the following recent books (Zhong 1993; Laursen 2002; Wriggers 2002).

Concerning the friction modeling, it can be performed in the framework of the thermodynamics of irreversible processes similar to the constitutive equations developed above. Various time-independent and time-dependent friction constitutive equations can be formulated in this framework (Zhong 1993; Laursen 2002; Wriggers 2002; Saanouni 2012a, b). Here, only the formulation of time-independent friction model from which the well-known Coulomb friction model can be obtained as a particular case (Saanouni 2012a, b).

For metallic materials, contact surfaces always contain asperities, the size (roughness) of which greatly influences friction conditions during slip between these surfaces. Indeed, the contact between two very smooth surfaces is governed by the asperities (state of the surface) present on each of the contacting surfaces. If these two contact solids are loaded, the asperities are deformed during the relative displacement between the pairs of contacting surfaces. If the imposed load does not exceed a certain threshold, the asperities only undergo reversible (elastic) deformation while remaining interlocked, so that if the solids are unloaded, then the asperities return to their original configuration. This is referred as reversible relative slip between the two solids. However, if the applied load exceeds a certain

threshold, then the asperities undergo irreversible deformation so that after unloading the solids can no longer return to their initial preload configuration which is called irreversible slip.

Based on these observed facts, the normal ($\dot{\vartheta}_N$) and tangential ($\vec{\dot{\vartheta}}_\tau$) relative slip rates at any material point lying in the contact interfaces are additively decomposed into reversible and irreversible parts according to

$$\dot{\vartheta}_N = \dot{\vartheta}_N^{re} + \dot{\vartheta}_N^{ir}, \vec{\dot{\vartheta}}_\tau = \vec{\dot{\vartheta}}_\tau^{re} + \vec{\dot{\vartheta}}_\tau^{ir} \tag{52}$$

In practice, the tangential reversible slip rate is termed the **adhesion** or **friction** rate between two solids at the contact point ($\vec{\dot{\vartheta}}_\tau^{re} = \vec{\dot{\vartheta}}_\tau^{fr}$), while the tangential irreversible slip rate is the true **slip** rate between the surfaces ($\vec{\dot{\vartheta}}_\tau^{ir} = \vec{\dot{\vartheta}}_\tau^{sl}$). In a similar way to cumulative inelastic strain rate (see Eq. 38), the following cumulative tangential slip rate is defined:

$$\left\| \vec{\dot{\vartheta}}_\tau \right\| = \sqrt{\vec{\dot{\vartheta}}_\tau^{ir} \cdot \vec{\dot{\vartheta}}_\tau^{ir}} = \sqrt{\vec{\dot{\vartheta}}_\tau^{sl} \cdot \vec{\dot{\vartheta}}_\tau^{sl}} = \dot{\vartheta}_\tau \tag{53}$$

Following the thermodynamical approach, the following couples of state variables are introduced at each point from the contact interface:

- $\left(\vec{\vartheta}_\tau^{fr}, \vec{F}_\tau^{fr}\right)$: The adhesion $\vec{\vartheta}_\tau^{fr}$ associated with the tangential friction force $\vec{F}_\tau^{fr}$
- $\left(\vec{\vartheta}_\tau^{sl}, \vec{F}_\tau^{sl}\right)$: The tangential slip $\vec{\vartheta}_\tau^{sl}$ associated with the irreversible part of the slip tangential force $\vec{F}_\tau^{sl}$ known as "tear" force
- $(\vartheta_\tau, F_\tau^w)$: The cumulative tangential slip ϑ_τ associated with the tangential wear force F_τ^w.

The friction (or adhesion) force $\vec{F}_\tau^{fr}$ and the slip (or tear) force $\vec{F}_\tau^{sl}$ (vectors) introduce a kind the friction anisotropy, i.e., dependence on directions in space (similar to the kinematic hardening for the plastic flow), whereas the wear force F_τ^w is a scalar which governs the size of the friction domain (similar to the isotropic hardening in plasticity). Thus, a formal analogy can be established between the adhesion $\vec{F}_\tau^{fr}$, tear $\vec{F}_\tau^{sl}$, and wear F_τ^w forces and the force-like state variables $\underline{\sigma}, \underline{X}$, and R which are associated with plastic flow, kinematic hardening, and isotropic hardening, respectively.

According to this analogy, a friction state potential is postulated while ignoring the strong coupling with both temperature and damage which only act as simple parameters:

$$\rho\psi_{fr}\left(\vec{\vartheta}_{\tau}^{fr},\vec{\vartheta}_{\tau}^{sl},\vartheta_{\tau};T,D,\ldots\right)=\frac{1}{2}\vec{\vartheta}_{\tau}^{fr}.\left[\underline{C}^{fr}(T,D,\ldots)\right].\vec{\vartheta}_{\tau}^{fr}+\frac{1}{2}\vec{\vartheta}_{\tau}^{sl}.\left[\underline{C}^{sl}(T,D,\ldots)\right].\vec{\vartheta}_{\tau}^{sl}+\frac{1}{2}\left[Q^{w}(T,D,\ldots)\right](\vartheta_{\tau})^{2} \tag{54}$$

Hence, the relative contact–friction state relationships are simply deduced from the friction state potential according to

$$\vec{F}_{\tau}^{fr}=\rho\frac{\partial\psi_{fr}}{\partial\vec{\vartheta}_{\tau}^{fr}}=\left[\underline{C}^{fr}(T,D,\ldots)\right].\vec{\vartheta}_{\tau}^{fr} \tag{55}$$

$$\vec{F}_{\tau}^{sl}=\rho\frac{\partial\psi_{fr}}{\partial\vec{\vartheta}_{\tau}^{sl}}=\left[\underline{C}^{sl}(T,D,\ldots)\right].\vec{\vartheta}_{\tau}^{sl} \tag{56}$$

$$F_{\tau}^{w}=\rho\frac{\partial\psi_{fr}}{\partial\vartheta_{\tau}}=\left[Q^{w}(T,D,\ldots)\right]\vartheta_{\tau} \tag{57}$$

where $\underline{C}^{fr}(T,D,\ldots)$ is the "penalty" second-rank symmetric tensor which represents the elasticity of the asperities of contact surfaces, $\underline{C}^{sl}(T,D,\ldots)$ is a "roughness" symmetric second-rank tensor which characterizes the tear of asperities in the three directions of the space, and $Q^{w}(T,D,\ldots)$ is a contact surface roughness "modulus." All of these operators are dependent on temperature and other physical phenomena, such as, for example, damage caused by microcracking, wear, flaking, and so on (Saanouni 2012a, b).

Similarly, complementary relationships which define the flux variables $\dot{\vec{\vartheta}}_{\tau}^{fr}$, $\dot{\vec{\vartheta}}_{\tau}^{sl}$, and $\dot{\vartheta}_{\tau}$ may be defined from a "friction" yield criterion $f_f\left(\vec{F}_{\tau}^{fr},\vec{F}_{\tau}^{sl},F_{\tau}^{w};T,D,\ldots\right)$ and a "friction" potential $F_f\left(\vec{F}_{\tau}^{fr},\vec{F}_{\tau}^{g},F_{\tau}^{w};T,D,\ldots\right)$ in the context of a nonassociative theory. These are chosen under the following form (Saanouni 2012a, b):

$$f_f\left(\vec{F}_{\tau}^{fr},\vec{F}_{\tau}^{sl},F_{\tau}^{w};T,D,\ldots\right)=\left\|\vec{F}_{\tau}^{fr}-\vec{F}_{\tau}^{sl}\right\|_f-\eta F_N^{fr}-F_{\tau}^{w}-F_y\leq 0 \tag{58}$$

$$F_f\left(\vec{F}_{\tau}^{fr},\vec{F}_{\tau}^{sl},F_{\tau}^{w};T,D,\ldots\right)=f_f+\frac{a^{sl}}{2}\vec{F}_{\tau}^{sl}.\left(\underline{C}^{sl}\right)^{-1}.\vec{F}_{\tau}^{sl}+\frac{b^{w}}{2Q^{w}}\left(F_{\tau}^{w}\right)^{2} \tag{59}$$

where F_y is a threshold tangential force (yield limit force), a^{sl} is a nonlinearity coefficient of the tear, and b^{w} is the nonlinearity parameter of wear. The norm of the tangential effective forces may be chosen as quadratic isotropic or anisotropic:

$$\left\|\vec{F}_{\tau}^{fr}-\vec{F}_{\tau}^{sl}\right\|_f=\begin{cases}\sqrt{\frac{1}{2}\left(\vec{F}_{\tau}^{fr}-\vec{F}_{\tau}^{sl}\right).\left(\vec{F}_{\tau}^{fr}-\vec{F}_{\tau}^{sl}\right)} & \text{if} \quad \text{isotropic friction}\\ \sqrt{\left(\vec{F}_{\tau}^{fr}-\vec{F}_{\tau}^{sl}\right).\underline{\chi}.\left(\vec{F}_{\tau}^{fr}-\vec{F}_{\tau}^{sl}\right)} & \text{if} \quad \text{anisotropic friction}\end{cases} \tag{60}$$

in which the second-rank symmetric tensor $\underline{\chi}(T,D,\ldots)$ defines the anisotropy of the friction "flow" depending in the absolute temperature and the ductile local damage among other phenomena. The application of the normality rule enables the derivation of the evolution relationships associated with friction, tear, and wear:

$$\dot{\vec{\vartheta}}_{\tau}^{fr}=\dot{\lambda}_f\frac{\partial F_f}{\partial\left(\vec{F}_{\tau}^{fr}\right)}=\dot{\lambda}_f\frac{\partial f_f}{\partial\left(\vec{F}_{\tau}^{fr}\right)}=\dot{\lambda}_f\vec{n}_f \tag{61}$$

$$\dot{\vec{\vartheta}}_{\tau}^{sl}=-\dot{\lambda}_f\frac{\partial F_f}{\partial\left(\vec{F}_{\tau}^{sl}\right)}=-\dot{\lambda}_f\left(\frac{\partial f_f}{\partial\left(\vec{F}_{\tau}^{sl}\right)}+a^{sl}\left(\overline{\underline{C}}^{sl}\right)^{-1}.\vec{F}_{\tau}^{sl}\right)=\dot{\lambda}_f\left(\vec{n}_f-a^{sl}\overline{\vec{\vartheta}}_{v}^{sl}\right) \tag{62}$$

$$\dot{\vartheta}=-\dot{\lambda}_f\frac{\partial F_f}{\partial\left(F_{\tau}^{w}\right)}=-\dot{\lambda}_f\left(\frac{\partial f_f}{\partial\left(F_{\tau}^{w}\right)}+\frac{b^w}{Q^w}F_T^w\right)=\dot{\lambda}_f(1-b^w\vartheta_{\tau}) \tag{63}$$

where the outward normal to the friction surface is given by

$$\vec{n}_f=\frac{\partial f_f}{\partial\left(\vec{F}_{\tau}^{fr}\right)}=\begin{cases}\dfrac{\left(\vec{F}_{\tau}^{fr}-\vec{F}_{\tau}^{sl}\right)}{\left\|\vec{F}_{\tau}^{fr}-\vec{F}_{\tau}^{sl}\right\|_f} & \text{if} \quad \text{isotropic friction}\\ \dfrac{\underline{\chi}.\left(\vec{F}_{\tau}^{fr}-\vec{F}_{\tau}^{sl}\right)}{\left\|\vec{F}_{\tau}^{fr}-\vec{F}_{\tau}^{sl}\right\|_f} & \text{if} \quad \text{anisotropic friction}\end{cases} \tag{64}$$

Finally, the consistency condition applied to the friction criterion, $\dot{f}_f=0$ if $f_f=0$, allows obtaining the analytic expression of the friction multiplier $\dot{\lambda}_f$ as a solution of the following equation:

$$\frac{\partial f_f}{\partial\left(\vec{F}_{\tau}^{fr}\right)}.\dot{\vec{F}}_{\tau}^{fr}+\frac{\partial f_f}{\partial\left(\vec{F}_{\tau}^{sl}\right)}.\dot{\vec{F}}_{\tau}^{sl}+\frac{\partial f_f}{\partial\left(F_{\tau}^{w}\right)}\dot{F}_{\tau}^{w}+\frac{\partial f_f}{\partial\left(F_N^{fr}\right)}\dot{F}_N^{fr}+\frac{\partial f_f}{\partial T}\dot{T}+\frac{\partial f_f}{\partial D}\dot{D}+\ldots=0 \tag{65}$$

The development of this advanced friction model is stopped here while pointing out that the well-known Coulomb friction model can be found as a particular case of

the present model when the "hardening" phenomena are neglected leading to a perfect friction model described in the associative theory framework by the following friction surface convex function of the normal F_N and tangential $\vec{F}_\tau$ forces (Saanouni 2012a, b):

$$f_f\left(F_N, \vec{F}_\tau; T, \ldots\right) = \left\|\vec{F}_\tau\right\|_f - \eta(T, D, \ldots)F_N - F_y(T, D, \ldots) \leq 0 \qquad (66)$$

where $\eta(T, D, \ldots)$ is a material parameter which characterizes the friction of the contact interfaces and $F_y(T, D, \ldots)$ is a material parameter which characterizes the adhesion threshold. Both are temperature and damage dependent among others.

Numerical Aspects

The overall balance equations (6), (47), (48), and (49) are nothing but the strong forms of a highly nonlinear and fully coupled initial and boundary value problem (IBVP) governing the equilibrium of the micromorphic solid. In this section, the numerical aspects to solve the multifunctional IBVP are briefly discussed on the framework of the finite element code ABAQUS/Explicit® based exclusively on updated Lagrangian formulation.

Weak Form of the IBVP

Let Ω be the volume of the micromorphic solid at a typical time t and Γ its boundary relative to the deformed configuration. The position of each particle from this micromorphic body is depicted by the spatial Cartesian coordinates x, y, z. In the absence of the micromorphic body and contact forces, the solid under concern is supposed to be subject to the classical body force $\vec{f}^u$, the displacement or velocity field on the boundary Γ_u, and the force $\vec{F}^u$ on the boundary Γ_F with the usual conditions $\Gamma_u \cup \Gamma_F = \Gamma$ and $\Gamma_u \cap \Gamma_F = \varnothing$ (all the micromorphic body and surface forces being neglected).

By using the classical weighted residual method together with the Galerkin assumption, and after performing the required integration by part taking into account the Neumann boundary conditions, the weak forms of the above defined IBVP can be easily obtained under the following forms (Hamed 2012; Saanouni 2012a, b; Saanouni and Hamed 2013):

$$
\left\{
\begin{aligned}
&J_u\left(\vec{u},\delta\vec{u}\right)=\int_\Omega \rho\vec{\ddot{u}}.\delta\vec{u}d\Omega+\int_\Omega \underline{\sigma}:\left(\vec{\nabla}\ \delta\vec{u}\right)^{sym}d\Omega-\int_\Omega \vec{f}^{\,u}.\delta\vec{u}d\Omega-\\
&\qquad\int_{\Gamma_F}\vec{F}^{u}.\delta\vec{u}d\Omega-\int_{\Gamma_c}\vec{F}^{c}.\delta\vec{u}d\Omega=0\quad\forall\delta\vec{u}\ K.A.\quad\text{(a)}\\
&J_{\breve{D}}\left(\breve{D},\delta\breve{D}\right)=\int_\Omega \rho\frac{\breve{\zeta}_d}{\breve{H}}\ddot{\breve{D}}\delta\dot{\breve{D}}\,d\Omega+\int_\Omega \ell_d^2\left(\vec{\nabla}\ \breve{D}\right).\left(\vec{\nabla}\ \delta\dot{\breve{D}}\right)d\Omega-\\
&\qquad\int_\Omega\left(D-\breve{D}\right)\delta\dot{\breve{D}}\,d\Omega=0\qquad\forall\delta\dot{\breve{D}}\ K.A.\quad\text{(b)}\\
&J_{\breve{r}}\left(\breve{r},\delta\dot{\breve{r}}\right)=\int_\Omega\frac{\rho\breve{\zeta}_{\breve{r}}}{\breve{Q}}\ddot{\breve{r}}\delta\dot{\breve{r}}\,d\Omega-\int_\Omega l_{\breve{r}}^2\left(\vec{\nabla}\ \breve{r}\right).\left(\vec{\nabla}\ \delta\dot{\breve{r}}\right)d\Omega-\\
&\qquad\int_\Omega\left(\left(\sqrt{1-D}\right)r-\breve{r}\right)\delta\dot{\breve{r}}\,d\Omega=0\qquad\forall\delta\dot{\breve{r}}\ K.A.\quad\text{(c)}\\
&J_{\underline{\breve{\alpha}}}\left(\underline{\breve{\alpha}},\delta\underline{\breve{\alpha}}\right)=\int_\Omega\frac{\rho\breve{\zeta}_{\underline{\alpha}}}{\breve{C}}\ddot{\underline{\breve{\alpha}}}:\delta\dot{\underline{\breve{\alpha}}}d\Omega+\int_\Omega l_{\underline{\alpha}}^2\vec{\nabla}\ \underline{\breve{\alpha}}\cdot\vec{\nabla}\ \delta\dot{\underline{\breve{\alpha}}}d\Omega-\\
&\qquad\frac{2}{3}\int_\Omega\left(\left(\sqrt{1-D}\right)\underline{\alpha}-\underline{\breve{\alpha}}\right):\delta\dot{\underline{\breve{\alpha}}}d\Omega=0\qquad\forall\delta\dot{\underline{\breve{\alpha}}}\ K.A.\quad\text{(d)}.
\end{aligned}
\right.
\tag{67}
$$

where $\delta\vec{u}$, $\delta\dot{\breve{D}}$, $\delta\dot{\breve{r}}$, and $\delta\dot{\underline{\breve{\alpha}}}$ are arbitrary kinematically admissible *K. A.* (i.e., fulfilling the associated Dirichlet boundary conditions) virtual velocity, virtual micromorphic damage rate, virtual isotropic hardening rate, and virtual kinematic hardening rate fields, respectively, while $\vec{F}^c$ in the last LHS of Eq. 27 stands for the contact force at the contact interface which is the solution of the contact–friction model discussed above.

Time and Space Discretization

The IBVP defined by Eq. 67 has to be solved on the space and time domain $\Omega \times [t_0, t_f]$ in which t_0 and t_f are the initial and final times of the interval during which the external loading path (purely mechanical) is applied.

The total time interval $I_t = [t_0, t_f]$ is thus discretized into (Nt) subintervals of nonconstant size Δt, so that the approximation $I_t \cong \bigcup_{n=0}^{Nt}[t_n, t_{n+1} = t_n + \Delta t]$ holds. For each of these time subintervals, one has to solve a nonlinear problem to determine all the unknowns of the IBVP at the time t_{n+1} while supposing fully known their values at t_n.

With this time discretization in hand, and considering the updated Lagrangian formulation, the last equilibrium configuration obtained at time t_n, called Ω_{t_n}, is taken as the reference configuration while seeking to determine completely the equilibrium configuration $\Omega_{t_{n+1}}$ at time t_{n+1} as well as the overall kinematic variables ($\vec{u}_{n+1}, \breve{D}_{n+1}, \breve{r}_{n+1}$, and $\breve{\underline{\alpha}}_{n+1}$) and local state variables ($\underline{\varepsilon}^p_{n+1}, \underline{\sigma}_{n+1}, \underline{X}_{n+1}\left(\underline{\alpha}_{n+1}\right), R_{n+1}$ $\left(r_{n+1}\right)$ and D_{n+1}) in that configuration.

For the space discretization, the standard (Galerkin-type) displacement-based finite element method is used to discretize the reference configuration Ω_{t_n} into a finite number (*Nte*) of subdomains (or finite elements) with the simplest geometric form called Ω_e in such a manner that the approximation $\Omega_{t_n} \cong \bigcup_{e=1}^{Nte} \Omega_e$ holds with enough precision. In each finite element Ω_e of the domain Ω_{t_n} defined by (*Nen*) nodes, the main unknowns of the IBVP (here the displacement vector, the micromorphic damage, the micromorphic isotropic hardening, and the micromorphic kinematic hardening fields) are approximated, based on a nodal approximation on the subdomain, using the appropriate Lagrange-type polynomial interpolation functions. When the reference frame defined by the natural coordinates $\vec{\xi}^e$ is used to define the reference element Ω_r associated to Ω_e, the nodal unknowns (together with their associated virtual unknowns) are approximated on Ω_r using the classical matrix notation according to

$$\begin{cases} \{u^e(x,t)\} = [N^e(\xi_i)]\{u^e_I(t)\} \\ \{\delta u^e(x,t)\} = [N^e(\xi_i)]\{\delta u^e_I(t)\} \end{cases} \tag{68}$$

$$\begin{cases} \breve{D}^e(x,t) = \left[\breve{N}^e_{\breve{D}}(\xi_i)\right]\left\{\breve{D}^e_I(t)\right\} \\ \delta\breve{D}^e(x,t) = \left[\breve{N}^e_{\breve{D}}(\xi_i)\right]\left\{\delta\breve{D}^e_I(t)\right\} \end{cases} \tag{69}$$

$$\begin{cases} \breve{r}^e(x,t) = \left[\breve{N}^e_{\breve{r}}(\xi_i)\right]\left\{\breve{r}^e_I(t)\right\} \\ \delta\breve{r}^e(x,t) = \left[\breve{N}^e_{\breve{r}}(\xi_i)\right]\left\{\delta\breve{r}^e_I(t)\right\} \end{cases} \tag{70}$$

$$\begin{cases} \left\{\breve{\alpha}^e(x,t)\right\} = \left[\breve{N}^e_{\breve{\alpha}}(\xi_i)\right]\left\{\breve{\alpha}^e_I(t)\right\} \\ \left\{\delta\breve{\alpha}^e(x,t)\right\} = \left[\breve{N}^e_{\breve{\alpha}}(\xi_i)\right]\left\{\delta\breve{\alpha}^e_I(t)\right\} \end{cases} \tag{71}$$

where $[N^e]$, $\left[\breve{N}^e_{\breve{r}}\right]$, $\left[\breve{N}^e_{\breve{\alpha}}\right]$ and $\left[\breve{N}^e_{\breve{D}}\right]$ are the matrices of the interpolation (or shape) functions for each of the four nodal unknowns. The index I stands for the total number of dofs for each element (i.e., the total number of nodes of the element *Nen* times the number of unknowns per node). In 3D, this leads to an element having 11 unknowns for each node. The first and second derivatives of Eqs. 68, 69, 70, and 71 with respect to time allow the easy calculation of the corresponding velocity and acceleration fields:

$$\begin{cases} \{\dot{u}^e(x,t)\} = [N^e(\xi_i)]\{\dot{u}_I^e(t)\} \\ \{\delta\dot{u}^e(x,t)\} = [N^e(\xi_i)]\{\delta\dot{u}_I^e(t)\} \end{cases} \tag{72}$$

$$\begin{cases} \overset{\cdot}{\underset{\sim}{D}}\, e(x,t) = \left[\underset{\sim}{N}{}^e_D(\xi_i)\right]\left\{\overset{\cdot}{\underset{\sim}{D}}{}^e_I(t)\right\} \\ \delta\overset{\cdot}{\underset{\sim}{D}}\, e(x,t) = \left[\underset{\sim}{N}{}^e_D(\xi_i)\right]\left\{\delta\overset{\cdot}{\underset{\sim}{D}}{}^e_I(t)\right\} \end{cases} \tag{73}$$

$$\begin{cases} \overset{\cdot}{\underset{\sim}{r}}{}^e(x,t) = \left[\underset{\sim}{N}{}^e_r(\xi_i)\right]\left\{\overset{\cdot}{\underset{\sim}{r}}{}^e_I(t)\right\} \\ \delta\overset{\cdot}{\underset{\sim}{r}}{}^e(x,t) = \left[\underset{\sim}{N}{}^e_r(\xi_i)\right]\left\{\delta\overset{\cdot}{\underset{\sim}{r}}{}^e_I(t)\right\} \end{cases} \tag{74}$$

$$\begin{cases} \left\{\overset{\cdot}{\underset{\sim}{\alpha}}\, e(x,t)\right\} = \left[\underset{\sim}{N}{}^e_\alpha(\xi_i)\right]\left\{\overset{\cdot}{\underset{\sim}{\alpha}}{}^e_I(t)\right\} \\ \left\{\delta\overset{\cdot}{\underset{\sim}{\alpha}}\, e(x,t)\right\} = \left[\underset{\sim}{N}{}^e_\alpha(\xi_i)\right]\left\{\delta\,\overset{\cdot}{\underset{\sim}{\alpha}}{}^e_I(t)\right\} \end{cases} \tag{75}$$

$$\{\ddot{u}^e(x,t)\} = [N^e(\xi_i)]\{\ddot{u}_I^e(t)\} \tag{76}$$

$$\ddot{\underset{\sim}{D}}{}^{(e)}(x,t) = \left[\underset{\sim}{N}{}^e_D(\xi_i)\right]\left\{\ddot{\underset{\sim}{D}}{}^e_I(t)\right\} \tag{77}$$

$$\ddot{\underset{\sim}{r}}{}^{(e)}(x,t) = \left[\underset{\sim}{N}{}^e_r(\xi_i)\right]\left\{\ddot{\underset{\sim}{r}}{}^e_I(t)\right\} \tag{78}$$

$$\left\{\ddot{\underset{\sim}{\alpha}}{}^{(e)}(x,t)\right\} = \left[\underset{\sim}{N}{}^e_\alpha(\xi_i)\right]\left\{\ddot{\underset{\sim}{\alpha}}{}^e_I(t)\right\} \tag{79}$$

On the other hand, the first gradient of each real and virtual variable comes directly from Eqs. 68 to 71:

$$\begin{cases} \left\{\vec{\nabla}\ (\dot{u}^e)\right\} = \{\dot{\varepsilon}^e\} = \left[\dfrac{\partial N^e}{\partial \vec{x}}\right]\{\dot{u}_I^e\} = \left[\dfrac{\partial N^e}{\partial \vec{\xi}}\right]\left[\dfrac{\partial \vec{\xi}}{\partial \vec{x}}\right]\{\dot{u}_I^e\} = [B^e]\{\dot{u}_I^e\} \\ \left\{\vec{\nabla}\ (\delta\dot{u}^e)\right\} = \{\delta\dot{\varepsilon}^e\} = \left[\dfrac{\partial N^e}{\partial \vec{x}}\right]\{\delta\dot{u}_I^e\} = \left[\dfrac{\partial N^e}{\partial \vec{\xi}}\right]\left[\dfrac{\partial \vec{\xi}}{\partial \vec{x}}\right]\{\delta\dot{u}_I^e\} = [B^e]\{\delta\dot{u}_I^e\} \end{cases} \tag{80}$$

$$\begin{cases} \left\{\vec{\nabla}\ \overset{\cdot}{\underset{\sim}{D}}{}^e\right\} = \left[\dfrac{\partial \underset{\sim}{N}{}^e_D}{\partial \vec{x}}\right]\left\{\overset{\cdot}{\underset{\sim}{D}}{}^e_I\right\} = \left[\dfrac{\partial \underset{\sim}{N}{}^e_D}{\partial \vec{\xi}}\right]\left[\dfrac{\partial \vec{\xi}}{\partial \vec{x}}\right]\left\{\overset{\cdot}{\underset{\sim}{D}}{}^e_I\right\} = \left[\underset{\sim}{B}{}^e_D\right]\left\{\overset{\cdot}{\underset{\sim}{D}}{}^e_I\right\} \\ \left\{\vec{\nabla}\ \left(\delta\overset{\cdot}{\underset{\sim}{D}}{}^e\right)\right\} = \left[\dfrac{\partial \underset{\sim}{N}{}^e_D}{\partial \vec{x}}\right]\left\{\delta\overset{\cdot}{\underset{\sim}{D}}{}^e_I\right\} = \left[\dfrac{\partial \underset{\sim}{N}{}^e_D}{\partial \vec{\xi}}\right]\left[\dfrac{\partial \vec{\xi}}{\partial \vec{x}}\right]\left\{\delta\overset{\cdot}{\underset{\sim}{D}}{}^e_I\right\} = \left[\underset{\sim}{B}{}^e_D\right]\left\{\delta\overset{\cdot}{\underset{\sim}{D}}{}^e_I\right\} \end{cases} \tag{81}$$

$$\begin{cases} \left\{\vec{\nabla}\ \breve{r}^e\right\} = \left[\dfrac{\partial \breve{N}^e_{\breve{r}}}{\partial \vec{x}}\right]\left\{\breve{r}^e_I\right\} = \left[\dfrac{\partial \breve{N}^e_{\breve{r}}}{\partial \vec{\xi}}\right]\left[\dfrac{\partial \vec{\xi}}{\partial \vec{x}}\right]\left\{\breve{r}^e_I\right\} = \left[\breve{B}^e_{\breve{r}}\right]\left\{\breve{r}^e_I\right\} \\ \left\{\vec{\nabla}\ \left(\delta\breve{r}^e\right)\right\} = \left[\dfrac{\partial \breve{N}^e_{\breve{r}}}{\partial \vec{x}}\right]\left\{\delta\breve{r}^e_I\right\} = \left[\dfrac{\partial \breve{N}^e_{\breve{r}}}{\partial \vec{\xi}}\right]\left[\dfrac{\partial \vec{\xi}}{\partial \vec{x}}\right]\left\{\delta\breve{r}^e_I\right\} = \left[\breve{B}^e_{\breve{r}}\right]\left\{\delta\breve{r}^e_I\right\} \end{cases} \tag{82}$$

$$\begin{cases} \left\{\vec{\nabla}\ \breve{\alpha}^e\right\} = \left[\dfrac{\partial \breve{N}^e_{\breve{\alpha}}}{\partial \vec{x}}\right]\left\{\breve{\alpha}^e_I\right\} = \left[\dfrac{\partial \breve{N}^e_{\breve{\alpha}}}{\partial \vec{\xi}}\right]\left[\dfrac{\partial \vec{\xi}}{\partial \vec{x}}\right]\left\{\breve{\alpha}^e_I\right\} = \left[\breve{B}^e_{\breve{\alpha}}\right]\left\{\breve{\alpha}^e_I\right\} \\ \left\{\vec{\nabla}\ \left(\delta\breve{\alpha}^e\right)\right\} = \left[\dfrac{\partial \breve{N}^e_{\breve{\alpha}}}{\partial \vec{x}}\right]\left\{\delta\breve{\alpha}^e_I\right\} = \left[\dfrac{\partial \breve{N}^e_{\breve{\alpha}}}{\partial \vec{\xi}}\right]\left[\dfrac{\partial \vec{\xi}}{\partial \vec{x}}\right]\left\{\delta\breve{\alpha}^e_I\right\} = \left[\breve{B}^e_{\breve{\alpha}}\right]\left\{\delta\breve{\alpha}^e_I\right\} \end{cases} \tag{83}$$

With the help of Eqs. 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, and 83, the weak forms taken from Eq. 67 can be written for, a typical finite element (e), under the matrix notation as follows:

$$J^e_u\left(\vec{u}^e, \delta\vec{u}^e\right) = \{\delta\ddot{u}^e\}^T\left([M^e]\{\ddot{u}^e_I\} + \{F^e_{\text{int}}\} - \{F^e_{ext}\}\right) \tag{84}$$

$$J^e_{\breve{D}}\left(\breve{D}^e, \delta\breve{D}^e\right) = \left\{\delta\dot{\breve{D}}^e\right\}^T\left(\left[\breve{M}^e_{\breve{D}}\right]\left\{\ddot{\breve{D}}^e_I\right\} + \left\{\breve{F}^e_{\text{int}\,\breve{D}}\right\} - \left\{\breve{F}^e_{ext\,\breve{D}}\right\}\right) \tag{85}$$

$$J^e_{\breve{r}}\left(\breve{r}^e, \delta\dot{\breve{r}}^e\right) = \left\{\delta\dot{\breve{r}}^e\right\}^T\left(\left[\breve{M}^e_{\breve{r}}\right]\left\{\ddot{\breve{r}}^e_I\right\} + \left\{\breve{F}^e_{\text{int}\,\breve{r}}\right\} - \left\{\breve{F}^e_{ext\,\breve{r}}\right\}\right) \tag{86}$$

$$J^e_{\underline{\breve{\alpha}}}\left(\underline{\breve{\alpha}}^e, \delta\underline{\dot{\breve{\alpha}}}^e\right) = \left\{\delta\dot{\breve{\alpha}}^e\right\}^T\left(\left[\underline{\breve{M}}^e_{\underline{\breve{\alpha}}}\right]\left\{\ddot{\breve{\alpha}}^e_I\right\} + \left\{\breve{F}^e_{\text{int}\,\underline{\breve{\alpha}}}\right\} - \left\{\breve{F}^e_{ext\,\underline{\breve{\alpha}}}\right\}\right) \tag{87}$$

where the mass matrices and the internal and external loading vectors for the reference element under concern are:

- The consistent mass matrices:

$$[M^e] = \int_{\Omega_r} \rho[N^e]^T[N^e]J^e_v d\Omega_r \tag{88}$$

$$\left[\breve{M}^e_{\breve{D}}\right] = \int_{\Omega_r} \frac{\rho\breve{\zeta}_{\breve{D}}}{\breve{H}}\left[N^e_{\breve{D}}\right]^T\left[N^e_{\breve{D}}\right]J^e_v d\Omega_r \tag{89}$$

$$\left[\breve{M}^e_{\breve{r}}\right] = \int_{\Omega_r} \frac{\rho\breve{\zeta}_{\breve{r}}}{\breve{Q}}\left[N^e_{\breve{r}}\right]^T\left[N^e_{\breve{r}}\right]J^e_v d\Omega_r \tag{90}$$

$$\left[\underset{\sim}{M}^{e}_{\underset{\sim}{\alpha}}\right] = \int_{\Omega_r} \frac{\rho\zeta_{\underline{\underline{\alpha}}}}{\underset{\sim}{C}} \left[N^{e}_{\underset{\sim}{\alpha}}\right]^{T} \left[N^{e}_{\underset{\sim}{\alpha}}\right] J^{e}_{v} d\Omega_r \tag{91}$$

- The internal force's vectors:

$$\left\{F^{e}_{\text{int}}\right\} = \int_{\Omega_r} [B^{e}]^{T} \{\sigma^{e}\} J^{e}_{v} d\Omega_r \tag{92}$$

$$\left\{\underset{\sim}{F}^{e}_{\text{int}\,\underset{\sim}{D}}\right\} = \int_{\Omega_r} \left[\underset{\sim}{\ell}^{2}_{d} \left[\underset{\sim}{B}^{e}_{\underset{\sim}{D}}\right]^{T} \left[\underset{\sim}{B}^{e}_{\underset{\sim}{D}}\right] + \left[\underset{\sim}{N}^{e}_{\underset{\sim}{D}}\right]^{T} \left[\underset{\sim}{N}^{e}_{\underset{\sim}{D}}\right]\right] \left\{\underset{\sim}{D}^{e}\right\} J^{e}_{v} d\Omega_r \tag{93}$$

$$\left\{\underset{\sim}{F}^{e}_{\text{int}\,\underset{\sim}{r}}\right\} = \int_{\Omega_r} \left[\underset{\sim}{\ell}^{2}_{r} \left[\underset{\sim}{B}^{e}_{\underset{\sim}{r}}\right]^{T} \left[\underset{\sim}{B}^{e}_{\underset{\sim}{r}}\right] + \left[\underset{\sim}{N}^{e}_{\underset{\sim}{r}}\right]^{T} \left[\underset{\sim}{N}^{e}_{\underset{\sim}{r}}\right]\right] \left\{\underset{\sim}{r}^{e}\right\} J^{e}_{v} d\Omega_r \tag{94}$$

$$\left\{\underset{\sim}{F}^{e}_{int\,\underline{\underline{\alpha}}}\right\} = \int_{\Omega_r} \left[\underline{\underline{\ell}}^{2}_{\underline{\underline{\alpha}}} \left[\underset{\sim}{B}^{e}_{\underline{\underline{\alpha}}}\right]^{T} \left[\underset{\sim}{B}^{e}_{\underline{\underline{\alpha}}}\right] + \frac{2}{3} \left[\underset{\sim}{N}^{e}_{\underline{\underline{\alpha}}}\right]^{T} \left[\underset{\sim}{N}^{e}_{\underline{\underline{\alpha}}}\right]\right] \left\{\underset{\sim}{\alpha}^{e}\right\} J^{e}_{v} d\Omega_r \tag{95}$$

- The external force's vectors:

$$\left\{F^{e}_{ext}\right\} = \int_{\Omega_r} [N^{e}]^{T} \{f^{u}\} J^{e}_{v} d\Omega_r + \int_{\Gamma_{Fu_r}} [N^{e}]^{T} \{F^{u}\} J^{e}_{s} d\Gamma_r + \int_{\Gamma_{C_r}} \{F^{c}\} J^{e}_{s} d\Gamma_r \tag{96}$$

$$\left\{\underset{\sim}{F}^{e}_{ext\,\underset{\sim}{D}}\right\} = \int_{\Omega_r} D^{e} \left[\underset{\sim}{N}^{e}_{\underset{\sim}{D}}\right]^{T} J^{e}_{v} d\Omega_r \tag{97}$$

$$\left\{\underset{\sim}{F}^{e}_{ext\,\underset{\sim}{r}}\right\} = \int_{\Omega_r} \sqrt{1 - D^{e}}\, r^{e} \left[\underset{\sim}{N}^{e}_{\underset{\sim}{r}}\right]^{T} J^{e}_{v} d\Omega_r \tag{98}$$

$$\left\{\underset{\sim}{F}^{e}_{ext,\,\underline{\underline{\alpha}}}\right\} = \frac{2}{3} \int_{\Omega_r} \sqrt{1 - D^{e}} \{\alpha\} \left[\underset{\sim}{N}^{e}_{\underline{\underline{\alpha}}}\right]^{T} J^{e}_{v} d\Omega_r \tag{99}$$

Proceeding with the standard assembly over the total number (Nte) of elements of the structure $\left(\text{symbolized by the operator } \underset{e=1}{\overset{Nte}{\mathrm{A}}}\right)$ and keeping in mind that the algebraic system in Eq. 67 should be fulfilled for any *K.A.* virtual fields $\{\delta\dot{U}\}$, $\left\{\delta\,\underset{\sim}{\dot{r}}\right\}$, $\left\{\delta\,\underset{\sim}{\dot{r}}\right\}$, and $\left\{\delta\,\underset{\sim}{\dot{D}}\right\}$, the following fully discretized (in space and in time) system in Eq. 100 is obtained (here written at t_{n+1} end of the current time interval $[t_n,\ t_{n+1} = t_n + \Delta t]$):

$$\begin{cases} J_u^{n+1} = \underset{e=1}{\overset{N^{te}}{\mathrm{A}}} J^e_{n+1} = [M]_{n+1}\{\ddot{U}\}_{n+1} + \{F_{\text{int}}\}_{n+1} - \{F_{ext}\}_{n+1} = \{0\} \quad \text{(a)} \\ J_{\breve{D}}^{n+1} = \underset{e=1}{\overset{N^{te}}{\mathrm{A}}} J^e_{\breve{D},n+1} = \left[\breve{M}_{\breve{D}}\right]_{n+1}\left\{\breve{\ddot{D}}\right\}_{n+1} + \left\{\breve{F}_{\text{int}\,\breve{D}}\right\}_{n+1} - \left\{\breve{F}_{ext,\breve{D}}\right\}_{n+1} = \{0\} \quad \text{(b)} \\ J_{\breve{r}}^{n+1} = \underset{e=1}{\overset{N^{te}}{\mathrm{A}}} J^e_{\breve{r},n+1} = \left[\breve{M}_{\breve{r}}\right]_{n+1}\left\{\breve{\ddot{r}}\right\}_{n+1} + \left\{\breve{F}_{\text{int}\,\breve{r}}\right\}_{n+1} - \left\{\breve{F}_{ext,\breve{r}}\right\}_{n+1} = \{0\} \quad \text{(c)} \\ J_{\underline{\breve{\alpha}}}^{n+1} = \underset{e=1}{\overset{N^{te}}{\mathrm{A}}} J^e_{\underline{\breve{\alpha}},n+1} = \left[\breve{M}_{\breve{D}}\right]_{n+1}\left\{\breve{\underline{\ddot{\alpha}}}\right\}_{n+1} + \left\{\breve{F}_{\text{int},\,\underline{\breve{\alpha}}}\right\}_{n+1} - \left\{\breve{F}_{int,\,\underline{\breve{\alpha}}}\right\}_{n+1} = \{0\} \quad \text{(d)} \end{cases} \tag{100}$$

where the global mass matrices are $[M] = \underset{e=1}{\overset{Nte}{\mathrm{A}}}[M^e]$, $\left[\breve{M}_{\breve{D}}\right] = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\left[\breve{M}^e_{\breve{D}}\right]$, $\left[\breve{M}_{\breve{r}}\right] = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\left[\breve{M}^e_{\breve{r}}\right]$, and $\left[\breve{M}_{\underline{\breve{\alpha}}}\right] = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\left[\breve{M}^e_{\underline{\breve{\alpha}}}\right]$; the global nodal accelerations vectors are $\{\ddot{U}\} = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\{\ddot{u}^e\}$, $\left\{\breve{\ddot{D}}\right\} = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\left\{\breve{\ddot{D}}^e\right\}$, $\left\{\breve{\ddot{r}}\right\} = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\left\{\breve{\ddot{r}}^e\right\}$, and $\left\{\breve{\underline{\ddot{\alpha}}}\right\} = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\left\{\breve{\underline{\ddot{\alpha}}}^e\right\}$; the global internal forces vectors are $\{F_{\text{int}}\} = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\{F^e_{\text{int}}\}$, $\left\{\breve{F}_{\text{int}\,\breve{D}}\right\} = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\left\{\breve{F}^e_{\text{int}\,\breve{D}}\right\}$, $\left\{\breve{F}_{\text{int}\,\breve{r}}\right\} = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\left\{\breve{F}^e_{\text{int}\,\breve{r}}\right\}$, and $\left\{\breve{F}_{\text{int}\,\underline{\breve{\alpha}}}\right\} = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\left\{\breve{F}^e_{\text{int}\underline{\breve{\alpha}}}\right\}$; and finally the global external forces vectors are $\{F_{\text{ext}}\} = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\{F^e_{\text{ext}}\}$, $\left\{\breve{F}_{\text{ext}\,\breve{D}}\right\} = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\left\{\breve{F}^e_{\text{ext}\,\breve{D}}\right\}$, $\left\{\breve{F}_{\text{ext}\,\breve{r}}\right\} = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\left\{\breve{F}^e_{\text{ext}\,\breve{r}}\right\}$, and $\left\{\breve{F}_{\text{ext}\,\underline{\breve{\alpha}}}\right\} = \underset{e=1}{\overset{Nte}{\mathrm{A}}}\left\{\breve{F}^e_{\text{ext}\,\underline{\breve{\alpha}}}\right\}$.

Many finite elements (iso-parametric or sub-parametric with or without Hourglass control) can be constructed with, for each node, 11 unknowns for 3D problems: three components of the displacement vector, one micromorphic damage, one micromorphic isotropic hardening, and six components for the micromorphic kinematic hardening second-rank tensor. For the sake of shortness of this chapter, the description of the finite elements having additional micromorphic dofs is skipped and the reader is referred to Hamed (2012), Saanouni (2012a, b), and Saanouni and Hamed (2013).

Global Resolution Schemes

The fully discretized algebraic system in Eq. 100 is a hyperbolic, highly nonlinear system to be solved numerically over each typical time increment of size $\Delta t = t_{n+1} - t_n$ as discussed in section "Time and Space Discretization." The solution methodology consists of assuming to be known the complete solution $\mathbb{S}_n = \left\{U_n, \breve{D}_n, \breve{r}_n, \breve{\alpha}_n\right\}$ of this hyperbolic IBVP at the time t_n and of seeking a complete approximated solution $\mathbb{S}_{n+1} = \left\{U_{n+1}, \breve{D}_{n+1}, \breve{r}_{n+1}, \breve{\alpha}_{n+1}\right\}$ at time t_{n+1} that fulfill the Eq. 100 over the current time interval $[t_n,\ t_{n+1} = t_n + \Delta t]$.

Briefly speaking, this involves the linearization of the system in Eq. 100 thanks to appropriate time discretization and the solving of a linear or nonlinear problem over each time interval $[t_n,\ t_{n+1} = t_n + \Delta t]$ using iterative or direct non-iterative methods.

Two resolution schemes are often used: the static implicit (SI) scheme and the dynamic explicit (DE) or dynamic implicit (DI) schemes. Several solution schemes can be used according to the type and severity of the nonlinearities of the IBVP and depending on the imposed constraints, such as incompressibility conditions or contact conditions with our without friction (Oden 1972; Zienkiewicz and Taylor 1967; Owen and Hinton 1980; Crisfield 1991, 1997; Hinton 1992; Bonnet and Wood 1997; Ladeveze 1998; Simo and Hughes 1997; Belytschko et al. 2000; Reddy 2004; Ibrahimbegovic 2006; De Souza et al. 2008; Wriggers 2008). In general, the procedure of obtaining the numerical solutions of a nonlinear IBVP such as in Eq. 100 is drawn either from explicit non-iterative schemes or from implicit iterative schemes. However, two resolution schemes are often used: the static implicit (SI) scheme and the dynamic explicit (DE) scheme.

Outlines of the SI Scheme

If the inertia effects are neglected (quasi-static problems), the system in Eq. 100 transforms to an elliptical nonlinear algebraic system of the form in Eq. 101, the solution of which can be performed using the iterative Newton-Raphson scheme (Hamed 2012; Saanouni 2012a, b; Saanouni and Hamed 2013).

$$\begin{cases} \{\mathfrak{R}_u\}_{n+1} = \{F_{\text{int}}\}_{n+1} - \{F_{ext}\}_{n+1} = \{0\} & \text{(a)} \\ \left\{\mathfrak{R}_{\breve{D}}\right\}_{n+1} = \left\{\breve{F}_{\text{int}\,\breve{D}}\right\}_{n+1} - \left\{\breve{F}_{ext,\breve{D}}\right\}_{n+1} = \{0\} & \text{(b)} \\ \left\{\mathfrak{R}_{\breve{r}}\right\}_{n+1} = \left\{\breve{F}_{\text{int}\,\breve{r}}\right\}_{n+1} - \left\{\breve{F}_{ext,\breve{r}}\right\}_{n+1} = \{0\} & \text{(c)} \\ \left\{\mathfrak{R}_{\breve{\underline{\alpha}}}\right\}_{n+1} = \left\{\breve{F}_{int,\breve{\underline{\alpha}}}\right\}_{n+1} - \left\{\breve{F}_{ext,\breve{\underline{\alpha}}}\right\}_{n+1} = \{0\} & \text{(d)} \end{cases} \tag{101}$$

When linearized using the Newton–Raphson scheme, this system can arranged under the following formal matrix form:

$$\begin{Bmatrix} \{\mathfrak{R}_u\} \\ \{\mathfrak{R}_{\breve{D}}\} \\ \{\mathfrak{R}_{\breve{r}}\} \\ \{\mathfrak{R}_{\breve{\alpha}}\} \end{Bmatrix}^i_{n+1} + \begin{bmatrix} \frac{\partial\{\mathfrak{R}_u\}}{\partial\{U\}} & \frac{\partial\{\mathfrak{R}_u\}}{\partial\breve{D}} & \frac{\partial\{\mathfrak{R}_u\}}{\partial\breve{r}} & \frac{\partial\{\mathfrak{R}_u\}}{\partial\{\breve{\alpha}\}} \\ \frac{\partial\{\mathfrak{R}_{\breve{D}}\}}{\partial\{U\}} & \frac{\partial\{\mathfrak{R}_{\breve{D}}\}}{\partial\breve{D}} & \frac{\partial\{\mathfrak{R}_{\breve{D}}\}}{\partial\breve{r}} & \frac{\partial\{\mathfrak{R}_{\breve{D}}\}}{\partial\{\breve{\alpha}\}} \\ \frac{\partial\{\mathfrak{R}_{\breve{r}}\}}{\partial\{U\}} & \frac{\partial\{\mathfrak{R}_{\breve{r}}\}}{\partial\breve{D}} & \frac{\partial\{\mathfrak{R}_{\breve{r}}\}}{\partial\breve{r}} & \frac{\partial\{\mathfrak{R}_{\breve{r}}\}}{\partial\{\breve{\alpha}\}} \\ \frac{\partial\{\mathfrak{R}_{\breve{\alpha}}\}}{\partial\{U\}} & \frac{\partial\{\mathfrak{R}_{\breve{\alpha}}\}}{\partial\breve{D}} & \frac{\partial\{\mathfrak{R}_{\breve{\alpha}}\}}{\partial\breve{r}} & \frac{\partial\{\mathfrak{R}_{\breve{\alpha}}\}}{\partial\{\breve{\alpha}\}} \end{bmatrix}^i_{n+1} \begin{Bmatrix} \delta U \\ \delta\breve{D} \\ \delta\breve{r} \\ \delta\breve{\underline{\alpha}} \end{Bmatrix}_{n+1} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix} \tag{102}$$

where (i) is the number of the current iteration and the operator δ defines the corrections of nodal unknowns during two successive iterations according to

$$\begin{cases} (\delta\{U\})_{n+1} = \{U\}_{n+1}^{i+1} - \{U\}_{n+1}^{i} \\ \left(\delta\breve{D}\right)_{n+1} = \left(\breve{D}\right)_{n+1}^{i+1} - \left(\breve{D}\right)_{n+1}^{i} \\ \left(\delta\breve{r}\right)_{n+1} = \left(\breve{r}\right)_{n+1}^{i+1} - \left(\delta\breve{r}\right)_{n+1}^{i} \\ (\delta\{U\})_{n+1} = \left\{\breve{\alpha}\right\}_{n+1}^{i+1} - \left\{\breve{\alpha}\right\}_{n+1}^{i} \end{cases} \tag{103}$$

Without going into details, the iterative solution procedure is the following: starting from an initial solution which is the one at t_n for the first iteration, the system in Eq. 102 is solved with respect to the unknowns $(\delta\{U\})_{n+1}$, $\left(\delta\breve{D}\right)_{n+1}$, $\left(\delta\breve{r}\right)_{n+1}$, and $(\delta\{U\})_{n+1}$. These allow the calculation of the new solution at the next iteration thanks to Eq. 103. This new solution is checked for the system in Eq. 101 using appropriate convergence criterion until the final convergence. The convergence rate depends on the mathematical properties of the tangent matrix in the system in Eq. 102 (Dautray and Lions 1984). Note that each term of this tangent matrix gives rise to several terms coming from the derivatives of the integrals found in Eqs. 92, 93, 94, 95, 96, 97, 98, and 99 in which the state variables are discretized by the time discretization scheme (see later). For the sake of shortness, the overall terms of this tangent matrix are not computed here and can be found in Hamed (2012), Saanouni (2012a, b), and Saanouni and Hamed (2013). It is simply indicated that the computation of the overall terms of the tangent stiffness matrix is not so trivial under large plastic strains (geometrical nonlinearities), plastic flow with nonlinear hardening, damage-induced softening, and contact with friction commonly observed in metal forming of metallic structures (see the last term of the RHS of Eq. 96).

Outlines of the DE Scheme

When the inertia effects are accounted for (dynamic problems), the nonlinear hyperbolic system in Eq. 100 can be rewritten on the typical time increment $[t_n, t_{n+1} = t_n + \Delta t]$ with the help of Eq. 101, under the following form:

$$\begin{cases} [M]_{n+1}\left\{\ddot{U}\right\}_{n+1} + \{\mathfrak{R}_u\}_{n+1} = \{0\} & \text{(a)} \\ \left[\breve{M}_D\right]_{n+1}\left\{\ddot{\breve{D}}\right\}_{n+1} + \left\{\mathfrak{R}_{\breve{D}}\right\}_{n+1} = \{0\} & \text{(b)} \\ \left[\breve{M}_r\right]_{n+1}\left\{\ddot{\breve{r}}\right\}_{n+1} + \{\mathfrak{R}_{\breve{r}}\}_{n+1} = \{0\} & \text{(c)} \\ \left[\breve{M}_D\right]_{n+1}\left\{\ddot{\breve{\alpha}}\right\}_{n+1} + \left\{\mathfrak{R}_{\underline{\breve{\alpha}}}\right\}_{n+1} = \{0\} & \text{(d)} \end{cases} \tag{104}$$

The most used DE resolution schemes of IBVP with several weak forms as Eq. 104 consist in solving sequentially and successively, over the same typical time increment $[t_n \quad t_{n+1} = t_n + \Delta t]$, the four equations of system Eq. 104. Explicit schemes express the solution at t_{n+1} exclusively in terms of the quantities entirely

known at the preceding instants t_n, t_{n-1}, etc. This has the benefit of great simplicity of numerical implementation and avoids the computation of the tangent stiffness matrix. However, the size of the time step Δt is controlled by appropriate stability and precision criteria and can be severely limited giving rise to a large CPU time. If, in addition, the lumped mass matrices can be easily obtained, then an explicit dynamic global resolution scheme is preferred (see Zienkiewicz and Taylor (1967); Hughes (1987); Bathe (1996); Belytschko et al. (2000); Wriggers (2008) for more details on the sequential solving strategy).

Briefly speaking, first the mechanical problem in Eq. 104 is solved to determine $\{U\}_{n+1}$ using the values of the other dofs at t_n, i.e., $\left\{\breve{D}\right\}_n$, $\left\{\breve{r}\right\}_n$, and $\left\{\breve{\alpha}\right\}_n$. The problem in Eq. 104 is then solved with respect to $\left\{\breve{D}\right\}_{n+1}$ using $\{U\}_{n+1}$, $\left\{\breve{r}\right\}_n$, and $\left\{\breve{\alpha}\right\}_n$. Next the problem in Eq. 104 is solved to determine $\left\{\breve{r}\right\}_{n+1}$ using $\{U\}_{n+1}$, $\left\{\breve{D}\right\}_{n+1}$, and $\left\{\breve{\alpha}\right\}_n$. Finally, the problem in Eq. 104 is solved on the same time increment to determine $\left\{\breve{\alpha}\right\}_{n+1}$ knowing $\{U\}_{n+1}$, $\left\{\breve{D}\right\}_{n+1}$, and $\left\{\breve{r}\right\}_{n+1}$. To illustrate this solving scheme, only the main steps of solving Eq. 104 are outlined hereafter.

Since the solution $\mathbb{S}_n = \left\{U_n, \breve{D}_n, \breve{r}_n, \breve{\alpha}_n\right\}$ is fully known at t_n, the equation in Eq. 104 written at t_n takes the form ($[M_L]$ stands for the lumped mass matrix):

$$[M_L]_n\left\{\ddot{U}\right\}_n + \{\mathfrak{R}_u\}_n = [M_L]_n\left\{\ddot{U}\right\}_n + \{F_{\text{int}}\}_n - \{F_{ext}\}_n = \{0\} \tag{105}$$

This equation is solved in order to compute the acceleration vector at t_n:

$$\left\{\ddot{U}\right\}_n = [M_L]_n^{-1}\left\{\{F_{ext}\}_n - \{F_{int}\}_n\right\} \tag{106}$$

With the help of truncated Taylor expansion, the velocity and the middle of the time increment defined by $t_{n+1/2} = (t_n + t_{n+1})/2$ are obtained (Δt is the current time step, i.e., Δt_{n+1}):

$$\left\{\dot{U}\right\}_{n+\frac{1}{2}} \simeq \left\{\dot{U}\right\}_{n-\frac{1}{2}} + \frac{\Delta t + \Delta t_n}{2}\left\{\ddot{U}\right\}_n + \ldots \tag{107}$$

Finally, the displacement vector at the end of the time increment (t_{n+1}) is also obtained using the similar truncated Taylor expansion:

$$\{U\}_{n+1} \simeq \{U\}_n + \Delta t\left\{\dot{U}\right\}_{n+\frac{1}{2}} + \ldots \tag{108}$$

The stability condition can be approximated by (see ABAQUS User's manual):

$$\Delta t \leq \min\left(h^e\sqrt{\frac{\rho}{\lambda_e + 2\mu_e}}\right) \tag{109}$$

in which h^e is the size of the smallest finite element and $\sqrt{\rho/(\lambda_e + 2\mu_e)}$ is the inverse of the speed of a stress wave traveling through the elastic solid.

The same type of DE solving scheme is applied to solve successively equations Eq. 104b–d associated time steps are calculated (see Hamed 2012; Saanouni 2012a, b; Saanouni and Hamed 2013). The smallest value of the four computed time steps is selected to increase the time and to go to the next step.

Local Integration Scheme: Computation of the State Variables at Each Gauss Point

Regardless of the global resolution scheme (static implicit or dynamic explicit), it requires the calculation of internal and external force vectors given by the integrals in Eqs. 92, 93, 94, 95, 96, 97, 98, and 99. The numerical computation of these integrals using the quadrature Gauss method and the values of some state variables at each Gauss point of each element are required. This is the case for the stress tensor $\underline{\sigma}_{n+1}$ in Eq. 92; the contact force vector $\vec{F}^C_{n+1}$ in Eq. 96; the local damage D_{n+1} in Eqs. 97, 98, and 99; the local isotropic hardening strain r_{n+1} in Eq. 98; and the kinematic hardening strain tensor $\underline{\alpha}_{n+1}$ in Eq. 99. The computation of these state variables passes through the local integration of the fully coupled constitutive equations defined by Eqs. 20, 21, 22, 23, 24, 25, 26, 27, 29, and 33 on the typical time interval $[\,t_n \quad t_{n+1} = t_n + \Delta t\,]$.

This will be performed using the well-known return-mapping algorithm based on the elastic prediction and plastic correction method for general nonassociative elastoplasticity models (see, e.g., Simo and Hughes 1997; Belytschko et al. 2000; Saanouni 2012a, b). In the presence of the nonlinear isotropic and kinematic hardenings, it has been shown in Saanouni and Chaboche (2003), Badreddine et al. (2010), and Saanouni (2012a, b) that combining the asymptotic scheme (Walker and Freed 1991) with the return-mapping algorithm leads to an efficient and robust unconditionally stable integration scheme in presence of the ductile damage.

First of all, let us mention that the first-order ordinary differential equations (29), (30), (31), and (32) can be classified into two classes: Eqs. 29 and 32 have the form

$$\forall t \in [t_n\ ,\ t_{n+1}] \quad \begin{cases} \dot{y} = \varphi(y,t) \\ y(t) = y_n \quad \text{for} \quad t = t_n \end{cases} \tag{110}$$

while Eqs. 30 and 31 have the form

$$\forall t \in [t_n,\ t_{n+1}] \quad \begin{cases} \dot{y} = \varphi(y,t)[\phi(y,t) - y] \\ y(t) = y_n \quad \text{for} \quad t = t_n \end{cases} \tag{111}$$

The solution of Eq. 110 obtained by the classical θ-method is

$$y_{n+\theta} = y_n + \Delta t\left(\theta \dot{y}_{n+1} + (1-\theta)\dot{y}_n\right) \quad \text{for} \quad 0 \le \theta \le 1 \tag{112}$$

while the solution of Eq. 111 has the following form:

$$y_{n+\theta} = y_n \exp\left(-\theta\varphi\left(y_{n+\theta}\right)\Delta t\right) + \left[1 - \exp\left(-\theta\varphi\left(y_{n+\theta}\right)\Delta t\right)\right]\phi\left(y_{n+\theta}\right) \quad \text{for} \quad 0 \le \theta \le 1 \tag{113}$$

Applying the solutions (112) and (113) with fully implicit case ($\theta = 1$) allows rewriting the main micromorphic fully coupled constitutive equations, at the end of the time step ($t_{n+1} = t_n + \Delta t$), under the following discretized form (with $\underline{Z}_{n+1} = \underline{\sigma}^{dev}_{n+1} - \underline{X}_{n+1}$):

$$\underline{\varepsilon}^p_{n+1} = \underline{\varepsilon}^p_n + \Delta\lambda^p \frac{\widetilde{\underline{n}}_{n+1}}{\sqrt{(1-D_{n+1})}} \text{ where } \widetilde{\underline{n}}_{n+1} = \frac{3}{2}\frac{\left(\underline{\sigma}^{dev}_{n+1} - \underline{X}_{n+1}\right)}{\|\underline{\sigma} - \underline{X}\|_{n+1}} = \frac{3}{2}\frac{\underline{Z}_{n+1}}{\left\|\underline{Z}_{n+1}\right\|_{n+1}} \tag{114}$$

$$\begin{aligned} r_{n+1} = {} & r_n \exp\left(-b\Delta\lambda^p\left(\frac{Q+\breve{Q}}{Q}\right)\right) \\ & + \left(1 - \exp\left(-b\Delta\lambda^p\left(\frac{Q+\breve{Q}}{Q}\right)\right)\right)\left(\frac{1}{b\sqrt{(1-D_{n+1})}}\left(\frac{Q}{Q+\breve{Q}}\right)\right. \\ & \left. + \frac{\breve{r}_n}{\sqrt{(1-D_{n+1})}}\left(\frac{\breve{Q}}{Q+\breve{Q}}\right)\right) \end{aligned} \tag{115}$$

$$\begin{aligned} \underline{\alpha}_{n+1} = {} & \underline{\alpha}_n \exp\left(-\Delta\lambda^p a\left(\frac{C+\breve{C}}{C}\right)\right) \\ & + \left(1 - \exp\left(-\Delta\lambda^p a\left(\frac{C+\breve{C}}{C}\right)\right)\right)\left(\frac{\widetilde{\underline{n}}_{n+1}}{a\sqrt{(1-D_{n+1})}}\left(\frac{C}{C+\breve{C}}\right)\right. \\ & \left. + \frac{\breve{\underline{\alpha}}_n}{\sqrt{(1-D_{n+1})}}\left(\frac{\breve{C}}{C+\breve{C}}\right)\right) \end{aligned} \tag{116}$$

$$D_{n+1} = D_n + \frac{\Delta\lambda^p}{(1-D_{n+1})^\beta}\left(\frac{\langle Y_{n+1} - Y_0\rangle}{S}\right)^s = D_n + \Delta\lambda^p Y^*_D n + 1 \tag{117}$$

where only the von Mises stress norm $\|\underline{\sigma} - \underline{X}\|$ is used to be brief (no problem to use any other quadratic or non-quadratic stress norm). To these differential equations should be added the yield function and the overall force-like variables written at t_{n+1}:

$$f_{n+1} = \frac{\left\|\underline{\sigma}_{n+1} - \underline{X}_{n+1}\right\|}{\sqrt{1-D_{n+1}}} - \frac{R_{n+1}}{\sqrt{1-D_{n+1}}} - \sigma_y \tag{118}$$

$$\underline{\sigma}_{n+1} = (1-D_{n+1})\underline{\underline{\Lambda}} : \underline{\varepsilon}^e_{n+1} = (1-D_{n+1})\underline{\underline{\Lambda}} : \left(\underline{\varepsilon}^e_n + \Delta\underline{\varepsilon} - \Delta\lambda^p \frac{\widetilde{\underline{n}}_{n+1}}{\sqrt{(1-D_{n+1})}}\right) \tag{119}$$

$$\underline{X}_{n+1} = \frac{2}{3}\sqrt{(1-D_{n+1})}\left(\sqrt{(1-D_{n+1})}\left(C+\breve{C}\right)\underline{\alpha}_{n+1} - \breve{C}\breve{\underline{\alpha}}_n\right) \tag{120}$$

$$R_{n+1} = \sqrt{(1-D_{n+1})}\left(\sqrt{(1-D_{n+1})}\left(Q+\breve{Q}\right)r_{n+1} - \breve{Q}\breve{r}_n\right) \tag{121}$$

$$\begin{aligned} Y_{n+1} = \frac{1}{2}\underline{\varepsilon}^e{}_{n+1} &: \underline{\underline{\Lambda}} : \underline{\varepsilon}^e{}_{n+1} + \frac{1}{2}Qr^2_{n+1} + \frac{1}{3}C\underline{\alpha}_{n+1} \\ &: \underline{\alpha}_{n+1} + \frac{1}{2}\breve{Q}\left(\sqrt{(1-D_{n+1})}r_{n+1} - \breve{r}_n\right)\frac{r_{n+1}}{\sqrt{1-D_{n+1}}} \\ &+ \frac{1}{3}\breve{C}\left(\sqrt{(1-D_{n+1})}\underline{\alpha}_{n+1} - \breve{\underline{\alpha}}_n\right) : \frac{\underline{\alpha}_{n+1}}{\sqrt{1-D_{n+1}}} - \breve{H}\left(D_{n+1} - \breve{D}_n\right) \end{aligned} \tag{122}$$

In Eqs. 120, 121, and 122, $\underline{\alpha}_{n+1}$ is given by Eq. 116, r_{n+1} is given by Eq. 115, and D_{n+1} is given by Eq. 117, while the dofs $(\breve{D}_n, \breve{r}_n, \breve{\underline{\alpha}}_n)$ are provided by the convergent solution of the previous loading step.

Given an applied total strain increment $\Delta\underline{\varepsilon}$ computed from the increment of the deformation gradient imposed over the time increment $[t_n,\ t_{n+1}]$, so that the total strain at t_{n+1} is known and given by $\underline{\varepsilon}_{n+1} = \underline{\varepsilon}_n + \Delta\underline{\varepsilon}$, and knowing all the state variables $(\underline{\sigma}_n, \underline{\alpha}_n, \underline{\varepsilon}^p_n, \underline{\alpha}_n, r_n, D_n)$ as well as the nodal variables or dofs $(\breve{D}_n, \breve{r}_n, \breve{\underline{\alpha}}_n)$ at the beginning of the time interval t_n, their values $(\underline{\sigma}_{n+1}, \underline{\alpha}_{n+1}, \underline{\varepsilon}^p_{n+1}, \underline{\alpha}_{n+1}, r_{n+1}, D_{n+1})$ at t_{n+1} have to be computed so that the yield condition, i.e., $f_{n+1}\left(\underline{\sigma}_{n+1}, \underline{X}_{n+1}, R_{n+1}, D_{n+1}\right) = 0$ (see Eq. 115), is fulfilled.

The well-known elastic prediction – plastic correction algorithm – is used to solve the problem using a reduced number of equations as is summarized hereafter (see Saanouni (2012a, b) for more details).

Elastic prediction: If $\Delta\underline{\varepsilon}$ is assumed purely elastic without producing any dissipative phenomenon (i.e., $\Delta\lambda^p = 0$), then from Eq. 116 a trial stress tensor is deduced for the isotropic elasticity ($\underline{\varepsilon}^{trial}_{n+1} = \underline{\varepsilon}^e_n + \Delta\underline{\varepsilon}$ is the fully known assumed elastic strain):

$$\underline{\sigma}^{trial}_{n+1} = (1-D_n)\underline{\underline{\Lambda}} : \underline{\varepsilon}^{trial}_{n+1} = (1-D_n)\left(\lambda_e trace\left(\underline{\varepsilon}^{trial}_{n+1}\right)\underline{1} + 2\mu_e\underline{\varepsilon}^{trial}_{n+1}\right) \tag{123}$$

The trial yield function is then computed from Eq. 116 and writes:

$$f_{n+1}^{trial}\left(\underline{\sigma}_{n+1}^{trial},\underline{X}_n,R_n;D_n\right)=\frac{\left\|\underline{\sigma}_{n+1}^{trial}-\underline{X}_n\right\|-R_n}{\sqrt{(1-D_n)}}-\sigma_y \tag{124}$$

If $f_{n+1}^{trial}\left(\underline{\sigma}_{n+1}^{trial},\underline{X}_n,R_n;D_n\right)<0$, then the elastic train assumption holds and the solution at the end of this step is given by $\underline{\sigma}_{n+1}=\underline{\sigma}_{n+1}^{trial},\underline{X}_{n+1}=\underline{X}_n,\underline{\varepsilon}_{n+1}^{p}=\underline{\varepsilon}_n^{p},R_{n+1}=R_n,Y_{n+1}=Y_n$ and $D_{n+1}=D_n$, and the next loading step is performed. This happens typically when elastic unloading takes place.

Plastic correction: If $f_{n+1}^{trial}\left(\underline{\sigma}_{n+1}^{trial},\underline{X}_n,R_n;D_n\right)>0$, then the step under concern is plastic and the variables $\underline{\sigma}_{n+1}^{trial}\left(\underline{\varepsilon}_{n+1}^{trial}\ or\ \underline{\varepsilon}_n^{p}\right),\underline{X}_n(\underline{\alpha}_n),R_n(r_n),D_n(Y_n)$ should be iteratively corrected in order to get their final values $\underline{\sigma}_{n+1}\left(\underline{\varepsilon}_{n+1}^{p}\right),\underline{X}_{n+1}\left(\underline{\alpha}_{n+1}\right),R_{n+1}(r_{n+1}),D_{n+1}(Y_{n+1})$ fulfilling the plastic admissibility condition at t_{n+1} (i.e.,$f_{n+1}^{trial}\left(\underline{\sigma}_{n+1},\underline{X}_{n+1},R_{n+1};D_{n+1}\right)=0$). To do that the constitutive equations will be reduced to only two nonlinear equations with two independent unknowns $\Delta\lambda^p$ and D_{n+1}.

By using Eq. 123 the stress tensor, Eq. 119 can be expressed function of the trial stress:

$$\underline{\sigma}_{n+1}=(1-D_{n+1})\left(\frac{\underline{\sigma}_{n+1}^{trial}}{1-D_n}-\frac{2\mu^e\Delta\lambda^p}{\sqrt{(1-D_{n+1})}}\widetilde{\underline{n}}_{n+1}\right) \tag{125}$$

With the help of Eqs. 114 to 122 and 125 and after some simple algebraic transformations, the problem is reduced to the two following nonlinear equations (see Saanouni 2012a, b; Hamed 2012; Saanouni and Hamed 2013):

$$\begin{cases}\bar{f}_{n+1}=\left\|\underline{Z}_{n+1}^{*}\right\|-\dfrac{1}{\sqrt{1-D_{n+1}}}\left(3\mu\Delta\lambda^p+\left(1-\exp\left(-\Delta\lambda^p a\left(\dfrac{C+\breve{C}}{C}\right)\right)\right)\dfrac{C}{a}\right)+\dfrac{\left(\breve{Q}\breve{r}_n-\sigma_y\right)}{\sqrt{1-D_{n+1}}}-\\ \left(Q+\breve{Q}\right)\left[\begin{array}{l}r_n\exp\left(-b\Delta\lambda^p\left(\dfrac{Q+\breve{Q}}{Q}\right)\right)+\left(1-\exp\left(-b\Delta\lambda^p\left(\dfrac{Q+\breve{Q}}{Q}\right)\right)\right)\times\\ \left(\dfrac{1}{b\sqrt{(1-D_{n+1})}}\left(\dfrac{Q}{Q+\breve{Q}}\right)+\dfrac{\breve{r}_n}{\sqrt{(1-D_{n+1})}}\left(\dfrac{\breve{Q}}{Q+\breve{Q}}\right)\right)\end{array}\right]=0\\ and\\ g_{n+1}=D_{n+1}-D_n-\Delta\lambda^p Y_{n+1}^{*}=D_{n+1}-D_n-\dfrac{\Delta\lambda^p}{(1-D_{n+1})^{\beta}}\left\langle\dfrac{Y_{n+1}(D_{n+1},\Delta\lambda^p)-Y_0}{S}\right\rangle^{s}=0\end{cases} \tag{126}$$

These two highly nonlinear equations are linearized and iteratively solved thanks to the Newton-Raphson scheme according to

$$\left\{ \begin{matrix} \bar{f}_{n+1} \\ g_{n+1} \end{matrix} \right\}^{s} + \left[\begin{matrix} \frac{\partial \bar{f}_{n+1}}{\partial \Delta\lambda^{p}} & \frac{\partial \bar{f}_{n+1}}{\partial D_{n+1}} \\ \frac{\partial g_{n+1}}{\partial \Delta\lambda^{p}} & \frac{\partial g_{n+1}}{\partial D_{n+1}} \end{matrix} \right]^{s} \left\{ \begin{matrix} \delta\Delta\lambda^{p} \\ \delta D_{n+1} \end{matrix} \right\} + \ldots = 0 \tag{127}$$

where (s) is the iteration number and $\delta\Delta\lambda^{p}$, δD_{n+1} stand for the correction of the two unknowns between two successive iterations according to

$$\left\{ \begin{matrix} \delta\Delta\lambda^{p} = (\Delta\lambda^{p})^{s+1} - (\Delta\lambda^{p})^{s} \\ \delta D = (D_{n+1})^{s+1} - (D_{n+1})^{s} \end{matrix} \right. \text{from which} \left\{ \begin{matrix} (\Delta\lambda^{p})^{s+1} = (\Delta\lambda^{p})^{s} + \delta\Delta\lambda^{p} \\ (D_{n+1})^{s+1} = (D_{n+1})^{s} + \delta D \end{matrix} \right. \tag{128}$$

At each iteration, first Eq. 127 is solved with respect to the two unknowns $\delta\Delta\lambda^{p}$, δD, and then using Eq. 128 the solution for the iteration $(s+1)$ is deduced and checked for Eq. 126 using the appropriate convergence criterion. When the convergence is reached, the final values of $\delta\Delta\lambda^{p}$, δD stand for the final solution for the current step and are used to determine all the other state variables using the discretized equation given above. This iterative computation is performed for each Gauss point of each finite element during each loading increment (see Saanouni (2012a, b) for more details).

Finally let us mention that the computation of the rotation tensor at the end of each time increment $\underline{Q}_{n+1}$ required to rotate the overall tensorial variables (see section "Extension to Large Irreversible Strains Framework") is performed via the integration of Eq. 43 using a fully implicit backward Euler scheme (Hughes and Winget 1980) together with the assumption of constant velocity gradient over the time increment (linear kinematics), in order to ensure the incremental objectivity when the corotational frame is used (see Badreddine et al. 2010 for the Green–Naghdi total and plastic frames). For this corotational frame, one can easily obtain

$$\underline{Q}_{n+1} = \left[\underline{1} - \frac{\Delta t}{2} \underline{\Omega}_{n+1/2} \right]^{-1} \cdot \left[\underline{1} + \frac{\Delta t}{2} \underline{\Omega}_{n+1/2} \right] . \underline{Q}_{n} \tag{129}$$

where $\underline{\Omega}_{n+1/2}$ is the material rotation rate at the middle of the time increment. Once the rotation tensor is computed using Eq. 91, it is used to rotate all the tensorial variables according to the discussion in section "Extension to Large Irreversible Strains Framework." This allows rotating the actual deformed fictive configuration in order to have the same Lagrangian orientation as the unreformed initial configuration in which the local integration procedure discussed above is performed.

On the Numerical Aspects of the Contact with Friction

The contact with friction is an important issue in metal forming (Kobayashi et al. 1989; Rowe et al. 1991; Wagoner and Chenot 2001; Dixit and Dixit 2008; Saanouni 2012a, b). The theoretical formulation (see section "Modeling of the Contact with Friction") and numerical treatment of the contact with friction are

deeply discussed in the recent books and the references given there (Zhong 1993; Laursen 2002; Wriggers 2002; Saanouni 2012a, b). Here, only the numerical treatment of the contact/friction is shortly discussed in the framework of the penalty method using the master/slave surfaces method, widely used in virtual metal forming, and described in detail in the literature.

As shown by the last term of the RHS of the external forces vector Eq. 96, the contact forces vectors $\vec{F}^c = F_N^c \vec{n} + \vec{F}_\tau^c$ at all the nodes lying in the contact interfaces are required. In the general 3D cases, the search of the nodes concerned by the contact (i.e., the determination of the unknown contact interface) during the typical time interval $[t_n \quad t_{n+1} = t_n + \Delta t]$ is not an easy task (see Zhong (1993), Laursen (2002), Wriggers (2002), and Saanouni (2012a, b) for more details in this field). Once the elements (then nodes) concerned by the contact are determined, the contact force at each contact node is determined. This necessitates the numerical integration of the friction constitutive equations which are discussed in section "Modeling of the Contact with Friction." Due to the mathematical similarity between the plastic constitutive equations (inside the material volume) and the friction constitutive equations only valid for the material points lying in the contact interfaces, the predictor/corrector integration scheme can be applied.

If the Coulomb model (perfect friction without hardening) defined by the fiction yield condition in Eq. 66 with a constant friction parameter η and a constant yield limit force F_y is considered, the friction constitutive equations can be summarized as follows:

$$\begin{cases} f_f = \left\|\vec{F}_\tau^c\right\|_f - \eta\left(F_N^c\right) - F_y \le 0 \quad with \quad \left\|\vec{F}_\tau^c\right\|_f = \sqrt{\vec{F}_\tau^c.\vec{F}_\tau^c} & (a) \\ \dot{\vec{F}}_\tau^c = p_\tau\left(\dot{\vec{\vartheta}}_\tau - \dot{\lambda}_f \vec{n}_f\right) & (b) \\ \vec{n}_f = \dfrac{\vec{F}_\tau^c}{\left\|\vec{F}_\tau^c\right\|_f} & (c) \\ \dot{\lambda}_f \ge 0 \ , \dot{\lambda}_f f_f = 0 & (d) \end{cases} \tag{130}$$

If these equations are discretized using backward (fully implicit $\theta = 1$) Euler scheme, they can be written at t_{n+1} as

$$\begin{cases} f_{n+1}^f = \left\|\vec{F}_{\tau,n+1}^c\right\|_f - \eta\left(F_{N,n+1}^c\right) - F_y \le 0 \text{ with } \left\|\vec{F}_{\tau,n+1}^c\right\|_f = \sqrt{\vec{F}_{\tau,n+1}^c.\vec{F}_{\tau,n+1}^c} & (a) \\ \vec{F}_{\tau,n+1}^c = \dfrac{\vec{F}_{\tau,n}^c + p_\tau\left(\vec{\vartheta}_{n+1}^\tau - \vec{\vartheta}_n^\tau\right)}{1 + p_\tau \dfrac{\Delta\lambda_f}{\left\|\vec{F}_{\tau,n+1}^c\right\|_f}} \quad \text{with} \quad \Delta\lambda_f \ge 0 \ , \ \Delta\lambda_f f_{n+1}^f = 0 & (b) \end{cases} \tag{131}$$

where $\Delta\lambda_f$ is the increment of the friction Lagrange multiplier and p_τ being the regularization factor (see Saanouni 2012a, b). The "trial" state is obtained by supposing contact with adhesion and without sliding (i.e., $\Delta\lambda_f = 0$ in Eq. 130b), which allows us to write the "trial" friction criterion as

$$\begin{cases} f_{n+1}^{f,trial} = \left\| \vec{F}_{n+1}^{\tau,trial} \right\|_f - \eta F_{N,n+1}^c - F_y & \text{(a)} \\ \text{with} \\ \vec{F}_{n+1}^{\tau,trial} = \vec{F}_n^{\tau} + p_\tau \left(\vec{\vartheta}_{n+1}^{\tau} - \vec{\vartheta}_n^{\tau} \right) & \text{(b)} \end{cases} \tag{132}$$

Following the same approach as in isotropic plasticity (see section "Local Integration Scheme: Computation of the State Variables at Each Gauss Point"), there are no friction if $f_{n+1}^{f,trial} < 0$ confirming the trial state. While if $f_{n+1}^{f,trial} > 0$, the friction takes place and the friction correction should be performed by imposing that $f_{n+1}^{f} = 0$ at t_{n+1}.

If Eq. 131a is zero (friction condition), and with the help of Eq. 130c, an expression of the tangential force vector $\vec{F}_{\tau,n+1}^{c}$ can be easily deduced. Equating it with Eq. 131b allows to deduce the following equation which is linear with respect to the main unknown $\Delta\lambda_f$:

$$\left\| \vec{F}_{n+1}^{\tau,trial} \right\|_f - \eta \left(F_{N,n+1}^c \right) - F_y - p_\tau \Delta\lambda_f = f_{n+1}^{f,trial} - p_\tau \Delta\lambda_f = 0 \tag{133}$$

Giving the simple closed form solution,

$$\Delta\lambda_f = \frac{\left\| \vec{\overline{F}}_{n+1}^{\tau,essai} \right\|_f + \eta_{n+1} \left(F_{n+1}^N \right) + F_{n+1}^y}{p_\tau} = \frac{f_{n+1}^{f,essai}}{p_\tau} \tag{134}$$

Thus, the tangential friction force is calculated by replacing Eq. 134 in Eq. 131b to get

$$\vec{F}_{\tau,n+1}^{c} = \begin{cases} \vec{F}_{n+1}^{\tau,trial} = \vec{F}_{\tau,n}^{c} + p_\tau \left(\vec{\overline{\vartheta}}_{n+1}^{\tau} - \vec{\overline{\vartheta}}_n^{\tau} \right) & \text{if } f_{n+1}^{f,trial} \le 0 \text{ (adhesion)} \quad \text{(a)} \\ \left(\eta \left(F_{N,n+1}^c \right) + F_y \right) \dfrac{\vec{F}_{n+1}^{\tau,trial}}{\left\| \vec{F}_{n+1}^{\tau,trial} \right\|_f} & \text{if } f_{n+1}^{f,trial} > 0 \quad \text{(sliding)} \quad \text{(b)} \end{cases} \tag{135}$$

This simple case without "friction hardening" leads to an exact solution in $\Delta\lambda_f$, and there is thus no need to use an iterative procedure, as in plasticity with nonlinear hardening, to compute the friction multiplier (the case with friction hardening can be found in Saanouni (2012a, b)).

Implementation into ABAQUS/Explicit®

The micromorphic elastoplastic constitutive equations with nonlinear isotropic and kinematic hardening fully coupled with ductile damage developed in section "Proposed Fully Coupled Micromorphic Constitutive Equations" and discretized in section "Local Integration Scheme: Computation of the State Variables at Each Gauss Point" have been implemented into ABAQUS/Explicit® using the user's developed subroutine VUMAT. Also, some 2D (T3 and Q4) and 3D (T4) with additional dofs which are the micromorphic kinematic variables have been developed using the subroutine VUEL. All the practical aspects related to this implementation can be found in Hamed (2012), Saanouni (2012a, b), and Saanouni and Hamed (2013). Here, only some information concerning the specific treatment of the fully damaged Gauss points and fully damaged elements are presented in order to describe the macroscopic cracks initiation and their propagation inside the deformed part. Also the adaptive remeshing methodology used in conjunction with ABAQUS/Explicit® in order to perform adaptive analyses mainly in 2D cases is shortly discussed.

During the analysis, each Gauss point of each element is checked with respect to the critical value of the local damage $D_{cr} \geq 0.99$. When this condition is fulfilled, the overall stress-like variables are near zero in that point. Accordingly, all the stress-like variables are put to zero and $D = 1.0$ indicating that this Gauss point is fully damaged (microvoid) and consequently the point is excluded from the integration domain for the remaining time after the strain-like variables in that point are stored to keep their values at the ultimate fracture. Similarly, if all the Gauss points of an element are fully damaged, the element is excluded from the structure and a new mesh is performed with the new geometry. This works very well if the mesh inside the damaged zones is adequately fine or if a fully adaptive remeshing procedure is used for which the size of elements are very small inside the area where the damage is near D_{cr}. The detailed description of this fully adaptive remeshing methodology can be found in Labergère et al. (2011) and Saanouni (2012a, b).

Some Typical Examples of Virtual Metal Forming Processes

The advanced fully coupled constitutive equations presented above when implemented into a general purpose finite element code lead to a helpful adaptive numerical methodology for metal forming simulation called "virtual" metal forming. The interested reader can find a comprehensive presentation of this kind of numerical methodology in the recent book dedicated to damage mechanics in metal forming (Saanouni 2012a, b) where, in addition to the modeling aspects, many applications to sheet and bulk metal forming using various versions of the constitutive equations either in plasticity or viscoplasticity and low or high temperature can be found. In this chapter, some typical examples dealing with sheet and bulk metal forming are briefly presented and discussed without going into the

details of the calculations. The objective is to show how the use of constitutive equations with high predictive capabilities which can account for multiphysical phenomena greatly improves, both qualitatively and quantitatively, the numerical results of calculations in industrial simulation of various metal forming processes. Particularly, it will be shown that, using the same numerical methodology of virtual metal forming, it is possible, thanks to a series of numerical simulations, to optimize the technical parameters of a given process in order to:

- Minimize the occurrence of ductile damage in formed parts in order to obtain defects-free components (i.e., without localization zones, wrinkles, cracks, etc.) as in deep drawing of thin sheets, hydroforming tubes or sheets, or cold forging or extrusion, etc.
- Favor the ductile damage occurrence in formed parts by controlling its intensity and direction to reproduce various material cutting processes such as stamping of thin or thick parts by sheering or slitting and guillotining thin sheets.

To be used in virtual metal forming, these fully coupled constitutive equations have to be identified for each used material in order to determine the overall material parameters present in theses equations. This can be done by using appropriate inverse approach which consists in minimizing the difference between the reference experimental results and the numerically predicted results using a wide range of applied loading paths under a representative range of loading conditions (temperature, loading rate, etc.). A specific identification methodology is described in Chap. 4 of Saanouni (2012a, b). Note that the identification procedure is performed based on experimental results obtained for the material under concern loaded with certain loading paths until the final fracture. The best values of the material parameters should be used for the numerical simulation of any forming process of the same material. Note that when the micromorphic variables are zero, i.e., when the model is fully local, the identification should made using the specimens meshed with a given mesh size. The same mesh size should be used inside the fully damaged zone during the subsequent forming processes simulations.

Hereafter, some numerical results obtained for various sheet and bulk metal forming exclusively under room temperature without considering thermal coupling are given. For each example after a brief description of the forming process itself as well as of the model used, some typical results are presented together with experimental pictures if available. All the numerical solution are obtained using ABAQUS/Explicit® together with our VUMAT and VUEL subroutines and for some 2D examples using appropriate adaptive remeshing procedure described in Labergère et al. (2011) and Saanouni (2012a, b).

Now, let us give an idea on the micromorphic model and its ability to regularize the IBVP characterized by strong damage-induced softening. For that, the uniaxial tensile test for a hypothetical metallic material is loaded in tension until the final fracture due to ductile damage initiation, growth, and coalescence.

The specimen with 50 mm large and 150 mm length is fixed at the bottom side, and a displacement with constant velocity is imposed on the opposite (top) side.

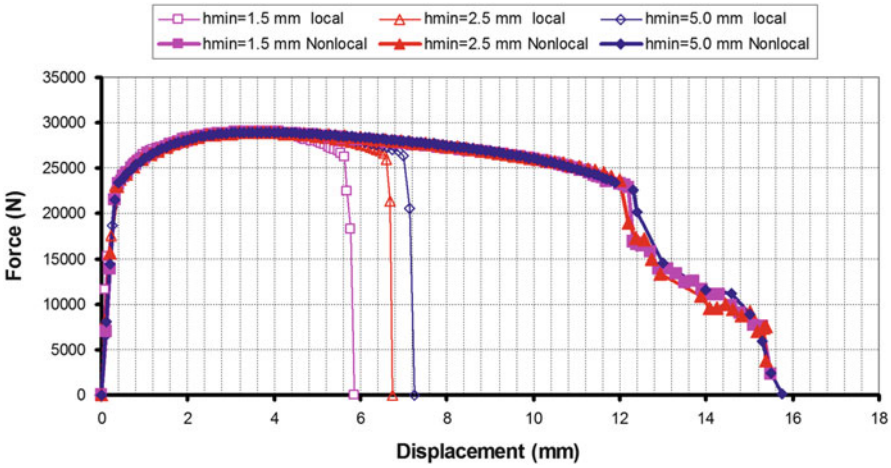


Fig. 5 Force–displacement curves for the three mesh sizes. Comparisons between local and nonlocal (micromorphic damage) models ($\ell_{\breve{d}} = 1.0$ mm)

The specimen is meshed using the triangular linear element (T3) using three different mesh sizes constant inside the central part of the specimen and defined by $h_{min} = 5.0$ mm (662 elements, 372 nodes), $h_{min} = 2.5$ mm (1,512 elements, 806 nodes), and $h_{min} = 1.5$ mm (3,242 elements, 1,705 nodes). Except for the coarse mesh which is quasi-homogeneous throughout the specimen, the other two meshes are refined only inside the central area (gage length) of the specimen in order to save the CPU time.

The version of the model used contains isotropic and kinematic hardening with only micromorphic damage defined by the following values of the material constants: $E = 210.0$ GPa, $\nu = 0.29$, $\sigma_y = 400.0$ MPa, $Q = 1,000.0$ MPa, $b = 50.0$, $C = 10,000.0$ MPa, $a = 100.0$, $\beta = 1.0$, $S = 1.2$, $s = 1.0$, $Y_0 = 0.0$, $\breve{H}^g = \breve{H} = 10^3$ giving the internal length *related to the micromorphic damage of* $\ell_{\breve{d}} = 1.0$ mm *and finally* $\zeta_{\breve{d}} = 1.0$ mm^2. Note that for the local model $\breve{H}^g = \breve{H} = \zeta_{\breve{d}} = 0$ hold.

Figure 5 compares the force–displacement curves predicted by the fully local model (no micromorphic phenomena called local model) and the model with micromorphic damage (nonlocal model). Clearly, the solution provided by the local model is highly mesh dependent in the softening stage, while the solution provided by the nonlocal model is much more mesh independent during the softening stage. The nonlocal model predicts a more large softening stage giving a displacement to fracture $u_f \approx 15.6$ mm for the three mesh sizes, while the displacement to fracture predicted with the fully local model is mesh dependent and is $u_f \approx 5.8$ mm for the fine mesh, $u_f \approx 6.7$ mm for the medium mesh, and

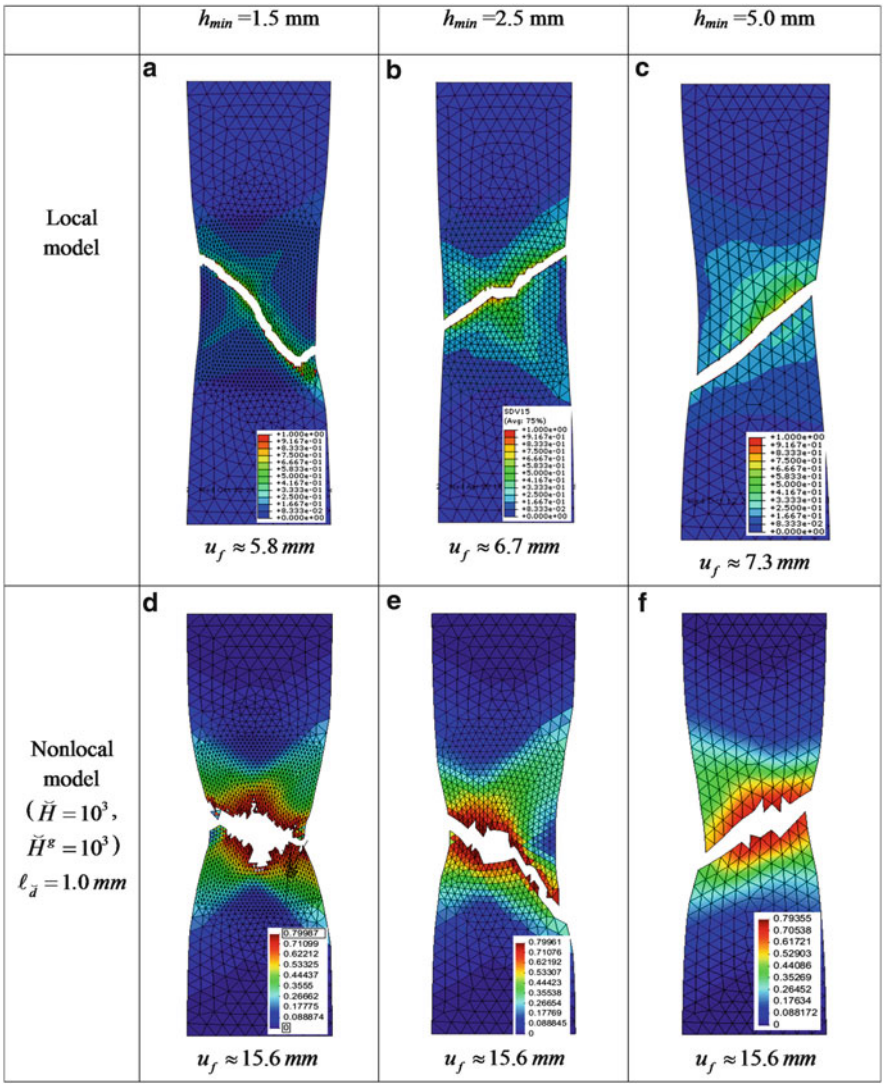


Fig. 6 Distribution of the local damage for the three mesh sizes for both the fully local model and the nonlocal model (micromorphic damage only)

$u_f \approx 7.3$ mm for the coarse mesh. Also the last stage of the softening curves predicted by the nonlocal model has somewhat spurious aspect (Fig. 6d–f) due to the fact that the width of the fully damaged zone contains a large number of elements with a heterogeneous fracture, while for the local model, this stage is clean since only one row of elements forms the fully damaged zone (Fig. 6a–c).

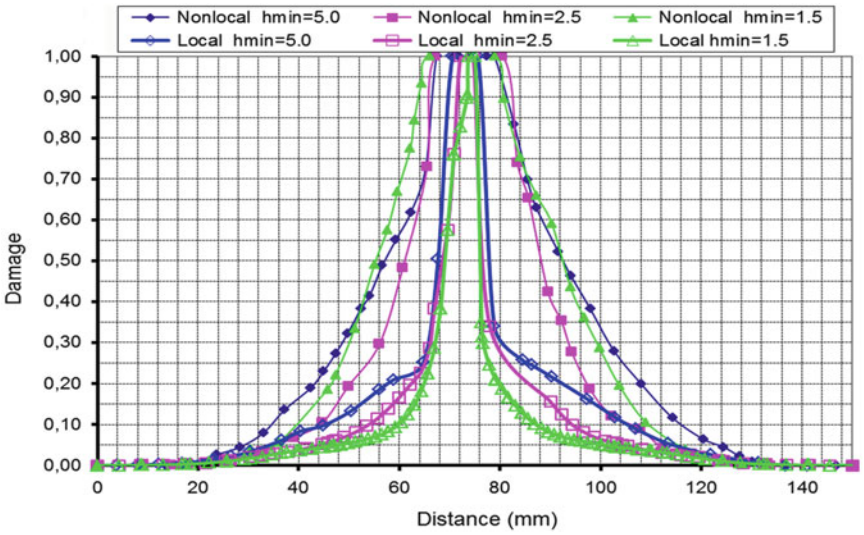


Fig. 7 Damage distribution along the central axis of the specimen for the three mesh sizes for both the fully local model and the nonlocal model (micromorphic damage only)

Also from this figure, it is noted that the orientation of the shear band as well as the macroscopic crack is dependent on the mesh size for the local model, while it seems to keep the same orientation for the nonlocal model when the mesh is refined enough, i.e., at convergence (see Fig. 6d, e).

Finally, Fig. 7 summarized the plot of the local damage distribution along the central axis of the specimen. Clearly, the damage is more homogenized with the nonlocal model which gives the crack thickness or width independent from the mesh size. However, for the local model, the damage is more localized and the crack width is clearly correlated to the mesh size.

Let us now discuss some examples concerning the damage prediction in sheet metal forming. All the calculations were performed in 3D using the fully local model with different version of the elastoplasticity fully coupled with isotropic and/or anisotropic damage. The blank is meshed using C3D8R solid elements, and the tools are meshed using rigid elements R3D4 both from the ABAQUS element library.

The first example concerns the hydrobulging of thin anisotropic sheet using an elliptical matrix schematized in Fig. 8 where a circular blank of 133 mm of diameter is maintained between the matrix and the blank-holder and an increasing pressure is injected between the blank-holder and the sheet surface. Under the effect of internal pressure, the sheet moves through the elliptical hole contained in the die.

Figure 9 gives the comparison between the experimentally observed crack (Fig. 9c) and the numerically predicted cracks using the isotropic damage

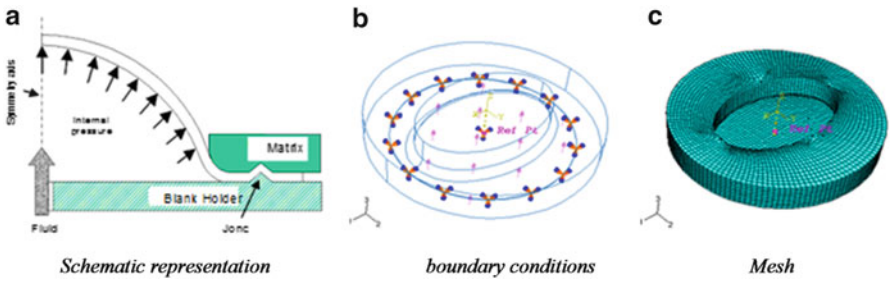


Fig. 8 Hydrobulging of a thin orthotropic sheet

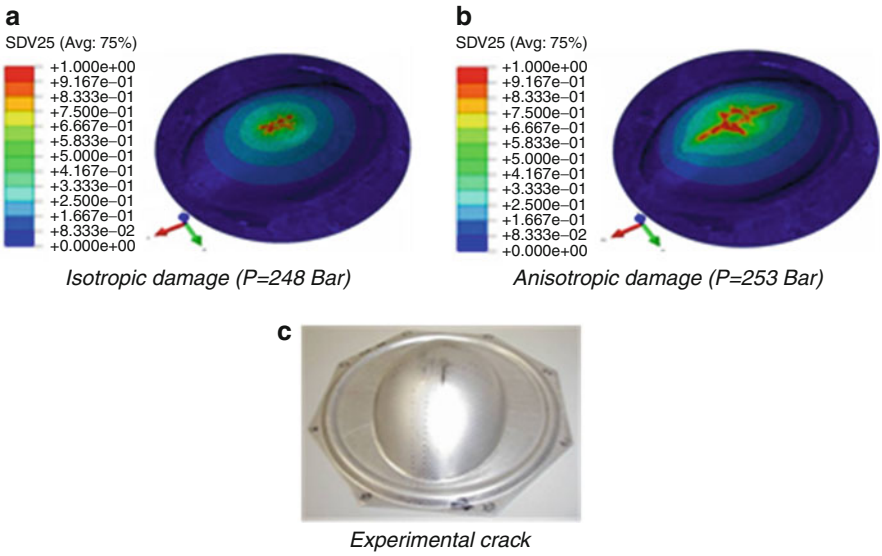


Fig. 9 Predicted versus experimentally observed cracks in hydrobulging of anisotropic sheet

(Fig. 9a) and the anisotropic damage (Fig. 9b) as can be found in Nguyen (2012). Clearly, the anisotropic model predicts a macroscopic crack oriented along the main axis of the ellipsoid close to the experimental crack observed for at a displacement of the dome $u = 34$ mm for $P = 225$ Bar.

The second sheet metal example concerns the trimming process using a cut rolling shear guillotine schematized in Fig. 10 and fully described in detail in Ghozzi et al. (2012).

Figure 11 illustrates the distribution of the von Mises equivalent stress at two different instants during the cutting operation.

The final example concerns the process dealing the assembly of a tube (25.1 mm of external diameter and 1.5 mm thickness) with a plate containing an initial hole of

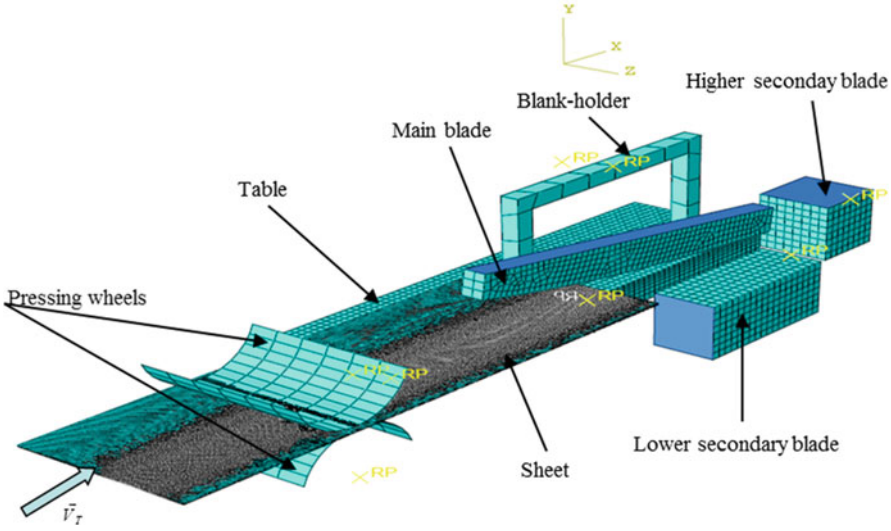


Fig. 10 Schematization of the trimming process for a thin sheet (Saanouni 2012a, b)

25 mm diameter called tube-end forming. The distribution of the accumulated plastic strain at the end of the assembly process is shown in Fig. 12 where the maximum of the plastic strain (around 140 %) is, as expected, located in the core area of the deformed tube.

Also Fig. 13 presents an experimental view zooming on the core area (Fig. 13a) and a zoom of the same area showing the distribution of the ductile damage (Fig. 13b) where the maximum of damage does not exceed 35 % located inside the plastic zone.

Concerning the application to the bulk metal forming processes, two typical examples are given here to illustrate the virtual metal forming methodology. All the calculations were performed using the isotropic thermo-elasto-viscoplasticity model fully coupled with isotropic damage for high-temperature bulk metal forming.

The first example concerns the hot forging of a billet presented in Fig. 14 where it is clearly seen that the damage localizes rapidly at the head of the billet under the biggest die head.

The last illustrative example concerns the metal machining (orthogonal cutting) by chip formation and segmentation. This simple orthogonal machining problem, schematized in Fig. 15, was modeled in 2D with both the part and the tool meshed using CPE4RT element taken from the ABAQUS library for thermomechanical problems. The anisothermal thermo-elasto-viscoplastic constitutive equations were used together with a specific adaptive mesh methodology described in Labergère et al. (2011) and Saanouni (2012a, b).

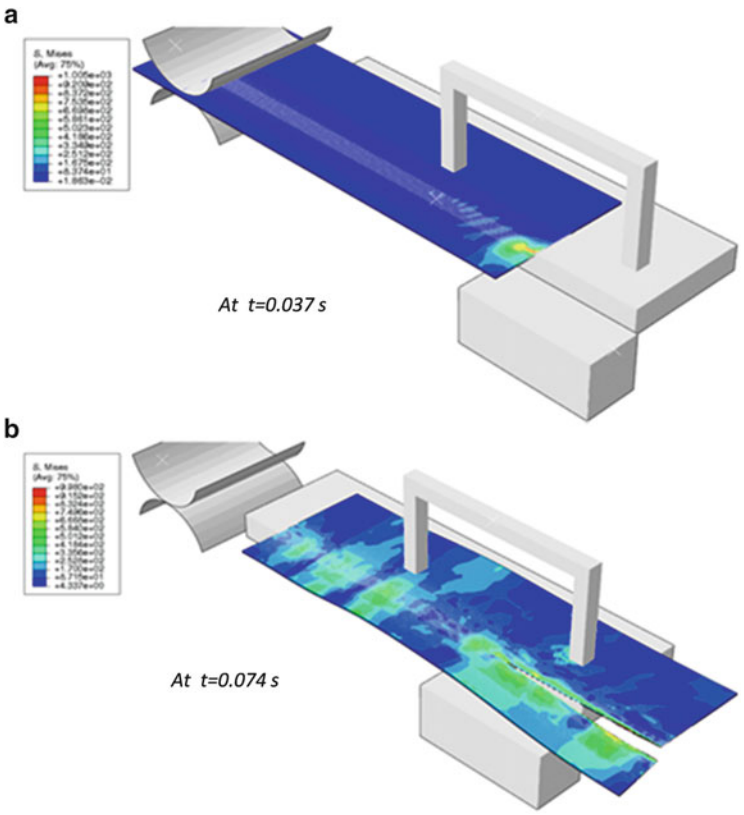


Fig. 11 Equivalent stress distribution at two different instants of the trimming process (Saanouni 2012a, b)

A typical result of this simulation is shown in Fig. 16 where the temperature distribution is shown in Fig. 16a and the adaptive remeshing is shown in Fig. 1b. It is worth noting that starting from the room temperature (25 °C) at the initial configuration, the temperature increases during the chip formation and exceeds 1,000 °C with a maximum located at the primary and secondary adiabatic shear bands. Inside the primary intensive adiabatic shear band, the accumulated plastic strain reaches 683 %, while the strain rate exceeds 1.2 10^{+06} s^{-1} and the temperature is about 1,250 °C (see Issa et al. (2012) for more details).

In conclusion, the overall presented examples show clearly the excellent predictive capabilities of the virtual metal forming processes methodology as long as it is based on advanced fully coupled and multiphysical constitutive equations.

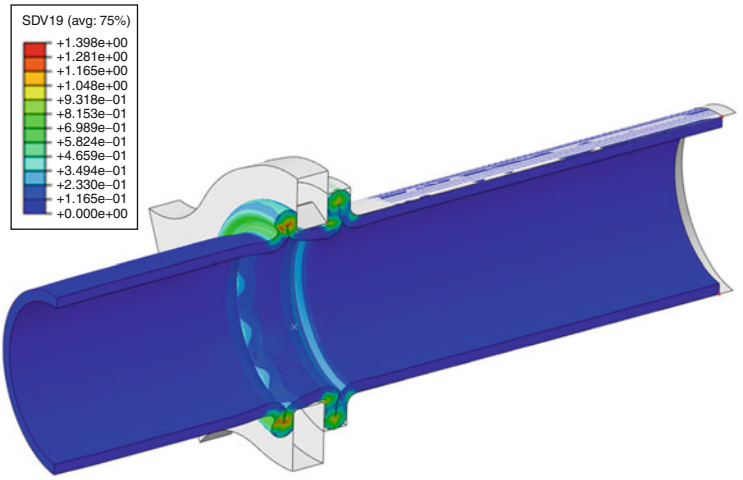


Fig. 12 Equivalent plastic strain distribution

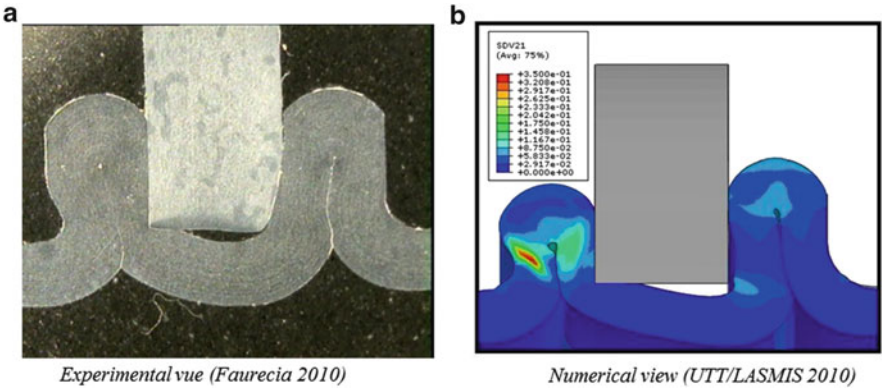


Fig. 13 The final view (zoom) of the tube-end process: (**a**) experimental view, (**b**) ductile damage distribution inside the core area

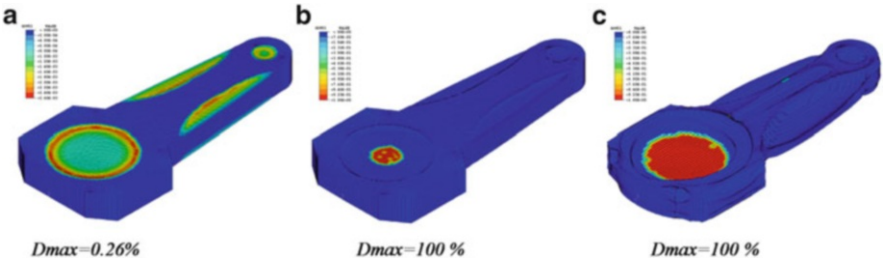


Fig. 14 Ductile damage distribution at three different instants of the forging process

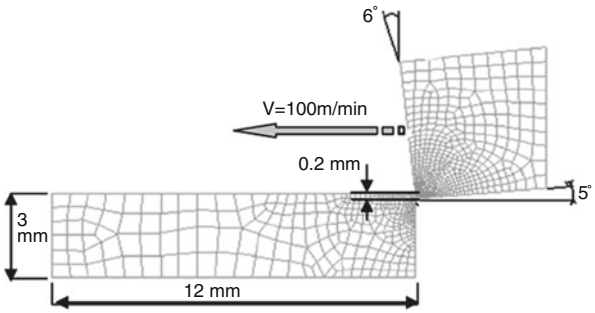


Fig. 15 Schematic representation of the orthogonal cutting problem (initial configuration at the first contact between the tool and the part)

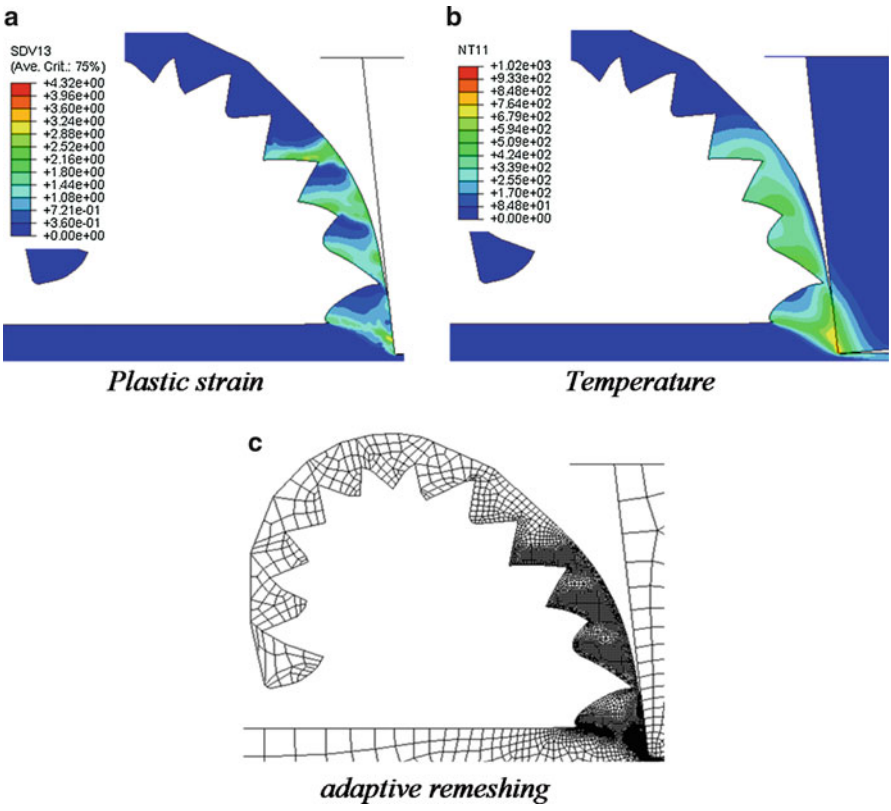


Fig. 16 Numerically predicted (**a**) accumulated plastic strain, (**b**) temperature, (**c**) adaptive remeshing at the tool displacement of 6.75 mm

Conclusions and Main Perspectives

In this chapter, an attempt has been made in order to show how the use of constitutive equations with high predictive capabilities greatly improves, qualitatively and quantitatively, the numerical prediction of the ductile damage occurrence

in sheet and bulk metal forming processes. This necessitates not only advanced macro or micro–macro (see ▶ Chap. 4, "Evolution of Fabric Tensors in Continuum Damage Mechanics of Solids with Micro-cracks: Studying the Effects of Length and Orientation" of this volume) constitutive equations to describe the thermomechanical behavior of the deformed material in the volume and at the contact interfaces, but also robust and adaptive numerical methods together with an appropriate efficient identification methodology in order to determine the pertinent values of the overall material parameters entering those fully coupled constitutive equations.

This kind of numerical methodology can be helpfully used to optimize any forming or machining process in order to determine the best forming or machining plan which minimizes the undesirable defects under a low economical cost and a low carbon emission according to the demand of the new environmental requirements. Accordingly, the use of highly predictive constitutive equations representing the "realistic" physical phenomena allows to:

- Minimize the ductile damage occurrence in formed parts in order to obtain a save component without any defects (i.e., cracks, localization zones, wrinkle zones, etc.) as in deep drawing of thin sheets, hydroforming tubes or sheets, hot or cold forging or extrusion, etc.
- Maximize the occurrence of ductile damage in formed parts by controlling it in intensity and direction to reproduce various material cutting processes such as stamping of thin or thick parts by sheering or slitting and guillotining thin sheets or machining by chip formation.
- Minimize various kinds of damage (wear, friction, cracks, etc.) inside forming tools (matrices, dies, cutting tools, etc.) and at the metal/tool contact interfaces in order to increase their life span.

These minimization and/or maximization tasks can be subjected to some constraints related to the reduction of the structure weight, the energy consumption, the market cost, or the environment criteria as the minimization of the carbon emission. More information can be found in the recent book dedicated to the ductile damage prediction in metal forming processes (Saanouni 2012a, b).

References

R. Arrieux, Determination and use of the forming limit stress diagrams in sheet metal forming. J. Mater. Process. Technol. **53**, 47–56 (1995)

R. Arrieux, M. Boivin, Determination of the forming limit stress curve for anisotropic sheets. CIRP Ann. Manuf. Technol. **16**(1), 195–198 (1987)

R. Arrieux, C. Bedrin, M. Boivin, Determination of an intrinsic forming limit stress diagram for isotropic metal sheets, in *Proceedings of 12th IDDRG*, (1982), pp. 61–71

R. Asaro, V. Lubards, *Mechanics of Solids and Materials* (Cambridge University Press, Cambridge, 2006)

A. Assempour, R. Hashemi, K. Abrinia, M. Ganjiani, E. Masoumi, A methodology for prediction of forming limit stress diagrams considering the strain path effect. Comput. Mater. Sci. **45**, 195–204 (2009)

H. Badreddine, K. Saanouni, A. Dogui, A. Gahbich, Elastoplasticité anisotrope non normale en grandes déformations avec Endommagement. Application à la mise en forme de tôles minces. Revue Européenne de Mécanique numérique **16**(6–7), 913–940 (2007)

H. Badreddine, K. Saanouni, A. Dogui, On non-associative anisotropic finite plasticity fully coupled with isotropic ductile damage for metal forming. Int. J. Plast. **26**, 1541–1575 (2010)

K.J. Bathe, *Finite Element Procedures* (Prentice Hall, Upper Saddle River, 1996)

T. Belytschko, W.K. Liu, B. Moran, *Nonlinear Finite Elements for Continua and Structures* (Wiley, NewYork, 2000)

J. Besson, G. Cailletaud, J.L. Chaboche, S. Forest, *Mécanique Non-linéaire des Matériaux* (Hermes, Paris, 2001)

J. Bonnet, R.D. Wood, *Nonlinear Continuum Mechanics for Finite Element Analysis* (Cambridge University Press, Cambridge, 1997)

N. Bontcheva, R. Iankov, Numerical investigation of the damage process in metal forming. Eng. Frac. Mech. **40**, 387–393 (1991)

M. Bornert, T. Brethau, P. Gilormini (eds.), *Homogénéisation en mécanique des matériaux*, vol. 1 et 2 (Hermès, Paris, 2001)

M. Boudifa, K. Saanouni, J.L. Chaboche, A micromechanical model for inelastic ductile damage prediction in polycrystalline metals. Int. J. Mech. Sci. **51**, 453–464 (2009)

J.D. Bressan, J.A. Williams, The use of a shear instability criterion to predict local necking in sheet metal deformation. Int. J. Mech. Sci. **25**, 155–168 (1983)

M. Brunet, F. Morestin, Experimental and analytical necking studies of anisotropic sheet metals. J. Mater. Process. Technol. **112**, 214–226 (2001)

M. Brunet, F. Sabourin, S. Mguil-Touchal, The prediction of necking and failure in 3D sheet forming analysis using damage variable. J. Geophys. Res. **6**, 473–482 (1996)

M. Brunet, F. Morestin, H. Walter-Leberre, Failure analysis of anisotropic sheet-metals using a nonlocal plastic damage model. J. Mater. Process. Technol. **170**(1-2), 457–470 (2005)

H.J. Bunge, *Texture Analysis in Materials Science – Mathematical Methods* (Butterworths, London, 1982)

J. Carbonnière, S. Thuillier, F. Sabourin, M. Brunet, P.Y. Manach, Comparison of the work hardening of metallic sheets in bending-unbending and simple shear. Int. J. Mech. Sci. **51**, 122–130 (2009)

J.M.A. Cesar de Sa, P.M.A. Areias, C. Zheng, Damage modelling in metal forming problems using an implicit non-local gradient model. Comput. Methods Appl. Mech. Eng. **195**, 6646–6660 (2006)

J.L. Chaboche, M. Boudifa, K. Saanouni, A CDM approach of ductile damage with plastic incompressibility. Int. J. Frac. **137**, 51–75 (2006)

A. Cherouat, K. Saanouni, Numerical simulation of sheet metal blanking process using a coupled finite elastoplastic damage modelling. Int. J. Form. Process **6**(1), 7–32 (2003)

A. Cherouat, K. Saanouni, Y. Hammi, Improvement of forging process of a 3D complex part with respect to damage occurrence. J Mater. Process. Technol. **142**(2), 307–317 (2002a)

A. Cherouat, K. Saanouni, Y. Hammi, Numerical improvement of thin tubes hydroforming with respect to ductile damage. Int. J. Mech. Sci. **44**, 2427–2446 (2002b)

W.Y. Chien, J. Pan, S.C. Tang, A combined necking and shear localization analysis for aluminum sheets under biaxial stretching conditions. Int. J. Plast. **20**, 1953–1981 (2004)

C.C. Chu, An analysis of localized necking in punch stretching. Int. J. Solids Struct. **16**, 913–921 (1980)

C.C. Chu, A. Needleman, Voids nucleation effects in biaxially stretched sheets. J. Eng. Mater. Technol. **102**, 249–256 (1980)

J.P. Cordebois, P. Ladevèze, Necking criterion applied in sheet metal forming, in *Plastic Behaviour of Anisotropic Solids*, ed. by J.P. Boehler (Editions CNRS, 1985)

E. Cosserat, F. Cosserat, Notes sur la théorie des corps déformables, in *Traité de Physique, t.2*, ed. by O.D. Chwolson (Hermann Librairie Scientifique, Paris, 1909), pp. 953–1173
E. Cosserat, F. Cosserat, *Théorie des Corps Déformables* (Hermann Editeurs, Paris, 2009). ISBN 978 27056 6920 1
M.A. Crisfield, *Nonlinear Finite Element Analysis of Solids and Structures*. Essentials, vol. 1 (Wiley, Chichester, 1991)
M.A. Crisfield, *Nonlinear Finite Element Analysis of Solids and Structures*. Advanced topics, vol. 2 (Wiley, Chichester, 1997)
R. Dautray, J.J. Lions, *Analyse mathématique et calcul numérique pour les sciences et les techniques: tomes 1, 2 et 3* (Dunaud, Paris, 1984)
N.E.A. De Souza, D. Peric, D.R.J. Owen, *Computational Methods for Plasticity: Theory and Applications* (Wiley, Chichester, 2008)
P.M. Dixit, U.S. Dixit, *Modeling of Metal Forming and Machining Processes by Finite Element and Soft Computing Methods* (Springer, London, 2008)
R. Engelen, M.G.D. Geers, F. Baaijens, Nonlocal implicit gradient-enhanced elasto-plasticity for the modelling of softening behaviour. Int. J. Plast. **19**, 403–433 (2003)
A.C. Eringen, *Microcontinuum Field Theories: Foundation and Solids* (Springer, New York, 1999)
A.C. Eringen, *Nonlocal Continuum Field Theories* (Springer, New York, 2002)
S. Forest, *Milieux Continus Généralisés et Milieux Hétérogènes* (Presses de l'Ecole des Mines, Paris, 2006)
S. Forest, Micromorphic approach for gradient elasticity, viscoplasticity and damage. ASCE J. Eng. Mech. **135**(3), 117–131 (2009)
S. Forest, E.C. Aifantis, Some links between recent gradient thermo-elasto-plasticity theories and the thermomechanics of generalized continua. Int. J. Solids Struct. **47**, 3367–3376 (2010)
S. Forest, R. Sievert, Elastoviscoplastic constitutive frameworks for generalized continua. Acta Mech. **160**, 71–111 (2003)
W. Gambin, *Plasticity and Textures* (Kluwer Academic, Dordrecht, 2001)
M.G.D. Geers, Finite strain logarithmic hyperelasto-plasticity with softening: a strongly non-local implicit gradient framework. Comput. Methods. Appl. Mech. Eng. **193**, 3377–3401 (2004)
M.G.D. Geers, R. Ubachs, R. Engelen, Strongly non-local gradient-enhanced finite strain elastoplasticity. Int. J. Numer. Methods Eng. **56**, 2039–2068 (2003)
J.C. Gelin, Finite element analysis of ductile fracture and defects formations in cold and hot forging. Ann. CIRP **39**, 215–218 (1990)
J.C. Gelin, J. Oudin, Y. Ravalard, An imposed finite element method for the analysis of damage and ductile fracture in cold metal forming processes. Ann. CIRP **34**(1), 209–213 (1985)
P. Germain, *Cours de mécanique des milieux continus* (Masson, Paris, 1973)
A.K. Ghosh, J.V. Laukonis, The influence of the strain path changes on the formability of sheet steel, in *9th Congress of IDDRG* (ASM Publication, 1976)
Y. Ghozzi, C. Labergère, K. Saanouni, Modelling and numerical simulation of thick sheets slitting using continuum damage mechanics. in *1st International Conference on Damage Mechanics (ICDM'2012)*, Belgrade, June 2012, pp. 25–27
A.F. Graf, W.F. Hosford, Calculations of forming limit diagrams for changing strain paths. Met. Trans. A **24**, 2497–2501 (1993)
M. Hamed, Formulations micromorphiques en élastoplasticité non-locale avec endommagement en transformations finies, Ph.D., University of Technology of Troyes, 2012
P. Hartley, S.E. Clift, J. Salimi, C.E.N. Sturgess, I. Pillinger, The prediction of ductile fracture initiation in metal forming using a finite element method and various fracture criteria. Res. Mech. **28**, 269–293 (1989)
K.S. Havner, *Finite Plastic Deformation of Crystalline Solids* (Cambridge University Press, Cambridge, 1992)
E. Hinton, *Introduction to Nonlinear Finite Element Analysis* (NAFEMS, Glasgow, 1992)

Hora P., Tong L., 2009, Prediction of failure under complex 3D-stress conditions, in *Proceedings of Forming Technology Forum 2009*, IVP, ETH Zurich, 5–6 May 2009, pp. 133–138

T.J.R. Hughes, *The Finite Element Method. Linear Static and Dynamic Finite Element Analysis* (Dover Publications, Mineola, 1987)

T.J.R. Hughes, J. Winget, Finite rotation effects in numerical integration of rate-constitutive equations arising in large-deformation analysis. Int. J. Numer. Methods Eng. **15**, 1862–1867 (1980)

A. Ibrahimbegovic, *Mécanique non linéaire des solides déformables* (Hermes, Paris, 2006)

M. Issa, K. Saanouni, C. Labergère, A. Rassineux, Prediction of serrated chip formation in orthogonal metal cutting by advanced adaptive 2D numerical methodology. Int. J. Mach. Machinab. Mater. **9**(3/4), 295–315 (2011)

M. Issa, C. Labergère, K. Saanouni, A. Rassineux, Numerical prediction of thermomechanical fields localization in orthogonal cutting. CIRP J. Manuf. Sci. Technol. **5**, 175–195 (2012)

S. Kobayashi, S.I. Oh, T. Altan, *Metal Forming and the Finite Element Method* (Oxford University Press, Oxford, 1989)

U.F. Kocks, C.N. Tomé, H.R. Wenk, *Texture and Anisotropy: Preferred Orientations in Polycrystals and Their Effect on Material Properties* (Cambridge University Press, Cambridge, 1998)

C. Labergère, A. Rassineux, K. Saanouni, 2D adaptive mesh methodology for the simulation of metal forming processes with damage. Int. J. Mater. Form. **4**(3), 317–328 (2011)

P. Ladevèze, *Non Linear Computational Structural Mechanics* (Springer, New York, 1998)

T.A. Laursen, *Computational Contact and Impact Mechanics: Fundamentals of Modelling Interfacial Phenomena in Nonlinear Finite Element Analysis* (Springer, Berlin, 2002)

N. Le Maoût, S. Thuillier, P.Y. Manach, Aluminium alloy damage evolution for different strain paths- application to hemming process. Eng. Frac. Mech. **76**, 1202–1214 (2009)

H. Lee, K.E. Peng, J. Wang, An anisotropic damage criterion for deformation instability and its application to forming limit analysis of metal plates. Eng. Frac. Mech. **21**(5), 1031–1054 (1985)

J. Lemaitre, J.L. Chaboche, A. Benallal, R. Desmorat, *Mécanique des matériaux solides*, 3rd edn. (Dunod, Paris, 2009)

P. Lestriez, K. Saanouni, J.F. Mariage, A. Cherouat, Numerical prediction of damage in metal forming process including thermal effects. Int. J. Damage Mech. **13**(1), 59–80 (2004)

P. Lestriez, K. Saanouni, A. Cherouat, Simulation numérique de la coupe orthogonale par couplage thermique-comportement-endommagement en transformations finies. Mécanique Indust. **6**, 297–307 (2005)

J.F. Mariage, K. Saanouni, P. Lestriez, A. Cherouat, Numerical simulation of an hexnut forming process including damage effect. Int. J. Form. Process. **8**(2), 291–310 (2005)

Z. Mariciniack, K. Kunczynski, Limit strain in the processes of stretch forming sheet steel. J. Mech. Phys. Solids **1**, 609–620 (1967)

Z. Mariciniack, K. Kunczynski, T. Pokora, Influence of plastic properties of a material on the forming limit diagram for sheet metal in tension. Int. J. Mech. Sci. **15**, 789–803 (1973)

K. Mathur, P. Dawson, Damage evolution modelling in bulk forming processes, in *Computational Methods for Predicting Material Processing Defects*, ed. by M. Predeleanu (Elsevier, 1987)

P.H. Matin, L.M. Smith, S. Petrusevski, A method for stress space forming limit diagram construction for aluminium alloys. J. Mat. Proc. Techn. **174**, 258–265 (2006)

T. Mura, *Micromechanics of Defects in Solids* (Martinus Nijhoff Publishers, Dordrecht, 1987)

A. Needleman, N. Triantafyllidis, Void growth and local necking in biaxially stretched sheets. J. Eng. Mater. Technol. **100**, 164–172 (1980)

S. Nemat-Nasser, *Plasticity. A Treatise on Finite Deformation of Heterogeneous Inelastic Materials* (Cambridge University Press, Cambridge, 2004)

S. Nemat-Nasser, M. Hori, *Micromechanics: Overall Properties of Heterogeneous Materials* (Elsevier, Amsterdam, 1993)

T.D. Nguyen, Anisotropie de l'endommagement et simulations numériques en mise en forme par grandes déformations plastiques, Ph.D., University of Technology of Troyes, 2012
J.T. Oden, *Finite Elements of Nonlinear Continua* (McGraw-Hill, New York, 1972)
E. Onate, M. Kleiber, Plastic and viscoplastic flow of void containing metal – applications to axisymmetric sheet forming problem. Int. J. Numer. Methods Eng. **25**, 237–251 (1988)
D.R.J. Owen, E. Hinton, *Finite Elements in Plasticity: Theories and Practice* (Pineridge Press, Swansea, 1980)
P. Picart, O. Ghouati, J.C. Gelin, Optimization of metal forming process parameters with damage minimization. J. Mater. Process. Technol. **80–81**, 597–601 (1998)
D. Raabe, *Computational Material Science: The Simulation of Materials Microstructures and Properties* (Wiley-VCH, Weinheim, 1998)
J.N. Reddy, *An Introduction to Nonlinear Finite Element Analysis* (Oxford University Press, Oxford, 2004)
G.H. Rowe, C.E.N. Sturguess, P. Hartley, I. Pillinger, *Finite Element Plasticity and Metal Forming Analysis* (Cambridge University Press, Cambridge, 1991)
K. Saanouni, Virtual metal forming including the ductile damage occurrence, actual state of the art and main perspectives. J. Mater. Process. Technol. **177**, 19–25 (2006)
K. Saanouni, On the numerical prediction of the ductile fracture in metal forming. Eng. Frac. Mech. **75**, 3545–3559 (2008)
K. Saanouni, *Damage Mechanics in Metal Forming. Advanced Modeling and Numerical Simulation* (ISTE John Wiley, London, 2012a). ISBN 978-1-8482-1348-7
K. Saanouni, *Modélisation et simulation numériques en formage Virtuel* (Hermes, Paris, 2012b). ISBN 978-2-7462-3225-9
K. Saanouni, M. Hamed, Micromorphic approach of finite gradient-elastoplasticity fully coupled with ductile damage. Formulation and computational approaches. Int. J Solids Struct. **50**, 2289–2309 (2013)
K. Saanouni, P. Lestriez, Modelling and numerical simulation of ductile damage in bulk metal forming. Steel Res. Int. **80**(9), 645–657 (2009)
K. Saanouni, C. Forster, F. Benhatira, On the anelastic flow with damage. Int. J. Damage Mech. **3** (2), 140–169 (1994)
K. Saanouni, K. Nesnas, Y. Hammi, Damage modelling in metal forming processes. Int. J. Damage Mech. **9**, 196–240 (2000)
K. Saanouni, A. Cherouat, Y. Hammi, Numerical aspects of finite elastoplasticity with isotropic ductile damage for metal forming. Revue Européenne des E. F **10**(2-3-4), 327–351 (2001)
K. Saanouni, J.L. Chaboche, Computational damage mechanics. Application to metal forming', chapter 3.06, in *Comprehensive Structural Integrity*, ed. by I. Milne, R.O. Ritchie, B. Karihaloo. Numerical and Computational Methods (editors: R. de Borst, H. A. Mang), vol. 3 (Elsevier Ltd, Oxford, 2003), pp. 321–376. ISBN: 0-08-043749-4
K. Saanouni, J.F. Mariage, A. Cherouat, P. Lestriez, 2004, Numerical prediction of discontinuous central bursting in axisymmetric forward extrusion by continuum damage mechanics. Comput. Struct. **82**, 2309–2332 (2004)
K. Saanouni, H. Badreddine, M. Ajmal, Advances in virtual metal forming including the ductile damage occurrence, application to 3D sheet metal deep drawing. J. Eng. Mater. Technol. **130**, 021022-1–021022-1 (2008)
K. Saanouni, N. Belamri, P. Autesserre, Finite element simulation of 3D sheet metal guillotining using advanced fully coupled elasto-plastic damage constitutive equations. J. Finite Elem Anal. Des **46**, 535–550 (2010)
K. Saanouni, P. Lestriez, C. Labergère, 2D adaptive simulations in finite thermo-elasto-viscoplasticity with ductile damage : application to orthogonal metal cutting by chip formation and breaking. Int. J. Damage Mech. **20**(1), 23–61 (2011)
E. Schmid, W. Boas, *Plasticity of Cristals* (Chapman and Hall, London, 1968)
F. Sidoroff, A. Dogui, Some issues about anisotropic elastic–plastic models at finite strain. Int. J. Solids Struct. **38**, 9569–9578 (2001)

J.C. Simo, T.J.R. Hughes, *Computational Inelasticity* (Springer, New York, 1997)
D. Sornin, K. Saanouni, About elastoplastic non-local formulations with damage gradients. Int. J. Damage Mech. **20**(6), 845–875 (2011)
C. Soyarslan, A.E. Tekkaya, U. Akyüz, Application of continuum damage mechanics in crack propagation problems: forward extrusion chevron predictions. Z. Angew. Math. Mech. **88**(6), 436–453 (2008)
S. Storen, J.R. Rice, Localized necking in sheets. J. Mech. Phys. Solids **23**, 421–441 (1975)
T. Stoughton, Stress-based forming limits in sheet metal forming. J. Eng. Mater. Technol. **123**, 417–422 (2001)
W. Swift, Plastic instability under plane stress. J. Mech. Phys. Solids **1**, 1–18 (1952)
C. Truesdell, W. Noll, *The Nonlinear Filed Theories of Mechanics*, 1st edn. (Springer, New York, 1965)
C. Truesdell, W. Noll, *The Nonlinear Filed Theories of Mechanics*, 3rd edn. (Springer, New York, 2004)
P. Villon, H. Borouchaki, K. Saanouni, Transfert de champs plastiquement admissibles. CRAS, Mécanique **330**, 313–318 (2002)
R.H. Wagoner, J.L. Chenot, *Metal Forming Analysis* (Cambridge University Press, Cambridge, 2001)
K. Walker, A. Freed, *Asymptotic Integration Algorithm for Nonhomogeneous, Nonlinear First Order ODEs* (Engineering Science Software, NASA Technical Memorendum, 1991)
P. Wriggers, *Computational Contact Mechanics* (Wiley, Hoboken, 2002)
P. Wriggers, *Nonlinear Finite Element Methods* (Springer, Berlin, 2008)
W. Yang, W.B. Lee, *Mesoplasticity and Its Applications* (Springer, Berlin, 1993)
K. Yoshida, T. Kuwabara, Effect of strain hardening, behaviour on forming limit stresses of steel tube subjected to nonproportional loading paths. Int. J. Plast. **23**, 1260–1284 (2007)
M. Yoshida, F. Yoshida, H. Konishi, K. Fukumoto, Fracture limits of sheet metals under stretch bending. Int. J. Mech. Sci. **47**, 1885–1896 (2005)
Z.H. Zhong, *Finite Element Procedures for Contact-Impact Problems* (Oxford University Press, Oxford, 1993)
Y.Y. Zhu, S. Cescotto, The finite element prediction of ductile fracture initiation in dynamic metal forming processes. J. Phys. III **1**, 751–757 (1991)
Y.Y. Zhu, S. Cescotto, A fully coupled elasto-visco-plastic damage theory for anisotropic materials. Int. J. Solids Struct. **32**(11), 1607–1641 (1995)
Y.Y. Zhu, S. Cescotto, A.M. Habraken, A fully coupled elastoplastic damage modelling and fracture criteria in metal forming processes. J. Mater. Process. Technol. **32**, 197–204 (1992)
O.C. Zienkiewicz, R.L. Taylor, *The Finite Element Method for Solids and Structural Mechanics*, 1st edn. (Elsevier, Burlington, 1967) (6th edition in 2005)

Ductile Failure Modeling: Stress Dependence, Non-locality and Damage to Fracture Transition

26

J. M. A. Cesar de Sa, F. M. A. Pires, F. X. C. Andrade, L. Malcher, and M. R. R. Seabra

Contents

J.M.A.C. de Sa (✉) • F.M.A. Pires • M.R.R. Seabra
Faculty of Engineering, University of Porto, Porto, Portugal
e-mail: cesarsa@fe.up.pt; fpires@fe.up.pt; marianas@fe.up.pt

F.X.C. Andrade
DYNAmore GmbH, Stuttgart, Germany
e-mail: filipe.andrade@dynamore.de

L. Malcher
Department of Mechanical Engineering, University of Brasilia, Asa Norte - Brasília, DF, Brazil
e-mail: malcher@unb.br

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_39

Abstract

In this chapter, some recent developments and proposals for improvement of material models at the constitutive level to deal with ductile damage at large plastic strains are addressed. Numerical tests are carried out to test their performance on shear-dominated stress states where their main differences lie. Subsequently, aspects of the use of nonlocal models for the regularization of the numerical values associated with damage models, namely, discretization dependency, are reviewed. Different approaches on the choice of the regulation variable or variables are tested at different stress states characterized by different values of triaxiality and third invariant of the deviatoric stress tensor. Finally, a simple strategy on how to handle the transition from damage to fracture by means of the extended finite element method is described.

Introduction

Metal forming processes have an enormous industrial importance as they are involved in the manufacture of a huge variety of structural parts in a diversity of industrial sectors, such as automotive, aeronautics, and consumption goods, to name a few. These processes encompass significant changes in the shape of the material in the solid state, from an initial component to the final product. The material flow, forced by tools and very often controlled by contact and friction against dies, generally involves large plastic strains and depends on various factors such as load conditions; temperature; geometry of preforms, tools, and dies; lubrication of contact zones; material properties; and forming limits, to name only the most important ones. A crucial issue in the design, development, and optimization of metal forming processes is to be able to control most of these parameters. Numerical modelling plays here an important role, and in the last decades, powerful commercial codes have become indispensable tools in the industry. The great development and evolution on both theoretical and numerical abilities, allied to the fast development in computing facilities, allow us today, to a large extent, to predict deformation, strain and stresses at critical points, changes in material properties, the influence of tools geometry, lubrication conditions, etc.

In large deformation of metals, when plastic deformation reaches a threshold level, which may depend on the loading, the fatigue limit, and the ultimate stress, a ductile damage process may occur concomitantly with the plastic deformation due to the nucleation, growth, and coalescence of micro-voids. Therefore, once the main features of large deformation modelling were mastered, the need to accurately predict material formability, under complex loading paths, became a decisive feature to accomplish, in order to be able to avoid defective parts in forming processes or to describe processes in which fracture is a part of the process itself, as in sheet blanking or metal cutting.

In the design of bulk forming processes, at the industrial level, it is still common to utilize fracture criteria, based on the computational evaluation of functions of

some state variables that depend on the deformation history. Those criteria may be, generally, categorized into two groups: one based on micromechanics which utilizes as primer state variable the total plastic work (e.g., Freudenthal 1950), the maximum plastic shear work, or the equivalent plastic strain (e.g., Datsko 1966), and another based on the growth of defects which includes geometric aspects (e.g., McClintock 1968; Rice and Tracey 1969), growth mechanisms dependent on principal stresses (e.g., Cockcroft and Latham 1968) or hydrostatic pressure (e.g., Norris et al. 1978; Atkins 1981), or material behavior coupling (e.g., Oyane et al. 1978; Tai and Yang 1987; Lemaître 1986).

However, these are, commonly, a posteriori fracture indicators that are not always reliable. In complex deformation paths, more and more present in the production of many components in the industry, very often they fail to give the appropriate information. Situations in which the damage localizes away from the sites where the maximum equivalent plastic deformation is concentrated or where damage evolves differently for different compression or traction stress sates, different triaxialities or diverse shear stress states are hardly handled by these criteria.

More robust models were therefore required so that damage evolution is taken into account through the deformation process. The pioneering work of Kachanov (1958) and Rabotnov (Rabotnov et al. 1963) paved the way for numerous significant contributions and new theories proposed for the description of material progressive internal degradation. One route is based on Continuum Damage Mechanics (CDM) and the thermodynamics of irreversible processes, mainly coupling elastoplasticity and damage at the constitutive level (e.g., Lemaitre 1985a, b, 1996). Another direction is based on micromechanical grounds, i.e., coupling damage and plasticity at the constitutive level (e.g., Gurson 1977; Tvergaard and Needleman 1984; Xue et al. 2007).

However, a large number of those models are rooted on the assumption of the so-called local continuum. In local media, the behavior of the material is completely represented by a point-wise constitutive law, which is independent of the influence of surrounding material points. As a matter of fact, the local theory assumes that the material is continuous at any scale and, therefore, size effects are inherently neglected. Nevertheless, the softening induced by the standard implementation of those models in finite element solutions, within the local theory, leads to mesh and orientation dependence. This fact is associated to the local change of the underlying type of differential equations representing the problem whenever a negative stiffness is locally included due to softening. As a consequence, localization effects are not correctly dealt with by mesh refinement. One of the solutions for this problem is the use of nonlocal models (e.g., Pijaudier-Cabot and Bažant 1987; De Borst and Mühlhaus 1992; De Vree et al. 1995; Strömberg and Ristinmaa 1996; Polizzotto et al. 1998; Borino et al. 1999; Jirásek and Rolshoven 2003; Cesar de Sa et al. 2006; Jirásek 2007; Andrade et al. 2009). Nonlocal models include some length scale information, related with localization effects due to microstructure heterogeneity, in order to average an internal variable effect associated with dissipative process. Two types of models are usually assumed for this purpose: integral and gradient models.

Nowadays a lot of research is being invested on multi-scales models aiming to gain an important insight on how the damage mechanisms at lower scales reflect themselves at the macroscale and how that can be replicated by the phenomenological laws usually adopted at this level. Nevertheless, these models require a huge amount of computer capabilities which, by now, limits their application at the industry level, and therefore, in this chapter we will not address those issues.

After this small introduction, initially recent developments and proposals for improvement of material models at the constitutive level to deal with ductile damage at large plastic strains are addressed. Some numerical tests are accomplished to test their performance on shear-dominated stress states where their main differences lie. Subsequently, some aspects of the use of nonlocal models for the regularization of the numerical values associated with damage models, namely, discretization dependency, are reviewed. Different approaches on the choice of the regulation variable or variables are tested in various situations at different stress states characterized by the different values of triaxiality and third invariant of the deviatoric stress tensor. Finally, a simple strategy on how to handle the transition from damage to fracture by means of the extended finite element method is described.

Constitutive Models for Ductile Failure Under High and Low Triaxiality

Introduction

Experimental evidence has shown that the nucleation and growth of voids and micro-cracks, which accompany large plastic flow, cause a reduction of the elastic modulus, induce a softening effect in the material, and can be strongly influenced by the level of stress triaxiality (McClintock 1968; Rice and Tracey 1969; Hancock and Mackenzie 1976). The equivalent plastic strain at fracture and the level of stress triaxiality were initially employed to characterize material ductility in engineering applications (Bridgman 1952; McClintock 1968; Rice and Tracey 1969; Johnson and Cook 1985). A simple exponential expression for the evolution of the equivalent strain with stress triaxiality was established by McClintock (1968) and Rice and Tracey (1969) based on the analysis of void growth under hydrostatic loads, which is usually referred to as the two-dimensional fracture loci. The work performed by Mirza et al. (1996) on pure iron, mild steel, and aluminum alloy BS1474 over a wide range of strain rates confirmed the strong dependence of the equivalent strain to crack formation with the level of stress triaxiality.

Recently, several researchers (Kim et al. 2003, 2004; Bao and Wierzbicki 2004; Gao et al. 2005; Gao and Kim 2006; Kim et al. 2007; Barsoum and Faleskog 2007a, b; Bai and Wierzbicki 2008; Brünig et al. 2008; Gao et al. 2009) have shown that the Lode angle, which is associated to the third invariant of the deviatoric stress tensor, is an essential parameter in the characterization of the effect of the stress state on material yielding and on ductile fracture. In particular, Bai and Wierzbicki (2008) have suggested a three-dimensional fracture loci on the space of equivalent strain,

stress triaxiality, and Lode angle. This fracture surface is clearly different for materials weakly or strongly dependent on both pressure and Lode angle and can be calibrated by means of conventional and butterfly specimens. Mirone and Corallo (2010) have proposed a local viewpoint for evaluating the influence of the stress triaxiality and Lode angle on ductile failure, analyzing three theories, namely, the Tresca criteria and two models proposed by Wierzbicki. According to Mirone and Corallo (2010), the phenomenon of ductile failure is influenced by the relation with the variables from the stress-strain characterization, and failure predictions are better described by plastic strain, stress triaxiality, and Lode angle parameters. An experimental program to study the influence of the stress tensor invariants in ductile failure was presented by Driemeier et al. (2010). This methodology can be seen as an efficient tool to investigate the effects of the stress intensity, stress triaxiality, and Lode angle. Gao et al. (2011) have proposed a new elastoplastic model, which is a function of the hydrostatic stress as well as the second and third invariants of the stress deviator, and carried out tests in specimens with a high level of stress triaxiality showing the dependence of the plastic flow rule of both stress triaxiality and Lode angle.

Some Constitutive Models for Ductile Failure

The governing equations of three constitutive models are briefly reviewed in this section. Firstly, the Gurson–Tvergaard–Needleman (GTN) model (Gurson 1977; Tvergaard and Needleman 1984) is presented, then the Lemaitre model (Lemaitre 1985a), and finally Bai and Wierzbicki's model (Bai and Wierzbicki 2008). In addition, a shear mechanism proposed by Xue (2008) is described and incorporated in the GTN model as well as Bao's fracture indicator (Bao et al. 2003), which is used in conjunction with Bai and Wierzbicki's model to allow the prediction of damage with this model.

The Gurson–Tvergaard–Needleman Model

Inspired by the work of Gurson (1977), Tvergaard and Needleman (1984) have proposed a model for the description of damage and fracture in ductile materials. The original Gurson model introduces a strong coupling between plastic strain and damage (Chaboche et al. 2006), and the presence of micro-voids in the formulation leads to a yield surface that depends on both the hydrostatic pressure and porosity. The material degradation is measured through a parameter called the void volume fraction, which is represented by the variable f. This parameter is defined by the ratio between the volume of micro-voids, V_{voids}, and the representative volume element, V_{RVE}

$$f = \frac{V_{\text{voids}}}{V_{\text{RVE}}} \tag{1}$$

The Gurson–Tvergaard–Needleman (GTN) model, which is one of the most well-known extensions of the Gurson model, assumes both isotropic hardening and

damage. Nevertheless, the damage variable in this model is represented by an effective porosity f^*. The flow potential is generalized into the form:

$$\Phi(\boldsymbol{\sigma}, r, f^*) = q - \frac{1}{3}\left\{1 + q_3 f^{*2} - 2q_1 f^* \cosh\left(\frac{q_2 3p}{2\sigma_y}\right)\right\}\sigma_y^2 \tag{2}$$

where q represents the von Mises equivalent stress and σ_y is the radius of the yield surface function of the isotropic hardening. The parameters q_1, q_2, and q_3 are introduced into the yield surface definition in order to bring the model predictions into closer agreement with full numerical analyses of a periodic array of voids and p represents the hydrostatic pressure. The evolution of spherical voids can be reproduced by three simultaneous or successive steps: nucleation, growth, and coalescence of voids (Tvergaard and Needleman 1984). The effective porosity is determined by the following bilinear function:

$$f^* = \begin{cases} f, & f < f_c \\ f_c + \left(\frac{1}{q_1} - f_c\right)\frac{(f - f_c)}{\left(f_f - f_c\right)}, & f \geq f_c \end{cases} \tag{3}$$

where the parameter f represents the porosity, the constant f_c is the porosity to trigger coalescence, and the parameter f_f represents the porosity at fracture. The evolution of the porosity is given by the sum of both the nucleation and growth mechanisms, as

$$\dot{f} = \dot{f}^N + \dot{f}^G \tag{4}$$

The nucleation mechanism is driven by the plastic strain and can be represented as

$$\dot{f}^N = \frac{f^{\mathrm{N}}}{s_{\mathrm{N}}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon^p_{eq} - \varepsilon_{\mathrm{N}}}{s_{\mathrm{N}}}\right)^2\right]\dot{\varepsilon}^p_{eq} \tag{5}$$

where f^{N} represents the volume fraction of all second-phase particles (see Fig. 1) with potential for micro-void nucleation and ε_{N} and s_{N} are the mean strain for void nucleation and its standard deviation. The variable ε^p_{eq} represents the equivalent plastic strain and $\dot{\varepsilon}^p_{eq}$ is the rate of the equivalent plastic strain.

The most significant contribution to the evolution of spherical voids is the growth mechanism, obtained from the condition of plastic incompressibility of the matrix material, which can be expressed by

$$\dot{f}^G = (1 - f)\mathrm{tr}(\dot{\boldsymbol{\varepsilon}}^p) = (1 - f)\dot{\varepsilon}^p_v \tag{6}$$

where $\dot{\boldsymbol{\varepsilon}}^p$ represents the rate of the plastic strain tensor and $\dot{\varepsilon}^p_v$ is the rate of the volumetric plastic strain. In this work, the GTN model implementation includes

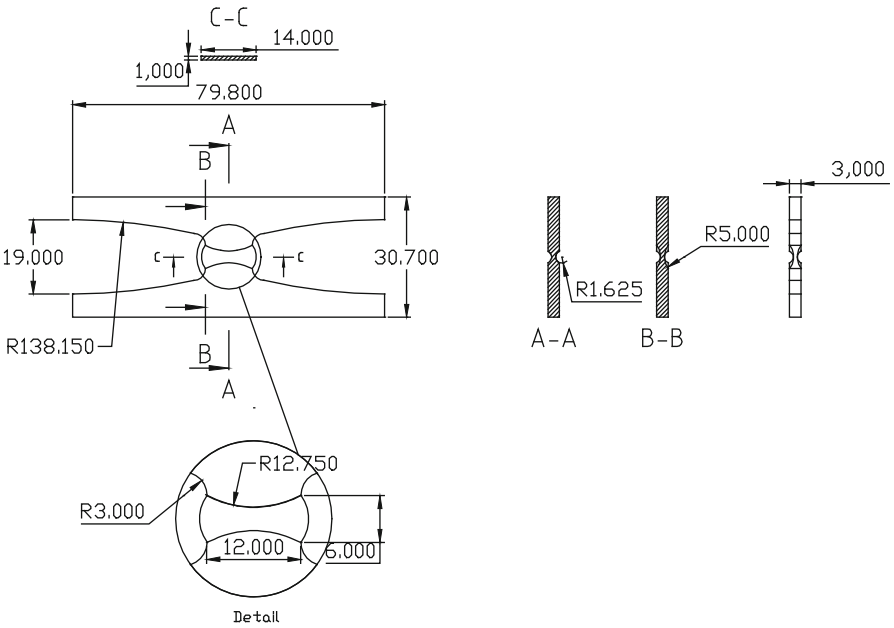


Fig. 1 Geometry of the butterfly specimen (dimensions in mm) (The specimen was reproduced from Bai et al. (2008)). Reprinted from L. Malcher et al., An Assessment of Isotropic Constitutive Models for Ductile Fracture under High and Low Stress Triaxiality. Int. J. Plast. **30–31**, 81–115 (2012), with permission from Elsevier

both nucleation and growth of micro-voids. The coalescence effect was not addressed since our main objective is the prediction of fracture onset.

One important limitation associated with Gurson-based models is that shear effects are not considered in the formulation, which excludes the possibility of predicting shear localization and fracture under conditions of low triaxiality. Therefore, in order to improve the GTN model predictive ability, under both zero and low levels of stress triaxialities, Xue (2008) has proposed the introduction of a shear mechanism. The mechanism is based on geometrical considerations of a unit cell structure, containing a circular void at the center, which is subjected to a simple shear strain (Xue 2008). The evolution of shear damage, according to the author, depends on the porosity, the equivalent strain, and the Lode angle. After some straightforward algebraic manipulations, the rate of this mechanism can be mathematically expressed by (Xue 2008)

$$\dot{f}^{\text{Shear}} = q_4 f^{q_5} g_o \varepsilon_{eq} \dot{\varepsilon}_{eq} \tag{7}$$

where q_4 and q_5 are parameters related to two- or three-dimensional problems. For two-dimensional problems, $q_4 = 1.69$ and $q_5 = 1/2$, and for three-dimensional problems, $q_4 = 1.86$ and $q_5 = 1/3$. The variable f represents the porosity, ε_{eq} is the equivalent strain, and g_o is a parameter that introduces the Lode angle

dependence in the shear mechanism. If the Lode angle function g_o is different from zero, the mechanism is triggered and shear effects are taken into account. However, if g_o is zero, the shear mechanism has no effect on the damage evolution, and only the nucleation and growth mechanisms are active. The Lode angle function, g_o, can be defined by

$$g_o = 1 - \frac{6|\theta|}{\pi} \tag{8}$$

where θ is the Lode angle that is determined according to

$$\theta = \tan^{-1}\left\{\frac{1}{\sqrt{3}}\left[2\left(\frac{S_2 - S_3}{S_1 - S_3}\right) - 1\right]\right\}. \tag{9}$$

The scalars S_1, S_2 and S_3 are the components of the deviatoric stress tensor in the principal plane. The shear mechanism proposed by Xue (2008) can be included in the GTN model, which already features the mechanisms of nucleation and growth of micro-voids. Thus, the evolution of the porosity originally expressed by Eq. 8, for this model, is redefined as

$$\dot{f} = \dot{f}^N + \dot{f}^G + \dot{f}^{\text{Shear}} \tag{10}$$

The evolution of damage in the material inevitably reduces the overall elastic properties. However, this effect is small when compared to the influence of damage on the plastic behavior. Therefore, the evolution of damage due to shear effects, employed in this work, will neglect the influence of damage on elasticity as is usually done in this type of model. The shear damage evolution law is redefined as a function of both the accumulated plastic strain and the rate of the accumulated plastic strain instead of the total strain and total strain rate (see Eq. 7):

$$\dot{f}^{\text{Shear}} = q_4 f^{q_5} g_o \varepsilon^p_{eq} \dot{\varepsilon}^p_{eq} \tag{11}$$

The Lode angle function can also be rewritten as a function of the normalized third invariant, such as

$$g_o = 1 - \left|\overline{\theta}\right| \tag{12}$$

where $\overline{\theta}$ represents the normalized Lode angle that is a function of the normalized third invariant, such as

$$\overline{\theta} = 1 - \frac{6\theta}{\pi} = 1 - \frac{2}{\pi}\arccos\xi \tag{13}$$

where ξ represents the normalized third invariant that is calculated by

$$\xi = \frac{27}{2}\frac{\det \boldsymbol{\varepsilon}^e_d}{\left(\frac{3}{2}\boldsymbol{\varepsilon}^e_d : \boldsymbol{\varepsilon}^e_d\right)^{3/2}} \tag{14}$$

where $\boldsymbol{\varepsilon}^e_d$ represents the deviatoric elastic strain tensor.

Lemaitre's Damage Model

The constitutive equations for ductile damage, described in this section, have been proposed by Lemaitre (1985a). Based on the concept of effective stress and the hypothesis of strain equivalence, the Lemaitre model includes the evolution of internal damage as well as nonlinear isotropic and kinematic hardening in the description of the behavior of ductile materials. The constitutive formulation starts from the definition of the Helmholtz specific free energy that can be taken as the state potential of the material and is a function of all state variables. The free energy can be expressed as a function of the set $\{\boldsymbol{\varepsilon}^e, r, D\}$ of state variables:

$$\psi = \psi(\boldsymbol{\varepsilon}^e, r, D) \tag{15}$$

where ψ represents the specific free energy, $\boldsymbol{\varepsilon}^e$ is the elastic strain tensor, r is the isotropic hardening internal variable, and D represents the isotropic damage internal variable.

Under the hypothesis of decoupling between elasticity–damage and plastic hardening, the specific free energy is assumed to be given by the sum:

$$\psi = \psi^{ed}(\boldsymbol{\varepsilon}^e, D) + \psi^p(r) \tag{16}$$

where ψ^{ed} represents the elastic–damage contribution and ψ^p is the plastic contribution to the free energy. The elastic–damage contribution for the free energy can be postulated by the following expression (Lemaitre 1985a):

$$\overline{\rho}\psi^{ed}(\boldsymbol{\varepsilon}^e, D) = \frac{1}{2}(1-D)\boldsymbol{\varepsilon}^e : \mathbf{D}^e : \boldsymbol{\varepsilon}^e \tag{17}$$

where $\mathbf{D}^e$ represents the isotropic elasticity tensor. The plastic potential can be represented by the isotropic hardening contribution as (if we disregard kinematic hardening):

$$\overline{\rho}\psi^p(r) = \overline{\rho}\psi^I(r) \tag{18}$$

The elasticity law is obtained by performing the derivative of the elastic–damage potential (Eq. 17) in order to the elastic strain tensor, as

$$\boldsymbol{\sigma} = \overline{\rho}\frac{\partial \psi^{ed}}{\partial \boldsymbol{\varepsilon}^e} = (1-D)\mathbf{D}^e : \boldsymbol{\varepsilon}^e \tag{19}$$

The thermodynamic forces conjugated with damage and isotropic hardening internal variable are obtained, respectively, by performing the derivative of the elastic–damage contribution, $\overline{\rho}\psi^{ed}(\boldsymbol{\varepsilon}^e, D)$ (Eq. 17), with regard to the damage variable, D, and by taking the derivative of the plastic potential, $\overline{\rho}\psi^p(r)$ (Eq. 18), with regard to the isotropic hardening variable, R, respectively (Lemaitre and Desmorat 2005):

$$-Y \equiv -\overline{\rho}\frac{\partial \psi^{ed}}{\partial D} = \frac{q^2}{6G(1-D)^2} + \frac{p^2}{2K(1-D)^2} \tag{20}$$

$$R \equiv -\overline{\rho}\frac{\partial \psi^{I}}{\partial r} = R(r) \tag{21}$$

where Y represents the thermodynamic force associated with damage, q is the von Mises equivalent stress, p is the hydrostatic pressure, G is the shear elasticity modulus, K is the elastic compressibility modulus, and R represents the thermodynamic force associated with the isotropic hardening variable.

The evolution of the internal variable can be obtained by assuming the existence of the flow potential, Ψ, given by

$$\Psi = \Phi + \frac{S}{(1-D)(s+1)}\left(\frac{-Y}{S}\right)^{s+1} \tag{22}$$

where the parameters S and s are damage evolution constants and Φ represents the yield function, which is defined as

$$\Phi = \frac{q}{(1-D)} - \sigma_{y_0} - R(r) \tag{23}$$

where σ_{y_0} is the initial uniaxial yield stress. According to the hypothesis of generalized normality, the plastic flow is given by

$$\dot{\boldsymbol{\varepsilon}}^{p} = \dot{\gamma}\frac{\partial \Phi}{\partial \boldsymbol{\sigma}} = \dot{\gamma}\boldsymbol{N} \tag{24}$$

$$\boldsymbol{N} = \sqrt{\frac{3}{2}}\frac{\boldsymbol{S}}{\|\boldsymbol{S}\|}\frac{1}{(1-D)} \tag{25}$$

where $\dot{\gamma}$ is the plastic multiplier, $\boldsymbol{N}$ represents the flow vector, and $\boldsymbol{S}$ is the deviatoric stress tensor. The evolution law for damage and for the isotropic hardening internal variable can be established by performing, firstly, the derivative of the flow potential (Eq. 22) with regard to the thermodynamic force associated with damage, Y, and, secondly, with regard to the isotropic hardening variable, r, respectively:

$$\dot{D} \equiv \dot{\gamma}\frac{\partial \Psi}{\partial Y} = \dot{\gamma}\frac{1}{(1-D)}\left(\frac{-Y}{S}\right)^{s} \tag{26}$$

$$\dot{\gamma} \equiv \dot{\gamma}\frac{\partial \Psi}{\partial R} = \dot{\gamma} \tag{27}$$

The complementary law of the rate-independent plasticity also needs to be fulfilled:

$$\dot{\gamma} \geq 0, \Phi \leq 0, \dot{\gamma}\Phi = 0. \tag{28}$$

Bai and Wierzbicki's Model

Bai and Wierzbicki (2008) have recently proposed an elastoplastic model that includes the effect of both the hydrostatic pressure, through the stress triaxiality, and the effect of the third deviatoric stress invariant, through the Lode angle. These effects are introduced in the well-established von Mises elastoplastic model by redefining the hardening rule of the material. While in the classic von Mises model the hardening rule is only a function of the accumulated plastic strain, $\sigma_y(\varepsilon_{eq}^p)$, in Bai and Wierzbicki's model, the hardening rule is a function of the accumulated plastic strain, the stress triaxiality, and the parameter μ, which is a function of the Lode angle, $\sigma_y(\varepsilon_{eq}^p, \eta, \mu)$. Thus, the new definition for the hardening rule can be expressed by

$$\sigma_y\left(\varepsilon_{eq}^p, \eta, \mu\right) = \sigma_y\left(\varepsilon_{eq}^p\right)\left[1 - C_\eta(\eta - \eta_0)\right]\left[C_\theta^s + \left(C_\theta^{ax} - C_\theta^s\right)\left(\mu - \frac{\mu^{m+1}}{m+1}\right)\right] \tag{29}$$

where C_η, C_θ^s, C_θ^{ax}, and m are material constants, η_0 is the reference value for the stress triaxiality, and μ is a Lode angle function that is expressed by

$$\mu = \frac{\cos(\pi/6)}{1 - \cos(\pi/6)}\left[\frac{1}{\cos(\theta - \pi/6)} - 1\right] \tag{30}$$

The effects of the stress triaxiality and Lode angle are included on the hardening rule through the terms $[1 - C_\eta(\eta - \eta_0)]$ and $\left[C_\theta^s + \left(C_\theta^{ax} - C_\theta^s\right)\left(\mu - \frac{\mu^{m+1}}{m+1}\right)\right]$, respectively. The new yield criterion replaces the standard hardening rule $\sigma_y(\varepsilon_{eq}^p)$ by $\sigma_y(\varepsilon_{eq}^p, \eta, \mu)$ on the J_2 theory, such that the yield function can be rewritten as

$$\Phi = q - \sigma_y\left(\varepsilon_{eq}^p\right)AB \tag{31}$$

where q is the von Mises equivalent stress (see section "Constitutive Models for Ductile Failure Under High and Low Triaxiality"), A is the function that incorporates the effect of stress triaxiality, and B is the function which includes the Lode angle effect. They are defined according to

$$A = \left[1 - C_\eta(\eta - \eta_0)\right] \tag{32}$$

$$B = \left[C_\theta^s + \left(C_\theta^{ax} - C_\theta^s\right)\left(\mu - \frac{\mu^{m+1}}{m+1}\right)\right] \tag{33}$$

The influence of the experimental parameters (C_η, C_θ^s, C_θ^{ax}, η_0, m) on the behavior of the constitutive model can be analyzed as follows. The parameter C_η describes the hydrostatic pressure effect on the material plastic flow. If C_η is equal to zero, then the model loses the dependence of the stress triaxiality.

The stress triaxiality reference, η_0, depends on the type of test undertaken and on the geometry of the specimen. It can assume three different values: for a smooth bar subjected to a tensile test, η_0 takes the value equal to 1/3; for a cylindrical specimen subjected to a compression test, η_0 takes the value equal to $-1/3$; and finally for both torsion and shear tests, η_0 takes the value equal to zero. The experimental parameter, C_θ^{ax}, which is related to the Lode angle effect, can assume one of two forms according to the type of loading (tension/compression) applied:

$$C_\theta^{ax} = \begin{cases} C_\theta^t & for \quad \overline{\theta} \geq 0 \\ C_\theta^c & for \quad \overline{\theta} < 0 \end{cases} \tag{34}$$

where $\overline{\theta}$ represents the normalized Lode angle (see Eq. 16). The parameter C_θ^s (Eq. 36) also depends on the type of test. For example, if a smooth bar is used in a tensile test, $C_\theta^t = 0$; if a torsion test is conducted, $C_\theta^s = 1$; and if a cylinder specimen is used in a compressive test $C_\theta^c = 1$. The range of the parameter μ is between $0 \leq \mu \leq 1$. When $\mu = 0$, it corresponds to plane strain or shear condition, and when $\mu = 1$, it corresponds an axisymmetric problem. The introduction of the term $\mu^{m+1}/m + 1$ is done to ensure the smoothness of yield surface and its differentiability with respect to Lode angle around 1.

Due to the fact that Bai and Wierzbicki's model (Bai and Wierzbicki 2008) does not include a damage variable in the constitutive formulation, we will use in our comparisons with the previously described damage models a fracture indicator that was proposed by Bao et al. (2003). This fracture indicator is a post-processed variable, which was developed after conducting a thorough experimental investigation on the behavior of ductile crack formation, expressed by

$$D = \int_0^{\varepsilon_f} \frac{p}{q} d\varepsilon_{eq} = \eta_{av} \varepsilon_{f,eq} \dot{\varepsilon}_v^p \tag{35}$$

where ε_{eq} represents the equivalent strain, $\varepsilon_{f,eq}$ is the equivalent strain to fracture, and η_{av} is the so-called stress triaxiality average. The Lode angle average, θ_{av}, is also a parameter widely used to represent the three-dimensional fracture locus, and both parameters can be expressed by

$$\eta_{av} = \frac{1}{\varepsilon_{f,eq}} \int_0^{\varepsilon_f} \frac{p}{q} d\varepsilon_{eq} \quad \theta_{av} = \frac{1}{\varepsilon_{f,eq}} \int_0^{\varepsilon_f} \theta d\varepsilon_{eq} \tag{36}$$

More details about this fracture indicator can be obtained in references (Bai et al. 2008; Bai and Wierzbicki 2008).

Numerical Implementation and Results

Solution Strategy

Stress update procedures, which are based on the so-called operator split concept (see Simo and Hughes 1998; De Souza Neto et al. 2008), are especially suitable for the numerical integration of the evolution problem and have been widely used in

computational plasticity (e.g., Simo and Hughes 1998; De Souza Neto et al. 2008). This method, which was used in our developments, consists in splitting the problem in two parts: an elastic predictor, where the problem is assumed to be elastic, and a plastic corrector, in which the system of residual equations comprising the elasticity law, plastic consistency, and the rate equations is solved, taking the results of the elastic predictor stage as initial conditions. In the case of violation of the yield condition, the plastic corrector stage has to be initiated and the Newton–Raphson procedure is used to solve the discretized set of equations. The Newton–Raphson procedure was chosen motivated by the quadratic rates of the convergence achieved, which results in computationally efficient return mapping procedures (see Simo and Hughes 1998; De Souza Neto et al. 2008).

Geometry and Calibration

In order to compare, both qualitatively and quantitatively, the constitutive models, based on different levels of stress triaxiality, a butterfly specimen, which was initially proposed by Bai and Wierzbicki (see Bai et al. 2008; Bai and Wierzbicki 2008), was selected. The material properties, the stress–strain curve, and the damage parameters employed by the constitutive models are conveniently listed in Table 1.

The geometry of the specimen, called "butterfly specimen," is illustrated in Fig. 1. A three-dimensional finite element mesh of 3,392 twenty noded elements, with eight Gauss integration points, is used amounting to 17,465 nodes (see Fig. 2).

The strategy employed to determine the undamaged stress–strain curves and the critical damage values for the constitutive models was the following: having at hand the displacement to fracture ($u_f = 6.65$ mm) together with the force–displacement curve for a smooth bar tensile specimen, which were experimentally obtained by Bao and Wierzbicki (2004), an inverse and iterative methodology was conducted. The objective is to identify the stress–strain curve for each constitutive model such that the force–displacement curve is as close as possible to the experimental one. Figure 3a shows the load curves obtained for all the constitutive models after the application of the inverse identification method. It was possible to obtain a close agreement for all constitutive models.

The critical value for the damage variable, of each constitutive model, was also obtained from the simulation of the stretching of the smooth bar. The value of the critical damage variable, of each constitutive model, is set to the value of the internal variable, which is used on the numerical simulation, when the numerical displacement is equal to the experimental displacement to fracture. The critical damage values obtained are listed in Table 2.

The results of the calibration procedure for the stress–strain curves of all models can be observed in Fig. 4. The undamaged stress–strain curve obtained for the Lemaitre model has a more pronounced hardening than the GTN model and both are notably different. It is worth mentioning that the stress–strain curve used in Bai and Wierzbicki's model (Bai and Wierzbicki 2008) which is depicted in Fig. 4 and labelled as "uncoupled damage model" is the curve that includes the effect of damage in the hardening.

Table 1 Material properties for the 2024-T351 aluminum alloy

Description	Symbol	Value	Reference
Density	ρ	2.7×10^3 (Kg/m^3)	Bao and Wierzbicki (2004)
Elastic modulus	E	72.400 (MPa)	Bao and Wierzbicki (2004)
Poisson's ratio	ν	0.33	Bao and Wierzbicki (2004)
Initial yield stress	σ_{y_0}	352 (MPa)	Bao and Wierzbicki (2004)
Damage data (exponent)	s	1	Teng (2008)
Damage data (denominator)	S	6 (MPa)	Teng (2008)
GTN material parameter	q_1	1.5	Xue et al. (2007)
GTN material parameter	q_2	1.0	Xue et al. (2007)
GTN material parameter	q_3	2.25	Xue et al. (2007)
Xue shear mechanism parameter	q_4	1.69 (2D)/1.86 (3D)	Xue et al. (2007)
Xue shear mechanism parameter	q_5	0.50 (2D)/0.33 (3D)	Xue et al. (2007)
Volume fraction of void nucleation	f_N	0.04	Xue et al. (2007)
Stand. dev. of the pl. strain of void nuc.	s_N	0.1	Xue et al. (2007)
Mean plas. strain of the dist. of void nuc.	ε_N	0.2	Xue et al. (2007)
Bai pressure parameter	C_η	0.09	Bai and Wierzbicki (2008)
Triaxiality reference	η_0	0.33	Bai and Wierzbicki (2008)
Bai tensile parameter	C_θ^t	1.0	Bai and Wierzbicki (2008)
Bai compression parameter	C_θ^c	0.9	Bai and Wierzbicki (2008)
Bai shear parameter	C_θ^s	0.855	Bai and Wierzbicki (2008)
Bai exponent parameter	m	6	Bai and Wierzbicki (2008)

It is important to mention that the material properties, the stress–strain curve, and the damage parameters employed by Lemaitre and GTN constitutive models can be obtained from one single experimental test, which is the stretching of a smooth round specimen. On the other hand, the parameters needed by the uncoupled model proposed by Bai and Wierzbicki, which are listed in Table 1, require four types of experimental tests (Bai and Wierzbicki 2008): a smooth round bar tensile test, a notched round bar tensile test, a tensile test of flat grooved plate, and an upsetting test.

It is noteworthy that the determination of the material hardening curves and critical damage parameters was based on the stretching of a smooth bar specimen.

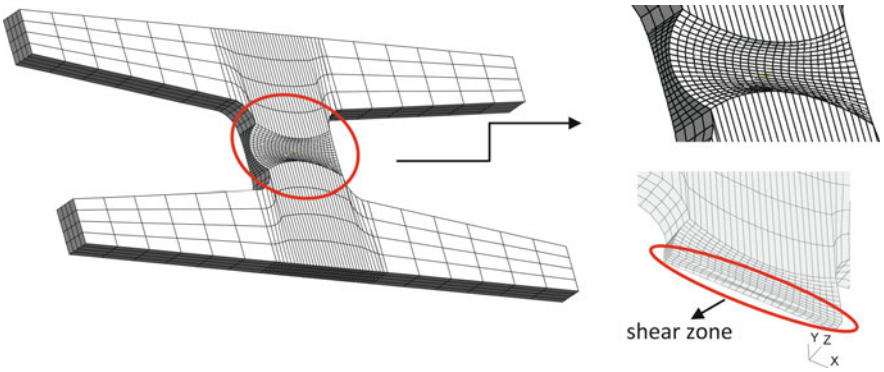


Fig. 2 Finite elements mesh of the butterfly specimen and shear zone to fracture. Reprinted from L. Malcher et al., An Assessment of Isotropic Constitutive Models for Ductile Fracture under High and Low Stress Triaxiality. Int. J. Plast. **30–31**, 81–115 (2012), with permission from Elsevier

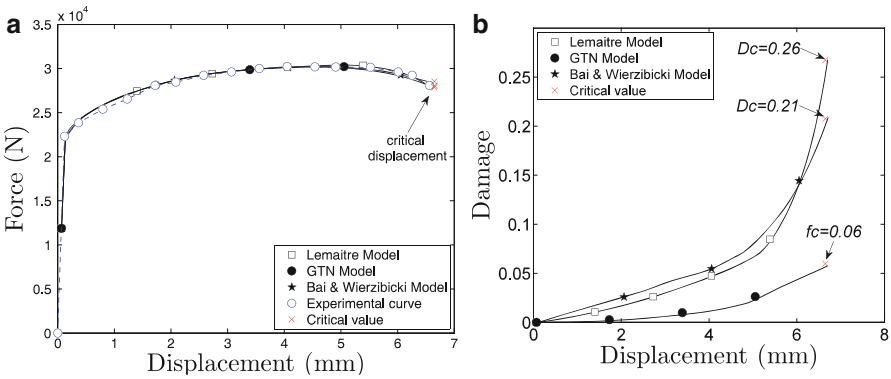


Fig. 3 (**a**) Force versus displacement curve for all models and experimental results. (**b**) Critical damage parameter calibrated for the experimental displacement to fracture ($u_f = 6.65$ mm). Reprinted from L. Malcher et al., An Assessment of Isotropic Constitutive Models for Ductile Fracture under High and Low Stress Triaxiality. Int. J. Plast. **30–31**, 81–115 (2012), with permission from Elsevier

Table 2 Critical values for damage obtained by calibrating the models with a smooth bar specimen

Model	Critical value
Lemaitre	$D_c = 0.26$
GTN	$f_c = 0.06$
Bao	$D_c = 0.21$

Nevertheless, different results would be obtained from the inverse method if the calibration were carried out based on a different specimen. The authors have used this calibration point because it is the most widely used calibration point in the literature for damage models. The impact that the change on calibration strategy

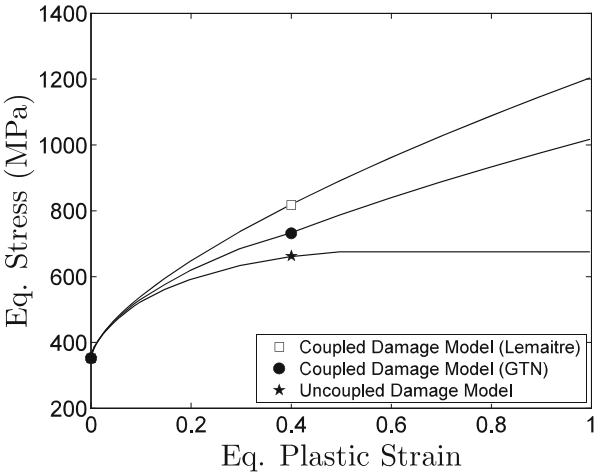


Fig. 4 Stress–strain curves for all constitutive models. Reprinted from L. Malcher et al., An Assessment of Isotropic Constitutive Models for Ductile Fracture under High and Low Stress Triaxiality. Int. J. Plast. **30–31**, 81–115 (2012), with permission from Elsevier

Table 3 Reference values for different specimens of the 2024-T351 (Bao et al. 2003; Bai et al. 2008)

Specimen	u_f (mm)	η_0	$\overline{\theta_0}$	η_{av}	θ_{av}	$\varepsilon_{f,eq}$	Fracture location
Butterfly (pure shear)	–	0	0	0.01	0.04	0.21	Surface of shear zone
Butterfly (tension/shear, 10°)	–	0.11	0.22	0.12	0.34	0.26	Middle of thickness on shear zone

would have on the predicted ability of the models is a topic of research that deserves further analysis.

Numerical Results

The predictive ability of the constitutive models is assessed here using the properties listed in Table 1 together with the calibrated parameters. A summary of the results that will serve as reference for comparison is presented in Table 3. In particular, the displacement at fracture, u_f; the initial stress triaxiality, η_0; the initial Lode angle, $\overline{\theta_0}$; the averaged stress triaxiality, η_{av}; the averaged Lode angle, θ_{av}; and the equivalent strain at fracture, $\varepsilon_{f,eq}$, are listed for each specimen. The displacements to fracture of the butterfly specimen, in both pure shear and combined tensile/shear loading conditions, were not available in the literature. The expected location for crack formation, experimentally observed, is also included. The information presented in Table 3 was obtained from Bao (Bao et al. 2003) and Bai (Bai et al. 2008).

Before proceeding, it is important to describe how the reference values, which are listed in Table 3, were obtained. The initial stress triaxiality, η_0, and initial Lode angle, $\overline{\theta_0}$, values are obtained from analytical expressions, which can be derived for

Table 4 A summary of the numerical results obtained by the damage constitutive models studied on aluminum alloy 2024-T351. Specimens subjected to a low level of stress triaxiality

Specimen	Model	u_f (mm)	$\eta_{av}{}^{a}$	η_{av}	$\theta_{av}{}^{a}$	θ_{av}	$\varepsilon^{a}_{f,eq}$	$\varepsilon^{p}_{f,eq}$	Fracture location
Butterfly (pure shear)	Bai and Wierzbicki	0.700	0.01	0.00	0.04	0.00	0.21	1.40	Surface of shear zone
	Lemaitre	0.464		0.08		0.04		0.64	
	GTN original	–		0.02		0.06		–	
	GTN modified	0.348		0.02		0.04		0.31	
Butterfly (tensile/shear, 10°)	Bai and Wierzbicki	0.540	0.12	0.22	0.34	0.43	0.26	0.67	Middle of thickness on shear zone
	Lemaitre	0.408		0.34		0.19		0.60	
	GTN original	0.642		0.30		0.47		0.64	
	GTN modified	0.340		0.27		0.43		0.35	

[a]Represents reference values

each particular specimen (Bai et al. 2008). The determination of the averaged stress triaxiality, η_{av}; averaged Lode angle, θ_{av}; and equivalent strain at fracture, $\varepsilon_{f,eq}$, was done using a combined experimental numerical method. After performing the experiments, numerical simulations of all specimens were conducted by Bao (Bao et al. 2003) and also by Bai (Bai et al. 2008) using a von Mises model. The averaged stress triaxiality, η_{av}; averaged Lode angle, θ_{av}; and equivalent strain at fracture, $\varepsilon_{f,eq}$, are calculated by the finite element simulation at the critical location for the measured displacement at fracture. Since the stress triaxiality and Lode angle parameters are variable during the loading process, Bao (Bao et al. 2003) and Bai (Bai et al. 2008) used averaged values according to Eq. 36.

All the numerical results obtained in this work were conducted following the same strategy. The simulation was performed until the damage variable of the particular constitutive model, at any point in the specimen, reached the critical value that is listed in Table 2. Therefore, the value of the displacement and effective plastic strain variables calculated from the finite element simulation, when the damage variable reaches its critical value, are understood as the displacement at fracture and effective strain at fracture from the numerical simulations. After attaining the critical damage value at a particular element, the finite element analysis is conducted as before without changing the fully damaged element. The focus here is to study the behavior of the constitutive models, previously described, under a low level of stress triaxiality and verify their ability to predict the correct fracture location. The butterfly specimen was simulated in both pure shear and combined tension/shear (10° with the x-axis) conditions. In Table 4, the numerical results, obtained by the finite element simulation, for the displacement at fracture, stress triaxiality average, Lode angle average, and equivalent plastic strain at fracture can be examined. In both loading scenarios, the prescribed displacement

was imposed until the damage variable of the particular constitutive model, at any point in the specimen, reached its critical value, previously calibrated (see Table 2). The value of the displacement and effective plastic strain variables calculated from the finite element simulation, when the damage variable reaches its critical value, are understood as the displacement at fracture and effective strain at fracture from the numerical simulations.

The results obtained with the original GTN model, under shear-dominated loading conditions, clearly emphasize the limitation of the model for predicting fracture under conditions of low stress triaxiality. In Table 4, it is possible to see that according to this model, the critical damage value would never be reached for pure shear stress states. Under combined tension/shear (10° with the x-axis), the predicted displacement to fracture is very high, when compared with the other models, since damage evolution is only due to the volumetric growth of voids. The displacements to fracture predicted by the Lemaitre model, $u_f = 0.464$ mm, and the GTN modified model, $u_f = 0.348$ mm, are more or less close for pure shear. There is a slightly better agreement between the predicted displacements to fracture for these two models for combined tension/shear stress states: the Lemaitre model predicts $u_f = 0.408$ mm, and the GTN modified model predicts $u_f = 0.34$ mm. Nevertheless, there is a marked difference between the levels of predicted equivalent plastic strain obtained with the Lemaitre model and the GTN modified model for both loading conditions, which are clearly different from the reference value of the effective plastic strain listed in Table 3. Through the analysis of the results obtained by Bai and Wierzbicki's model coupled with Bao's fracture indicator (see Table 4), it is possible to conclude that the overall prediction is not satisfactory. In particular, for pure shear loading conditions, the model predicts for both parameters, displacement and equivalent plastic strain at fracture, very high values ($u_f = 0.7$ mm; $\varepsilon^p_{f,eq} = 1.4$) that are different from the reference values listed in Table 3. These results clearly suggest that Bao's damage fracture indicator coupled with Bai and Wierzbicki's model might not be a good parameter to predict fracture under low level of stress triaxiality.

The evolution of the damage parameter, at the point where the damage variable reaches the maximum value, can be examined in Fig. 5. The critical damage value for each model, which is listed in Table 2, is reached at different levels of displacement.

The evolution of the damage variable for the GTN original model under shear loading illustrates its limitation to predict shear localization and fracture under conditions of low triaxiality. After an initial increase of the damage variable, which is due to void nucleation, there is no further evolution of damage. Under a combined shear/tensile loading, this model predicts damage evolution. Nevertheless, since this growth is only due to volume void growth, the overall damage evolution is slow and the model predicts a high value for the displacement to fracture, which is not in agreement with experimental evidence. The inclusion of shear effects on the formulation of the model, here labelled modified GTN model, clearly improves the ability of the model to predict damage growth under shear and combined shear/tensile (10° with the x-axis)

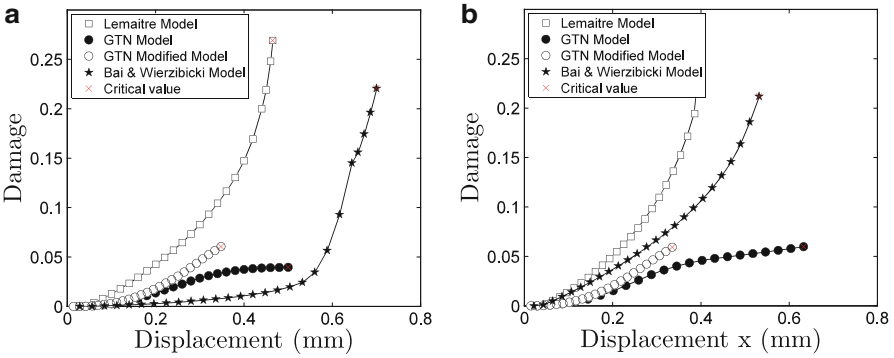


Fig. 5 Evolution of damage in (**a**) pure shear and (**b**) combined shear/tensile loading conditions. Reprinted from L. Malcher et al., An Assessment of Isotropic Constitutive Models for Ductile Fracture under High and Low Stress Triaxiality. Int. J. Plast. **30–31**, 81–115 (2012), with permission from Elsevier

loading conditions since the distortion of voids and inter-void linking are taken into account in the model (see Fig. 5a and b). It is important to observe that the Lemaitre model can predict the evolution of damage under conditions of low stress triaxiality. In addition, Bao's damage fracture indicator coupled with Bai and Wierzbicki's model is also able to predict the evolution of damage.

The damage distribution for each constitutive model, when the critical damage is attained, can be seen on Fig. 6 for pure shear loading. Experimental evidence has shown that the potential zone for crack formation occurs on the surface of the shear zone. Both the Lemaitre and GTN modified models, depicted in Fig. 6b and d, have been able to predict the correct location of fracture onset. On the other hand, Bao's damage fracture indicator coupled with Bai and Wierzbicki's model, depicted in Fig. 6a, has predicted fracture at the middle of the thickness on the critical zone, which is wrong. The original GTN model predicts damage over the central region of the critical zone, never reaching the critical value (see Fig. 6c).

The damage variable field obtained in the numerical simulation, for a combined tensile/shear loading condition, is illustrated by the contour plots shown in Fig. 7. For this loading condition, fracture onset is experimentally observed at the center of the shear zone. Therefore, from the analysis of Fig. 7, it is possible to conclude that Bao's damage fracture indicator coupled with Bai and Wierzbicki's model is able to predict the correct fracture location. The same happens with the GTN original model that also predicts fracture onset at the center of the specimen. However, the damage evolution for these two models is relatively slow and consequently they predict a large displacement at fracture. In contrast, the Lemaitre model and the GTN modified model have predicted fracture onset at the surface of the critical zone, which is in disagreement with experimental evidence. Therefore, these two models have not been able to predict the correct location of fracture under combined tensile/shear loading conditions.

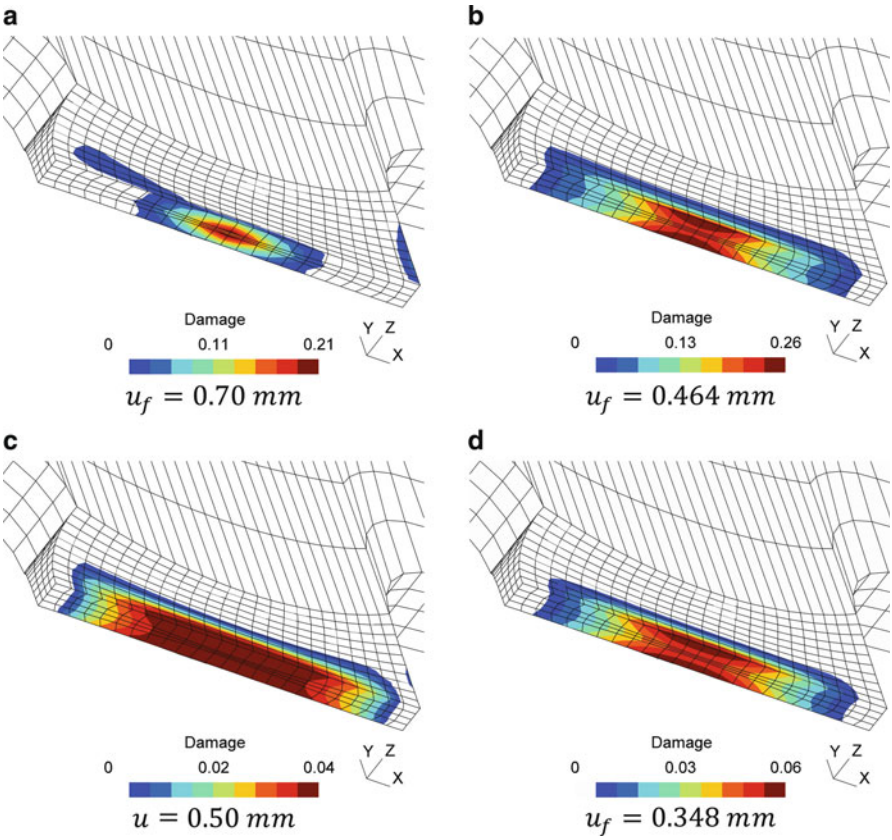


Fig. 6 Damage contours for the butterfly specimen under pure shear conditions. (**a**) Bai and Wierzbicki's model, (**b**) Lemaitre model, (**c**) GTN model, and (**d**) GTN modified model. Reprinted from L. Malcher et al., An Assessment of Isotropic Constitutive Models for Ductile Fracture under High and Low Stress Triaxiality. Int. J. Plast. **30–31**, 81–115 (2012), with permission from Elsevier

The inclusion of shear effects on the GTN model has a significant impact on the evolution of the equivalent plastic strain. Due to the strong coupling between plastic flow and damage, which exists on the modified GTN model, an increase of overall damage due to the combination of void growth with the distortion of voids leads to an increase of the equivalent plastic strain. This enhances the model that predicts a level of equivalent plastic strain at fracture close to the expected value. In order to discuss the predictive ability of the Lemaitre model, different values for the critical damage D_c were critically selected and the damage variable field obtained from the numerical simulation is illustrated by the contour plots shown in Fig. 8. It is important to remark that this is merely an exercise and the authors have not performed any additional calibration procedure.

From the analysis of the results depicted in Fig. 8, it is possible to conclude that if the critical value of damage is increased, the location of fracture onset moves

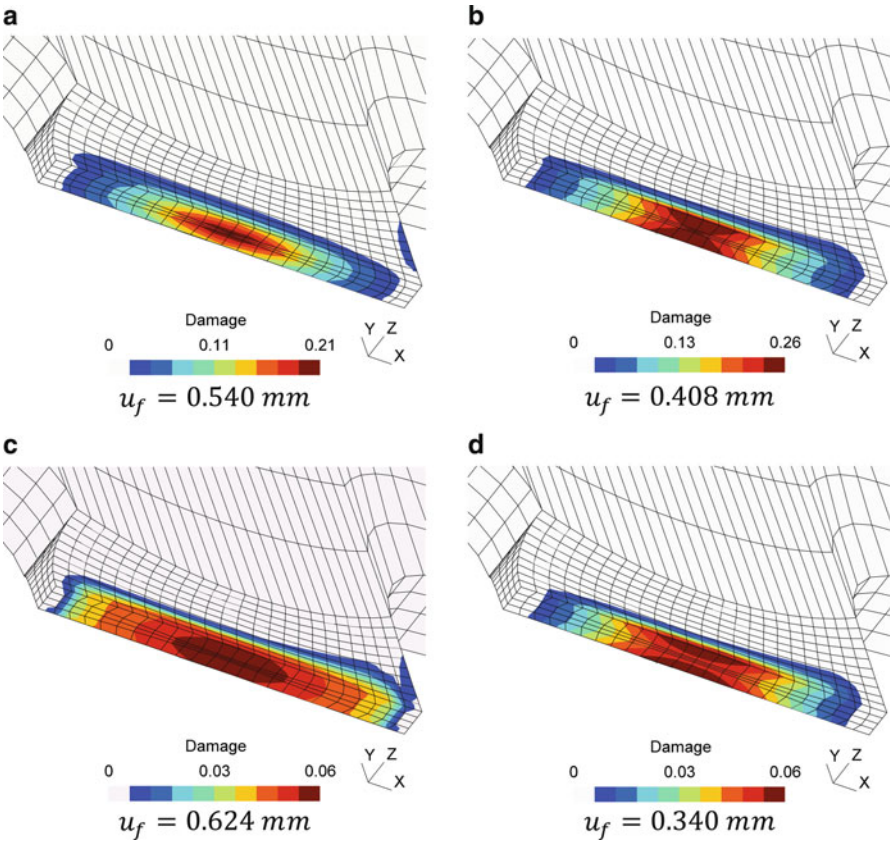


Fig. 7 Damage contours for a butterfly specimen under combined shear/tensile load. (**a**) Bai and Wierzbicki, (**b**) Lemaitre (**c**) GTN original, and (**d**) GTN modified models. Reprinted from L. Malcher et al., An Assessment of Isotropic Constitutive Models for Ductile Fracture under High and Low Stress Triaxiality. Int. J. Plast. **30–31**, 81–115 (2012), with permission from Elsevier

from the surface of the shear zone to the center of the shear zone. Therefore, for a high value of critical damage, $D_c = 0.50$, the prediction of fracture onset of the Lemaitre model would be in agreement with experimental observations.

Nonlocal Models

Introduction

In local models, the behavior of the material is completely represented by a point-wise constitutive law, which is independent of the influence of surrounding material points assuming that the material is continuous at any scale neglecting any size effects.

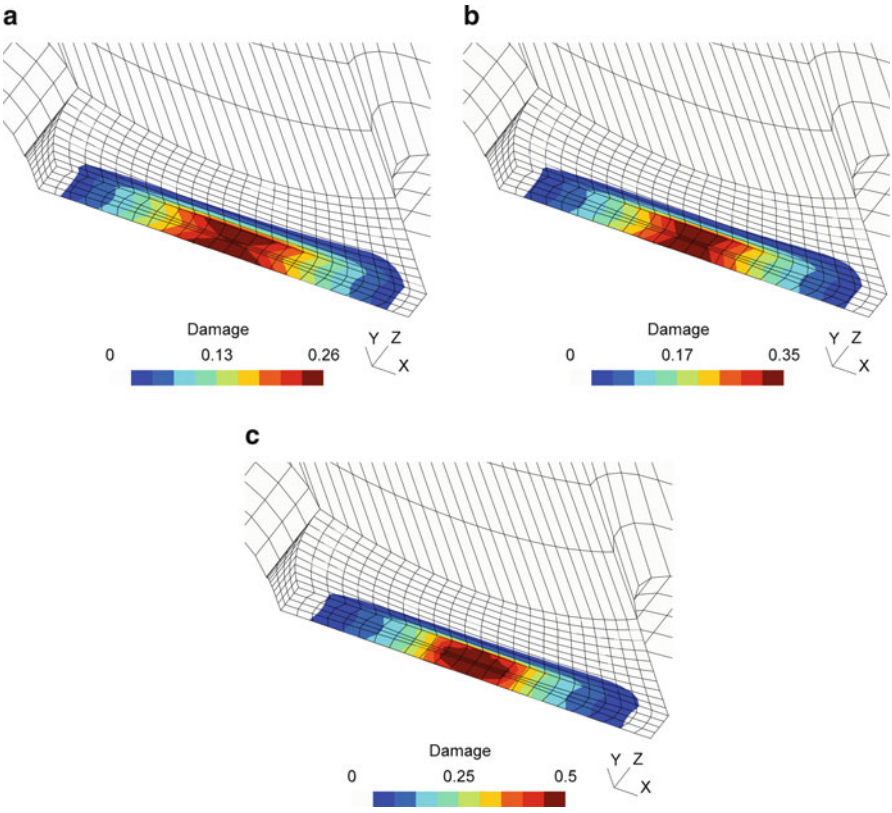


Fig. 8 Lemaitre's damage contour for the butterfly specimen under combined shear/tensile loading condition. (**a**) $D_c = 0.26$, (**b**) $D_c = 0.35$, and (**c**) $D_c = 0.50$. Reprinted from L. Malcher et al., An Assessment of Isotropic Constitutive Models for Ductile Fracture under High and Low Stress Triaxiality. Int. J. Plast. **30–31**, 81–115 (2012), with permission from Elsevier

Alternatively the nonlocal theory incorporates an intrinsic length into the traditional continuum theory, trying to mimic those size effects at the constitutive level and as a "side effect." If conveniently formulated, they alleviate or solve numerical problems associated with local models, by means of either gradient-enhanced or integral-type formulations.

In fact nonlocal models have been successfully used in the last two decades to avoid or alleviate pathological geometrical discretization dependency of numerical solutions. This problem is prone to occur whenever local strain-softening is triggered by a particular constitutive model. It may be explained, for example, in a quasi-static problem, by the fact that locally the underlying partial differential equilibrium equations lose ellipticity, "bringing to surface" associated numerical problems.

Since the pioneering work of Pijaudier-Cabot and Bazant (Pijaudier-Cabot and Bažant 1987), who applied the nonlocal formulation of Edelen and Laws (1971;

Edelen et al. 1971) in the context of an elastic–damage model, many reliable and efficient algorithms for the implementation of the nonlocal theory have been put forward both in an integral or differential counterparts (De Borst and Mühlhaus 1992; Strömberg and Ristinmaa 1996; Peerlings et al. 1996; Benvenuti and Tralli 2003; Engelen et al. 2003; Cesar de Sa et al. 2006; Polizzotto 2009; Voyiadjis et al. 2010; Andrade et al. 2011). It is noteworthy that both formulations give, in general, equivalent solutions in similar circumstances.

The derivation of any nonlocal theory requires the choice of the variable or variables to be enhanced by nonlocality. Typical choices are, among others, the regularization of variables related to kinematics (such as the strain tensor), regularization of internal variables (e.g., scalar measurements of the amount of plastic strain or damage), and regularization of thermodynamic forces power-conjugated with internal variables (for instance, the elastic energy release rate in damage models). In fact, the choice of the nonlocal variable depends on the kind of material to be modelled and on the nature of the problem to be solved. In the particular case of elastoplastic damaging ductile solids, the internal degradation of the material, which in the CDM theory is usually treated by means of some damage measurement as an internal variable, is closely linked to the localization phenomenon. Therefore, the choice of damage as the nonlocal variable emerges as natural in this context and will be chosen here, using an integral approach.

A nonlocal variable can be expressed from its local counterpart, in the integral-type formulation, by means of the spatially weighted averaging integral as in (Eq. 37):

$$\overline{\varphi}(x) = \int_V \beta(\boldsymbol{x}, \boldsymbol{\xi}) \varphi(\boldsymbol{\xi}) dV(\boldsymbol{\xi}) \tag{37}$$

where φ and $\overline{\varphi}$ are, respectively, the local and nonlocal damage variables and $\beta(\boldsymbol{x}, \boldsymbol{\xi})$ is a weighted averaging operator. It is noteworthy that $\overline{\varphi}(\mathbf{x})$ represents the measurement of nonlocal variable at a material point $\boldsymbol{x}$, which has been averaged over a finite volume V, whose size is somewhat related to a constitutive parameter, the intrinsic length ℓ_r.

The averaging operator, $\beta(\boldsymbol{x}, \boldsymbol{\xi})$, satisfies the normalizing condition

$$\int_V \beta(\boldsymbol{x}, \boldsymbol{\xi}) dV(\boldsymbol{\xi}) = 1 \tag{38}$$

The normalization may be accomplished by defining the operator as

$$\beta(\boldsymbol{x}, \boldsymbol{\xi}) = \frac{\alpha(\boldsymbol{x}, \boldsymbol{\xi})}{\Omega_r(\boldsymbol{x})} \tag{39}$$

where $\Omega_r(\boldsymbol{x})$ is a representative volume given by

$$\Omega_r(\boldsymbol{x}) = \int_V \alpha(\boldsymbol{x}, \boldsymbol{\xi}) dV(\boldsymbol{\xi}) \tag{40}$$

and $\alpha(x, \xi)$ is a prescribed weighting function which depends only on the distance r.

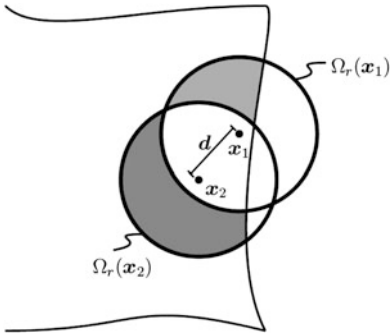


Fig. 9 Schematic illustration of the lack of symmetry of the nonsymmetric averaging operator. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

This weighting function should satisfy some basic characteristics in order to obtain the diffusive effect expected from the nonlocal theory, i.e., it should have its maximum at the origin and the decrease as the distance of neighboring points increases. For that purpose we adopt here the bell-shaped function, which is frequently employed in nonlocal theories of integral-type, as

$$\alpha(\boldsymbol{x},\boldsymbol{\xi}) = \left\langle 1 - \frac{\|\boldsymbol{x}-\boldsymbol{\xi}\|^2}{\ell_r^2(\boldsymbol{x})} \right\rangle^2 \tag{41}$$

where $\langle\cdot\rangle$ are the Macaulay brackets.

For a material point sufficiently far from the boundary $\Omega_r(\boldsymbol{x})$ is assumed to be constant and is usually denoted by Ω_∞. In practice this representative volume is computed as the area of a circle (2D problems) or the volume of a sphere (3D problems). As illustrated in Fig. 9, near the boundaries, because the size of the representative volume may vary, the averaging operator loses its symmetry, i.e., $\beta(\boldsymbol{x}, \boldsymbol{\xi}) \neq \beta(\boldsymbol{\xi}, \boldsymbol{x})$ due to the fact that

$$\frac{\alpha(\boldsymbol{x}_1,\boldsymbol{x}_2)}{\Omega_r(\boldsymbol{x}_1)} \neq \frac{\alpha(\boldsymbol{x}_1,\boldsymbol{x}_2)}{\Omega_r(\boldsymbol{x}_1)} \tag{42}$$

The use of a nonsymmetric operator is disadvantageous from a purely computational point of view due to the additional amount of memory necessary for storage when compared to the case of a symmetric operator (Andrade 2011).

Besides, the definition and use of a symmetric averaging operator are also appealing for the concise theoretical elaboration of nonlocal models which are consistent with thermodynamics (Borino et al. 1999). For this purpose, Borino et al. (1999) have proposed a symmetric averaging operator, which is written as

$$\beta(\boldsymbol{x},\boldsymbol{\xi}) = \left(1 - \frac{\Omega_r(x)}{\Omega_\infty}\right)\delta(\boldsymbol{x},\boldsymbol{\xi}) + \frac{1}{\Omega_\infty}\alpha(\boldsymbol{x},\boldsymbol{\xi}) \tag{43}$$

where $\delta(\boldsymbol{x}, \boldsymbol{\xi})$ is the Dirac delta function. With other authors and for the sake of clarity, we adopt the following notation to define a nonlocal variable:

$$\mathcal{R}(\varphi) = \overline{\varphi}(\boldsymbol{x}) = \int_V \beta(\boldsymbol{x}, \boldsymbol{\xi}) \varphi(\boldsymbol{\xi}) dV(\boldsymbol{\xi}) \tag{44}$$

where $\mathcal{R}(\cdot)$ denotes the integral regularization operator.

The first step for the nonlocal enhancement of a previously existing local material model is to choose where the integral operator should be incorporated in the constitutive equations. This step can be interpreted, to some extent, as "the choice of the nonlocal variable."

Since every single variable of a given constitutive model can be enhanced by nonlocality, a large number of possibilities seem to exist. Of course, some options are at first glance more promising than others; however, a simple guess does not guarantee a constitutive model free of spurious mesh dependence. Among the limited number of contributions who worked in this direction, Bazant and Chang (Chang and Bazant 1984) have concluded that the internal variables are the best candidates for nonlocal quantities when compared to other quantities such as the stress and the total strain tensors. However, which internal variable should be selected among the many possibilities is, to a larger extent, uncertain for most models. Other important contributions on the issue of the choice of the nonlocal variable are due to Jirasek (1998) and Jirasek and Rolshoven (2003), who assessed many different nonlocal formulations in order to verify their regularizing properties, as mentioned above. Several options have been considered and tested here for both Lemaitre- and Gurson-based damage models.

Nonlocal Formulation of a Lemaitre-Based Damage Model

The Lemaitre-based damage model, as described above, is firmly rooted in the Continuum Damage Mechanics and the thermodynamics of irreversible processes. Its extension to include the derivation of the nonlocal constitutive is possible (Andrade 2011). Nevertheless, most nonlocal theories are formulated within a classical approach, i.e., they are merely ad hoc extensions of previously existing local constitutive models for which one or more local variables are replaced by their nonlocal counterparts. Although less theoretically appealing, the classical version of the present nonlocal damage model is thermodynamically admissible, i.e., it does not generate negative dissipations (Andrade 2011). Furthermore it has been proven that in practical terms their implementation leads to very similar results (Andrade 2011) and therefore the classical approach in Cesar de Sa et al. (2010) was here favored as its implementation is substantially simplified.

For the case of Lemaitre-based models, the following four options as potential nonlocal variables have been considered:

- Regularization of damage, D
- Regularization of the isotropic hardening variable, R
- Regularization of elastic energy release rate, Y
- Simultaneous regularization of damage, D and the isotropic hardening variable, R

Table 5 Lemaitre-based nonlocal models

Associated variable	Symbol	Evolution equation	Reference
Damage	D	$\dot{\bar{D}} = \mathcal{R}(\dot{D})$	L-D
Isotropic hardening	R	$\dot{\bar{R}} = \mathcal{R}(\dot{R})$	L-R
Energy release rate	Y	$\dot{D} = \frac{\dot{\gamma}}{(1-D)} \left(\frac{\mathcal{R}\{-Y\}}{S}\right)^s$	L-Y
Damage, isotropic hardening	D, R	$\dot{\bar{D}} = \mathcal{R}\left\{\frac{\dot{\gamma}}{(1-D)} \left(\frac{-Y}{S}\right)^s\right\}$	L-DR
		$\dot{\bar{R}} = \mathcal{R}(\dot{R})$	

The choices listed above modify the Lemaitre local model in such manner that for each case one (or more) evolution equation of the local model needs to be modified. The necessary modifications in the constitutive model are directly summarized in Table 5.

Nonlocal Formulation of a Gurson-Based Damage Model

Some contributions have tried to tackle the issue of mesh dependence on Gurson-based constitutive models. For instance, Tvergaard and Needleman (1995) have proposed a nonlocal model of integral type by regularizing the porosity variable. In their approach, they have avoided the definition of an "exact" nonlocal formulation by only approximating the nonlocal rate equation. This approach has also been adopted by Feucht (1999) who has also enriched the Gurson model with a gradient-dependent theory. Reusch et al. (2003a) have also employed a gradient-enhanced nonlocal formulation where a new damage variable, which was related to the local porosity through a gradient equation, has been incorporated in the yield function proposed by Tvergaard and Needleman (1984). The model was later extended to finite strains (Reusch et al. 2003b). Hakansson et al. (2006) provided a thermomechanical constitutive theory for porous materials by including an additional equilibrium equation containing the gradient of the porosity variable. In their formulation, the local mechanical, the thermal, and the nonlocal problems were solved in an uncoupled fashion. Enakousta et al. (2007) have considered a nonlocal porosity rate, defined through an averaging integral, which was explicitly integrated. In their numerical strategy, the nonlocal porosity was computed only after the material problem was solved for a fixed value of porosity.

More recently, Samal et al. (2008) have enhanced Rousselier's constitutive model by adopting a similar approach to the one proposed by Reusch et al. (2003a), i.e., a new damage variable has been incorporated through a gradient implicit formulation.

However, none of the aforementioned contributions has focused in developing a Gurson-based nonlocal model within a "full" integral framework. The integral-type nonlocal formulation has the advantage of being completely defined on the material level, avoiding the definition of additional structural variables in the global system of equations. In addition, many of the advantages of the constitutive modelling at

Table 6 Gurson-based nonlocal models

Associated variable	Symbol	Evolution equation	Reference
Void volume fraction	$\overline{f}$	$\dot{\overline{f}} = \mathcal{R}\left(\dot{f}^N + \dot{f}^G + \dot{f}^{\text{Shear}}\right)$	G-F
Isotropic hardening	$\overline{R}$	$\dot{\overline{R}} = \mathcal{R}(\dot{R})$	G-R
Equivalent plastic strain	$\overline{\varepsilon}^p_{eq}$	$\dot{f}^N = \frac{f_N}{s_N\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\varepsilon^p_{eq}-\varepsilon_N}{s_N}\right)^2\right]\mathcal{R}\left(\dot{\varepsilon}^p_{eq}\right)$	G-EP
Void volume fraction, isotropic hardening	$\overline{f}, \overline{R}$	$\dot{\overline{f}} = \mathcal{R}\left(\dot{f}^N + \dot{f}^G + \dot{f}^{\text{Shear}}\right)$ $\dot{\overline{R}} = \mathcal{R}(\dot{R})$	G-FR

finite strains within the local framework can be straightforward extended to the nonlocal case.

Similar to the case of the Lemaitre model, four candidate variables have been chosen for the Gurson-based damage model described as follows:

- Regularization of damage, i.e., void volume fraction, f
- Regularization of the isotropic hardening variable, R
- Regularization of the equivalent plastic strain, $\overline{\varepsilon}^p_{eq}$
- Simultaneous regularization of void volume fraction, f, and the isotropic hardening variable, R

Again, each choice requires modifications in one or more evolution equations that are associated with the chosen constitutive variable, where the necessary modifications are summarized in Table 6 for convenience.

Assessment and Comparisons of Nonlocal Models

In this section different choices for nonlocal variables are numerically assessed for different combinations of stress triaxiality ratio, η, and normalized third invariant, ξ. As pointed out in section "Constitutive Models for Ductile Failure Under High and Low Triaxiality," ductile materials may have very different behavior upon fracture, with less or more dependency on those parameters. Certain materials exhibit more strain-driven softening than others, so the need of nonlocal modelling is, in quantitative terms, dependent on the specific material whose behavior one aims to simulate. Our intention is to numerically illustrate the performance of the different nonlocal models presented rather than reproducing experimental evidence with accuracy. Therefore, we will adopt generic material properties that very much resemble those typically observed and calibrated from experimental testing of high strength steel alloys (see Tables 7, 8, and 9). The same material properties will be used in all simulations for consistency.

Three different mesh refinements will be used in each case, for which the following line pattern has been adopted in the XY plots: the coarse, medium, and fine meshes are, respectively, represented by a full (−), a dashed (− −), and a dotted (· · ·) line.

Table 7 Basic material properties employed in all simulations

Property	Value
Elastic modulus	$E = 220$ Gpa
Poisson's ratio	$\nu = 0.3$
Hardening function	$\tau_y(R) = 700 + 300R^{0.3}$ MPa

Table 8 Lemaitre-related material properties

Property	Value
Lemaitre damage exponent	$s = 1.0$
Lemaitre damage denominator	$S = 3.0$ MPa

Table 9 Gurson-related material properties

Property	Value
Micro-void volumetric fraction for nucleation	$f_N = 0.04$
Mean strain for void nucleation	$\varepsilon_N = 0.2$
Standard deviatoric strain for nucleation	$s_N = 0.04$
Critical damage	$f_c = 0.06$
Damage at fracture	$f_c = 0.22$
Shear-damage factor	$k = 3.0$

Analysis at High Stress Triaxiality

The predominant internal degradation mechanism for ductile materials at high triaxialities, $\eta > 1/3$, is governed by the phenomena of nucleation, growth, and coalescence of micro-voids.

To a major extent, the Lemaitre and Gurson ductile damage models have been particularly conceived to capture the aforementioned phenomena. Indeed, this fact was the main motivation for the inclusion of the additional damage mechanism in the Gurson model in section "Constitutive Models for Ductile Failure Under High and Low Triaxiality," since without the adopted modification no damage evolution would take place in shear-dominated stress states with that model. In the case of Lemaitre, damage does evolve when subjected to both high and low stress triaxialities.

For the analysis at high triaxialities, two specimens will be considered in the present assessment: an axisymmetric notched bar and a grooved plate under plane strain condition (see Figs. 10, 11, and 12).

The mean values of triaxiality are, in both cases, very similar ($\eta \cong 0.8$ for the axisymmetric case and $\eta \cong 0.7$ for the plane strain case). On the contrary, the mean values of the normalized third invariant are quite different ($\xi = 1$ for the axisymmetric case and $\xi = 0$ for the plane strain case) characterizing different stress states.

Three mesh refinements have been considered in order to capture the effects of mesh dependency where only one quarter of the geometry has been simulated in both cases. Different nonlocal intrinsic lengths have been assigned so that the minimum necessary of elements (and their associated integration points) were spanned in order to activate the effects of the nonlocal formulation ($\ell_r = 0.6$ mm for the axisymmetric case and $\ell_r = 0.35$ mm for the plane strain case).

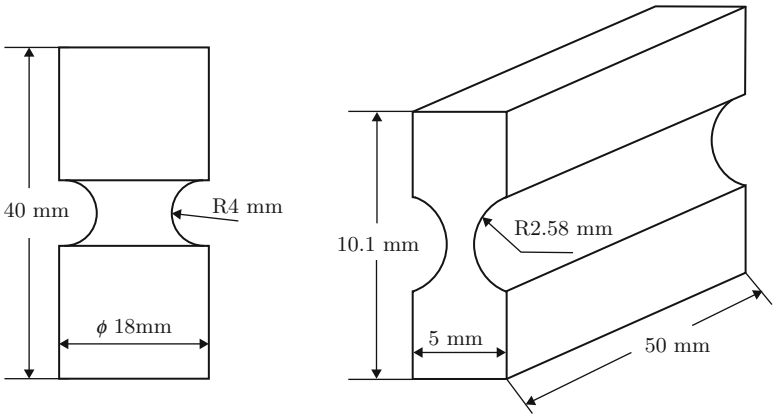


Fig. 10 Geometry for the axisymmetric (*left*) and plane strain (*right*) specimens. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

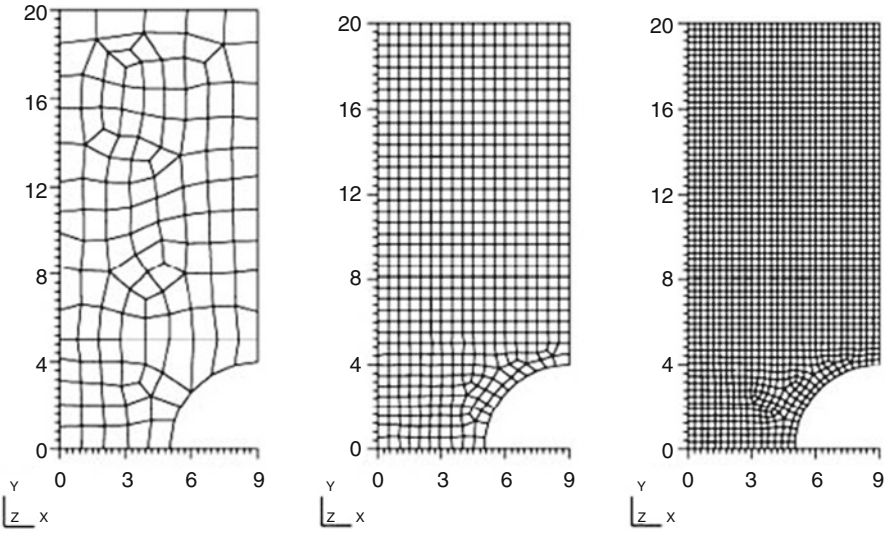


Fig. 11 Geometry for the axisymmetric (*left*) and plane strain (*right*) specimens. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

Despite the fact that the underlying assumptions are very different for both ductile damage models, mesh dependency has taken place in both cases in the local case, albeit more pronounced in the plane strain case, as it can be observed when the maximum value of damage evolution is plotted against the applied displacement (see Fig. 13). Although those curves may not reveal a strong mesh

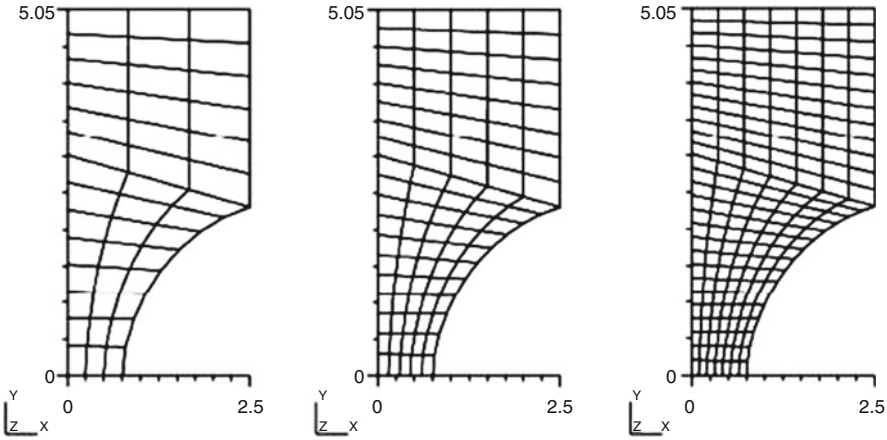


Fig. 12 Mesh refinement for the plane strain specimen. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

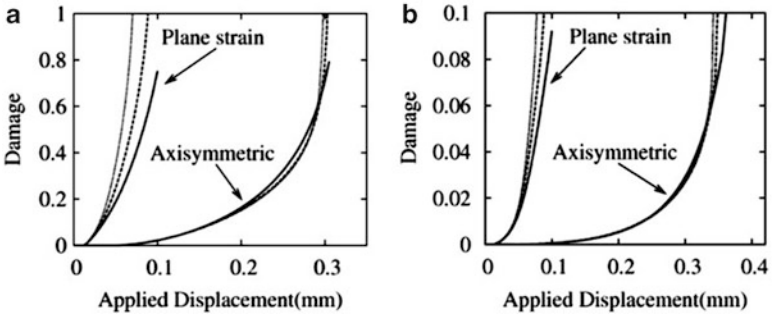


Fig. 13 Evolution of damage in the local case. Results for the (**a**) Lemaitre- and (**b**) Gurson-based models. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

dependency for the axisymmetric case, the damage values and contours depicted in Figs. 14 and 15 clearly reveal it. In the plane strain case, the damage contours show a more evident mesh dependency (see Figs. 16 and 17).

The evolution of damage for several nonlocal Lemaitre-based formulations is given in Fig. 18. For the axisymmetric case, the formulations L-D, L-Y, and L-DR have provided mesh-independent solutions, but for the plane strain solution, only L-D and L-DR have eliminated the pathological mesh dependency. The formulations L-Y and L-R are far from being adequate solutions for nonlocal enhancement of the Lemaitre-based model. In particular, the L-Y has provided slightly different results when compared to the local case for the plane strain specimen. In fact, a

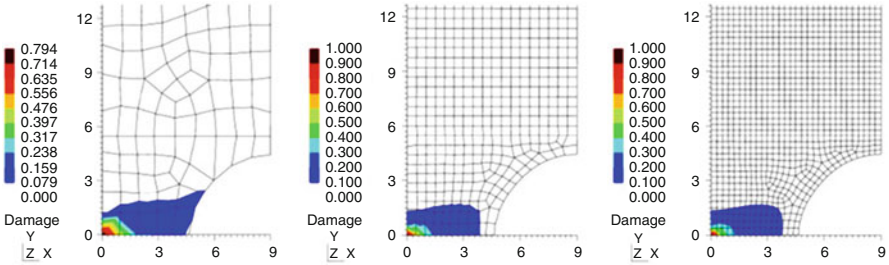


Fig. 14 Damage contours for the Lemaitre-based model in the local case. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

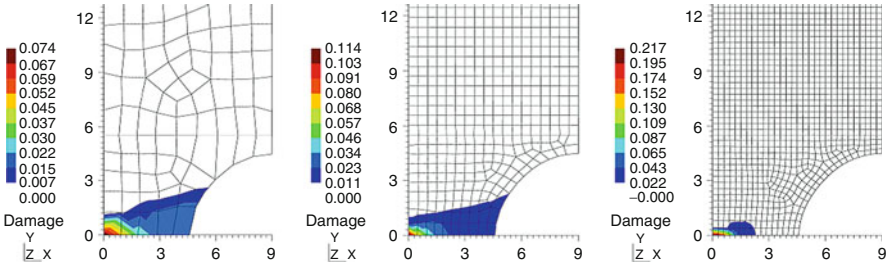


Fig. 15 Damage contours for the Gurson-based model in the local case. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

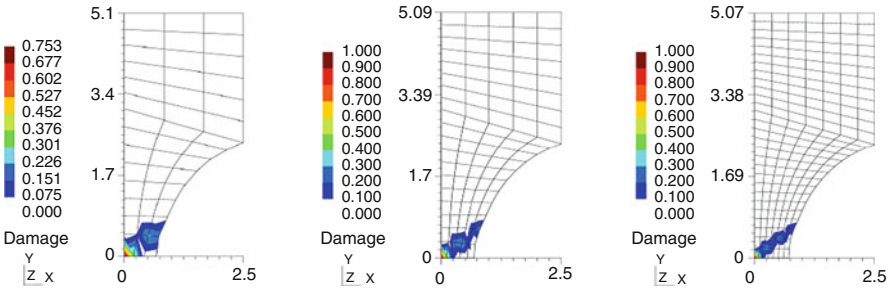


Fig. 16 Damage contours for the Lemaitre-based model in the local case. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

certain tendency to alleviate the pathological mesh dependency was observed. However, such tendency was substantially small, not enough to eliminate the issues of spurious mesh dependency. Further numerical experimentation has demonstrated that by increasing significantly the characteristic length, the solution tended to

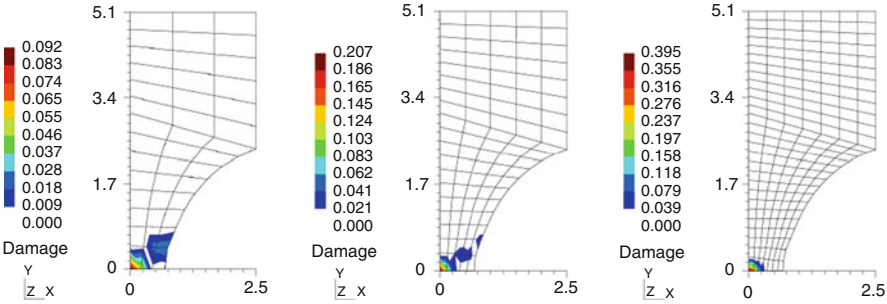


Fig. 17 Damage contours for the Gurson-based model in the local case. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

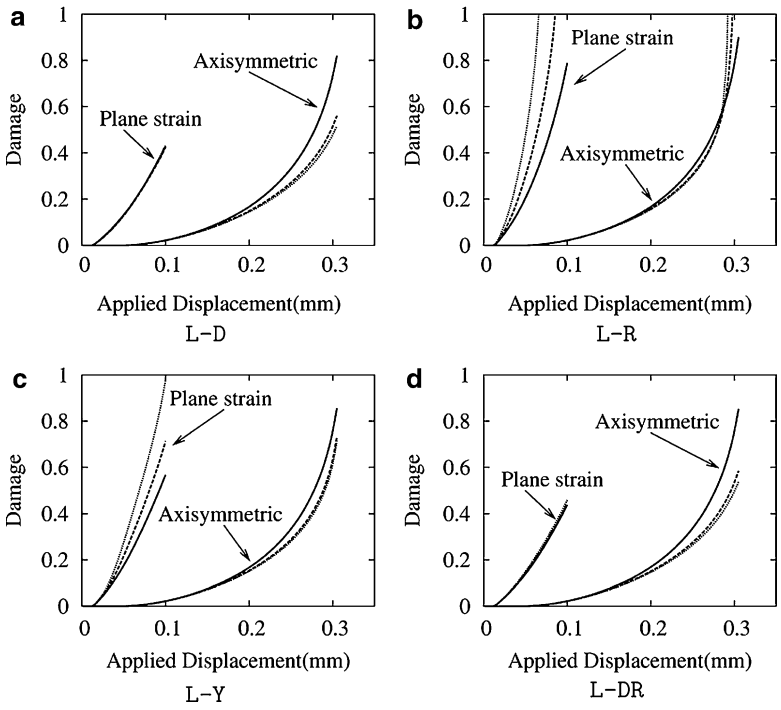


Fig. 18 Damage evolution for the Lemaitre-based nonlocal models. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

become mesh-independent but with the cost of triggering unwanted numerical instabilities and over-stiff responses. Another aspect worth mentioning regarding the conclusion that L-Y is not a good candidate for nonlocal variable in an implicit damage model is that seems to be not the case for nonlocal explicit damage models as reported by Jirasek and Rolshoven (Jirásek and Rolshoven 2003).

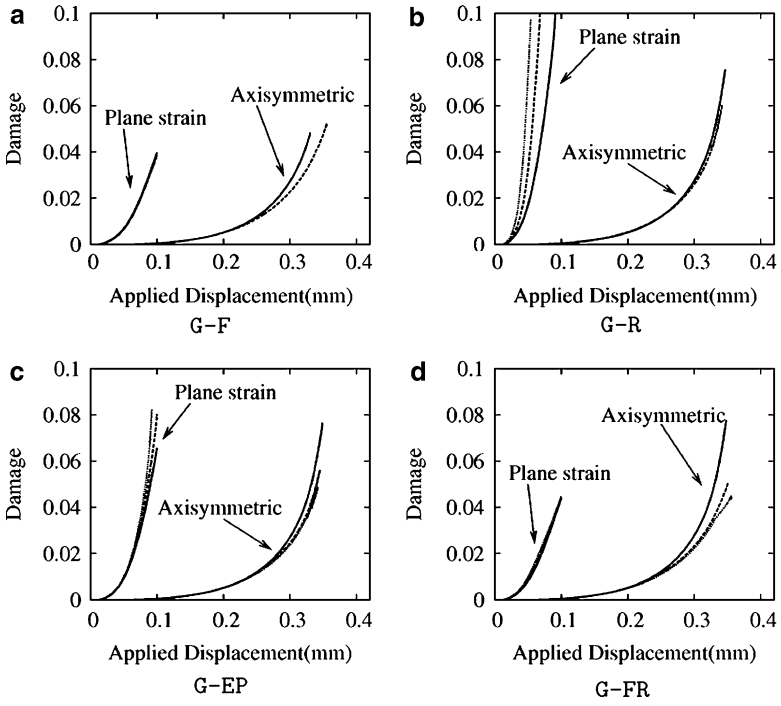


Fig. 19 Damage evolution for the Gurson-based nonlocal models. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

The evolution of damage for several Gurson-based nonlocal formulations is given in Fig. 19. Similar to the case of the Lemaitre-based nonlocal models, the regularization of damage (G-F) has eliminated pathological mesh dependency for both the axisymmetric and plane strain cases. In a similar manner, the simultaneous regularization of both damage and the hardening variables (G-FR) has clearly provided mesh-independent results in the plane strain case; however, inspection of Fig. 19d reveals that a slightly spurious behavior has appeared for the more refined mesh after a certain displacement was applied in the axisymmetric case. Unexpectedly, G-R has effectively diminished the effects of the loss of ellipticity in the plane strain case. Nonetheless, this option has revealed itself inadequate since it fails to avoid mesh dependency in the plane strain case. Finally, the formulation denoted by G-EP could substantially alleviate the effects of spurious mesh dependency for both tests, where in the latter case the regularization effects seem not to have been enough. Since the choice of damage (G-F) as nonlocal variable has presented superior results and recalling that the evolution of equivalent plastic strain is embedded in the evolution of damage itself, these results suggest that the evolution of volumetric plastic strain, $\dot{\varepsilon}_v^p$, also plays a significant role in the issue of pathological mesh dependency. Therefore, the regularization of both evolutions through the regularization of damage seems to be the optimal choice (Figs. 20–23).

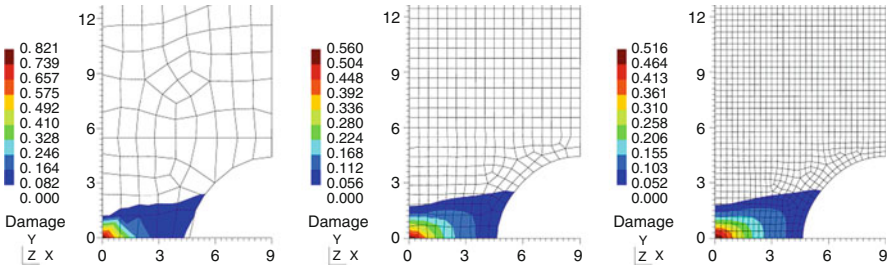


Fig. 20 Damage contours for the Lemaitre-based model nonlocal case (L-D). Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

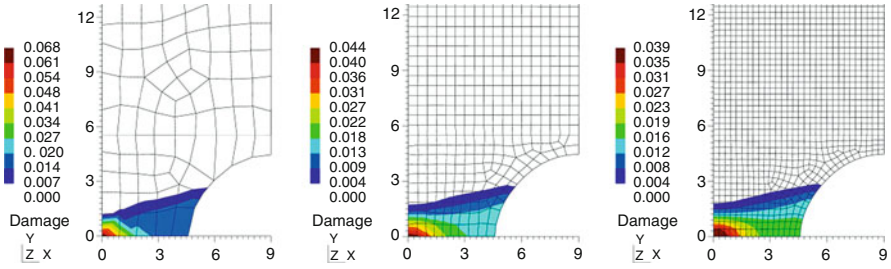


Fig. 21 Damage contours for the Gurson-based model nonlocal case (G-F). Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

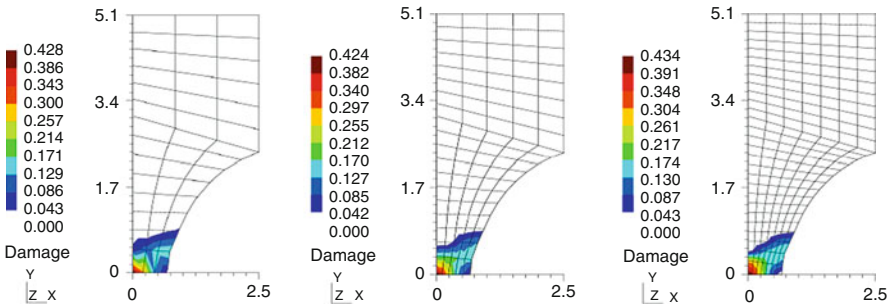


Fig. 22 Damage contours for the Lemaitre-based model nonlocal case (L-D). Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

Analysis at Moderate Stress Triaxiality

In order to assess the different nonlocal models defined in this chapter under the triaxiality of $\eta = 1/3$, a perforated plate specimen under traction force is simulated using three different meshes (see Fig. 24).

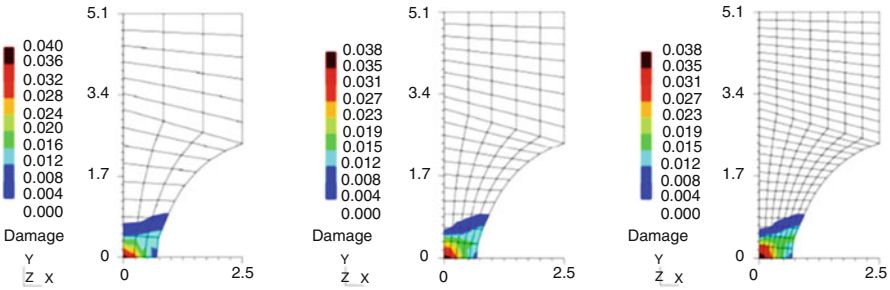


Fig. 23 Damage contours for the Gurson-based model nonlocal case (G-F). Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

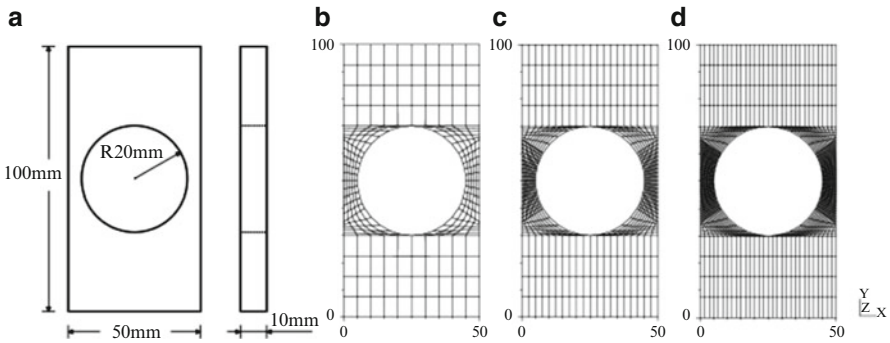


Fig. 24 Geometry and different mesh refinements for the perforated plate. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

Firstly the plate is simulated considering the local case. Figure 25 depicts the evolution of damage against the applied displacement for both Lemaitre- and Gurson-based models. The contours of damage are also plotted in Figs. 26 and 27. Clearly, pathological mesh dependency has taken place in both cases.

In Figs. 28 and 29, the evolution of damage for the nonlocal case is presented. Observing the results for the Lemaitre-based nonlocal models, we conclude that the best results were achieved with formulations L-D and L-DR with almost no differences between them. As expected, L-R was not able to regularize the problem, still giving solutions that are very sensitive to spatial discretization.

In the case of the Gurson-based models, all the four nonlocal formulations, G-F, G-R, G-EP, and G-FR, have eliminated the effects of spurious mesh dependency.

Analysis at Low Stress Triaxiality

In order to assess the different nonlocal models under low stress triaxiality, we shall limit ourselves to observe the behavior of these models under pure shear stress

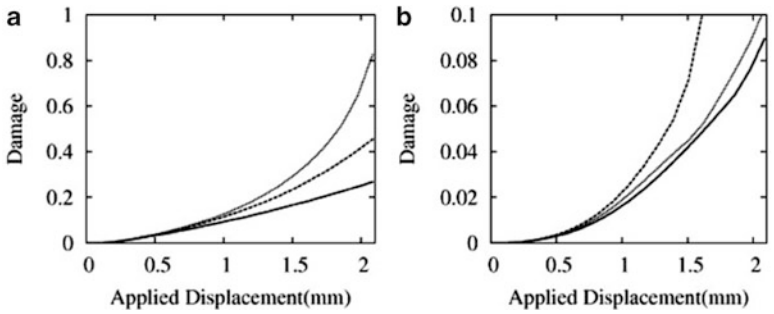


Fig. 25 Damage evolution in the local case. Results for the (**a**) Lemaitre- and (**b**) Gurson-based models. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

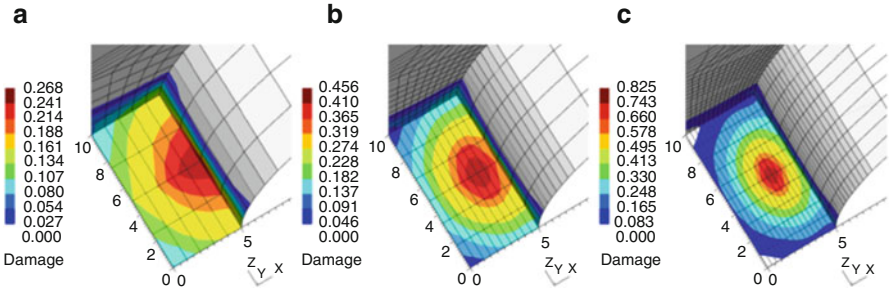


Fig. 26 Damage contours for the local case at the critical region: Lemaitre-based models. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

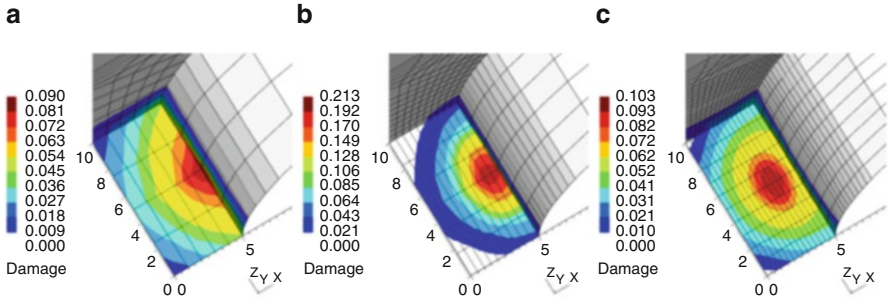


Fig. 27 Damage contours for the local case at the critical region: Gurson-based models. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

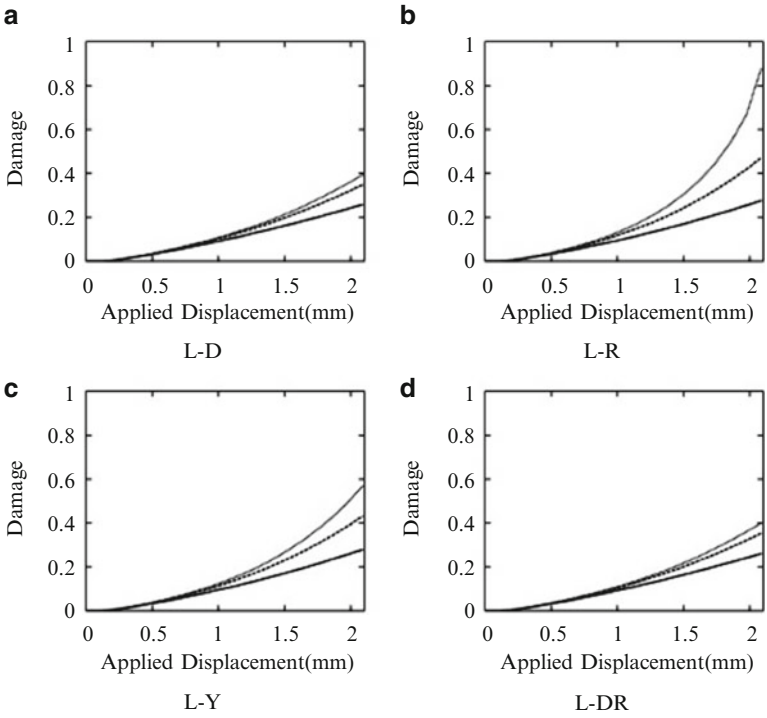


Fig. 28 Damage evolution for the Lemaitre-based nonlocal models. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

states (i.e., $\eta = 0$) and no analysis under compressive loadings will be carried out. The shear specimen used by Brunig et al. (2008) will be adopted for the present assessment where the geometry of the specimen and the three meshes adopted at the critical zone are shown in Figs. 30 and 31, respectively.

Damage evolution and contours for the local models are shown for both models in Figs. 32–34. It is clear that the same tendency of the previous cases in what regards mesh dependence is again verified but with a lesser extent for the Gurson model, which is perhaps due to both the facts that this model accounts for a more realistic modelling of the material behavior under shear stress state and that experiments tend to show that, in this case, localization is not as evident as in tension. Nevertheless, in the Gurson model although the maximum damage value does not change much upon mesh refinement, the failure zone is wrongly defined and concentrating in a very small region.

Some of the nonlocal models effectively eliminate the pathological mesh sensitivity as it may be seen in the results shown in Figs. 35 and 36. For the Lemaitre model's case, the solutions with L-D and L-DR have been able to regularize the solution, meanwhile L-Y has presented an enhanced solution (if compared to the local case) but not completely free of spurious dependency on the spatial discretization.

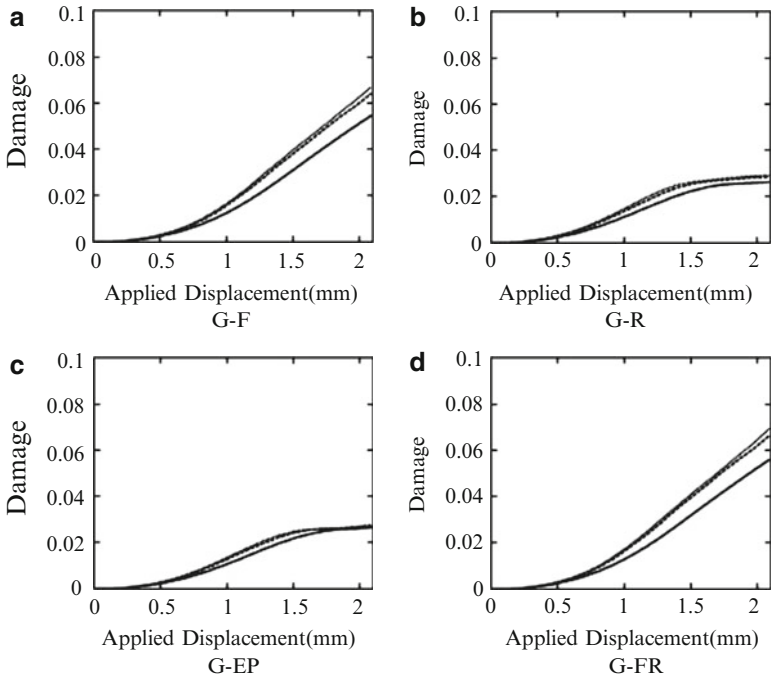


Fig. 29 Damage evolution for the Gurson-based nonlocal models. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

It is also clear that in the case of Gurson models, only G-F and G-FR could achieve mesh-insensitive solutions upon mesh refinement.

The contours of damage are given for the Lemaitre- (L-D) and Gurson-based (G-F) nonlocal models in Figs. 37 and 38, respectively. Clearly, the effects of pathological mesh dependency have been consistently eliminated since the damage contours practically remain constant and distributed over a finite area as the mesh is refined. Again, this undoubtedly demonstrates the effectiveness of these two nonlocal models (L-D and G-F) where it is worth mentioning that they have been able to tackle the issues of spurious mesh sensitivity in all cases analyzed in this section, regardless of the stress state.

Summary of Results

The results of the previous assessment are schematically summarized in Table 10. Reference values for the initial triaxiality and the average normalized third invariant have been given for convenience. The results suggest, to some extent, that the third invariant may have more influence in the issue of pathological mesh dependency than the triaxiality. Furthermore, the solutions in which damage has been regularized have been effective in all cases. Although L-D, L-DR, G-F, and G-FR have resulted in most cases, from a numerical point of view, the regularization of

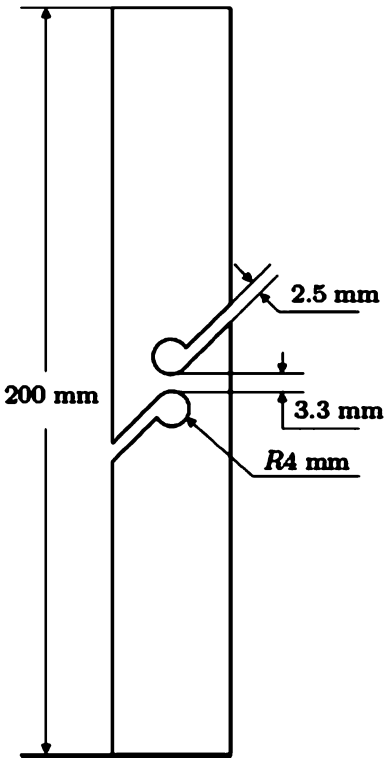


Fig. 30 Geometry of the shear specimen. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

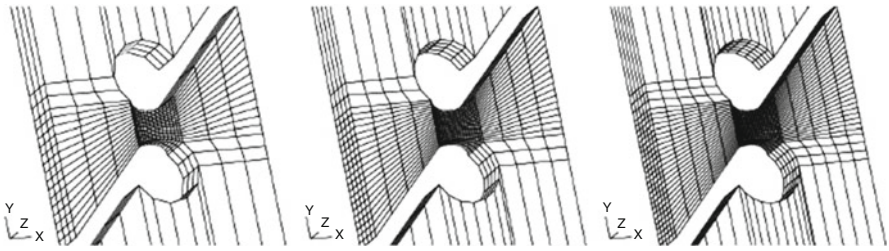


Fig. 31 Shear specimen: different mesh refinements at the critical region. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

damage only seems more advantageous than regularizing both damage and the isotropic hardening variable simultaneously. In the latter case, a dual integral averaging has to be carried out, being more computationally costly.

We remark again that the conclusion that damage is the preferred variable to be regularized in the case of Lemaitre- and Gurson-based models is significantly important as it seems that this is an inherent characteristic of implicit damage

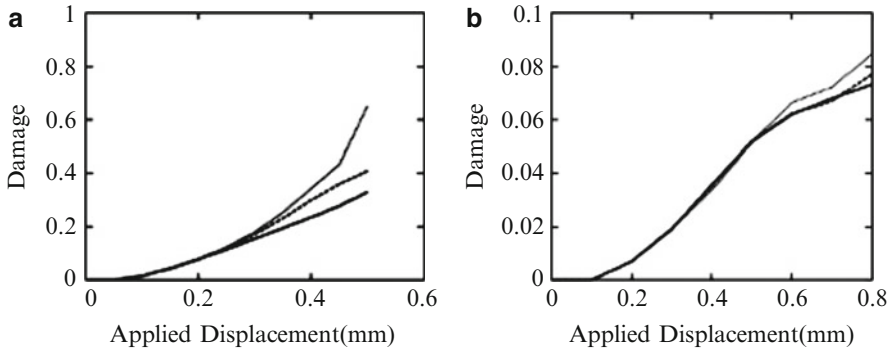


Fig. 32 Damage evolution in the local case. Results for the (**a**) Lemaitre- and (**b**) Gurson-based models. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

models. In the case of explicit damage models, often used to model quasi-brittle materials, the conclusion is completely different. As reported by Jirasek and Rolshoven (2003), the damage variable is a bad candidate for nonlocal variable for such explicit damage models and should be avoided. Again, this utterly implies that one has to be very careful when choosing the nonlocal variable for a given constitutive model; in particular, it should be analyzed if the damage formulation of such material model is done in an implicit or explicit fashion.

Damage to Fracture Transition

Introduction

Continuum models successfully describe most stages of material behavior. Nevertheless, when it comes to the final stages of failure, these models are not able to represent the initiation and propagation of macro-cracks within a structure. To correctly address surface decohesion and avoid spurious damage growth, the use of a discontinuous approach becomes imperative.

So far, the most successful simulations of ductile fracture processes, in a finite element method framework, lay in strategies that involve relatively fine meshes and continuous remeshing (Vaz and Owen 2001; Mediavilla et al. 2006; Mediavilla 2005; Areias et al. 2011; Saanouni 2008; Belytschko et al. 2000; Bouchard et al. 2000). Nevertheless, this method is still burdensome and extremely expensive in terms of computational costs. In addition, errors may be introduced due to the need of projecting the field variables.

Remeshing may be simplified by applying the element deletion technique (Song et al. 2008; Beissel et al. 1998), where the elements in which the fracture criterion is met are simply deleted from the mesh. However, this technique is highly dependent

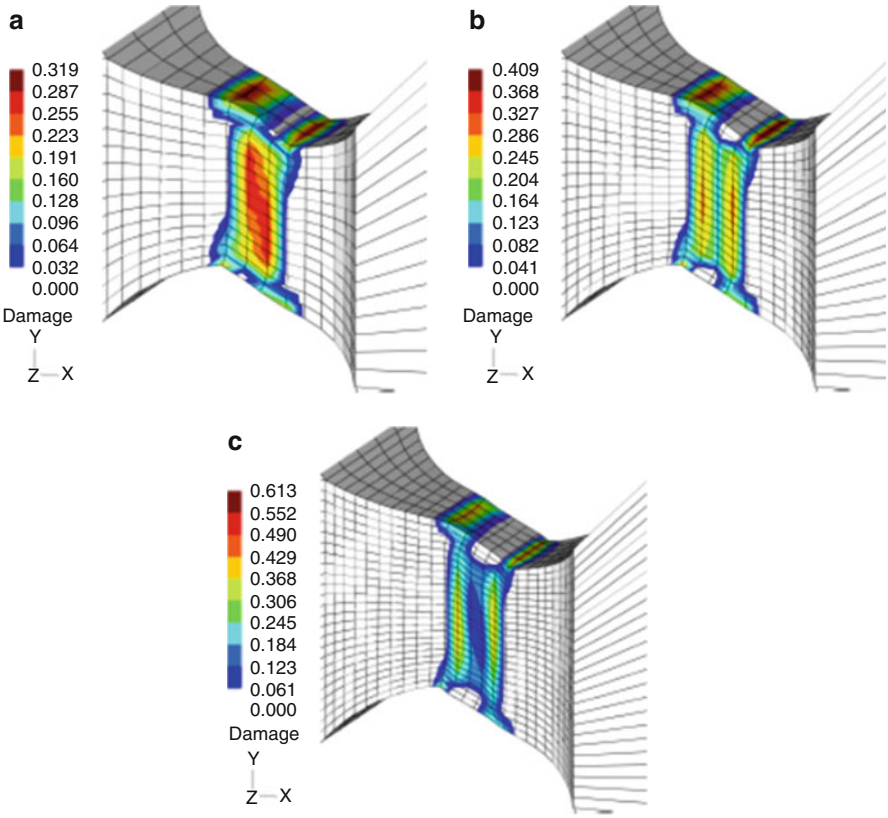


Fig. 33 Damage contours for the local case at the critical region: Lemaitre-based model. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

on the element size and orientation. Models based on the smeared crack model (Jirasek and Zimmermann 1998), in which the effects of a discontinuity are incorporated in the stress field and not at the displacement or strain field level, overcome the disadvantages of remeshing but, due to poor kinematic representation, are being replaced by approaches in which displacement or strain level discontinuities are positioned intra-element.

A large number of intra-element discontinuity models falls within the embedded discontinuities class (Jirasek 2000). These models are in general characterized by the introduction of new deformation modes in the standard finite element. These deformation modes are able to represent discontinuities with an arbitrary orientation both at the strain level (weak discontinuity) (e.g., Ortiz et al. 1987) or at the displacement level (strong discontinuity) (e.g., Belytschko et al. 1988; Simo et al. 1993). Extensions to large strain plasticity (e.g., Armero and Garikipati 1996) and large strain isotropic damage with application to ductile fracture

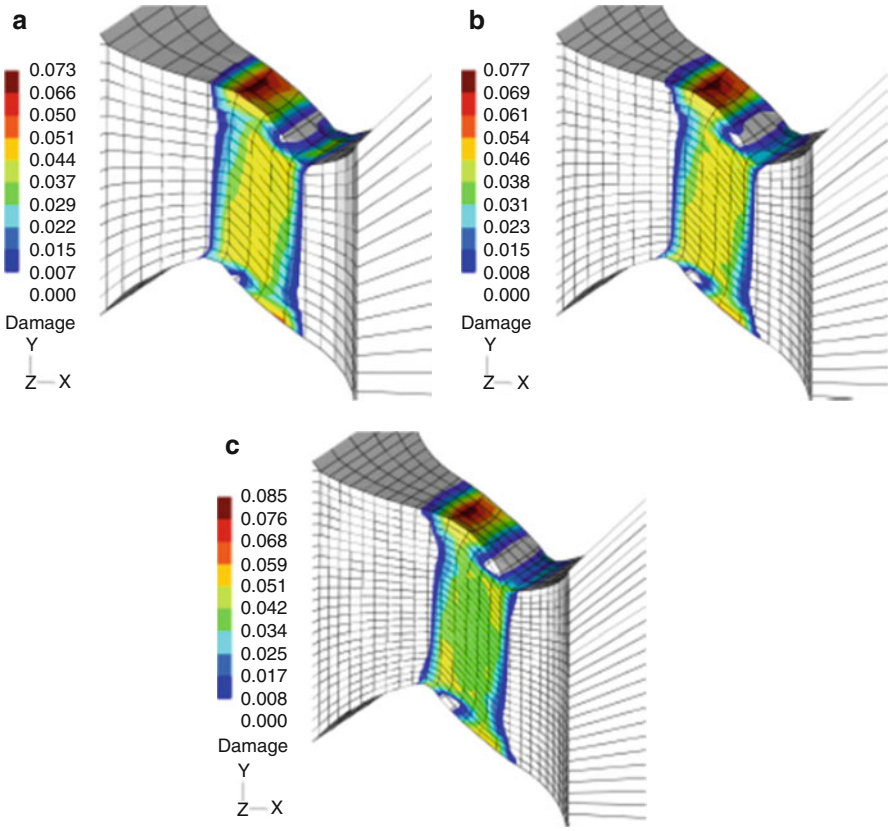


Fig. 34 Damage contours for the local case at the critical region: Gurson-based model. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

(e.g., Sanchez et al. 2008; Huespe et al. 2012) are also available. Although the embedded discontinuity approach allows the efficient modelling of regions with highly localized strains, the two parts of the element separated by the discontinuity are not totally independent and often the deformation state needs to be taken to be constant across the discontinuity band.

Another powerful technique to represent discontinuities intra-element is the eXtended Finite Element Method (XFEM) (Belytschko and Black 1999) in which the standard displacement field approximation is enriched with functions which are able to capture decohesion between two surfaces. Extra degrees of freedom are added to the nodes of the elements containing the discontinuity, allowing free propagation through the mesh. Although most of the existing literature on the XFEM is focused on brittle fracture, the method possesses the desired flexibility, at reasonable computer costs, to be employed in ductile fracture. Nevertheless, the developments in the use of the XFEM in ductile fracture problems are still limited, especially when dealing with large strain yielding.

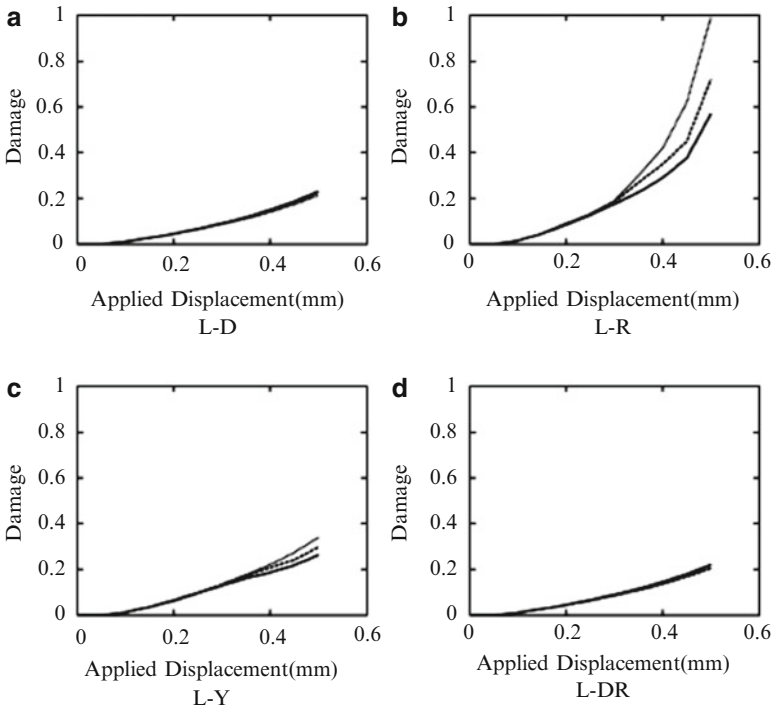


Fig. 35 Damage evolution for the Lemaitre-based nonlocal models. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

Initially, the XFEM works dealing with cracks in elastic media were combined with cohesive zone models to describe crack propagation in elastic-damageable materials (Alfaiate et al. 2002; Simone et al. 2003; Benvenuti 2008). Cazes et al. studied the conditions in which the transition from a continuous model to a continuous–discontinuous model is acceptable, firstly in elastic-damageable materials (Cazes et al. 2009) and later in elastoplastic-damageable materials (Cazes et al. 2010). However in those studies the methodology is only fully developed for 1D cases and the location of the crack has to be assumed a priori. It is also worth mentioning the extension of these works to a thermomechanical framework by Fagerström et al. (Fagerstrom and Larsson 2008) and the thick level set approach by Möes et al. (2011), which incorporates the damage variable in the level set to model the behavior of elastic-damageable solids in a new way.

The XFEM possesses interesting characteristics to develop a successful simulation of ductile failure processes, and a simple strategy on how to apply it is briefly described in this section. More sophisticated studies will be required for its full assessment. In the remainder of this section, it will be shown how the XFEM may be combined with continuum models, in particular the Lemaitre damage model, to describe ductile fracture in a more realistic way. Therefore, we start by briefly

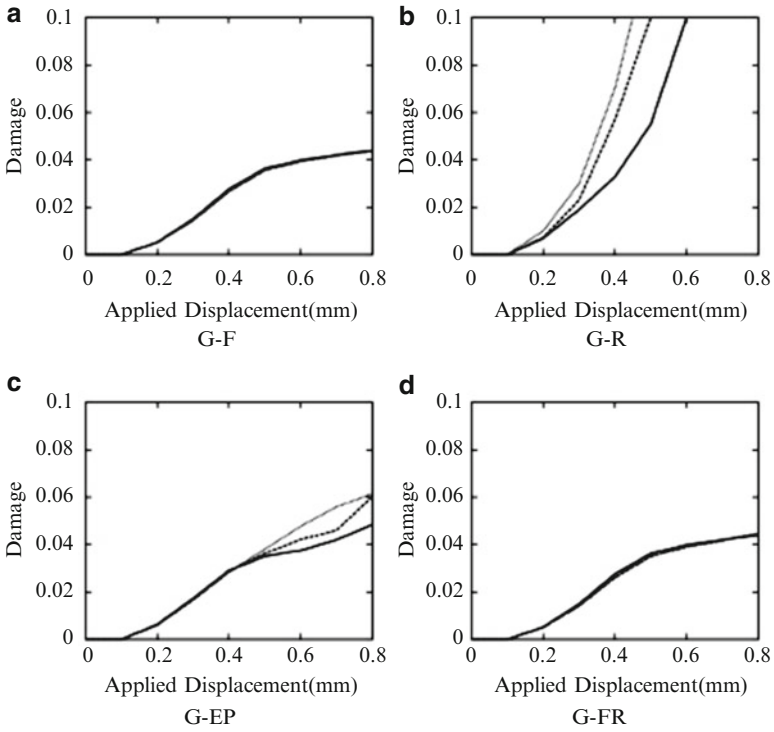


Fig. 36 Damage evolution for the Gurson-based nonlocal models. Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

presenting the XFEM enrichment functions used in a ductile fracture problem, then a short description of a possible and simplified transition criterion from damage to fracture is given and finally some numerical examples are used to test its features.

The Extended Finite Element Method (XFEM)

In the XFEM the standard problem field approximation is enriched with additional functions able to capture a certain desired feature. A large number of references are available on the subject (e.g. Belytschko and Black 1999; Möes et al. 1999; Abdelaziz and Hamouine 2008; Fries and Belytschko 2010), and therefore, this section will focus on the essential aspects of the enrichment functions.

In the particular case of fracture, the enrichment function should represent decohesion between the two surfaces of a crack. Discontinuous functions such as the following Heaviside function, $H(\hat{\eta})$, (Eq. 45) defined in terms of the coordinate orthogonal to the crack plane, $\hat{\eta}$, (Fig. 39), are particularly suitable for this purpose.

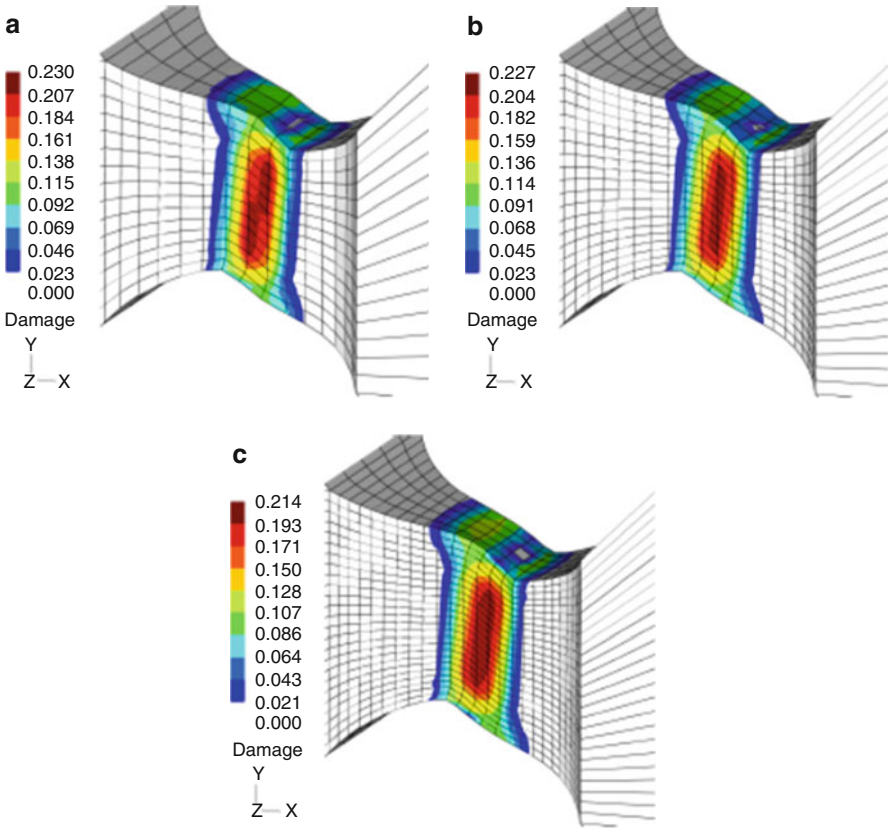


Fig. 37 Damage contours for the nonlocal case at the critical region: Lemaitre-based model (L-D). Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

$$H(\hat{\eta}) = \begin{cases} 1 & if\ \hat{\eta} \geq 0 \\ -1 & if\ \hat{\eta} < 0 \end{cases} \tag{45}$$

Enriching finite elements only with the Heaviside function always causes the extension of the crack to the edge of the element. Therefore, to represent cracks truly mesh independent, it is desirable to introduce a crack tip function. In this work, the following function denoted by T is employed:

$$T(\hat{\xi}) = \begin{cases} 1 & if\ \hat{\xi} \leq 0 \\ 0 & if\ \hat{\xi} > 0 \end{cases} \tag{46}$$

This function is particularly suitable for ductile fracture problems, where plasticity is not only confined to the region around the crack tip but widely spread. Furthermore, it vanishes outside the elements containing discontinuities, which is not the case if other types of enrichments are employed (Fries and Belytschko 2010; Chessa et al. 2003; Fries 2008).

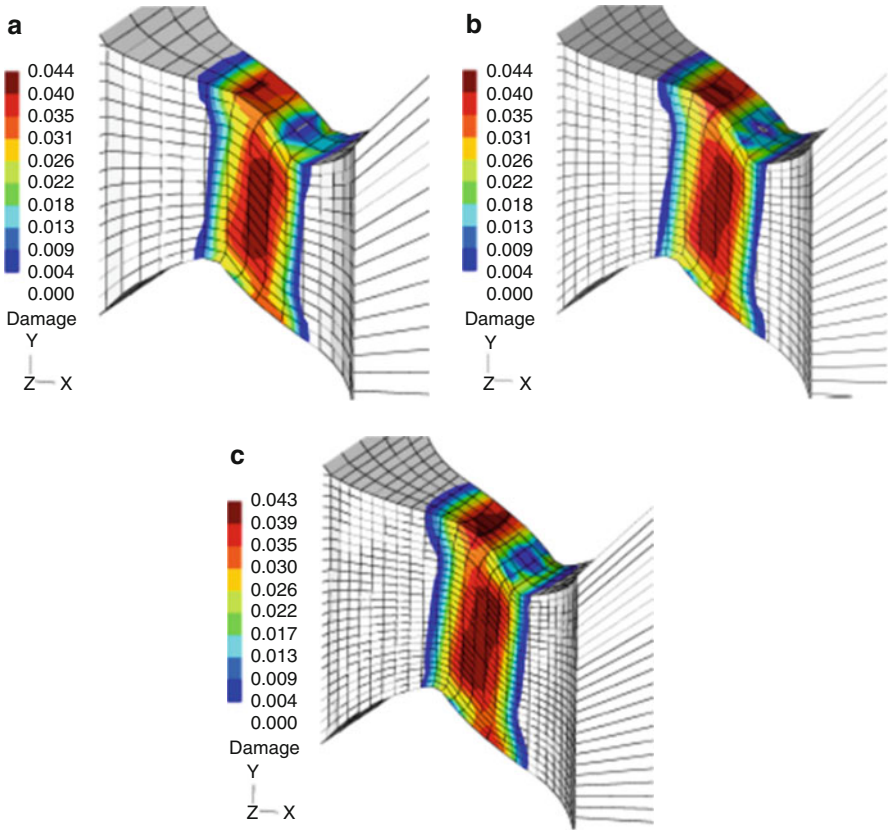


Fig. 38 Damage contours for the nonlocal case at the critical region: Gurson-based model (G-F). Reprinted from F.X.C. Andrade et al., Assessment and Comparison of Non-local Integral Models for Ductile Damage. Int. J. Damage Mech **23**(2), 261–296 (2014), by permission of SAGE Publications

After defining the enrichment functions, the displacement field approximation may be written as follows:

$$u(\mathbf{x}) = \sum_{i=1}^{n} N_i \boldsymbol{u}_i + \sum_{j=1}^{n_{split}} N_j \left[H(\boldsymbol{x}) - H\left(\boldsymbol{x}_j\right)\right] \boldsymbol{a}_j + \sum_{k=1}^{n_{tip}} N_k [T(\boldsymbol{x}) - T(\boldsymbol{x}_k)][H(\boldsymbol{x}) - H(\boldsymbol{x}_k)]\boldsymbol{b}_k \tag{47}$$

where N_i represents the standard element shape functions, $\boldsymbol{u}_i$ represents the nodal displacements, and $\boldsymbol{a}_j$ and $\boldsymbol{b}_k$ represent, respectively, the extra degrees of freedom associated with the elements totally crossed by a crack (*split elements*) and the extra degrees of freedom associated with the elements containing a crack tip

Table 10 Summary of results

Cases	η_0	ξ_{avg}	L-D	L-R	L-Y	L-DR	G-F	G-R	G-EP	G-FR
Axisymmetric	0.8	1.0	++	−−	++	++	++	++	++	++
Plane strain	0.7	0.0	++	−−	−	++	++	−−	+	++
Plate	1/3		++	−−	−	++	++	++	+	++
Shear	0.0	0.0	++	−−	+	++	++	−−	+	++

++ Full Regularization
+ Partial but acceptable regularization
− Poor/little regularization
−− No regularization

(*tip elements*). The enrichment functions are shifted relatively to the nodes to prevent them from spanning to neighbor elements.

Transition from Damage to Fracture

In a problem discretized through the finite element method, the values of the damage variable are stored at each Gauss point, from which the damage distribution pattern follows directly. This information is used to determine the crack characteristics, namely, its initiation point, direction, and length.

In the numerical model, fracture is triggered when a critical damage value, D_c, is reached and subsequently a crack is inserted in the model through the XFEM. The crack initiation point lies in the region where damage firstly reached the critical value. The elements in which damage is higher than the critical value are selected and grouped into clouds. The point of a cloud with the highest damage is the crack initiation point.

To determine the maximum damage point of the domain, it is essential to define a strategy to calculate the damage value in an arbitrary point of the domain. Therefore, a cubic B-spline (de Boor 1978; Piegl 1993; Cottrell et al. 2009) will be used as interpolation function. To build this function, we will start by considering the following polynomial basis defined within a set of four control points:

$$\begin{aligned} N_1 &= \frac{1}{8}(1-\xi)^3; \qquad N_2 = \frac{3}{8}(1-\xi)^2(1+\xi) \\ N_3 &= \frac{3}{8}(1-\xi)(1+\xi)^2; \qquad N_4 = \frac{1}{8}(1+\xi)^3 \end{aligned} \tag{48}$$

The value of function f at a point ξ may be obtained by

$$f(\xi) = \sum_{i=1}^{4} N_i f_i \tag{49}$$

where f_i is the value of f at the respective control point. To extend this procedure to a 2D problem, a B-spline patch may be considered (de Boor 1978), combining those

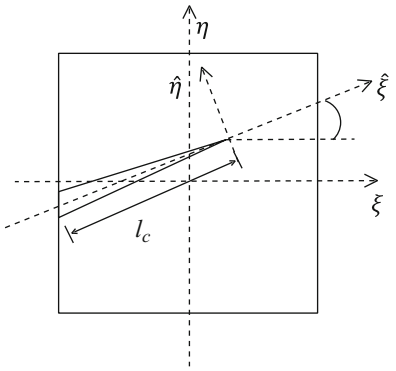


Fig. 39 Crack tip coordinates. Reprinted from M.R.R. Seabra, Damage Driven Crack Initiation and Propagation in Ductile Metals using XFEM. Comput. Mech. **52**(1), 161–179 (2013), with kind permission from Springer Science+Business Media

B-spline functions in two different directions. In particular, the set of control points may include points of the element in which the damage is to be determined and points in the adjacent elements, as illustrated in Fig. 40a.

Therefore, considering a coordinate system (ξ, η), originating in the central element of Fig. 40a, the following basis for the B-spline patch may be defined:

$$
\begin{aligned}
&\textit{for } \xi \leq 0 \\
N_\xi &= \left\{ \frac{1}{8}(1-\xi)^3, \frac{3}{8}(1-\xi)^2(1+\xi), \frac{3}{8}(1-\xi)(1+\xi)^2, \frac{1}{8}(1+\xi)^3, 0, 0, 0 \right\} \\
&\textit{for } \xi > 0 \\
N_\xi &= \left\{ 0, 0, 0, \frac{1}{8}(3-\xi)^3, \frac{3}{8}(3-\xi)^2(\xi-1), \frac{3}{8}(3-\xi)(\xi-1)^2, \frac{1}{8}(\xi-1)^3 \right\} \\
&\textit{for } \eta \leq 0 \\
N_\eta &= \left\{ \frac{1}{8}(1-\eta)^3, \frac{3}{8}(1-\eta)^2(1+\eta), \frac{3}{8}(1-\eta)(1+\eta)^2, \frac{1}{8}(1+\eta)^3, 0, 0, 0 \right\} \\
&\textit{for } \eta > 0 \\
N_\eta &= \left\{ 0, 0, 0, \frac{1}{8}(3-\eta)^3, \frac{3}{8}(3-\eta)^2(\eta-1), \frac{3}{8}(3-\eta)(\eta-1)^2, \frac{1}{8}(\eta-1)^3 \right\}
\end{aligned}
\tag{50}
$$

Then, the damage value at point $\boldsymbol{p}$, D_p, may be calculated using the relation:

$$
D_p = \boldsymbol{N}_\xi \boldsymbol{D}_{\xi\eta} \boldsymbol{N}_\eta \tag{51}
$$

where $\boldsymbol{D}_{\xi\eta}$ is a matrix containing the damage values at the control points.

This methodology to determine the damage value at an arbitrary point of the domain features several advantages. Unlikely Lagrange polynomials, B-splines do not oscillate and therefore do not introduce new maxima in the distribution (Cottrell et al. 2009). This insures that damage does not grow artificially due to the interpolation technique. Moreover, a B-spline patch of dimension *d* is itself a B-spline patch of dimension *d-1*, which means that a variable at point at the boundary of a B-spline surface/volume may be interpolated with the same functions as a point lying inside the domain.

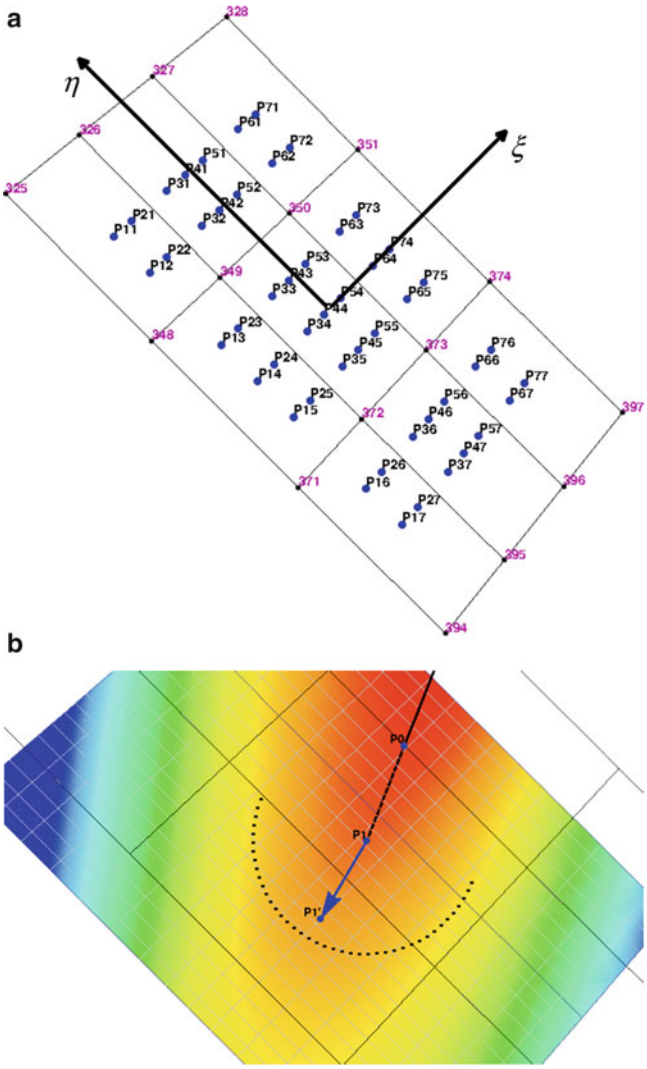


Fig. 40 (**a**) Set of control points for defining the interpolatory Bezier patch of the central element (**b**) Selection of points to determine the crack growth direction. Reprinted from M.R.R. Seabra, Damage Driven Crack Initiation and Propagation in Ductile Metals using XFEM. Comput. Mech. **52**(1), 161–179 (2013), with kind permission from Springer Science+Business Media

After determining the crack initiation point, the maximum damage growth direction is calculated. In terms of numerical implementation, the crack initiation algorithm is a particular case of the crack propagation algorithm in which the starting point of the crack is unknown.

To obtain a new segment of crack direction, a set of points is selected along a circumference, centered at the last crack tip, which contains all the elements where

Table 11 Material and geometrical properties

Property	Value
Elastic modulus	$E = 206.9$ GPa
Poisson's ratio	$\upsilon = 0.29$
Damage exponent	$s = 1.0$
Damage denominator	$r = 1.25$ MPa
Hardening function	$\tau_y(R) = 450 + 129.24R + 265(1 - e^{-16.93R})$ MPa

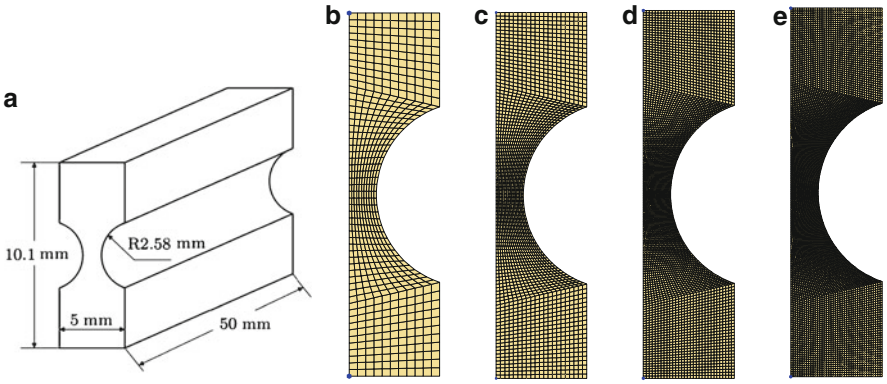


Fig. 41 (**a**) Plane strain specimen and finite element meshes with (**b**) 11, (**c**) 21, (**d**) 31, and (**e**) 41 elements per side. Reprinted from M.R.R. Seabra, Damage Driven Crack Initiation and Propagation in Ductile Metals using XFEM. Comput. Mech. **52**(1), 161–179 (2013), with kind permission from Springer Science+Business Media

the critical damage value has been reached in at least one Gauss point, as illustrated in Fig. 40b. As it is unlikely that the crack will snap back, the points behind the previous crack segment are excluded, and subsequently, the point with the highest damage value should be determined. The crack direction is then obtained by joining the last crack tip to the point with the highest damage. Moreover, for a good accuracy, two more circumferences, with a slightly higher and slightly smaller radius, are tested and then the three directions are averaged. Finally, the furthest point with $D = D_c$ is searched along the crack growth direction, using a midpoint algorithm. A detailed description of this process may be found in reference (Seabra et al. 2012).

Having the material model, the transition from damage to fracture criterion, and the crack representation strategy defined, the complete ductile fracture model is tested with some numerical examples.

Numerical Examples

The efficiency of the complete model for crack initiation and propagation in ductile metals is evaluated in this section, through some numerical examples. Material

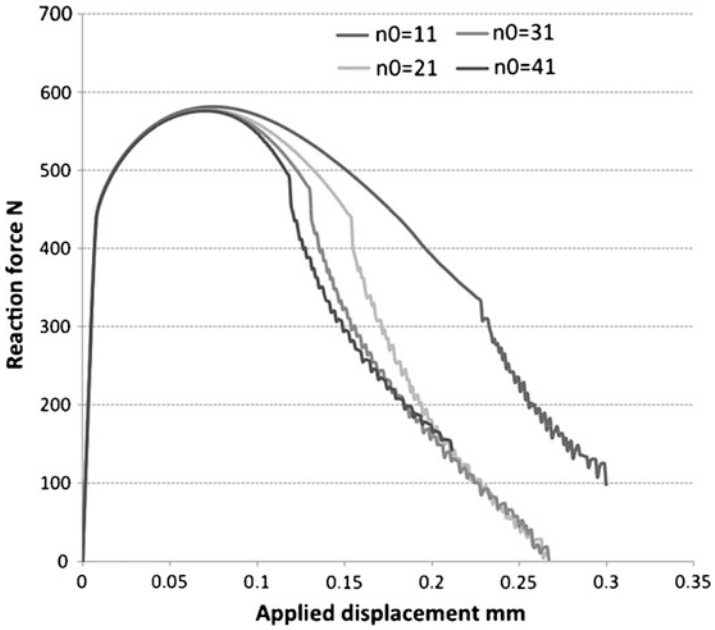


Fig. 42 Reaction force as a function of the applied displacement of the top nodes of the plane strain specimen. Reprinted from M.R.R. Seabra, Damage Driven Crack Initiation and Propagation in Ductile Metals using XFEM. Comput. Mech. **52**(1), 161–179 (2013), with kind permission from Springer Science+Business Media

behavior is described by the Lemaitre model for ductile damage, which is applicable to a wide range of ductile metals such as steel, aluminum, or copper (Lemaitre 1985a). In the following simulations, the chosen material is steel, whose properties are summarized in Table 11.

The Lemaitre model is implemented according to a nonlocal formulation; therefore, it requires a value for the nonlocal regularization length, ℓ_r, which is set to 1.6 mm in the following examples.

Finally, it is necessary to prescribe a value for the critical damage, D_c, which triggers crack initiation. Following the work of Lemaitre, (1985a), in real parts and structures, the critical damage value is not 1, which would correspond to theoretical fully damaged material, but is rather located between 0.2 and 0.5. In the following examples, the value $D_c = 0.5$ was adopted.

Plane Strain Specimen

The first example assessed is a plane strain specimen, whose geometry is represented in Fig. 41. Vertical displacements are applied to the top and bottom edges in order to produce traction-like loading conditions. The problem is discretized with four different FEM meshes, with 11, 21, 31, and 41, which are represented in Fig. 41 as well.

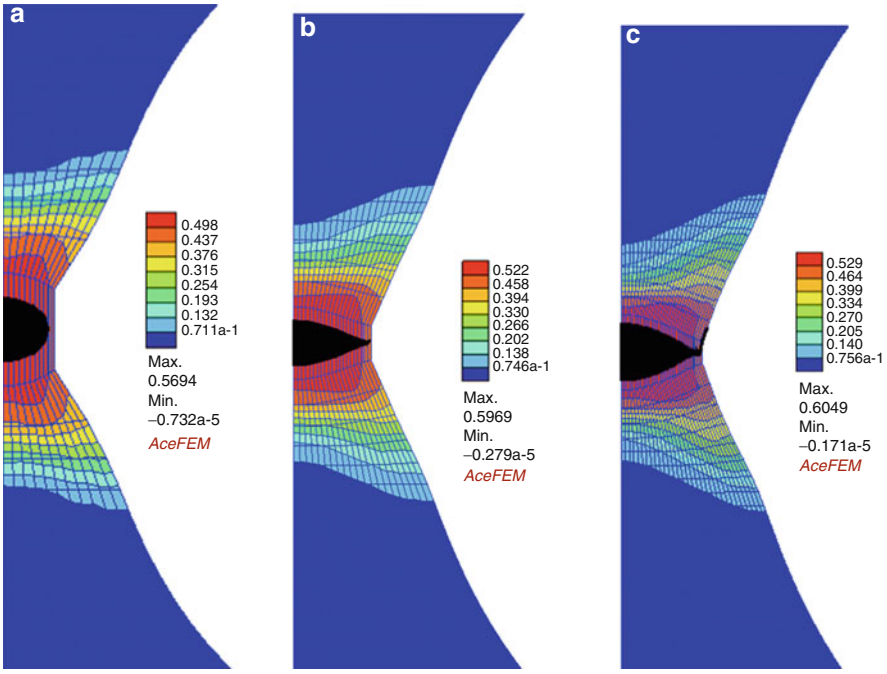


Fig. 43 Damage contour and final crack for mesh with (**a**) 11, (**b**) 21, and (**c**) 31 elements per side. Reprinted from M.R.R. Seabra, Damage Driven Crack Initiation and Propagation in Ductile Metals using XFEM. Comput. Mech. **52**(1), 161–179 (2013), with kind permission from Springer Science+Business Media

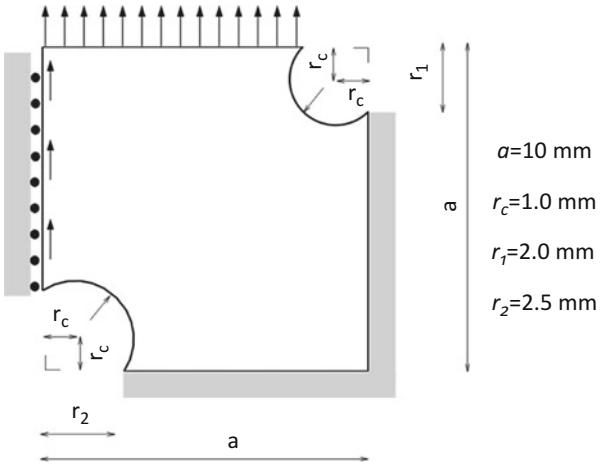


Fig. 44 Double-notched specimen geometry and boundary conditions. Reprinted from M.R.R. Seabra, Damage Driven Crack Initiation and Propagation in Ductile Metals using XFEM. Comput. Mech. **52**(1), 161–179 (2013), with kind permission from Springer Science+Business Media

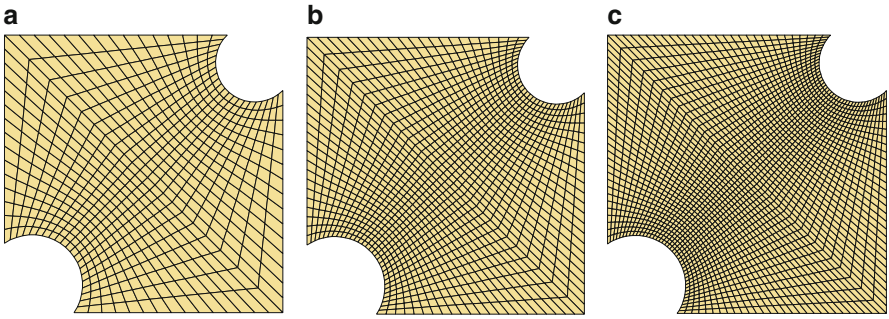


Fig. 45 FEM meshes used in the double-notched specimen with (**a**) 16 nodes per side, (**b**) 23 nodes per side, and (**c**) 30 nodes per side. Reprinted from M.R.R. Seabra, Damage Driven Crack Initiation and Propagation in Ductile Metals using XFEM. Comput. Mech. **52**(1), 161–179 (2013), with kind permission from Springer Science+Business Media

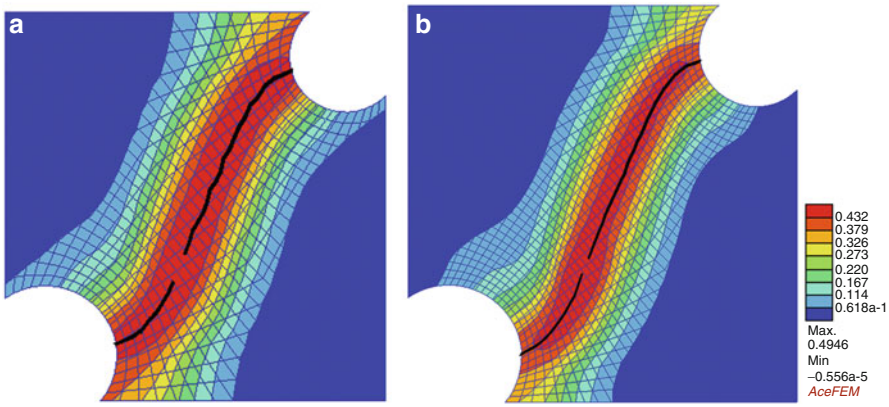


Fig. 46 Damage contour and final crack for mesh (**a**) and mesh (**b**). Reprinted from M.R.R. Seabra, Damage Driven Crack Initiation and Propagation in Ductile Metals using XFEM. Comput. Mech. **52**(1), 161–179 (2013), with kind permission from Springer Science+Business Media

In Fig. 42 the reaction force–displacement curves obtained are represented. A crack is included via XFEM when damage reaches its critical value according to the methodology described above. All the stages of material behavior, including hardening and softening up to failure, are clearly represented. Results are also convergent upon mesh refinement.

Under a plane strain condition, the nonlocal integral model also avoids pathological mesh dependence and spurious damage localization, as may be observed in the damage distribution contours illustrated in Fig. 43.

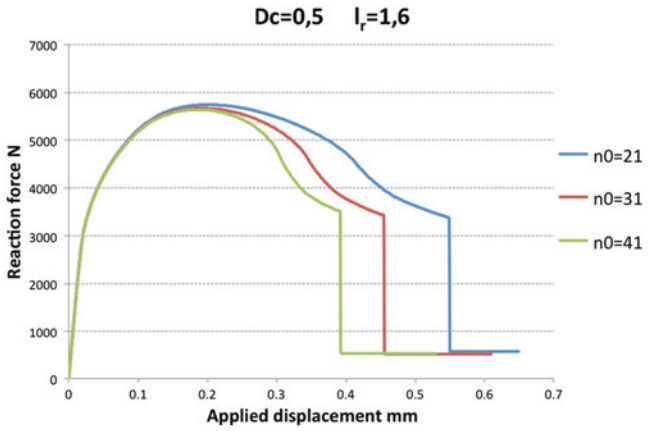


Fig. 47 Reaction force as a function of the applied displacement of the top nodes of the double-notched specimen. Reprinted from M.R.R. Seabra, Damage Driven Crack Initiation and Propagation in Ductile Metals using XFEM. Comput. Mech. **52**(1), 161–179 (2013), with kind permission from Springer Science+Business Media

Shear Specimen

This example consists of a double-notched specimen, illustrated in Fig. 44, and has two main objectives: on one hand, to evaluate the response of the model in the presence of shear loadings and, on the other hand, to compare the crack path obtained in a more challenging situation than the previous example.

The specimen is loaded in such a way that a shear-like failure mode will occur. The analysis is performed for the three meshes represented in Fig. 45.

In Fig. 46 the crack paths obtained for the two of the FEM meshes are represented. It can be observed that the paths are nearly the same for both cases, i.e., the crack path can be considered as mesh independent.

The nonlocal integral model avoids spurious localization of damage, resulting in similar damaged areas for different mesh refinements (Fig. 46). The subsequent insertion of the crack through the XFEM respects these damage contours independently of the mesh, indicating that this methodology is adequate to mimic failure due to crack formation and evolution.

Nevertheless, when observing the reaction force–applied displacement curve (Fig. 47), it is clear that the regularization effect of the nonlocal model is not as strong in this case as in the previous example. As referred above, the Lemaitre damage model is not the most adequate to capture material failure behavior when shear is its main cause, and therefore, some mesh dependence is still perceptible (Andrade 2011).

Conclusions

In phenomenological terms, the initiation of a crack in ductile metals is connected to the evolution of damage, which may be described by a continuum model.

Nevertheless, for a complete description of the failure process, if a local post-failure analysis is needed, once a critical damage value is met a transition to a discontinuous model is necessary.

This transition from a continuous to discontinuous model may be successfully dealt with by combining the XFEM with a plastic-damageable material model. Furthermore, the proposed methodology features the following advantages:

- Crack characteristics are determined directly from the continuous model and therefore there is no need of identifying additional material parameters for the fracture. Moreover, crack initiation and propagation are dealt with in a unified way.
- The crack initiation locus does not need to be known in advance, and the crack progresses independently from the mesh, saving computational resources when compared with competing approaches such as remeshing.
- Results are mesh independent upon a certain mesh refinement.

By resorting to nonlocal formulations, mesh dependence pathologies, typically associated to continuum softening models, are attenuated before and after the insertion of a crack. This simplified model, which may be a good approximation in practical terms to deal with the transition from a damage zone to a tension-free crack, could be, nevertheless, enhanced by adding a cohesive law to the discontinuity region so that the existing energy gap, due to the fact that the transition is made for a damage level, the critical damage, is less than the theoretical failure value.

Final Remarks and Some Perspectives

In a progressively more competitive industrial scenario, the need for more reliable predictions of ductile deformation and material failure is fundamental. Only so it is possible to substantially reduce the use of unnecessary material, thus achieving robust designs. A precise modelling of the ductile behavior inevitably involves the description of observed phenomena like plastic straining and strain-driven softening, often requiring highly nonlinear constitutive models. The introduction of new effects, both in the plastic flow rule of the material and in the evolution law for internal variables like damage, are among the topics most discussed and improved in the last decade. These scientific developments bring to the industry, in general, competitive gains in relation to the development of more efficient and durable mechanical components as well as the improvement of manufacturing processes. Thus, in this work, a study of different elastoplastic models with damage as internal variable was undertaken. The importance of the influence of the third invariant of the deviatoric stress tensor in the mechanical behavior of metallic materials was assessed mainly in shear-dominated stress states.

It is well known that the use of strain-softening laws inevitably leads to pathological mesh dependency if the standard local continuum theory is considered.

To solve this issue of spurious mesh sensitivity, a nonlocal approach of integral type was used, to a certain extent, to unfold the question of which constitutive variable should be regularized to avoid spurious mesh dependency. In order to properly answer this question, a comprehensive assessment of several nonlocal models has been carried out where these models have been subjected to different values of stress triaxiality and third invariant of the deviatoric stress tensor. Since the stress state plays a fundamental role in the behavior and fracture of ductile materials, it was the main concern to observe how the different nonlocal models respond under distinct external load conditions. It has been found that some options that are able to tackle spurious mesh sensitivity in a certain stress state fail completely when subjected to another stress triaxiality or third invariant. Furthermore, the conclusions reported by other authors in the context of explicit damage models are completely different from those drawn here, where implicit damage models have been considered. This utterly implies that one has to be very careful when choosing the nonlocal variable of a given constitutive model. After a very careful scrutiny, the assessment has shown that, in the case of implicit damage models, the damage variable seems to be the optimal choice.

A main assumption underlying this work was that the history of deformation has no influence on the nonlocal averaging operator, meaning that the nonlocal intrinsic length remains constant as the body undergoes deformation. However, there is no experimental evidence neither supporting nor contradicting this hypothesis. Apparently, a more realistic modelling would consider an evolving nonlocal intrinsic length, as a function of other variables that are known to significantly influence the dissipative fracturing process (e.g., the plastic strain, the stress state, the damage, or the history of deformation itself). However, the consideration of a nonlocal theory with a nonconstant intrinsic length as well as an averaging operator which is dependent on the history of deformation (or some other constitutive quantity) is still very challenging, both from the theoretical and the computational point of view. Nevertheless, the development of such enhanced theories would help to better understand material failure and also would widen the application of the nonlocal theory to a larger number of materials.

A continuum model is adequate to describe the underlying microscopic mechanisms which trigger fracture initiation. Nevertheless, a discontinuous approach is more suitable to describe the last stage of failure, which involves the propagation of macro-cracks. To combine these advantages of both continuous and discontinuous approaches, in order to build a model able to handle simultaneously large strains, damage localization, and crack propagation, the XFEM was inserted in the simulation model, in a simplified fashion. The damage variable distribution is used to define the size and the orientation of the evolving crack. Though the inclusion and propagation of a traction-free crack were governed by a critical damage level, improvements can be made by resorting to a cohesive law at the transition zone from damage to fracture, which should be calibrated by energetic consistency issues.

References

Y. Abdelaziz, A. Hamouine, A survey of the extended finite element. Comput. Struct. **86**(11–12), 1141–1151 (2008)

J. Alfaiate, G. Wells, L. Sluys, On the use of embedded discontinuity elements with crack path continuity for mode-I and mixed-mode fracture. Eng. Fract. Mech. **69**(6), 661–686 (2002)

F.X.C. Andrade, Non-local modelling of ductile damage: formulation and numerical issues. PhD Thesis, Faculty of Engineering, University of Porto, Porto, Portugal, 2011

F.X.C. Andrade, F.M. Andrade Pires, J.M.A. Cesar de Sa, L. Malcher, Nonlocal integral formulation for a plasticity-induced damage model. Comput. Methods Mater. Sci. **9**(1), 49–54 (2009)

F. Andrade, J. Cesar de Sa, F. Andrade Pires, A ductile damage nonlocal model of integral-type at finite strains: formulation and numerical issues. Int. J. Damage Mech. **20**, 515–557 (2011b)

P. Areias, N. Van Goethem, E. Pires, A damage model for ductile crack initiation and propagation. Comput. Mech. **47**, 641–656 (2011)

F. Armero, K. Garikipati, An analysis of strong discontinuities in multiplicative finite strain plasticity and their relation with the numerical simulation of strain localization in solids. Int. J. Solids Struct. **33**(20–22), 2863–2885 (1996)

A.G. Atkins, Possible explanation for unexpected departures in hydrostatic tension-fracture strain relations. Metal Sci. **15**, 81–83 (1981)

Y. Bai, Effect of loading history on necking and fracture. PhD Thesis, Massachusetts Institute of Technology, 2008

Y. Bai, T. Wierzbicki, A new model of metal plasticity and fracture with pressure and Lode dependence. Int. J. Plast. **24**, 1071–1096 (2008)

Y. Bao, Prediction of ductile crack formation in uncracked bodies. PhD Thesis, Massachusetts Institute of Technology, 2003

Y. Bao, T. Wierzbicki, On fracture locus in the equivalent strain and stress triaxiality space. Int. J. Mech. Sci. **46**(81), 81–98 (2004)

I. Barsoum, J. Faleskog, Rupture in combined tension and shear: experiments. Int. J. Solids Struct. **44**, 1768–1786 (2007a)

I. Barsoum, J. Faleskog, Rupture in combined tension and shear: micromechanics. Int. J. Solids Struct. **44**, 5481–5498 (2007b)

S.R. Beissel, G.R. Johnson, C.H. Popelar, An element-failure algorithm for dynamic crack propagation in general directions. Eng. Fract. Mech. **61**(3–4), 407–425 (1998)

T. Belytschko, T. Black, Elastic crack growth in finite elements with minimal remeshing. Int. J. Numer. Methods Eng. **45**, 601–620 (1999)

T. Belytschko, J. Fish, B.E. Engelmann, A finite element with embedded localization zones. Comput. Methods Appl. Mech. Eng. **70**(1), 59–89 (1988)

T. Belytschko, W.K. Liu, B. Moran, *Nonlinear Finite Elements for Continua and Structures* (Wiley, Chichester, 2000)

E. Benvenuti, A regularized XFEM framework for embedded cohesive interfaces. Comput. Methods Appl. Mech. Eng. **197**, 4367–4607 (2008)

E. Benvenuti, A. Tralli, Iterative LCP solver for non-local loading unloading conditions. Int. J. Numer. Methods Eng. **58**, 2343–2370 (2003)

G. Borino, P. Fuschi, C. Polizzotto, A thermodynamic approach to nonlocal plasticity and related variational principles. J. Appl. Mech. **66**, 952–963 (1999)

P.O. Bouchard, F. Bay, Y. Chastel, I. Tovena, Crack propagation modelling using an advanced remeshing technique. Comput. Methods Appl. Mech. Eng. **189**(3), 723–742 (2000)

P.W. Bridgman, *Studies in Large Plastic Flow and Fracture* (McGraw-Hill Book, New-York, 1952)

M. Brünig, S. Berger, H. Obrecht, Numerical simulation of the localization behavior of hydrostatic-stress-sensitive metals. Int. J. Mech. Sci. **42**, 2147–2166 (2008)

M. Brunig, O. Chyra, D. Albrecht, L. Driemeier, M. Alves, A ductile damage criterion at various stress triaxialities. Int. J. Plast. **24**, 1731–1755 (2008)

F. Cazes, M. Coret, A. Combescure, A. Gravouil, A thermodynamic method for the construction of a cohesive law from a non local damage model. Int. J. Solids Struct. **46**, 1476–1490 (2009)
F. Cazes, A. Simatos, M. Coret, A. Combescure, A cohesive zone model which is energetically equivalent to a gradient-enhanced coupled damage-plasticity model. Eur. J. Mech. A/Solids **29**, 976–998 (2010)
J.M.A. Cesar de Sa, P.M.A. Areias, C. Zheng, Damage modelling in metal forming problems using an implicit non-local gradient model. Comput. Methods Appl. Mech. Eng. **195**, 6646–6660 (2006)
J.M.A. Cesar de Sa, F.M. Andrade Pires, F.X.C. Andrade, Local and nonlocal modeling of ductile damage, in *Advanced Computational Materials Modelling: From Classical to Multi-Scale Techniques*, ed. by M. Vaz Jr., E.A. De Souza Neto, P.A. Muñoz-Rojas (Wiley-VCH, Weinheim, 2010)
J.L. Chaboche, M. Boudifa, K.A. Saanouni, CDM approach of ductile damage with plastic compressibility. Int. J. Fract. **137**, 51–75 (2006)
T.P. Chang, Z.P. Bazant, Instability of nonlocal continuum and strain averaging. J. Eng. Mech. ASCE **110**, 1441–1450 (1984)
J. Chessa, H. Wang, T. Belytschko, On the construction of blending elements for local partition of unity enriched finite elements. Int. J. Numer. Methods Eng. **57**, 1015–1038 (2003)
M.G. Cockcroft, D.J. Latham, Ductility and workability of metals. J. Inst. Metals **96**, 33–39 (1968)
J.A. Cottrell, T. Hughes, Y. Bazilevs, *Isogeometric Analysis – Toward Integration of CAD and FEA* (Wiley, Chichester, 2009)
J. Datsko, *Material Properties and Manufacturing Process* (Wiley, New York, 1966)
C. de Boor, *A Practical Guide to Splines* (Springer, New York, 1978)
R. De Borst, H. Mühlhaus, Gradient-dependent plasticity: formulation and algorithmic aspects. Int. J. Numer. Methods Eng. **35**, 521–539 (1992)
E.A. De Souza Neto, D. Peric, D.R.J. Owen, *Computational Methods for Plasticity: Theory and Applications* (Wiley, Chichester, 2008)
J.H.P. De Vree, W.A.M. Brekelmans, M.A.J. van Gils, Comparison of nonlocal approaches in continuum damage mechanics. Comput. Struct. **4**, 581–588 (1995)
L. Driemeier, M. Brünig, G. Micheli, M. Alves, Experiments on stress-triaxiality dependence of material behavior of aluminum alloys. Mech. Mater. **42**(2), 207–217 (2010)
D. Edelen, N. Laws, On the thermodynamics of systems with nonlocality. Arch. Ration. Mech. Anal. **43**, 24–35 (1971)
D. Edelen, A. Green, N. Laws, Nonlocal continuum mechanics. Arch. Ration. Mech. Anal. **43**, 36–44 (1971)
K. Enakousta, J.B. Leblond, G. Perrin, Numerical implementation and assessment of a phenomenological nonlocal model of ductile rupture. Comput. Methods Appl. Mech. Eng. **196**, 1946–1957 (2007)
R. Engelen, M. Geers, R. Ubachs, Nonlocal implicit gradient-enhanced elasto-plasticity for the modelling of softening behaviour. Int. J. Plast. **19**(4), 403–433 (2003)
M. Fagerstrom, R. Larsson, A thermo-mechanical cohesive zone formulation for ductile fracture. J. Mech. Phys. Solids **56**(10), 3037–3058 (2008)
M. Feucht, Ein gradientenabhangiges Gursonmodell zur Beschreibung duktiler Schadigung mit Entfestigung. PhD Thesis, Technische Universitat Darmstadt, 1999
A.M. Freudenthal, *The Inelastic Behaviour of Engineering Materials and Structures* (Wiley, New York, 1950)
T.-P. Fries, A corrected XFEM approximation without problems in blending elements. Int. J. Numer. Methods Eng. **75**, 503–532 (2008)
T.-P. Fries, T. Belytschko, The extended/generalized finite element method: an overview of the method and its applications. Int. J. Numer. Methods Eng. **84**, 253–304 (2010)

X. Gao, J. Kim, Modeling of ductile fracture: significance of void coalescence. Int. J. Solids Struct. **43**, 6277–6293 (2006)
X. Gao, T. Wang, J. Kim, On ductile fracture initiation toughness: effects of void volume fraction, void shape and void distribution. Int. J. Solids Struct. **42**, 5097–5117 (2005)
X. Gao, G. Zhang, C. Roe, A study on the effect of the stress state on ductile fracture. Int. J. Damage Mech. **19**, 75–94 (2009)
X. Gao, T. Zhang, J. Zhou, S.M. Graham, M. Hayden, C. Roe, On stress-state dependent plasticity modeling: significance of the hydrostatic stress, the third invariant of stress deviator and the non-associated flow rule. Int. J. Plast. **27**(2), 217–231 (2011)
A.L. Gurson, 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**, 2–15 (1977)
P. Hakansson, M. Wallin, M. Ristinmaa, Thermomechanical response of non-local porous material. Int. J. Plast. **22**, 2066–2090 (2006)
J.W. Hancock, A.C. Mackenzie, On the mechanisms of ductile failure in high-strength steels subjected to multi-axial stress-states. J. Mech. Phys. Solids **24**, 147–160 (1976)
A. Huespe, A. Needleman, J. Oliver, P.J. Sanchez, A finite strain, finite band method for modeling ductile fracture. Int. J. Plast. **28**(1), 53–69 (2012)
M. Jirasek, Nonlocal models for damage and fracture: comparison of approaches. Int. J. Solids Struct. **35**, 4133–4145 (1998)
M. Jirasek, Comparative study on finite elements with embedded discontinuities. Comput. Methods Appl. Mech. Eng. **188**(1–3), 307–330 (2000)
M. Jirásek, Nonlocal damage mechanics. Revue Européene de Génie Civil **11**, 993–1021 (2007)
M. Jirásek, S. Rolshoven, Comparison of integral-type nonlocal plasticity models for strain-softening materials. Int. J. Eng. Sci. **41**, 1553–1602 (2003)
M. Jirasek, T. Zimmermann, Analysis of rotating crack model. J. Eng. Mech., ASCE **124**, 842–851 (1998)
G.R. Johnson, W.H. Cook, Fracture characteristics of three metals subjected to various strains, strain rates, temperatures and pressures. Eng. Fract. Mech. **21**(1), 31–48 (1985)
L.M. Kachanov, Time of the rupture process under creep condition. Izv. Akad. Nauk. SSSR, Otd. Tekhn. Nauk **8**, 26–31 (1958)
J. Kim, X. Gao, T.S. Srivatsan, Modeling of crack growth in ductile solids: a three-dimensional analysis. Int. J. Solids Struct. **40**, 7357–7374 (2003)
J. Kim, X. Gao, T.S. Srivatsan, Modeling of void growth in ductile solids: effects of stress triaxiality and initial porosity. Eng. Fract. Mech. **71**, 379–400 (2004)
J. Kim, G. Zhang, X. Gao, Modeling of ductile fracture: application of the mechanism-based concepts. Int. J. Solids Struct. **44**, 1844–1862 (2007)
J. Lemaitre, A continuous damage mechanics model for ductile fracture. J. Eng. Mater. Technol. **107**, 83–89 (1985a)
J. Lemaitre, Coupled elasto-plasticity and damage constitutive equations. Comput. Methods Appl. Mech. Eng. **51**, 31–49 (1985b)
J. Lemaître, Local Approach of fracture. Eng. Fract. Mech. **25**, 523–537 (1986)
J. Lemaitre, *A Course on Damage Mechanics* (Springer, New York, 1996)
J. Lemaitre, R. Desmorat, *Engineering Damage Mechanics* (Springer, Berlin, 2005)
F.A. McClintock, A criterion for ductile fracture by growth of holes. J. Appl. Mech. **35**, 363–371 (1968)
J. Mediavilla, Continuous and discontinuous modeling of ductile fracture. PhD Thesis, Technische Universiteit Eindhoven, 2005
J. Mediavilla, R.H.J. Peerlings, M.G.D. Geers, A robust and consistent remeshing-transfer operator for ductile fracture simulations. Comput. Struct. **84**(8–9), 604–623 (2006)
G. Mirone, D. Corallo, A local viewpoint for evaluating the influence of stress triaxiality and Lode angle on ductile failure and hardening. Int. J. Plast. **26**(3), 348–371 (2010)

M.S. Mirza, D.C. Barton, P. Church, The effect of stress triaxiality and strain rate on the fracture characteristics of ductile metals. J. Mater. Sci. **31**, 453–461 (1996)
N. Möes, J. Dolbow, T. Belytschko, A finite element method for crack growth without remeshing. Int. J. Numer. Methods Eng. **46**, 131–150 (1999)
N. Möes, C. Stolz, P.-E. Bernard, N. Chevaugeon, A level set based model for damage growth: the thick level set approach. Int. J. Numer. Methods Eng. **86**(3), 358–380 (2011)
D.M. Norris, J.E. Reaugh, B. Moran, D.F. Quiñones, A plastic-strain, mean-stress criterion for ductile fracture. J. Eng. Mater. Technol., Trans ASME **100**, 279–286 (1978)
M. Ortiz, Y. Leroy, A. Needleman, A finite element method for localized failure analysis. Comput. Methods Appl. Mech. Eng. **61**(2), 189–214 (1987)
M. Oyane, S. Shima, T. Tabata, Considerations of basic equations and their application in the forming of metal powders and porous metals. J. Mech. Tech. **1**, 325–341 (1978)
R. Peerlings, R. De Borst, W.A.M. Brekelmans, J.H.P. De Vree, Gradient-enhanced damage for quasi-brittle materials. International Journal for Numerical Methods in Engineering. **39**, 1512–1533 (1996)
L. Piegl, *Fundamental Developments of Computer Aided Geometric Design* (Academic, San Diego, 1993)
G. Pijaudier-Cabot, Z.P. Bažant, Nonlocal damage theory. J. Eng. Mech. **113**(10), 1512–1533 (1987)
C. Polizzotto, A nonlocal strain gradient plasticity theory for finite deformations. Int. J. Plast. **25** (7), 1280–1300 (2009)
C. Polizzotto, G. Borino, P. Fuschi, A thermodynamic consistent formulation of nonlocal and gradient plasticity. Mech. Res. Commun. **25**(1), 75–82 (1998)
Y.N. Rabotnov, On the equations of state for creep, in *Progress in Applied Mechanics, Prager Anniversary Volume*, New York: MacMillan, pp 307–315 (1963)
F. Reusch, B. Svendsen, D. Klingbeil, A non-local extension of Gurson based ductile damage modeling. Comput. Mater. Sci. **26**, 219–229 (2003a)
F. Reusch, B. Svendsen, D. Klingbeil, Local and non-local Gurson based ductile damage and failure modelling at large deformation. Eur. J. Mech. A/Solids **22**, 779–792 (2003b)
J.R. Rice, D.M. Tracey, On the ductile enlargement of voids in triaxial stress fields. J. Mech. Phys. Solids **17**, 201–217 (1969)
K. Saanouni, On the numerical prediction of the ductile fracture in metal forming. Eng. Fract. Mech. **75**(11), 3545–3559 (2008)
M.K. Samal, M. Seidenfuss, E. Roos, B.K. Dutta, H.S. Kushwaha, Finite element formulation of a new nonlocal damage model. Finite Elem. Anal. Des. **44**, 358–371 (2008)
P.J. Sanchez, A.E. Huespe, J. Oliver, On some topics for the numerical simulation of ductile fracture. Int. J. Plast. **24**(6), 1008–1038 (2008)
M. Seabra, P. Sustaric, J. Cesar de Sa, T. Rodic, Damage driven crack initiation and propagation in ductile metals using XFEM. Comput. Mech. (2012). doi:10.1007/s00466-012-0804-9
J.C. Simo, T.J.R. Hughes, *Computational Inelasticity* (Springer, New York, 1998)
J.C. Simo, J. Oliver, F. Armero, An analysis of strong discontinuities induced by strain-softening in rate-independent inelastic solids. Comput. Mech. **12**, 277–296 (1993)
A. Simone, G. Wells, L. Sluys, From continuous to discontinuous failure in a gradient-enhanced continuum damage model. Comput. Methods Appl. Mech. Eng. **192**, 4581–4607 (2003)
J.-H. Song, H. Wang, T. Belytschko, A comparative study on finite element methods for dynamic fracture. Comput. Mech. **42**(2), 239–250 (2008)
L. Strömberg, M. Ristinmaa, FE-formulation of a nonlocal plasticity theory. Comput. Methods Appl. Mech. Eng. **136**, 127–144 (1996)
W. Tai, B.X. Yang, A new damage mechanics criterion for ductile fracture. Eng. Fract. Mech. **27**, 371–378 (1987)
X. Teng, Numerical prediction of slant fracture with continuum damage mechanics. Eng. Fract. Mech. **75**, 2020–2041 (2008)

V. Tvergaard, A. Needleman, Analysis of the cup-cone fracture in a round tensile bar. Acta Metall. **32**, 157–169 (1984)
V. Tvergaard, A. Needleman, Effects of nonlocal damage in porous plastic solids. Int. J. Solids Struct. **32**(8/9), 1063–1077 (1995)
M. Vaz, D.R.J. Owen, Aspects of ductile fracture and adaptive mesh re-finement in damaged elasto-plastic materials. Int. J. Numer. Methods Eng. **50**(1), 29–54 (2001)
G. Voyiadjis, G. Pekmezi, B. Deliktas, Nonlocal gradient-dependent modeling of plasticity with anisotropic hardening. Int. J. Plast. **26**, 1335–1356 (2010)
L. Xue, *Ductile Fracture Modeling – Theory, Experimental Investigation and Numerical Verification* (Massachusetts Institute of Technology, 2007)
L. Xue, Constitutive modeling of void shearing effect in ductile fracture of porous materials. Eng. Fract. Mech. **75**, 3343–3366 (2008)

Micromechanical Models of Ductile Damage and Fracture

27

A. Amine Benzerga

Contents

A.A. Benzerga (✉)
Department of Aerospace Engineering, Texas A&M University, College Station, TX, USA

Department of Materials Science and Engineering, Texas A&M University, College Station, TX, USA
e-mail: benzerga@aero.tamu.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_38

Abstract

Two classes of micromechanics-based models of void enlargement are presented succinctly with their fundamental hypotheses and synopsis of derivation highlighted. The first class of models deals with conventional void growth, i.e., under conditions of generalized plastic flow within the elementary volume. The second class of models deals with void coalescence, i.e., an accelerated void growth process in which plastic flow is highly localized. The structure of constitutive relations pertaining to either class of models is the same but their implications are different. With this as basis, two kinds of integrated models are presented which can be implemented in a finite-element code and used in ductile fracture simulations, in particular for metal forming processes. This chapter also describes elements of material parameter identification and how to use the integrated models.

Nomenclature

Symbol	Definition	Components
$\mathbf{I}$	Second-order identity tensor	δ_{ij}
$\mathbb{I}$	(Symmetric) fourth-order identity tensor	$\frac{1}{2}\left(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}\right)$
$T_{\mathrm{m}} \equiv 1/3\ \mathrm{tr}\ \mathbf{T}$	Mean part of tensor $\mathbf{T}$	
$\mathbf{T}' \equiv \mathbf{T} - T_{\mathrm{m}}\mathbf{I}$	Deviator of tensor $\mathbf{T}$	$T_{ij} - \frac{1}{3}T_{kk}\delta_{ij}$
$\mathbb{J} \equiv \mathbb{I} - \frac{1}{3}\mathbf{I}\otimes\mathbf{I}$	Deviatoric projector, e.g., $\mathbb{J} : \boldsymbol{\sigma} = \boldsymbol{\sigma}'$	$I_{ijkl} - \frac{1}{3}\delta_{ij}\delta_{kl}$
$\|\|\mathbf{T}\|\| \equiv (3/2\ \mathbf{T}' : \mathbf{T}')^{1/2}$	von Mises norm of tensor $\mathbf{T}$	
$\|\|\mathbf{T}\|\|_{\mathrm{H}} \equiv (3/2\ \mathbf{T}' : \mathbb{h} : \mathbf{T}')^{1/2}$	Hill norm of tensor $\mathbf{T}$	
Ω	Domain occupied by RVE	
ω	Domain occupied by voids	
$f \equiv \lvert\omega\rvert/\lvert\Omega\rvert$	Void volume fraction (porosity)	
W	Void aspect ratio (>1 if prolate)	
$\mathbf{e}_3$	Common axis of aligned spheroidal voids	
λ	Void spacing ratio (>1 if axial spacing is greatest)	
χ	Ligament parameter (=1 if no ligament left)	
Λ	Plastic multiplier	
$\boldsymbol{\sigma}$, $\mathbf{d}$	Microscopic Cauchy stress and rate of deformation	
$\boldsymbol{\Sigma}$, $\mathbf{D}$	Macroscopic Cauchy stress and rate of deformation	
$\Sigma_{\mathrm{eq}} \equiv \|\|\boldsymbol{\Sigma}\|\|$	von Mises equivalent stress (similar for σ_{eq})	
$D_{\mathrm{eq}} = 2/3\|\|\mathbf{D}\|\|$	Equivalent strain rate (similar for d_{eq})	
$\overline{\sigma}$	Microscopic yield stress	
$\mathbb{P}$	Hill's anisotropy tensor	
$\mathbb{h}$	Anisotropy tensor in the space of stress deviators	
$\hat{\mathbb{h}}$	Formal inverse of $\mathbb{h}$	

(*continued*)

Symbol	Definition	Components
$h_i i = 1{,}6$	Components of $\mathbb{h}$ after Voigt's condensation	
f_N	Volume fraction of void-nucleating particles	
ϵ_N	Average nucleation strain	
s_N	Standard deviation	

Introduction

Understanding the fundamental mechanisms of fracture in metal forming is of considerable importance, but the development of physically sound models with a quantitative predictive ability still poses a challenge. First, the plastic strains involved in forming are quite large, and this excludes the use of classical linear elastic fracture mechanics concepts. In addition, nonlinear fracture mechanics has essentially dealt with the fundamental problem of a body containing one or more initial cracks. In metal forming, there usually are no initial cracks. Also, significant microstructural evolutions take place at large plastic deformations. The latter include extreme grain elongations, texture development/evolution as well as the nucleation and growth of microvoids, either on second-phase particles or at stress risers.

The nonproportional loading paths that are inherent to forming operations make the prediction of failure even more challenging. Engineering tools and guidelines based on forming limit diagrams or stress-based criteria that do not embody a set of internal variables are of limited scope. They fail to capture, even qualitatively, the inherent path dependence of failure loci, in whichever way such loci are defined. There is a great potential for metal forming to rely on rational material design, i.e., based on sound physical models that possess the ability to connect processing parameters and microstructural variables to the mechanical properties of interest.

Of paramount importance to some forming operations is the intrinsic ductility of the material. Ductility is often understood as the ability of a material piece to withstand some amount of plastic or viscoplastic strain before the onset of structural instabilities. Necking of a bar under simple tension is the classical example. Viewed as a material property, the necking strain carries more the signature of the hardening capacity of the material than its intrinsic ductility. Yet qualitative correlations are commonly drawn between the two. This may be sufficient in some, but certainly not all forming operations, especially those involving load path changes.

The fundamental mechanisms and mechanics of ductile fracture have recently been reviewed by Benzerga and Leblond (2010); also see Besson (2010) for models. When the microscopic mechanisms involve microvoid growth to coalescence, micromechanical models have the capability to deliver physically sound predictions of structure-property relationships. Even since the last 2010 reviews, significant developments have taken place in this area. The objective of this chapter is to present a synthesis of ductile fracture models that are implementable

in finite-element programs for solving metal forming initial- and boundary-value problems. The experimental aspects of the subject are omitted and may be consulted in the above monograph. Also, void nucleation is not addressed for brevity. This lays the focus on the large-deformation phenomena that are void growth and void coalescence. In all models a representative volume element (RVE) is considered according to the classical acception in homogenization theory. The microscale refers to that of the matrix while the macroscale refers to matrix and voids. Situations where either scale separation does not hold or fracture is affected by extreme statistics are not within the scope of this synthesis. The general framework is that of porous metal plasticity with the difference in treatment between void growth and void coalescence lying in the assumed boundary conditions and propensity for microscopic localization.

Structure of Constitutive Relations

All porous metal plasticity models to be presented below share a common structure in their derivation. The macroscopic yield surface is parametrically defined as

$$\Sigma_{ij} = \frac{\partial \Pi}{\partial D_{ij}}(\mathbf{D}) \tag{1}$$

where $\Pi(\mathbf{D})$ is the macroscopic plastic dissipation associated with $\mathbf{D}$ (see "Nomenclature"):

$$\Pi(\mathbf{D}) = \inf_{\mathrm{d}\in\mathcal{K}(\mathbf{D})} \left\langle \sup_{\boldsymbol{\sigma}^*\in\mathcal{C}} \sigma^*_{ij} d_{ij} \right\rangle \Omega \tag{2}$$

Here, $\mathcal{C}$ denotes the microscopic (convex) domain of reversibility (the elasticity domain in small transformations) and $\mathcal{K}(\mathbf{D})$ the set of kinematically admissible microscopic deformations. If uniform strain-rate boundary conditions are assumed, then

$$\begin{aligned} \mathcal{K}(\mathbf{D}) = \Big\{ \mathbf{d} | \forall \mathbf{x} \in \Omega \backslash \omega, d_{kk} = 0 \quad \text{and} \quad \exists \mathbf{v}, \forall \mathbf{x} \in \Omega, \\ d_{ij} = \frac{1}{2}\left(v_{ij} + v_{j,i}\right) \text{ and } \forall \mathbf{x} \in \partial\Omega, \ \ v_i = D_{ij} x_{ij} \Big\} \end{aligned} \tag{3}$$

and localized modes of deformation within Ω are precluded. To account for these, other types of boundary conditions must be used. The former are typically employed in constructing void growth models, the latter in void-coalescence models.

To obtain expressions for Π in closed form, trial velocity fields are used. Therefore, the basic elements of a micromechanical model are:

(i) The geometry of the RVE
(ii) A microscale plasticity model, i.e., the boundary of $\mathcal{C}$ with the flow rule being necessarily associative
(iii) Kinematically admissible microscale velocity fields defining a subset of $\mathcal{K}(\mathbf{D})$

Bounding properties of the macroscopic dissipation are available, as discussed by Benzerga and Leblond (2010). Elimination of **D** from Eq. 1 leads to a macroscopic yield criterion of the form $\Phi(\Sigma; \text{ISVs}) = 0$ with an associated flow rule. Here "ISVs" refers to a collection of internal state variables with a definite microstructural significance.

Microstructural variables, e.g., the porosity f, enter Π, hence Φ, by way of homogenization. Void growth, rotation, and coalescence result from the evolution of these variables.

For example, the time rate of change of f results from plastic incompressibility at the microscale:

$$\dot{f} = (1-f)D_{kk} = (1-f)\Lambda\frac{\partial\Phi}{\partial\Sigma_{\mathrm{m}}}, \tag{4}$$

so that $\dot{f}$ derives directly from the yield criterion by normality. For anisotropic models with one or more void shape parameters, additional evolution equations are required for the void shapes and orientations. Void-coalescence models are inherently anisotropic.

Void Growth

Gurson Model

Gurson (1977) used a different method to obtain his yield function. The same yield function can be arrived at using the approach outlined above:

(i) Geometry:
The RVE is a hollow sphere containing a concentric spherical void. (A variant of the model exists for a cylindrical RVE.) The porosity f is the only microstructural variable entering the model.

(ii) Plasticity model:
An associated J_2 flow theory is used for the matrix with the yield criterion and flow rule written as

$$\sigma_{\mathrm{eq}} \equiv \|\boldsymbol{\sigma}\| = \overline{\sigma}, \qquad \mathbf{d} = \frac{3}{2}\frac{d_{\mathrm{eq}}}{\overline{\sigma}}\boldsymbol{\sigma}' \tag{5}$$

(iii) Velocity fields:

$$\forall \mathbf{x} \in \Omega\backslash\omega, \quad v_i(\mathbf{x}) = Av_i^A(\mathbf{x}) + \beta_{ij}x_j, \quad \mathbf{v}^A(\mathbf{x}) = \frac{1}{r^2}\mathbf{e}_r \tag{6}$$

where a mix of Cartesian and spherical coordinates is used for convenience. Scalar A and symmetric tensor $\boldsymbol{\beta}$ are parameters (with $\beta_{kk} = 0$).

On that basis, obtain a bounding dissipation function as

$$\Pi(\mathbf{D}) = \overline{\sigma}\left[2D_{\mathrm{m}}\sinh^{-1}\left(\frac{2D_{\mathrm{m}}x}{D_{\mathrm{eq}}}\right) - \sqrt{4D_{\mathrm{m}}^2 + \frac{D_{\mathrm{eq}}^2}{x^2}}\right]_{x=1}^{x=1/f} \tag{7}$$

Elimination of $\mathbf{D}$ from Eq. 1 then leads to the well-known Gurson yield function:

$$\mathbf{\Phi}^{\mathrm{Gurson}}(\mathbf{\Sigma};f) \equiv \frac{\Sigma_{\mathrm{eq}}^2}{\overline{\sigma}^2} + 2f \quad \cosh\left(\frac{3}{2}\frac{\Sigma_{\mathrm{m}}}{\overline{\sigma}}\right) - \left(1+f^2\right) \tag{8}$$

In the limit $f \to 0$ (dense matrix) criterion (8) reduces to the von Mises yield criterion $(5)_1$.

Incorporating Plastic Anisotropy

Benzerga and Besson (2001) generalized the Gurson model to a class of plastically anisotropic solids (the cylindrical case was also treated):

(i) Geometry:
The RVE is the hollow sphere model so that the porosity f is the only void-related microstructural variable entering the model.

(ii) Plasticity model:
The matrix is taken to obey Hill's quadratic associated yield criterion:

$$\sigma_{\mathrm{eq}} \equiv \sqrt{\frac{3}{2}\boldsymbol{\sigma} : \mathbb{p} : \boldsymbol{\sigma}} = \sqrt{\frac{3}{2}\boldsymbol{\sigma}' : \mathbb{h} : \boldsymbol{\sigma}'} = \overline{\sigma}, \quad \mathrm{d} = \frac{3}{2}\frac{d_{\mathrm{eq}}}{\overline{\sigma}}\mathbb{p} : \boldsymbol{\sigma}, \quad d_{\mathrm{eq}} \equiv \sqrt{\frac{3}{2}\mathrm{d} : \hat{\mathbb{h}} : \mathrm{d}} \tag{9}$$

where $\overline{\sigma}$ is the yield stress of the material in some reference direction and

$$\mathbb{p} = \mathbb{J} : \mathbb{h} : \mathbb{J}, \quad \hat{\mathbb{p}} \quad \mathbb{J} : \hat{\mathbb{h}} : \mathbb{J}, \quad \mathbb{p} : \hat{\mathbb{p}} = \hat{\mathbb{p}} : \mathbb{p} = \mathbb{J} \tag{10}$$

Fourth-rank tensors $\mathbb{h}$ and $\hat{\mathbb{h}}$ are symmetric, positive definite.

(iii) Velocity fields:

$$\forall \mathbf{x} \in \Omega\backslash\omega, \quad v_i(\mathbf{x}) = Av_i^A(\mathbf{x}) + \beta_{ij}x_j, \quad \mathbf{v}^A(\mathbf{x}) = \frac{1}{r^2}\mathbf{e}_r \tag{11}$$

where A and $\boldsymbol{\beta}$ ($\beta_{kk} = 0$) are parameters. These are the same velocity fields used by Gurson; cf. Benzerga and Leblond (2010) for a discussion.

An upper bound of the dissipation potential is then

$$\prod(\mathbf{D}) = \overline{\sigma}\left[hD_{\mathrm{m}}\sinh^{-1}\left(\frac{hD_{\mathrm{m}}x}{D_{\mathrm{eq}}}\right) - \sqrt{h^2D_{\mathrm{m}}^2 + \frac{D_{\mathrm{eq}}^2}{x^2}}\right]_{x=1}^{x=1/f} \tag{12}$$

so that Benzerga and Besson's yield function reads

$$\Phi^{\mathrm{BB}}(\Sigma; f, \mathbb{h}) \equiv \frac{3}{2}\frac{\Sigma' : \mathbb{h} : \Sigma'}{\overline{\sigma}^2} + 2f\cosh\left(\frac{3}{h}\frac{\Sigma_{\mathrm{m}}}{\overline{\sigma}}\right) - \left(1 + f^2\right) \tag{13}$$

h being an invariant of tensor $\mathbb{h}$. In axes pointing toward the principal directions of matrix orthotropy, h admits the following expression (Benzerga and Besson 2001):

$$h = 2\left[\frac{2}{5}\frac{h_1 + h_2 + h_3}{h_1h_2 + h_2h_3 + h_3h_1} + \frac{1}{5}\left(\frac{1}{h_4} + \frac{1}{h_5} + \frac{1}{h_6}\right)\right]^{\frac{1}{2}} \tag{14}$$

For an isotropic matrix, $\mathbb{h} = \mathbb{I}$ and $h = 2$ so that the yield function reduces to that of Gurson. For a dense matrix ($f = 0$) criterion (13) reduces to Hill's quadratic criterion.

Quite recently, Stewart and Cazacu (2011) generalized the above model to a class of anisotropic materials exhibiting tension-compression asymmetry, e.g., hexagonal closed-packed polycrystals. The matrix was taken to obey an associated quadratic yield criterion of a general family of non-quadratic criteria (Cazacu et al. 2006):

$$\sigma_{\mathrm{eq}} \equiv \sqrt{(|\hat{\sigma}_i| - k\hat{\sigma}_i)(|\hat{\sigma}_i| - k\hat{\sigma}_i)} = \overline{\sigma}, \quad \hat{\boldsymbol{\sigma}} = \mathbb{L} : \boldsymbol{\sigma}' \tag{15}$$

where $\overline{\sigma}$ as above and $\mathbb{L}$ an invertible tensor invariant with respect to the orthotropy group satisfying major and minor symmetries such that $L_{iikl} = \text{const}$ for $k = l$. Under axisymmetric loadings, their approximate macroscopic yield function takes the same form as Eq. 13 with the quadratic term replaced with $\Sigma_{\mathrm{eq.}}$ defined as in Eq. 15 and a variant of coefficient h in Eq. 14 appears in the exponential term.

Void Shape Effects: Case of Spheroids

Gologanu and coworkers incorporated the anisotropy due to void shape in a series of models: for prolate voids (Gologanu et al. 1993) and for oblate ones (Gologanu et al. 1994). Later, Gologanu et al. (1997) developed an improved model whose general lines are recalled next:

(i) Geometry:
The RVE is a hollow spheroid containing a confocal spheroidal cavity. In addition to the porosity f, the model involves one void aspect ratio, W, and the common void axis, e_3, as microstructural variables.

(ii) Plasticity model:
The isotropic, associated J_2 flow theory is used for the matrix; cf. Eq. 5.

(iii) Velocity fields:

$$\forall \mathbf{x} \in \Omega/\omega, \qquad v_i(\mathrm{x}) = A v_i^A(\mathrm{x}) + \beta_{ij} x_j, \tag{16}$$

where $\mathbf{v}^A$ gives rise to a nonuniform deformation field. It is given by four terms of the axisymmetric expansion field derived by Lee and Mear (1992) involving associated Legendre functions of the first and second kinds. As above, scalar A and symmetric tensor $\boldsymbol{\beta}$ are parameters (with $\beta_{kk} = 0$).

On that basis, an estimate of the dissipation function is obtained in an implicit form. After a series of approximations, the GLD yield function $\boldsymbol{\Phi}^{\mathrm{GLD}}(\boldsymbol{\Sigma}; f, W, \mathrm{e}_3)$ is given by

$$\Phi^{\mathrm{GLD}} = C\frac{\|\Sigma' + \eta\Sigma_h \mathbf{Q}\|^2}{\overline{\sigma}^2} + 2(g+1)(g+f)\cosh\left(\kappa\frac{\boldsymbol{\Sigma}\!:\mathbf{X}}{\overline{\sigma}}\right) - (g+1)^2 - (g+f)^2 \tag{17}$$

Here, $\mathbf{Q}$ and $\mathbf{X}$ are transversely isotropic tensors given by

$$\mathbf{X} \equiv \alpha_2(\mathrm{e}_1 \otimes \mathrm{e}_1 + \mathrm{e}_2 \otimes \mathrm{e}_2) + (1 - 2\alpha_2)\mathrm{e}_3 \otimes \mathrm{e}_3 \tag{18}$$

$$\mathbf{Q} \equiv -\frac{1}{3}(\mathrm{e}_1 \otimes \mathrm{e}_1 + \mathrm{e}_2 \otimes \mathrm{e}_2) + \frac{2}{3}\mathrm{e}_3 \otimes \mathrm{e}_3 \tag{19}$$

$\Sigma_h \equiv \Sigma : \mathbf{X}$ is a weighted average of the normal stresses along the principal axes of the void and e_1, e_2 are arbitrarily chosen transverse unit base vectors. Also, κ, α_2, g, C and n are scalar-valued functions of microstructural parameters f and W. In the limit of a spherical void $W \to 1$, Eq. 17 reduces to Gurson's yield function (8), whereas for $\mathrm{W} \to \infty$ it reduces to Gurson's criterion for cylindrical cavities. The von Mises yield criterion is obtained when setting $f = 0$ for $\mathrm{W} > 1$ (prolate voids). In the case of oblate voids, the limit $f \to 0$ corresponds to a material with a distribution of penny-shaped cracks.

The evolution of porosity is obtained by specializing Eq. 4 to $\Phi = \Phi^{\mathrm{GLD}}$. Void shape evolution is governed by

$$\begin{aligned}\dot{S} = \frac{3}{2}\left[1 + \left(\frac{9}{2} - \frac{T^2 + T^4}{2}\right)\left(1 - \sqrt{f}\right)\frac{\alpha_1 - \alpha_1^{\mathrm{G}}}{1 - 3\alpha_1}\right]\mathrm{e}_3 \cdot \\ \mathbf{D}'^{\mathrm{p}} \cdot \mathrm{e}_3 + \left(\frac{1 - 3\alpha_1}{f} + 3\alpha_2 - 1\right)\mathbf{I} : \mathbf{D}^{\mathrm{p}}\end{aligned} \tag{20}$$

where $S = \ln W$, T is the stress triaxiality ratio and $\alpha_1(f, W)$ and $\alpha_1^{\mathrm{G}}(f, W)$ are given in Appendix A. The evolution of the void axis e_3 is given by

$$\dot{\mathrm{e}}_3 = \mathbf{W} \cdot \mathrm{e}_3 \tag{21}$$

which assumes that the voids rotate with the material, $\mathbf{W}$ being the total material spin. This is clearly an approximation. An improved representation may be found in (Keralavarma and Benzerga 2010) on the basis of earlier work by Kailasam and Ponte Castaneda (1998).

Void Shape Effects: Case of Ellipsoids

More recently, Madou and Leblond (2012a, b) have implemented the homogenization approach of section "Structure of Constitutive Relations" to general ellipsoids:

(i) Geometry:
The RVE is an ellipsoidal volume containing a confocal ellipsoidal void. In addition to the porosity f, the model involves two void aspect ratios, W_1 and W_2, and the common void axes as microstructural variables.
(ii) Plasticity model:
The isotropic, associated J_2 flow theory is used for the matrix; cf. Eq. 5.
(iii) Velocity fields:
The authors used the fields discovered by Leblond and Gologanu (2008) provided in ellipsoidal coordinates and involving elliptic integrals.

The outcome of their analyses is a general yield function whose expression is omitted here. Even more recently, Madou and Leblond (2012a, b) have developed evolution laws for the microstructural variables of the model. They proposed heuristic corrections to the evolution of the void strain-rate and void axes. Their corrections are based on a large series of computationally efficient limit analyses.

Combined Plastic Anisotropy and Void Shape Effects

The homogenization problem combining the two kinds of anisotropies has been addressed by a number of authors in recent years. Thus, Monchiet et al. (2006, 2008) developed a solution based on consideration of the velocity fields used by Gologanu et al. (1993, 1994) in their earlier versions of the GLD model, and Keralavarma and Benzerga (2008) developed an improved solution using the richer Lee-Mear fields used by Gologanu et al. (1997). The latter model is, however, restricted to axisymmetric loadings and microstructures for which the void axis is aligned with one direction of material orthotropy.

Keralavarma and Benzerga (2010) developed a porous plasticity model for materials containing spheroidal voids embedded in a Hill matrix thus generalizing

the GLD model to plastically anisotropic matrices. Their model is also a generalization of Benzerga and Besson's (2001) model accounting for void shape effects:

(i) Geometry:
The RVE is a hollow spheroid containing a confocal spheroidal cavity. Porosity f, void aspect ratio, W, and void axis, e_3, are microstructural variables.
(ii) Plasticity model:
The orthotropic, associated Hill flow theory is used for the matrix; cf. Eq. 9. The orthotropy axes are not necessarily aligned with the voids, L, T, and S referring to the principal directions.
(iii) Velocity fields:

$$\forall \mathrm{x} \in \Omega/\omega, \qquad v_i(\mathbf{x}) = A v_i^A(\mathbf{x}) + \beta_{ij} x_j, \tag{22}$$

where the inhomogeneous part $\mathbf{v}^A$ of Eq. 6_2 is replaced with four terms of the axisymmetric expansion field due to Lee and Mear (1992) involving associated Legendre functions of the first and second kinds. As above, scalar A and symmetric tensor $\boldsymbol{\beta}$ are parameters (with $\beta_{kk} = 0$). Here, $\boldsymbol{\beta}$ is not necessarily axisymmetric if one admits the ensuing approximations.

Their approximate yield function $\Phi^{\mathrm{KB}}(\Sigma; f, W, \mathrm{e}_3, \mathbb{h})$, applicable to non-axisymmetric loadings, reads

$$\boldsymbol{\Phi}^{\mathrm{KB}} = C\frac{3}{2}\frac{\Sigma : \mathbb{H} : \Sigma}{\overline{\sigma}^2} + 2(g+1)(g+f)\cosh\left(\kappa\frac{\Sigma : \mathbf{X}}{\overline{\sigma}}\right) - (g+1)^2 - (g+f)^2 \tag{23}$$

where the macroscopic anisotropy tensor $\mathbb{H}$ is given by

$$\mathbb{H} \quad \mathbb{J} : \mathbb{h} : \mathbb{J} + \eta(\mathbf{X} \quad \mathbf{Q} + \mathbf{Q} \quad \mathbf{X}) \tag{24}$$

Here, $\mathbf{X}$ and $\mathbf{Q}$ are defined as in Eq. 19 and criterion parameters κ, C, and η are scalar-valued functions of microstructural parameters (f and W) and of $\mathbb{h}$, whereas α_2 and g are only functions of f and W; cf. Appendix B.

For example, a simplified expression of κ is

$$\kappa = \begin{cases} \dfrac{3}{h}\left\{1 + \dfrac{h_{\mathrm{t}}}{h^2 \mathrm{In}\, f}\mathrm{In}\dfrac{1-e_2^2}{1-e_1^2}\right\}^{-1/2} & \text{(p)} \\ \dfrac{3}{h}\left\{1 + \dfrac{\left(g_f - g_1\right) + \frac{4}{5}\left(g_f^{5/2} - g_1^{5/2}\right) - \frac{3}{5}\left(g_f^5 - g_1^5\right)}{\mathrm{In}\left(g_f/g_1\right)}\right\}^{-1} & \text{(o)} \end{cases} \tag{25}$$

where (p) and (o) stand for prolate and oblate, respectively, and $g_x = g/(g + x)$. The dependence of the criterion parameters upon anisotropy tensor $\mathbb{h}$ enters through one invariant, h, and two transversely isotropic invariants, h_{t} and h_{q}, of that tensor.

When expressed in the basis associated with the principal directions of orthotropy (in the context of this section, this means replacing indices 1 to 6 in Eq. 14 with L, T, S, TS, SL, and LT, respectively.), invariant h is given by Eq. 14, while h_t and h_q are given by

$$h_t = \frac{1}{5}\left[-\frac{13}{12}\left(\hat{h}_L + \hat{h}_T\right) + \frac{8}{3}\hat{h}_S + 4\left(\hat{h}_{TS} + \hat{h}_{SL}\right) - \frac{7}{2}\hat{h}_{LT}\right] \tag{26}$$

and

$$h_q \quad \frac{2}{3}\mathbf{Q} : \hat{\mathbb{h}} : \mathbf{Q} \tag{27}$$

Here, the $\hat{h}_i$ are the components of $\mathbb{h}$ expressed using Voigt's condensation. h_q only appears in the expressions of C and η (it was denoted $\hat{h}_q$ in Keralavarma and Benzerga (2010)).

In the special case of an isotropic von Mises matrix ($\mathbb{h} = \hat{\mathbb{h}} = \mathbb{I}$), the yield condition (23) reduces to the GLD criterion. In the case of spherical voids in a Hill matrix, one obtains $\lim_{W\to 1} \alpha_2 = 1/3,\ C = 1,\ \eta = 0$ and $\eta = 0$. Also, Eq. 25 reduces to $\kappa^{BB} = 3/h$ and the upper-bound yield criterion of Benzerga and Besson (2001) is recovered. In particular, the Gurson yield function is obtained in the limit of spherical voids in an isotropic matrix since $\mathbb{h} = \mathbb{I}$ implies $\kappa^{BB} = 3/2$. In the limit of cylindrical voids in a Hill matrix with $\mathbf{e}_S = \mathbf{e}_3$, we have $\lim_{W\to\infty} \alpha_2 = 1/2,\ C = 1,\ \eta = 0$ and Eq. 25 reduces to

$$\kappa^{cyl} = \sqrt{3}\left[\frac{1}{4}\frac{h_L + h_T + 4h_S}{h_L h_T + h_T h_S + h_S h_L} + \frac{1}{2h_{LT}}\right]^{-\frac{1}{2}} \tag{28}$$

which is the result obtained by Benzerga and Besson (2001). In particular, the Gurson yield function for cylindrical cavities in a von Mises matrix is recovered with $\kappa^{cyl} = \sqrt{3}$ in that case.

Keralavarma and Benzerga (2010) supplemented yield criterion (23) with evolution laws for the microstructural variables f, W, and the void axis $\mathbf{e}_3$. The first two are in essence similar to those used in the GLD model, but the latter one employs an Eshelby concentration tensor for the spin following a proposal by Kailasam and Ponte Castaneda (1998).

Void Coalescence

If void growth could proceed until failure (complete loss of stress carrying capacity) as modeled in the previous section, then a good estimate of void coalescence would be when the lateral void size (along x_3) has reached the lateral void spacing. The void size relative to its initial value is typically what a void growth model delivers. The current void spacing can directly be inferred from the initial void spacing and deformation history. This approach would lead to a considerable overestimation of

ductility and other fracture properties. This holds even if anisotropic void growth models are employed. To illustrate this, recall that typical values of the critical void growth ratio needed in failure models (Beremin 1981; Johnson and Cook 1985) fall between 1.2 and 2.0. On the other hand, typical values of the ratio of initial void spacing to void size are within the range 10–100, possibly larger. Even after considering the deformation-induced decrease in lateral spacing, an important gap remains. The reason for this is that the void growth models of Section 3 assume that plastic flow takes place in the whole RVE. It is now established that certain modes of localized plastic deformation would deliver lower values of the plastic dissipation Π, hence are more likely to prevail after sufficient microstructural evolution.

In this context, microstructure evolution refers to changes in the geometrical configuration of voids, as described by their relative size and spacing. Void coalescence is an inherently directional void growth process. In all the models presented below, void coalescence is assumed to take place in the x_1–x_2 plane with the major applied normal stress being along the $x3$ direction.

Coalescence Under Predominately Tensile Loads

Thomason's Model

Thomason (1985) posed the following limit-analysis problem:

(i) Geometry:
The RVE is a square-prismatic cell containing a cylindrical void with a square basis, the height of the void being smaller than the cell's height. This geometry is determined by the void aspect ratio W (i.e., the height to breadth ratio), the cell aspect ratio λ, and a relative ligament size χ, which is the ratio of void breadth to cell breadth, the latter representing the void spacing transverse to the major stress.

(ii) Plasticity model:
The isotropic, associated J_2 flow model (5) is used for the matrix, but only in the central region Ω_{lig} containing the intervoid ligament. The regions above and below the void are modeled as rigid.

(iii) Velocity fields: (in the intervoid ligaments only)

$$\forall \mathbf{x} \in \Omega_{\text{lig}}/\omega, \quad \mathbf{v} = \frac{A}{2}\left[\left(\frac{L^2}{x_1} - x_1\right)\mathbf{e}_1 + x_2\left(\frac{L^2}{x_1^2} - 1\right)\mathbf{e}_2 + 2x_3\mathbf{e}_3\right] \tag{29}$$

where A is a constant set by the boundary conditions. The above velocity field gives rise to a state of uniaxial extension of the cell ($D_{11} = D_{22} = 0$ and $D_{33} \neq 0$). As a consequence, the dissipation is only a function of D_{33} so that the yield criterion only depends on Σ_{33}. Note that under such circumstances, the criterion will be insensitive to variations in λ.

Thomason did not solve the above problem in closed form. He obtained numerical solutions to which he proposed an empirical fit. His yield function may be expressed as follows:

$$\boldsymbol{\Phi}^{\text{Thom}}(\Sigma; W, \chi) = \frac{\Sigma_{33}}{\overline{\sigma}} - \left(1 - \chi^2\right)\left(0.1\left(\frac{\chi^{-1} - 1}{W}\right)^2 + 1.2\sqrt{\chi^{-1}}\right) \equiv \frac{\Sigma_{33}}{\overline{\sigma}} - \frac{\Sigma_{33}^{\text{T}}}{\overline{\sigma}} \tag{30}$$

In the limit $\chi \to 1$ the square void fills the ligament and criterion (30) reduces to $\Sigma_{33} = 0$ so that all stress-bearing capacity vanishes. In the limit $W \to 0$ (flat void) $\Sigma_{33}^{\text{T}} \to \infty$ and the criterion is never met. This deficiency is believed to have limited consequences in materials failing after some significant void growth.

Thomason did not supplement his yield criterion with evolution equations for the microstructural variables W and χ. As a consequence, his yield criterion has essentially been used as a criterion for the onset of void coalescence, which is often sufficient to estimate strains to failure as a function of loading parameters, such as the stress-state triaxiality (Lassance et al. 2007). For ductile fracture simulations, however, criterion (30) must be supplemented with evolution laws for W and χ. This task was undertaken by Pardoen and Hutchinson (2000) and Benzerga (2002) who have proposed additional heuristic extensions of the above criterion.

A Complete Void-Coalescence Model

Benzerga (2002) posed the following limit-analysis problem:

(i) Geometry:
The RVE is a cylindrical cell containing a spheroidal void. This geometry is determined by the void aspect ratio W, the cell aspect ratio λ, and the relative ligament size, χ, which is the ratio of void diameter to cell diameter. The latter represents the void spacing transverse to the major stress.

(ii) Plasticity model:
The isotropic, associated J_2 flow model (5) is used for the matrix, but only in the central region Ω_{lig} containing the intervoid ligament. The regions above and below the void are modeled as rigid.

(iii) Velocity fields: (in the intervoid ligaments only)

$$\forall \mathbf{x} \in \Omega_{\text{lig}} \backslash \omega, \qquad v_i(\mathbf{x}) = A v_i^A(\mathbf{x}) + \beta_{ij} x_j, \tag{31}$$

where $\mathbf{v}^A$ contains the same four terms of the axisymmetric Lee-Mear field, Eq. 16. Here, the constants $\boldsymbol{A}$ and $\boldsymbol{\beta}$ are in principle determined by the boundary conditions, which are not of the homogeneous kind.

The above problem is mathematically more involved than Thomason's. Benzerga used some numerical solutions to which he proposed an empirical fit.

The numerical results were taken from Gologanu (1997) who assumed the GLD model to hold in the central porous layer. Benzerga's approximate yield function reads

$$\begin{aligned}\boldsymbol{\Phi}^{\text{Benz}}(\Sigma; W, \chi) &= \frac{\Sigma_{33}}{\overline{\sigma}} - \left(1-\chi^2\right)\left(\alpha\left(\frac{\chi^{-1}-1}{W^2+0.1\chi^{-1}+0.02\chi^{-2}}\right)^2 + \beta\sqrt{\chi-1}\right) \\ &\equiv \frac{\Sigma_{33}}{\overline{\sigma}} - \frac{\Sigma_{33}^{\text{B}}}{\overline{\sigma}}\end{aligned} \tag{32}$$

with $\alpha = 0.1$ and $\beta = 1.3$. This approximation is better than Thomason's for $W < 0.5$ and removes the deficiency in the limit $W \to 0$ (penny-shape crack) since Σ_{33}^{B} admits a finite limit in that case. This correction is believed to have important consequences in materials failing after some limited void growth. Pardoen and Hutchinson (2000) proposed another heuristic extension of criterion (30) in which the factors α and β were taken to vary with the strain-hardening exponent. Such a fit was based on a series of finite-element cell model calculations.

Benzerga (2002) also derived (the velocity fields used in the limit analysis were not used to derive the evolution equations) evolution equations for the state variables W and x on the basis of matrix incompressibility, boundary conditions, and cell model phenomenology. A shape factor γ was introduced in addition to W. The void shape was taken to evolve from spheroidal ($\gamma = 1/2$) at the onset of internal necking ($\chi = \chi_c$) to conical ($\gamma = 1$) at complete coalescence ($\chi = 1$). The evolution equations are as follows:

$$\dot{\chi} = \frac{3}{4}\frac{\lambda}{W}\left[\frac{3\gamma}{\chi^2} - 1\right]D_{\text{eq}} + \frac{\chi}{2\gamma}\dot{\gamma}, \tag{33}$$

$$\dot{W} = \frac{9}{4}\frac{\lambda}{\chi}\left[1 - \frac{\gamma}{\chi^2}\right]D_{\text{eq}} - \frac{W}{2\gamma}\dot{\gamma}, \tag{34}$$

$$\dot{\gamma} = \frac{1}{2(1-\chi_c)}\dot{\chi} \tag{35}$$

where λ represents the current value of the void spacing ratio, which is updated through

$$\dot{\lambda} = \frac{3}{2}\lambda\ D_{\text{eq}}. \tag{36}$$

The void and cell axes were tacitly taken to rotate with the material as per Eq. 21.

Thomason's Model Revisited

Recently, Benzerga and Leblond (2014) have revisited Thomason's analysis by considering a circular cylindrical geometry and a velocity field appropriate for the constrained plastic flow configuration. They obtained a fully analytical expression

for the effective yield criterion. Their closed-form expression can be used instead of Thomason's empirical relation (30). It also constitutes a first step toward some useful generalizations to other geometries and general loadings, which are lacking to date.

Coalescence Under Combined Tension and Shear

Very recently, Tekoglu et al. (2012) have proposed a void-coalescence model applicable under combined shear and tension:

(i) Geometry:
The RVE is a "sandwich" made of three superposed planar layers. Only the central layer contains some porosity. The void shape need not be specified in this model. This geometry is determined by the volume fraction of the porous layer, c, and the porosity within it, f_b. If the geometry is further specified as that considered by Thomason or Benzerga and Leblond (2014), then $f_b = \chi^2$ and $c = W\chi/\lambda$ using the same notations as above.

(ii) Plasticity model:
The isotropic, associated J_2 flow model (5) is used for the matrix but only in the central region Ω_{lig} containing the intervoid ligament. The top and bottom layers are modeled as rigid.

(iii) Velocity fields: (in the intervoid ligaments only)

$$\forall \mathbf{x} \in \Omega_{\text{lig}} \backslash \omega, \quad v_i(\mathbf{x}) = A v_i^A(\mathbf{x}) + \beta_{ij} x_j, \tag{37}$$

where the second field accommodates shear deformation with β being a constant, traceless symmetric tensor. The only nonzero components of β are

$$\beta_{13} = 2cD_{13}; \qquad \beta_{23} = 2cD_{23}$$

Above, v^A is the field that would prevail under pure triaxial tension (no shear). It is not explicitly specified but could be taken as Thomason's field, Eq. 29, if the geometry is made explicit. The constant A is fully determined by the boundary conditions, which are not of the homogeneous kind.

Under such circumstances and without specifying the field v^A, Tekoglu et al. (2012) obtain an approximate yield function of the quadratic type

$$\mathbf{\Phi}^{\text{TLP}}(\Sigma; W, \chi) = \frac{\Sigma_{33}}{\Sigma_{33}^A} + \frac{3\left(\Sigma_{13}^2 + \Sigma_{23}^2\right)}{(1-f_b)^2\overline{\sigma}^2} - 1 \tag{38}$$

where Σ_{33}^A refers to either Σ_{33}^{T} in Eq. 30 or Σ_{33}^{B} in Eq. 32. Hence, in the absence of any shear loading, the criterion (38) reduces to either Eqs. 30 or 32.

Description of Two Integrated Models

GTN Model

The most widely used model of ductile damage is the Gurson-Tvergaard-Needleman (GTN) model. It is based on the Gurson model with some heuristic, often micromechanically motivated extensions to incorporate hardening and viscous flow, void interactions, void nucleation, and void coalescence. Within a convective representation of finite deformation viscoplasticity, additive decomposition of the total rate of deformation **D** is assumed with the plastic part $\mathbf{D}^{\mathrm{p}}$ obtained from the flow potential (Gurson 1977; Pan et al. 1983)

$$\mathcal{F}^{\mathrm{GTN}} = \frac{\Sigma_{\mathrm{eq}}^2}{\overline{\sigma}^2} + 2q_1 f^* \cosh\left(\frac{3q_2\Sigma_{\mathrm{m}}}{2\overline{\sigma}}\right) - 1 - (q_1 f^*)^2 = 0 \tag{39}$$

by assuming equality of macroscopic plastic work rate and matrix dissipation

$$\mathbf{D}^{\mathrm{p}} = \left[\frac{(1-f)\overline{\sigma}\,\dot{\overline{\epsilon}}}{\boldsymbol{\Sigma} : \dfrac{\partial \mathcal{F}}{\partial \boldsymbol{\Sigma}}}\right] \frac{\partial \mathcal{F}}{\partial \boldsymbol{\Sigma}} \tag{40}$$

Here, $\overline{\sigma}$ is the matrix flow strength and q_1 and q_2 are parameters introduced by Tvergaard (1981). The function $f^*(f)$ was introduced by Tvergaard and Needleman (1984) to account for the effects of rapid void coalescence at failure

$$f^* = \begin{cases} f & f < f_c \\ f_c + \left(f_u^* - f_c\right)(f - f_c)/\left(f_f - f_c\right) & f \geq f_c \end{cases} \tag{41}$$

The constant $f_u^* = 1/q_1$ is the value of f^* at zero stress. As $f \to f_f$ and $f^* \to f_u^*$, the material loses all stress carrying capacity. Equation 41 is a phenomenological description of coalescence involving two parameters, f_c and f_f, which are both material dependent and stress-state dependent (Koplik and Needleman 1988). Appropriate values for f_c and f_f can be derived using predictive micromechanical models (Benzerga et al. 1999; Benzerga 2002), which assume that coalescence occurs through an internal necking mechanism. Strain-rate effects can be accounted for through a relation $\dot{\overline{\epsilon}}(\overline{\sigma}, \overline{\epsilon})$, e.g., Benzerga et al. (2002b), where $\dot{\overline{\epsilon}}$ is the effective strain rate and $\overline{\epsilon} = \int \dot{\overline{\epsilon}}\, \mathrm{d}t$ is the effective plastic strain.

To account for void nucleation, the rate of increase of the void volume fraction is given by

$$\dot{f} = \dot{f}_{\mathrm{growth}} + \dot{f}_{\mathrm{nucleation}} \tag{42}$$

where the first term accounts for the growth of existing voids through Eq. 4 and the second term represents the contribution from void nucleation. For example, the

nucleation of voids by a strain controlled mechanism is modeled using (Chu and Needleman 1980)

$$\dot{f}_{\text{nucleation}} = \mathcal{D}\dot{\overline{\epsilon}} \tag{43}$$

with

$$\mathcal{D} = \frac{f_N}{s_N\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\overline{\epsilon}-\epsilon_N}{s_N}\right)^2\right] \tag{44}$$

Proposed Model

Within a finite deformation framework, a corotational formulation of the constitutive equations can be used (Benzerga et al. 2004). The total rate of deformation $\mathbf{D}$ is written as the sum of an elastic and a plastic part:

$$\mathbf{D} = \mathbf{D}^{\text{e}} + \mathbf{D}^{\text{p}} \tag{45}$$

Elasticity is included through a hypoelastic law:

$$\mathbf{D}^{\text{e}} = \mathbb{C}^{-1} : \dot{\mathbf{P}} \tag{46}$$

where $\mathbb{C}$ is the rotated tensor of elastic moduli and $\mathbf{P}$ is the rotated stress:

$$\mathbf{P} = J\boldsymbol{\Omega}^T \cdot \Sigma \cdot \boldsymbol{\Omega} \tag{47}$$

Here, $\boldsymbol{\Omega}$ is an appropriate rotation tensor; it is identified with the rotation $\mathbf{R}$ resulting from the polar decomposition of the deformation gradient $\mathbf{F}$ if the Green-Naghdi rate of Σ is used and $\dot{\boldsymbol{\Omega}} \cdot \boldsymbol{\Omega}^T = \mathbf{W}$ if the Jaumann rate is used, $\mathbf{W}$ being the spin tensor. Also, $J = \det \mathbf{F}$.

To account for rate dependence, the plastic part of the rate of deformation, $\mathbf{D}^{\text{p}}$, is obtained by normality from the gauge function:

$$\phi = \sigma_\star - \overline{\sigma}(\overline{\varepsilon}) \tag{48}$$

where $\overline{\sigma}$ is the matrix flow stress, $\overline{\varepsilon}$ is the effective plastic strain, and $\sigma_\star$ is an effective matrix stress which is implicitly defined through an equation of the type $\mathcal{F}(\Sigma; \text{ISVs}, \sigma_\star) = 0$ where "ISVs" refers to a collection of internal state variables with a definite microstructural significance. For a rate-independent material (standard plasticity), plastic flow occurs for $\phi = 0$ and $\phi = 0$. For a rate-dependent material (viscoplasticity), one has $\phi > 0$ during plastic flow.

Prior to the onset of void coalescence, the potential $\mathcal{F}$ admits an expression of the type $\mathbb{h}$ where the relevant ISVs are the porosity f, the void aspect ratio W, the void axis $\mathbf{e}_3$, and Hill's tensor $\mathbb{h}$, which describes the effect of the current texture. With the KB model as reference (cf. Eq. 23), the following expression may be used:

$$\mathcal{F}^{(c-)} = C\frac{3}{2}\frac{\Sigma : \mathbb{H} : \Sigma}{\sigma_\star^2} + 2q_w(g+1)(g+f)\cosh\left(\kappa\frac{\Sigma : \mathbf{X}}{\sigma_\star}\right) - (g+1)^2 - q_w^2(g+f)^2 \tag{49}$$

where $\mathbb{H}$ is given by Eq. 24, $\mathbf{X}$ and $\mathbf{Q}$ by Eq. 19, κ by Eq. 25, h by Eq. 14, and h_t by Eq. 26 and the remaining criterion parameters $C(f, W, \mathbb{h})$, $\eta(f, W, \mathbb{h})$, $\alpha_2(f, W)$, and $g(f, W)$ are given in Appendix B. Also, q_W is a heuristic void-shape-dependent factor that was determined by Gologanu et al. (1997) to fit unit-cell results:

$$q_w = 1 + (q-1)/\cosh S \tag{50}$$

where $q = 1.6$ is the value taken by q_w for a spherical void. The evolution laws of f, W and e_3 are given by Eqs. 4, 20, and 21, respectively, where $\mathbf{D}$ should be replaced with $\mathbf{D}^\mathrm{p}$ and $\mathbf{W}$ with $\dot{\mathbf{\Omega}}\cdot\mathbf{\Omega}^T$, $\mathbf{\Omega}$ being the rotation tensor used in Eq. 47.

After the onset of void coalescence, the flow potential is given by

$$\mathcal{F}^{(c+)}(\Sigma, \chi, W, \mathbb{h}, \sigma_\star) = \frac{||\Sigma||_\mathrm{H}}{\sigma_\star} + \frac{1}{2}\frac{|\mathbf{I} : \Sigma|}{\sigma_\star} - \frac{3}{2}\Sigma_{33}^\mathrm{B}(\chi, W) \tag{51}$$

where Σ_{33}^B is given by Eq. 32. For an arbitrary void shape between a spheroid and a cone, χ is exactly related to the void spacing ratio, λ, through a shape factor γ as

$$\chi = \begin{cases} \left[3\gamma\frac{f}{W}\lambda\right]^{1/3} & (\mathrm{P}) \\ W\left[3\gamma\frac{f}{W}\lambda\right]^{1/3} & (\mathrm{T}) \end{cases} \tag{52}$$

where (P) and (T) are a shorthand notation for parallel and transverse loading, respectively. As $\chi \to 1$ the material loses all stress carrying capacity. At the onset of coalescence, we have $\mathcal{F}^{(\mathrm{c}-)} = \mathcal{F}^{(\mathrm{c}+)} = 0$. The evolution laws of the microstructural variables are given by Eqs. 33, 34, 35, and 36 along with Eq. 21.

A variant of the integrated model was implemented in a finite-element code and used to model ductile fracture in notched bars (Benzerga et al. 2004) and slant fracture in plane strain (Benzerga et al. 2002a). A fully implicit time integration procedure was used for the local behavior in conjunction with an iterative Newton-Raphson method. The consistent tangent matrix was computed as detailed in the case of prolate voids by Benzerga et al. (2002a).

Identification of Material Parameters

GTN Model

The following parameters enter the GTN model: q_1, q_2, f_c, f_f, f_N, s_N and ϵ_N. There is no straightforward procedure for identifying these parameters using standard experiments. In addition, some of the above parameters play interdependent roles.

In practice, the first four of these parameters could be fixed based on micromechanical models or cell model calculations. Other model parameters pertain to the elastoplastic behavior of the matrix. The hardening response of the matrix is determined using uniaxial testing with appropriate large-strain corrections.

Proposed Model

One advantage of using the proposed model is that it involves fewer parameters, void nucleation set aside. Examination of the constitutive equations of section "Proposed Model" reveals actually no adjustable fracture parameter, unless the value of $q = 1.6$ is treated as a free parameter. When this model is used to predict crack initiation, only deformation-related parameters need to be calibrated on experiments. Account for plastic anisotropy may be necessary depending on the material. Anisotropic response is common in metal forming applications although it is often restricted to two-dimensional measurements. Within the confines of the integrated model above, this first step will deliver the basic hardening curve $\overline{\sigma}(\overline{\varepsilon})$ as well as the anisotropy tensor $\mathbb{h}$.

With the basic flow properties of the matrix calibrated, the next step is to determine the volume fraction, aspect ratio and relative spacing of damage initiation sites (inclusions, precipitates, etc.) in the average sense. Practically, this can be achieved by examining three perpendicular cross sections in optical microscopy, carrying out the needed two-dimensional measurements using digital image analysis, and finally operating standard stereology transformations to infer their 3D counterparts. The outcome of this step is the set of parameters f_0, W_0, and λ_0 needed to initialize the state of the microstructure in constitutive Eqs. 48, 49, 50, 51, 52, 33, 34, 35, and 36. Other details on how to account for 3D aspects and void nucleation may be found in the review by Benzerga and Leblond (2010).

How to Use the Model

The proposed model as well as the GTN model may be used to model the initiation of a crack in an initially crack-free specimen or to model crack growth. In a first step, it is recommended to assume that void nucleation is instantaneous and occurs at a fixed value of the effective strain. Therefore, in simulations of ductile fracture, voids are considered to be present from the outset of plastic deformation. With all deformation-related quantities calibrated on experiments, model predictions can be compared with a sufficiently discriminating set of experiments (notched bars, CT specimens, plane strain bars, etc.) If the model predicts larger than measured fracture properties, one should attempt a threshold strain for nucleation. The hypothesis of a delayed or continuous nucleation could then be checked by metallographic examinations. If the latter do not corroborate the hypothesis, a parameter sensitivity analysis should be carried out by varying the initial values of microstructural parameters.

When failure by void coalescence is predicted at a material point, the material loses its stress-bearing capacity at this point. In a finite-element simulation, it is typical to represent this in terms of an element-vanish technique whereby the stresses and stiffness in the element are subsequently disregarded. This procedure allows computational simulations of crack growth. Obtaining mesh-independent predictions requires a length scale to be incorporated in the problem formulation. Several approaches with varying levels of refinement have been developed to that end but remain underutilized in fracture simulations for metal forming applications.

Acknowledgments This research was supported by NPRP grant No 4-1411-2-555 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the author. Partial support from the National Science Foundation (Grant Number DMR-0844082) is gratefully acknowledged.

Appendix A. GLD Criterion Parameters

There are six parameters which depend on the microstructural variables f and w: C, g, κ, η, and α_2, listed by order of appearance in criterion (17), and α_1, which mainly appears in the evolution law of w:

$$g = 0 \quad \text{(p)}; \qquad g = \frac{e_2^3}{\sqrt{1-e_2^2}} = f\frac{e_1^3}{\sqrt{1-e_1^2}} = f\frac{(1-w^2)^{\frac{3}{2}}}{w}\,\text{(o)} \tag{53}$$

where (p) and (o) are a shorthand notation for prolate and oblate, respectively. We recall that e_1 and e_2 are the eccentricities of the void and the outer boundary of the RVE, respectively. Both are implicit functions of f and w.

$$\kappa = \begin{cases} \left[\dfrac{1}{\sqrt{3}} + \dfrac{1}{\ln f}\left(\left(\sqrt{3}-2\right)\ln\dfrac{e_1}{e_2}\right)\right]^{-1} & \text{(p)} \\ \dfrac{3}{2}\left[1 + \dfrac{\left(g_f - g_1\right) + \frac{4}{5}\left(g_f^{5/2} - g_1^{5/2}\right) - \frac{3}{5}\left(g_f^5 - g_1^5\right)}{\ln\frac{g_f}{g_1}}\right]^{-1} & \text{(o)} \end{cases} \tag{54}$$

where

$$g_f \equiv \frac{g}{g+f}, \qquad g_1 \equiv \frac{g}{g+1}$$

$$\alpha_2 = \begin{cases} \dfrac{\left(1+e_2^2\right)}{\left(1+e_2^2\right)^2 + 2\left(1-e_2^2\right)^2} & \text{(p)} \\ \dfrac{\left(1-e_2^2\right)\left(1-2e_2^2\right)}{\left(1-2e_2^2\right)^2 + 2\left(1-e_2^2\right)} & \text{(o)} \end{cases} \tag{55}$$

$$\eta = -\frac{2}{3}\frac{\kappa Q^*(g+1)(g+f)\mathrm{sh}}{(g+1)^2+(g+f)^2+(g+1)(g+f)[\kappa H^*\mathrm{sh}-2\mathrm{ch}]},$$
$$C = -\frac{2}{3}\frac{\kappa(g+1)(g+f)\mathrm{sh}}{\left(Q^*+\frac{3}{2}\eta H^*\right)\eta}, \qquad \mathrm{sh}\equiv\sinh(\kappa H^*), \qquad \mathrm{ch}\equiv\cosh(\kappa H^*) \tag{56}$$

where $H^* = 2(\alpha_1 - \alpha_2)$ and $Q^* \equiv (1-f)$.

$$\alpha_1 = \begin{cases} \left[e_1 - \left(1-e_1^2\right)\tanh^{-1}e_1\right]/\left(2e_1^3\right) & \text{(p)} \\ \left[-e_1\left(1-e_1^2\right) + \sqrt{1-e_1^2}\sin^{-1}e_1\right]/\left(2e_1^3\right) & \text{(o)} \end{cases} \tag{57}$$

Finally, the parameter α_1^{G} which enters the evolution law (20) of the void shape parameter is given by

$$\alpha_1^{\mathrm{G}} = \begin{cases} 1/\left(3-e_1^2\right) & \text{(p)} \\ \left(1-e_1^2\right)/\left(3-2e_1^2\right) & \text{(o)} \end{cases} \tag{58}$$

Appendix B. KB Criterion Parameters

There are six parameters which depend on the microstructural variables f and w and on the anisotropy tensor $\mathbb{h}$: C, g, κ, η, and α_2, listed by order of appearance in criterion (23) and α_1, which appears in the evolution law of W:

$$g = 0 \quad \text{(p)}; \qquad g = \frac{e_2^3}{\sqrt{1-e_2^2}} = f\frac{e_1^3}{\sqrt{1-e_1^2}} = f\frac{(1-w^2)^{\frac{3}{2}}}{w} \quad \text{(o)} \tag{59}$$

We recall that e_1 and e_2 are the eccentricities of the void and the outer boundary of the RVE, respectively. Both are implicit functions of f and w. Next, the full expression of κ was provided by Keralavarma and Benzerga (2010) but can be simplified into

$$\kappa = \begin{cases} \dfrac{3}{h}\left\{1+\dfrac{h_{\mathrm{t}}}{h^2\mathrm{In}f}\mathrm{In}\dfrac{1-e_2^2}{1-e_1^2}\right\}^{-1/2} & \text{(p)} \\ \dfrac{3}{h}\left\{1+\dfrac{\left(g_f-g1\right)+\frac{4}{5}\left(g_f^{5/2}-g_1^{5/2}\right)-\frac{3}{5}\left(g_f^5-g_1^5\right)}{\mathrm{In}\left(g_f/g_1\right)}\right\}^{-1} & \text{(o)} \end{cases} \tag{60}$$

where shorthand notations are used for

$$g_f \equiv \frac{g}{g+f}, \qquad g_1 \equiv \frac{g}{g+1}$$

$$\alpha_2 = \begin{cases} \dfrac{\left(1+e_2^2\right)}{\left(1+e_2^2\right)^2 + 2\left(1-e_2^2\right)} & \text{(p)} \\ \dfrac{\left(1-e_2^2\right)\left(1-2e_2^2\right)}{\left(1-2e_2^2\right)^2 + 2\left(1-e_2^2\right)} & \text{(o)} \end{cases} \tag{61}$$

$$\begin{aligned} \eta &= -\frac{2}{3h_{\mathrm{q}}} \frac{\kappa Q^*(g+1)(g+f)\mathrm{sh}}{(g+1)^2 + (g+f)^2 + (g+1)(g+f)[\kappa H^* \mathrm{sh} - 2\mathrm{ch}]}, \\ C &= -\frac{2}{3} \frac{\kappa(g+1)(g+f)\mathrm{sh}}{\left(Q^* + \dfrac{3}{2} h_{\mathrm{q}} \eta H^*\right)\eta}, \qquad \mathrm{sh} \equiv \sinh(\kappa H^*), \qquad \mathrm{ch} \equiv \cosh(\kappa H^*) \end{aligned} \tag{62}$$

where $H^* \equiv 2\sqrt{h_{\mathrm{q}}}(\alpha_1 - \alpha_2)$ and $Q^* \equiv \sqrt{h_{\mathrm{q}}}(1-f)$.

$$\alpha_1 = \begin{cases} \left[e_1 - \left(1-e_1^2\right)\tanh^{-1} e_1\right] / \left(2e_1^3\right) & \text{(p)} \\ \left[-e_1\left(1-e_1^2\right) + \sqrt{1-e_1^2}\sin^{-1} e_1\right] / \left(2e_1^3\right) & \text{(o)} \end{cases} \tag{63}$$

Note that the expressions of α_2 and α_1 are identical to those given by Gologanu et al. (1997) for isotropic matrices.

References

A.A. Benzerga, Micromechanics of coalescence in ductile fracture. J. Mech. Phys. Solids **50**, 1331–1362 (2002)

A.A. Benzerga, J. Besson, Plastic potentials for anisotropic porous solids. Eur. J. Mech. **20A**, 397–434 (2001)

A.A. Benzerga, J. Besson, R. Batisse, A. Pineau, Synergistic effects of plastic anisotropy and void coalescence on fracture mode in plane strain. Model. Simul. Mater. Sci. Eng. **10**, 73–102 (2002a)

A.A. Benzerga, J. Besson, A. Pineau, Coalescence – controlled anisotropic ductile fracture. J. Eng. Mater. Tech. **121**, 221–229 (1999)

A.A. Benzerga, J. Besson, A. Pineau, Anisotropic ductile fracture. Part II: theory. Acta Mater. **52**, 4639–4650 (2004)

A.A. Benzerga, J.-B. Leblond, Ductile fracture by void growth to coalescence. Adv. Appl. Mech. **44**, 169–305 (2010)

A.A. Benzerga, J.-B. Leblond, Effective yield criterion accounting for microvoid coalescence. J. Appl. Mech. 81, 031009 (2014)

A.A. Benzerga, V. Tvergaard, A. Needleman, Size effects in the Charpy V-notch test. Int. J. Fract. **116**, 275–296 (2002b)

F.M. Beremin, Experimental and numerical study of the different stages in ductile rupture: application to crack initiation and stable crack growth, in *Three-Dimensional Constitutive relations of Damage and Fracture*, ed. by S. Nemat-Nasser (Pergamon press, North Holland, 1981), pp. 157–172

J. Besson, Continuum models of ductile fracture: a review. Int. J. Damage Mech. **19**, 3–52 (2010)

O. Cazacu, B. Plunkett, F. Barlat, Orthotropic yield criterion for hexagonal closed packed metals. Int. J. Plasticity **22**, 1171–1194 (2006)

C. Chu, A. Needleman, Void nucleation effects in biaxially stretched sheets. J. Eng. Mater. Technol. **102**, 249–256 (1980)
M. Gologanu, Etude de quelques problèmes de rupture ductile des métaux. Ph.D. thesis, Université Paris 6, 1997
M. Gologanu, J.-B. Leblond, J. Devaux, Approximate models for ductile metals containing non-spherical voids – case of axisymmetric prolate ellipsoidal cavities. J. Mech. Phys. Solids **41**(11), 1723–1754 (1993)
M. Gologanu, J.-B. Leblond, J. Devaux, Approximate models for ductile metals containing non-spherical voids – case of axisymmetric oblate ellipsoidal cavities. J. Eng. Mater. Technol. **116**, 290–297 (1994)
M. Gologanu, J.-B. Leblond, G. Perrin, J. Devaux, Recent extensions of Gurson's model for porous ductile metals, in *Continuum Micromechanics*, ed. by P. Suquet. CISM Lectures Series (Springer, New York, 1997), pp. 61–130
A.L. Gurson, 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**, 2–15 (1977)
G.R. Johnson, W.H. Cook, Fracture characteristics of three metals subjected to various strains, strain rates, temperatures and pressures. Eng. Fract. Mech. **21**(1), 31–48 (1985)
M. Kailasam, P. Ponte Castaneda, A general constitutive theory for linear and nonlinear particulate media with microstructure evolution. J. Mech. Phys. Solids **46**(3), 427–465 (1998)
S.M. Keralavarma, A.A. Benzerga, An approximate yield criterion for anisotropic porous media. Comptes Rendus Mecanique **336**, 685–692 (2008)
S.M. Keralavarma, A.A. Benzerga, A constitutive model for plastically anisotropic solids with non-spherical voids. J. Mech. Phys. Solids **58**, 874–901 (2010)
J. Koplik, A. Needleman, Void growth and coalescence in porous plastic solids. Int. J. Solids. Struct. **24**(8), 835–853 (1988)
D. Lassance, D. Fabrègue, F. Delannay, T. Pardoen, Micromechanics of room and high temperature fracture in 6xxx Al alloys. Progr. Mater. Sci. **52**, 62–129 (2007)
J.-B. Leblond, M. Gologanu, External estimate of the yield surface of an arbitrary ellipsoid containing a confocal void. Compt Rendus Mecanique **336**, 813–819 (2008)
B.J. Lee, M.E. Mear, Axisymmetric deformation of power–law solids containing a dilute concentration of aligned spheroidal voids. J. Mech. Phys. Solids **40**(8), 1805–1836 (1992)
K. Madou, J.-B. Leblond, A Gurson-type criterion for porous ductile solids containing arbitrary ellipsoidal voids – I: limit-analysis of some representative cell. J. Mech. Phys. Solids **60**, 1020–1036 (2012a)
K. Madou, J.-B. Leblond, A Gurson-type criterion for porous ductile solids containing arbitrary ellipsoidal voids – II: Determination of yield criterion parameters. J. Mech. Phys. Solids **60**, 1037–1058 (2012b)
K. Madou, J.-B. Leblond, Numerical studies of porous ductile materials containing arbitrary ellipsoidal voids – I: Yield surfaces of representative cells. Eur. J. Mech. **42**, 480–489 (2013)
K. Madou, J.-B. Leblond, Numerical studies of porous ductile materials containing arbitrary ellipsoidal voids – II: Evolution of the magnitude and orientation of the void axes. Eur. J. Mech. Volume **42**, 490–507 (2013)
V. Monchiet, O. Cazacu, E. Charkaluk, D. Kondo, Macroscopic yield criteria for plastic anisotropic materials containing spheroidal voids. Int. J. Plasticity **24**, 1158–1189 (2008)
V. Monchiet, C. Gruescu, E. Charkaluk, D. Kondo, Approximate yield criteria for anisotropic metals with prolate or oblate voids. Comptes Rendus Mecanique **334**, 431–439 (2006)
J. Pan, M. Saje, A. Needleman, Localization of deformation in rate sensitive porous plastic solids. Int. J. Fracture **21**, 261–278 (1983)
T. Pardoen, J.W. Hutchinson, An extended model for void growth and coalescence. J. Mech. Phys. Solids **48**, 2467–2512 (2000)
J.B. Stewart, O. Cazacu, Analytical yield criterion for an anisotropic material containing spherical voids and exhibiting tension-compression asymmetry. Int. J. Solids. Struct. **48**, 357–373 (2011)

C. Tekoglu, J.-B. Leblond, T. Pardoen, A criterion for the onset of void coalescence under combined tension and shear. J. Mech. Phys. Solids **60**, 1363–1381 (2012)
P.F. Thomason, Three-dimensional models for the plastic limit-loads at incipient failure of the intervoid matrix in ductile porous solids. Acta Metall. **33**, 1079–1085 (1985)
V. Tvergaard, Influence of voids on shear band instabilities under plane strain conditions. Int. J. Fracture **17**, 389–407 (1981)
V. Tvergaard, A. Needleman, Analysis of the cup-cone fracture in a round tensile bar. Acta Metall. **32**, 157–169 (1984)

Micromechanical Polycrystalline Damage-Plasticity Modeling for Metal Forming Processes

28

Benoit Panicaud, Léa Le Joncour, Neila Hfaiedh, and Khemais Saanouni

Contents

Abstract

This chapter deals with the presentation of micromechanical modeling of the elastoplastic material behavior exhibiting ductile damage together with

B. Panicaud (✉) • L. Le Joncour • N. Hfaiedh
ICD/LASMIS, STMR UMR-CNRS 6279, University of Technology of Troyes, Troyes Cedex, France
e-mail: benoit.panicaud@utt.fr; lea.le_joncour@utt.fr; neilahfaiedh@yahoo.fr

K. Saanouni
ICD/LASMIS, STMR UMR CNRS 6281, University of Technology of Troyes, Troyes, France
e-mail: khemais.saanouni@utt.fr

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_40

microstructural evolution in terms of grain rotation and phase transformation, under large inelastic strains. A description of the main experimental methods is proposed and multiscale measurements are discussed. For the mesoscopic scale, diffraction techniques are presented as well as microscopy's results for a specific material. For the macroscopic scale, techniques of tensile test coupled with digital image correlation are described. This allows the damage measurement at different scales. Micromechanical modeling aspects based on the thermodynamics of irreversible processes with state variables defined at different scales are discussed. A non-exhaustive review of several possible models is given. These models depend on the hypothesis for the energy or strain equivalence and on the smallest scale considered. Two particular models are then detailed with their associated constitutive equations and the corresponding numerical aspects. Application is made to two different materials to test the ability of the model to be used for metal forming simulations.

Abbreviations and Notations

CDM	Continuous damage mechanics
CRSS	Critical resolved shear stress
DIC	Digital image correlation
DSS	Duplex stainless steel(s)
EBSD	Electron backscatter diffraction
FEA	Finite element analysis
ODF	(Crystalline) orientation distribution function
RVE	Representative volume element
SEM	Scanning electron microscopy
X	Zero-rank tensor = scalar variable
XRD	X-ray diffraction
$\vec{X}$	One-rank tensor = vector variable
$\underline{X}$	Second-rank tensor
$\underline{\underline{X}}$	Fourth-rank tensor
$\underline{X} \cdot \underline{Y}$	Contraction between the second-rank tensors $\underline{X}$ and $\underline{Y}$
$\underline{X} : \underline{Y}$	Double contraction between the second-rank tensors $\underline{X}$ and $\underline{Y}$
$\underline{X} \otimes \underline{Y}$	Tensorial product between the second-rank tensors $\underline{X}$ and $\underline{Y}$
$\langle\langle X \rangle\rangle$	Macaulay brackets which means the positive part of a scalar X
$(\underline{X})^T$ or $\left(\underline{\underline{X}}\right)^T$	Transpose of X (second-rank or fourth-rank tensor)
$\|\underline{X}\| = \sqrt{\underline{X} : \underline{X}/3}$	Euclidean norm of a second-rank tensor $\underline{X}$
$\left\|\vec{X}\right\| = \sqrt{\vec{X} \cdot \vec{X}}$	Euclidean norm of a vector $\vec{X}$
$\langle x \rangle$	Average of the quantity x
	Capital letters are for macroscopic or part quantities, whereas minuscule letters are for mesoscopic or microscopic ones.

Introduction

As described in the introduction Ductile damage in metal forming: Advanced macroscopic modeling and numerical simulation, it is an undisputable fact that metallic materials, when formed at room or high temperature, exhibit large inelastic strains. These strains correspond to the observable consequence of various physical phenomena at the origin of high material nonlinearities that strongly depend on the microstructure of the materials. It is then necessary to consider phenomena such as plastic flow, hardening of various types, initial and induced anisotropies, textural evolutions, micro-defects initiation growth and coalescence (ductile damage), or phase transformations. In most cases, forming processes are further associated to severe environmental conditions (high temperature, important loading velocity, severe chemical conditions, etc.). From a modeling point of view, it seems thus essential to consider the initial microstructure of the material, its evolution under loading conditions, as well as its effects on the inelastic flow.

The natural way to account for microstructural evolutions in relation with material inelastic flow is the use of the so-called multiscale approach, briefly introduced in §2.2.2 of Chapter 1 and schematized in Fig. 4 of that chapter. From a theoretical point of view, a considerable number of articles have been dedicated to the micromechanical description of the metallic material deformation, based on the microstructural composition of each material. The reader may refer to several books representative of the work done in this field (Schmid and Boas 1968; Bunge 1982; Mura 1987; Havner 1992; Nemat-Nasser and Hori 1993; Yang and Lee 1993; Kocks et al. 1998; Raabe 1998; Bornert et al. 2001; Gambin 2001; Nemat-Nasser 2004; Asaro and Lubards 2006). Other references are specifically dedicated to the modeling and simulation of metal forming processes using the multiscale constitutive equations, for example, Sellard (1990), Chen et al. (1992), Kalidindi et al. (1992), Shercliff and Lovatt (1999), Furu et al. (1999), Gottstein et al. (2000), Zhu and Sellars (2000), Duan and Sheppard (2003), Grugicic and Batchu (2002), Dawson et al. (2003), Wilkinson et al. (2006), Cho and Dawson (2008), Boudifa et al. (2009), Hfaiedh et al. (2009), Inal et al. (2010), and Nedoushan et al. (2012).

This chapter is dedicated to the presentation of micromechanical modeling for transformation exhibiting large inelastic strains. The main mechanical phenomena taken into account are ductile damage and microstructural evolution in terms of grain rotation, while other second-order mechanisms are forsaken, for example, phase transformation phenomena. In the first Experimental aspects of this chapter, a description of the main experimental methods is proposed and multiscale measurements are discussed. For the mesoscopic scale, diffraction techniques are presented as well as microscopy's results, for a specific material (UR45N). For the macroscopic scale, techniques of tensile test coupled with digital image correlation are described. This allows the evaluation of damage at different scales and, thus, leads to a damage scenario that can be used to guide the modeling approach.

The second section is devoted to the micromechanical modeling aspects. The principles and fundamentals on which are based the formulation of constitutive

equations for metallic materials are briefly reviewed, and assumptions on micromechanics are given. It is worth noting that the present micromechanical modeling is also based on thermodynamics of irreversible processes with state variables introduced at different scales. This micromechanical approach is also described and compared with the macroscopic approach (see Chapter 1 of this volume). Some requirements on finite transformations are highlighted. Different mechanical constitutive models are introduced. Simple examples are presented for elastoplasticity, in order to illustrate the previous approach. Ductile damage is then introduced in the framework of the Continuous Damage Mechanics (CDM) approach. Modeling metal forming processes with a micromechanical approach while accounting for damage is not a simple task, due to the lack of fine experimental observations. A non-exhaustive review of several possible models is given. These models depend on the hypothesis for the energy or strain equivalence and on the smallest scale considered. Two particular models are then detailed with their associated constitutive equations and the corresponding numerical aspects.

Two particular materials are also considered. First, copper is studied to illustrate the possibility offered by the proposed approach. Numerical results are also given to test the ability of the model to be used for metal forming simulations. Second, a duplex stainless steel (UR45N) is investigated. Such a biphasic material is interesting as it allows the illustration of both the experimental and numerical approaches.

Experimental Aspects

Multiscale Measurements

Different techniques can be used to measure mechanicals fields (strains, damages, etc.). In the present chapter, the use of micromechanics implies that quantities have to be considered with experiments at the appropriate scales. It is thus necessary to define the different required scales for the considered metallic materials:

- The microscopic scale corresponds to slip systems.
- The mesoscopic scale corresponds to a grain (i.e., one crystallite).
- If an M-phase material is considered, a pseudo-macroscopic scale can be defined that represents each group of grains of the same phase M (M is an integer).
- The macroscopic scale is a RVE, corresponding to an aggregate of grains (generally hundreds to several thousands of grains, depending on the grain size).
- Then, taking account of the geometry of the structure, there is the part scale.

In this paragraph, both the micro-/mesoscopic scales and macroscopic/part scales are investigated considering coupled measurements (diffraction techniques and tensile tests). However, due to the finite strains that occur, some specific treatments and analyses are developed.

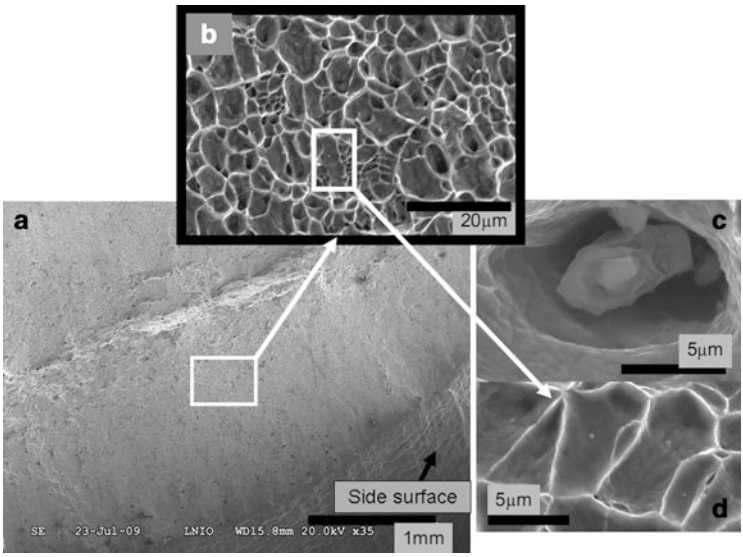


Fig. 1 Picture of the fracture surface of a DSS sample by SEM (Le Joncour et al. 2011)

The presence of damage leads to difficulties, mainly related to the softening stage that follows the hardening stage. Indeed, the measure of damage is still a largely open problem (Montheillet and Moussy 1988; François 2004). It is first necessary to agree on the definition of ductile damage. It is considered here that damage corresponds to any fault/deviation from the crystallinity of the metallic material (excluding plasticity). For example, microcracks and micro-voids will be considered as ductile damage. Precipitates will not, as long as they retain consistency and adherence with the embedded matrix. For simplicity, it is considered that there is damage when a fracture surface occurs between the microscopic atomic groups. This geometric definition of damage leads to consider experimentally damage as the density of "voids" that is present in the RVE, at a given time. For example, a qualitative geometric measure of the damage can be done simply by electron microscopy, as shown in Fig. 1. A more direct approach in 3D is possible thanks to the recent development of X-ray tomography applied to metallic materials (Maire et al. 2005).

The difficulty encountered with this kind of techniques is that it gives a snapshot of the geometry of damage, without necessarily considering the physical and/or kinetics aspects of the damage. For example, interaction between micro-voids or microcracks may exist but are not visible. This interaction alters the mechanical state of the material that in turn leads to a change in the value of damage. It can be then questioned whether this direct measure of damage leads to a representative value. In the current state of knowledge, the answer seems to be negative. A detailed analysis of the problem is presented in Montheillet and Moussy (1988) and François (2004). To directly or indirectly evaluate damage, other experimental techniques are available:

- Measurements at a microstructural scale. This may be done directly (density of microcracks and micro-voids) or indirectly (strain evolution measured by diffraction). These measures evaluate microscopic changes that can be integrated on a RVE using mathematical techniques of homogenization. The properties of the damaged RVE are then obtained.
- Global and physical measurements (density, electrical resistivity, acoustic emission, etc.). This requires the definition of a suitable model for converting data into mechanical characteristics.
- Global and mechanical measurements (evolution of elasticity, plasticity, or viscoplasticity moduli; strain fields; etc.). These measures are easier to interpret with damage variables, using the concept of effective stress (Lemaitre and Chaboche 2001).

Most of the time, damage is not directly accessible from experimental measurements. Like any physical quantity, its quantitative assessment is related to the definition of the variable chosen to represent the phenomenon. The previous list is obviously not exhaustive. The last two types of experiments lead to accurate results qualitatively, but are often limited quantitatively, showing a limit in current devices. It is necessary to distinguish the scales of measurement. At the macroscopic scale, mechanical measurements give better results than physical ones. At the microscopic scale, to avoid relying on a purely geometric approach, it is considered especially the diffraction methods to obtain the damage experimentally. These two techniques will be discussed further.

Neutron Diffraction

In this section, the methodology based on neutron diffraction in the crystalline lattice of a metallic polycrystal is mainly presented. The "time-of-flight" method (Santisteban 2001) is used in our tests, where neutrons cover a distance along a beam line. The configuration of the experiment is shown in Fig. 2. The time-of-flight t_{vol} is directly related to the interplanar spacing d_{hkl}, using the following relationship:

$$t_{vol} = \frac{2m_N L \sin\theta_B}{h} d_{hkl} \tag{1}$$

where m_N is the mass of the neutron, h is the Planck's constant, and θ_B is the Bragg angle fulfilling the Bragg diffraction condition. One of the main quantities measured by neutron diffraction (also valid for the X-ray diffraction) is the interplanar spacing d_{hkl}. When neutron beam or X-rays encounter a material, the periodicity properties of the crystal are the main cause of diffraction. The influence of the mechanical behavior can be then predicted on the results obtained by diffraction methods, listed below:

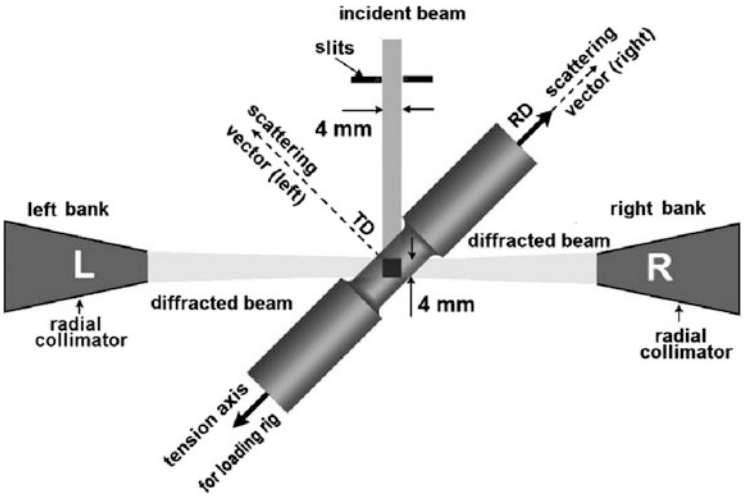


Fig. 2 Schematic principle of measurement by neutron diffraction

- The homogeneous elastic strains significantly alter the distance between atoms in the entire crystalline lattice.
- Because the structure of the lattice remains unchanged, plastic strains resulting from a complete slip of dislocations have no direct influence on the lattice parameter. However, multiple slides in a grain may have indirect consequences: the rotation of subgrains, their split, and the creation of incompatibilities between crystallites. In addition, one could expect stress relaxation due to shifts and leading to a decrease in the elastic strains.
- Measurement by diffraction on a polycrystal is an average measure. Defects, incompatibilities, and rotations of grains also translate the corresponding peaks, while other associated peaks corresponding to other grains do not vary. The consequence is a broadening of the total peak, as the sum of the diffraction of several grains with different behaviors.
- It is difficult to predict the influence of damage on such measurements. However, it can be assumed a priori that the formation and growth of micro-defects induce stress relaxation, leading to a shift of the peak. This will be discussed in a next paragraph (Fig. 3).

During neutron diffraction measurements, in situ tensile load is applied on the sample. The loading is stopped after that a set number of neutron counts is reached, to maximize the duration of the test while achieving satisfactory counting statistics. The traction machine is driven by stress during the elastic domain of the curve and then driven by displacement for the remainder of the tensile test. A greater number of measures are made during the elastic/plastic transition in order to better account for the microscopic mechanisms involved in the phases. After the beginning of plasticity and to avoid breaking of the strain gauge, it was withdrawn. The imposed

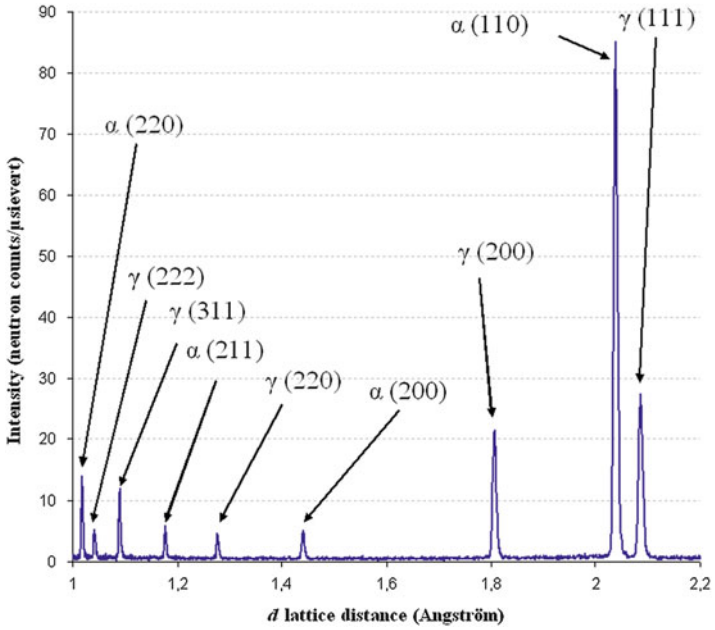


Fig. 3 Partial diffractogram obtained on a DSS UR45N (50 % ferrite + 50 % austenite)

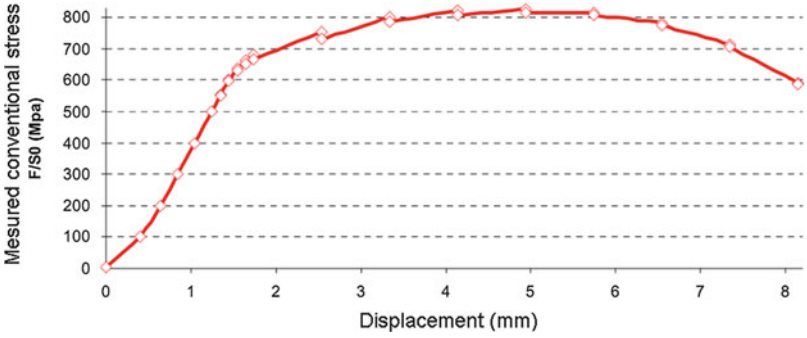


Fig. 4 Diagram of tensile stress imposed for measurement by neutron diffraction (Le Joncour et al. 2011)

loading is shown in Fig. 4. It is important to note that the method was carried out up to the rupture of the specimen. Through a calibration step, it is then possible to extract information to relatively high strains (globally 50 %, locally 250 % in the area of necking for a DSS). In published work, the interpretation of the measurements has been performed with strains up to 10 % (Clausen et al. 1999; Dakhlaoui et al. 2006). Several publications have achieved higher strains: 17 % for Dawson et al. (Dawson et al. 2001) and 30 % for Neil et al. (2010).

Measurements of Elastic Strains

Strictly speaking, the measurements obtained by neutron diffraction are not completely mesoscopic. Indeed, the response is obtained from a family of planes fulfilling the diffraction conditions within the gauge volume probed by the neutron beam. Several grains can be "switch on," but the contribution in an angular domain of the diffraction pattern (peak on diffractogram) is only due to one crystallographic phase (superposition of several peaks are possible). Therefore, the advantage of such a method is the selectivity in terms of phases. The diffractogram of each strain state is analyzed to extract the parameters of each peak (Bragg position, line broadening, and integrated intensity). It is first possible to identify the phases corresponding to each peak using previous studies on similar materials. Then, one can perform a statistical fitting (e.g., with a pseudo-Voigt function (Pecharsky and Zavalij 2005)) to identify the parameters from the observed peaks. The positions of the peaks are related to the elastic strain ε_{elas} (component along the loading direction) with the following relationship:

$$\varepsilon_{elas} = \ln\left(\frac{d_{hkl}}{d_{hkl,0}}\right) \tag{2}$$

To calculate the true elastic strain, the interplanar spacing of each family $\{hkl\}$ is divided by the interplanar spacing $d_{hkl,0}$, corresponding to the neutral configuration of the structure (assuming or checking that the internal residual stresses are zero).

Calibration Procedure

For high strains, the stress is affected by the necking of the sample (geometric change at macroscopic scale, due to plastic strain localization). An innovative method that does not require precise knowledge of the geometry of the sample during the test is proposed here. It allows the evaluation of the average stress that is present in the gauge volume irradiated by neutrons. This calibration is based on a number of assumptions detailed later and enables to expand the rank of the studied strains, until failure of the sample (up to globally 50 %, locally 250 % on the necking zone for a DSS).

Several methods have been proposed for the evaluation of elastic strains in the gauge volume irradiated by neutrons (Daymond 2004). The purpose of these investigations is to establish an average value for these elastic strains, due to the actual macroscopic stress and the effect of intergranular interactions. However, these methods have never been applied until failure. In order to calculate this macroscopic stress, a calibration method is used to treat diffraction data. This method is based on the calculation of the average of all the microscopic elastic strains. An arithmetic mean of the elastic strains on $\{hkl\}$ plans in all phases is performed. The choice of $\{hkl\}$ plans used in the arithmetic mean is delicate, the ultimate goal being to maximize the accuracy of the results. The selection for the peaks included in the computation of the average is done according to different criteria:

- It is necessary that the fit with the pseudo-Voigt functions be reliable from the beginning to the end of the test. Therefore, the mesoscopic elastic strain curves have to show no abnormal values.
- Only the first-order plans are taken into account to ensure that elastic strains related to these plans are counted only once.
- It is useful to use a transition scales model for checking the representativeness of the selected plans.

The details of the model to justify the calibration procedure are presented in the Ph.D. thesis of Lea Le Joncour et al. (2011). The main hypotheses are next presented that justify the proposed method.

- The threshold for which the stress heterogeneity becomes important also corresponds to a more pronounced triaxiality. The modeling performed in a more critical case than reality (the most deformed sample subjected to the maximal true strain) can be used to determine the stresses in the transverse (TD) and normal (ND) directions of the sample. If it does not exceed a small fraction of the strain in the loading direction (RD), then it is possible to assume that triaxiality is negligible.
- In the case of neutron diffraction experiments, the evolution is monitored of the elastic constants with the knowledge of the evolution of texture (ODF). This texture implemented in a self-consistent model shows that the elastic stiffness has very small influence.
- In addition, according to the experimental parameters (sensors in horizontal and vertical angular interval commonly around $\pm$ 20°) and measures of the orientation distribution functions, it is also possible to quantify the proportion of the diffracting volume used for the average calibration. The selected $\{hkl\}$ plans should at least represent 50 % of the irradiated volume. It is assumed that it ensures sufficient representativeness for neglecting strain incompatibilities.

  Therefore, a relationship of proportionality between the "true" macroscopic stress Σ_{RD} (i.e., corrected by the calibration method) and the average elastic strains (in all phases) $\langle\langle\varepsilon_{RD}\rangle_{hkl}\rangle_{tot}$ is proposed to end up as (Le Joncour et al. 2011)

$$\underline{\sigma}^g = \underline{\underline{c}}^g : \underline{\varepsilon}^g_{elas} = \underline{\underline{p}}^g : \underline{\Sigma} + \underline{\underline{q}}^g : \left(\underline{E}_{plas} - \underline{\varepsilon}^g_{plas}\right) \tag{3}$$

$$\Rightarrow \Sigma_{RD} \approx k\left\langle\langle\varepsilon_{RD}\rangle_{hkl}\right\rangle_{tot} \tag{4}$$

where $\underline{\underline{c}}^g$ represents the elastic stiffness tensor at mesoscopic scale, i.e., for a grain "g," and $\underline{\underline{p}}$ and $\underline{\underline{q}}$ are fourth-rank tensors related to the microstructure of the material (Baczmanski 2005). The subscript *tot* in the final average is an average over all the M phases. The constant k is identified with the slope of the beginning of the curve (part of the test without necking). This coefficient is generally very close to the Young's modulus of the material. This enables to check the representativeness of

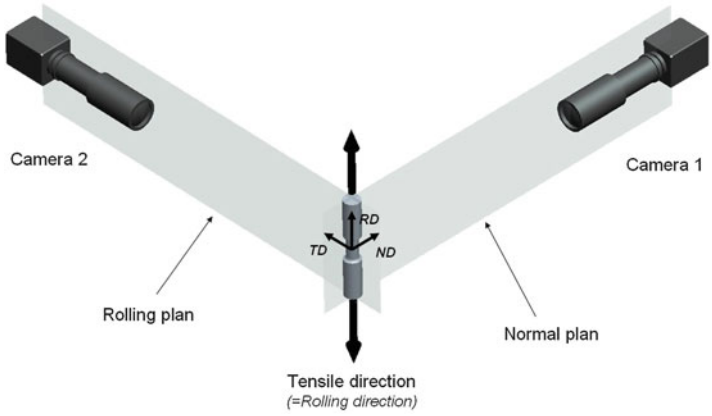


Fig. 5 Configuration for DIC (Le Joncour et al. 2011)

the selected average in the elastic range. The calibration can also be verified by comparison with a self-consistent model. This approach enables to correct the stress value.

Loading Tests with Digital Image Correlation

In addition to neutron diffraction, a tensile test can be performed and instrumented. For this, a simultaneous measurement of changes in the geometry of the specimens is carried out with two cameras positioned perpendicularly. This is a double digital image correlation (DIC), as presented in Fig. 5.

For example, a set of pictures is obtained such as those shown in Fig. 6. These pictures were taken at a speed of three frames per second, stored, and analyzed. To correct the macroscopic stress from necking effect, a Matlab program allows extracting the evolution of the section in the studied area by detecting the contour of the specimen. Strains are evaluated from the displacements of the points of a speckle image correlation. A toolbox of Matlab (Eberl et al. 2006) is launched in the same software program to perform this correlation. Particular attention has been paid to the sensitivity of the grid size as well as the initial speckle pattern, in order to compare the strains between cameras and areas.

The use of DIC enables to extract two kinds of information in order to correct the macroscopic geometric and kinematic effects and to have the intrinsic stress–strain curve of the material:

1. One needs to know the evolution of the cross section over time in order to overcome the necking effect due to the localization of plastic flow.
2. One needs to know the evolution of the true local strain with time.

This issue of this method is now common and details can be found in Le Joncour et al. (2010).

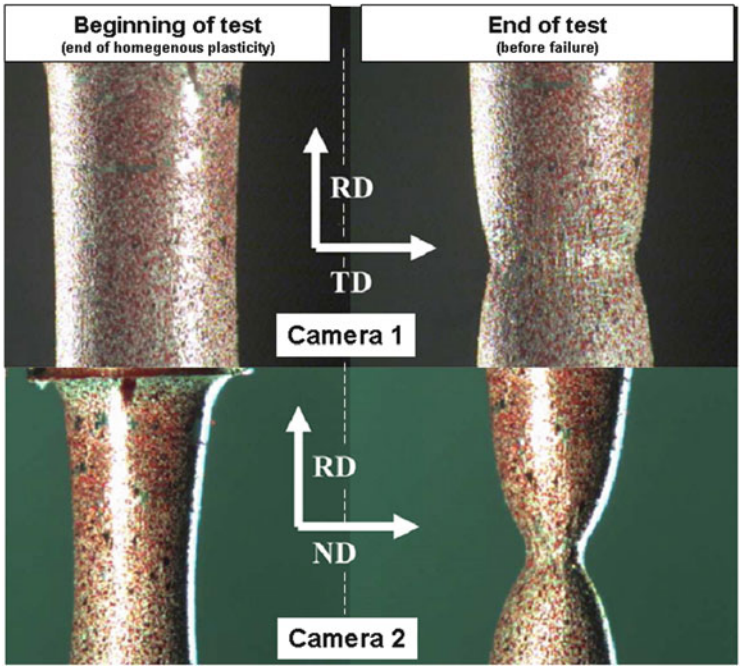


Fig. 6 Pictures of the sample UR45N in tension test with the two cameras at the beginning of the test, and at the end of the test, just before the break (Le Joncour et al. 2011)

Results on a DSS

Presentation of the Material

The duplex stainless steel UR45N was chosen for several reasons. The chemical composition of this steel is given in Table 1. On the one hand, it is a multiphase material ($M = 2$, ferrite 50 % and austenite 50 %), which is of obvious interest for industrial applications in highly corrosive process, such as chemical, petrochemical, offshore, nuclear, or paper industries, and is yet frequently studied. Second, it can be compared with experiments already performed on austenitic–ferritic steels from the literature (Bugat et al. 2000, Mcirdi et al. 2000, Dakhlaoui et al. 2006, El Bartali et al. 2007). Finally, tests performed by neutron diffraction showed that the steel was particularly suited to the measurement of ductile damage.

The studied material was obtained by continuous casting and then hot rolled down to 15 mm sheet thickness. The characteristic microstructure of this steel consists of austenitic islands elongated along the rolling direction and embedded in a ferritic matrix. EBSD observations have shown that all crystallites of ferritic phase have almost the same orientation, while austenitic islands are divided into smaller grains with different orientations of the lattice (Wroński et al. 2007).

Table 1 Composition of an austenitic–ferritic steel UR45N

Designation	C	Mn	Cr	Ni	Mo	Cu	S	N	Fe
*X*2 Cr Ni Mo 22.5.3 (UR45N)	0.015	1.6	22.4	5.4	2.9	0.12	0.001	0.17	Bal.

The sheet was annealed during 1,000 h at a temperature of 400 °C and next cooled at ambient temperature. It is well known (Desestret and Charles 1990; Mateo et al. 1997) that, at this temperature of aging (lower than 475 °C), the mechanism of spinodal decomposition of ferrite occurs. Transformations in ferrite are mainly decomposition of α/α′ (into Cr-poor α and Cr-rich α′ domains) and precipitation of an intermetallic phase rich in Ni, Si, and Mo (named as the G phase). The role of α′ and the G phases in hardening and embrittlement of ferrite is widely discussed in the literature, and the majority of authors (Calonne et al. 2001) agree that hardening is attributed essentially to the α′ phase. Indeed, the coherence shift between the lattice parameters of α and α′ phases introduces internal stresses reducing the dislocations mobility. The G particles have very small sizes (between 1 and 10 nm generally and up to 50 nm occasionally), and they precipitate, more or less uniformly, in the ferritic grains depending on the chemical composition of the steel. The largest particles are preferentially formed in the vicinity of a defect: the others are formed in the α/α′ and austenite/ferrite interfaces. Some microstructural transformations may be present in the austenitic phase, but they do not change the overall mechanical properties of the material.

It is therefore useful to observe the mechanisms of strain and damage of a two-phase material with the use of neutron diffraction. However, a number of constraints must be respected in order to perform correctly the experiments, especially by neutron diffraction:

- The material chosen should not be too textured in order to have a maximum number of peaks on the diffractograms and thus have the opportunity to observe the elastic strains for a maximum of plans.
- The grain size of the chosen polycrystal should not be too large so that the behavior of the irradiated gauge volume is representative.
- The distribution of stresses in both phases is investigated throughout the tensile test. It is therefore interesting to study two phases with different plastic behavior while taking great care to ensure that the elastic limits are sufficiently important to decrease the measurement errors.
- The diffraction pattern of the material should not have too many superposed peaks. Indeed, indexation would be too difficult.
- The fraction of both phases should be close to maintain sufficient peak intensity.

The selected material UR45N presents these characteristics. In addition, a previous study to lower stress and without aging treatment has provided a number of results on which this work can be supported (Dakhlaoui et al. 2006).

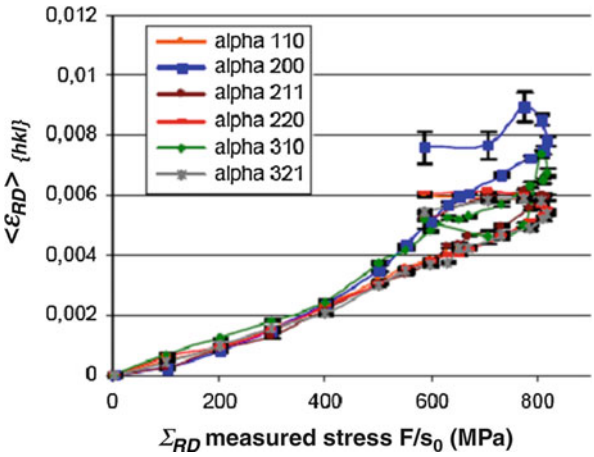


Fig. 7 Elastic strain of the diffracting volume in the loading direction as a function of the macroscopic stress applied to the ferrite in different $\{hkl\}$ families (Baczmanski et al. 2011)

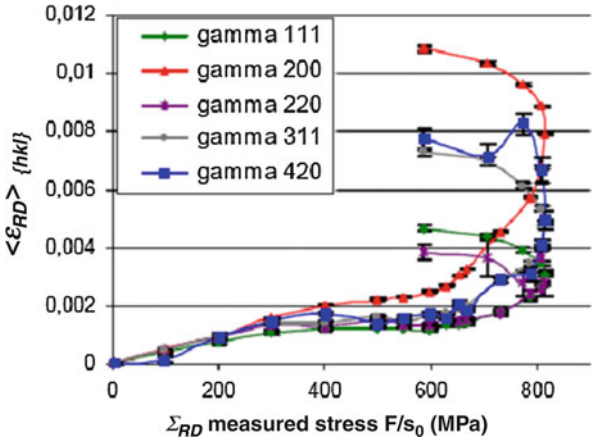


Fig. 8 Elastic strain of the diffracting volume in the loading direction as a function of the macroscopic stress applied to the austenite in different $\{hkl\}$ families (Baczmanski et al. 2011)

Neutron Diffraction Results at Mesoscopic Scale

According to the previous methodology (paragraph 2.2), the following results are obtained without calibration (i.e., without correction of stress from necking effect), as shown in Figs. 7 and 8.

By using the calibration, it is possible to correct the macroscopic stress, and this is presented in Figs. 9 and 10.

In the domain $E\alpha, \gamma$ from 0 to about 200 MPa (in Figs. 9 and 10), the two phases deform elastically: a linear progression is observed with a slope that is directly related to the radiocrystallographic elastic constants. The slope difference between different $\{hkl\}$ planes results from the anisotropy of the elastic properties of the material across the diffracting volume (intermediary between the grain scale and the macroscopic scale). The domain $E\alpha P\gamma$ (in Figs. 9 and 10) corresponds to the beginning of plasticity in the austenite. The yield stress of austenite is indeed lower

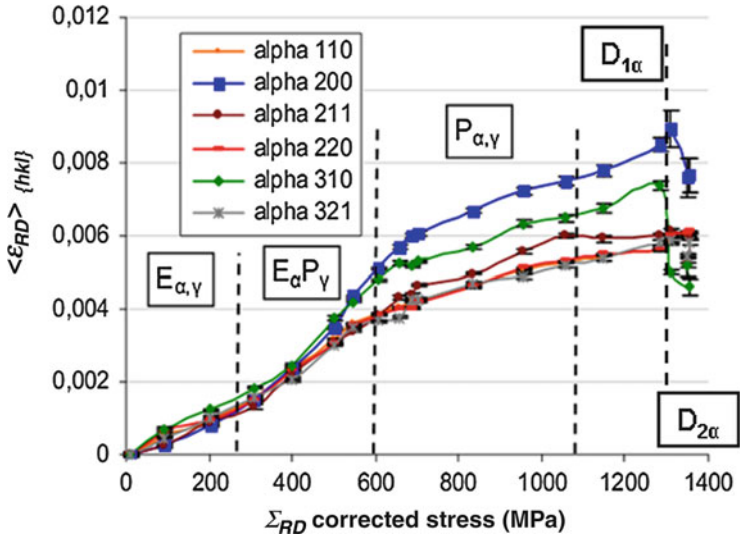


Fig. 9 Elastic strain of the diffracting volume in the loading direction as a function of the **corrected** macroscopic stress applied to the ferrite in different {*hkl*} families (Baczmanski et al. 2011)

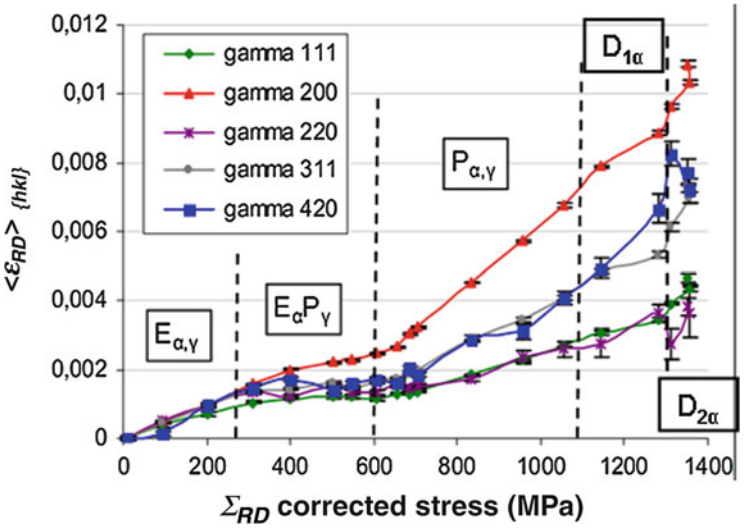


Fig. 10 Elastic strain of the diffracting volume in the loading direction as a function of the **corrected** macroscopic stress applied to the austenite in different {*hkl*} families (Baczmanski et al. 2011)

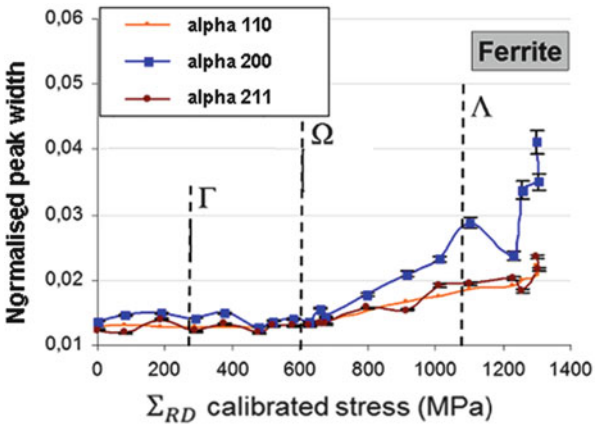


Fig. 11 Normalized width integral as a function of the **corrected** macroscopic stress applied to the ferrite in different $\{hkl\}$ families (Le Joncour et al. 2011)

than the yield stress of ferrite, that itself continues to deform elastically. Thus, there is a simultaneous change of slope for all austenite planes. The elastic strains in the austenite, for the same increment of stress, increase much less than in the elastic domain, because of the plasticity: a small slope remains due to the hardening. Therefore, an increase of elastic strain occurs in ferrite that balances the stress, but the linearity of the curves is preserved. This is an important observation that justifies a linear hardening model at the microscopic scale. In the domain $P\alpha$, γ (in Figs. 9 and 10), the two phases are plastically deformed. From that moment, the elastic strains in ferrite are less important than in austenite for the same increment of stress. It is interesting to note once again the linearity of the elastic strain evolution of the two phases with respect to the stress, except in the last part of the graph where a relaxation in ferrite is observed. These results are in good agreement with those found by Dakhlaoui et al. (2006), for smaller strains (i.e., not requiring the calibration step).

Characterization of Plastic Hardening

It is difficult to obtain precise information on hardening from the previous curves, because a direct access to plastic strain mechanisms (slip system activity) is not possible. Only the continuous increase of stress during the test indicates a hardening. Moreover, in heterogeneous materials, access to pieces of information on heterogeneities must be done with special attention to the so-called hardening. For example, when ferrite is still elastic at the macroscopic scale, the equivalent homogeneous material hardens strongly but mainly because of the incompatibility between the two phases. While at the mesoscopic scale, none of the two phases significantly harden (in fact, ferrite remains elastic and does not hardened at all in this domain $E\alpha P\gamma$).

Another possibility to extract information on microscopic hardening effects is to focus on the peak width. After normalization of the peaks, Figs. 11 and 12 are obtained.

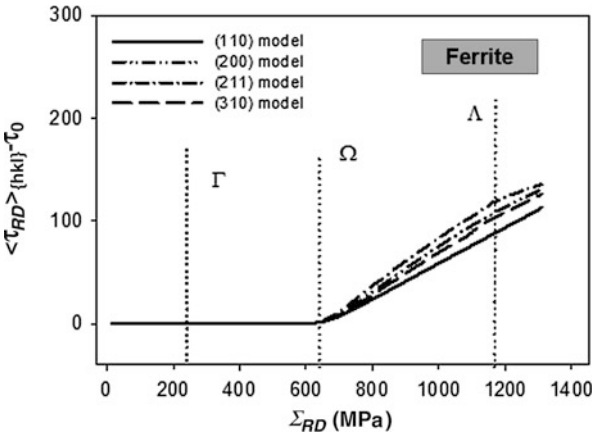


Fig. 12 Critical resolved shear stress resolved as a function of the **corrected** macroscopic stress applied to the austenite in different $\{hkl\}$ families (Le Joncour et al. 2011)

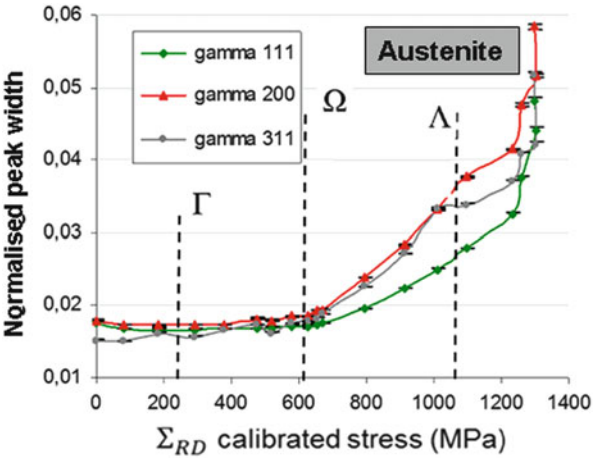


Fig. 13 Normalized width integral as a function of the **corrected** macroscopic stress applied to the ferrite in different $\{hkl\}$ families (Le Joncour et al. 2011)

There are also curves from an undamaged elastoplastic model. These theoretical curves represent the evolution of the shear stress limit (critical resolved shear stress) acting on the slip systems. A good correlation between the measured and calculated values is found. This experiment thus gives access to a measure of hardening across the slip systems (Le Joncour et al. 2011). It is then directly possible to connect this quantity with the dislocation density and experimentally check the evolution of the dislocation density. This is one of the variables in the model, which will be introduced and detailed in the following paragraphs. This hardening is limited during the plastification of austenite, as can be seen in Fig. 13 (stage $E\alpha P\gamma$), which is confirmed by the model. This can be explained by considering that internal mechanical forces are eventually balanced by the ferritic phase. This trend becomes significant only during the plasticity of ferrite where forces are more significant in austenite (with a higher tangent plastic modulus; Dakhlaoui et al. 2006; Fig. 14).

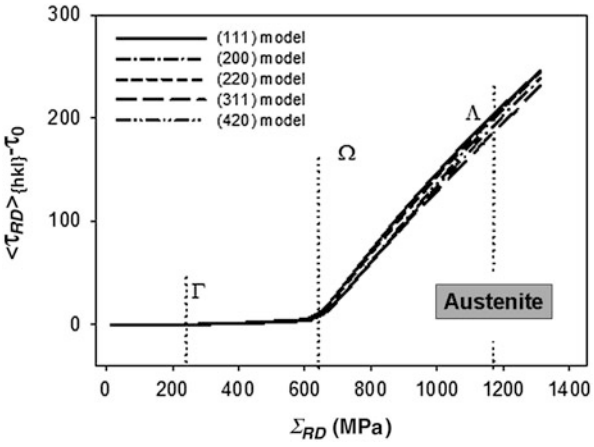


Fig. 14 Critical resolved shear stress resolved as a function of the **corrected** macroscopic stress applied to the austenite in different $\{hkl\}$ families (Le Joncour et al. 2011)

Characterization of Damage

In Fig. 9, some softening in ferrite can be seen after the nominal maximal stress is reached. This is specifically observed on some curves of elastic strain, for the ferritic phase. This softening occurs in two stages, first the elastic strain of the {211} plan loses its linear evolution, this step has been called $D1\alpha$. Second, in the step $D2\alpha$, two other peaks exhibit a relaxation of the elastic strains. Several hypotheses can be proposed to explain these two stages of relaxation of the elastic strains, in several planes of the ferritic phase:

1. A loss of linearity of the hardening
2. A change in texture
3. The growth of micro-voids or microcracks corresponding to a microscopic damage

All those hypotheses were tested by comparing the evolution of different quantities as cumulative plastic strain, orientation distribution function, etc., with an elastoplastic model without damage. It has been shown that the hypothesis of damage can explain all the observed phenomena. The two others cannot alone explain the relaxation of the elastic strain peaks at the end of the test.

This raises a question concerning the validity of such a measure. The effect of damage on diffraction measurements has been analyzed in Le Joncour et al. (2010), by considering a simplified Voigt-like model. It establishes several interesting results with a number of assumptions:

1. There is a direct influence of damage on the elastic strain, measured with a diffraction method: damage causes a decrease in the elastic strain.
2. The average elastic strain over the gauge volume increases less at the end of the test, for a given increment of displacement, because of damage.

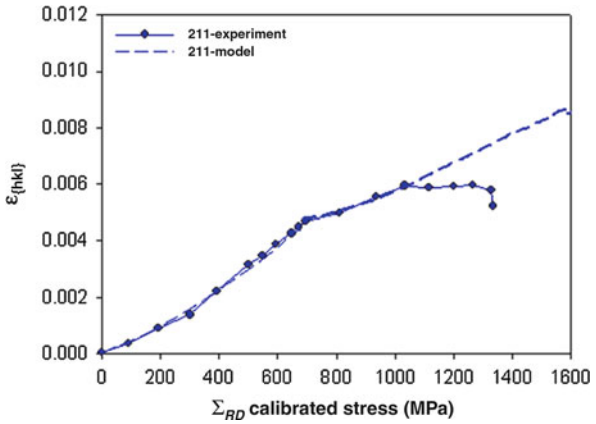


Fig. 15 Evolution of elastic strain along tensile direction in family {211} for experimental and modeling, as a function of corrected stress

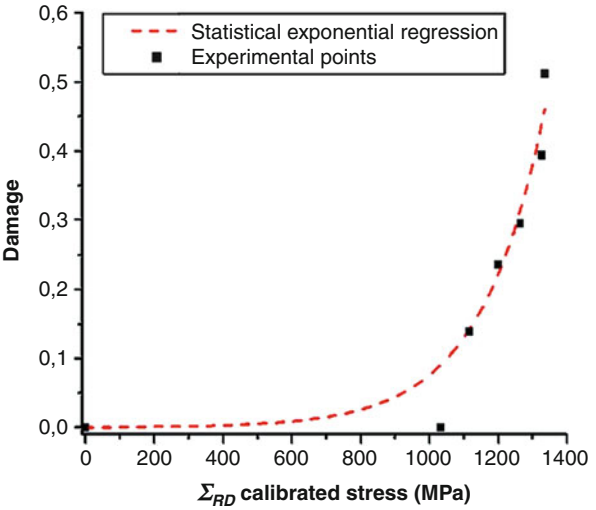


Fig. 16 Evolution of damage in family {211} extracted from experimental and modeling results

The possibility to indirectly extract damage from these experiments is therefore considered. For this, the relation is used that will be detailed later in this work:

$$\underline{\tilde{\varepsilon}}^g = \underline{\varepsilon}^g \sqrt{1 - d^g} \tag{5}$$

It compares the measured strains $\underline{\tilde{\varepsilon}}^g$ related to the damaged material to the calculated strains $\underline{\varepsilon}^g$ obtained using an elastoplastic model without damage. A curve is obtained, the evolution of the damage versus the macroscopic stress in the rolling direction (Le Joncour et al. 2011; Fig. 15), as shown in Fig. 16.

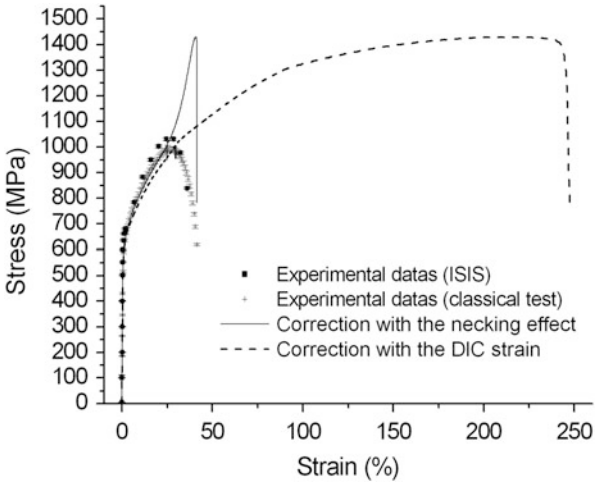


Fig. 17 Macroscopic stress versus macroscopic strain (Le Joncour et al. 2011)

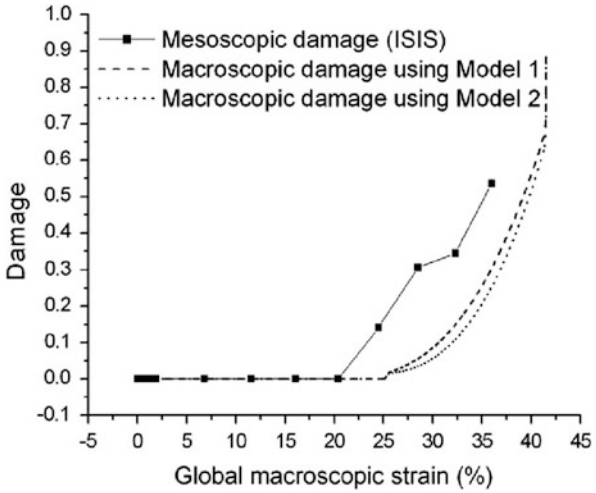


Fig. 18 Comparison of damage at different scales (Le Joncour et al. 2011)

Tensile Test with DIC Results at Macroscopic Scale

According to the previous approach, it is possible to link diffraction measurements with instrumented tensile test. Then, Fig. 17 (for stress–strain curve) and Fig. 18 (for damage–strain curve) are obtained.

On these figures, the tests performed during neutron diffraction (ISIS) as well as those made with laboratory goniometer are shown. A good repeatability between the two tests can be noted. In addition, it can also be seen that the difference between corrected curves is significant only for large strains (above 30 %). The black solid curve of Fig. 17 allows to identify the mechanical behavior in the necking area, from the plastification step to the damage step. Thus, in the same way as for mesoscopic results, it is possible to extract a macroscopic damage by

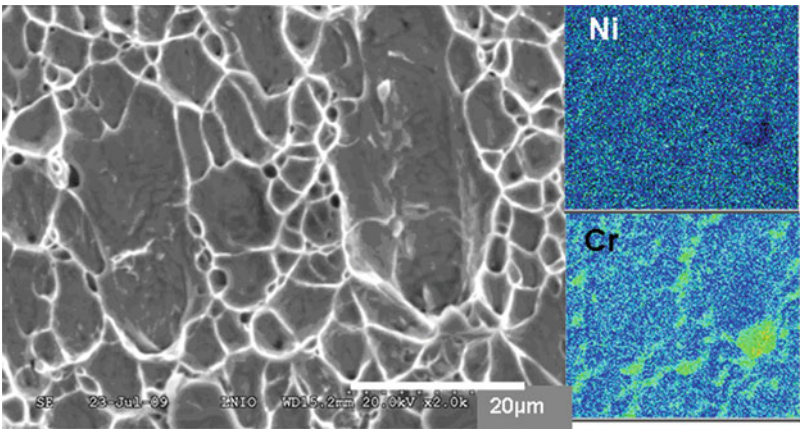


Fig. 19 Picture and mapping of elements in the fracture surface of the UR45N sample taken SEM equipped with an X-ray spectrometer (Le Joncour et al. 2011)

comparing the experimental curve with an undamaged elastoplastic model at macroscopic scale, as can be seen in Fig. 18. Two kinds of hardening have been considered. Model 1 corresponds to a nonlinear isotropic hardening, Prandtl–Reuss type. Model 2 corresponds to a linear isotropic hardening increased by a secondary surface (nonassociated plasticity), which describes a saturation, so a nonlinearity (Le Joncour et al. 2010). Both models lead to almost the same trend of damage. The representation is performed as a function of macroscopic global strain. If these curves are converted into local strain (in the necking area where mainly damage occurs), then the strain reaches 250 %. The damage is not activated until the very end of the tensile test. This tends to show that the damage coupled to plasticity is somewhat "brittle."

Other Measurements

Micrographies may be performed to complete the previous measurements. For example, the fracture surfaces of the samples were observed by SEM. The observations mainly show ductile damage with voids of various sizes (Fig. 1). At the center of the sample, the roughness is more pronounced. This can be explained by greater stress triaxiality when necking occurs. Numerous inclusions/precipitates of different sizes have been observed in many voids or deeper holes. Despite the presence of an X-ray fluorescence spectrometer using dispersive energy, the chemical composition of these inclusions could not be determined due to lack of sufficient signal. Indeed, the larger inclusions were located too deeply in the sample, and the other inclusions were not large enough to be detected. The fracture surface of the sample has also several areas similar to Fig. 19. Mapping of chemical elements performed on these areas appears to have more chromium content in the bottom of the voids and on the contrary an increase of the nickel content. It can be then assumed that the most important holes observed on fracture surface have grown in the ferritic phase.

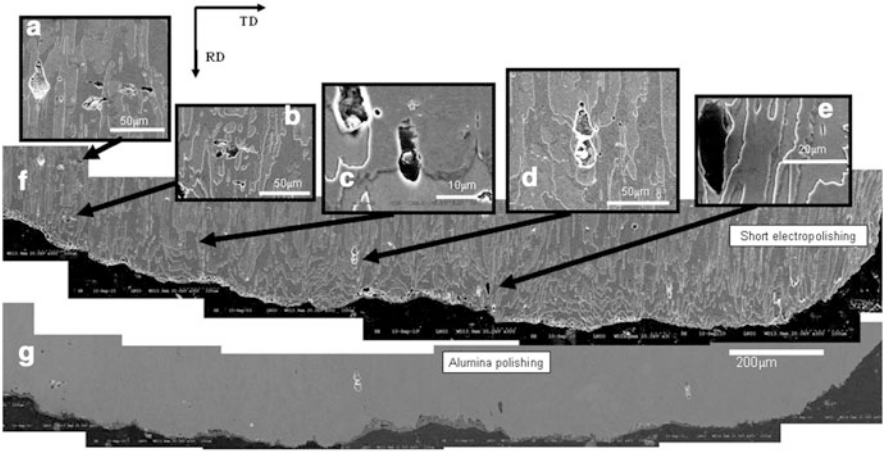


Fig. 20 SEM fractography taken along the rolling plane of the fracture surface of the UR45N sample (Le Joncour et al. 2011)

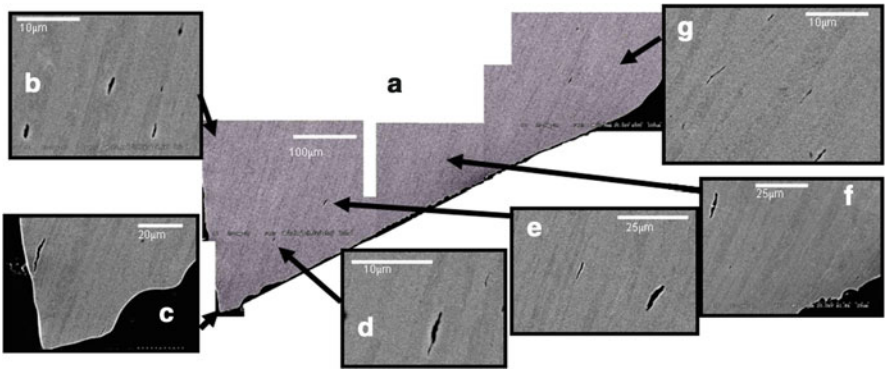


Fig. 21 SEM fractography taken along the transverse plane of the fracture surface of the UR45N sample (Le Joncour et al. 2011)

To observe the damage following the other two planes of the part, fractographies were made on the samples (Figs. 20 and 21). In Fig. 20, a sectional view along the plane that is normal to the rolling direction is presented. The pictures are taken before electropolishing to ensure that voids are not due to polishing. Moreover, mechanical polishing has been made in different directions. Finally, fractographies have been made using the same process without causing the same damage, while the observed damage varies; this supports the hypothesis that the damage observed in these pictures really comes from the tensile stress. In these figures, the phases can be recognized to their colors: dark for ferrite and clear for austenite, because of the electropolishing. The damage appears to be in the ferritic phase, perpendicular to

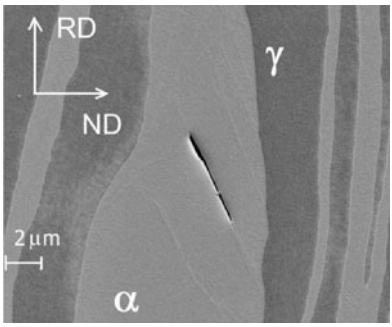


Fig. 22 Damage observed in ferrite along a slip system (Baczmanski et al. 2011)

the direction of traction loading. Figure 21 shows mainly an interfacial damage in the rolling/tensile direction.

Scenario for Damage Evolution

Damage in Slip Systems

The results presented in this study show that it is possible to extract damage from mesoscopic scale, based on an analysis performed with neutron diffraction. This analysis must be coupled with a multiscale and undamaged elastoplastic model. In the case of our austenitic–ferritic steel, the use of diffraction has confirmed the presence of damage in ferrite and revealed that the damage is link with the {211} plans. Further analysis of the samples shows more directly the effect of slip systems damaged, as shown in Fig. 22, carried out by SEM. Even if the corresponding slip planes are not strictly identified, the existence of this phenomenon has to be considered at the microscopic scale, along the slip systems of ferrite.

It is also known that the defects involved in the strain of the ferritic phase are dislocations with a Burgers vector $\frac{a}{2}\langle 111\rangle$, but the slip plane can belong to several families such as {110}, {211}, and {123} (with increasing atomic density) (Mahajan 1975, Louchet 1979). Mechanical twinning can also occur in the ferrite. Twinning occurs there only on {211} slip system (Christian 1970, Vitek 1970). Whatever the expected mechanism, the damage seems strongly coupled with elastoplastic strain of {211} slip system. The resulting microcracks are probably intracrystalline. Local analysis with EBSD and/or transmission electron microscopy near microcracks would support the actual mechanism, responsible for the damage occurrence in these steels.

Damage Scenario in UR45N

By merging pieces of information, a scenario of the chronology of damage can be then built. Three kinds of damage mechanisms appear possibly in the studied material:

1. The damage due to debonding phases around inclusions. However, this mechanism does not appear at an advanced stage, as it has been observed only in a few section views.

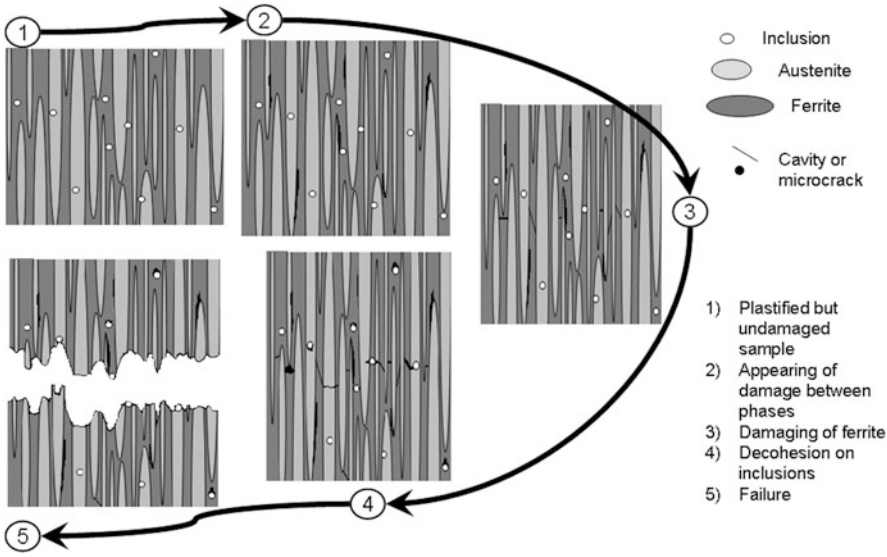


Fig. 23 Scenario of damage in UR45N leading to rupture of the RVE (Le Joncour et al. 2011)

2. The damage in ferrite due to sliding, observed by neutron diffraction. The damage does not necessarily occur along the plane for which the relaxation is measured by diffraction. This may be along other more "brittle" slip systems. However, when microcracks appear along this slip system, a relaxation occurs inside the grain.
3. The interfacial debonding due to strain incompatibilities between the two phases and the accumulation of dislocations at the interfaces. This mechanism is certainly not the one leading to rupture because the direction of the cracks is along the loading direction. A sample containing this kind of damage can certainly still resist the stress loading.

Numerous publications on other aged DSS (Bugat et al. 2000, Calonne et al. 2001, El Bartali et al. 2007, and Mcirdi et al. 2000) reported that damage in the ferrite is quite brittle, due to the aging treatment. Figure 23 shows a synthesis of the damage mechanisms and proposes a scenario for the development of damage in the studied material. The timeline is deducted from damage observed on various section views and elastic relaxations obtained by neutron diffraction. The five stages of this reconstructed history are detailed as follows:

1. The present diagram begins after the elastic strain, when the plastic flow takes place inside the two phases. It is supposed that (ductile) damage occurs only after plastification of the two phases.
2. The interfacial debonding is easily observed in Fig. 21. It is widespread and close to the fracture surface, but can also be observed in less distorted areas. This is the reason why it could be the first appearance of damage.

3. Microcracks in the ferrite along the slip systems (visible in Fig. 22) may account for the relaxation of the elastic strain in the same phase observed by neutrons from the necking event. These microcracks can be caused by:
   – A local but very large plastic strain: it is known that austenite tends to slide along several slip systems, while slides in the aged ferrite are fewer but more intensive (Bugat et al. 2000).
   – A decohesion around a defect in the ferrite, due to thermal aging increasing the presence of precipitates.
   – A combination of the above mechanisms is also possible.
4. The fracture surfaces consist mostly in voids with a large amount of inclusions of different sizes. This is probably one of the main causes in the rupture of the sample. It is chronologically positioned just before the rupture.
5. Fracture occurs during this last step, by coalescence of different voids inside the material. Then, the three stages of ductile fracture are found: nucleation, growth, and coalescence of voids (François et al. 1995).

Principles and Fundaments of the Modeling

Hypotheses on Micromechanics and Thermodynamics

In order to predict the previous experimental results, different modeling approaches may be used to take damage into account. Several assumptions have been considered to illustrate the present modeling:

- Monotonic loading
- Loading controlled by the total strain
- Isothermal conditions

Micromechanics Principles

Micromechanics consists in using microscopic description of a material continuum to obtain information at a macroscopic scale. However, depending on the considered behaviors, several scales can be used. For metallic materials, the possible scales have already been detailed in section Multiscale Measurements. However, in the present modeling, only three scales will be explicitly considered:

– The microscopic scale corresponds to the slip systems.
– The mesoscopic scale corresponds to a grain (i.e., one crystallite).
– The macroscopic scale is a RVE, corresponding to an aggregate of grains (generally several thousand, depending on the grain size).

To position the models in a solving approach, the main steps are proposed. A general algorithm for such a micromechanical approach (only the two higher scales are considered for the purpose of illustration) is considered below:

$$\underline{E}_n = \underline{E}_{n-1} + \Delta\underline{E} \tag{6}$$

$$\underline{\varepsilon}_n^g = \underline{\underline{A}}_n^g : \underline{E}_n \tag{7}$$

$$\underline{\sigma}_n^g = \underline{\underline{l}}_n^g(\underline{\varepsilon}^g) : \underline{\varepsilon}_n^g + \underline{\sigma}_{n-1}^g \tag{8}$$

$$\underline{\Sigma}_n = \left\langle \underline{\sigma}_n^g \right\rangle_g \tag{9}$$

where $\underline{E}, \underline{\varepsilon}$ represent the strain tensor, respectively, at macroscopic and mesoscopic scales and $\underline{\Sigma}, \underline{\sigma}$ represent the stress tensor, respectively, at macroscopic and mesoscopic scales. n corresponds to an iterative step of the calculation. "g" is the label of the mesoscopic entity. The previous system of equations must be inserted in a loop. These equations describe the main stages of a 2-scale transition (Bornert et al. 2001). Interpretation of the different stages is given below:

- *Localization*: allows passing from the macroscopic scale to control the mesoscopic scale, where $\underline{\underline{A}}$ is the tensor of strain localization (Eq. 7)
- *Description*: allows giving the mesoscopic mechanical behavior of the material, using the modulus tensor $\underline{\underline{l}}$, for example, and considering possible initial/residual stresses (Eq. 8)
- *Homogenization*: allows going back to the macroscopic scale by an appropriate average (Eq. 9)

If another scale has to be taken into account, corresponding stages of localization and homogenization (similar to Eqs. 7 and 9) have to be added to the present algorithm.

Thermodynamics Principles

The proposed models are also based on thermodynamics principles. It is assumed that the medium is always in state that is close to equilibrium. Using the thermodynamics of irreversible processes (Lemaitre and Chaboche 2001), the method is as follows:

- Definition of the internal and external state variables
- Definition of a Helmholtz-specific free energy, from which the state relations are derived
- Definition of dissipation potentials and yield functions, from which the evolution equations of dissipative phenomena are derived

This scheme is used for the description of both the mesoscopic and microscopic scales. It means that the Helmholtz free energy has to be defined considering phenomena occurring on two scales.

Finite Transformation and Objectivity

Mechanical constitutive models have to satisfy the general principles of physics: causality, determinism, material simplicity, Curie principle, objectivity, etc. (Truesdell and Noll 2003). The latter is of great importance. Objectivity is a specific constraint also named as frame indifference. However, this concept is quite ambiguous as detailed in Frewer (2009), Panicaud et al. (2013), and Rouhaud et al. (2013). It means that tensors and tensorial equations have to respect several criteria depending on the tensor rank. It can be demonstrated that such criteria correspond equivalently either to an invariance through any observer changes or to an indifference to the rigid body motion superposition in a three-dimensional space. In the classical approach, any Lagrangian quantities can be used in constitutive laws, whereas Eulerian quantities have to be tested because only some of them are objective. Derivatives also depend on the chosen frame of reference and have to be corrected leading to convected rates (Truesdell) or rotation rates (Jaumann, Green–Naghdi, etc.). They do not all represent a differentiation with respect to time (Rouhaud et al. 2013).

Frame indifference (i.e., objectivity) is of particular importance for finite transformations, where strain tensor can reach high values. There are different ways to take into account this kinematics problem, depending on the considered scale. For a pure macroscopic method, either a rate formulation (tangent approach) or a secant formulation can be envisaged. For a micromechanical method, the same two formulations could be used. However, when using a tangent approach, the global nonlinear behavior is separated in a succession of elementary linear behaviors. If load increments are small enough, the material response can be roughly estimated using classical kinematics tensor. When using the micromechanics approach, this aspect has to be completed by updating the grain orientations for each step load, to take into account the induced texture effects. This approximation is used in the present chapter (except in the description of elastic behavior). However, two limitations have to be kept in mind:

- First, the rate quantities should be corrected, because an objective rate is not strictly a derivative operation, even for small transformations. A correct derivative would give a different trend. For small *but finite* transformations, this objective derivative would lead to a different response as the classical objective rates (Jaumann), also different from the uncorrected classical derivative (considered here for presentation) (Rouhaud et al. 2013).
- Second, some modeling for finite transformation exists (Lipinski 1989). It should be verified that the limit of such a modeling, for a small transformation and rate formulation, is consistent with the chosen linear modeling.

Consequently, the actual presentation will be written for small transformations (except in the next paragraph to illustrate elasticity), as it has been proven as sufficient at first approximation. However, if another way is considered, the

previous limitations and assumptions have to be addressed and the present methodology can be adapted by replacing the corresponding quantities with objective ones.

Mechanical Models

Three specific behaviors have to be considered while modeling metal process forming: elasticity, plasticity, and damage. As its coupling is particular, damage will be detailed in the last paragraph. Low temperature forming processes and behavior will only be considered.

Elastic Constitutive Relation

Elastic behavior corresponds to reversible strain occurring at an atomic scale. However, due to the considered scale, the approach is limited to the mesoscopic scale (i.e., grain scale). To illustrate the micromechanical approach, the following hypotheses valid only in the present paragraph are adopted:

- The material is isotropic at the macroscopic scale (i.e., considering a random texture); this assumption is not of primary importance and can be updated by changing the average in Eq. 11.
- The material is anisotropic at the microscopic scale.
- The behavior is linear elastic at the microscopic level.

In addition to Eqs. 6 and 7, Eqs. 8 and 9 have to be replaced by Eqs. 10–12:

$$\underline{\underline{\sigma}}_n^g = \underline{\underline{c}}_n^g : \underline{\underline{\varepsilon}}_n^g \tag{10}$$

$$\underline{\underline{C}}_n = \left\langle \underline{\underline{c}}_n^g : A_n^g \right\rangle_g \tag{11}$$

$$\underline{\Sigma}_n = \underline{\underline{C}}_n : \underline{E}_n \tag{12}$$

where $\underline{\underline{c}}^g$ and $\underline{\underline{C}}$ are, respectively, the mesoscopic and macroscopic elastic compliance tensors. Only elastic strains have to be considered and initial residual stresses have been neglected.

To illustrate the thermodynamical approach and following the previous method (see section "Hypotheses on Micromechanics and Thermodynamics"), state variables have to be defined. The definition for stress and elastic strain will be considered at the mesoscopic scale. A secant formulation is chosen, so that finite transformations have to be carefully taken into account. Using a Lagrangian formulation, a Helmholtz-specific free energy ${\psi_0}^g$ can be written as follows:

$${\rho_0}^g {\psi_0}^g = \frac{1}{2} \underline{\underline{\varepsilon}}_L^g : \underline{\underline{c}}^g : \underline{\underline{\varepsilon}}_L^g \tag{13}$$

where $\underline{\underline{\varepsilon}}_L^g$ is the Green–Lagrange strain tensor. Because no dissipation occurs, the dissipation functions Φ and dissipation potential are equal to zero. Using the

Gibbs–Duhem relation, it is possible to prove classically that (Lemaitre and Chaboche 2001)

$$\Phi^g = \underline{\sigma}_L^g : \dot{\underline{\varepsilon}}_L^g - {\rho_0}^g \frac{\partial {\psi_0}^g}{\partial t} = 0 \Rightarrow \forall \dot{\underline{\varepsilon}}_L^g, \underline{\sigma}_L^g = {\rho_0}^g \frac{\partial {\psi_0}^g}{\partial \underline{\varepsilon}_{L\,\Omega}} = \underline{\underline{c}}^g : \underline{\varepsilon}_E^g \tag{14}$$

The last equality is obtained using Eq. 13. Equation 14 is similar to Eq. 10, except that it takes into account kinematics nonlinearities. If the Cauchy (Eulerian) stress tensor is required, it can be obtained by using the following relation:

$$\underline{\sigma}_E^g = J^{-1} \underline{F}^g \cdot \underline{\sigma}_L^g \cdot : \underline{F}^{gT} = J^{-1} \underline{F}^g \cdot \underline{\underline{c}}^g : \underline{\varepsilon}_E^g \cdot : \underline{F}^{gT} \tag{15}$$

where J is the determinant of $\underline{F}^g$, the transformation gradient tensor at mesoscopic scale. Equation 14 or 15 should now replace Eq. 10.

It is also possible to take material nonlinearities into account. Choosing to retain the particular assumption of material *isotropy* at microscopic scale, a potential of the form can be chosen:

$${\rho_0}^g {\psi_0}^g(\underline{\varepsilon}_L^g) = {\rho_0}^g {\psi_0}^g(\varepsilon_I, \varepsilon_{II}, \varepsilon_{III}) \tag{16}$$

where ε_I, ε_{II}, ε_{III} are the invariants of $\underline{\varepsilon}_L^g$. Representative theorem for isotropic functions is used to write the right terms of Eq. 17 (Boehler 1978). Using the Gibbs–Duhem relation, it is now possible to prove that

$$\Phi^g = 0 \Rightarrow \forall \dot{\underline{\varepsilon}}_L^g, \underline{\sigma}_L^g = {\rho_0}^g \sum_{i=I}^{III} \frac{\partial {\psi_0}^g}{\partial \varepsilon_i} \frac{\partial {\varepsilon_i}^g}{\partial \underline{\varepsilon}_L} = a_0 \underline{1} + a_1 \underline{\varepsilon}_L^g + a_2 \underline{\varepsilon}_L^g \cdot \underline{\varepsilon}_L^g \tag{17}$$

where a_0, a_1, a_2 are three functions depending on ε_I, ε_{II}, ε_{III}. An explicit free energy is required to specify those functions. For Eulerian representation, a relation similar to Eq. 15 can be established. It is then possible to replace Eq. 10. It can be noticed that the Lagrangian to Eulerian transport is usually performed at macroscopic scale.

Plastic Constitutive Relations

Plasticity of Metals

Plastic strains correspond to irreversible transformations for the considered material (Lemaitre and Chaboche 2001). This definition is based on macroscopic observations, such as those that can be performed during a monotonic tensile test. But it is known that similar mechanisms may occur locally, such as during cyclic loading by fatigue (Catalao et al. 2005).

The mechanisms at work are now clearly identified. It corresponds to irreversible movements of dislocations (François et al. 1995; Hull and Bacon 1995). The literature is quite prolix on the subject, and micromechanical models based on these phenomena are now common (Paquin et al. 2001). Dislocations are linear defects that form within the material and able to move (essentially by sliding at low

temperature). They interact with other microscopic components, possibly with blocking, annihilations, etc. For example, it has been demonstrated and verified that the change in yield stress of materials is mainly due to the work of dislocations (François et al. 1995).

State of the Art in Micromechanical Plasticity Models

There are several approaches to describe the behavior of plastic materials (especially a metal or a metal alloy). For a detailed comparison of these different approaches, see the Ph.D. thesis of Boudifa (Boudifa et al. 2006):

- The physical microscopic approach is based on a realistic analysis of the behavior of each single crystal/grain in terms of dislocation density. Schmid relationship is used to calculate the flow of active systems. However, when considering hardening, different relationships have been proposed (Teodosiu and Sidoro 1976; Franciosi 1985; Pilvin 1994). This approach helped to validate the concept of interaction matrix between dislocations. Finally, to complete the system of equations, one has to define a rate equation for the density of dislocations. A verification of these relationships is possible in particular by transmission electron microscopy (Keller et al. 2010). Research is currently done with the dynamics of dislocations theory, which should enable a priori to calculate the various quantities, such as the interaction matrix (Devincre et al. 2001).
- The phenomenological microscopic approach is a variant of the previous one, but with less physical content. It no longer involves the density of dislocations, because it considers directly the mechanical quantities such as sliding. This approach has the advantage of conveniency, but suffers from a more difficult identification with direct microscopic observations. In addition, the coupling between the variables is only performed by analogy with macroscopic methods, at least historically.
- To be consistent in the previous approach requires the introduction of thermodynamic concepts that ensures a correct modeling. This is the phenomenological microscopic approach with internal variables (Boudifa et al. 2006; Hfaiedh et al. 2009; Le Joncour et al. 2011). For this, internal state variables are introduced that have to be linked as correctly as possible. These variables are supposed to reflect the effects of plasticity. The advantage of such a micromechanical approach is to be thermodynamically compatible and thus as predictive as possible. The problem, once again, is that identification based on experimental observations is not always straightforward. Nevertheless, this has been chosen to describe the elastoplastic materials (from a micromechanical approach), in the present chapter.
- The macroscopic approach neglects completely the microscopic origins of plasticity. The most robust approach currently proposed is built using a thermodynamic framework with internal variables (Lestriez et al. 2003; Mariage et al. 2003), as it has been demonstrated in applications of metal forming processes.

Continuous Damage Mechanics

Before considering the detailed approaches, damage should be mathematically defined. The definition of a mechanical damage is a difficult problem since, macroscopically, nothing distinguishes an undamaged volume element from a damaged volume element. Therefore, an internal variable representing the state of deterioration of the material can be imagined. Several approaches are then possible depending on the kind of considered damage. The CDM approach is used. To adequately describe the overall behavior of a damaged structure, it is necessary to introduce an internal variable of this damage, D_{dam}. This is the approach suggested by Rabotnov (Lemaitre and Chaboche 2001). In a mechanically loaded RVE, the damage using the ratio of the effective area S_{eff} on the geometric one S_{geo} (associated with normal $\vec{n}$) can be defined:

$$D_{\text{dam}} = 1 - \frac{S_{\text{eff}}}{S_{\text{geo}}} \tag{18}$$

This justifies the geometric interpretation of this phenomenon. D_{dam} can be interpreted as a surface density of defects associated with the considered damage. At first approximation, the isotropy of this parameter is assumed. This assumption is only valid if the defects are uniformly and randomly oriented in all directions (with normal $\vec{n}$). Montheillet and Moussy (1988) give the interpretation of different values taken by D_{dam}:

- $D_{\text{dam}} = 0$: undamaged state in the RVE, i.e., $S_{\text{eff}} = S_{\text{geo}}$.
- $D_{\text{dam}} = 1$: totally damaged state in the RVE, i.e., $S_{\text{eff}} = 0$.
- $0 < D_{\text{dam}} < 1$: thus characterizes the progressive state of damage in the RVE.

If the force applied to the sample is constant, this implies that the stress is affected by the presence of damage. The damage variable is thus taken into account through the concept of effective stress. In other words, only a surface fraction allows maintaining cohesion of the solid and resists mechanically to the solicitation. This defines the effective stress by

$$\Sigma_{\text{eff}} = \Sigma_{\text{geo}} \frac{S_{\text{geo}}}{S_{\text{eff}}} = \frac{\Sigma_{\text{geo}}}{1 - D_{\text{dam}}} \tag{19}$$

$$\Leftrightarrow D_{\text{dam}} = 1 - \frac{\Sigma_{\text{geo}}}{\Sigma_{\text{eff}}} \tag{20}$$

It is assumed that the damage occurs only by this single change, which demonstrates its indirect link with the mechanical stress field. In models, it will then be sufficient to replace the stress by the effective stress. For example, for linear elasticity, damage coupling will eventually result in a change in Young's modulus of the damaged material. This technique can be applied a priori to all kinds of behaviors. It is equivalent to replace the damaged material, featuring local

discontinuities by using an equivalent homogeneous material, describable in the approach of continuum mechanics: this is the CDM. The present definition corresponds to the strain equivalence. There is also an energy equivalence definition that is described in the next paragraph.

Strain Versus Energy Equivalence

Two possible definitions of the coupling between mechanical fields and damage can be used. To illustrate the damage coupling, a pure elastic behavior is considered. Let us first introduce a scalar damage variable d^g_{dam}. Two configurations are envisaged, at the mesoscopic scale: the damaged material and the equivalent undamaged material. This equivalence is possible because the damage description is based on the CDM approach. In the first configuration, the elastic stiffness tensor $\underline{\underline{\widetilde{c}}}^g$ is damaged (noted with a ~) and the material is under loading $\left(\underline{\underline{\sigma}}^g, \underline{\underline{\varepsilon}}^g_{\text{elas}}\right)$. The elastic behavior is given by

$$\underline{\underline{\sigma}}^g = \underline{\underline{\widetilde{c}}}^g : \underline{\underline{\varepsilon}}^g_{\text{elas}} \tag{21}$$

In the second configuration, the elastic stiffness tensor $\underline{\underline{c}}^g$ is undamaged, but the load is modified: $\left(\underline{\underline{\widetilde{\sigma}}}^g, \underline{\underline{\varepsilon}}^g_{\text{elas}}\right)$. The strain is the same as in the first configuration; this is the strain equivalence. The elastic behavior is now given by

$$\underline{\underline{\widetilde{\sigma}}}^g = \underline{\underline{c}}^g : \underline{\underline{\varepsilon}}^g_{\text{elas}} \tag{22}$$

According to the previous definition of damage, the effective stress (second configuration) and the real/true stress (first configuration) are related by

$$\underline{\underline{\widetilde{\sigma}}}^g = \frac{\underline{\underline{\sigma}}^g}{1 - d^g_{\text{dam}}} \tag{23}$$

$$\Rightarrow \underline{\underline{\widetilde{c}}}^g = \left(1 - d^g_{\text{dam}}\right)\underline{\underline{c}}^g \tag{24}$$

For the energy equivalence, it is assumed that configuration 2 has to be modified, with the load $\left(\underline{\underline{\widetilde{\sigma}}}^g, \underline{\underline{\widetilde{\varepsilon}}}^g_{\text{elas}}\right)$. Elastic work in the damaged configuration and undamaged one is supposed to be equivalent, so equations are obtained:

$$\underline{\underline{\widetilde{\sigma}}}^g : \underline{\underline{\widetilde{\varepsilon}}}^g_{\text{elas}} = \underline{\underline{\sigma}}^g : \underline{\underline{\varepsilon}}^g_{\text{elas}} \tag{25}$$

$$\Rightarrow \begin{cases} \underline{\underline{\widetilde{\sigma}}} = \dfrac{\underline{\underline{\sigma}}^g}{\zeta\left(d^g_{\text{dam}}\right)} = \dfrac{\underline{\underline{\sigma}}^g}{\sqrt{1 - d^g_{\text{dam}}}} \\ \underline{\underline{\widetilde{\varepsilon}}}^g_{\text{elas}} = \zeta\left(d^g_{\text{dam}}\right)\underline{\underline{\varepsilon}}^g_{\text{elas}} = \sqrt{1 - d^g_{\text{dam}}}\,\underline{\underline{\varepsilon}}^g_{\text{elas}} \end{cases} \tag{26}$$

Let us now introduce a second-rank tensor damage $\underline{d}^g_{\text{dam}}$. The energy equivalence is directly considered that will be used in further derivations. To ensure the symmetry of the elastic stiffness tensor, it is possible to define a fourth-rank tensorial damage operator, such as

$$\underline{\underline{\widetilde{c}}}^g = \underline{\underline{\zeta}}\left(\underline{d}^g_{\text{dam}}\right) : \underline{\underline{c}}^g : \underline{\underline{\zeta}}^T\left(\underline{d}^g_{\text{dam}}\right) \tag{27}$$

It leads to

$$\begin{cases} \underline{\widetilde{\sigma}} = \underline{\underline{\zeta}}^{-1}\left(\underline{d}^g_{\text{dam}}\right) : \underline{\sigma}^g \\ \underline{\widetilde{\varepsilon}}^g_{\text{elas}} = \underline{\varepsilon}^g_{\text{elas}} : \underline{\underline{\zeta}}\left(\underline{d}^g_{\text{dam}}\right) \end{cases} \tag{28}$$

The last step requires to define the damage operator $\underline{\underline{\zeta}}\left(\underline{d}^g_{\text{dam}}\right)$. Different coupling are possible. For example, a choice could be $\underline{\underline{\zeta}}\left(\underline{d}^g_{\text{dam}}\right) = \left(1 - \left\|\underline{d}^g_{\text{dam}}\right\|\right)^{+\frac{1}{2}}\underline{\underline{1}}$. With this choice, one loses the effect of damage anisotropy at grain scale (mesoscopic), during homogenization of the RVE scale (macroscopic). It could be also $\underline{\underline{\zeta}}\left(\underline{d}^g_{\text{dam}}\right) = \left(\underline{\underline{1}} - \underline{d}^g_{\text{dam}} \otimes \underline{d}^g_{\text{dam}}\right)^{+\frac{1}{4}}$. Other choices can be used as detailed in Montheillet and Moussy (1988). The energy equivalence can be also performed with elastoplastic behavior, as it will be done for further applications.

Scale Definition of Damage for $N = 2$ and $N = 3$

According to experimental results (section Experimental Aspects), damage can be defined either at mesoscopic scale and/or at microscopic scale. In the previous paragraph, the strain and energy equivalences are presented using a damage variable defined at mesoscopic scale.

The framework is based on a phenomenological microscopic approach with internal variables, describing plasticity (using thermodynamics of irreversible processes). Plasticity is defined over three scales: macroscopic (RVE), mesoscopic (grain), and microscopic (slip system; although strictly speaking, this is not really a scale of description but a projection step).

Thermodynamic state variables d^s and/or $\underline{d}^g$ can be defined that represent, respectively, damage at either microscopic or mesoscopic scale. It can be wondered how many scales have to be really considered. Indeed, there are not necessarily as many scales as for plasticity. There are two possible modeling for damage, either at mesoscopic scale (i.e., $N = 2$) or at microscopic scale (i.e., $N = 3$). The latter can be interpreted as the result of these inclusions or precipitates on slip systems as shown in Fig. 24.

Damage Coupling

The different behaviors can be linked with damage, using the total energy equivalence, either directly/heuristically by considering the relations between damaged and undamaged fields (assuming a relation such as Eq. 28) or indirectly by

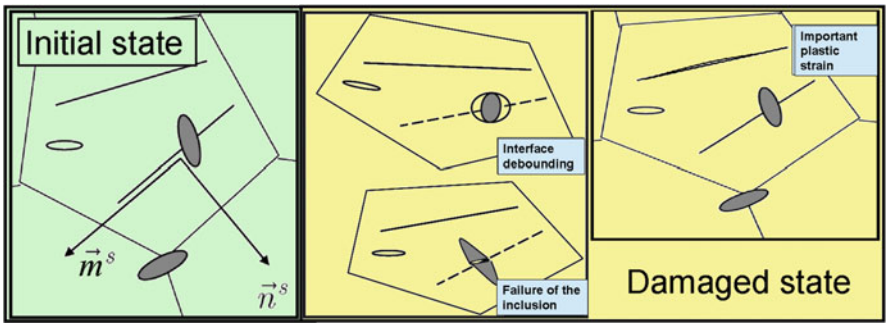


Fig. 24 Damage mechanisms corresponding to the microscopic damage variable

considering the thermodynamics methodology. These two approaches can be easily compared and lead to the same results. In the present case, the thermodynamics methodology is used.

A damage rate relation has to be added, for example, the one proposed by Lemaître and Chaboche (Lemaitre and Chaboche 2001) (for each slip systems):

$$\dot{d}^s = \frac{\dot{\lambda}}{(1-d^s)^{\alpha_1}} \left\langle \left\langle \frac{y^s - \alpha_2}{\alpha_3} \right\rangle \right\rangle^{\alpha_4} \tag{29}$$

where α_i are material constants and y^s is the thermodynamic damage force, dual variable of the damage variable d^s.

Other variant models are possible. In particular, an evolution of Eq. 29 is proposed, in agreement with experimental results (e.g., as seen in UR45N DSS).

Micromechanical Modeling of an (Elasto)plastic Model with N-Scale Damage

General Case

Table 2 lists the chosen state variables.

Then, a Helmholtz-specific free energy should be defined, sum of an elastic contribution and plastic contributions – an isotropic hardening and a kinematic hardening (with damage coupling):

$$\rho^g \psi^g = \frac{1}{2} \underline{\varepsilon}^g_{\text{elas}} : \underline{\underline{\tilde{c}}}^g : \underline{\varepsilon}^g_{\text{elas}} + \sum_{s,t \in g} \frac{1}{2} \tilde{H}^{st} r^s r^t + \frac{C_C}{3} \underline{\beta}^g : \underline{\beta}^g + Cste \tag{30}$$

where $\underline{\underline{\tilde{c}}}^g$ is the damaged elastic stiffness tensor and $\tilde{H}^{st}$ is the damaged interaction matrix between dislocations. In this phenomenological approach, the dislocation

Table 2 State variables of the explicit model for $N = 3$

Scale mesoscopic	Physical meaning elasticity in grain Ω	Flux variable $\underline{\varepsilon}^{\Omega}_{\text{elas}}$	Force variable $\underline{\sigma}^{\Omega}$
Mesoscopic	Intergranular kinematic hardening	$\underline{\beta}^{\Omega}$	$\underline{X}^{\Omega}$
Microscopic	Intragranular isotropic hardening	R^5	R_c^5
Mesoscopic	Damage in grains	$-\underline{d}^{\Omega}$	$\underline{y}^{\Omega}$
Microscopic	Damage in slip systems	d^5	y^s

density is replaced by the hardening variable r^s. In addition, this formulation is currently valid for any mesoscopic constituent "g" (whatever the number of phases).

The coupling between the elastic moduli and the damage variable is done using Eq. 24, while plastic parameter can be coupled considering the same approach (energy equivalence) with

$$\tilde{H}^{st} = \sqrt{1-d^s}\sqrt{1-d^t}H^{st} \tag{31}$$

A plasticity/yield criterion has to be established. This is now well known and can be written as

$$\tilde{f}^{s\in g} = \left|\left(\underline{\widetilde{\sigma}}^g - \underline{\widetilde{X}}^g\right) : \underline{M}^s_{\text{sym}}\right| - \left(\tilde{R}^s_c + R^s_{c0}\right) \leq 0 \tag{32}$$

where $\underline{M}^s_{\text{sym}}$ is the Schmid tensor which defines the direction of slip systems and R^s_{c0} is the initial critical resolved shear stress. $\underline{\widetilde{\sigma}}^g, \underline{\widetilde{X}}^g, \tilde{R}^s_c$ are coupled with damage using equations similar to Eq. 28. It remains to define an intrinsic dissipation and dissipation potential that can be established using the Gibbs–Duhem relation:

$$\begin{aligned}\Phi^g &= \Phi^g_{\text{plas}} + \Phi^g_{\text{dam}} \\ &= \left(\underline{\sigma}^g : \dot{\underline{\varepsilon}}^g_{\text{plas}} - \underline{X}^g : \dot{\underline{\beta}}^g - \sum_{s\in g} R^s_c \dot{r}^s\right) + \left(\underline{y}^g : \dot{\underline{d}}^g - \sum_{s\in g} y^s \dot{d}^s\right)\end{aligned} \tag{33}$$

$$\begin{aligned}F^{s\in g} = F^{s\in g}_{\text{plas}} + F^{s\in g}_{\text{dam}} &= \left(\tilde{f}^s + C^s_A r^s \tilde{R}^s_c + C^g_B \underline{\beta}^g : \underline{X}^g\right) \\ &+ \frac{\alpha^s_3}{\left(\alpha^s_4+1\right)\left(1-d^s\right)^{\alpha^s_1}} \left\langle\left\langle \frac{\left(\underline{y}^g \cdot \underline{M}^s_{sym}\right) : \underline{M}^s_{sym} - \left(y^s + y^s_0\right)}{\alpha^s_3} \right\rangle\right\rangle^{\alpha^s_4+1}\end{aligned} \tag{34}$$

Equation 33 is quite original. Usually, the dissipation due to damage is defined additively. The damage between the two scales is then related using a micromechanics relation that is added a posteriori (Boudifa et al. 2006). There are two sets of state variables, but only one independent physical degree of freedom; the link is made in an arbitrary manner (e.g., arithmetic average; Hfaiedh

et al. 2009). On the contrary, with the above approach, no corresponding equations are added. Because of the particular choice, two scales are associated but also two damage mechanisms. One is occurring along the slip systems, while the other takes place more broadly across the grain. This description of the damage is similar to the one used for plasticity, meaning that an equivalent "yield damage surface" (compared to the yield plasticity surface) with an "isotropic hardening" of damage is obtained. The methodology used in plasticity to transpose to the damage is copied on purpose, although strictly there is no flow/yield of damage.

Finally, the damage rate evolution in Eq. 33 is also original. This is a proposed modification of the equation model of Lemaitre et Chaboche (2001). The latter is modified also by analogy with plasticity criteria. The term in Macaulay brackets can be seen as the damage criterion. The choice of the term $\left(\underline{y}^g \cdot \underline{\underline{M}}^s_{\text{sym}}\right) : \underline{\underline{M}}^s_{\text{sym}}$ rather than an expression $\underline{y}^g : \underline{\underline{M}}^s_{\text{sym}}$ will be discussed later. In this formulation, the damage at the mesoscopic scale can be obtained directly from damage in the slip systems, whereas this was not possible in previous studies.

The plastic multiplier can be calculated through different assumptions. Here, the Cailletaud approach is considered for a time-dependent plastic material (i.e., with viscoplasticity). Thus, the plastic multiplier is given by (Besson et al. 2001)

$$\dot{\lambda}^s = \left\langle \left\langle \frac{f^s}{C^s_K} \right\rangle \right\rangle^{C^s_N} \tag{35}$$

The case of time-independent plasticity can be obtained by a specific choice of the material parameters in Eq. 35.

By adding the equations that govern the macroscopic scale (Eq. 7 for localization and Eq. 9 for homogenization, which can be more or less complex (Bornert et al. 2001)), a complete system of differential and algebraic equations is obtained. Entering an increment of strain, the evolution of all unknown variables (at the considered scales) is obtained. The problem lies in the number of material parameters used in this model that has to be experimentally identified. In addition to the presented parameters in Eqs. 30–35, one has to define the number of grains, the number of activated slip systems, the initial critical resolved shear stress, and the ODF. The latter can be obtained by texture measurements (XRD or EBSD). Other parameters can be identified either on a macroscopic mechanical loading test or using diffraction methods coupled with loading test, as described in paragraph 2.

Model for One Phase ($M = 1$) at Three Scale ($N = 3$)

Using the thermodynamics methodology, state and dissipation relations can be detailed for the case of one phase ($M = 1$):

$$\underline{\sigma}^g = \rho^g \frac{\partial \psi^g}{\partial \underline{\varepsilon}^g_{\text{elas}}} = \underline{\underline{\tilde{c}}}^g : \underline{\varepsilon}^g_{\text{elas}} = \underline{\underline{\zeta}}(\underline{d}^g) : \underline{\underline{c}}^g : \underline{\varepsilon}^g_{\text{elas}} : \underline{\underline{\zeta}}(\underline{d}^g) \tag{36}$$

$$R_c^{s\in g} = \rho^g \frac{\partial \psi^g}{\partial r^{s\in g}} = \sum_t \tilde{H}^{st\in g} r^{t\in g} = \sum_t \sqrt{1-d^{s\in g}}\sqrt{1-d^{t\in g}} H^{st\in g} r^{t\in g} \tag{37}$$

$$\underline{X}^g = \rho^g \frac{\partial \psi^g}{\partial \underline{\beta}^g} = \frac{2C_C}{3}\underline{\beta}^g \tag{38}$$

$$\dot{r}^{s\in g} = -\dot{\lambda}^{s\in g}\frac{\partial F^{s\in g}}{\partial \tau_c^{s\in g}} = \frac{\dot{\lambda}^{s\in g}}{\sqrt{1-d^{s\in g}}}\left(1 - C_A^{s\in g} r^{s\in g}\right) \tag{39}$$

$$\dot{\underline{\beta}}^g = -\sum_s \dot{\lambda}^{s\in g}\frac{\partial F^{s\in g}}{\partial \underline{X}^g} = \dot{\underline{\varepsilon}}^g_{\text{plas}} - C_B^g \underline{\beta}^g \sum_s \dot{\lambda}^{s\in g} \tag{40}$$

$$\dot{\underline{\varepsilon}}^g_{\text{plas}} = \sum_s \dot{\lambda}^{s\in g}\frac{\partial F^{s\in g}}{\partial \underline{\sigma}^g} = \underline{\underline{\zeta}}^{-1}\left(\underline{d}^g_{\text{dam}}\right) : \sum_s \dot{\lambda}^{s\in g}\underline{M}^{s\in g}_{\text{sym}} \ \text{sign}\left(\left(\widetilde{\underline{\sigma}}^g - \widetilde{\underline{X}}^g\right) : \underline{M}^{s\in g}_{\text{sym}}\right) \tag{41}$$

where $\dot{\lambda}^{s\in g}$ is the plastic multiplier. As previously, the latter can be calculated using Eq. 35. Relations related to the damage variables are also obtained:

$$\underline{y}^g = -\rho^g \frac{\partial \psi^g}{\partial \underline{d}^g} = -\frac{1}{2}\underline{\varepsilon}^g_{\text{elas}} : \frac{\partial \widetilde{\underline{\underline{c}}}^g_{\text{elas}}}{\partial \underline{d}^g} : \underline{\varepsilon}^g_{\text{elas}} \simeq \frac{1}{3}\frac{\underline{d}^g}{\|\underline{d}\|^g} E^{g,\,und}_{\text{elas}} \overset{\left(\underline{d}^g \approx d^g \underline{1}\right)}{\approx} \frac{E^{g,\,und}_{\text{elas}}}{3}\underline{1} \tag{42}$$

$$y^{s\in g} = \rho^g \frac{\partial \psi^g}{\partial d^{s\in g}} = -\sum_t \frac{\sqrt{1-d^{t\in g}}}{2\sqrt{1-d^{s\in g}}} H^{st\in g} r^{s\in g} r^{t\in g} \overset{(\forall s,t,\, d^s \approx d^t)}{\approx} -E^{s\in g,\,\text{und}}_{\text{plas}} \tag{43}$$

$$\dot{\underline{d}}^g = \sum_s \dot{\lambda}^{s\in F}\frac{\partial F^{s\in F}_{\text{dam}}}{\partial \underline{y}^F} = \sum_s \dot{d}^{s\in g}\underline{M}^{s\in g}_{\text{sym}} \cdot \underline{M}^{s\in g}_{\text{sym}} \tag{44}$$

$$\dot{d}^{s\in g} = \dot{\lambda}^{s\in g}\frac{\partial F^{s\in g}_{\text{dam}}}{\partial y^{s\in g}} = \frac{\dot{\lambda}^{s\in g}}{\left(1-d^{s\in g}\right)^{\alpha_1^{s\in g}}}\left\langle\left\langle \frac{\left(\underline{y}^g \cdot \underline{M}^{s\in g}_{\text{sym}}\right) : \underline{M}^{s\in g}_{\text{sym}} - \left(y^{s\in g} + y_0^{s\in g}\right)}{\alpha_3^{s\in g}}\right\rangle\right\rangle^{\alpha_4^{s\in g}} \tag{45}$$

$$\approx \frac{\dot{\lambda}^{s\in g}}{\left(1-d^{s\in g}\right)^{\alpha_1^{s\in g}}}\left\langle\left\langle \frac{E^{g,\,und}_{\text{elas}}\left\|\underline{M}^{s\in g}_{\text{sym}}\right\|^2 + E^{s\in g,\,\text{und}}_{\text{plas}} - y_0^{s\in g}}{\alpha_3^{s\in g}}\right\rangle\right\rangle^{\alpha_4^{s\in g}} \tag{46}$$

where $E^{g,\text{und}}_{\text{elas}}$ is the elastic energy in an undamaged grain and $E^{s\in g,\text{und}}_{\text{plas}}$ refers to the energy accumulated along an undamaged slip system. In Eq. 41, as an illustration, it has been chosen to introduce the following relation for the damage operator

$\underline{\underline{\zeta}}(\underline{d}^g) = (1 - \|\underline{d}^g\|)^{+\frac{1}{2}}\underline{\underline{1}}$. It leads to a damage force $\underline{y}^F$ that depends linearly on $\underline{d}^F$ on the direction of the grain scale and is proportional to the elastic energy stored in the grain. It corresponds to a first mechanism responsible for damage. A simple coupling function $\underline{\underline{\zeta}}(\underline{d}^g)$ (e.g., with the trace) would lead to an isotropic damage force $\underline{y}^F$, even for an anisotropic damage $\underline{d}^F$. More complicated function with anisotropic coupling (e.g., with a tensor product) would lead to a damage force $\underline{y}^F$ that depends nonlinearly on $\underline{d}^F$, enhancing the anisotropy of damage. Our choice is eventually based on a compromise between the complexity of the description at grain scale and the simplicity of the effect of anisotropy at the same scale.

In Eq. 43, the damage force at the microscopic level that depends linearly on $y^{s\in g}$ may be identified approximately with the plastic energy stored in the system by the slide of dislocations. This is a second mechanism responsible for damage.

Using the thermodynamics potentials, the relationship between mesoscopic and microscopic damage is directly obtained by the thermodynamic formulation in Eq. 44, by substituting Eq. 45. The latter corresponds to the damage evolution in slip systems, but depends simultaneously on plastic energy (at the same scale) and elastic energy (at the upper scale), meaning that a combined effect of the two mechanisms has been introduced, which was not the case in existing approaches.

Model for Two Phases ($M = 2$) at Three Scales ($N = 3$)

The previous equations could be extended for two phases. Grains of phase A and of B are considered. For example, it is assumed that damage can appear in one of the phases (the phase B); it leads to the same equations as Eqs. 36–46 for the phase B, whereas only the Eqs. 36–41 are still valid.

Numerical Aspects and Discussions

The numerical implementation of the micro–macro polycrystalline plasticity model into FEA code follows the classical way. Each integration (or Gauss) point of each macroscopic finite element is represented by an aggregate of a finite number of grains. This aggregate corresponds to a typical representative volume element (RVE) to which the macroscopic strain tensor obtained from the classical FEA is homogeneously applied. The mechanical answer of this RVE in terms of macroscopic total strain tensor is computed following the self-consistent model presented above (see, e.g., Anand 2004; Cailletaud et al. 2003; Habraken and Duchene 2004; Miehe and Schotte 2004; Raabe and Roters 2004). In this work, the model developed above has been implemented in the Z-MAT package connected to the FE code ZéBuLoN (ZéBuLoN 2008) according to the description given in Cailletaud et al. (2003).

For the global resolution scheme, the ZéBuLoN solver with static implicit resolution procedure is used. For more complex calculations (e.g., metal forming), the dynamic explicit resolution strategy is used, with ABAQUS/explicit FE code

connected with the material models in Z_MAT using the interface Z_ABA that connects ABAQUS to Z_MAT (see ZéBuLoN user's manual (ZéBuLoN 2008)). The dynamic explicit resolution scheme is preferred since it avoids the iteration process as well as the calculation of the consistent tangent matrix (see Saanouni 2012).

Applications and Numerical Results

Tests for a Single-Phase Material

Parametric Study of the Proposed Model

The constitutive equations developed in previous sections have been implemented into the general purpose finite element code ZéBuLoN (2008) in order to perform appropriate numerical simulations. In the present paper, only comments on numerical results obtained with the proposed model are given. This is done through a relatively detailed parametric study conducted on both the RVE as a material point as well as tensile specimen modeled as 3D solid. Different number and initial orientation of grains have been investigated.

The constitutive equations and their numerical integration, using 4th-order Runge–Kutta integration scheme, have been implemented with a subroutine as indicated in ZéBuLoN users' manual (ZéBuLoN 2008). Each integration (or Gauss) point of each finite element is modeled as an aggregate of N^g grains. In each grain, $N^{s\in g} \leq 12$ crystallographic slip systems can be potentially active. A loop is performed over the total number of the aggregate grains, and the stress tensor $\underline{\sigma}^g$ for the grain under concern is calculated using an adapted localization equation (Hfaiedh 2009). For each of these grains, a new loop over the 12 slip systems is performed in order to calculate the resolved shear stress using $\tau^s = \underline{\widetilde{\sigma}}^g : \underline{M}^s_{\text{sym}}$, as well as all the related mechanical fields including hardening and damage evolution (see Eqs. 36–46). The rotation tensor of each grain is also computed allowing the update of the texture during the loading path. For this evolution problem with N^e elements each having N^G Gauss points in the structure, the number of unknowns to be computed at each load increment is around $N^e \times N^G \times (7 + 45N^g)$.

Application to the RVE

This section is devoted to studying the effect of the number of grains N^g, as well as their initial orientations, on the numerical results predicted by the proposed model. Each investigated aggregate is submitted to an imposed total strain in one direction. In other words, the macroscopic strain component E_{11} is imposed, and all the stress components are fixed to zero, except the component Σ_{11}. Results are analyzed in terms of the evolution of the macroscopic von Mises equivalent stress $\Sigma_{eq} = \sqrt{3\underline{\Sigma}^{\text{dev}} : \underline{\Sigma}^{\text{dev}}}/2$ versus the macroscopic accumulated plastic strain $P = \sqrt{3\underline{E}^{\text{dev}} : \underline{E}^{\text{dev}}}/2$, the macroscopic damage D_{dam} versus P, and the Euler angles $(\varphi_1, \phi, \varphi_2)$ versus P.

Table 3 Elastoplastic parameters of the steel material for the RVE

$E(MPa)$, ν	$C_C(MPa)$	$C_A^{s\in g}$	$C_N^{s\in g}$
200000, 0.3	30067	26,7	25
$C_K^{s\in g}$(MPa)	$R_{c0}^{s\in g}$(MPa)	$H^{st\in g}$ (MPa)	C_B^g
50	145	50	74.4
$\alpha_3^{s\in g}$(MPa)	$\alpha_1^{s\in g}$	$\alpha_4^{s\in g}$	$y_0^{s\in g}$(MPa)
0.84	2	60	0.001

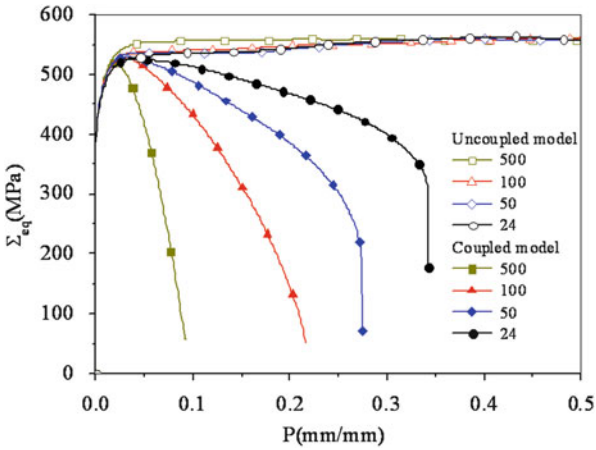


Fig. 25 Effect of N^g on the macroscopic equivalent stress–equivalent plastic strain curves, for uncoupled and fully coupled models, with texture evolution (Hfaiedh et al. 2009)

Concerning the values of material parameters, they are supposed to be the same for all the aggregates, and their numerical values are taken from Hfaiedh (2009) (defined for $N^g = 24$), as summarized in Table 3.

Let us examine the effect of N^g in both the fully coupled (i.e., with damage effect) and uncoupled (without damage effect) calculations by highlighting the effect of the damage on the textural evolution. For the considered material point (or aggregate), the four values of N^g: N^g = 24, 50, 100, and 500 grains are investigated.

Figure 25 shows the comparison between the stress–strain curves for the four values of N^g and for both the uncoupled and fully coupled solutions. From this figure, it appears that for the uncoupled model, the solution is quasi-independent from N^g when the hardening is fully saturated. As expected, the best solution is obtained with the highest number of grains ($N^g = 500$), while the lower number of grains ($N^g = 24$) gives the same saturated equivalent stress $\Sigma_{eq} = 550$ MPa for $P \geq 30$ %. However, for the fully coupled model, the numerical solutions are highly dependent on N^g . Indeed, the higher is the value of N^g, the lower is the value of the material ductility (i.e., the macroscopic accumulated plastic strain at fracture). This means that the damage rate is higher for the aggregates with higher number of grains as clearly confirmed in Fig. 26, where the total macroscopic damage D_{dam} is plotted versus the macroscopic accumulated plastic strain P.

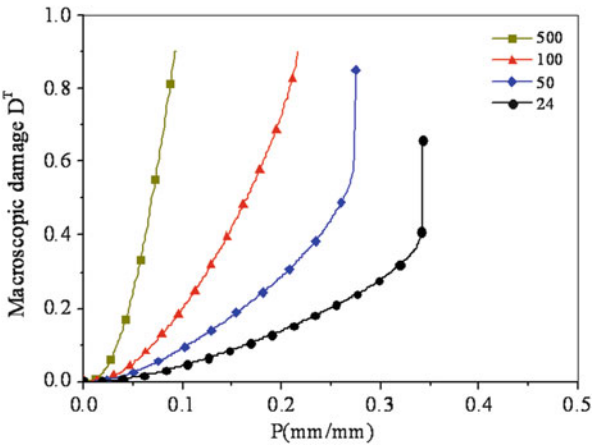


Fig. 26 Effect of N^g on the macroscopic damage evolution in fully coupled model with texture evolution (Hfaiedh et al. 2009)

Application to a Uniaxial Tensile Test

This section is devoted to the application of the proposed model to a tensile test in which the specimen is modeled in 3D. The specimen with dimensions of $36 \times 12 \times 1$ mm^3 is loaded with an imposed constant displacement rate of $\dot{u}_1 = 0$ mm/s, $\dot{u}_2 = 0.1$ mm/s, and $\dot{u}_3 = 0$ mm/s. It is applied to the top side of the specimen, while its bottom side is fully clamped ($u_1 = u_2 = u_3 = 0$). The specimen is meshed with 432 elements ($36 \times 12 \times 1$) of size $\Delta x = 1$ mm and type C3d20R, taken from the ZÉBULON element library (ZéBuLoN 2008). In order to minimize the CPU time, only one element is put through the specimen thickness. C3d20R is a 3D (solid) quadratic hexahedral element with 20 nodes and a reduced integration with only four quadrature or Gauss points. The values of material parameters are given in Table 3. To illustrate the influence of the initial microstructure on the numerically predicted results with coupled and uncoupled models, several aggregates with respectively $N^g = 1$, 12, 24, and 100 grains were considered. To save the CPU time, the effect of the texture evolution has been investigated using only the aggregate with $N^g = 24$ grains.

Effect of the Aggregate and Texture on Plastic Localization

To investigate the effect of the polycrystalline aggregate on the mode of localization (leading to the formation of the well-known shear bands), two different initial orientations of grains (texture 1 and texture 2) inside three aggregates having $N^g = 1$, 24 and 100 grains are considered. The texture evolution has been taken into account.

Figure 27 shows the distribution of the macroscopic accumulated plastic strain inside the specimen for two different values of the imposed displacement. It corresponds to the prediction of the uncoupled model but accounting for the texture evolution. The plastic strain localization mode is sensitive to the number of grains inside each aggregate as well as on the initial orientation of the grains. The width of the shear band is smaller for the small number of grains ($N^g = 24$) as seen from the

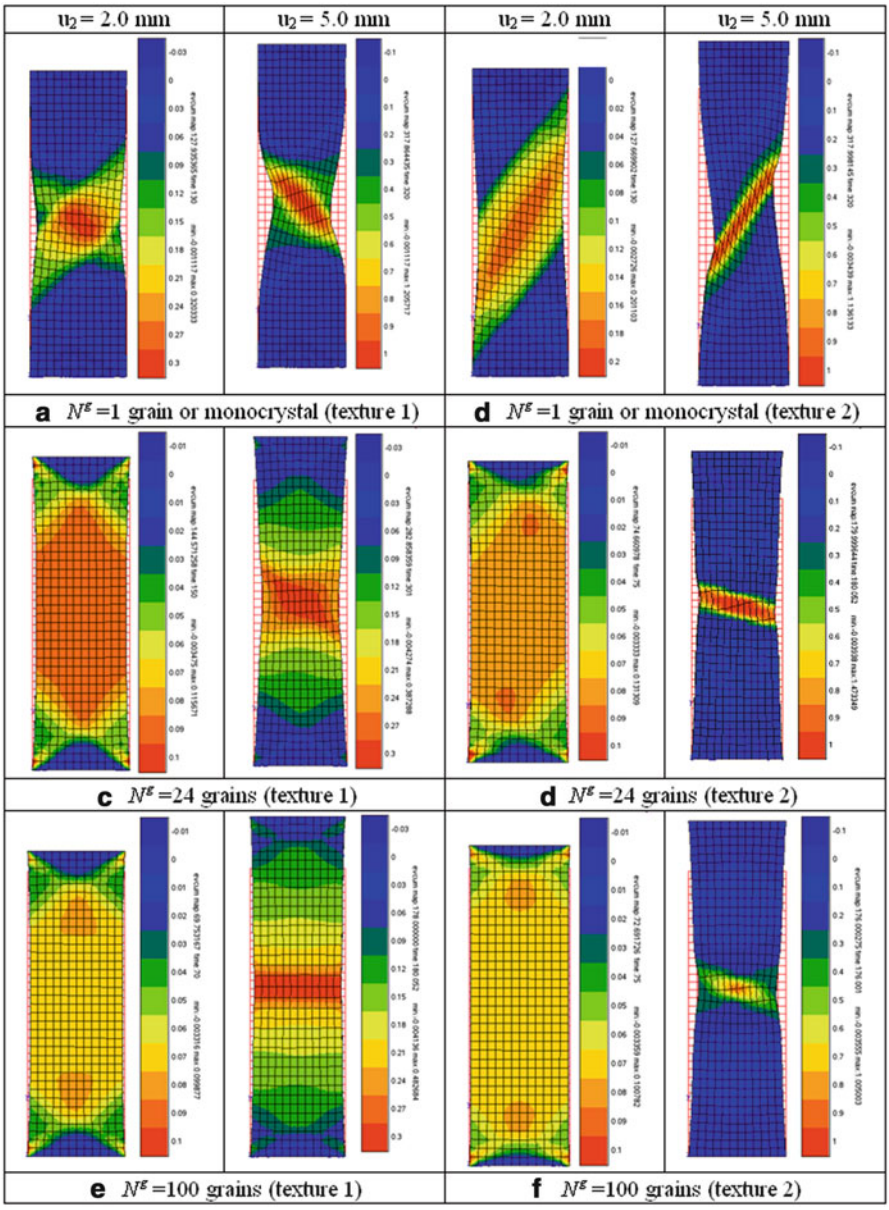


Fig. 27 Distributions of the macroscopic plastic strain at two displacement values for various N^g (Hfaiedh et al. 2009)

comparison of the monocrystalline aggregate (Fig. 27a and b). Differences occur also for the aggregate with $N^g = 24$ (Fig. 27c and d) and with $N^g = 100$ (Fig. 27e and f). Also the localization of plastic strain seems more severe for the texture 2 as seen from Fig. 27b, d, and f. It is worth noting that the angle made by the shear

bands with respect to the specimen axis (or the loading direction) is higher for the high values of N^g as observed from these figures. Particularly, for $N^g = 100$ grains, this angle is exactly 90° for the texture 1 (Fig. 27e) and near 80° for the texture 2 (Fig. 27f).

The same remarks apply for the distribution of the macroscopic damage presented in Fig. 28, where the localization of the damage inside the shear bands seems more severe than the plastic strain. Note also from Fig. 28 that the final fracture occurs early for the high values of the grain number: $u_2^{fr} \geq 5.0$ mm for the two initial textures of the monocrystalline aggregate (Fig. 28a and b), $3.25 \leq u_2^{fr} \leq 3.95$ mm for the aggregate with $N^g = 24$ (Fig. 28c and d), and $2.40 \leq u_2^{fr} \leq 2.58$ mm for the aggregate with $N^g = 100$ (Fig. 28e and f).

Figure 29 shows the global specimen responses in terms of force-displacement curves predicted by the fully coupled and the uncoupled models for the three aggregates defined by $N^g = 12$, 24 and 100 grains with the same initial orientation (texture 1). Note that the force-displacement curves are approximately similar for the uncoupled calculations for the three aggregates for which the observed softening is only due to the diffuse necking. However, for the fully coupled solution, the final fracture occurs as early as the aggregate has a high number of grains as can be expected. The strong softening observed in these cases is due to the combined effect of necking and damage (macroscopic crack propagation inside the shear bands).

Now consider the important case of a "monocrystal" (defined here as the aggregate of several grains with initially the same orientation). Taking into account, the texture evolution enabled to highlight the transformation of a single-oriented polycrystalline aggregate into a multiple-oriented polycrystalline aggregate as the plastic strain increases. Indeed, if a calculation is performed starting from a single-oriented polycrystalline aggregate, one observes that the macroscopic plastic strain increases. Disorientations occur due to the high strain gradient inside the shear bands where the plastic strain is highly localized as can be observed in Fig. 30. In fact, the orientation of 24 grains located around the shear band has been determined for different increasing values of the macroscopic plastic strain. Figure 30 summarizes the poles figures for the crystallographic planes {200} and {220}. It is worth noting how a material point with initially a single crystalline orientation (Fig. 30a for $E_{22} = 0.0$ %) is transformed into a polycrystal as the macroscopic plastic strain increases as indicated in Fig. 30b for $E_{22} = 50.0$ % and Fig. 30c for $E_{22} = 100.0$ %. Additional investigations would make it possible to see if this phenomenon is systematic for different initial orientations of the "monocrystal" and if this mode of transformation corresponds to the experimental observations.

Effect of the Texture Evolution

In this section, the effect of the texture evolution in comparisons with calculations performed with a constant texture in both fully coupled and uncoupled models is investigated. For the sake of shortness and to save the CPU time, only the aggregate with $N^g = 24$ is investigated.

For the uncoupled model (Figs. 31 and 32), one can note that if the texture does not evolve (no grain rotation), then the softening induced by the necking appears

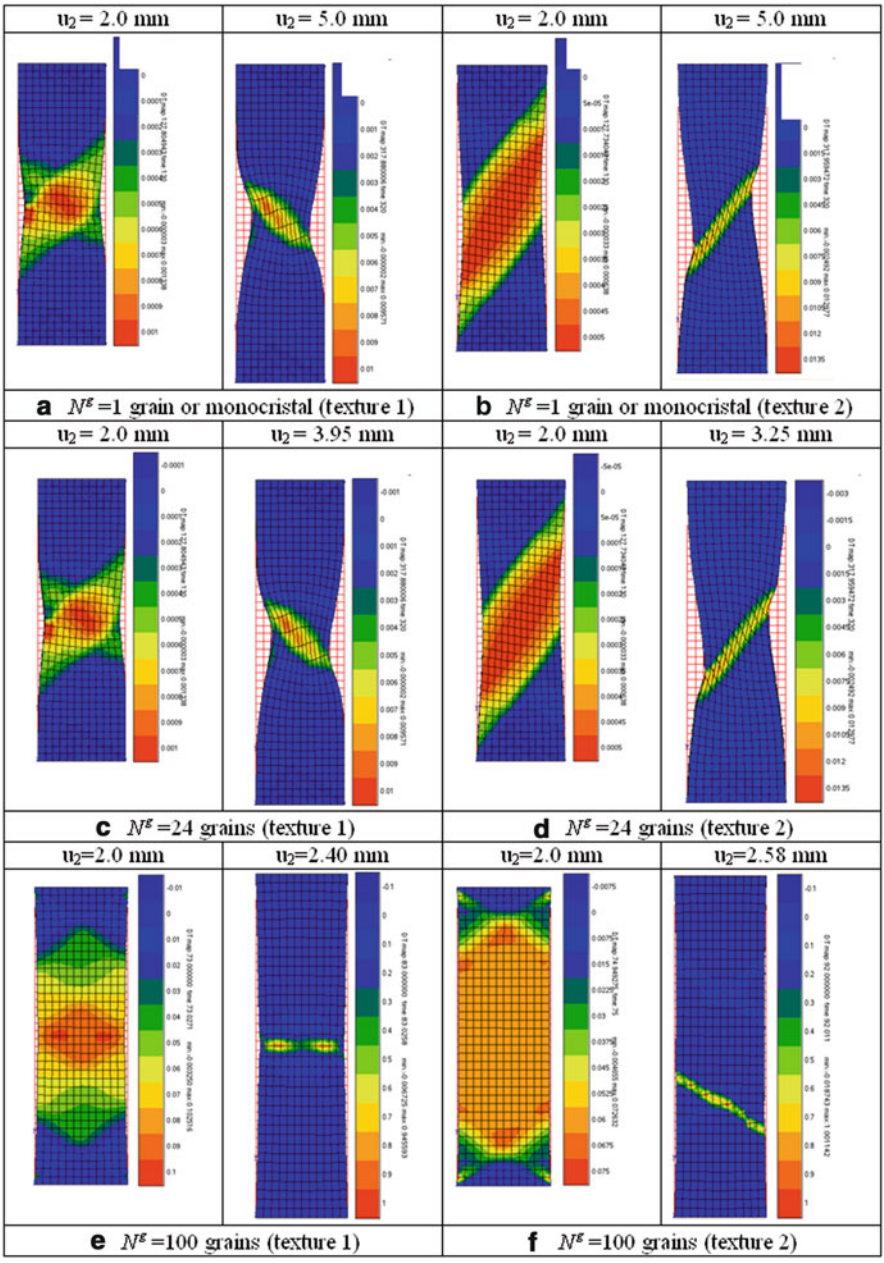


Fig. 28 Distribution of the macroscopic damage at different displacement values of N^g (Hfaiedh et al. 2009)

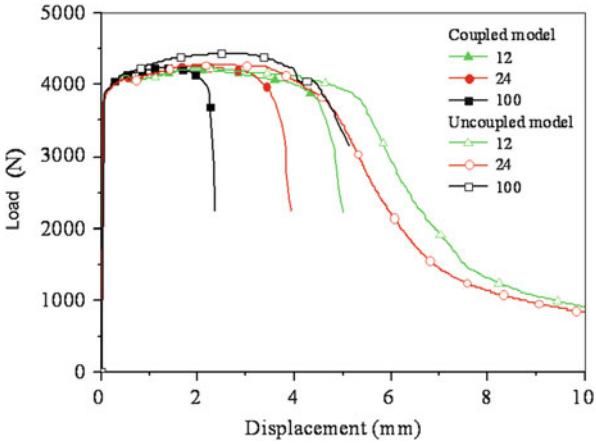


Fig. 29 Comparison of the total answers in terms of force-displacement curves for aggregates with various numbers of grains and various initial textures in coupled and uncoupled models (Hfaiedh et al. 2009)

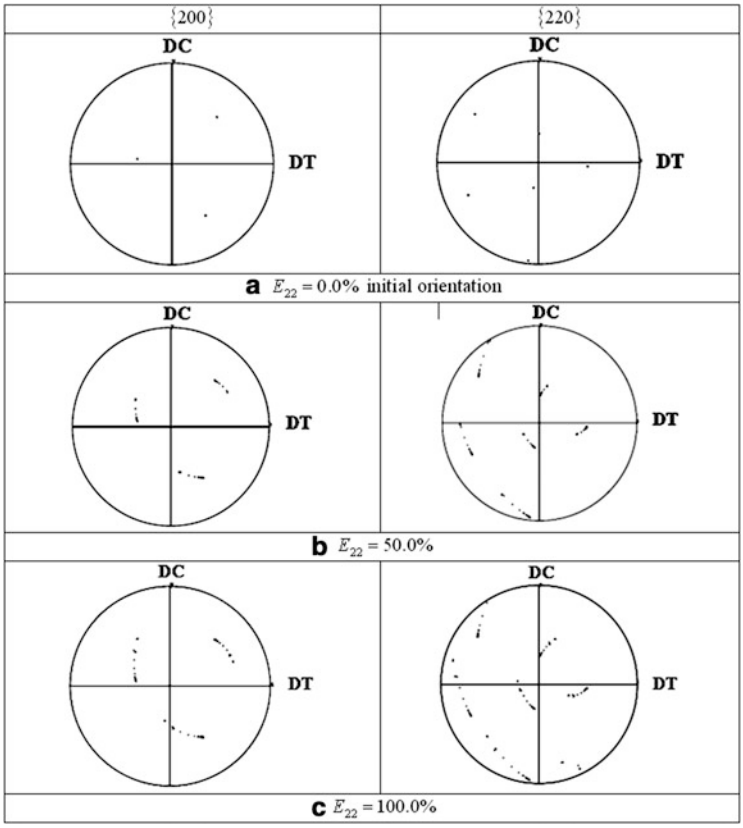


Fig. 30 Poles figures for the two planes {200} and {220}, obtained from 24 grains with the same initial orientation (Hfaiedh et al. 2009)

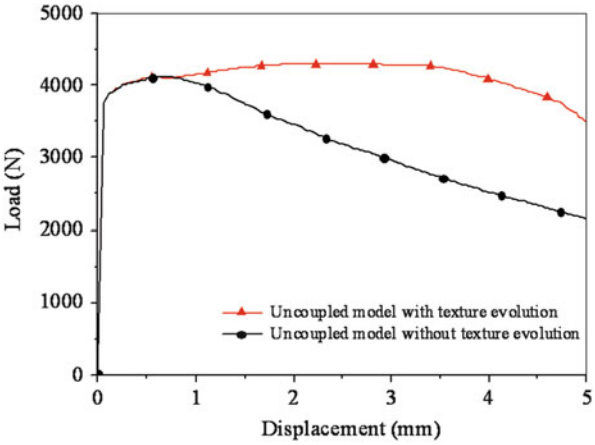


Fig. 31 Comparison of the force-displacement curves for an aggregate of 24 grains with random texture, predicted by the uncoupled model (Hfaiedh et al. 2009)

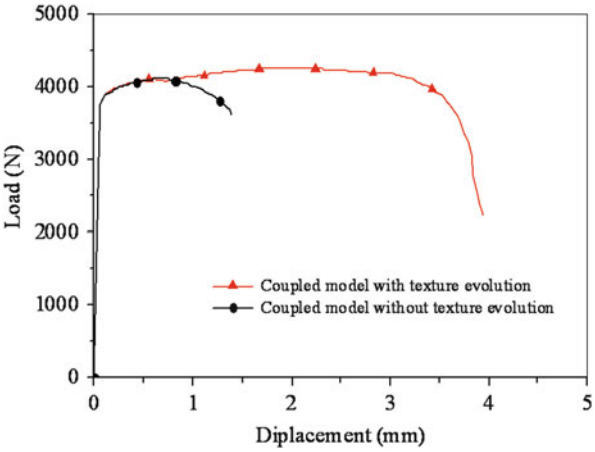


Fig. 32 Comparison of the force-displacement curves for an aggregate of 24 grains with a random texture, predicted by the fully coupled model (Hfaiedh et al. 2009)

rather early around $u = 0.7$ mm. However, when the grain rotation is accounted for, a clear delay in the appearance of the softening stage is observed and the maximum of the force is observed for around $u = 3.5$ mm. This delay can be explained by the increase of the strain hardening due to the multiplicity of active systems related to the rotation of the grains. The same observations can be made for the fully coupled model. It predicts a fracture occurring at the displacement $u = 1.5$ mm when the texture evolution is neglected, while the fracture occurs later at $u = 4.0$ mm when the texture evolution is taken into account. This is again due to the enhancement of the hardening stage when the grains rotate giving the opportunity to more slip systems to be active. These remarks are confirmed by the distribution of the macroscopic plastic strain at fracture (or ductility) as shown in Fig. 33. It can be clearly seen that $P \approx 33$ % when the texture evolution is accounted for and $P \approx 8.25$ % when the rotation of the grains is neglected.

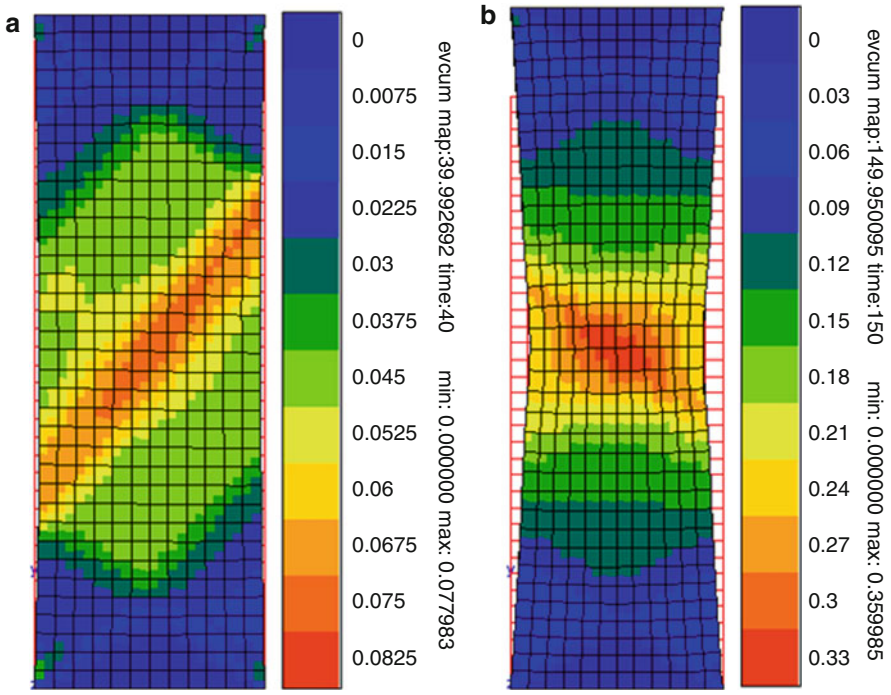


Fig. 33 Macroscopic equivalent plastic strain distribution for an initially random aggregate of 24 grains, predicted by the fully coupled model: (**a**) without texture evolution and (**b**) with texture evolution (Hfaiedh et al. 2009)

Application to Copper

In this last section, a comparison of the numerical results with experimental (macroscopic and mesoscopic) data obtained for copper is proposed. The macroscopic data includes force-displacement curves obtained in uniaxial tensile tests (see Fig. 34). The values of the material parameters for copper are summarized in Table 4.

Figure 34 shows the global specimen responses in terms of force-displacement curves, predicted by the fully coupled and the uncoupled models and experiment for the copper. The results for the coupled model approximate the shape of the experimental curve in the critical stage. The effect of $\alpha_3^{s\in g}$ parameter is important to obtain the same curves.

The comparison between the experimental poles figures and the numerical simulation at 23 % of equivalent plastic strain is presented for the two crystallographic planes {111} and {220} in Fig. 35. It leads to a relatively good agreement between experiment and calculation, especially the numerical model presents with a small number of grains ($N^g = 24$).

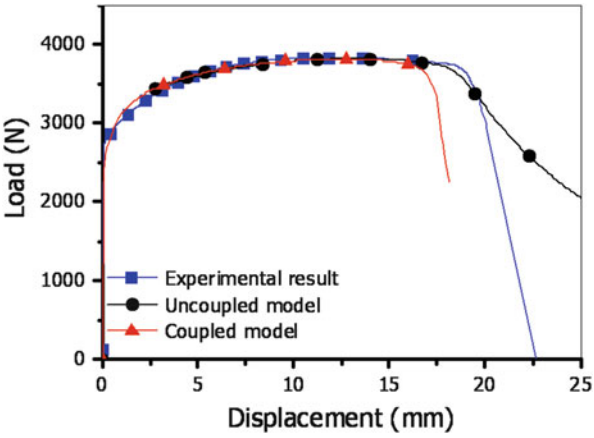


Fig. 34 Comparison of force-displacement curves for the numerical and experimental results (Hfaiedh et al. 2009)

Table 4 Elastoplastic parameters of copper

E (MPa), ν	C_C (MPa)	$C_A^{s\in g}$	$C_N^{s\in g}$
125000., 0.33	*4730*	*15*	*25*
$C_K^{s\in g}$(MPa)	$R_{c0}^{s\in g}$(MPa)	$H^{st\in g}$ (MPa)	C_B^g
50	*51..3*	*85*	*4.4*
$\alpha_3^{s\in g}$(MPa)	$\alpha_1^{s\in g}$	$\alpha_4^{s\in g}$	$y_0^{s\in g}$(MPa)
4	*7.2*	*60*	0.001

Test for Two-Phase Material

This section concerns the simulation of duplex stainless steel behavior. Before the identification stage of final material parameters, a study of their influence has been performed. The comparison between experimental results from Experimental aspects and the following curves is only qualitative, but can show the capacity of the previous model to represent the behavior of a two-phase material. The point of this work is to illustrate that not only the macroscopic behavior can be reproduced, but smaller scales may also be accurately represented with this modeling approach.

Behavior Before Damage

A study of the influence of plastic parameters on the macroscopic and mesoscopic behaviors of an aggregate of $N_g = 20$ grains is performed before identification with the real material. Several material parameters have been fixed, including the macroscopic and grain elasticity tensors that have been obtained from Le Joncour et al. (2011).

The aggregate is composed of 50 % of each phase, and it is chosen to take austenite (phase 1) and ferrite (phase 2) slip systems for further identifications. In austenite, 12 slip systems (FCC: <110> {111}) are chosen and 24 in ferrite (BCC: <111> {211}, <111> {110}).

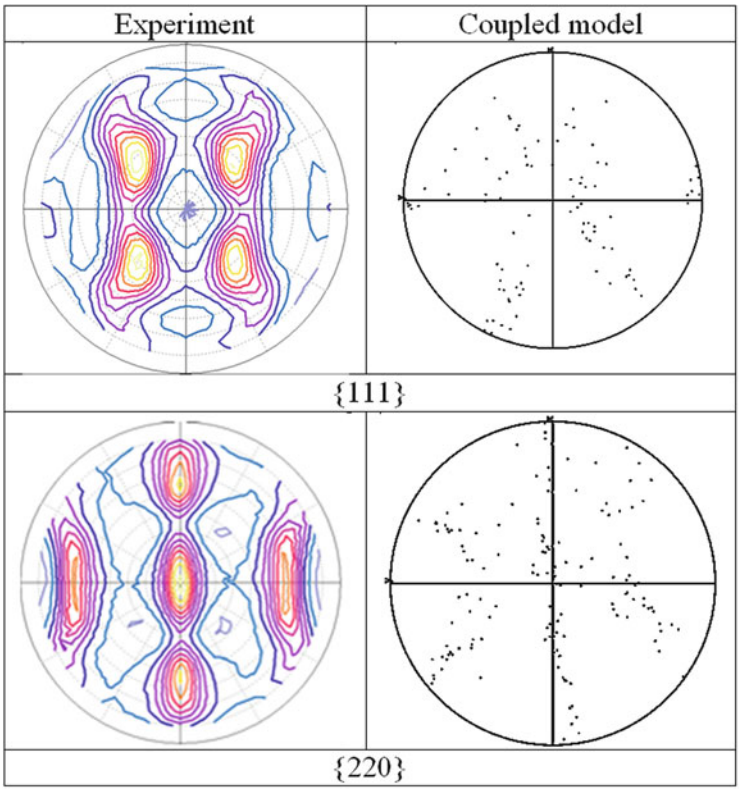


Fig. 35 Comparison between experimental poles figures and coupled model at 23 % of equivalent plastic strain (Hfaiedh et al. 2009)

Table 5 Elastoplastic parameters of austenite

Cijkl(MPa)	C_C*(MPa)*	$C_A^{s\in g}$	$C_N^{s\in g}$
LeJoncour (2011)	Evrard (2008)	*1*, *2.5*, and *5**	*25*
$C_K^{s\in g}$(MPa)	$R_{c0}^{s\in g}$(MPa)	$H^{st\in g}$ (MPa)	C_B^g
50	*75*	*175*,*225*,*275*	0.5, 1, 2, 3, 5*

The elastoplastic parameters chosen for the two phases are presented in Tables 5 and 6.

C_C parameter can be extracted from shear modulus μ of ferrite and austenite, given in the work of Evrard (2008). $C_N^{s\in g}$ and $C_K^{s\in g}$ are taken to minimize the influence of time dependence and calculus time of simulations. The parameter $R_{c0}^{s\in g}$ is easily identified from mesoscopic behavior thresholds and depends on each phase plastic transition and residual stresses (these have been measured and given in Le Joncour et al. (2011)). To avoid damage effect, $y_0^{s\in g}$ is firstly chosen very important (around 10^{11} MPa).

Table 6 Elastoplastic parameters of ferrite

Cijkl (MPa)	C_C (MPa)	$C_A^{s\in g}$	$C_N^{s\in g}$
LeJoncour(2011)	Evrard (2008)	*1*, *2.5*, and *5**	*25*
$C_K^{s\in g}$(MPa)	$R_{c0}^{s\in g}$(MPa)	$H^{st\in g}$ (MPa)	C_B^g
50	*350*	*60*, *110*, *260*	0.5, 1, 2, 3, 5*

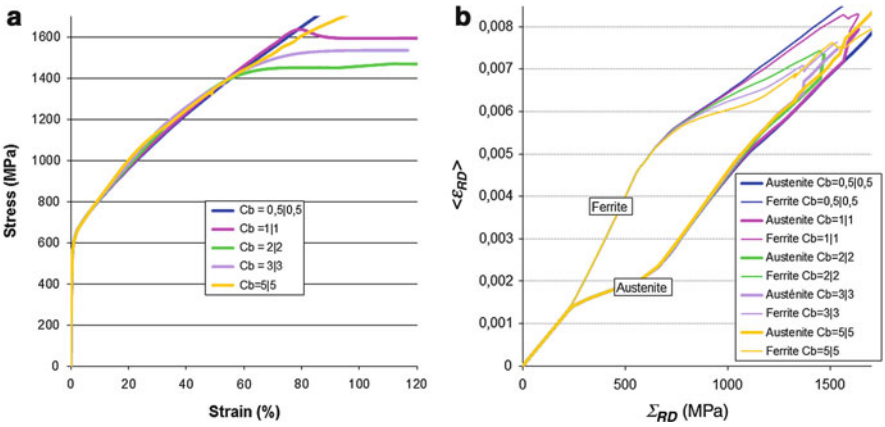


Fig. 36 (**a**) Macroscopic and (**b**) mesoscopic behavior for different values of $C_B^g = 0.5$, 1, 2, 3, and 5

Finally, gray parameters in Tables 5 and 6 have different values; their influence can be observed below in Figs. 36, 37, and 38.

Plasticity Parameter Influence

The parameter C_B^g controls the nonlinearity of kinematics hardening, and this causes, for the aggregate of few grains, the frame of the macroscopic curve (Fig. 36a). This influences the elastic strains of grains versus macroscopic stress in the same way as Fig. 36b, increasing the nonlinearity of the curves of both phases.

The $H^{st\in g}$ interaction matrix components have been taken equal to each other; therefore, this parameter is setting the intensity of isotropic hardening. Figure 37a shows this influence; by increasing H, the slope of macroscopic curves during plasticity increases and saturated stress is then heightened. This influences the mesoscopic curves (Fig. 37b) by changing their slopes after plastification and the final stresses in both phases.

In Fig. 38, influence of $C_A^{s\in g}$ parameters which sets the speed of saturation of hardening can be seen. The increase of this parameter causes the decrease of final stress in ferrite and austenite on curves at phases scale.

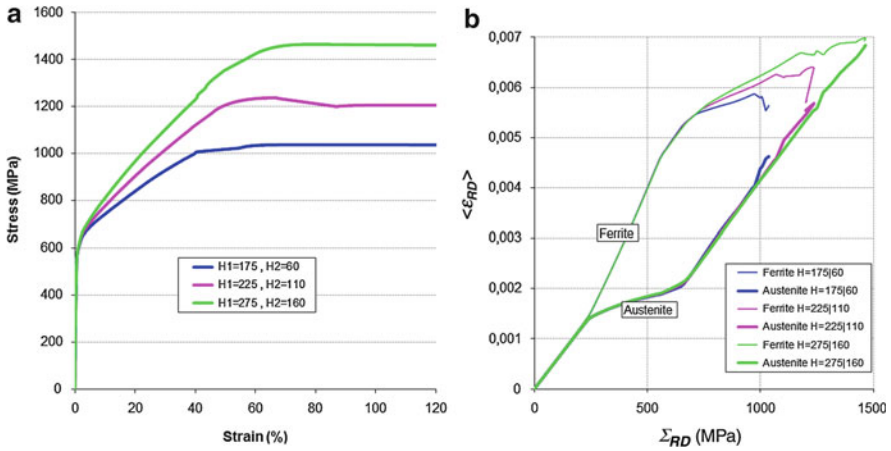


Fig. 37 (**a**) Macroscopic and (**b**) mesoscopic behavior for different values of $H^{st\in g}$, Phase 1: 175, 225, 275, and Phase 2: 60, 110, 260

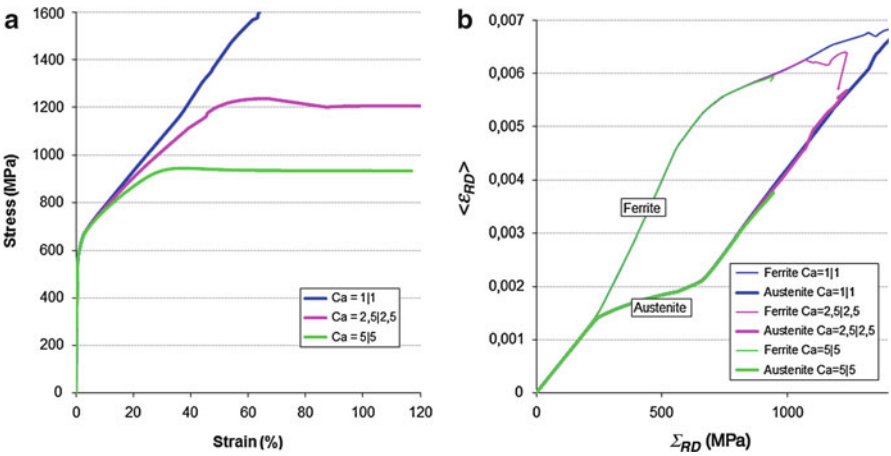


Fig. 38 (**a**) Macroscopic and (**b**) mesoscopic behavior for different values of $C_A^{s\in g} = 1, 2.5, 5$

Table 7 Plastic parameters for simulation of Fig. 39a and b, for the damage parametric study

Austemte		
C_B^g	$H^{st\in g}$ (MPa)	$C_A^{s\in g}$
30	*260*	*1.3*
Ferrite		
C_B^g	$H^{st\in g}$ (MPa)	$C_A^{s\in g}$
30	*130*	*1.5*

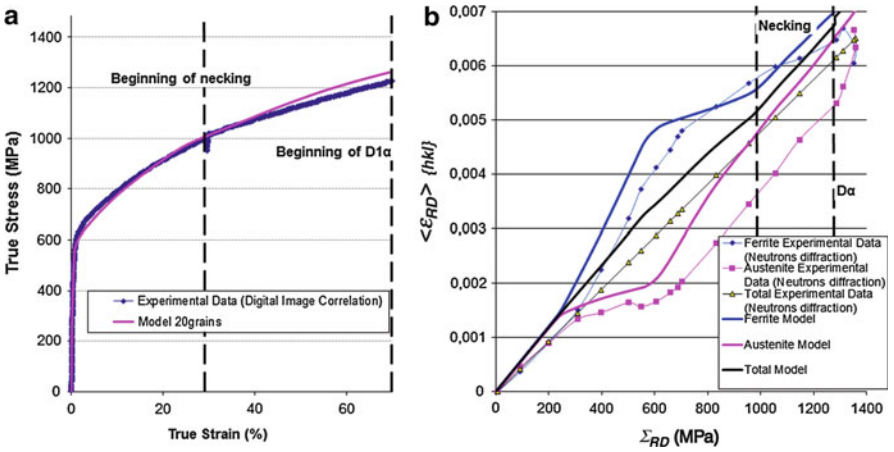


Fig. 39 (**a**) Macroscopic and (**b**) mesoscopic behavior, comparison between simulation and experimental data

Table 8 Damage parameters in ferrite

$\alpha_3^{s\in g}$(MPa)	$\alpha_1^{s\in g}$	$\alpha_4^{s\in g}$	$y_0^{s\in g}$(MPa)
0.5, 5, 25, 50, 60, 100, 1,000	1, 3, 6, 10. 60	1, 2, 3, 5, 10	5, 10, 25, 40, 50, 100

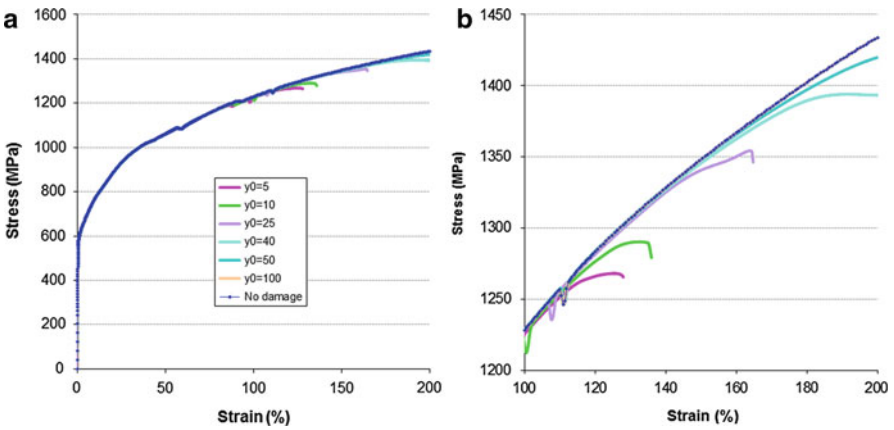


Fig. 40 (**a**) Macroscopic curve and (**b**) magnification of the behavior with different values of damage parameter $y_0^{s\in g}$

Comparison with Experimental Data

After the observation of plastic parameters, values have been chosen to obtain macroscopic local curve of UR45N presented in section II, for an aggregate of 20 grains. These parameters are given in Table 7.

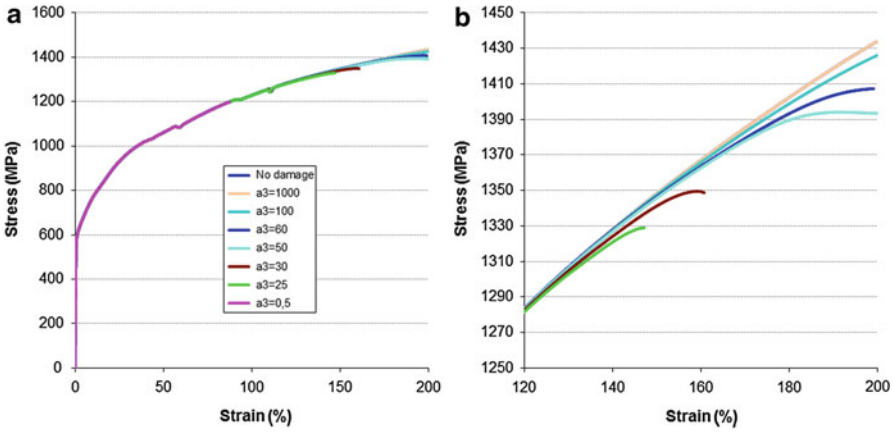


Fig. 41 (**a**) Macroscopic curve and (**b**) magnification of the behavior with different values of damage parameter $\alpha_3^{s \in g}$

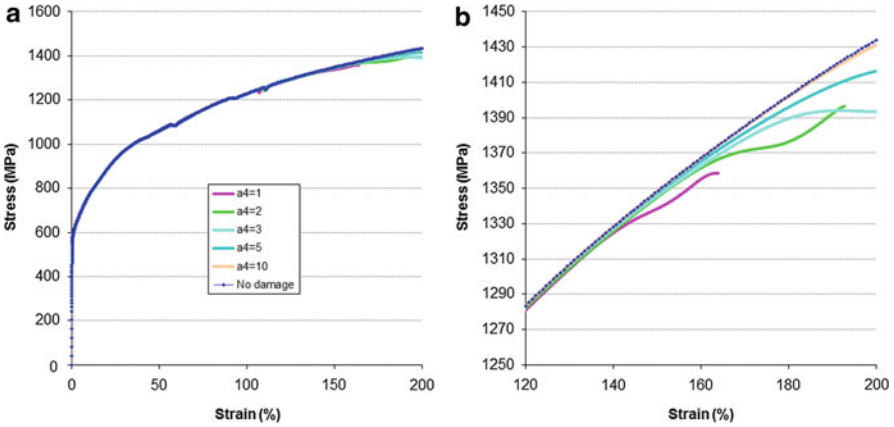


Fig. 42 (**a**) Macroscopic curve and (**b**) magnification of the behavior with different values of damage parameter $\alpha_4^{s \in g}$

On Fig. 39a, a good correlation between experiment and simulation for the macroscopic behavior can be seen. Only the beginning of the curve until 70 % of strain has been used for the comparison because of the use of plasticity without damage. After this threshold, consequences of damage on elastic strains per lattice plans have been observed on the experiments; therefore, a modeling of damage is necessary after this limit.

On Fig. 39b, it is shown that correlation of experimental and simulated curves for phase scale is not as good as for the macroscopic scale. Better material parameters could probably be found, and calculations have to be launched on an

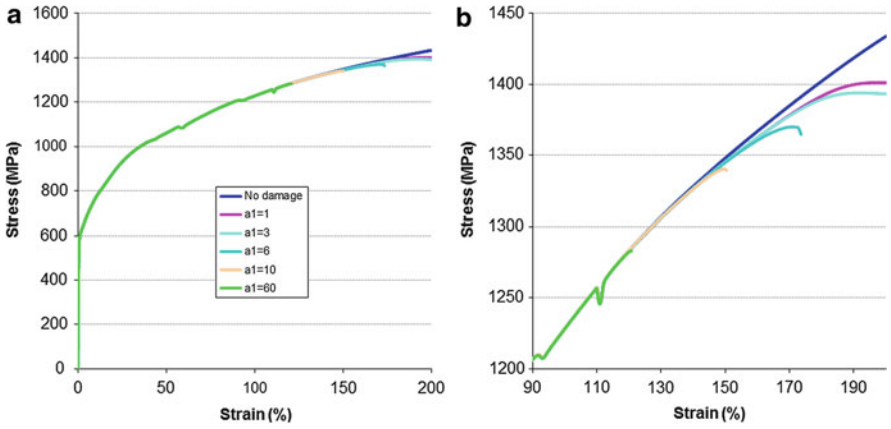


Fig. 43 (**a**) Macroscopic curve and (**b**) magnification of the behavior with different values of damage parameter $\alpha_1^{s\in g}$

aggregate with more grains to be representative. Still, evolutions and trends are qualitatively the same. The balance between phases and thresholds corresponding to yield stress is observed.

Damage Parameter Influence

In this section, a parametric study of damage is proposed. Each of the four parameters has been changed considering the same set of elastoplastic parameters found in the latest paragraph for UR45N duplex stainless steel. Simulation below has been performed on an aggregate of $N_g = 10$ grains to limit the time of calculation. The ferrite is the only phase that can damage; therefore, $y_0^{s\in g}$ of austenite is fixed as before and ferrite damage parameters are changing as presented below in Table 8.

Conclusions and Perspectives

In the present chapter, a general methodology has been presented and illustrated with applications to different materials. It has been shown that ductile damage can be investigated using micromechanics. However, it is necessary to emphasize the different difficulties that have been encountered and the associated solutions.

First, to focus on the microscopic scale, experimental techniques have been investigated through the possibility to obtain constitutive models. Diffraction methods seem to be the most efficient when it is necessary to focus on the microscopic scale. Indeed, it enables to determine the material behavior selectively for each phase. However, the possibility to extract damage is not directly possible, and specific data treatments are required. The comparison of damage at macroscopic scale is then necessary to elaborate some damage evolution models, through the construction of damage scenario at several scales.

Second, to obtain efficient modeling for the simulation of virtual process forming, it is mandatory to have a permanent exchange with experiments. The experimental methods proposed in the present chapter have proven their usefulness. The modeling approach is mainly based on a thermodynamics approach that enables to ensure the validity of the physical laws. In addition to the elastic and plastic behavior, ductile damage is introduced through intern variables, whose kinematic contributions represent the fraction of defects in the RVE. The number of scales used, as well as the number of phases, determines the type of model that is necessary. From experimental observations, at least three scales have to be taken into account for the modeling of ductile damage. Moreover, it represents a coherent approach, provided the scale transitions are properly done. The proposed modeling method extends the classical formulation, which include some phenomenological relations, by replacing equations with thermodynamically compatible relations. However, the restriction is limited and several are still possible, which justify the exchange with experiments.

Third, the proposed models have been tested on several materials (monophasic such as copper and biphasic such as duplex stainless steels). Provided the formulation is built as objective and texture is updated through the evolution of the Eulerian angles, it is possible to reproduce quantitatively the mechanical behavior of materials up to failure. It can thus be applied to the simulation of metal forming processes.

Acknowledgments Authors would like to thank gratefully the different people that are indirectly participants of this work through their collaboration to the theoretical, numerical, and experimental aspects: Manuel François, Arjen Roos, Andrzej Baczmanski, and Chedly Braham. We would also thank Emmanuelle Rouhaud for the time that she generously spent to correct this entire chapter. All these collaborations have allowed progressing on this particular domain and making the present authors as efficient as possible.

References

L. Anand, Single-crystal elasto-viscoplasticity: application to texture evolution in polycrystalline metals at large strains. Comput. Meth. Appl. Mech. Eng. **193**(48–51), 5359–5383 (2004)

R. Asaro, V. Lubards, *Mechanics of Solids and Materials* (Cambridge University Press, Cambridge, 2006)

A. Baczmanski, Habilitation thesis: stress field in polycrystalline materials studied using diffraction and self-consistent modeling, Ph.D. thesis, Informatyki Stosowanej Akademia Gorniczo-Hutnicza Krakow, 2005

A. Baczmanski, L. Le Joncour, B. Panicaud, M. Francois, C. Braham, A.M. Paradowska, S. Wronski, S. Amara, R. Chiron, Neutron time-of-flight diffraction used to study aged duplex stainless steel at small and large strain until sample fracture. J. Appl. Crystallogr. **44**, 966–982 (2011)

J. Besson, G. Cailletaud, J-L. Chaboche, S. Forest, Mécanique non linéaire des Matériaux, 2001

J.-P. Boehler. Lois de comportement anisotrope des milieux continus. Journal de Mécanique, 17 :153190, 1978

M. Bornert, T. Bretheau, P. Gilormini, *Homogénéisation en Mécanique des Matériaux, Vols. 1 et 2* (ISTE USA, Newport Beach, 2001)

M. Boudifa, Modélisation macro et micro-macro des matériaux polycristallins endommageables avec compressibilité induite, Ph.D. thesis, Université de Technologie de Troyes, 2006.

M. Boudifa, K. Saanouni, J.L. Chaboche, A micromechanical model for inelastic ductile damage prediction in polycrystalline metals. Int. J. Mech. Sci. **51**, 453–464 (2009)

B. Bugat, Comportement et endommagement des aciers austéno-ferritiques vieillis: une approche micromécanique, Ph.D. thesis, Ecole Nationale Supérieure des Mines de Paris, 2000.
H.J. Bunge, *Texture Analysis in Materials Science – Mathematical Methods* (Butterworths, London, 1982)
G. Cailletaud, S. Forest, D. Jeulin, F. Feyel, I. Galliet, V. Mounoury, S. Quilici, Some elements of microstructural mechanics. Comput. Mater. Sci. **27**(3), 351–374 (2003)
V. Calonne, Propagation de fissures par fatigue dans les aciers austéno-ferritiques moulés; influence de la microstructure, du vieillissement et de la température d'essai, Ph.D. thesis, Ecole des mines de Paris, 2001
S. Catalao, X. Feaugas, P. Pilvin, M.-T. Cabrillata, Dipole heights in cyclically deformed polycrystalline aisi 316l stainless steel. Mater. Sci. Eng. A **400–401**, 349–352 (2005)
B.K. Chen, P.F. Thomson, S.K. Choi, Computer modeling of microstructure during hot flat rolling of aluminium. Mater. Sci. Technol. **8**(1), 72–77 (1992)
J.H. Cho, P.R. Dawson, Modeling texture evolution during friction stir welding of stainless steel with comparison to experiments. J. Eng. Mater. Technol. **130**, 1–12 (2008)
J.W. Christian, Plastic deformation of bcc metals, in *International Conference on the Strength of Metals and Alloys* (1970)
B. Clausen, T. Lorentzen, M.A.M. Bourke, M.R. Daymond, Lattice strain evolution during uniaxial tensile loading of stainless steel. Mater. Sci. Eng. A **259**, 17–24 (1999)
R. Dakhlaoui, Analyse du comportement mécanique des constituants d'un alliage polycristallin multiphasé par diffraction des rayons X et neutronique, Ph.D. thesis, Ecole Nationale Supérieur des Arts et Métier de Paris, 2006
P. Dawson, D. Boyce, S. MacEwen, R. Rogge, On the influence of crystal elastic moduli on computed lattice strains in aa-5182 following plastic straining. Mater. Sci. Eng. A **313**, 123–144 (2001)
P.R. Dawson, S.R. MacEwen, P.D. Wu, Advances in sheetmetal forming analyses: dealing with mechanical anisotropy from crystallographic texture. Int. Mater. Rev. **48**(2), 86–122 (2003)
M.R. Daymond, The determination of a continuum mechanics equivalent elastic strain from the analysis of multiple diffraction peaks. J. Appl. Phys. **96**, 4263–4272 (2004)
A. Desestret, J. Charles, Les aciers inoxydables austéno-ferritiques. Les aciers inoxydables, Les éditions de la physique 31–677 (1990)
B. Devincre, L.P. Kubin, C. Lemarchand, R. Madec, Mesoscopic simulations of plastic deformation. Mater. Sci. Eng. A **309–310**, 211–219 (2001)
X. Duan, T. Shepard, Simulation and control of microstructure evolution during hot extrusion of hard aluminium alloy. Mater. Sci. Eng. A **351**(1/2), 282–292 (2003)
C. Eberl, R. Thompson, D. Gianola, Digital image correlation and tracking (2006). http://www.mathworks.com/matlabcentral/fileexchange/12413
A. El Bartali, Apport des mesures de champs cinématiques à l'étude des micromécanismes d'endommagement en fatigue plastique d'un acier inoxydable duplex, Ph.D. thesis, Ecole Centrale de Lille, 2007
P. Evrard, Modélisation polycrystalline du comportement élastoplastique d'un acier inoxydable austéno-ferritique, Thèse de doctorat, Ecole Centrale de Lille, 2008
P. Franciosi, The concept of latent hardening and strain hardening in metallic single crystals. Acta Metall. **33**, 1601–1612 (1985)
D. François, A. Pineau, A. Zaoui, Comportement mécanique des matériaux: viscoplasticité, endommagement. mécanique de la rupture et mécanique du contact (1995)
D. François, Endommagement et rupture des matériaux (2004)
M. Frewer, More clarity on the concept of material frame-indifference in classical continuum mechanics. Acta Mech. **202**, 213–246 (2009)
T. Furu, H.R. Shercliff, G.J. Baxter, C.M. Sellars, Influence of transient deformation conditions on recrystallization during thermomechanical processing of an Al-1% Mg alloy. Acta Mater. **47**(8), 2377–2389 (1999)
W. Gambin, *Plasticity and Textures* (Kluwer, Dordrecht, 2001)

G. Gottstein, V. Marx, R. Sebald, Integration of physically-based models into FEM and application in simulation of metal forming processes. Model. Simul. Mater. Sci. Eng. **8**(6), 881–891 (2000)
M. Grugicic, S. Batchu, Crystal plasticity analysis of earing in deep-drawn OFHC copper cups. J. Mater. Sci. **37**, 753–764 (2002)
A.M. Habraken, L. Duchene, Anisotropic elasto-plastic finite element analysis using a stress-strain interpolation method based on a polycrystalline model. Int. J. Plasticity **20**(8–9), 1525–1560 (2004)
K.S. Havner, *Finite Plastic Deformation of Crystalline Solids* (Cambridge University Press, Cambridge, 1992)
N. Hfaiedh, K. Saanouni, M. Francois, A. Roos, Self-consistent intragranular ductile damage modeling in large plasticity for FCC polycrystalline materials. Proc. Eng. **1**, 229–232 (2009)
N. Hfaiedh, Modélisation micromécanique des polycristaux – couplage plasticité, texture et endommagement, Ph.D. thesis, Université de Technologie de Troyes, 2009
D. Hull, D.J. Bacon, *Introduction to Dislocations* (Butterworth–Heinemann, Oxford, 1995)
K. Inal, R.K. Mishra, O. Cazacu, Forming simulation of aluminum sheets using an anisotropic yield function coupled with crystal plasticity theory. Int. J. Solids Struct. **47**, 2223–2233 (2010)
S.R. Kalidindi, C.A. Bronkhorst, L. Anand, Crystallographic texture evolution in bulk deformation processing of FCC metals. J. Mech. Phys. Solids **40**, 537–569 (1992)
C. Keller, E. Hug, R. Retoux, X. Feaugas, TEM study of dislocation patterns in near-surface and core regions of deformed nickel polycrystals with few grains across the cross section. Mech. Mater. **42**, 44–54 (2010)
U.F. Kocks, C.N. Tomé, H.R. Wenk, *Texture and Anisotropy: Preferred Orientations in Polycrystals and Their Effect on Material Properties* (Cambridge University Press, Cambridge, 1998)
L. Le Joncour, B. Panicaud, A. Baczmański, M. François, C. Braham, A. Paradowska, S. Wroński, R. Chiron, Duplex steel studied at large deformation until damage at mesoscopic and macroscopic scales. Mech. Mater. **42**, 1048–1058 (2010)
L. Le Joncour, Analyses expérimentales et modélisation multi-échelles de l'endommagement d'un acier UR45N laminé vieilli, Ph.D. thesis, Université de Technologie de Troyes, 2011
J. Lemaitre, J.-L. Chaboche, *Mécanique Des Matériaux Solides* (Cambridge University Press, Cambridge, 2001)
P. Lestriez, Modélisation numérique du couplage thermo-mécanique endommagement en transformations finies. Application à la mise en forme, Ph.D. thesis, Université de Technologie de Troyes, 2003
P. Lipinski, M. Berveiller Int. J. Plasticity, 5, 149–172 (1989)
F. Louchet, Plasticité des métaux de structure cubique centrée. Dislocations et déformation plastique (1979)
S. Mahajan, Interrelationship between slip and twinning in bcc crystals. Acta Metall. **23**, 671–684 (1975)
E. Maire, C. Bordreuil, J.-C. Boyer, L. Babouta, Damage initiation and growth in metals. Comparison between modeling and tomography experiments. J. Mech. Phys. Solids **53**, 2411–2434 (2005)
J.-F. Mariage, Simulation numérique de l'endommagement ductile en formage de pièces massives, Ph.D. thesis, Université de Technologie de Troyes, 2003.
A. Mateo, L. Lianes, M. Anglade, A. Redjaimia, G. Metauer, Characterization of the intermetallic G-phase in an AISI 329 duplex stainless steel. J. Mater. Sci. **32**(12), 4533–4540 (1997)
L. Mcirdi, Comportement et endommagement sous sollicitation mécanique d'un acier austéno-ferritique moulé vieilli, Ph.D. thesis, Ecole Nationale Supérieur des Arts et Métier de Paris, 2000
C. Miehe, J. Schotte, Anisotropic finite elastoplastic analysis of shells: Simulation of earing in deep-drawing of single- and polycrystalline sheets by Taylor-type micro-to-macro transitions. Comput. Meth. Appl. Mech. Eng. **193**(1–2), 25–57 (2004)
F. Montheillet, F. Moussy, *Physique et Mécanique de L'endommagement* (les Ed. De Physique, Les Ulis, 1988)
T. Mura, *Micromechanics of Defects in Solids* (Martinus Nijhoff Publishers, Dordrecht, 1987)

R.J. Nedoushan, M. Farzin, M. Mashayekh, D. Banabic, A microstructure-based constitutive model for superplastic forming. Metal. Mater. Trans. A **43**, 4266–4280 (2012)
C.J. Neil, J.A. Wollmershauser, B. Clausen, C.N. Tomé, R. Agnew, Modeling lattice strain evolution at finite strains and experimental verification for copper and stainless steel using in situ neutron diffraction. Int. J. Plasticity **26**(12), 1772–1791 (2010)
S. Nemat-Nasser, M. Hori, *Micromechanics: Overall Properties of Heterogeneous Materials* (Elsevier, Amsterdam, 1993)
S. Nemat-Nasser, *Plasticity. A Treatise on Finite Deformation of Heterogeneous Inelastic Materials* (Cambridge University Press, Cambridge, 2004)
B. Panicaud, E. Rouhaud, A frame-indifferent model for a thermo-elastic material beyond the three-dimensional eulerian and lagrangian descriptions, Cont. Mech. Thermodyn. (2013, in press)
A. Paquin, S. Berbenni, V. Favier, X. Lemoine, M. Berveiller, Micromechanical modeling of the elastic-viscoplastic behavior of polycrystalline steels. Int. J. Plasticity **17**, 1267–1302 (2001)
V.K. Pecharsky, P.Y. Zavalij, *Fundamentals of Powder Diffraction and Structural Characterization of Materials* (Springer, New York, 2005)
P. Pilvin, The contribution of micromechanical approaches to the modeling of inelastic behavior of polycrystals. Soc. Fr. Métall. Matér. **1**, 31–45 (1994)
D. Raabe, *Computational Material Science: The Simulation of Materials Microstructures and Properties* (Wiley-VCH, Weinheim, 1998)
D. Raabe, F. Roters, Using texture components in crystal plasticity finite element simulations. Int. J. Plasticity **20**(3), 339–361 (2004)
E. Rouhaud, B. Panicaud, R. Kerner, Canonical frame-indifferent transport operators with the four-dimensional formalism of differential geometry, Comput. Mater. Sci. (2013, in press)
K. Saanouni, *Damage Mechanics in Metal Forming. Advanced Modeling and Numerical Simulation* (ISTE John Wiley, London, 2012). ISBN 978-1-8482-1348-7
J. R. Santisteban, L. Edwards, A. Steuwer, P. J. Withers, Time-of-flight neutron transmission diffraction, J. Appl. Crystallogr. (2001). ISSN 0021-8898. http://www.isis.stfc.ac.uk/instruments/engin-x/documents/engin-x-a-third-generation-neutron-strain-scanner10390.pdf
E. Schmid, W. Boas, *Plasticity of Crystals* (Chapman and Hall, London, 1968)
C.M. Sellard, Modeling microsctructural development during hot rolling. Mater. Sci. Technol. **6**, 1072–1081 (1990)
H.R. Shercliff, A.M. Lovatt, Modeling of microstructure evolution in hot deformation. Philos. Trans. R. Soc. Lond. **357**, 1621–1643 (1999)
C. Teodosiu, F. Sidoro, Theory of finite elastoviscoplasticity of single crystals. Int. J. Eng. Sci. **14**, 165–176 (1976)
C. Truesdell, W. Noll, *The Non-Linear Field Theories of Mechanics*, 3rd edn. (Springer, New York, 2003)
V. Vitek, The core structure of 1/2[1 1 1] screw dislocations in bcc crystals. Philos. Mag. **21**, 1049–1073 (1970)
D.S. Wilkinson, X. Duan, J. Kang, M. Jain, J.D. Embury, Modeling the role of microstructure on shear instability with reference to the formability of aluminum alloys, Mater. Sci. Forum 519/521, 183–190 (2006)
S. Wroński, A. Baczmański, R. Dakhlaoui, C. Braham, K. Wierzbanowski, E.C. Oliver, Determination of stress field in textured duplex steel using TOF neutron diffraction method. Acta Mater. **55**, 6219–6233 (2007)
W. Yang, W.B. Lee, *Mesoplasticity and Its Applications* (Springer, Berlin, 1993)
Zebulon, *Zset/Zebulon: Developer Manual* (2008)
Q. Zhu, C.M. Sellars, Microstructural evolution of Al-Mg alloys during thermomechanical processing. Mater. Sci. Forum **331**(1), 409–420 (2000)

Section VII

Damage Micromechanics of Composite Materials, and Coupled Elastoplastic Damage-Healing Mechanics of Granular Materials

Fiber Cracking and Elastoplastic Damage Behavior of Fiber Reinforced Metal Matrix Composites

29

Yu-Fu Ko and Jiann-Wen Woody Ju

Contents

Abstract

To predict the overall transverse mechanical behavior and damage evolutions of *cylindrical* fiber-reinforced ductile composites with fiber cracking, an innovative micromechanical multi-level elastoplastic evolutionary damage framework is proposed. Progressively cracked fibers are modeled by using the *double-inclusion theory*. The effective elastic moduli of 3-phase composites,

Y.-F. Ko (✉)
Department of Civil Engineering and Construction Engineering Management, California State University, Long Beach, CA, USA
e-mail: Yu-Fu.Ko@csulb.edu

J.-W.W. Ju
Department of Civil and Environmental Engineering, University of California, Los Angeles CA, USA
e-mail: juj@ucla.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_12

consisting of a matrix, *randomly* located yet monotonically aligned cylindrical uncracked fibers and cracked fibers, are derived by using a micromechanical formulation. A micromechanical effective yield criterion is derived based on the ensemble-area averaging process and the first-order effects of eigenstrains to characterize the homogenized elastoplastic behavior. The resulting effective yield criterion, together with the overall associative plastic flow rule and the hardening law, constitutes the analytical framework for the estimation of effective transverse elastoplastic damage responses of ductile composites containing both uncracked and cracked fibers. An evolutionary fiber cracking process, governed by the internal stresses and the fracture strength of fibers, is incorporated into the proposed work. The Weibull's probabilistic distribution is employed to describe the varying probability of fiber cracking. Further, systematic numerical simulations are presented to illustrate the potential of the proposed methodology.

Introduction

Composite materials consist of two or more phases combined together on a macroscopic scale to form specific new materials with certain desirable material properties and enhanced performance. The "inclusions" in composites can be of various forms such as fibers, whiskers, and particulates, which can be made of alumina, silicon carbide, silicon nitride, boron, and graphite, etc. The other primary phase in composites is called the "matrix." The matrix materials usually serve as the binder materials not only to support and protect the inclusions but also to transfer stresses between perfectly bonded and partially debonded/broken inclusions under three-dimensional complex loadings. The matrix materials can be made of polymers, metals, ceramics, or carbons, etc. In order to enhance the material performance and meet the demands from the aerospace (commercial or military aircraft), automobile industries, civil infrastructures, and sporting goods, etc., extensive research and development of improved high-performance structural materials have been conducted over the past several decades around the world. Composite materials can significantly improve such properties of a material as strength, stiffness, corrosion resistance, wear resistance, attractiveness, weight, fatigue life, temperature-dependent behavior, thermal insulation, thermal conductivity, and acoustical insulation, etc. In general, not all of these properties can be improved at the same time. The primary objective is to create a new heterogeneous material that has certain desired characteristics needed to perform a particular engineering mission or task.

There are four commonly acknowledged types of composite materials. These include (1) fibrous composite materials that consist of fibers in a matrix; (2) laminated composite materials that consist of layers of various materials; (3) particulate composite materials that are composed of particles in a matrix; and (4) combinations of some or all of the first three types. Composite materials can also be classified

according to their matrix phase. There are, for example, polymer matrix composites (PMCs), ceramic matrix composites (CMCs), and metal matrix composites (MMCs). Materials within these categories are often called advanced composite materials. Advanced materials combine the properties of high strength, high stiffness, low weight, corrosion resistance, high temperature resistance, and in some cases special electrical properties. This combination of properties makes advanced composite materials very attractive in the applications of next generation aircraft and aerospace structural parts. Advanced composite materials are originally developed primarily for the applications of aerospace and defense industry. To date, they are also widely adopted in civil engineering applications. For example, carbon fiber and fiberglass have been used as prestressing materials, particularly for areas where there is concern about corrosion and stress embrittlement of conventional prestressed steel. A considerable body of research has been conducted to establish the effectiveness of column retrofit, using jackets of composite materials such as fiberglass and carbon fiber, generally bonded together and to the column with epoxy (Priestley et al. 1996).

In recent years, metal-matrix composites (MMCs) have emerged as a promising class of new materials with excellent ductility, formability, low density, high strength, and high thermal conductivity. Metal-matrix composites, in general, consist of at least two components. One is the metal matrix and the other is the reinforcement. The matrix is defined as metal in all cases, but pure metal is rarely used as the matrix. Instead, an alloy is generally used as the matrix material. In the production of the composites, the matrix and the reinforcements are mixed together in a random fashion or a periodic array.

More recently, fiber-reinforced metal matrix composite materials that have high strength-to-weight and stiffness-to-weight ratios have become important in weight-sensitive applications such as buildings, bridges, aircraft, space vehicles, and sporting goods, etc. For example, continuous fiber-reinforced metal matrix composites (CFRMMCs) have been increasingly used in engineering due to their improved mechanical properties. CFRMMCs are composed of at least two phases. The matrix materials of CFRMMCs are made of ductile metals or alloys with high strain capability, such as aluminum, steel, or titanium. The inclusions of CFRMMCs can be made of carbon, boron, or glass fibers. The shape of inclusions could be circular or elliptical cylinder. These unidirectional aligned and impenetrable fibers disperse (usually randomly) in a matrix and behave in an elastic manner.

In recent decades, fiber-reinforced metal matrix composite materials that have high strength-to-weight and stiffness-to-weight ratios have gained importance in weight-sensitive applications such as buildings, bridges, aircraft, space vehicles, and sporting goods, etc. For example, continuous fiber-reinforced metal matrix composites (CFRMMCs) have been increasingly used in engineering due to their improved mechanical properties. CFRMMCs are composed of at least two phases. The matrix materials of CFRMMCs are made of ductile metals or alloys with high strain capability, such as aluminum, steel, or titanium. The inclusions of CFRMMCs can be made of alumina, silicon carbide, silicon nitride, carbon,

boron, or glass fibers. The shape of inclusions can be circular or elliptical cylinder. These unidirectional aligned and impenetrable fibers disperse (usually randomly) in a matrix and behave in an elastic manner.

In general, the filament packing array of CFRMMCs could be either a *periodic regular array* or a *randomly distributed one*. However, during the fabrication process, a periodic regular array of filament packing may be hard to achieve. A randomly distributed array of filament packing is often observed.

With the increasing applications of CFRMMCs, numerous efforts have been devoted to the theoretical prediction of mechanical and thermal responses of CFRMMCs. To facilitate the development of the state-of-the-art theoretical frameworks, it is important to understand the manufacturing process of MMCs. Many physical and chemical phenomena can be observed during the fabrication process of MMCs.

The manufacturing process of MMCs generally involves two stages: (1) the fabrication of the composite materials from base metal and fiber reinforcement, and (2) the subsequent fabrication of laminates from the composite materials.

The first stage involves five main processing techniques (Eluripati et al. 2003; Kaczmar et al. 2000): (i) Liquid-State Processing; (ii) Solid-State Processing; (iii) Physical Vapor Processing (PVD); (iv) Powder Metallurgy; and (v) Direct Processing/Spray Deposition. The second stage involves fabrication of laminates from the composite materials. Taking CFRMMCs as an example, the three principal lay-up processes are winding, laying, and molding. Sheet molding compound (SMC), roll-forming process, and pultrusion process are usually used to produce various forms and shapes of structural parts.

A short overview of Solid-State Processing method used to fabricate Titanium Matrix Composites (TMCs)/Aluminum Matrix Composites (AMCs) is discussed below.

The Solid-State Processing is further divided into three stages: (a) Pre-Processing of composites; (b) Primary Processing; and (c) Secondary Processing.

Pre-Processing, in general, includes the surface treatment of ingredient materials and preform fabrication. A preform is a shaped porous assembly of ingredient material elements such as fibers, whiskers or particles. This is followed by the stacking of alternate layers of matrix foils and the fiber preforms.

Primary-Processing combines ingredient materials (e.g., the matrix foils and fibers). Diffusion bonding is the main process used for manufacturing continuous fiber- reinforced metal matrix composites (CFRMMCs) during current process. A typical high-temperature consolidation process called HIPing (Hot Isostatic Pressing) utilizes diffusion bonding for manufacturing the composites. The stacks of matrix foils and fibers are encapsulated in an insulated chamber of a pressure vessel inside a furnace. A cooling jacket provided with the furnace is attached to a heat exchanger to control the temperature inside the furnace when required. A vacuum pump connected to the insulated chamber evacuates the chamber. Argon gas is used to apply pressure on the specimen. The required pressure levels are obtained by using a compressor, and the required amount of gas is provided from a liquid argon storage tank by means of a liquid pump and a vaporizer. Temperature, pressure, and

process time are controlled by a computer controller. When the temperature inside the chamber exceeds the matrix re-crystallization temperature, the molten matrix material infiltrates into the porous fiber preform with application of hydrostatic pressure by means of argon gas. However, it is not necessary to obtain a final shape or final microstructure by primary processing.

It is noted that during the consolidation stage, the composites are cooled down to the room temperature and the constituents of the composites contract independently according to their Coefficient of Thermal Expansion (CTE). Residual stresses are generated in the composites due to the CTE mismatch of the constituents. Therefore, the "as-processed" composite materials at room temperature are actually subjected to a state of stress without any external forces.

Secondary Processing is further divided into two steps. The first step includes shaping and forming, and the second step involves the final composition. As the names suggest, the first step aims to alter the shape and microstructure of the material with processes such as forging, rolling, heat treatment, etc. The second step involves machining and joining operations, which include grinding, drilling, welding, etc. The final composite is obtained at this stage.

After understanding the manufacturing process of CFRMMCs, we will proceed to discuss the factors that should be taken into account in the development of the theoretical frameworks to model the mechanical behavior of CFRMMCs.

As a result of manufacturing process and continuous external loading, internal defects as well as evolutionary defects within fiber-reinforced composites play a significant role in the damage and fracture mechanism. Some of the most common defects have been studied exclusively such as crystal defects and planar cracks. While other forms of defects need to be investigated comprehensively such as the fiber-matrix debonding, matrix-cracking, fiber breakage, fiber cracking, fiber pull-out, and shear-sliding between fibers and the matrix. However, due to different damage mechanisms of various types of defects and the tremendous difficulty in considering them simultaneously, it is wise and reasonable to deal with the predominant defect first. It should be noted that the dominant damage mechanisms is intimately related to the strength of reinforcement, the composite matrix and its interface, the reinforcement shape and concentration, and the loading mode. The present study is confined to the reinforcing *fiber cracking* mechanism.

In traditional continuum mechanics analysis, the study of composite material behavior is at the level of macromechanics. The material is presumed to be homogeneous, and the effects of the constituent materials are detected only as averaged apparent macroscopic properties. Macroscopic theories of materials would not give as accurate solutions as micromechanics theories do. Therefore, the theories based upon micromechanics must be developed.

Micromechanics is a method of analytically predicting basic composite material properties as functions of the properties of the constituent materials and their micro-geometrical and microstructural relationships among one another, as opposed to the analysis of members fabricated from such composites (i.e., "macromechanics").

Traditional continuum mechanics deals with idealized materials. It assumes that: (1) the elastic properties of a solid at a given point are the same in every direction

(isotropy); and (2) the material property is the same at all points within the solid (homogeneity). These two assumptions render the uniform stress and strain distribution within an infinitesimal material element. However, from the microscopic and realistic points of view, an infinitesimal material element has its own complex microstructure. Therefore, the stress and strain field within such a material element is not uniform at the micro-scale level. It is the major objective of micromechanics to characterize such infinitesimal material element and its neighborhood with a rigorous theoretical framework. Optical microscopy shows that the microscopic structure of a solid is complex. It consists of inclusions, grains separated by grain boundaries, microcavities, cracks, and dislocations, for example. The RVE for a material point of a continuum mass defines a material volume which statistically represents the infinitesimal material neighborhood of that material point (Nemat-Nasser and Hori 1993).

In essence, the concept of RVE in classical micromechanics is a mathematical paradigm. It has no fixed length scale associated with each level. The length scales associated with macro-level and micro-level are relative. If one studies the effective material properties of a heterogeneous metal, the length scale of the micro-level may extend from a few nm to μm, and the length scale of macro-level may be from a few mm to a centimeter. For example, in a study of the stiffness of a dam, the length scale if micro-level could be from centimeters, whereas the length scale at macro-level could be meters. In classical continuum mechanics, at macro-level, material properties are always assumed to be homogeneous but unknown, whereas at micro-level, i.e., inside the RVE, the material properties are heterogeneous but known. At micro-level, the heterogeneous micro-structure and the physical laws are known. Micromechanics used information of microstructure to discover homogeneous material properties at macro-level, which are often called overall material properties or effective material properties.

The methodology to estimate the effective material properties is named "homogenization." Here, the term "homogenization" means the statistical and volume averaging within an RVE. Mathematically, such procedure is related to the ensemble-volume averaging process, and leads to overall governing constitutive equations.

Emanating from the eigenstrain concept (Eshelby 1957) and the micromechanical framework (Ju and Chen 1994a, b, c; Ju and Tseng 1996, 1997; Ju and Sun 2001; Sun and Ju 2001; Ju and Zhang 1998, 2001; Lin and Ju 2009; Feng and Yu 2010; Ju and Yanase 2010, 2011; Ko and Ju 2012, 2013), the effective elastoplastic damage behavior of fiber/particle-reinforced composites due to interfacial complete/partial debonding (Ju and Lee 2000, 2001; Sun et al. 2003a, b; Liu et al. 2004, 2006; Ju et al. 2006, 2008, 2009; Paulino et al. 2006; Ju and Ko 2008; Lee and Ju 2007, 2008; Ju and Yanase 2009; Okabe et al. 2010, particle-cracking (Sun et al. 2003b), and thermal residual effect (Liu and Sun 2004; Ju and Yanase 2008; Zhang and Wang 2010) have been investigated extensively in the literature. In addition, a number of studies regarding the effects of *particle-cracking* in particle-reinforced metal matrix composites (PRMMCs) have been conducted

under the unit-cell micromechanical framework and the finite element analysis. Bao (1992) studied the influence of damage evolution paths on the brittle particle-reinforced metal matrix composites and the reduction in composite limit flow stress for aligned spherical and cylindrical particles in an elastic-perfectly plastic matrix subject to tensile loading normal to the plane of particle cracks. Brockenbrough and Zok (1995) examined the flow response of particle-reinforced metal matrix composites by employing unit cells containing either intact or cracked particles in a plastically hardening solid with a power law hardening to determine the corresponding asymptotic flow strengths. The effects of the hardening exponent and the elastic mismatch between the particles and the matrix on the flow response are also investigated. Llorca et al. (1997) introduced the Weibull statistics to simulate the damage evolution of particle-cracking with their finite element models. To consider composites with random reinforcement distributions, Ghosh and Moorthy (1998) developed a microstructure-based *Voronoi Cell Finite Element Model* (VCFEM) to analyze the particle-cracking and splitting effect. Moreover, Steglich et al. (1999) reported on an investigation of crack propagation in a metal matrix composite (MMC) where the dominant failure mechanism observed was particle cracking. The Gurson-Tvergaard-Needleman model (Tvergaard and Needleman 1984) was used to study crack propagation in the material; the particle cracking was directly taken into account by means of a cohesive zone model. Finally, Wang et al. (2008) studied the fracture of a fiber embedded in a matrix of finite radius under the axial extension and residual temperature change of the composite medium. The periodic array of cracks in the fiber along the central axis of the medium is assumed. Singular integral equation technique and fracture mechanics analysis were employed in their studies.

However, the studies on predictions of the overall elastoplastic-damage behavior of CFRMMCs under transverse tensile loadings due to *fiber cracking* have been rather limited to date. The main objective of this chapter is to micromechanically predict the elastoplastic-damage behavior of CFRMMCs under the transverse *uniaxial* tensile loading due to *fiber cracking* by considering the mechanical properties of constituent phases, fiber volume fractions, random spatial distributions of fibers, and critical fracture strengths of fibers, etc.

Eshelby's Micromechanical Theory

Based on micromechanics, when a material contains inhomogeneities of different material properties such as voids, cracks, or precipitates, etc., it is subjected to an internal stress (eigenstress) field even if it is free from external load. Such stress field is caused by the eigenstrain inside the inhomogeneities due to the misfit and phase transformation. Eshelby (1957) first pointed out that the stress perturbation in an applied stress due to the presence of an inhomogeneity can be simulated by an eigenstress caused by an inclusion when the eigenstrain is chosen properly.

Let us now consider an infinitely extended material domain D (elastic modulus $\mathbf{C}^{(0)}$) with an inclusion (inhomogeneity) domain Ω (elastic modulus $\mathbf{C}^{(1)}$). Suppose that the applied stress at infinity is $\boldsymbol{\sigma}^{o}(\mathbf{x})$ and the corresponding strain is $\boldsymbol{\varepsilon}^{o}(\mathbf{x})$. Furthermore, the perturbed stress field and perturbed strain field are denoted by $\boldsymbol{\sigma}'(\mathbf{x})$ and $\boldsymbol{\varepsilon}'(\mathbf{x})$, respectively. Therefore, the Hooke's law takes the form

$$\sigma_{ij}^{o}(\mathbf{x}) + \sigma'_{ij}(\mathbf{x}) = C_{ijkl}^{(1)}\left[\varepsilon_{kl}^{o}(\mathbf{x}) + \varepsilon'_{kl}(\mathbf{x})\right] \quad \text{in } \Omega \tag{1}$$

$$\sigma_{ij}^{o}(\mathbf{x}) + \sigma'_{ij}(\mathbf{x}) = C_{ijkl}^{(0)}\left[\varepsilon_{kl}^{o}(\mathbf{x}) + \varepsilon'_{kl}(\mathbf{x})\right] \quad \text{in } D - \Omega \tag{2}$$

Eshelby's equivalent eigenstrain (eigenstress) principle is a homogenization method. It establishes the equivalency between an eigenstrain (eigenstress) field and an inhomogeneity distribution, such that distribution of inhomogeneities may be replaced by the eigenstrain field with the equivalent mechanical effect. This equivalency mapping process translates the heterogeity of material into an added non-uniform strain distribution, while making the material properties become homogeneous again. In other words, the so-called *Eshelby's equivalent eigenstrain (eigenstress) principle* is to replace the inhomogeneity with a homogenized inclusion, with which an eigenstrain field is prescribed, such that the homogenized field is mechanically equivalent to the original inhomogeneous field.

By introducing an eigenstrain $\boldsymbol{\varepsilon}^{*}(\mathbf{x})$ within this inclusion domain Ω and applying the Eshelby's equivalence principle, the Hooke's law yields

$$\sigma_{ij}^{o}(\mathbf{x}) + \sigma'_{ij}(\mathbf{x}) = C_{ijkl}^{(0)}\left[\varepsilon_{kl}^{o}(\mathbf{x}) + \varepsilon'_{kl}(\mathbf{x}) - \varepsilon_{kl}^{*}(\mathbf{x})\right] \quad \text{in } \Omega \tag{3}$$

$$\sigma_{ij}^{o}(\mathbf{x}) + \sigma'_{ij}(\mathbf{x}) = C_{ijkl}^{(0)}\left[\varepsilon_{kl}^{o}(\mathbf{x}) + \varepsilon'_{kl}(\mathbf{x})\right] \quad \text{in } D - \Omega \tag{4}$$

Apparently, the necessary and sufficient condition for the equivalency of the stresses and strains in Eqs. 3 and 4 is

$$C_{ijkl}^{(1)}\left[\varepsilon_{kl}^{o}(\mathbf{x}) + \varepsilon'_{kl}(\mathbf{x})\right] = C_{ijkl}^{(0)}\left[\varepsilon_{kl}^{o}(\mathbf{x}) + \varepsilon'_{kl}(\mathbf{x}) - \varepsilon_{kl}^{*}(\mathbf{x})\right] \tag{5}$$

In the case of uniform stress $\boldsymbol{\sigma}^{o}(\mathbf{x})$ at the far field, the eigenstrain $\boldsymbol{\varepsilon}^{*}(\mathbf{x})$ is also uniform within inclusion domain Ω.

Upon the solution of the local stress/strain field, one typically performs an ensemble-averaging process (homogenization) within the aforementioned mesoscopic representative volume element (RVE) in order to obtain overall (effective) constitutive equations and properties of the heterogeneous composites. In present framework, to avoid the truncation errors of *Green's functions* outside the domain of an RVE, an RVE itself is embedded in an *infinite* and identical matrix material within present framework. The entire assembly is subjected to specified far-field stresses $\boldsymbol{\sigma}^{o}$ or strains $\boldsymbol{\varepsilon}^{o}$. Furthermore, all inclusions are assumed to be non-intersecting and impenetrable (Ju and Chen 1994a, b, c).

The volume-averaged stress tensor is defined as

$$\overline{\boldsymbol{\sigma}} \equiv \frac{1}{V}\int_{V} \overline{\boldsymbol{\sigma}}(\mathbf{x})\, d\mathbf{x} = \frac{1}{V}\left[\int_{V_0} \overline{\boldsymbol{\sigma}}(\mathbf{x})\, d\mathbf{x} + \sum_{r=1}^{n} \int_{V_r} \overline{\boldsymbol{\sigma}}(\mathbf{x})\, d\mathbf{x}\right] \tag{6}$$

Similarly, the volume-averaged strain tensor is defined as

$$\overline{\boldsymbol{\varepsilon}} \equiv \frac{1}{V}\int_{V} \overline{\boldsymbol{\varepsilon}}(\mathbf{x})\, d\mathbf{x} = \frac{1}{V}\left[\int_{V_0} \overline{\boldsymbol{\varepsilon}}(\mathbf{x})\, d\mathbf{x} + \sum_{r=1}^{n} \int_{V_r} \overline{\boldsymbol{\varepsilon}}(\mathbf{x})\, d\mathbf{x}\right] \tag{7}$$

where V is the volume of an RVE, V_0 is the volume of the matrix, V_r is the volume of the *r*th-phase inhomogeneities, and n denotes the number of inclusion phases of different material properties excluding the matrix.

Based upon Eshelby's equivalence principle and the ensemble-averaging method, effective constitutive relations as well as elastoplastic-damage behavior of fiber- reinforced ductile metal matrix composites are derived in the following secgtions along with the accurate formulation of the eigenstrains.

Derivation of Two-Dimensional Interior-Point Eshelby Tensor S

Suppose that the uniform eigenstrain $\boldsymbol{\varepsilon}^*(\mathbf{x})$ is prescribed in the inhomogeneous inclusion domain Ω and, at the same time, the entire medium is loaded by the far-field uniform applied stress $\boldsymbol{\sigma}^o$. Eshelby used a fourth-rank tensor **S** which is traditionally termed the *interior-point Eshelby's tensor* to describe the strain and stress fields in the inclusion domain.

The two-dimensional interior-point Eshelby tensor **S** is defined as

$$\mathbf{S} \equiv \int_{\Omega_i} \mathbf{G}(\mathbf{x}-\mathbf{x}')d\mathbf{x}'; \quad \mathbf{x}, \quad \mathbf{x}' \in \Omega_i \tag{8}$$

The components of **S** depend upon Poisson's ratio of the matrix (ν_0) and the shape of the fiber cross-sectional domain Ω_i.

Derivations: according to the Eshelby's solution, the elastic displacement field due to inclusion in an isotropic infinite body reads

$$u_i(\mathbf{x}) = -C_{jkmn}\varepsilon^*_{mn}\int_{\Omega} G_{ij,k}\left(\mathbf{x}-\mathbf{x}'\right)d\mathbf{x}' \tag{9}$$

where the second rank plane-strain Green's function is given by Mura (1987):

$$G_{ij}(\mathbf{x}-\mathbf{x}') = \frac{1}{8\pi(1-\nu_0)\mu_0}\left[\frac{\left(x_i-x_i'\right)\left(x_j-x_j'\right)}{\|\mathbf{x}-\mathbf{x}'\|^2} - (3-4\nu_0)\delta_{ij}\ln\|\mathbf{x}-\mathbf{x}'\|\right] \tag{10}$$

By taking the derivative of $G_{ij}(\mathbf{x}-\mathbf{x}')$ in Eq. 10 with respect to x_k and substituting the result into Eq. 9, we arrive at

$$u_i(\mathbf{x}) = -\frac{\varepsilon^*_{jk}}{4\pi(1-\nu_0)}\int_{\Omega} g_{ijk}(\mathbf{l})\frac{d\mathbf{x}'}{\|\mathbf{x}-\mathbf{x}'\|} \tag{11}$$

where

$$g_{ijk}(\mathbf{l}) = (1-2\nu_0)\left(\delta_{ij}l_k + \delta_{ik}l_j - \delta_{jk}l_i\right) + 2l_il_jl_k \tag{12}$$

and

$$\mathbf{l} \equiv \frac{(\mathbf{x}'-\mathbf{x})}{\|\mathbf{x}'-\mathbf{x}\|} \tag{13}$$

When the point $\mathbf{x}$ is located inside the inclusion, the strain and stress fields become uniform for the interior points. Moreover, Eq. 11 can be integrated explicitly. The differential element $d\mathbf{x}'$ can be written as

$$d\mathbf{x}' = rdrd\theta \tag{14}$$

where $r = \|\mathbf{x}' - \mathbf{x}\|$ and $d\theta$ is the differential angle element centered at point $\mathbf{x}(x_1, x_2)$. Upon integration with respect to r, Eq. 11 becomes

$$u_i(\mathbf{x}) = -\frac{\varepsilon^*_{jk}}{4\pi(1-\nu_0)}\int_0^{2\pi} r(\mathbf{l})g_{ijk}(\mathbf{l})d\theta \tag{15}$$

Here, $r(\mathbf{l})$ is the positive root of the following equation

$$(x_1 + rl_1)^2 + (x_2 + rl_2)^2 = a^2 \tag{16}$$

where a is the radius of a single fiber. Therefore, we have

$$r(\mathbf{l}) = -\frac{f}{h} + \sqrt{\frac{f^2}{h^2} + \frac{e}{h}} \tag{17}$$

where

$$h = \frac{l_1^2 + l_2^2}{a^2}, \quad f = \frac{l_1x_1 + l_2x_2}{a^2}, \quad e = 1 - \frac{x_1^2 + x_2^2}{a^2} \tag{18}$$

Substituting Eq. 17 into Eq. 15, we find that the integration involving the term of $\sqrt{f^2/h^2 + e/h}$ vanishes because it is even in $\mathbf{l}$ while g_{ijk} is odd in $\mathbf{l}$. Consequently, we obtain

$$u_i(\mathbf{x}) = \frac{x_j\varepsilon^*_{mn}}{4\pi(1-\nu_0)}\int_0^{2\pi} \frac{\lambda_j g_{imn}(\mathbf{l})}{h}d\theta \tag{19}$$

and

$$\varepsilon_{ij}(\mathbf{x}) = \frac{\varepsilon^*_{kl}}{8\pi(1-\nu_0)}\int_0^{2\pi} \frac{\lambda_i g_{jkl} + \lambda_j g_{ikl}}{h}d\theta \tag{20}$$

Where

$$\lambda_i \equiv \frac{l_i}{a^2} \tag{21}$$

According to the definition of Eshelby's tensor $\varepsilon_{ij} = S_{ijkl}\varepsilon^*_{kl}$, we then arrive at

$$S_{ijkl} = \frac{1}{8\pi(1-\nu_0)} \int_0^{2\pi} \frac{\lambda_i g_{jkl} + \lambda_j g_{ikl}}{h} d\theta \tag{22}$$

Finally, the plane-strain Eshelby's tensor for a circular inclusion can be expressed as

$$S_{ijkl} = \frac{1}{8(1-\nu_0)} \left[(4\nu_0 - 1)\delta_{ij}\delta_{kl} + (3 - 4\nu_0)\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) \right] \tag{23}$$

where ν_0 is the Poisson's ratio for the matrix material.

Damage Theory of Composites

Unidirectionally aligned yet randomly distributed circular fiber-reinforced ductile matrix composites are considered. The initially perfectly bonded two-phase composites consists of an elastic matrix (phase 0) and unidirectionally aligned yet *randomly located* cylindrical fibers (phase 1); see Fig. 1. Under the applied external transverse tensile loading, some cylindrical fibers may initiate fiber cracking (phase 2) once the *local principal stress* within the fibers reaches certain critical fracture strength; cf. Fig. 2.

Since our major objective is to model the *overall* mechanical properties of composites through the local accurate stress/strain fields, simplification is made that those imperfect (cracked) isotropic fibers can be modeled by the *double-*

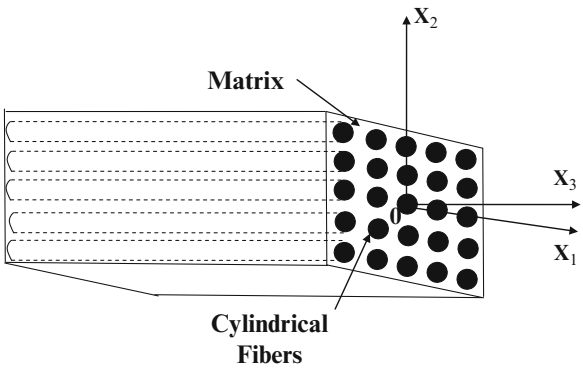


Fig. 1 The schematic drawing of unidirectionally aligned yet randomly distributed circular fiber-reinforced composites

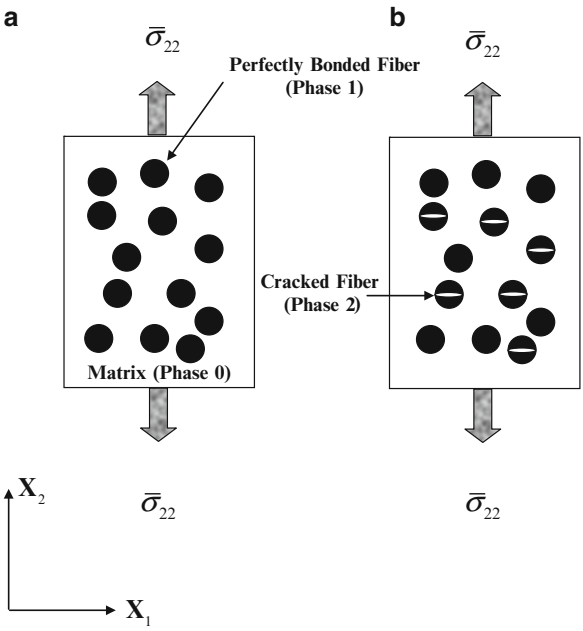


Fig. 2 A schematic diagram of a fiber composite subjected to transverse uniaxial tensile loading along the X_2-direction only: (**a**) the initial state (undamaged); (**b**) the damaged state

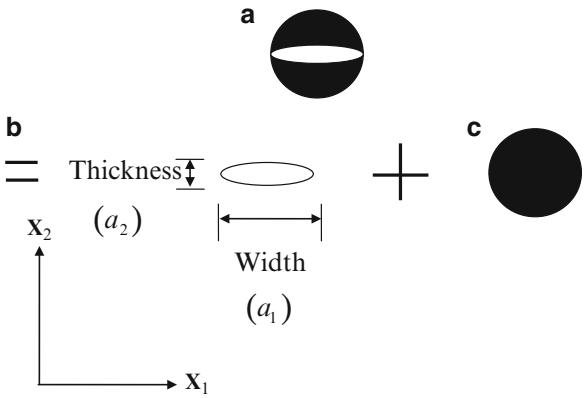


Fig. 3 The double-inclusion theory: (**a**) a cracked fiber; (**b**) an elliptical void (crack); (**c**) a perfectly bonded isotropic cylindrical fiber

inclusion theory (Hori and Nemat-Nasser (1993); Shodja and Sarvestani (2001); see Fig. 3. The cracked fiber (phase 2) is modeled by an elliptical void with the aspect ratio $\alpha = \frac{a_1}{a_2} \to \infty$ or $\rho = \frac{a_2}{a_1} \to 0$, embedded in a perfectly bonded isotropic cylindrical fiber.

The equivalent eigenstrain in a perfectly bonded isotropic cylindrical fiber is defined as

$$\boldsymbol{\varepsilon}_*^{(1)} = -\mathbf{K}^{-1} : \boldsymbol{\varepsilon}^{\mathrm{o}} = -\left[\mathbf{S}^{(1)} + \left(\mathbf{C}^{(1)} - \mathbf{C}^{(0)}\right)^{-1} \bullet \mathbf{C}^{(0)}\right]^{-1} : \boldsymbol{\varepsilon}^{\mathrm{o}} \tag{24}$$

The equivalent eigenstrain in an elliptical void can be derived as

$$\boldsymbol{\varepsilon}_*^{(2)} = -\mathbf{J}^{*-1} : \boldsymbol{\varepsilon}^{\mathrm{o}} \tag{25}$$

in which,

$$\mathbf{J}^* \equiv \left\{\mathbf{I} + \left(\mathbf{S}^{(1)} - \mathbf{S}^{(2)}\right) \bullet \left[\mathbf{S}^{(2)} + \left(\mathbf{C}^{(1)} - \mathbf{C}^{(0)}\right)^{-1} \bullet \mathbf{C}^{(0)}\right]^{-1}\right\} \bullet \mathbf{E}^* \tag{26}$$

with $\mathbf{E}^* \equiv (\mathbf{S}^{(2)} - \mathbf{I})$. Here, $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ are the interior-point Eshelby tensors for the cylindrical fiber and the elliptical void, respectively. In addition, $\mathbf{C}^{(0)}$ and $\mathbf{C}^{(1)}$ are the matrix and the fiber elasticity tensors, respectively. Therefore, the overall (averaged) eigenstrain in a cracked cylindrical fiber can be derived as

$$\langle \boldsymbol{\varepsilon}^* \rangle_{(2)} = f\, \boldsymbol{\varepsilon}_*^{(2)} + (1-f)\boldsymbol{\varepsilon}_*^{(1)} = \rho\left[-\mathbf{J}^{*-1} + \mathbf{K}^{-1}\right] : \boldsymbol{\varepsilon}^{\mathrm{o}} - \mathbf{K}^{-1} : \boldsymbol{\varepsilon}^{\mathrm{o}} \tag{27}$$

where

$$f = \frac{\text{Area of an elliptical} - \text{shaped void}}{\text{Area of a cylindrical fiber}} = \left(\frac{\pi(a_1)(a_2 = a_1/\alpha)}{\pi(a_1)^2}\right)_{\alpha\to\infty}$$
$$= \left(\frac{1}{\alpha}\right)_{\alpha\to\infty} = (\rho)_{\rho\to 0} \tag{28}$$

and the $\langle\rangle_{(2)}$ denotes the overall eigenstrain in a cracked isotropic cylindrical fiber (phase 2).

Effective Elastic-Damage Moduli of Composites

When the small strain is considered, the total macroscopic strain $\overline{\boldsymbol{\varepsilon}}$ consists of two parts:

$$\overline{\boldsymbol{\varepsilon}} = \overline{\boldsymbol{\varepsilon}}^e + \overline{\boldsymbol{\varepsilon}}^p \tag{29}$$

where $\overline{\boldsymbol{\varepsilon}}^e$ and $\overline{\boldsymbol{\varepsilon}}^p$ denote the overall elastic and plastic strains of composites, respectively. The relationship between the macroscopic stress $\overline{\boldsymbol{\sigma}}$ and macroscopic elastic strain $\overline{\boldsymbol{\varepsilon}}^e$ reads

$$\overline{\boldsymbol{\sigma}} = \overline{\mathbf{C}} : \overline{\boldsymbol{\varepsilon}}^e \tag{30}$$

After lengthy derivations, the effective elastic stiffness of composite can be determined as (cf. Ju and Chen (1994a, b); Ko and Ju 2012, 2013)

$$\begin{aligned}\overline{\mathbf{C}} = \mathbf{C}^{(0)} + \phi\mathbf{C}^{(0)} \bullet \left[\mathbf{K} - \phi\mathbf{S}^{(1)} - \phi^{(2)}\mathbf{S}^{(1)} \bullet \mathbf{J}^{-1} \bullet \mathbf{K}\right]^{-1} \\ + \phi^{(2)}\mathbf{C}^{(0)} \bullet \left[-\phi^{(2)}\mathbf{S}^{(1)} + \left(\mathbf{I} - \phi\mathbf{S}^{(1)} \bullet \mathbf{K}^{-1}\right) \bullet \mathbf{J}\right]^{-1}\end{aligned} \tag{31}$$

where

$$\mathbf{K} = \mathbf{S}^{(1)} + \left(\mathbf{C}^{(1)} - \mathbf{C}^{(0)}\right)^{-1} \bullet \mathbf{C}^{(0)} \tag{32}$$

$$\mathbf{J} = \frac{\mathbf{J}^*}{\rho} \equiv \left\{\mathbf{I} + \left(\mathbf{S}^{(1)} - \mathbf{S}^{(2)}\right) \bullet \left[\mathbf{S}^{(2)} + \left(\mathbf{C}^{(1)} - \mathbf{C}^{(0)}\right)^{-1} \bullet \mathbf{C}^{(0)}\right]^{-1}\right\} \bullet \mathbf{E} \tag{33}$$

Furthermore, in component form,

$$\mathbf{J} = \mathrm{J}_{\mathrm{IK}}^{(1)}\delta_{ij}\delta_{kl} + \mathrm{J}_{\mathrm{IJ}}^{(2)}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right), \quad i,j,k,l = 1,2 \tag{34}$$

with $\mathbf{E} = \frac{\mathbf{E}^*}{\rho} = \lim_{\rho\to 0}\left(\mathbf{S}^{(2)} - \mathbf{I}\right)/\rho$. Here, $\mathbf{I}$ is the fourth-order identity tensor, "•" denotes the tensor multiplication, ϕ denotes the total volume fraction of fibers in the composites, and $\phi^{(2)}$ signifies the volume fraction of the cracked fibers.

The components of the fourth-rank *interior-point* Eshelby tensor $\mathbf{S}^{(1)}$ for a cylindrical fiber and $\mathbf{S}^{(2)}$ for an elliptical void are given by Mura (1987), Ju and Sun (2001), Sun and Ju (2001), and Ju et al. (2009) by setting $\alpha = 1$ and $\alpha \to \infty$, respectively. Here, $\alpha = a_1/a_2$ is defined as the aspect ratio of the elliptical void (cf. Fig. 3).

For the overall effective *orthotropic* elastic materials,

$$\{\overline{\sigma}_i\}_{6\mathrm{x}1} = \left[\overline{C}_{ij}\right]_{6\mathrm{x}6}\left\{\overline{\varepsilon}_j^e\right\}_{6\mathrm{x}1}, \quad i = 1 \text{ to } 6, \quad j = 1 \text{ to } 6 \tag{35}$$

where

$$\{\overline{\sigma}_i\} = \{\overline{\sigma}_{11}, \overline{\sigma}_{22}, \overline{\sigma}_{33}, \overline{\sigma}_{12}, \overline{\sigma}_{23}, \overline{\sigma}_{31}\}^{\mathbf{T}}, i = 1 \text{ to } 6 \tag{36}$$

$$\left\{\overline{\varepsilon}_j^e\right\} = \{\overline{\varepsilon}_{11}^e, \overline{\varepsilon}_{22}^e, \overline{\varepsilon}_{33}^e, 2\overline{\varepsilon}_{12}^e, 2\overline{\varepsilon}_{23}^e, 2\overline{\varepsilon}_{31}^e\}^{\mathbf{T}}, j = 1 \text{ to } 6 \tag{37}$$

Therefore, there are only *nine* independent elastic constants for the overall *orthotropic* composite. Since the composite consists of a ductile (elastoplastic) matrix and *randomly* located yet unidirectionally aligned cylindrical fibers, the plane-strain condition governs here. Furthermore,

$$\overline{C}_{ijkl} = \mathrm{C}_{\mathrm{IK}}^{(1)}\delta_{ij}\delta_{kl} + \mathrm{C}_{\mathrm{IJ}}^{(2)}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right), i,j,k,l = 1,2,3 \tag{38}$$

Moreover, for the overall *orthotropic* materials,

$$\left[\overline{C}_{ij}\right]_{6\times6} = \begin{bmatrix} \overline{C}_{11} & \overline{C}_{12} & \overline{C}_{13} & 0 & 0 & 0 \\ \overline{C}_{12} & \overline{C}_{22} & \overline{C}_{23} & 0 & 0 & 0 \\ \overline{C}_{13} & \overline{C}_{23} & \overline{C}_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & \overline{C}_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \overline{C}_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & \overline{C}_{66} \end{bmatrix}_{6\times6} \tag{39}$$

In other words,

$$\left[\overline{C}_{ij}\right]_{6\times6} = \begin{bmatrix} \mathrm{C}_{11}^{(1)} + 2\mathrm{C}_{11}^{(2)} & \mathrm{C}_{12}^{(1)} & \mathrm{C}_{13}^{(1)} & 0 & 0 & 0 \\ \mathrm{C}_{21}^{(1)} & \mathrm{C}_{22}^{(1)} + 2\mathrm{C}_{22}^{(2)} & \mathrm{C}_{23}^{(1)} & 0 & 0 & 0 \\ \mathrm{C}_{31}^{(1)} & \mathrm{C}_{32}^{(1)} & \mathrm{C}_{33}^{(1)} + 2\mathrm{C}_{33}^{(2)} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{C}_{12}^{(2)} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathrm{C}_{23}^{(2)} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathrm{C}_{31}^{(2)} \end{bmatrix}_{6\times6} \tag{40}$$

Overall effective elastic moduli associated with such orthotropic materials take the form:

$$\overline{E}_{11} = \frac{\overline{C}_{33}\overline{C}_{12}^2 - 2\overline{C}_{13}\overline{C}_{23}\overline{C}_{12} + \overline{C}_{13}^2\overline{C}_{22} + \overline{C}_{11}\left(\overline{C}_{23}^2 - \overline{C}_{22}\overline{C}_{33}\right)}{\overline{C}_{23}^2 - \overline{C}_{22}\overline{C}_{33}} \tag{41}$$

$$\overline{E}_{22} = \frac{\overline{C}_{33}\overline{C}_{12}^2 - 2\overline{C}_{13}\overline{C}_{23}\overline{C}_{12} + \overline{C}_{13}^2\overline{C}_{22} + \overline{C}_{11}\left(\overline{C}_{23}^2 - \overline{C}_{22}\overline{C}_{33}\right)}{\overline{C}_{13}^2 - \overline{C}_{11}\overline{C}_{33}} \tag{42}$$

$$\overline{E}_{33} = \frac{\overline{C}_{22}\overline{C}_{13}^2 - 2\overline{C}_{12}\overline{C}_{23}\overline{C}_{13} + \overline{C}_{23}^2\overline{C}_{11}}{\overline{C}_{12}^2 - \overline{C}_{11}\overline{C}_{22}} + \overline{C}_{33} \tag{43}$$

$$\overline{\nu}_{12} = \frac{\overline{C}_{12}\overline{C}_{33} - \overline{C}_{13}\overline{C}_{23}}{\overline{C}_{22}\overline{C}_{33} - \overline{C}_{23}^2}, \overline{\nu}_{21} = \frac{\overline{C}_{12}\overline{C}_{33} - \overline{C}_{13}\overline{C}_{23}}{\overline{C}_{11}\overline{C}_{33} - \overline{C}_{13}^2}, \overline{\nu}_{13} = \frac{\overline{C}_{13}\overline{C}_{22} - \overline{C}_{12}\overline{C}_{23}}{\overline{C}_{22}\overline{C}_{33} - \overline{C}_{23}^2} \tag{44}$$

$$\overline{\nu}_{31} = \frac{\overline{C}_{13}\overline{C}_{22} - \overline{C}_{12}\overline{C}_{23}}{\overline{C}_{11}\overline{C}_{22} - \overline{C}_{12}^2}, \overline{\nu}_{23} = \frac{\overline{C}_{11}\overline{C}_{23} - \overline{C}_{12}\overline{C}_{13}}{\overline{C}_{11}\overline{C}_{33} - \overline{C}_{13}^2}, \overline{\nu}_{32} = \frac{\overline{C}_{11}\overline{C}_{23} - \overline{C}_{12}\overline{C}_{13}}{\overline{C}_{11}\overline{C}_{22} - \overline{C}_{12}^2} \tag{45}$$

$$\overline{\mu}_{12} = \overline{C}_{44}, \overline{\mu}_{23} = \overline{C}_{55}, \overline{\mu}_{31} = \overline{C}_{66} \tag{46}$$

$$\frac{\overline{\nu}_{12}}{\overline{E}_{11}} = \frac{\overline{\nu}_{21}}{\overline{E}_{22}}, \frac{\overline{\nu}_{23}}{\overline{E}_{22}} = \frac{\overline{\nu}_{32}}{\overline{E}_{33}}, \frac{\overline{\nu}_{13}}{\overline{E}_{11}} = \frac{\overline{\nu}_{31}}{\overline{E}_{33}} \tag{47}$$

Therefore, there are only *nine* independent elastic constants for the overall *orthotropic* composite materials. Here, effective Poisson's ratio $\overline{\nu}_{ij}$ is defined as the ratio of strain shrinkage in the j^{th} direction over the strain extension in the i^{th} direction when a tensile stress is applied in the i^{th} direction.

In addition, the effective elastic compliance matrix becomes

$$\left[\overline{D}_{ij}\right]_{6\times 6} = \left[\overline{C}_{ij}\right]^{-1}{}_{6\times 6} = \begin{bmatrix} \frac{1}{\overline{E}_{11}} & \frac{-\overline{\nu}_{21}}{\overline{E}_{22}} & \frac{-\overline{\nu}_{31}}{\overline{E}_{33}} & 0 & 0 & 0 \\ \frac{-\overline{\nu}_{12}}{\overline{E}_{11}} & \frac{1}{\overline{E}_{22}} & \frac{-\overline{\nu}_{32}}{\overline{E}_{33}} & 0 & 0 & 0 \\ \frac{-\overline{\nu}_{13}}{\overline{E}_{11}} & \frac{-\overline{\nu}_{23}}{\overline{E}_{22}} & \frac{1}{\overline{E}_{33}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\overline{\mu}_{12}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\overline{\mu}_{23}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\overline{\mu}_{31}} \end{bmatrix}_{6\times 6} \tag{48}$$

The stress component $\overline{\sigma}_{33}$ along the longitudinal fiber direction thus reads

$$\overline{\sigma}_{33} = \eta_1 \, \overline{\sigma}_{11} + \eta_2 \, \overline{\sigma}_{22} \tag{49}$$

where

$$\eta_1 = \frac{\overline{C}_{13} \, \overline{C}_{22} - \overline{C}_{12} \, \overline{C}_{23}}{\Xi}, \eta_2 = \frac{\overline{C}_{11} \, \overline{C}_{23} - \overline{C}_{12} \, \overline{C}_{13}}{\Xi}, \Xi = \overline{C}_{11} \, \overline{C}_{22} - \overline{C}_{12}^2 \tag{50}$$

To illustrate the influences of the volume fraction of cracked fibers upon the effective elastic moduli of two-phase composites under the external transverse uniaxial tensile loading along the X_2-direction, the material properties of boron fibers and the 2024 aluminum alloy matrix are considered. The elastic moduli of the 2024 aluminum alloy matrix and boron fibers, respectively, are: $E_0 = 8,100$ ksi, $\nu_0 = 0.32$, $E_1 = 55{,}000$ ksi, and $\nu_1 = 0.2$. It is noted that the proposed framework can predict the effective Young's moduli, effective shear moduli, and effective Poisson's ratios with varying volume fractions of cracked fibers; see section 5.2.2 of Ko 2005, pp. 263–266) for more details. For illustration purposes, only the effective Young's moduli with varying volume fractions of the cracked fibers are demonstrated here. In Fig. 4, in the undamaged state ($\phi^{(2)} = 0$), the composite is overall *transversely isotropic*. It is observed that the effective Young's moduli E_1^* and E_2^* are equal, while the effective Young's modulus E_3^* along the longitudinal direction of fibers is the largest. As $\phi^{(2)}$ increases, it is also noticed that both E_2^* and E_3^* decrease. However, E_2^* decreases drastically as $\phi^{(2)}$ increases. By contrast, E_1^* increases as $\phi^{(2)}$ increases due to the Poisson's ratio effect.

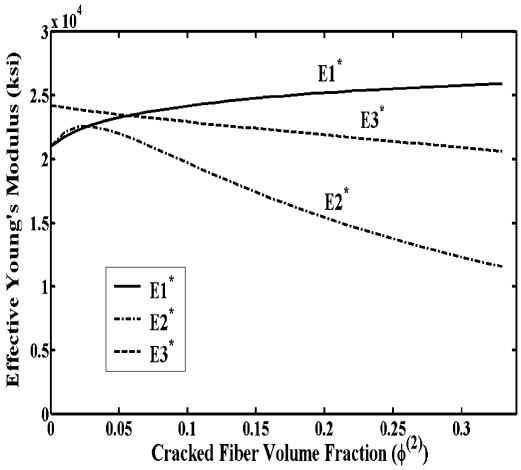


Fig. 4 The effective Young's modulus versus the cracked fiber volume fraction ($\phi^{(2)}$)

Elastoplastic-Damage Behavior of 3-Phase Composites

The multi-level overall elastoplastic-damage responses of 3-phase fiber-reinforced composites are now considered where the cylindrical fibers are elastic and the matrix is elastoplastic. Accordingly, at any matrix point, the stress $\boldsymbol{\sigma}$ and the equivalent plastic strain $\overline{e}^p$ must satisfy the following yield function:

$$F(\boldsymbol{\sigma}, \overline{e}^p) \;=\; \sqrt{H(\boldsymbol{\sigma})} - K(\overline{e}^p) \;\leq\; 0 \tag{51}$$

in which $K(\overline{e}^p)$ is the isotropic hardening function of the matrix-only material. Furthermore, $H(\boldsymbol{\sigma}) \equiv \boldsymbol{\sigma} : \mathbf{I}_{\mathrm{d}} : \boldsymbol{\sigma}$ denotes the square of the deviatoric stress form, where $\mathbf{I}_{\mathrm{d}}$ signifies the deviatoric part of the fourth-rank identity tensor $\mathbf{I}$. To obtain the effective (overall) elastoplastic-damage behaviors of the composites, the homogenization procedure is usually performed within a mesoscopic representative area element (RAE). Here, a framework is proposed in which an ensemble-averaged yield criterion is constructed for the composites.

Following Ju and Chen (1994c), Ju and Lee (2000), and Ju and Zhang (2001), $H(\mathbf{x} \mid g)$ denotes the square of the "current stress norm" at the local matrix point $\mathbf{x}$, which determines the plastic strain in a composite for a given phase configuration g. $\langle H \rangle_m(\mathbf{x})$ is defined as the ensemble average of $H(\mathbf{x}| g)$ over all possible realization where $\mathbf{x}$ is in the matrix phase. Here, the angled bracket $< \cdot >$ signifies the ensemble average operator. Let $P(g_q)$ be the probability density function for finding the q-phase ($q = 1, 2$) configuration g_q in the composite. Hence, $\langle H \rangle_m$ ($\mathbf{x}$) is obtained as

$$\langle H \rangle_m = H^{\mathrm{o}} + \int_{g_1} \{H(\mathbf{x}|\, g_1) - H^{\mathrm{o}}\}\, P(g_1)\, \mathrm{d}g + \int_{g_2} \{H(\mathbf{x}|\, g_2) - H^{\mathrm{o}}\}\, P(g_2)\, \mathrm{d}g \quad (52)$$

where H^{o} is the square of the far-field stress norm in the matrix:

$$H^{\mathrm{o}} = \boldsymbol{\sigma}^{\mathrm{o}} : \mathbf{I}_{\mathrm{d}} : \boldsymbol{\sigma}^{\mathrm{o}} \quad (53)$$

The stress perturbation due to the presence of single uncracked fiber centered at $\mathbf{x}^{(1)}$ is derived as

$$\boldsymbol{\sigma}'\left(\mathbf{x}|\mathbf{x}^{(1)}\right) = \left[\mathbf{C}^{(0)} \bullet \overline{\mathbf{G}}^{(1)}\left(\mathbf{x} - \mathbf{x}^{(1)}\right)\right] :\ \boldsymbol{\varepsilon}_*^{(1)} \quad (54)$$

Where

$$\boldsymbol{\varepsilon}_*^{(1)} = -\mathbf{K}^{-1} : \boldsymbol{\varepsilon}^{\mathrm{o}} = -\left[\mathbf{S}^{(1)} + \left(\mathbf{C}^{(1)} - \mathbf{C}^{(0)}\right)^{-1} \bullet \mathbf{C}^{(0)}\right]^{-1} : \boldsymbol{\varepsilon}^{\mathrm{o}} \quad (55)$$

$$\overline{\mathbf{G}}^{(1)}\left(\mathbf{x} - \mathbf{x}^{(1)}\right) \equiv \int_{\Omega^{(1)}} \mathbf{G}(\mathbf{x} - \mathbf{x}')\, \mathrm{d}\mathbf{x}' \quad (56)$$

From the *double-inclusion* theory, the local perturbed stress $\boldsymbol{\sigma}'(\mathbf{x})$ due to a cracked fiber centered at $\mathbf{x}^{(2)}$ can be estimated as

$$\boldsymbol{\sigma}'\left(\mathbf{x}|\mathbf{x}^{(2)}\right) = \left[\mathbf{C}^{(0)} \bullet \overline{\mathbf{G}}^{(1)}\left(\mathbf{x} - \mathbf{x}^{(1)}\right)\right] :\ \boldsymbol{\varepsilon}_*^{(1)} + \left[\mathbf{C}^{(0)} \bullet \overline{\mathbf{G}}^{(2)}\left(\mathbf{x} - \mathbf{x}^{(2)}\right)\right] :\ \rho\left(\boldsymbol{\varepsilon}_*^{(2)} - \boldsymbol{\varepsilon}_*^{(1)}\right) \quad (57)$$

where

$$\boldsymbol{\varepsilon}_*^{(2)} = -\mathbf{J}^{*-1} : \boldsymbol{\varepsilon}^{\mathrm{o}} \quad (58)$$

$$\overline{\mathbf{G}}^{(2)}\left(\mathbf{x} - \mathbf{x}^{(2)}\right) \equiv \int_{\Omega^{(2)}} \mathbf{G}(\mathbf{x} - \mathbf{x}')\, \mathrm{d}\mathbf{x}' \quad (59)$$

Here, $\boldsymbol{\varepsilon}^{\mathrm{o}}$ is the elastic strain field induced by the far-field loading, and $\mathbf{x}'$ resides in either an uncracked fiber or a cracked fiber. Further, A is the statistically representative area element (RAE). It is realized that Eqs. 40 and 43 represent the method of Green's function. The components of the fourth-rank tensor $\mathbf{G}$ can be found in Ju et al. (2006, 2008, 2009) and Ju and Ko (2008). In the first-order approximation approach, a matrix point simply collects the perturbations from all noninteracting fibers one by one. In the absence of actual microstructural evidences, for simplicity, $P(\mathbf{x}_1^{(1)})$ and $P(\mathbf{x}_2^{(1)})$ are assumed to be statistically homogeneous, isotropic and uniform. That is, it is assumed that the probability density functions

take the form $P\left(\mathbf{x}_1^{(1)}\right) = \frac{N_1}{A}$ and $P\left(\mathbf{x}_2^{(1)}\right) = \frac{N_2}{A}$, where N_1 and N_2 are the total numbers of uncracked fibers and cracked fibers, respectively, dispersed in a representative area A.

Using the two identities (Eqs. 34 and 35) in Ju and Zhang (1998) and the perturbed stress given in Eqs. 38 and 41, the ensemble-averaged current stress norm at any matrix point is obtained

$$\langle H \rangle_m(\mathbf{x}) = \boldsymbol{\sigma}^{\mathrm{o}} : \mathbf{T} : \boldsymbol{\sigma}^{\mathrm{o}} \tag{60}$$

The components of the positive definite fourth-rank tensor $\mathbf{T}$ read

$$T_{ijkl} = \mathrm{T}_{\mathrm{IK}}^{(1)} \delta_{ij}\delta_{kl} + \mathrm{T}_{\mathrm{IJ}}^{(2)} \left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right), \quad i,\ j,\ k,\ l = 1,\ 2 \tag{61}$$

where the (current) volume fraction of the cracked fibers is defined as $\phi^{(2)} \equiv \pi a_1 a_2 \frac{N_2}{A}$; see section "Overall Elastoplastic-Damage Stress–Strain Responses" for more details.

The general relationship between the applied far-field stress $\boldsymbol{\sigma}^{\mathrm{o}}$ and macroscopic (ensemble volume averaged) stress $\overline{\boldsymbol{\sigma}}$ is given by (cf. Ju and Chen 1994c)

$$\boldsymbol{\sigma}^{\mathrm{o}} = \mathbf{P} : \overline{\boldsymbol{\sigma}} \tag{62}$$

where the fourth-rank tensor $\mathbf{P}$ reads

$$\mathbf{P} = \left\{ \mathbf{C}^{(0)} \bullet \left[\mathbf{I} + \left(\mathbf{I} - \mathbf{S}^{(1)} \right) \bullet \mathbf{Y} \right] \bullet \mathbf{C}^{(0)\,-1} \right\}^{-1} \tag{63}$$

$$\mathbf{Y} = \phi \left[\mathbf{S}^{(1)} + \left(\mathbf{C}^{(1)} - \mathbf{C}^{(0)} \right)^{-1} \bullet \mathbf{C}^{(0)} \right]^{-1} \tag{64}$$

Combination of Eqs. 44 and 46 leads to an alternative expression of the ensemble averaged square of the current stress norm as

$$\langle H \rangle_m(\mathbf{x}) = \overline{\boldsymbol{\sigma}} : \overline{\mathbf{T}} : \overline{\boldsymbol{\sigma}} \tag{65}$$

where the fourth-rank tensor $\overline{\mathbf{T}} \equiv \mathbf{P}^{\mathbf{T}} : \mathbf{T} : \mathbf{P}$. Furthermore, Eq. 49 reduces to the classical J_2-invariant for $\phi = \phi^{(2)} = 0$; i.e., the matrix only material; see Ko (Ko et al. 2005), Ju et al. (2006, 2008, 2009), and Ju and Ko (2008) for more details.

The ensemble-averaged "current stress norm" for any matrix point $\mathbf{x}$ in a 3-phase fiber-reinforced composite can be defined as:

$$\sqrt{\langle H \rangle\ (\mathbf{x})} = \left(1 - \phi^{(1)}\right) \sqrt{\overline{\boldsymbol{\sigma}} : \overline{\mathbf{T}} : \overline{\boldsymbol{\sigma}}} \tag{66}$$

where $\phi^{(1)}$ is the current uncracked fiber volume fraction. Therefore, the effective yield function for the 3-phase fiber-reinforced ductile matrix composites can be proposed as

$$\overline{F} = \left(1 - \phi^{(1)}\right) \sqrt{\overline{\boldsymbol{\sigma}} : \overline{\mathbf{T}} : \overline{\boldsymbol{\sigma}}} - K(\overline{e}^{\mathrm{p}}) \tag{67}$$

with the isotropic hardening function $K(\overline{e}^{\mathrm{p}})$ for the 3-phase composite materials. For illustration, it is assumed that the overall flow rule for the composite is associative. Therefore, the effective ensemble-averaged plastic strain rate for the composite reads

$$\dot{\overline{\boldsymbol{\varepsilon}}}^{p} = \dot{\lambda} \frac{\partial \overline{F}}{\partial \overline{\boldsymbol{\sigma}}} = \left(1 - \phi^{(1)}\right) \dot{\lambda} \frac{\overline{\mathbf{T}} : \overline{\boldsymbol{\sigma}}}{\sqrt{\overline{\boldsymbol{\sigma}} : \overline{\mathbf{T}} : \overline{\boldsymbol{\sigma}}}} \tag{68}$$

where $\dot{\lambda}$ denotes the plastic consistency parameter.

The effective equivalent plastic strain rate for the composite is defined as

$$\dot{\overline{e}}^{p} \equiv \sqrt{\frac{2}{3} \dot{\overline{\boldsymbol{\varepsilon}}}^{p} : \overline{\mathbf{T}}^{-1} : \dot{\overline{\boldsymbol{\varepsilon}}}^{p}} = \sqrt{\frac{2}{3}} \left(1 - \phi^{(1)}\right) \dot{\lambda} \tag{69}$$

The $\dot{\lambda}$ together with the yield function $\overline{F}$ must obey the *Kuhn-Tucker* conditions:

$$\dot{\lambda} \geq 0, \qquad \overline{F} \leq 0, \qquad \dot{\lambda} \overline{F} = 0, \qquad \dot{\lambda} \dot{\overline{F}} = 0 \tag{70}$$

The simple power-law type isotropic hardening function is employed as an example:

$$K(\overline{e}^{p}) = \sqrt{\frac{2}{3}} \left\{ \overline{\sigma}_{y} + h(\overline{e}^{p})^{q} \right\} \tag{71}$$

where $\overline{\sigma}_{y}$ is the initial yield stress, and h and q signify the linear and exponential isotropic hardening parameters, respectively, for the 3-phase composites.

Evolutionary Fiber Cracking

The progressive fiber cracking may occur under increasing deformations and influence the overall stress–strain behavior of composites. After the fiber cracking, the cracked fibers lose the load carrying capacity and are modeled by the *double-inclusion* theory. Within the context of the first-order approximation, the stresses inside fibers are uniform. For convenience, following Tohgo and Weng (1994) and Zhao and Weng (1996, 1997), the average internal stress of a fiber as the controlling

factor is employed. The probability of fiber cracking is modeled as a two-parameter Weibull process. Assuming that the Weibull (1951) statistics governs, the cumulative probability distribution function of fiber cracking, P_d, for the plane-strain uniaxial tensile loading can be expressed as follows. For transverse uniaxial tensile loading along the X_2-axis,

$$P_d = P_2 = \left\{ 1 - \exp\left[-\left(\frac{\overline{\sigma}_{22}^{(1)} - \sigma_{cri}}{S_0} \right)^M \right], \qquad \overline{\sigma}_{22}^{(1)} \geq \sigma_{cri}\, 0, \quad \overline{\sigma}_{22}^{(1)} < \sigma_{cri}. \right. \tag{72}$$

Here $\overline{\sigma}_{22}^{(1)}$ is the average internal stress of fibers in the X_2-direction. In addition, the Weibull parameters M and S_0 denote the evolution rate of the volume fractions of cracked fibers and the average fracture strength of fibers, respectively. The parameter σ_{cri} represents the critical local fracture strength of fibers.

The current volume fraction of cracked fibers $\phi^{(2)}$ at given level of $\overline{\sigma}_{22}^{(1)}$ is given by

$$\phi^{(2)} = \phi \;\; P_2 \tag{73}$$

and the current volume fraction of uncracked fibers can be written as

$$\phi^{(1)} = \phi - \phi^{(2)} \tag{74}$$

The averaged internal stresses of fibers can be expressed as

$$\begin{aligned} \overline{\boldsymbol{\sigma}}^{(1)} = \Bigg[\mathbf{C}^{(0)} \bullet \Bigg\{ \mathbf{I} + \left(\mathbf{I} - \mathbf{S}^{(1)} \right) \bullet \left[\mathbf{S}^{(1)} + \left(\mathbf{C}^{(1)} - \mathbf{C}^{(0)} \right)^{-1} \bullet \mathbf{C}^{(0)} \right]^{-1} \Bigg\} \\ \bullet \left[\mathbf{I} + \left(\mathbf{I} - \mathbf{S}^{(1)} \right) \bullet \mathbf{Y} \right]^{-1} \bullet \mathbf{C}^{(0)\,-1} \Bigg] : \overline{\boldsymbol{\sigma}} \end{aligned} \tag{75}$$

where

$$\mathbf{Y} = \sum_{\beta=1}^{2} \phi^{(\beta)} \left[\mathbf{S}^{(1)} + \left(\mathbf{C}^{(\beta)} - \mathbf{C}^{(0)} \right)^{-1} \bullet \mathbf{C}^{(0)} \right]^{-1}, \quad \beta = 1, 2 \tag{76}$$

Overall Elastoplastic-Damage Stress–Strain Responses

In order to illustrate the proposed micromechanics-based elastoplastic-damage model for cylindrical fiber-reinforced ductile matrix composites, the examples of the transverse *uniaxial* tensile loading under the plane-strain condition are considered.

Transverse Uniaxial Loading Along the X_2-Axis

The applied macroscopic stress $\overline{\boldsymbol{\sigma}}$ can be written as

$$\overline{\sigma}_{22} > 0, \quad \overline{\sigma}_{33} = \eta_2 \overline{\sigma}_{22}, \quad \text{all other } \overline{\sigma}_{ij} = 0. \tag{77}$$

With the isotropic hardening law described by Eq. 55, the overall yield function becomes

$$\overline{F}(\overline{\sigma}_{22}, \overline{e}^p) = \left(1 - \phi^{(1)}\right) \sqrt{\overline{\boldsymbol{\sigma}} : \overline{\mathbf{T}} : \overline{\boldsymbol{\sigma}}} - \sqrt{\frac{2}{3}} \left\{\overline{\sigma}_y + h(\overline{e}^p)^q\right\} \tag{78}$$

Substituting Eq. 61 into Eq. 62, the effective yield function for the uniaxial loading reads

$$\overline{F}(\overline{\sigma}_{22}, \overline{e}^p) = \left(1 - \phi^{(1)}\right) \sqrt{\left(\overline{\mathrm{T}}_{22}^{(1)} + 2\overline{\mathrm{T}}_{22}^{(2)}\right)} \overline{\sigma}_{22} - \sqrt{\frac{2}{3}} \left\{\overline{\sigma}_y + h(\overline{e}^p)^q\right\} \tag{79}$$

The macroscopic incremental plastic strain rate defined by Eq. 52 then takes the form

$$\Delta \overline{\boldsymbol{\varepsilon}}^p = \left(1 - \phi^{(1)}\right) \frac{\Delta\lambda}{\sqrt{\left(\overline{\mathrm{T}}_{22}^{(1)} + 2\overline{\mathrm{T}}_{22}^{(2)}\right)}} \begin{bmatrix} \overline{\mathrm{T}}_{12}^{(1)} & 0 \\ 0 & \overline{\mathrm{T}}_{22}^{(1)} + 2\overline{\mathrm{T}}_{22}^{(2)} \end{bmatrix} \tag{80}$$

for any stress beyond the initial yielding where $\Delta\lambda$ is the incremental plastic consistency parameter. Similarly, the incremental equivalent plastic strain can be expressed as

$$\Delta \overline{e}^p = \sqrt{\frac{2}{3}} \left(1 - \phi^{(1)}\right) \Delta\lambda \tag{81}$$

The macroscopic incremental elastic is

$$\Delta \overline{\boldsymbol{\varepsilon}}^e = \begin{bmatrix} \overline{D}_{12} + \eta_2 \overline{D}_{13} & 0 \\ 0 & \overline{D}_{22} + \eta_2 \overline{D}_{23} \end{bmatrix} \Delta \overline{\sigma}_{22} \tag{82}$$

where $\overline{D}_{ij}$ is the effective elastic compliance of the composites and $\eta_2 = \frac{\overline{C}_{11}\overline{C}_{23} - \overline{C}_{12}\overline{C}_{13}}{\overline{C}_{11}\overline{C}_{22} - \overline{C}_{12}^{\,2}}$.

For the monotonic plane-strain uniaxial loading, the overall incremental macroscopic stress–strain relation can be obtained by summing Eqs. 64 and 66 as follows:

$$\Delta\overline{\boldsymbol{\varepsilon}} = \begin{bmatrix} \overline{D}_{12} + \eta_2 \overline{D}_{13} & 0 \\ 0 & \overline{D}_{22} + \eta_2 \overline{D}_{23} \end{bmatrix} \Delta\overline{\sigma}_{22} + \left(1 - \phi^{(1)}\right) \frac{\Delta\lambda}{\sqrt{\left(\overline{\mathrm{T}}_{22}^{(1)} + 2\overline{\mathrm{T}}_{22}^{(2)}\right)}} \begin{bmatrix} \overline{\mathrm{T}}_{12}^{(1)} & 0 \\ 0 & \overline{\mathrm{T}}_{22}^{(1)} + 2\overline{\mathrm{T}}_{22}^{(2)} \end{bmatrix} \tag{83}$$

where the positive parameter $\Delta\lambda$ is solved from the nonlinear equation obtained by enforcing the plastic consistency condition $\overline{F} = 0$:

$$\left(1 - \phi^{(1)}\right) \sqrt{\left(\overline{\mathrm{T}}_{22}^{(1)} + 2\overline{\mathrm{T}}_{22}^{(2)}\right)} \left(\overline{\sigma}_{22}\right)_{n+1} = \sqrt{\frac{2}{3}} \left\{\overline{\sigma}_y + h\left[\overline{e}_n^P + \Delta\overline{e}_{n+1}^P\right]^q\right\} \tag{84}$$

Equations 65 and 68 then lead to:

$$\left(1 - \phi^{(1)}\right) \sqrt{\left(\overline{\mathrm{T}}_{22}^{(1)} + 2\overline{\mathrm{T}}_{22}^{(2)}\right)} \left(\overline{\sigma}_{22}\right)_{n+1} = \sqrt{\frac{2}{3}} \left\{\overline{\sigma}_y + h\left[\overline{e}_n^P + \sqrt{\frac{2}{3}}\left(1 - \phi^{(1)}\right)\Delta\lambda\right]^q\right\} \tag{85}$$

Therefore, the expression for $\Delta\lambda$ reads

$$\Delta\lambda = \frac{1}{\sqrt{\frac{2}{3}}\left(1 - \phi^{(1)}\right)} \left\{ \left[\frac{\left(1 - \phi^{(1)}\right)\sqrt{\frac{3}{2}\left(\overline{\mathrm{T}}_{22}^{(1)} + 2\overline{\mathrm{T}}_{22}^{(2)}\right)} \left(\overline{\sigma}_{22}\right)_{n+1} - \overline{\sigma}_y}{h}\right]^{1/q} - \overline{e}_n^p \right\} \tag{86}$$

Numerical Simulations and Experimental Comparison

To assess the predictive capability of the present framework, a number of numerical examples are presented in this section. The numerical predictions of CFRMMCs with fiber cracking are validated by the three sets of General Dynamics experimental data of filament-cracking in continuous boron fiber-reinforced 2024 aluminum alloy metal matrix composites under the transverse uniaxial tensile loading (Adams 1970). As reported by Adams (1970), failure of the composites was associated with failure of the filaments, i.e., the typical fracture surface passed through the filaments along diametrical planes normal to the direction of applied stress, and not around each filament (i.e., interfacial debonding). The elastic moduli of the matrix and fibers are: $E_0 = 8$, 100 ksi, $\nu_0 = 0.32$ for the 2024 aluminum alloy metal matrix, and $E_1 = 55{,}000$ ksi, $\nu_1 = 0.2$ for the boron fiber. Moreover, the least-square parameter estimation procedure is employed to estimate the following plastic parameters: $\overline{\sigma}_y = 23$ ksi, $h = 120$ ksi, $q = 0.60$, and the Weibull parameters: $S_0 = 50$ ksi, $M = 5$, and $\sigma_{cri} = 5$ ksi based on the experimental data.

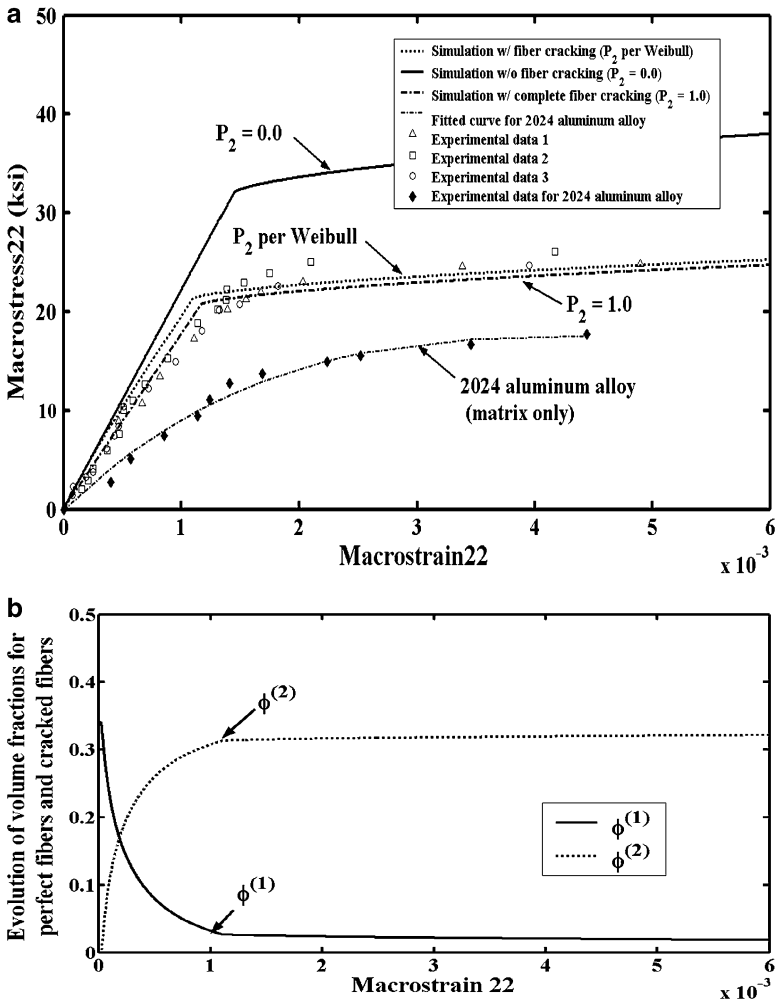


Fig. 5 (**a**) The comparison between the predicted and the measured transverse uniaxial stress–strain behavior of Boron/2024 aluminum alloy composites with the initial fiber volume fraction 34 % at room temperature under varying P_2; (**b**) the corresponding progressive damage of fibers

The composites are subjected to the transverse uniaxial tensile loading along the X_2-axis (cf. Fig. 2). As shown in Fig. 5a, our theoretical prediction of $\overline{\sigma}_{22}$ versus $\overline{\varepsilon}_{22}$ with damage evolution per Eq. 56 matches the three sets of experimental data very well. The corresponding progressive damage of fibers is shown in Fig. 5b. Two stages of mechanical behavior of the composites are well characterized by proposed framework. In the first stage, the minor damage is coupled with elasticity of metal matrix before the composites yields (around strain = 0.0023). In the second stage, after the composites yields, the behavior of the sudden stiffness degradation is

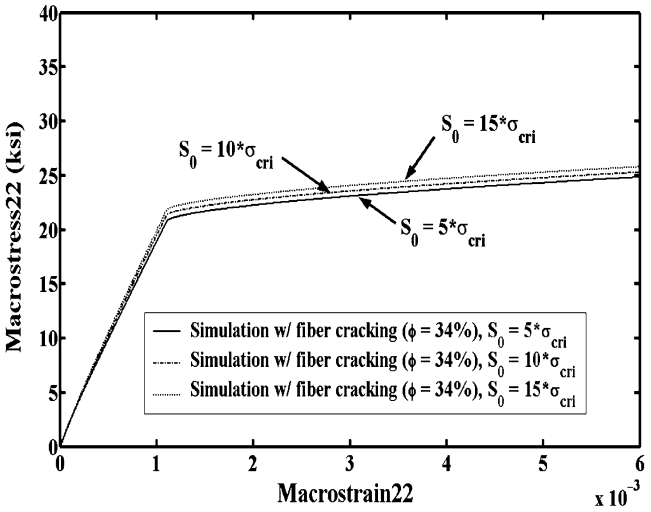


Fig. 6 Theoretical predictions of transverse uniaxial stress–strain behavior of Boron/2024 aluminum alloy composites with the initial fiber volume fraction 34 % at room temperature under varying S_0

mainly controlled by fiber-cracking where the maximum fiber volume fraction of cracked fibers reaches about 33 % as illustrated in Fig. 5b. Furthermore, the theoretical predictions with varying P_2 are also shown in Fig. 5a. It is observed that the prediction with $P_2 = 0.0$ demonstrates the highest stiffness in both the elastic and elastoplastic ranges of the stress–strain curve, as it corresponds to the stress–strain curve without any cracked fibers. On the other hand, the prediction with $P_2 = 1.0$ demonstrates the lowest stiffness in both the elastic and elastoplastic ranges of the stress–strain curve, as it corresponds to the stress–strain curve with fiber cracking inside all fibers in the composites before any loading is applied.

In Fig. 6, $\overline{\sigma}_{22}$ versus $\overline{\varepsilon}_{22}$ with varying S_0 is exhibited ($S_0 = 5\sigma_{cri}$; $S_0 = 10\sigma_{cri}$; $S_0 = 15\sigma_{cri}$). As displayed in Fig. 6, the effects of average fracture strength of fibers play an important role in the mechanical behavior of fiber-cracking especially after the composites yield. The higher S_0 value leads to higher post-yield stiffness of the composites.

Parameters in Equation 38 for Tensor T

$$T_{11}^{(1)} = \overline{a} + A_{q1} + A_{q2} - 2T_{11}^{(2)}, T_{22}^{(1)} = \overline{b} + B_{q1} + B_{q2} - 2T_{22}^{(2)} \tag{87}$$

$$T_{11}^{(2)} = \frac{1}{8(\nu_0 - 1)^2}\left[(5\overline{a} + 5\overline{b} + 3\overline{c} + 2)\left(\phi B_{11}^{(2)2} + B_{(2)11}^{(2)}{}^2\phi^{(2)}\right)\mu_0{}^2 + 4(\nu_0 - 2)\nu_0 + 4\right] \tag{88}$$

$$T_{22}^{(2)} = \frac{1}{8(\nu_0-1)^2}\left[\left(5\bar{a}+5\bar{b}+3\bar{c}+2\right)\left(\phi B_{22}^{(2)2}+B_{(2)22}^{(2)}{}^2\phi^{(2)}\right)\mu_0{}^2+4(\nu_0-2)\nu_0+4\right] \tag{89}$$

$$T_{21}^{(2)} = T_{12}^{(2)} = \frac{1}{4}\left(2+D_{q1}+D_{q2}\right), T_{21}^{(1)} = T_{12}^{(1)} = \frac{1}{2}\left(\bar{c}+C_{q1}+C_{q2}\right) \tag{90}$$

$$\begin{aligned} A_{q1} = \phi\Bigg\{ & \frac{1}{32(\nu_0-1)^2}\Big[\Big(\left(17\bar{a}+\bar{b}-\bar{c}+10\right)B_{11}^{(1)2}-2\left(B_{21}^{(1)}\left(\bar{a}+\bar{b}+7\bar{c}-6\right)\right. \\ & \left.-2B_{11}^{(2)}\left(17\bar{a}+\bar{b}-\bar{c}+10\right)\right)B_{11}^{(1)}+4B_{11}^{(2)2}\left(17\bar{a}+\bar{b}-\bar{c}+10\right) \\ & +B_{21}^{(1)2}\left(\bar{a}+17\bar{b}-\bar{c}+10\right)-4B_{21}^{(1)}B_{11}^{(2)}\left(\bar{a}+\bar{b}+7\bar{c}-6\right)\Big)\mu_0{}^2\Big]\Bigg\} \end{aligned} \tag{91}$$

$$\begin{aligned} A_{q2} = \phi^{(2)}\Bigg\{ & \frac{1}{32(\nu_0-1)^2}\Big[\Big(\left(17\bar{a}+\bar{b}-\bar{c}+10\right)B_{(1)11}^{(2)}{}^2-2\left(B_{(1)21}^{(2)}\left(\bar{a}+\bar{b}+7\bar{c}-6\right)\right. \\ & \left.-2B_{(2)11}^{(2)}\left(17\bar{a}+\bar{b}-\bar{c}+10\right)\right)B_{(1)11}^{(2)}+4B_{(2)11}^{(2)}{}^2\left(17\bar{a}+\bar{b}-\bar{c}+10\right) \\ & +B_{(1)21}^{(2)}{}^2\left(\bar{a}+17\bar{b}-\bar{c}+10\right)-4B_{(1)21}^{(2)}B_{(2)11}^{(2)}\left(\bar{a}+\bar{b}+7\bar{c}-6\right)\Big)\mu_0{}^2\Big]\Bigg\} \end{aligned} \tag{92}$$

$$\begin{aligned} B_{q1} = & \frac{1}{32(\nu_0-1)^2}\Big\{\Big[\left(\bar{b}-\bar{c}+10\right)B_{12}^{(1)2}-2\left(B_{22}^{(1)}+2\right)\left(\bar{b}+7\bar{c}-6\right)B_{12}^{(1)}+ \\ & \bar{a}\left(17B_{12}^{(1)2}-2\left(B_{22}^{(1)}+2\right)B_{12}^{(1)}+\left(B_{22}^{(1)}+2\right)^2\right)+\left(B_{22}^{(1)}+2\right)^2\left(17\bar{b}-\bar{c}+10\right)\Big]\mu_0{}^2\phi\Big\} \end{aligned} \tag{93}$$

$$\begin{aligned} B_{q2} = & \frac{1}{32(\nu_0-1)^2}\Big\{\Big[\left(17\bar{a}+\bar{b}-\bar{c}+10\right)B_{(1)12}^{(2)}{}^2-2\left(B_{(1)22}^{(2)}+2B_{(2)22}^{(2)}\right)\left(\bar{a}+\bar{b}+7\bar{c}-6\right)B_{(1)12}^{(2)} \\ & +\left(B_{(1)22}^{(2)}+2B_{(2)22}^{(2)}\right)^2\left(\bar{a}+17\bar{b}-\bar{c}+10\right)\Big]\mu_0{}^2\phi^{(2)}\Big\} \end{aligned} \tag{94}$$

$$\begin{aligned} C_{q1} = & \frac{1}{16(\nu_0-1)^2}\Big\{\Big[-B_{12}^{(1)}\left(B_{21}^{(1)}\left(\bar{a}+\bar{b}+7\bar{c}-6\right)-2B_{11}^{(2)}\left(17\bar{a}+\bar{b}-\bar{c}+10\right)\right) \\ & +\left(B_{22}^{(1)}+2B_{22}^{(2)}\right)\left(B_{21}^{(1)}\left(\bar{a}+17\bar{b}-\bar{c}+10\right)-2B_{11}^{(2)}\left(\bar{a}+\bar{b}+7\bar{c}-6\right)\right) \\ & +B_{11}^{(1)}\left(B_{12}^{(1)}\left(17\bar{a}+\bar{b}-\bar{c}+10\right)-\left(B_{22}^{(1)}+2B_{22}^{(2)}\right)\left(\bar{a}+\bar{b}+7\bar{c}-6\right)\right)\Big]\mu_0{}^2\phi\Big\} \end{aligned} \tag{95}$$

$$\begin{aligned} C_{q2} = & \frac{1}{16(\nu_0-1)^2}\Big\{\Big[-B_{(1)12}^{(2)}\left(B_{(1)21}^{(2)}\left(\bar{a}+\bar{b}+7\bar{c}-6\right)-2B_{(2)11}^{(2)}\left(17\bar{a}+\bar{b}-\bar{c}+10\right)\right) \\ & +\left(B_{(1)22}^{(2)}+2B_{(2)22}^{(2)}\right)\left(B_{(1)21}^{(2)}\left(\bar{a}+17\bar{b}-\bar{c}+10\right)-2B_{(2)11}^{(2)}\left(\bar{a}+\bar{b}+7\bar{c}-6\right)\right) \\ & +B_{(1)11}^{(2)}\left(B_{(1)12}^{(2)}\left(17\bar{a}+\bar{b}-\bar{c}+10\right)-\left(B_{(1)22}^{(2)}+2B_{(2)22}^{(2)}\right)\left(\bar{a}+\bar{b}+7\bar{c}-6\right)\right)\Big]\mu_0{}^2\phi^{(2)}\Big\} \end{aligned} \tag{96}$$

$$D_{q1} = \frac{1}{2(\nu_0 - 1)^2}\left[\left(5\overline{a} + 5\overline{b} + 3\overline{c} + 2\right)B_{12}^{(2)2}\mu_0{}^2\phi\right] \tag{97}$$

$$D_{q2} = \frac{1}{2(\nu_0 - 1)^2}\left[\left(5\overline{a} + 5\overline{b} + 3\overline{c} + 2\right)B_{(2)12}^{(2)}{}^2\mu_0{}^2\phi^{(2)}\right] \tag{98}$$

$$B_{MR}^{(1)} = -\frac{\left[\frac{1}{4}m\ \left(YYY_{M1}^{(1)} + YYY_{M2}^{(1)}\right) + \frac{1}{2}n\ \left(YYY_{MR}^{(1)}\right)\right]}{2\left(\mu^{(1)\prime} + \frac{S_{MM}^{(2)}}{4(1-\nu_0)}\right)} + \frac{\frac{1}{2}m}{4\left(\mu^{(1)\prime} + \frac{S_{MM}^{(2)}}{4(1-\nu_0)}\right)},$$
$$M, R = 1, 2 \tag{99}$$

$$B_{MN}^{(2)} = \frac{\frac{1}{2}n}{4\left(\mu^{(1)\prime} + \frac{S_{MN}^{(2)}}{4(1-\nu_0)}\right)}, \qquad M, N = 1, 2 \tag{100}$$

$$\begin{Bmatrix} YYY_{M1}^{(1)} \\ YYY_{M2}^{(1)} \end{Bmatrix} = \begin{bmatrix} \lambda^{(1)\prime} + 2\mu^{(1)\prime} + \frac{\left(S_{11}^{(1)} + 2S_{11}^{(2)}\right)}{4(1-\nu_0)} & \lambda^{(1)\prime} + \frac{S_{21}^{(1)}}{4(1-\nu_0)} \\ \lambda^{(1)\prime} + \frac{S_{12}^{(1)}}{4(1-\nu_0)} & \lambda^{(1)\prime} + 2\mu^{(1)\prime} + \frac{\left(S_{22}^{(1)} + 2S_{22}^{(2)}\right)}{4(1-\nu_0)} \end{bmatrix}^{-1}$$
$$\begin{Bmatrix} \lambda^{(1)\prime} + \frac{S_{M1}^{(1)}}{4(1-\nu_0)} \\ \lambda^{(1)\prime} + \frac{S_{M2}^{(1)}}{4(1-\nu_0)} \end{Bmatrix} \tag{101}$$

$$m = \frac{1}{\lambda_0 + \mu_0} - \frac{1}{\mu_0}, \quad n = \frac{1}{\mu_0}, \quad \lambda^{(1)\prime} = \frac{\lambda_0\mu_1 - \lambda_1\mu_0}{2(\mu_1 - \mu_0)[\lambda_1 - \lambda_0 + \mu_1 - \mu_0]}, \quad \mu^{(1)\prime} = \frac{\mu_0}{2(\mu_1 - \mu_0)} \tag{102}$$

$$B_{(1)}^{(2)}MR = \left[\sum_{i=1}^{2} A_{(1)Mi}^{(2)}\frac{1}{4}m + A_{(2)MM}^{(2)}\frac{1}{2}m + A_{(1)MR}^{(2)}\frac{1}{2}n\right], \quad M, R = 1, 2 \tag{103}$$

$$B_{(2)MN}^{(2)} = A_{(2)MN}^{(2)}\frac{1}{2}n, \qquad M, N = 1, 2 \tag{104}$$

$$A_{(1)IK}^{(2)} = \frac{-(-YJ_{IK})}{2J_{II}^{(2)}}, \quad A_{(2)IJ}^{(2)} = \frac{-1}{4J_{IJ}^{(2)}}, \qquad I, J, K = 1, 2 \tag{105}$$

$$\begin{Bmatrix} YJ_{I1} \\ YJ_{I2} \end{Bmatrix} = \begin{bmatrix} J_{11}^{(1)} + 2J_{11}^{(2)} & J_{21}^{(1)} \\ J_{12}^{(1)} & J_{22}^{(1)} + 2J_{22}^{(2)} \end{bmatrix}^{-1} \begin{Bmatrix} J_{I1}^{(1)} \\ J_{I2}^{(1)} \end{Bmatrix} \tag{106}$$

$$\overline{a} = \frac{2\eta_1^2}{9} - \frac{2\eta_1}{9} + \frac{1}{9}(2\eta_1 - 1)^2 + \frac{5}{9}, \quad \overline{b} = \frac{2\eta_2^2}{9} - \frac{2\eta_2}{9} + \frac{1}{9}(2\eta_2 - 1)^2 + \frac{5}{9} \tag{107}$$

$$\overline{c} = \frac{4\eta_2\eta_1}{9} - \frac{2\eta_1}{9} - \frac{2\eta_2}{9} + \frac{2}{9}(2\eta_1 - 1)(2\eta_2 - 1) - \frac{8}{9} \tag{108}$$

$$\eta_1 = \frac{\overline{C}_{13}\overline{C}_{22} - \overline{C}_{12}\overline{C}_{23}}{\overline{C}_{11}\overline{C}_{22} - \overline{C}_{12}^{\,2}}, \quad \eta_2 = \frac{\overline{C}_{11}\overline{C}_{23} - \overline{C}_{12}\overline{C}_{13}}{\overline{C}_{11}\overline{C}_{22} - \overline{C}_{12}^{\,2}} \tag{109}$$

Conclusions

In this chapter, a multi-level micromechanical elastoplastic-damage framework is developed to predict the overall elastoplastic behavior and damage evolution of cylindrical fiber-reinforced ductile matrix composites. Progressively cracked fibers are modeled by the *double-inclusion* theory. After the fibers are cracked, the composites become 3-phase composites with different constituent volume fractions at different loading time due to progressive damage evolutions. The effective elastic moduli of 3-phase composites under the plane-strain condition are derived by using the multi-phase micromechanical framework. In order to estimate the overall elastoplastic-damage behavior, an effective yield criterion is micromechanically derived based on the ensemble-area averaging procedure and the first-order effects of eigenstrains due to cylindrical inclusions. The effects of random dispersion of elastic inclusions are considered through the ensemble averaging process. The proposed overall yield criterion, in conjunction with the overall associative plastic flow rule and the hardening law, provides the analytical formulation for the estimation of effective elastoplastic-damage responses of ductile composites.

An evolutionary *fiber cracking* model is subsequently employed in accordance with the Weibull's probability function to characterize the varying probability of *fiber cracking*. The effective elastoplastic-damage behavior is influenced through the evolutionary volume fraction of uncracked fibers and cracked fibers during the damage process. The proposed elastoplastic-damage formulation is applied to the special case of the transverse uniaxial tensile loading under the plane-strain condition to predict the corresponding stress–strain responses. Efficient incremental micromechanical computational algorithms are also presented. Finally, the plane-strain uniaxial numerical simulations are performed. The numerical simulations of the effective elastoplastic-damage behavior of fiber-reinforced composites are presented to illustrate the potential of the proposed *fiber-cracking* damage model under the transverse uniaxial loading.

Future research is warranted to extend the present framework to accommodate the effects of *biaxial* loading on the overall responses of fiber-reinforced composites. The biaxial loading can be classified as *proportional* and *non-proportional* loading. Special attention is given to *non-proportional* loading where the direction and value of the maximum principal normal tensile stress within the fiber keep changing as the loading increases. The orientation-averaging technique is required to account for all possible fiber-cracking directions which are normal to the maximum principal normal tensile stress directions.

References

D.F. Adams, Inelastic analysis of a unidirectional composite subjected to transverse normal loading. J. Comput. Mater. **4**, 310–328 (1970)

G. Bao, Damage due to fracture of brittle reinforcements in a ductile matrix. Acta Metall. Mater. **40**(10), 2547–2555 (1992)

J.R. Brockenbrough, F.W. Zok, On the role of particle cracking in flow and fracture of metal matrix composites. Acta Metall. Mater. **43**(1), 11–20 (1995)

R. Eluripati, A multi-fiber unit cell for prediction of transverse properties in metal matrix composites, M.S. thesis at West Virginia University, 2003

J.D. Eshelby, The determination of the elastic field of an ellipsoidal inclusion, and related problems. Proc. R. Soc. Lond. **A241**, 376–396 (1957)

X.Q. Feng, S.W. Yu, Damage micromechanics for constitutive relations and failure of microcracked quasi-brittle materials. Int. J. Damage Mech. **19**(8), 911–948 (2010)

S. Ghosh, S. Moorthy, Particle fracture simulation in non-uniform microstructures of metal-matrix composites. Acta Mater. **46**(3), 965–982 (1998)

M. Hori, S. Nemat-Nasser, Double-inclusion model and overall moduli of multi-phase composites. Mech. Mater. **14**, 189–206 (1993)

J.W. Ju, T.M. Chen, Micromechanics and effective moduli of elastic composites containing randomly dispersed ellipsoidal inhomogeneities. Acta Mech. **103**, 103–121 (1994a)

J.W. Ju, T.M. Chen, Effective elastic moduli of two-phase composites containing randomly dispersed spherical inhomogeneities. Acta Mech. **103**, 123–144 (1994b)

J.W. Ju, T.M. Chen, Micromechanics and effective elastoplastic behavior of two-phase metal matrix composites. Trans. ASME, J. Eng. Mater. Tech. **116**, 310–318 (1994c)

J.W. Ju, H.K. Lee, A micromechanical damage model for effective elastoplastic behavior of ductile matrix composites considering evolutionary complete particle debonding. Comput. Method Appl. Mech. Eng. **183**, 201–222 (2000)

J.W. Ju, H.K. Lee, A micromechanical damage model for effective elastoplastic behavior of partially debonded ductile matrix composites. Int. J. Solid Struct. **38**, 6307–6332 (2001)

J.W. Ju, Y.F. Ko, Micromechanical elastoplastic damage modeling of progressive interfacial arc debonding for fiber reinforced composites. Int. J. Damage Mech. **17**, 307–356 (2008)

J.W. Ju, L.Z. Sun, Effective elastoplastic behavior of metal matrix composites containing randomly located aligned spheroidal inhomogeneities. Part I: micromechanics-based formulation. Int. J. Solid Struct. **38**(2), 183–201 (2001)

J.W. Ju, K.H. Tseng, Effective elastoplastic behavior of two-phase ductile matrix composites: a micromechanical framework. Int. J. Solid Struct. **33**, 4267–4291 (1996)

J.W. Ju, K.H. Tseng, Effective elastoplastic algorithms for ductile matrix composites. J. Eng. Mech. **123**, 260–266 (1997)

J.W. Ju, K. Yanase, Elastoplastic damage micromechanics for elliptical fiber composites with progressive partial fiber debonding and thermal residual stresses. Theor. Appl. Mech. **35**(1–3), 137–170 (2008)

J.W. Ju, K. Yanase, Micromechanical elastoplastic damage mechanics for elliptical fiber-reinforced composites with progressive partial fiber debonding. Int. J. Damage Mech. **18**(7), 639–668 (2009)
J.W. Ju, K. Yanase, Micromechanics and effective elastic moduli of particle-reinforced composites with near-field particle interactions. Acta Mech. **215**(1), 135–153 (2010)
J.W. Ju, K. Yanase, Micromechanical effective elastic moduli of continuous fiber-reinforced composites with near-field fiber interactions. Acta Mech. **216**(1), 87–103 (2011)
J.W. Ju, X.D. Zhang, Micromechanics and effective transverse elastic moduli of composites with randomly located aligned circular fibers. Int. J. Solid Struct. **35**(9–10), 941–960 (1998)
J.W. Ju, X.D. Zhang, Effective elastoplastic behavior of ductile matrix composites containing randomly located aligned circular fibers. Int. J. Solid Struct. **38**, 4045–4069 (2001)
J.W. Ju, Y.F. Ko, H.N. Ruan, Effective elastoplastic damage mechanics for fiber reinforced composites with evolutionary complete fiber debonding. Int. J. Damage Mech. **15**(3), 237–265 (2006)
J.W. Ju, Y.F. Ko, H.N. Ruan, Effective elastoplastic damage mechanics for fiber reinforced composites with evolutionary partial fiber debonding. Int. J. Damage Mech. **17**(6), 493–537 (2008)
J.W. Ju, Y.F. Ko, X.D. Zhang, Multi-level elastoplastic damage mechanics for elliptical fiber reinforced composites with evolutionary complete fiber debonding. Int. J. Damage Mech. **18** (5), 419–460 (2009)
J.W. Kaczmar, K. Pietrzak, W. Wlosinski, The production and application of metal matrix composite materials. J. Mater. Process. Technol. **106**, 58–67 (2000)
Y.F. Ko, Effective elastoplastic-damage model for fiber-reinforced metal matrix composites with evolutionary fibers debonding, Ph.D. Thesis, UCLA, Dec, 2005
Y.F. Ko, J.W. Ju, New higher-order bounds on effective transverse elastic moduli of three-phase fiber reinforced composites with randomly located and interacting aligned circular fibers. Acta Mech. **223**(11), 2437–2458 (2012)
Y.F. Ko, J.W. Ju, Effective transverse elastic moduli of three-phase hybrid fiber reinforced composites with randomly located and interacting aligned circular fibers of distinct elastic properties and sizes. Acta Mech. **224**(1), 157–182 (2013)
H.K. Lee, J.W. Ju, A three-dimensional stress analysis of a penny-shaped crack interacting with a spherical inclusion. Int. J. Damage Mech. **16**(3), 331–359 (2007)
H.K. Lee, J.W. Ju, 3-D micromechanics and effective moduli for brittle composites with randomly located interacting microcracks and inclusions. Int. J. Damage Mech. **17**(5), 377–417 (2008)
P.J. Lin, J.W. Ju, Effective elastic moduli of three-phase composites with randomly located and interacting spherical particles of distinct properties. Acta Mech. **208**, 11–26 (2009)
H.T. Liu, L.Z. Sun, Effects of thermal residual stresses on effective elastoplastic behavior of metal matrix composites. Int. J. Solid Struct. **41**, 2189–2203 (2004)
H.T. Liu, L.Z. Sun, J.W. Ju, An interfacial debonding model for particle-reinforced composites. Int. J. Damage Mech. **13**, 163–185 (2004)
H.T. Liu, L.Z. Sun, J.W. Ju, Elastoplastic modeling of progressive interfacial debonding for particle-reinforced metal matrix composites. Acta Mech. **181**(1–2), 1–17 (2006)
J. Llorca, J.L. Martínez, M. Elices, Reinforcement fracture and tensile ductility in sphere-reinforced metal matrix composites. Fatigue Fract. Eng. Mater. Struct. **20**(5), 689–702 (1997)
T. Mura, *Micromechanics of Defects in Solids*, 2nd edn. (Kluwer, Dordrecht, 1987)
S. Nemat-Nasser, M. Hori, *Micromechanics: Overall Properties of Heterogeneous Materials* (Elsevier Science, Amsterdam, 1993)
T. Okabe, M. Nishikawa, N. Takeda, Micromechanics on the rate-dependent fracture of discontinuous fiber-reinforced plastics. Int. J. Damage Mech. **19**(3), 339–360 (2010)
G.H.H. Paulino, H.M. Yin, L.Z. Sun, Micromechanics-based interfacial debonding model for damage of functionally graded materials with particle interactions. Int. J. Damage Mech. **15**(3), 267–288 (2006)
M.J.N. Priestley, F. Seible, G.M. Calvi, *Seismic Design and Retrofit of Bridges* (Wiley, New York, 1996)

H.M. Shodja, A.S. Sarvestani, Elastic fields in double inhomogeneity by the equivalent inclusion method. ASME J. Appl. Mech. **68**, 3–10 (2001)
D. Steglich, T. Siegmund, W. Brocks, Micromechanical modeling of damage due to particle cracking in reinforced metals. Comput. Mater. Sci. **16**, 404–413 (1999)
L.Z. Sun, J.W. Ju, Effective elastoplastic behavior of metal matrix composites containing randomly located aligned spheroidal inhomogeneities. Part II: applications. Int. J. Solid Struct. **38** (2), 203–225 (2001)
L.Z. Sun, J.W. Ju, H.T. Liu, Elastoplastic modeling of metal matrix composites with evolutionary particle debonding. Mech. Mater. **35**, 559–569 (2003a)
L.Z. Sun, H.T. Liu, J.W. Ju, Effect of particle cracking on elastoplastic behaviour of metal matrix composites. Int. J. Numer. Meth. Eng. **56**, 2183–2198 (2003b)
K. Tohgo, G.J. Weng, A progress damage mechanics in particle-reinforced metal-matrix composites under high triaxial tension. J. Eng. Mater. Technol. **116**, 414–420 (1994)
V. Tvergaard, A. Needleman, Analysis of the cup-cone fracture in a round tensile bar. Acta Metall. **32**(1), 157–169 (1984)
B.L. Wang, Y.G. Sun, H.Y. Zhang, Multiple cracking of fiber/matrix composites-analysis of normal extension. Int. J. Solid Struct. **45**, 4032–4048 (2008)
W. Weibull, A statistical distribution function of wide applicability. J. Appl. Mech. **18**, 293–297 (1951)
J. Zhang, F. Wang, Modeling of damage evolution and failure in fiber-reinforced ductile composites under thermomechanical fatigue loading. Int. J. Damage Mech. **19**(7), 851–875 (2010)
Y.H. Zhao, G.J. Weng, Plasticity of a two-phase composite with partially debonded inclusions. Int. J. Plast. **12**, 781–804 (1996)
Y.H. Zhao, G.J. Weng, Transversely isotropic moduli of two partially debonded composites. Int. J. Solid Struct. **34**, 493–507 (1997)

Micromechanical Elastoplastic Damage Modeling of Evolutionary Interfacial Arc Debonding for Fiber Reinforced Composites

30

Jiann-Wen Woody Ju and Yu-Fu Ko

Contents

Abstract

An innovative micromechanical elastoplastic evolutionary damage model is presented to predict the effective transverse mechanical behavior and interfacial arc debonding evolution of fiber-reinforced composites. The partial debonding

J.-W.W. Ju
Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, USA
e-mail: juj@ucla.edu

Y.-F. Ko (✉)
Department of Civil Engineering and Construction Engineering Management, California State University, Long Beach, CA, USA
e-mail: Yu-Fu.Ko@csulb.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_11

process at the fiber/matrix interfaces is represented by the growing debonding angles of arc microcracks. Three different types of debonded modes, i.e., perfectly bonded, partially debonded, and completely debonded modes, are considered. For debonded fibers, the elastic equivalency is constructed in terms of the equivalent *orthotropic* yet perfectly bonded elastic cylindrical fibers. The equivalent orthotropic elastic moduli are constructed to characterize the reduction of the load-transfer capacity in the debonded directions. The damage evolution is represented by the debonding angle that is dependent on the external loading condition. The effective elastic moduli of four-phase composites are derived by using a micromechanical formulation. In order to characterize the overall transverse elastoplastic-damage behavior, an effective yield criterion is derived on the basis of the ensemble-area averaging procedure and the first-order effects of eigenstrains upon yielding. Moreover, the varying probability of the evolutionary volume fractions of debonded fibers is characterized by Weibull's probabilistic approach. The proposed effective yield criterion, coupling with the overall plastic flow rule and the hardening law, comprises the analytical framework for the prediction of effective elastoplastic-damage responses of ductile matrix composites containing randomly located yet aligned cylindrical fibers. The proposed micromechanical elastoplastic-damage model is then applied to the transverse uniaxial and transverse biaxial tensile loading with varied stress ratios. Comparisons between the present predictions and available experimental data, as well as other numerical simulations, are performed to elucidate the potential of the proposed formulation.

Introduction

Man-made metal matrix composites (MMCs), containing reinforcements by continuous high-strength fibers, possess the highest strength and stiffness in the direction of the fibers. The overall mechanical behavior of these composites depends on microstructures including the heterogeneous constituents of reinforcement and matrix. Their deformation and damage failure mechanisms are generally different from the monolithic matrix materials. Several possible damage modes exist for fiber composites, such as the interfacial fiber/matrix debonding (Ju et al. 2006, 2008; Ju and Ko 2009; Ju and Lee 2000, 2001; Liu et al. 2004, 2006; Marshall et al. 1994; Nimmer et al. 1991; Pagano and Tandon 1990; Paulino et al. 2006; Sun et al. 2003a; Yanase and Ju 2012; Zhao and Weng 1997, 2002), Voyiadjis and Allen (1996), the matrix cracking (Lee and Mal 1998; Yanase and Ju 2013), the fiber cracking (Ko and Ju 2013a; Sun et al. 2003b), the fiber breakage (Case and Reifsnider 1996; Kim and Nairn 2002; Steif 1984), the fiber pullout (Hsueh 1990; Hutchinson and Jensen 1990; Li et al. 1990, 1993), the kink band formation of fibers (Steif 1990), and the shear sliding of fibers (He and Lim 2001; Xia et al. 1994). Specifically, the dominant damage mechanism in continuous unidirectional fiber-reinforced ductile composites under *transverse* loading is the initiation and

progressive interfacial partial fiber debonding (arc microcracks) followed by plastic yielding (Gundel and Miracle 1998; Ju 1991a, b, 1996; Ju and Lee 1991; Ju and Yanase 2009; Lee and Mal 1998; Lee and Ju 1991; Marshall et al. 1994; Nimmer et al. 1991; Yanase and Ju 2012). For example, the experimental evidence of separation is visible in the form of a ridge of acetate at the interface between the inner coating and the fiber. The existence of this acetate ridge is indicative of a gap between fiber and matrix (Nimmer et al. 1991). For transverse behavior of pseudo-elastic Kevlar single fibers under compression, we refer to Cheng and Chen (2006), and for particle volume fraction change (particle crowding) under indentation in metal matrix composites, we cite Pereyra and Shen (2005).

Moreover, Paulino et al. (2006) presented a micromechanics-based interfacial debonding model for damage of functionally graded materials with particle interactions based on Ju and Chen (1994a, b). Further, Ju et al. (2006, 2008; Ju and Ko 2009) considered the first-order elastoplastic-damage mechanics for fiber-reinforced composites with evolutionary complete or partial fiber debonding. Ju and Lee (1991), Lee and Ju (1991), and Ju and Zhang (1998b, c) also proposed micromechanical damage models for brittle solids under tensile and compressive loading; and Berryman (2006) proposed estimates and rigorous bounds on pore-fluid enhanced shear modulus in poroelastic media with hard and soft anisotropy. For dynamic damage wave in elastic-brittle materials, we refer to Lu et al. (2005); for low-speed impact damage of heterogeneous two-layer materials, we refer to Xu and Rosakis (2005); for high-strain-rate failure process of unidirectional SiC_f –Al composites, we refer to Zhou et al. (2005); and for high velocity impact finite strain plastic-damage model, we cite Abu Al-Rub and Voyiadjis (2006) and Voyiadjis and Abu Al-Rub (2006).

As a result of the interfacial fiber debonding, the strength and strain to failure become lower under transverse tensile loading. However, the extent of performance degradation depends on the evolutionary debonding geometry, size, and number density. To model the interfacial fiber debonding in composites, Jasiuk and Tong (1989), Pagano and Tandon (1990), Qu (1993), Yang and Mal (1995), and Sangani and Mo (1997) introduced either a linear spring layer with vanishing thickness or an interlayer with constant thickness. In their models, a different elastic constant of the spring layer/interlayer from the matrix and reinforcement is used to simulate the *loss* of load-transfer capability through the interface due to local fiber debonding. Since the spring layer/interlayer elastic properties in their models are not position dependent, their models are not applicable to partial debonding mechanisms.

Another physically appealing approach to model interfacial debonding is the "equivalent inclusion method" (cf. Zhao and Weng 1997; Wong and Ait-Kadi 1997; Ju and Sun 1999; Ju and Lee 2000, 2001; Sun et al. 2003a, b; Sun and Ju 2004; Liu et al. 2004; Ju et al. 2006, 2008) in which the isotropic *debonded* inclusions (particles or fibers) are replaced by the perfectly *bonded* inclusions with prescribed *equivalent* anisotropic stiffness tensor to characterize the reduction in load-transfer capacity of the debonded inclusion/matrix interface. Therefore, the celebrated Eshelby's inclusion theory and the micromechanics method can be applied here to deal with multiphase composites with interfacial damage. While providing simple yet attractive methodology, these analytical micromechanical

damage models assume that once interfacial debonding occurs in a certain direction, the capacity of interfacial load transfer is *completely* lost in that direction without considering the *progressive* arc debonding process from one direction to another. In reality, the interfacial debonding between the matrix and fibers is usually a progressive process in which the debonding arc microcrack (represented by the debonding angle) evolves with changing applied tensile loads.

In an effort to further improve the above micromechanical damage models by accounting for the effects of progressive debonding angles upon the overall behavior of partially debonded composites, Zhao and Weng (2002) and Liu et al. (2006) applied the volume ratio of a debonded particle directly beneath the interfacial microcrack as a measure for interfacial damage parameter. Their results show that the progressive partial interfacial debonding process has a significant impact on the overall elastic-damage moduli and elastoplastic-damage responses of composites.

To capture the progressive process of local interfacial fiber/matrix debonding, the evolution of interfacial damage, and the micromechanical transition among various fiber debonding modes, a new micromechanical elastoplastic-damage framework is developed in this chapter to simulate the interfacial debonding between the matrix and the fibers in continuous fiber-reinforced ductile matrix composites under *transverse* mechanical loading. An analytical solution based on direct micromechanical approach is presented. No costly Monte Carlo simulations are needed under the proposed analytical micromechanical framework.

In the present formulation, continuous fiber-reinforced ductile matrix composites consist of identical unidirectional elastic cylindrical fibers that are *randomly* dispersed in the elastoplastic matrix. The unit-cell models are not applicable here due to the inherent limitation of *periodic* microstructure. With increasing external loading, some *randomly*located cylindrical fibers in the composites begin to gradually debond, while the progressive debonding angles (arc microcracks) and the number of debonded fibers (the volume fraction) will also gradually increase during the transverse loading process. At the microstructural level, the local interfacial debonding mechanism is treated as a fracture process in which local stresses play a key role. For those partially debonded fibers, the anisotropic elastic equivalency is constructed by means of the equivalent stiffness tensor. Namely, the originally isotropic but now partially debonded fibers are substituted by the orthotropic yet perfectly bonded fibers embedded in the elastoplastic matrix. The volume fraction evolution of partially debonded fibers is characterized by Weibull's (1951) probabilistic distribution. Furthermore, the first-order micromechanical approximation method proposed by Ju and Chen (1994a, b, c) and Ju and Tseng (1996. 1997) [equivalent to the Mori–Tanaka method (Mori and Tanaka 1973)] is utilized in this chapter to estimate the effective elastic-damage stiffness tensors of resultant multiphase composites. For the second-order evolutionary microcrack interactions and effective elastic moduli, we cite Ju and Tseng (1992, 1995), Ju and Chen (1994d, e), and Lee and Ju (2007, 2008). The proposed evolutionary micromechanical elastoplastic-damage framework is developed under the general two-dimensional loading condition and the probabilistic ensemble-area averaging homogenization procedure. A computational algorithm is proposed to numerically

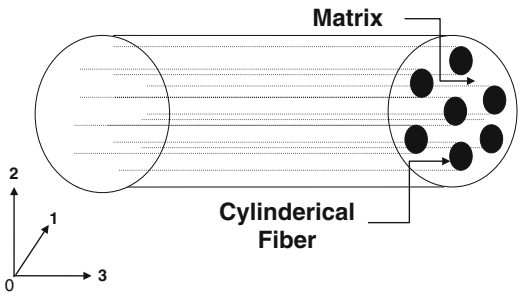


Fig. 1 A schematic plot of a composite reinforced by unidirectionally aligned yet randomly located cylindrical fibers

simulate the overall nonlinear stress–strain behaviors of continuous fiber-reinforced ductile matrix composites (cf. Fig. 1) with progressive probabilistic partial fiber debonding effects. Finally, we present a number of numerical examples under the *transverse* uniaxial and biaxial loading conditions and comparison between the model predictions and available experimental data to elucidate the potential of the proposed statistical-micromechanical elastoplastic-damage framework.

Progressive Fiber Debonding Modes

Under the applied external transverse tensile loading, some cylindrical fibers may initiate partial debonding from the matrix once the local stresses at the interface reach a certain critical level; see Fig. 2.

By employing the equivalent inclusion method (Zhao and Weng 1997, 2002; Sun et al. 2003a, b; Liu et al. 2004, 2006), the *debonded* fibers featuring various stages of partial debonding modes are treated as a different phase of perfectly *bonded* fibers with *anisotropic* elastic equivalency. Therefore, the micromechanical composite theories containing various phases of randomly located yet perfectly bonded inclusions can be employed here. According to Eshelby's inclusion theory (Eshelby 1957), without considering the direct strong interactions among fibers, the stresses inside the perfectly bonded fibers are uniform and are expressed as a function of the external loads as (Mura 1987)

$$\overline{\boldsymbol{\sigma}}^{(1)} = \left[\mathbf{C}^{(0)} \bullet \left\{\mathbf{I} + (\mathbf{I}-\mathbf{S}) \bullet \left[\mathbf{S} + \left(\mathbf{C}^{(\beta)} - \mathbf{C}^{(0)}\right)^{-1} \bullet \mathbf{C}^{(0)}\right]^{-1}\right\} \bullet [\mathbf{I} + (\mathbf{I}-\mathbf{S}) \bullet \mathbf{Y}]^{-1} \bullet \mathbf{C}^{(0)^{-1}}\right] : \overline{\boldsymbol{\sigma}}, \beta = 1,2,3 \tag{1}$$

with

$$\mathbf{Y} = \sum_{\beta=1}^{3} \phi^{(\beta)} \left[\mathbf{S} + \left(\mathbf{C}^{(\beta)} - \mathbf{C}^{(0)}\right)^{-1} \bullet \mathbf{C}^{(0)}\right]^{-1}, \quad \beta = 1,2,3 \tag{2}$$

where $\overline{\boldsymbol{\sigma}}$ is the macroscopic stress, $\mathbf{C}^{(0)}$ represents the elastic stiffness tensor of the matrix, $\mathbf{C}^{(\beta)}$ is the elastic stiffness tensor of the β-phase, $\mathbf{I}$ represents the fourth-rank

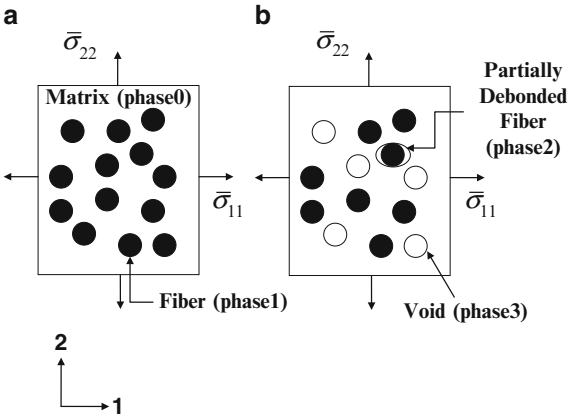


Fig. 2 A schematic diagram of a fiber composite subjected to biaxial transverse tensile loading: (**a**) the initial state (undamaged), (**b**) the damaged state

identity tensor, and $\mathbf{S}$ is the interior-point Eshelby's tensor of a cylindrical inclusion. When $\beta = 1$, it denotes the undamaged cylindrical fibers; when $\beta = 2$, it denotes the partially debonded cylindrical fibers; and when $\beta = 3$, it signifies the completely debonded cylindrical fibers (voids); see Fig. 2. Further, $\boldsymbol{\varepsilon}_*^{(\beta)\mathrm{o}}$ is the "first-order" eigenstrain tensor in the β-phase inclusions and is defined as

$$\boldsymbol{\varepsilon}_*^{(\beta)\mathrm{o}} = -\left[\mathbf{S} + \left(\mathbf{C}^{(\beta)} - \mathbf{C}^{(0)}\right)^{-1} \bullet \mathbf{C}^{(0)}\right]^{-1} : \boldsymbol{\varepsilon}^{\mathrm{o}}, \quad \beta = 1, 2, 3 \tag{3}$$

in which $\boldsymbol{\varepsilon}^{\mathrm{o}}$ is the corresponding strain field due to the far-field stress $\boldsymbol{\sigma}^{\mathrm{o}} = \mathbf{C}^{(0)} : \boldsymbol{\varepsilon}^{\mathrm{o}}$ and $\mathbf{C}^{(\beta)}$ is the elastic stiffness tensor of the β-phase. In Eqs. 1, 2 and 3, the interior-point Eshelby's tensor S_{ijkl} has the following explicit form for an isotropic and elastic *cylindrical* inclusion:

$$S_{ijkl} = \frac{1}{4(1-\nu_0)}\left\{ \mathrm{S}_{\mathrm{IK}}^{(1)}\delta_{ij}\delta_{kl} + \mathrm{S}_{\mathrm{IJ}}^{(2)}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right)\right\}, \quad i,j,k,l = 1,2,3 \tag{4}$$

where ν_0 denotes Poisson's ratio of the matrix and δ_{ij} signifies the Kronecker delta. All components of the second-rank tensors $\mathrm{S}_{\mathrm{IK}}^{(1)}$ and $\mathrm{S}_{\mathrm{IJ}}^{(2)}$ are expressed by

$$\mathrm{S}_{11}^{(1)} = \mathrm{S}_{22}^{(1)} = \mathrm{S}_{12}^{(1)} = \mathrm{S}_{21}^{(1)} = \frac{1}{2}(4\nu_0 - 1), \mathrm{S}_{13}^{(1)} = \mathrm{S}_{23}^{(1)} = 2\nu_0 \tag{5}$$

$$\mathrm{S}_{31}^{(1)} = \mathrm{S}_{32}^{(1)} = \mathrm{S}_{33}^{(1)} = 0, \mathrm{S}_{11}^{(2)} = \mathrm{S}_{22}^{(2)} = \mathrm{S}_{12}^{(2)} = \mathrm{S}_{21}^{(2)} = \frac{1}{2}(3 - 4\nu_0) \tag{6}$$

$$\mathrm{S}_{13}^{(2)} = \mathrm{S}_{23}^{(2)} = \mathrm{S}_{31}^{(2)} = \mathrm{S}_{32}^{(2)} = (1 - \nu_0), \mathrm{S}_{33}^{(2)} = 0 \tag{7}$$

Eshelby's micromechanical formulas given by Eqs. 4, 5, 6, and 7 are valid for composites in which cylindrical fibers are embedded in and perfectly bonded with the surrounding matrix. In the present damage model, when partial interfacial fiber

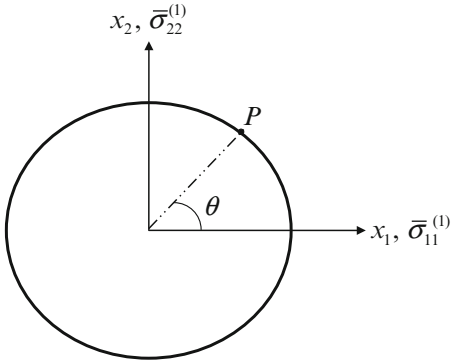


Fig. 3 The Eulerian angle θ representing a surface point P in the local coordinates of a fiber

debonding occurs, the equivalent inclusion method (Zhao and Weng 1997, 2002; Sun et al. 2003a, b; Liu et al. 2004, 2006) is introduced; i.e., the partially debonded isotropic cylindrical fibers are replaced by equivalent orthotropic yet perfectly bonded cylindrical fibers. Consequently, Eshelby's theory is still applicable here.

In what follows, the *local* Cartesian coordinate system is selected to coincide with the two principal directions of the local stress field $\overline{\boldsymbol{\sigma}}^{(1)}$ inside the perfectly bonded fiber. The two local principal radial normal stresses $(\overline{\sigma}_{11}^{(1)}, \overline{\sigma}_{22}^{(1)})$ can be computed accordingly from Eq. 1 and follow the conventional order of $\overline{\sigma}_{11}^{(1)} \geq \overline{\sigma}_{22}^{(1)}$. Tensile stress is taken as positive. For a certain surface point P on a cylindrical fiber, the unit normal direction can be expressed as $\mathbf{n} = \{\cos\theta, \sin\theta, 0\}$, where θ is the Eulerian angle as shown in Fig. 3. Hence, the radial normal stress at a point P can be derived as

$$\sigma^{\text{normal}} = \overline{\sigma}_{11}^{(1)} (\cos\theta)^2 + \overline{\sigma}_{22}^{(1)} (\sin\theta)^2 \tag{8}$$

Emanating from the assumption that the radial normal stress controls the local interfacial partial debonding initiation on the interface, as the (tensile) radial normal stress reaches a critical interfacial debonding strength, i.e., σ_{cri}, the partial debonding process would initiate at the point P. The partial fiber debonding criterion can be conveniently written as

$$\overline{\sigma}_{11}^{(1)} (\cos\theta)^2 + \overline{\sigma}_{22}^{(1)} (\sin\theta)^2 = \sigma_{\text{cri}} \tag{9}$$

Therefore, the relationship between the local principal radial normal stresses and the critical debonding strength results in the following three different types of interfacial debonding modes.

Mode 1: $\sigma_{\text{cri}} \geq \overline{\sigma}_{11}^{(1)} \geq \overline{\sigma}_{22}^{(1)}$ This is actually the undamaged state; i.e., all cylindrical fibers are perfectly bonded since none of the local principal stresses reach the critical debonding strength yet. No partial debonding process is activated.

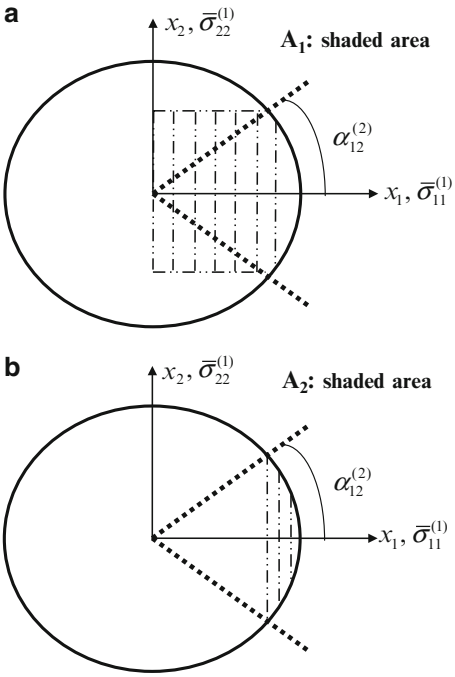


Fig. 4 The damage parameters for Mode 2: (**a**) projection of the damaged arc in the x^1-direction, (**b**) projection of the damaged arc in the x^2-direction

Mode 2: $\overline{\sigma}_{11}^{(1)} \geq \sigma_{\text{cri}} \geq \overline{\sigma}_{22}^{(1)}$ Only one local principal stress is greater than the critical interfacial debonding strength σ_{cri}. In this case, the interfacial partial debonding initiates from the local x_1-direction and propagates progressively toward the second principal direction. Figure 4 illustrates the cylindrical partial fiber debonding. The debonded arc microcrack length can be characterized by the debonding angle $\alpha_{12}^{(2)}$. Accordingly, the debonding angle for Mode 2 reads

$$\alpha_{12}^{(2)} = \sin^{-1}\sqrt{\frac{\overline{\sigma}_{11}^{(1)} - \sigma_{\text{cri}}}{\overline{\sigma}_{11}^{(1)} - \overline{\sigma}_{22}^{(1)}}} \tag{10}$$

Mode 3: $\overline{\sigma}_{11}^{(1)} \geq \overline{\sigma}_{22}^{(1)} \geq \sigma_{\text{cri}}$ All two local principal stresses exceed the critical debonding strength, which indicates that the normal stress at any point on a cylindrical fiber surface is greater than the critical strength. Therefore, the entire fiber/matrix interface is debonded; and Mode 3 implies the formation of a cylindrical void.

It is noted that, for all of the above three damage modes, the range of partial debonding angles is between 0 and $\pi/2$. The lower and upper bounds of the interfacial fiber debonding angles in a certain principal direction correspond to the perfect bonding and total debonding, respectively, in that direction.

The Equivalent Inclusion Method

The progressive interfacial fiber debonding process gradually reduces the overall elastic moduli of composites due to the loss in tensile load-carrying capacity in the randomly dispersed unidirectional cylindrical fibers. The progressive *partial* loss of load-transfer capacity in the debonding directions is adopted in the present framework to simulate the interfacial damage mechanism of fiber-reinforced composites. Therefore, it is needed to establish the relationship between the progressive interfacial debonding angle and the loss of tensile load-transfer capacity, which is manifested by a reduction in elastic moduli. Furthermore, from the preceding various debonding modes along the principal coordinates, the debonding profiles are dependent on the transverse axial directions so that load-transfer capabilities are different along distinct axes. Hence, the interfacially damaged isotropic cylindrical fibers behave *orthotropically*.

Zhao and Weng (2002) employed the volume of inclusion directly beneath the interfacial microcracks as a measure of the interfacial debonding damage, which can easily be obtained in 2-D cases. Under general two-dimensional loading conditions, there exist three distinct interfacial damage parameters D_i ($i = 1, 2$; $D_3 = 0$) serving as the measures of elastic stiffness reduction in certain directions. In accordance with the three different interfacial debonding modes presented in the preceding section, the local interfacial damage parameters can be developed as follows.

Mode 1 For perfectly bonded cylindrical fibers, all interfacial damage parameters are equal to 0 (i.e., no interfacial damage development). Accordingly, we have

$$D_1^{(1)} = D_2^{(1)} = D_3^{(1)} = 0 \tag{11}$$

Mode 2 Interfacial partial debonding starts normal to the first principal stress axis (the x_1-direction). In terms of the interfacial progressive debonding angle $\alpha_{12}^{(2)}$, the projection of the damaged arc microcrack in the x_1-direction (A_1 in Fig. 4a) is shown. The local interfacial damage parameter with respect to the x_1-direction is defined as

$$D_1^{(2)} \equiv \frac{\mathrm{A}_1}{\mathrm{A}} = \frac{2}{\pi}\left(\alpha_{12}^{(2)} + \sin\alpha_{12}^{(2)}\cos\alpha_{12}^{(2)}\right) \tag{12}$$

The other damage parameter in the x_2-direction (A_2 in Fig. 4b) can be derived as

$$D_2^{(2)} \equiv \frac{\mathrm{A}_2}{\mathrm{A}} = \frac{2}{\pi}\left(\alpha_{12}^{(2)} - \sin\alpha_{12}^{(2)}\cos\alpha_{12}^{(2)}\right), D_3^{(2)} = 0 \tag{13}$$

where A is the cross-sectional area of the circular fiber.

Mode 3 For the total interfacial fiber debonding mode, the cylindrical fibers cannot transfer any transverse tensile loads, which can be expressed as

$$D_1^{(3)} = D_2^{(3)} = 1, D_3^{(3)} = 0 \tag{14}$$

The interfacial damage vectors introduced in the above three distinct interfacial debonding modes reflect the reduction in tensile load-transfer capability in certain directions. Their numerical values vary from 0 to 1. A higher value of damage parameter signifies a more significant level of reduction in the local tensile load-transfer capacity.

As a result, the corresponding fourth-rank equivalent elastic stiffness tensor of cylindrical fibers follows the form

$$C_{ijkl}^{(\beta)} = \lambda_{\mathrm{IK}}^{(\beta)} \delta_{ij}\delta_{kl} + \mu_{\mathrm{IJ}}^{(\beta)} \left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right), \quad i,j,k,l = 1,2,3 \tag{15}$$

where

$$\begin{aligned} \lambda_{\mathrm{IK}}^{(\beta)} &= \lambda^{(1)} \left(1 - D_{\mathrm{I}}^{(\beta)}\right)\left(1 - D_{\mathrm{K}}^{(\beta)}\right), \quad \beta = 1,2,3 \\ \mu_{\mathrm{IJ}}^{(\beta)} &= \mu^{(1)} \left(1 - D_{\mathrm{I}}^{(\beta)}\right)\left(1 - D_{\mathrm{J}}^{(\beta)}\right), \quad \beta = 1,2,3 \end{aligned} \tag{16}$$

with $\lambda^{(1)}$, $\mu^{(1)}$ denoting the isotropic elastic Lamé constants of the original (perfectly bonded) fibers. Here, we follow Mura's (1987) tensorial indicial notation; i.e., the repeated lowercase indices are summed up from 1 to 3, whereas the uppercase indices take on the same numbers as the corresponding lowercase ones but are not summed up. This indicial expression facilitates the subsequent derivations and computations.

By using the elastic-damage equivalent inclusion treatment, all partially debonded fibers are replaced by perfectly bonded fibers with the aforementioned equivalent orthotropic elastic stiffness tensors. Therefore, multiphase micromechanical probabilistic approaches can be established to characterize the progressive interfacial debonding processes of fiber composites.

Evolution of Volume Fractions of Debonded Fibers

The present micromechanical interfacial damage model of fiber-reinforced composites assumes all elastic unidirectional cylindrical fibers to be randomly dispersed in the matrix, but not to initiate partial interfacial fiber debonding simultaneously. The evolution process concerning the volume fractions of partially debonded fibers depends on the external loading history, ranging from zero (no debonded fiber) to the total volume fraction of fibers ϕ^{Total} (all fibers debonded).

With increasing applied external transverse loads, progressive transverse interfacial fiber debonding may develop and transform from one damage mode to another. For example, under the transverse biaxial loading condition with different magnitudes of applied loads in the two transverse axes, the damage Mode 2 will initiate when the first local principal radial normal stress reaches the critical interfacial debonding strength. Further increasing the applied transverse loads can activate the damage Mode 3 when the second local principal radial normal stress also reaches the critical debonding strength. At this stage, there exist three types of

damage modes simultaneously, including Mode 1 with perfectly bonded fibers and the other two modes with partially and completely debonded fibers as explained previously. To quantify the volume fraction evolution of progressively debonded fibers, we denote $\phi^{(\beta)}$ ($\beta = 1, 2, 3$) as the volume fraction for a certain interfacial damage mode of fibers at the current loading stage. The volume fractions of the three various local interfacial damage modes $\phi^{(\beta)}(\beta = 1, 2, 3)$ are expressed as follows to characterize the evolution of interfacial partial fiber debonding and the transition among the three debonding modes:

$$\begin{aligned} \phi^{(3)} &= \phi^{\text{Total}} P_2 \\ \phi^{(2)} &= \phi^{\text{Total}} [P_1 - P_2] \\ \phi^{(1)} &= \phi^{\text{Total}} [1 - P_1] \end{aligned} \tag{17}$$

where ϕ^{Total} is the total volume fraction of all fibers in the composites. The probabilistic function P_i ($i = 1, 2$) controls the evolution of the volume fraction of perfectly bonded fibers in the *ith* principal stress direction and increases as the local principal stress $\overline{\sigma}_{ii}^{(1)}$ increases. For simplicity, a two-parameter Weibull's probability distribution function is adopted in this work to describe the evolutionary process (Tohgo and Weng 1994; Sun et al. 2003a; Liu et al. 2004, 2006). We write

$$P_i = \begin{cases} 1 - \exp\left[-\left(\frac{\overline{\sigma}_{ii}^{(1)} - \sigma_{\text{cri}}}{S_0} \right)^M \right] & , \overline{\sigma}_{ii}^{(1)} \geq \sigma_{\text{cri}} \\ 0 & , \overline{\sigma}_{ii}^{(1)} < \sigma_{\text{cri}} \end{cases} . \quad (i = 1, 2) \tag{18}$$

Here, Weibull's parameters M and S_0 denote the evolution rate of the volume fractions of debonded fibers and the average interfacial strength, respectively. The parameter σ_{cri} represents the critical local bonding strength.

Equation 17 is established on the basis of $\overline{\sigma}_{11}^{(1)} \geq \overline{\sigma}_{22}^{(1)}$, indicating that once the local principal radial normal stress along the second direction ($\overline{\sigma}_{22}^{(1)}$) reaches the critical strength, the first local principal radial normal stresses must be greater than the critical strength. Therefore, the second principal radial normal stress $\overline{\sigma}_{22}^{(1)}$ controls the Mode 3 debonding. When the first local principal radial normal stress $\overline{\sigma}_{11}^{(1)}$ in a certain fiber exceeds the critical bonding strength σ_{cri}, that fiber could be either in Mode 2 ($\overline{\sigma}_{22}^{(1)} < \sigma_{\text{cri}}$) or Mode 3 ($\overline{\sigma}_{22}^{(1)} \geq \sigma_{\text{cri}}$). Therefore, the net Mode 2 evolution rate is expected to be controlled by $(P_1 - P_2)$ while the remaining Mode 1 (undamaged) fibers are decreased by the percentage of $-P_1$.

In the case of transverse tensile loading/unloading condition, no further interfacial fiber debonding would occur during the tensile unloading. On the other hand, in the case of the fully reversed cyclic loading condition, all interfacial microcracks would be temporarily closed and deactivated; i.e., $\phi^{(2)} = \phi^{(3)} = 0$, under the reverse compressive loading. It is also recognized that local shear stresses acting on the cylindrical fiber interfaces could play a role in the interfacial debonding process of composites. However, large local interfacial shear stresses could potentially change the principal stress directions such that they are sufficiently different from

the principal axes of a cylindrical fiber, thus making the equivalent orthotropic inclusion treatment difficult. Future research is desired to accommodate the effects of combined local radial normal and shear stresses, possibly by introducing an energy-based mixed-mode interfacial debonding criterion and a new, generalized treatment of equivalent anisotropic inclusion that could handle arbitrarily changing principal stress directions.

Effective Elastic Moduli of Debonded Composites

Once the equivalent stiffness tensor in Eq. 15 for partially debonded fibers is established and the volume fractions associated with different phases of partially debonded fibers are categorized as proposed above, the effective elastic-damage moduli of fiber-reinforced composites can be estimated through the micromechanics approach and homogenization. Here, the first-order micromechanical approximation of Ju and Chen (1994a, b, c), Ju and Zhang (1998a), and Ko and Ju (2012, 2013b) is applied to obtain the effective elastic moduli of composites containing randomly dispersed aligned cylindrical fibers. It is noted that Ju and Chen's first-order elastic approximation method is actually equivalent to the Hashin–Shtrikman bounds (Hashin and Shtrikman 1962) and the Mori–Tanaka method (Mori and Tanaka 1973) for multiphase elastic composites. The explicit expression of the effective elastic moduli of the composites can be expressed as

$$\overline{C}_{ijkl} = C^{(0)}_{ijmn}\left[I_{mnkl} + \left(Y^{-1}_{mnkl} - S_{mnkl}\right)^{-1}\right], \quad i,j,k,l,m,n = 1,2,3 \tag{19}$$

with

$$Y_{mnkl} = \sum_{\beta=1}^{3} \phi^{(\beta)}\left[S_{mnkl} + \left(C^{(\beta)}_{mnpq} - C^{(0)}_{mnpq}\right)^{-1} C^{(0)}_{pqkl}\right]^{-1}, \quad k,l,m,n,p,q = 1,2,3 \tag{20}$$

where $C^{(\beta)}_{ijkl}$ is the equivalent stiffness tensor of the β-phase particles given by the Eq. 15. Moreover, for the overall *orthotropic* materials,

$$\{\overline{\sigma}_i\} = \left[\overline{C}_{ij}\right]\left\{\overline{\varepsilon}^e_j\right\}, \quad i,j = 1 \text{ to } 6 \tag{21}$$

where

$$\{\overline{\sigma}_i\} = \{\overline{\sigma}_{11}, \overline{\sigma}_{22}, \overline{\sigma}_{33}, \overline{\sigma}_{12}, \overline{\sigma}_{23}, \overline{\sigma}_{31}\}^{\mathbf{T}}, \quad i = 1 \text{ to } 6 \tag{22}$$

$$\left\{\overline{\varepsilon}^e_j\right\} = \left\{\overline{\varepsilon}^e_{11}, \overline{\varepsilon}^e_{22}, \overline{\varepsilon}^e_{33}, 2\overline{\varepsilon}^e_{12}, 2\overline{\varepsilon}^e_{23}, 2\overline{\varepsilon}^e_{31}\right\}^{\mathbf{T}}, \quad j = 1 \text{ to } 6 \tag{23}$$

Therefore, there are only 9 independent elastic constants for the overall *orthotropic* composite. Since the composite consists of a ductile (elastoplastic)

matrix and *randomly* located yet unidirectionally aligned cylindrical fibers, the plane-strain condition governs here. Furthermore,

$$\overline{C}_{ijkl} = \mathrm{C}_{\mathrm{IK}}^{(1)}\delta_{ij}\delta_{kl} + \mathrm{C}_{\mathrm{IJ}}^{(2)}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right), \quad i,j,k,l = 1,2,3 \tag{24}$$

Moreover, for the overall *orthotropic* materials,

$$\left[\overline{C}_{ij}\right]_{6\times 6} = \begin{bmatrix} \overline{C}_{11} & \overline{C}_{12} & \overline{C}_{13} & 0 & 0 & 0 \\ \overline{C}_{12} & \overline{C}_{22} & \overline{C}_{23} & 0 & 0 & 0 \\ \overline{C}_{13} & \overline{C}_{23} & \overline{C}_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & \overline{C}_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \overline{C}_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & \overline{C}_{66} \end{bmatrix}_{6\times 6} \tag{25}$$

In other words,

$$\left[\overline{C}_{ij}\right]_{6\times 6} = \begin{bmatrix} \mathrm{C}_{11}^{(1)} + 2\mathrm{C}_{11}^{(2)} & \mathrm{C}_{12}^{(1)} & \mathrm{C}_{13}^{(1)} & 0 & 0 & 0 \\ \mathrm{C}_{21}^{(1)} & \mathrm{C}_{22}^{(1)} + 2\mathrm{C}_{22}^{(2)} & \mathrm{C}_{23}^{(1)} & 0 & 0 & 0 \\ \mathrm{C}_{31}^{(1)} & \mathrm{C}_{32}^{(1)} & \mathrm{C}_{33}^{(1)} + 2\mathrm{C}_{33}^{(2)} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{C}_{12}^{(2)} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathrm{C}_{23}^{(2)} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathrm{C}_{31}^{(2)} \end{bmatrix}_{6\times 6} \tag{26}$$

Overall effective elastic moduli associated with such orthotropic materials take the form

$$\overline{E}_{11} = \frac{\overline{C}_{33}\overline{C}_{12}^2 - 2\overline{C}_{13}\overline{C}_{23}\overline{C}_{12} + \overline{C}_{13}^2\overline{C}_{22} + \overline{C}_{11}\left(\overline{C}_{23}^2 - \overline{C}_{22}\overline{C}_{33}\right)}{\overline{C}_{23}^2 - \overline{C}_{22}\overline{C}_{33}} \tag{27}$$

$$\overline{E}_{22} = \frac{\overline{C}_{33}\overline{C}_{12}^2 - 2\overline{C}_{13}\overline{C}_{23}\overline{C}_{12} + \overline{C}_{13}^2\overline{C}_{22} + \overline{C}_{11}\left(\overline{C}_{23}^2 - \overline{C}_{22}\overline{C}_{33}\right)}{\overline{C}_{13}^2 - \overline{C}_{11}\overline{C}_{33}} \tag{28}$$

$$\overline{E}_{33} = \frac{\overline{C}_{22}\overline{C}_{13}^2 - 2\overline{C}_{12}\overline{C}_{23}\overline{C}_{13} + \overline{C}_{23}^2\overline{C}_{11}}{\overline{C}_{12}^2 - \overline{C}_{11}\overline{C}_{22}} + \overline{C}_{33} \tag{29}$$

$$\overline{\nu}_{12} = \frac{\overline{C}_{12}\overline{C}_{33} - \overline{C}_{13}\overline{C}_{23}}{\overline{C}_{22}\overline{C}_{33} - \overline{C}_{23}^2}, \overline{\nu}_{21} = \frac{\overline{C}_{12}\overline{C}_{33} - \overline{C}_{13}\overline{C}_{23}}{\overline{C}_{11}\overline{C}_{33} - \overline{C}_{13}^2}, \overline{\nu}_{13} = \frac{\overline{C}_{13}\overline{C}_{22} - \overline{C}_{12}\overline{C}_{23}}{\overline{C}_{22}\overline{C}_{33} - \overline{C}_{23}^2} \tag{30}$$

$$\overline{\nu}_{31} = \frac{\overline{C}_{13}\overline{C}_{22} - \overline{C}_{12}\overline{C}_{23}}{\overline{C}_{11}\overline{C}_{22} - \overline{C}_{12}^2}, \overline{\nu}_{23} = \frac{\overline{C}_{11}\overline{C}_{23} - \overline{C}_{12}\overline{C}_{13}}{\overline{C}_{11}\overline{C}_{33} - \overline{C}_{13}^2}, \overline{\nu}_{32} = \frac{\overline{C}_{11}\overline{C}_{23} - \overline{C}_{12}\overline{C}_{13}}{\overline{C}_{11}\overline{C}_{22} - \overline{C}_{12}^2} \tag{31}$$

$$\overline{\mu}_{12} = \overline{C}_{44}, \overline{\mu}_{23} = \overline{C}_{55}, \overline{\mu}_{31} = \overline{C}_{66} \tag{32}$$

$$\frac{\overline{\nu}_{12}}{\overline{E}_{11}} = \frac{\overline{\nu}_{21}}{\overline{E}_{22}}, \frac{\overline{\nu}_{23}}{\overline{E}_{22}} = \frac{\overline{\nu}_{32}}{\overline{E}_{33}}, \frac{\overline{\nu}_{13}}{\overline{E}_{11}} = \frac{\overline{\nu}_{31}}{\overline{E}_{33}} \tag{33}$$

Therefore, there are only 9 independent elastic constants for the overall *orthotropic* composite materials. Here, effective Poisson's ratio $\overline{\nu}_{ij}$ is defined as the ratio of strain shrinkage in the *j*th direction over the strain extension in the *i*th direction when a tensile stress is applied in the *i*th direction.

In addition, the effective elastic compliance matrix becomes

$$\left[\overline{D}_{ij}\right]_{6\times 6} = \left[\overline{C}_{ij}\right]^{-1}{}_{6\times 6} = \begin{bmatrix} \frac{1}{\overline{E}_{11}} & \frac{-\overline{\nu}_{21}}{\overline{E}_{22}} & \frac{-\overline{\nu}_{31}}{\overline{E}_{33}} & 0 & 0 & 0 \\ \frac{-\overline{\nu}_{12}}{\overline{E}_{11}} & \frac{1}{\overline{E}_{22}} & \frac{-\overline{\nu}_{32}}{\overline{E}_{33}} & 0 & 0 & 0 \\ \frac{-\overline{\nu}_{13}}{\overline{E}_{11}} & \frac{-\overline{\nu}_{23}}{\overline{E}_{22}} & \frac{1}{\overline{E}_{33}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\overline{\mu}_{12}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\overline{\mu}_{23}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\overline{\mu}_{31}} \end{bmatrix}_{6\times 6} \tag{34}$$

The stress component $\overline{\sigma}_{33}$ along the longitudinal fiber direction thus reads

$$\overline{\sigma}_{33} = \eta_1\, \overline{\sigma}_{11} + \eta_2\, \overline{\sigma}_{22} \tag{35}$$

where

$$\eta_1 = \frac{\overline{C}_{13}\, \overline{C}_{22} - \overline{C}_{12}\, \overline{C}_{23}}{\Xi}, \eta_2 = \frac{\overline{C}_{11}\, \overline{C}_{23} - \overline{C}_{12}\, \overline{C}_{13}}{\Xi}, \Xi = \overline{C}_{11}\, \overline{C}_{22} - \overline{C}_{12}^2 \tag{36}$$

The Ensemble-Averaged Effective Yield Function

The ensemble-area averaging homogenization procedure is applied here to estimate the effective elastoplastic-damage behavior of composites, accounting for the effects of initial yielding and plastic hardening of the matrix. The averaging homogenization procedure is usually performed within a mesoscopic representative volume/area element; see, e.g., Nemat-Nasser and Hori (1993). At any local matrix material point $\mathbf{x}$, the microscopic stress $\boldsymbol{\sigma}(\mathbf{x})$ is assumed to satisfy the von Mises J_2-yield criterion

$$F\left(\boldsymbol{\sigma}\, \overline{e}_m^p\right) \;=\; \sqrt{\boldsymbol{\sigma} : \mathbf{I}_{\mathrm{d}} : \boldsymbol{\sigma}} - K\left(\overline{e}_m^p\right) \;\leq\; 0 \tag{37}$$

where $\overline{e}_m^p$ and $K\left(\overline{e}_m^p\right)$ are the equivalent plastic strain and the isotropic hardening of the matrix-only material, respectively. Moreover, $\mathbf{I}_\mathrm{d} \equiv \mathbf{I} - \frac{1}{3}\mathbf{1}\otimes\mathbf{1}$ denotes the deviatoric part of the fourth-rank identity tensor $\mathbf{I}$.

Following Ju and Chen (1994c), Ju and Tseng (1996, 1997), Ju and Sun (2001), Sun and Ju (2001), and Ju and Zhang (2001), we denote $H(\mathbf{x}|g) = \boldsymbol{\sigma}(\mathbf{x}|g) : \mathbf{I}_\mathrm{d} : \boldsymbol{\sigma}(\mathbf{x}|g)$ as the square of the "current stress norm" at a local point $\mathbf{x}$, which contributes to the initial yield criterion of composite for a given fiber configuration g (assembly). Further, $\langle H\rangle_m(\mathbf{x})$ defines the ensemble average of $H(\mathbf{x}|g)$ over all possible realization for a matrix point $\mathbf{x}$:

$$\langle H\rangle_m(\mathbf{x}) \;\cong\; H^\mathrm{o} + \int_g \{H(\mathbf{x}|g) - H^\mathrm{o}\}\, P(g)\, \mathrm{d}g \tag{38}$$

where $P(g)$ is the probability density function for finding a fiber configuration g in the composite and $H^\mathrm{o} = \boldsymbol{\sigma}^\mathrm{o} : \mathbf{I}_\mathrm{d} : \boldsymbol{\sigma}^\mathrm{o}$ is the square of the far-field stress norm applied on the ductile composite.

For the complete second-order formulation, any two-fiber would interact first, and then the matrix point collects the perturbations based on the results of near-field fiber interactions. In the absence of exact solution for the near-field fiber interactions problem, the first-order approximation provides a simple way to account for the perturbations on a matrix point from the fibers. In the first-order approximation approach, at a matrix point $\mathbf{x}_m$, the surrounding fibers are treated as isolated sources of perturbations. A local matrix material point simply collects the perturbations from all cylindrical fibers one by one. In the absence of exact solutions for statistical higher-order near-field fiber interaction problems, the first-order approximation provides a feasible way to account for the randomly located fiber-induced perturbations upon a local matrix point. Such first-order approximation is valid when the fiber volume fraction is moderate.

Hence, the expression of $\langle H\rangle_m(\mathbf{x})$ for the composites with three phases of fibers (corresponding to the three interfacial debonding modes) under the (generalized) plane-strain condition can be approximately obtained as

$$\begin{aligned}\langle H\rangle_m(\mathbf{x}) \;\cong\; H^\mathrm{o} &+ \int_{|\mathbf{x}-\mathbf{x}^{(1)}|>a} \left\{H\left(\mathbf{x}|\mathbf{x}^{(1)}\right) - H^\mathrm{o}\right\} P\left(\mathbf{x}^{(1)}\right)\, \mathrm{d}\mathbf{x}^{(1)} \\ &+ \int_{|\mathbf{x}-\mathbf{x}^{(2)}|>a} \left\{H\left(\mathbf{x}|\mathbf{x}^{(2)}\right) - H^\mathrm{o}\right\} P\left(\mathbf{x}^{(2)}\right)\, \mathrm{d}\mathbf{x}^{(2)} \\ &+ \int_{|\mathbf{x}-\mathbf{x}^{(3)}|>a} \left\{H\left(\mathbf{x}|\mathbf{x}^{(3)}\right) - H^\mathrm{o}\right\} P\left(\mathbf{x}^{(3)}\right)\, \mathrm{d}\mathbf{x}^{(3)}\end{aligned} \tag{39}$$

where $|\mathbf{x}-\mathbf{x}^{(1)}| > a$, $|\mathbf{x}-\mathbf{x}^{(2)}| > a$, and $|\mathbf{x}-\mathbf{x}^{(3)}| > a$ are the "exclusion zone" of $\mathbf{x}$ for the center location $\mathbf{x}^{(\eta)}$ of a fiber in the probability space, which is identical to the shape and size of a fiber. In addition, $\mathbf{x}^{(1)}$ represents a perfectly bonded fiber in

the probability space. Besides, $\mathbf{x}^{(2)}$ and $\mathbf{x}^{(3)}$ represent a partially debonded fiber and a completely debonded fiber in the probability space, respectively. The probabilistic exclusion zone states that $\mathbf{x}^{(\eta)}$ cannot be located within the exclusion zone domain because $\mathbf{x}$ must be located within the matrix phase. $P(\mathbf{x}^{(\eta)})$ is assumed to be statistically homogeneous, isotropic, and uniform. Further, $H(\mathbf{x}|\mathbf{x}^{(1)})$, $H(\mathbf{x}|\mathbf{x}^{(2)})$, and $H(\mathbf{x}|\mathbf{x}^{(3)})$ are the stress-norm collection of contributions from the undamaged fibers, the partially debonded fibers, and the completely debonded fibers, respectively. They can be estimated from Eshelby's micromechanics framework (e.g., Mura 1987; Ju and Sun 2001; Sun and Ju 2001) as

$$H^{(\beta)} = \boldsymbol{\sigma}^{(\beta)}(\mathbf{x}) : \mathbf{I}_{\mathrm{d}} : \boldsymbol{\sigma}^{(\beta)}(\mathbf{x}), \quad \beta = 1, 2, 3 \tag{40}$$

where the local stress tensor in the matrix due to the β-phase fiber is

$$\boldsymbol{\sigma}^{(\beta)}(\mathbf{x}) = \boldsymbol{\sigma}^{\mathrm{o}} + \mathbf{C}^{(0)} \bullet \overline{\mathbf{G}}(\mathbf{x}) : \boldsymbol{\varepsilon}_{*}^{(\beta)\mathrm{o}}, \quad \beta = 1, 2, 3 \tag{41}$$

and the "first-order" eigenstrain tensor $\boldsymbol{\varepsilon}_{*}^{(\beta)\mathrm{o}}$ in the β-phase fiber is

$$\left(\varepsilon_{*}^{(\beta)\mathrm{o}}\right)_{ij} = -\left[S_{ijmn} + \left(C_{ijkl}^{(\beta)} - C_{ijkl}^{(0)}\right)^{-1} C_{klmn}^{(0)}\right]^{-1} (\varepsilon^{\mathrm{o}})_{mn}, \quad i, j, k, l, m, n = 1, 2, \beta = 1, 2, 3 \tag{42}$$

in which $(\varepsilon^{\mathrm{o}})_{kl} = C_{ijkl}^{(0)-1}(\sigma^{\mathrm{o}})_{ij}$. Therefore, the perturbed stress for any matrix point $\mathbf{x}$ due to a typical isolated β-phase inhomogeneity centered at $\mathbf{x}^{(\beta)}$ takes the form

$$\boldsymbol{\sigma}'\left(\mathbf{x}|\mathbf{x}^{(\beta)}\right) = \left[\mathbf{C}^{(0)} \bullet \overline{\mathbf{G}}\left(\mathbf{x} - \mathbf{x}^{(\eta)}\right)\right] : \boldsymbol{\varepsilon}_{*}^{(\beta)\mathrm{o}} \tag{43}$$

$$\overline{\mathbf{G}}\left(\mathbf{x} - \mathbf{x}^{(\beta)}\right) \equiv \int_{\Omega^{(\beta)}} \mathbf{G}(\mathbf{x} - \mathbf{x}')\, \mathrm{d}\mathbf{x}' \tag{44}$$

for $\mathbf{x} \notin \Omega^{(\beta)}$ in which $\Omega^{(\beta)}$ is the single inhomogeneity domain centered at $\mathbf{x}^{(\beta)}$ in the β-phase. Alternatively, we can write

$$\overline{\mathbf{G}}(\mathbf{r}_{\beta}) = \frac{1}{8(1-\nu_0)} \left(\rho_{\beta}^2\, \mathbf{H}^1 + \frac{\rho_{\beta}^4}{2}\, \mathbf{H}^2\right) \tag{45}$$

The components of $\mathbf{H}^1$ and $\mathbf{H}^2$ are given by

$$\mathbf{H}_{ijkl}^1(\mathbf{r}_{\beta}) \equiv 2\mathbf{F}_{ijkl}(-8, 2\nu_0, 2, 2 - 4\nu_0, -1 + 2\nu_0, 1 - 2\nu_0) \tag{46}$$

$$\mathbf{H}_{ijkl}^2(\mathbf{r}_{\beta}) \equiv 2\mathbf{F}_{ijkl}(24, -4, -4, -4, 1, 1) \tag{47}$$

where $\rho_{\beta} = a/r_{\beta}$ and a is the radius of a cylindrical fiber or void.

The components of the fourth-rank tensor $\mathbf{F}$ – which depends on six scalar quantities B_1,B_2,B_3,B_4,B_5,B_6 – are defined by

$$\begin{aligned} F_{ijkl}(B_m) \;\equiv\; & B_1 n_i n_j n_k n_l + B_2\left(\delta_{ik} n_j n_l + \delta_{il} n_j n_k + \delta_{jk} n_i n_l + \delta_{jl} n_i n_k\right) \\ & + B_3 \delta_{ij} n_k n_l + B_4 \delta_{kl} n_i n_j + B_5 \delta_{ij} \delta_{kl} + B_6\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) \end{aligned} \tag{48}$$

with the unit normal vector $\mathbf{n} \equiv \mathbf{r}_\beta / r_\beta$ and index $m = 1$ to 6.

For simplicity, it is assumed that all fibers are uniformly randomly distributed in the matrix. Nevertheless, nonuniform probabilistic fiber distributions $P(\mathbf{x}^{(\eta)})$ within the matrix can be easily accommodated within the proposed framework. Accordingly, $P(\mathbf{x}^{(\beta)})$ can be expressed as $N^{(\beta)}/A$ ($\beta = 1, 2, 3$) where $N^{(\beta)}$ is the total number of the β-phase fibers uniformly dispersed in a representative area element A. After a series of lengthy yet straightforward derivations, the ensemble-averaged $\langle H \rangle_m$ can be evaluated as

$$\langle H \rangle_m = \boldsymbol{\sigma}^{\mathrm{o}} : \mathbf{T} : \boldsymbol{\sigma}^{\mathrm{o}} \tag{49}$$

where the components of the fourth-rank tensor $\mathbf{T}$ take the form

$$T_{ijkl} = \mathrm{T}^{(1)}_{\mathrm{IK}} \delta_{ij}\delta_{kl} + \mathrm{T}^{(2)}_{\mathrm{IJ}}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right), \quad i,j,k,l = 1,2 \tag{50}$$

with

$$\begin{aligned} \mathrm{T}^{(2)}_{12} = & \frac{1}{8}\left[4 + \frac{\left(2 + 5\hat{\mathrm{a}} + 5\hat{\mathrm{b}} + 3\hat{\mathrm{c}}\right)\phi^{(1)}}{\beta_1^2} + \frac{\left(2 + 5\hat{\mathrm{a}} + 5\hat{\mathrm{b}} + 3\hat{\mathrm{c}}\right)\phi^{(3)}}{\beta_2^2}\right. \\ & \left. + \frac{\left(2 + 5\hat{\mathrm{a}} + 5\hat{\mathrm{b}} + 3\hat{\mathrm{c}}\right)\mu_0{}^2 \overline{\mathrm{B}}^2_{1211\mathrm{d}}\phi^{(2)}}{\left(-1 + \nu_0\right)^2}\right] \end{aligned} \tag{51}$$

$$\mathrm{T}^{(1)}_{12} = \frac{1}{2}\left(\hat{\mathrm{c}} + \mathrm{C}_{\mathrm{q1}} + \mathrm{C}_{\mathrm{q2}} + \mathrm{C}_{\mathrm{q211}}\right) \tag{52}$$

$$\begin{aligned} \mathrm{T}^{(2)}_{11} = & \frac{1}{8}\left[4 + \frac{\left(2 + 5\hat{\mathrm{a}} + 5\hat{\mathrm{b}} + 3\hat{\mathrm{c}}\right)\phi^{(1)}}{\beta_1^2} + \frac{\left(2 + 5\hat{\mathrm{a}} + 5\hat{\mathrm{b}} + 3\hat{\mathrm{c}}\right)\phi^{(3)}}{\beta_2^2}\right. \\ & \left. + \frac{\left(2 + 5\hat{\mathrm{a}} + 5\hat{\mathrm{b}} + 3\hat{\mathrm{c}}\right)\mu_0{}^2 \overline{\mathrm{B}}^2_{1111\mathrm{d}}\phi^{(2)}}{\left(-1 + \nu_0\right)^2}\right] \end{aligned} \tag{53}$$

$$\begin{aligned} \mathrm{T}^{(1)}_{11} = & \hat{\mathrm{a}} + \mathrm{A}_{\mathrm{q1}} + \mathrm{A}_{\mathrm{q2}} + \mathrm{A}_{\mathrm{q211}} + \frac{1}{4}\left\{ - \frac{\left(2 + 5\hat{\mathrm{a}} + 5\hat{\mathrm{b}} + 3\hat{\mathrm{c}}\right)\phi^{(1)}}{\beta_1^2}\right. \\ & + \frac{1}{\beta_2^2\left(-1 + \nu_0\right)^2}\left[- \left(2 + 5\hat{\mathrm{a}} + 5\hat{\mathrm{b}} + 3\hat{\mathrm{c}}\right)\right. \\ & \left.\left.\left(-1 + \nu_0\right)^2 \phi^{(3)} + \beta_2^2\left(-4\left(-1 + \nu_{\mathrm{o}}\right)^2 - \left(2 + 5\hat{\mathrm{a}} + 5\hat{\mathrm{b}} + 3\hat{\mathrm{c}}\right)\mu_0^2 \overline{\mathrm{B}}^2_{1111\mathrm{d}}\phi^{(2)}\right)\right]\right\} \end{aligned} \tag{54}$$

$$\mathrm{T}_{22}^{(2)} = \frac{1}{8}\left[4 + \frac{\left(2+5\hat{a}+5\hat{b}+3\hat{c}\right)\phi^{(1)}}{\beta_1^2} + \frac{\left(2+5\hat{a}+5\hat{b}+3\hat{c}\right)\phi^{(3)}}{\beta_2^2} + \frac{\left(2+5\hat{a}+5\hat{b}+3\hat{c}\right)\mu_0{}^2\overline{\mathrm{B}}_{2211\mathrm{d}}^2\phi^{(2)}}{(-1+\nu_0)^2}\right] \tag{55}$$

$$\mathrm{T}_{22}^{(1)} = \hat{\mathrm{b}} + \mathrm{B}_{\mathrm{q}1} + \mathrm{B}_{\mathrm{q}2} + \mathrm{B}_{\mathrm{q}211} + \frac{1}{4}\left\{-\frac{\left(2+5\hat{a}+5\hat{b}+3\hat{c}\right)\phi^{(1)}}{\beta_1^2} + \frac{1}{\beta_2^2(-1+\nu_0)^2}\left[-\left(2+5\hat{a}+5\hat{b}+3\hat{c}\right)(-1+\nu_0)^2\phi^{(3)} + \beta_2^2\left(-4(-1+\nu_0)^2 - \left(2+5\hat{a}+5\hat{b}+3\hat{c}\right)\mu_0^2\overline{\mathrm{B}}_{2211\mathrm{d}}^2\phi^{(2)}\right)\right]\right\} \tag{56}$$

Here $\phi^{(1)}$, $\phi^{(2)}$, and $\phi^{(3)}$ are the volume fractions of the perfectly bonded, the partially debonded, and the completely debonded fibers, respectively; see the section "Detailed Derivation for Tensor T in Eq. 11" for further details.

The general relationship between the applied far-field stress $\boldsymbol{\sigma}^{\mathrm{o}}$ and the macroscopic (ensemble-area averaged) stress $\overline{\boldsymbol{\sigma}}$ is given by (cf. Ju and Chen 1994c)

$$\boldsymbol{\sigma}^{\mathrm{o}} = \mathbf{P} \;:\; \overline{\boldsymbol{\sigma}} \tag{57}$$

where the fourth-rank tensor $\mathbf{P}$ reads

$$\mathbf{P} = \left\{\mathbf{C}^{(0)} \bullet \left[\mathbf{I} + (\mathbf{I}-\mathbf{S}) \bullet \mathbf{Y}\right] \bullet \mathbf{C}^{(0)^{-1}}\right\}^{-1} \tag{58}$$

with

$$\mathbf{Y} = \sum_{\beta=1}^{3} \phi^{(\beta)} \left[\mathbf{S} + \left(\mathbf{C}^{(\beta)} - \mathbf{C}^{(0)}\right)^{-1} \bullet \mathbf{C}^{(0)}\right]^{-1} \tag{59}$$

We refer to the section "Detailed Derivation for Tensor P in Eq. 58" for detailed derivation of the tensor $\mathbf{P}$. Combining Eqs. 49 and 57 leads to an alternative expression of the ensemble-averaged square of the current stress norm as

$$\langle H \rangle_m(\mathbf{x}) = \overline{\boldsymbol{\sigma}} : \overline{\mathbf{T}} : \overline{\boldsymbol{\sigma}} \tag{60}$$

where the fourth-rank tensor $\overline{\mathbf{T}} \equiv \mathbf{P}^{\mathbf{T}} : \mathbf{T} : \mathbf{P}$. Furthermore, Eq. 60 will reduce to the classical J_2-invariant if $\phi^{(1)} = \phi^{(2)} = \phi^{(3)} = 0$, i.e., the matrix-only material. In the indicial form, we write

$$\langle H \rangle_m(\mathbf{x}) = \overline{\sigma}_{ij} : \overline{T}_{ijkl} : \overline{\sigma}_{kl}, i,j,k,l = 1,2 \tag{61}$$

where

$$\begin{aligned}\overline{T}_{ijkl} &= P_{mnij}T_{mnpq}P_{pqkl} \\ &= \overline{\mathrm{T}}^{(1)}_{\mathrm{IK}}\delta_{ij}\delta_{kl} + \overline{\mathrm{T}}^{(2)}_{\mathrm{IJ}}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right)\end{aligned} \tag{62}$$

and

$$\begin{aligned}\overline{\mathrm{T}}^{(1)}_{\mathrm{IK}} &= \mathrm{P}^{(1)}_{\mathrm{Im}}\mathrm{T}^{(1)}_{\mathrm{mn}}\mathrm{P}^{(1)}_{\mathrm{nK}} + 2\mathrm{P}^{(2)}_{\mathrm{II}}\mathrm{T}^{(1)}_{\mathrm{In}}\mathrm{P}^{(1)}_{\mathrm{nK}} + 2\mathrm{P}^{(1)}_{\mathrm{In}}\mathrm{T}^{(2)}_{\mathrm{nn}}\mathrm{P}^{(1)}_{\mathrm{nK}} + 4\mathrm{P}^{(2)}_{\mathrm{II}}\mathrm{T}^{(2)}_{\mathrm{II}}\mathrm{P}^{(1)}_{\mathrm{IK}} \\ &\quad + 2\sum_{\mathrm{m}=1}^{2}\mathrm{P}^{(1)}_{\mathrm{Im}}\mathrm{T}^{(1)}_{\mathrm{mK}}\mathrm{P}^{(2)}_{\mathrm{KK}} + 4\mathrm{P}^{(2)}_{\mathrm{II}}\mathrm{T}^{(1)}_{\mathrm{IK}}\mathrm{P}^{(2)}_{\mathrm{KK}} + 4\mathrm{P}^{(1)}_{\mathrm{IK}}\mathrm{T}^{(2)}_{\mathrm{KK}}\mathrm{P}^{(2)}_{\mathrm{KK}}\end{aligned} \tag{63}$$

$$\overline{\mathrm{T}}^{(2)}_{\mathrm{IJ}} = 4\mathrm{P}^{(2)}_{\mathrm{IJ}}\mathrm{T}^{(2)}_{\mathrm{IJ}}\mathrm{P}^{(2)}_{\mathrm{IJ}} \tag{64}$$

In the above equations, summation convention applies for the lowercase entities. By contrast, no summation convention applies for the uppercase entities.

The ensemble-averaged "current stress norm" for any point $\mathbf{x}$ in four-phase progressively debonded fiber-reinforced composite can be expressed as

$$\sqrt{\langle H\rangle\,(\mathbf{x})} = \left(1-\phi^{(1)}\right)\sqrt{\overline{\boldsymbol{\sigma}}:\ \overline{\mathbf{T}}\ :\ \overline{\boldsymbol{\sigma}}} \tag{65}$$

where $\phi^{(1)}$ is the *current* perfectly bonded (undamaged) fiber volume fraction. The parameter $(1-\phi^{(1)})$ reflects that the matrix is treated as elastoplastic and the fibers are assumed to be elastic only. It is noted that the debonded fibers partially lose the constraint to the matrix, and for simplicity, only the undamaged fibers (represented by $\phi^{(1)}$) are counted in Eq. 65. Therefore, the effective yield function for the four-phase composite can be characterized as

$$\overline{F} = \left(1-\phi^{(1)}\right)\sqrt{\overline{\boldsymbol{\sigma}}:\ \overline{\mathbf{T}}\ :\overline{\boldsymbol{\sigma}}} - K(\overline{e}^{\mathrm{p}}) \tag{66}$$

with the isotropic hardening function $K(\overline{e}^{\mathrm{p}})$ for the four-phase composite materials. For illustration, we assume that the overall flow rule for the composite is associative. In general, the overall flow rule of the composite may become *nonassociative* when fibers, microcracks, and voids exist according to dislocation dynamics analysis. Extension to the nonassociative flow rule can be constructed in a similar fashion but involving both the normal and tangential flow directions. It is also noted that the effective yield function is *pressure dependent* and not of the *von Mises* type anymore. Therefore, the effective ensemble-averaged plastic strain rate for the composite can be expressed as

$$\dot{\overline{\boldsymbol{\varepsilon}}}^{p} = \dot{\lambda}\frac{\partial\overline{F}}{\partial\overline{\boldsymbol{\sigma}}} = \left(1-\phi^{(1)}\right)\dot{\lambda}\,\frac{\overline{\mathbf{T}}\ :\overline{\boldsymbol{\sigma}}}{\sqrt{\overline{\boldsymbol{\sigma}}:\overline{\mathbf{T}}\ :\ \overline{\boldsymbol{\sigma}}}} \tag{67}$$

where $\dot{\lambda}$ denotes the plastic consistency parameter.

Inspired by the structure of the micromechanically derived stress norm, the effective equivalent plastic strain rate for the composites is expressed as

$$\dot{\bar{e}}^p \equiv \sqrt{\frac{2}{3}\,\dot{\bar{\boldsymbol{\varepsilon}}}^p : \overline{\mathbf{T}}^{-1} : \dot{\bar{\boldsymbol{\varepsilon}}}^p} = \sqrt{\frac{2}{3}}\left(1-\phi^{(1)}\right)\dot{\lambda} \tag{68}$$

The $\dot{\lambda}$ together with the yield function $\overline{F}$ obeys the *Kuhn–Tucker* loading/unloading conditions. The *Kuhn–Tucker* conditions can be displayed as

$$\dot{\lambda} \geq 0,\ \overline{F} \leq 0,\ \dot{\lambda}\overline{F} = 0,\ \dot{\lambda}\dot{\overline{F}} = 0 \tag{69}$$

The ensemble-averaged yield function in Eq. 66, the averaged plastic flow rule in Eq. 67, the equivalent plastic strain rate in Eq. 68, and the *Kuhn–Tucker* conditions completely characterize the effective plasticity formulation for a composite material with the isotropic hardening function $K(\bar{e}^{\mathrm{p}})$. Extension of the proposed model to accommodate the kinematic hardening is possible. Here, the simple power-law isotropic hardening function is employed as an example:

$$K(\bar{e}^p) = \sqrt{\frac{2}{3}}\left\{\bar{\sigma}_y + h(\bar{e}^p)^q\right\} \tag{70}$$

where $\bar{\sigma}_y$ is the initial yield stress and h and q signify the linear and exponential isotropic hardening parameters, respectively, for the four-phase composite.

Elastoplastic-Damage Responses for Composites

To elucidate the proposed micromechanical ensemble-area-homogenized transverse elastoplastic progressive damage formulation for composites, we consider here the transverse *uniaxial* and *biaxial* tensile loading under the plane-strain condition.

Elastoplastic-Damage Behavior under Transverse Uniaxial Tensile Loading

The applied macroscopic stress $\bar{\boldsymbol{\sigma}}$ can be written as

$$\bar{\sigma}_{11} > 0, \bar{\sigma}_{33} = \eta_1 \bar{\sigma}_{11},\ \text{all other}\ \bar{\sigma}_{ij} = 0 \tag{71}$$

With the simple isotropic hardening law described by Eq. 70, the overall yield function reads

$$\overline{F}(\bar{\sigma}_{11}, \bar{e}^p) = \left(1-\phi^{(1)}\right)\sqrt{\bar{\boldsymbol{\sigma}} : \overline{\mathbf{T}} : \bar{\boldsymbol{\sigma}}} - \sqrt{\frac{2}{3}}\left\{\bar{\sigma}_y + h(\bar{e}^p)^q\right\} \tag{72}$$

Substituting Eq. 71 into Eq. 72, the effective yield function for the special case of transverse uniaxial loading becomes

$$\overline{F}\left(\overline{\sigma}_{11},\overline{e}^{p}\right) = \left(1-\phi^{(1)}\right)\sqrt{\left(\overline{\mathrm{T}}_{11}^{(1)}+2\overline{\mathrm{T}}_{11}^{(2)}\right)}\,\overline{\sigma}_{11}-\sqrt{\frac{2}{3}}\left\{\overline{\sigma}_{y}+h(\overline{e}^{p})^{q}\right\} \tag{73}$$

The macroscopic incremental plastic strain rate defined by Eq. 67 then takes the form

$$\Delta\overline{\boldsymbol{\varepsilon}}^{p} = \left(1-\phi^{(1)}\right)\frac{\Delta\lambda}{\sqrt{\left(\overline{\mathrm{T}}_{11}^{(1)}+2\overline{\mathrm{T}}_{11}^{(2)}\right)}}\begin{bmatrix}\overline{\mathrm{T}}_{11}^{(1)}+2\overline{\mathrm{T}}_{11}^{(2)} & 0\\ 0 & \overline{\mathrm{T}}_{21}^{(1)}\end{bmatrix} \tag{74}$$

for any stress beyond the initial yielding. Here, $\Delta\lambda$ is the incremental effective plastic consistency parameter. Similarly, the incremental equivalent plastic strain becomes

$$\Delta\overline{e}^{p} = \sqrt{\frac{2}{3}}\left(1-\phi^{(1)}\right)\Delta\lambda \tag{75}$$

From the linear elasticity theory, the macroscopic incremental elastic strain takes the form

$$\Delta\overline{\boldsymbol{\varepsilon}}^{e} = \begin{bmatrix}\overline{D}_{11}+\eta_{1}\overline{D}_{13} & 0\\ 0 & \overline{D}_{21}+\eta_{1}\overline{D}_{23}\end{bmatrix}\Delta\overline{\sigma}_{11} \tag{76}$$

in which $\overline{D}_{ij}$ is the effective elastic compliance of the composite and η_1 is defined in Eq. 36.

For the monotonic plane-strain uniaxial loading, the overall incremental macroscopic stress–strain relation can be obtained by merging Eqs. 74 and 76 as follows:

$$\begin{aligned}\Delta\overline{\boldsymbol{\varepsilon}} = &\begin{bmatrix}\overline{D}_{11}+\eta_{1}\overline{D}_{13} & 0\\ 0 & \overline{D}_{21}+\eta_{1}\overline{D}_{23}\end{bmatrix}\Delta\overline{\sigma}_{11}\\ &+\left(1-\phi^{(1)}\right)\frac{\Delta\lambda}{\sqrt{\left(\overline{\mathrm{T}}_{11}^{(1)}+2\overline{\mathrm{T}}_{11}^{(2)}\right)}}\begin{bmatrix}\overline{\mathrm{T}}_{11}^{(1)}+2\overline{\mathrm{T}}_{11}^{(2)} & 0\\ 0 & \overline{\mathrm{T}}_{21}^{(1)}\end{bmatrix}\end{aligned} \tag{77}$$

where the positive parameter $\Delta\lambda$ is solved from the nonlinear equation obtained by enforcing the plastic consistency condition $\overline{F}=0$:

$$\left(1-\phi^{(1)}\right)\sqrt{\left(\overline{\mathrm{T}}_{11}^{(1)}+2\overline{\mathrm{T}}_{11}^{(2)}\right)}\left(\overline{\sigma}_{11}\right)_{n+1} = \sqrt{\frac{2}{3}}\left\{\overline{\sigma}_{y}+h\left[\overline{e}_{n}^{P}+\triangle\overline{e}_{n+1}^{P}\right]^{q}\right\} \tag{78}$$

Here, $\left(\overline{\sigma}_{11}\right)_{n+1}$ and $\triangle\overline{e}_{n+1}^{P}$ are the prescribed macroscopic stress along the 11-direction and the incremental equivalent plastic strain at the current time step, respectively. Further, $\overline{e}_{n}^{P}$ is the equivalent plastic strain at the previous load step. Equations 75 and 78 then result in

$$\left(1-\phi^{(1)}\right)\sqrt{\left(\overline{\mathrm{T}}_{11}^{(1)}+2\overline{\mathrm{T}}_{11}^{(2)}\right)}\,(\overline{\sigma}_{11})_{n+1}=\sqrt{\frac{2}{3}}\left\{\overline{\sigma}_y+h\left[\overline{e}_n^P+\sqrt{\frac{2}{3}}\left(1-\phi^{(1)}\right)\Delta\lambda\right]^q\right\} \tag{79}$$

Therefore, the expression for $\triangle\lambda$ arrives at

$$\triangle\lambda=\frac{1}{\sqrt{\frac{2}{3}}\,\left(1-\phi^{(1)}\right)}\left\{\left[\frac{\left(1-\phi^{(1)}\right)\sqrt{\frac{3}{2}\left(\overline{\mathrm{T}}_{11}^{(1)}+2\overline{\mathrm{T}}_{11}^{(2)}\right)}\,(\overline{\sigma}_{11})_{n+1}-\overline{\sigma}_y}{h}\right]^{1/q}-\overline{e}_n^p\right\} \tag{80}$$

Elastoplastic-Damage Behavior under Transverse Biaxial Tensile Loading

The applied macroscopic stress $\overline{\boldsymbol{\sigma}}$ can be rephrased as

$$\overline{\sigma}_{11}\;>\;0,\;\overline{\sigma}_{22}\;=\;\mathrm{R}\overline{\sigma}_{11},\overline{\sigma}_{33}\;=(\eta_1+\mathrm{R}\eta_2)\,\overline{\sigma}_{11},\;\text{ all other }\;\overline{\sigma}_{ij}\;=\;0. \tag{81}$$

Here, R is a parameter defining the loading *stress ratio*. Specifically, if $\mathrm{R}=0$, the transverse biaxial loading will reduce to the transverse uniaxial loading. Substituting Eq. 81 into Eq. 72, the effective yield function for the case of biaxial loading leads to

$$\overline{F}\,(\overline{\sigma}_{11},\overline{e}^p)\;=\;\left(1-\phi^{(1)}\right)\Phi(\mathrm{R})\,\overline{\sigma}_{11}-\sqrt{\frac{2}{3}}\left\{\overline{\sigma}_y+h(\overline{e}^p)^q\right\} \tag{82}$$

where

$$\Phi(\mathrm{R})=\sqrt{\left(\overline{\mathrm{T}}_{11}^{(1)}+2\overline{\mathrm{T}}_{11}^{(2)}\right)+\mathrm{R}^2\left(\overline{\mathrm{T}}_{22}^{(1)}+2\overline{\mathrm{T}}_{22}^{(2)}\right)+2\mathrm{R}\overline{\mathrm{T}}_{12}^{(1)}} \tag{83}$$

The macroscopic incremental plastic strain rate defined by Eq. 67 becomes

$$\Delta\overline{\boldsymbol{\varepsilon}}^p=\left(1-\phi^{(1)}\right)\frac{\Delta\lambda}{\Phi(\mathrm{R})}\begin{bmatrix}\overline{\mathrm{T}}_{11}^{(1)}+2\overline{\mathrm{T}}_{11}^{(2)}+\mathrm{R}\overline{\mathrm{T}}_{12}^{(1)} & 0\\ 0 & \overline{\mathrm{T}}_{21}^{(1)}+2\mathrm{R}\overline{\mathrm{T}}_{22}^{(2)}+\mathrm{R}\overline{\mathrm{T}}_{22}^{(1)}\end{bmatrix} \tag{84}$$

for any stress beyond the initial yielding. Similarly, the incremental equivalent plastic strain can be rephrased as

$$\Delta\overline{e}^p=\sqrt{\frac{2}{3}}\left(1-\phi^{(1)}\right)\Delta\lambda \tag{85}$$

From the linear elasticity theory, the macroscopic incremental elastic strain takes the form

$$\Delta\overline{\boldsymbol{\varepsilon}}^{e} = \begin{bmatrix} \overline{D}_{11} + \mathrm{R}\overline{D}_{12} + (\eta_1 + \mathrm{R}\eta_2)\overline{D}_{13} & 0 \\ 0 & \overline{D}_{21} + \mathrm{R}\overline{D}_{22} + (\eta_1 + \mathrm{R}\eta_2)\overline{D}_{23} \end{bmatrix} \Delta\overline{\sigma}_{11} \tag{86}$$

Again, $\overline{D}_{ij}$ is the effective elastic compliance of the composites, and η_1 and η_2 have been defined in Eq. 36.

For the monotonic plane-strain biaxial loading, the overall incremental macroscopic stress–strain relation can be obtained by summing Eqs. 84 and 86 as follows:

$$\begin{aligned} \Delta\overline{\boldsymbol{\varepsilon}} = & \begin{bmatrix} \overline{D}_{11} + \mathrm{R}\overline{D}_{12} + (\eta_1 + \mathrm{R}\eta_2)\overline{D}_{13} & 0 \\ 0 & \overline{D}_{21} + \mathrm{R}\overline{D}_{22} + (\eta_1 + \mathrm{R}\eta_2)\overline{D}_{23} \end{bmatrix} \Delta\overline{\sigma}_{11} \\ & + \left(1 - \phi^{(1)}\right) \frac{\Delta\lambda}{\Phi(\mathrm{R})} \begin{bmatrix} \overline{\mathrm{T}}_{11}^{(1)} + 2\overline{\mathrm{T}}_{11}^{(2)} + \mathrm{R}\overline{\mathrm{T}}_{12}^{(1)} & 0 \\ 0 & \overline{\mathrm{T}}_{21}^{(1)} + 2\mathrm{R}\overline{\mathrm{T}}_{22}^{(2)} + \mathrm{R}\overline{\mathrm{T}}_{22}^{(1)} \end{bmatrix} \end{aligned} \tag{87}$$

where the positive parameter $\Delta\lambda$ is solved from the nonlinear equation obtained by enforcing the plastic consistency condition $\overline{F} = 0$:

$$\left(1 - \phi^{(1)}\right) \Phi(\mathrm{R}) \left(\overline{\sigma}_{11}\right)_{n+1} = \sqrt{\frac{2}{3}} \left\{\overline{\sigma}_y + h\left[\overline{e}_n^P + \triangle\overline{e}_{n+1}^P\right]^q\right\} \tag{88}$$

Again, $\left(\overline{\sigma}_{11}\right)_{n+1}$ and $\triangle\overline{e}_{n+1}^P$ are the prescribed macroscopic stress along the 11-direction and the incremental equivalent plastic strain at the current time step, respectively. Here, $\overline{e}_n^P$ is the equivalent plastic strain at the previous load step. Equations 85 and 88 then render

$$\left(1 - \phi^{(1)}\right) \Phi(\mathrm{R}) \left(\overline{\sigma}_{11}\right)_{n+1} = \sqrt{\frac{2}{3}} \left\{\overline{\sigma}_y + h\left[\overline{e}_n^P + \sqrt{\frac{2}{3}}\left(1 - \phi^{(1)}\right)\triangle\lambda\right]^q\right\} \tag{89}$$

Therefore, the expression for $\triangle\lambda$ reads

$$\triangle\lambda = \frac{1}{\sqrt{\frac{2}{3}}\left(1 - \phi^{(1)}\right)} \left\{\left[\frac{\left(1 - \phi^{(1)}\right)\sqrt{\frac{3}{2}}\,\Phi(\mathrm{R})\left(\overline{\sigma}_{11}\right)_{n+1} - \overline{\sigma}_y}{h}\right]^{1/q} - \overline{e}_n^p\right\} \tag{90}$$

If there is no interfacial debonding damage, the orthotropic composite would recover the original *transversely isotropic* composite.

Numerical Simulations and Experimental Comparison

The experimental studies to characterize the damage evolution mechanisms in the fiber-reinforced metal matrix composites have been very limited in the literature to date. To assess the predictive capability of the present framework, comparisons are

made between the present theoretical predictions and the limited experimental data as quoted by Nimmer et al. (1991). The experimental data were observed at 23°, 315°, and 427°C for the SiC/Ti–6Al–4Vmetal matrix composites, with unidirectional silicon-carbide fiber (Textron SCS-6, with 32 % in fiber volume fraction) in a Ti–6Al–4V matrix under uniaxial transverse tensile normal loading. The experimental data recorded at 23 °C is selected for comparison here. Residual stresses occur by subjecting the composites to cooling from the processing temperature before application of mechanical loads. For simplicity, the silicon-carbide fibers are considered to be isotropic and homogeneous elastic properties. The elastic moduli of the matrix and fibers as reported by Nimmer et al. (1991) are $E_0 = 113.7 \times 10^3$ MPa and $\nu_0 = 0.3$ for the Ti - 6Al - 4V metal matrix at 21°C and $E_1 = 414 \times 10^3$ MPa and $\nu_1 = 0.3$ for the SiC fiber. Moreover, the least-square parameter estimation procedure is employed to estimate the following plastic parameters based on the experimental data at 23 °C: $\overline{\sigma}_y = 500$ MPa, $h = 700$ MPa, and $q = 0.1$. To estimate Weibull's parameters S_0 and M, Ju and Lee (2000, 2001) are followed, and $S_0 = 180$ MPa, $M = 3$, and $\sigma_{\text{cri}} = 170$ MPa are adopted. The residual stresses due to cooling process are in part considered in our estimation for S_0 and σ_{cri} here. Furthermore, the residual stress in the matrix is considered below the yield strength in the matrix. If the residual stress in the matrix (under tension) is greater than the yield strength in the matrix, then the matrix may yield even before the application of mechanical loading. In that event, it would be necessary to incorporate the prescribed uniform thermal eigenstrain inside the cylindrical fibers. If the temperature change during processing is sufficiently large, the prescribed thermal eigenstrain would cause the ductile matrix to yield and produce plastic flow. To account for the change of matrix properties due to the residual stress-induced plastic flow, the secant method (cf. Berveiller and Zaoui 1979; Tandon and Weng 1988) can be employed so that the secant Young's modulus and Poisson's ratio of the matrix become dependent upon the equivalent plastic strain (cf. Liu and Sun 2004, in the absence of damage).

Uniaxial Transverse Tensile Loading

The transverse uniaxial stress–strain responses are often referred to as important indicators of mechanical behaviors of composite materials. The comparison between the present predictions and experimental data on the overall *uniaxial* elastoplastic with and without damage behavior of fiber-reinforced ductile matrix composites is shown in Fig. 5a. As expected, the predictions without debonding models highly overestimate the behavior when compared with the experimental data. In general, the predictions based on the present model are in good agreement with experimental data. As depicted in Fig. 5a, the experimental data, at a load substantially less than the measured yield strength, shows a significant reduction in transverse modulus taking place. That is, the creation of a characteristic first "knee" in the transverse tensile stress–strain curve is observed. Physical evidence that the experimentally observed first "knee" is actually associated with such an interfacial separation event is provided by the edge replica experiments which identify "gaps"

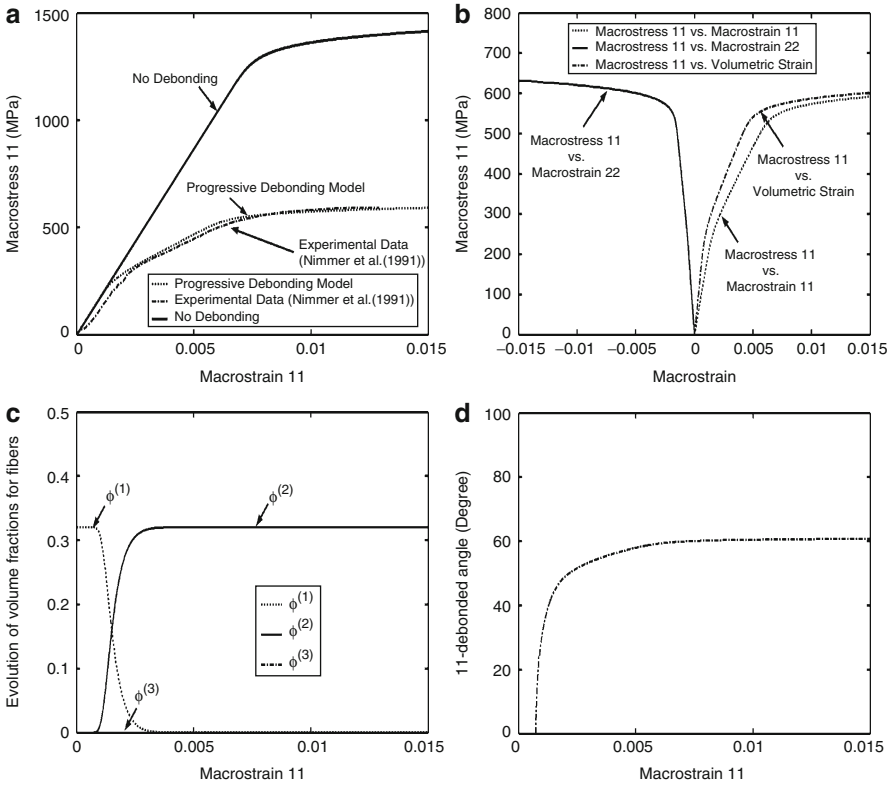


Fig. 5 (**a**) Comparison between the predicted and the measured transverse uniaxial stress–strain behavior of SiC/Ti–6Al–4V composite with the initial fiber volume fraction 32 %, at room temperature = 23 °C; (**b**) $\overline{\sigma}_{11}$ versus $\overline{\varepsilon}_{\mathrm{v}}$ (volumetric strain) and $\overline{\sigma}_{11}$ versus $\overline{\varepsilon}_{22}$; (**c**) the corresponding progressive damage of fibers; (**d**) the evolution of debonding angle versus $\overline{\varepsilon}_{11}$

between the fibers and matrix above the first "knee" in a transverse test (Nimmer et al. 1991). In order to separate the interfaces between the cylindrical fibers and the matrix, thus creating the first "knee," the compressive residual stresses must be overcome by the tensile mechanical loading. At still higher stress level, the matrix plasticity dominates and a maximum stress is reached. Therefore, the second "knee" is formed. The feature of trilinear-like stress–strain curve is well captured under the present formulation.

Figure 5b displays the predictions of $\overline{\sigma}_{11}$ versus $\overline{\varepsilon}_{\mathrm{v}}$ (the volumetric strain) and $\overline{\sigma}_{11}$ versus $\overline{\varepsilon}_{22}$ corresponding to Fig. 5a, respectively. The volumetric strain increases as the loading increases after the composite yields. Clearly, the effective yield function in the current framework is *pressure dependent* and not of the *von Mises* type. Due to Poisson's ratio effect, the $\overline{\varepsilon}_{22}$ strain under $\overline{\sigma}_{11}$ is negative. It is noted that due to the projection of the interfacial debonding arc onto the x_2-direction (cf. Fig. 4b), minor stiffness degradation along the x_2-direction is noted.

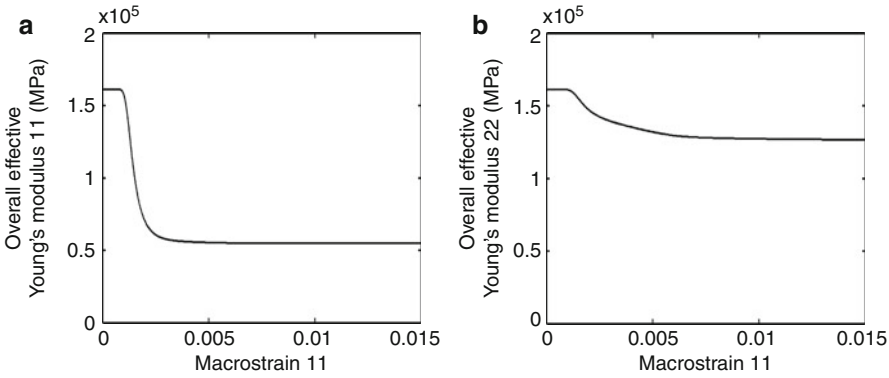


Fig. 6 Overall effective elastic modulus versus $\overline{\varepsilon}_{11}$ corresponding to Fig. 5: (**a**) $\overline{E}_{11}$ versus $\overline{\varepsilon}_{11}$, (**b**) $\overline{E}_{22}$ versus $\overline{\varepsilon}_{11}$

Moreover, Fig. 5c shows that the interfacial debonding damage does not occur until $\overline{\sigma}_{11} = 128$ MPa and $\overline{\varepsilon}_{11} = 0.000739$. The interfacial damage gradually increases as the strain increases up to $\overline{\varepsilon}_{11} = 0.0046697$. Following the next few incremental stress steps, the volume fraction of partially debonded fibers (Mode 2) reaches almost a constant around 32 %; i.e., nearly all fibers have partially debonded in the composite. In addition, according to the simulation results, $\overline{F}(\overline{\sigma}_{11}, \overline{e}^p) = 0$ in Eq. 72 occurs at $\overline{\varepsilon}_{11} = 0.0034439$ and $\overline{\sigma}_{11} = 378$ MPa. It also explains why the stress–strain curve associated with present prediction demonstrates the overall plastic hardening effect after $\overline{\varepsilon}_{11} = 0.0034439$ as shown in Fig. 5a. Under this loading condition, the direction of the first local principal stress coincides with the external uniaxial loading direction. From Eq. 1, the first local principal stress is tensile, and the second local principal stress is compressive. Therefore, only one debonding mode (Mode 2) – debonding along the loading direction – would occur.

In Fig. 5d, the debonding (arc microcrack)-angle progression as a function of the overall strain history is exhibited. It is observed that the debonding angle increases rapidly at the beginning stage and then become saturated as the overall deformation increases.

In addition, the overall effective transverse Young's modulus $\overline{E}_{11}$ versus $\overline{\varepsilon}_{11}$ and $\overline{E}_{22}$ versus $\overline{\varepsilon}_{11}$ is shown in Fig. 6a, b, respectively. Due to the nature of the loading along the x_1-direction (cf. Fig. 4), the overall effective transverse Young's modulus $\overline{E}_{11}$ demonstrates larger reduction when compared with the overall effective transverse Young's modulus $\overline{E}_{22}$ during the loading history.

Biaxial Transverse Tensile Loading

As discussed before, the transverse uniaxial loading causes only one tensile local principal stress within the fibers. Therefore, only one type of debonding mode

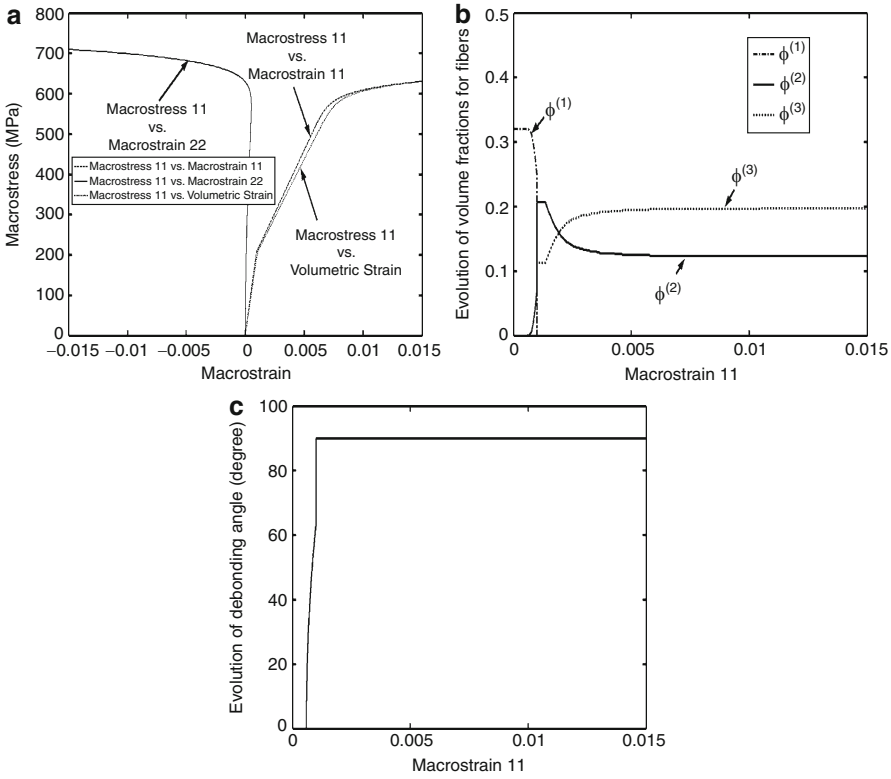


Fig. 7 The transverse biaxial tension simulation: (**a**) the overall transverse stress–strain relation $\overline{\sigma}_{11}$ versus $\overline{\varepsilon}_{11}$, $\overline{\sigma}_{11}$ versus $\overline{\varepsilon}_{\mathrm{v}}$ (volumetric strain), and $\overline{\sigma}_{11}$ versus $\overline{\varepsilon}_{22}$ of the SiC/Ti–6Al–4V composite with initial fiber volume fraction 32 %, at room temperature = 23°C for R = 0.5; (**b**) the corresponding volume fractions of progressive interfacial damages versus $\overline{\varepsilon}_{11}$; (**c**) the corresponding evolution of partial debonding angle versus $\overline{\varepsilon}_{11}$

(Mode 2) would be triggered. Under the transverse biaxial loading condition featuring different far-field stress components (e.g., $\overline{\sigma}_{11} = 2\overline{\sigma}_{22}$), more than one partial debonding modes can be activated. Figure 7a presents the elastoplastic-damage responses of $\overline{\sigma}_{11}$ versus $\overline{\varepsilon}_{11}$, $\overline{\sigma}_{11}$ versus $\overline{\varepsilon}_{22}$, and $\overline{\sigma}_{11}$ versus $\overline{\varepsilon}_{\mathrm{v}}$ (the volumetric strain). Due to Poisson's ratio effect, the value of $\overline{\varepsilon}_{22}$ is positive in the initial loading steps and becomes negative macroscopic strain as loading increases. Figure 7b exhibits the progressive volume fractions of interfacial damages over the strain history. With the increasing applied external loads, progressive interfacial debonding may develop and transform from one mode to another. It is noted that when $\alpha_{12}^{(2)}$ reaches $\pi/2$, i.e., the second local principal stress also reaches the critical debonding strength, part of the damage Mode 2 will transform into Mode 3, indicating that the fiber arc between the x_1 and x_2 axes is totally debonded; and the projected debonding area will evolve toward the x_2 axis from the x_1 axis (cf. Fig. 4).

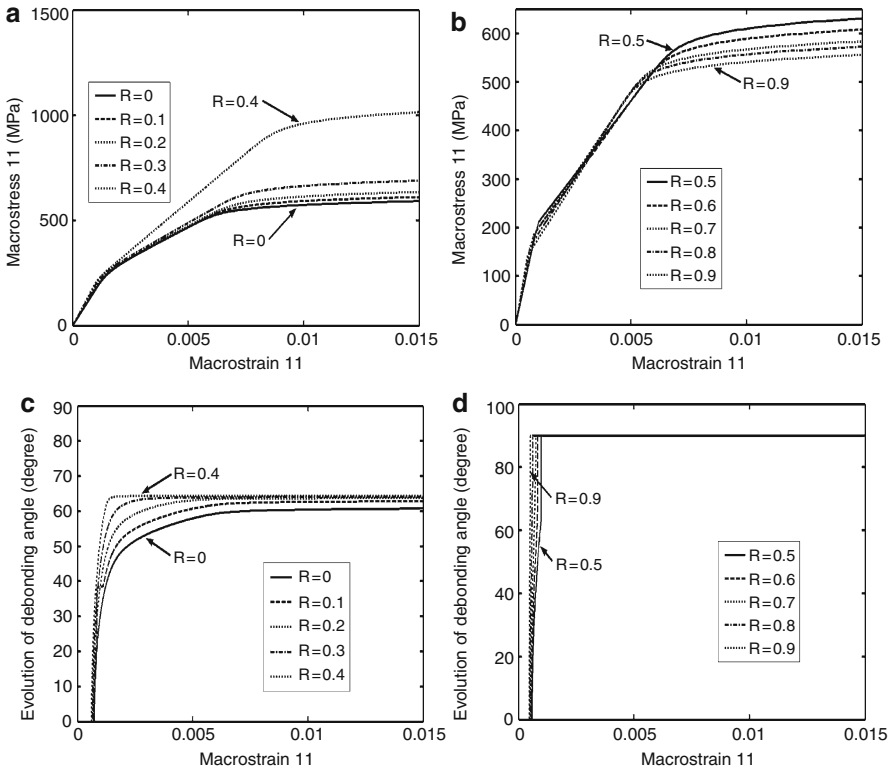


Fig. 8 The transverse biaxial stress–strain relation $\overline{\sigma}_{11}$ versus $\overline{\varepsilon}_{11}$ of the SiC/Ti–6Al–4V composite with initial fiber volume fraction 32 %, at room temperature = 23 °C (**a**) for varied R ($0 \sim 0.4$) and (**b**) for varied R ($0.5 \sim 0.9$); (**c**) the evolution of debonding angle versus $\overline{\varepsilon}_{11}$ corresponding to (**a**); (**d**) the evolution of debonding angle versus $\overline{\varepsilon}_{11}$ corresponding to (**b**)

In Fig. 7c, the debonding-angle progression as a function of the overall strain and stress history is observed. The debonding angle $\alpha_{12}^{(2)}$ quickly reaches $\pi/2$. Under the transverse biaxial loading condition, with the increase of the second local principal stress within the fibers, the partial fiber debonding propagates toward the second direction easily and quickly reaches its ultimate value.

In addition, in order to investigate the effects of the *stress ratio* R on the overall elastoplastic-damage behavior, the composites are subjected to the biaxial tensile loading with varying stress ratios, R. In Fig. 8a, the overall $\overline{\sigma}_{11}$ versus $\overline{\varepsilon}_{11}$ is rendered. In general, as R increases from 0 to 0.4, the overall response demonstrates higher stiffness in the elastic range and higher yielding strength of the composites. On the other hand, as shown in Fig. 8b, as R increases from 0.5 to 0.9, the overall response demonstrates higher stiffness in the elastic range yet lower yielding strength of the composites. This is mainly due to the rapid evolutionary debonding-angle progress as R increases from 0.5 to 0.9 (cf. Fig. 8c, d).

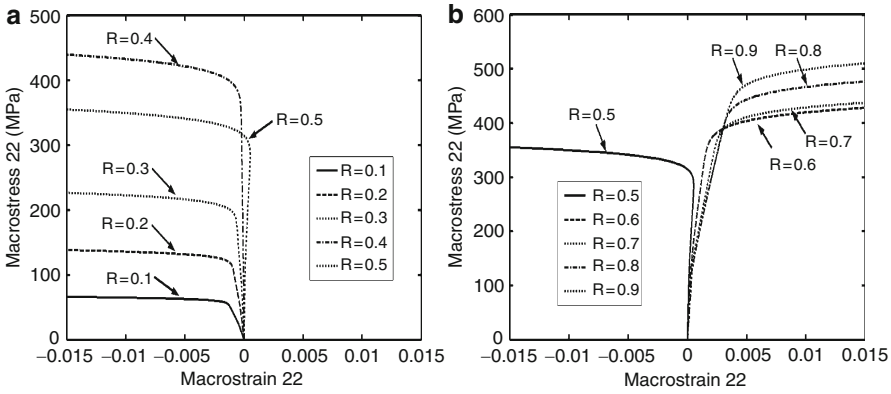


Fig. 9 The transverse biaxial stress–strain relation $\overline{\sigma}_{22}$ versus $\overline{\varepsilon}_{22}$ of the SiC/Ti–6Al–4V composite with initial fiber volume fraction 32 %, at room temperature = 23°C (**a**) for varied R between 0 and 0.5 and (**b**) for varied R between 0.5 and 0.9

Figure 9 displays the $\overline{\sigma}_{22}$ versus $\overline{\varepsilon}_{22}$ with varying stress ratios, R. When R changes from 0.1 to 0.5, the so-called "bend over" effect becomes more significant (cf. Fig. 9a). As a result, the values of $\overline{\varepsilon}_{22}$ are negative. However, when R increases from 0.6 to 0.9, the positive values of $\overline{\varepsilon}_{22}$ and higher $\overline{\sigma}_{22}$ versus $\overline{\varepsilon}_{22}$ responses are shown in Fig. 9b. This is mainly due to the reason that higher $\overline{\sigma}_{22}$ would overcome Poisson's ratio effect under $\overline{\sigma}_{11}$.

Detailed Derivation for Tensor T in Eq. 49

$$\mathrm{A_{q1}} = \frac{1}{2{\beta_1}^2(\alpha_1+\beta_1)^2}\left\{\left[\frac{1}{4}\left(5\hat{a}+5\hat{b}+3\hat{c}+2\right){\alpha_1}^2+\beta_1\left(\frac{9\hat{a}}{2}+\frac{\hat{b}}{2}+\frac{3\hat{c}}{2}\right.\right.\right.$$
$$\left.+4\left(\hat{b}-\hat{a}\right)\nu_0+1\right)\alpha_1+\frac{1}{4}{\beta_1}^2\left(16\left(\hat{a}+\hat{b}-\hat{c}+2\right){\nu_0}^2-16(2\hat{a}-\hat{c}+2)\nu_0$$
$$\left.\left.+17\hat{a}+\hat{b}-\hat{c}+10\right)\right]\phi^{(1)}\right\} \tag{91}$$

$$\mathrm{A_{q211}} = \phi^{(2)}\left\{\frac{1}{32(-1+\nu_0)^2}\left[{\mu_0}^2\left(\left(10+17\hat{a}+\hat{b}-\hat{c}\right)\hat{\mathrm{B}}^2_{1111\mathrm{d}}+\left(10+\hat{a}+17\hat{b}-\hat{c}\right)\hat{\mathrm{B}}^2_{2111\mathrm{d}}\right.\right.\right.$$
$$-4\left(-6+\hat{a}+\hat{b}+7\hat{c}\right)\hat{\mathrm{B}}_{2111\mathrm{d}}\overline{\mathrm{B}}_{1111\mathrm{d}}+4\left(10+17\hat{a}+\hat{b}-\hat{c}\right)\overline{\mathrm{B}}^2_{1111\mathrm{d}}$$
$$\left.\left.\left.+2\hat{\mathrm{B}}_{1111\mathrm{d}}\left(-\left(-6+\hat{a}+\hat{b}+7\hat{c}\right)\hat{\mathrm{B}}_{2111\mathrm{d}}+2\left(10+17\hat{a}+\hat{b}-\hat{c}\right)\overline{\mathrm{B}}_{1111\mathrm{d}}\right)\right)\right]\right\} \tag{92}$$

$$A_{q2} = \frac{1}{2{\beta_2}^2(\alpha_2+\beta_2)^2}\left\{\left[\frac{1}{4}(5\hat{a}+5\hat{b}+3\hat{c}+2){\alpha_2}^2+\beta_2\left(\frac{9\hat{a}}{2}+\frac{\hat{b}}{2}+\frac{3\hat{c}}{2}\right.\right.\right.$$
$$\left.+4(\hat{b}-\hat{a})\nu_0+1)\alpha_2+\frac{1}{4}{\beta_2}^2(16(\hat{a}+\hat{b}-\hat{c}+2){\nu_0}^2-16(2\hat{a}-\hat{c}+2)\nu_0$$
$$\left.\left.\left.+17\hat{a}+\hat{b}-\hat{c}+10)\right]\phi^{(3)}\right\} \tag{93}$$

$$B_{q1} = \frac{1}{2{\beta_1}^2(\alpha_1+\beta_1)^2}\left\{\left[\frac{1}{4}(5\hat{a}+5\hat{b}+3\hat{c}+2){\alpha_1}^2+\beta_1\left(\frac{\hat{a}}{2}+\frac{9\hat{b}}{2}+\frac{3\hat{c}}{2}\right.\right.\right.$$
$$\left.+4(\hat{a}-\hat{b})\nu_0+1)\alpha_1+\frac{1}{4}{\beta_1}^2(16(\hat{a}+\hat{b}-\hat{c}+2){\nu_0}^2-16(2\hat{b}-\hat{c}+2)\nu_0$$
$$\left.\left.\left.+\hat{a}+17\hat{b}-\hat{c}+10)\right]\phi^{(1)}\right\} \tag{94}$$

$$B_{q211} = \frac{1}{32(-1+\nu_0)^2}\left\{{\mu_0}^2\left[(10+17\hat{a}+\hat{b}-\hat{c})\hat{B}^2_{1211d}-2(-6+\hat{a}+\hat{b}+7\hat{c})\hat{B}_{1211d}\right.\right.$$
$$\left.\left.(\hat{B}_{2211d}+2\overline{B}_{2211d})+(10+\hat{a}+17\hat{b}-\hat{c})(\hat{B}_{2211d}+2\overline{B}_{2211d})^2\right]\phi^{(2)}\right\} \tag{95}$$

$$B_{q2} = \frac{1}{2{\beta_2}^2(\alpha_2+\beta_2)^2}\left\{\left[\frac{1}{4}(5\hat{a}+5\hat{b}+3\hat{c}+2){\alpha_2}^2+\beta_2\left(\frac{\hat{a}}{2}+\frac{9\hat{b}}{2}+\frac{3\hat{c}}{2}\right.\right.\right.$$
$$\left.+4(\hat{a}-\hat{b})\nu_0+1)\alpha_2+\frac{1}{4}{\beta_2}^2(16(\hat{a}+\hat{b}-\hat{c}+2){\nu_0}^2-16(2\hat{b}-\hat{c}+2)\nu_0$$
$$\left.\left.\left.+\hat{a}+17\hat{b}-\hat{c}+10)\right]\phi^{(3)}\right\} \tag{96}$$

$$C_{q1} = \frac{1}{{\beta_1}^2(\alpha_1+\beta_1)^2}\left\{\left[-\frac{1}{4}(5\hat{a}+5\hat{b}+3\hat{c}+2){\alpha_1}^2-\frac{1}{2}(5\hat{a}+5\hat{b}+3\hat{c}+2)\alpha_1\beta_1\right.\right.$$
$$\left.\left.+\frac{1}{4}{\beta_1}^2(16(\hat{a}+\hat{b}-\hat{c}+2){\nu_0}^2-16(\hat{a}+\hat{b}-\hat{c}+2)\nu_0-\hat{a}-\hat{b}-7\hat{c}+6)\right]\phi^{(1)}\right\} \tag{97}$$

$$C_{q211} = \frac{1}{16(-1+\nu_0)^2}\{{\mu_0}^2[\hat{B}_{1211d}(-(-6+\hat{a}+\hat{b}+7\hat{c})\hat{B}_{2111d}+2(10+17\hat{a}+\hat{b}-\hat{c})$$
$$\overline{B}_{1111d}+((10+\hat{a}+17\hat{b}-\hat{c})\hat{B}_{2111d}-2(-6+\hat{a}+\hat{b}+7\hat{c})\overline{B}_{1111d})$$
$$(\hat{B}_{2211d}+2\overline{B}_{2211d})+\hat{B}_{1111d}((10+17\hat{a}+\hat{b}-\hat{c})\hat{B}_{1211d}-(-6+\hat{a}+\hat{b}+7\hat{c})$$
$$(\hat{B}_{2211d}+2\overline{B}_{2211d}))\phi^{(2)}\} \tag{98}$$

$$\mathrm{C_{q2}} = \frac{1}{{\beta_2}^2(\alpha_2+\beta_2)^2}\left\{\left[-\frac{1}{4}\left(5\hat{\mathrm{a}}+5\hat{\mathrm{b}}+3\hat{\mathrm{c}}+2\right){\alpha_2}^2-\frac{1}{2}\left(5\hat{\mathrm{a}}+5\hat{\mathrm{b}}+3\hat{\mathrm{c}}+2\right)\alpha_2\beta_2 \right.\right.$$
$$\left.\left.+\frac{1}{4}{\beta_2}^2\left(16\left(\hat{\mathrm{a}}+\hat{\mathrm{b}}-\hat{\mathrm{c}}+2\right){\nu_0}^2-16\left(\hat{\mathrm{a}}+\hat{\mathrm{b}}-\hat{\mathrm{c}}+2\right)\nu_0-\hat{\mathrm{a}}-\hat{\mathrm{b}}-7\hat{\mathrm{c}}+6\right)\right]\phi^{(3)}\right\} \tag{99}$$

$$\begin{aligned}\alpha_1 &= (4\nu_0-1)+\frac{(\lambda_0\mu_1-\lambda_1\mu_0)[8(1-\nu_0)]}{2(\mu_1-\mu_0)(\lambda_1-\lambda_0+\mu_1-\mu_0)},\\ \beta_1 &= (3-4\nu_0)+4(1-\nu_0)\frac{\mu_0}{(\mu_1-\mu_0)}\end{aligned} \tag{100}$$

$$\alpha_2 = (4\nu_0-1), \beta_2 = -1,\ \ \mathrm{m} = \frac{1}{(\lambda_0+\mu_0)}-\frac{1}{\mu_0},\ \ \mathrm{n} = \frac{1}{4\mu_0} \tag{101}$$

$$\hat{\mathrm{B}}_{1111\mathrm{d}} = \frac{(\nu_0-1)[-4\nu_0\mathrm{n}+\mathrm{n}+\mathrm{m}(8\nu_0-5)][\mu_0(8\nu_0-5)+\mu_1(\nu_1+1)(\nu_0-1)(8\nu_0-1)]}{2(2\nu_0-1)[4\mu_1(\nu_1+1)(\nu_0-1)(4\nu_0-1)+3\mu_0(8\nu_0-5)]} \tag{102}$$

$$\hat{\mathrm{B}}_{2111\mathrm{d}} = -\frac{[\mu_0+\mu_1(\nu_1+1)(\nu_0-1)](\nu_0-1)[-\mu_1+5\mu_0+4(\mu_1-2\mu_0)\nu_0][-4\nu_0\mathrm{n}+\mathrm{n}+\mathrm{m}(8\nu_0-5)]}{2(2\nu_0-1)(-4\nu_0\mu_1+3\mu_1+\mu_0)[4\mu_1(\nu_1+1)(\nu_0-1)(4\nu_0-1)+3\mu_0(8\nu_0-5)]} \tag{103}$$

$$\begin{aligned}\hat{\mathrm{B}}_{2211\mathrm{d}} = \{&(\nu_0-1)\left(\mathrm{m}(\mu_0+\mu_1(\nu_1+1)(\nu_0-1))(8\nu_0-5)(-\mu_1+5\mu_0+4(\mu_1-2\mu_0)\nu_0)\right.\\ &+\mathrm{n}(-(\nu_1+1)(\nu_0-1)(4\nu_0-1)(12\nu_0-5){\mu_1}^2+\mu_0(-32(\nu_1+1){\nu_0}^3+4(31\nu_1+3){\nu_0}^2\\ &+4(31\nu_1+3){\nu_0}^2+(21-131\nu_1)\nu_0+39\nu_1-10)\mu_1+{\mu_0}^2(4\nu_0(8\nu_0-7)+5))]\}/\\ &\{2(2\nu_0-1)(4\nu_0\mu_1-3\mu_1-\mu_0)[4\mu_1(\nu_1+1)(\nu_0-1)(4\nu_0-1)+3\mu_0(8\nu_0-5)]\}\end{aligned} \tag{104}$$

$$\begin{aligned}\hat{\mathrm{B}}_{1211\mathrm{d}} = \{&(\nu_0-1)(8\nu_0-5)[\mathrm{n}(-4\nu_0\mu_0+\mu_0+\mu_1(\nu_1+1)(\nu_0-1)(4\nu_0+1))\\ &+\mathrm{m}(\mu_0(8\nu_0-5)+\mu_1(\nu_1+1)(\nu_0-1)(8\nu_0-1))]\}/\\ &\{2(2\nu_0-1)[4\mu_1(\nu_1+1)(\nu_0-1)(4\nu_0-1)+3\mu_0(8\nu_0-5)]\}\end{aligned} \tag{105}$$

$$\begin{aligned}\overline{\mathrm{B}}_{1111\mathrm{d}} &= \mathrm{n}(\nu_0-1),\ \ \overline{\mathrm{B}}_{2211\mathrm{d}} = \frac{\mathrm{n}(\mu_1-\mu_0)(\nu_0-1)}{4\nu_0\mu_1-3\mu_1-\mu_0},\\ \overline{\mathrm{B}}_{1211\mathrm{d}} &= \mathrm{n}(\nu_0-1),\ \ \overline{\mathrm{B}}_{2211\mathrm{d}} = \mathrm{n}(\nu_0-1)\end{aligned} \tag{106}$$

$$\begin{aligned}&\hat{\mathrm{a}} = \frac{2\eta_1^2}{9}-\frac{2\eta_1}{9}+\frac{1}{9}(2\eta_1-1)^2+\frac{5}{9},\ \ \hat{\mathrm{b}} = \frac{2\eta_2^2}{9}-\frac{2\eta_2}{9}+\frac{1}{9}(2\eta_2-1)^2+\frac{5}{9},\\ &\eta_1 = \frac{\overline{C}_{13}\overline{C}_{22}-\overline{C}_{12}\overline{C}_{23}}{\Xi}\end{aligned} \tag{107}$$

$$\begin{aligned}\hat{c} &= \frac{4\eta_2\eta_1}{9} - \frac{2\eta_1}{9} - \frac{2\eta_2}{9} + \frac{2}{9}(2\eta_1 - 1)(2\eta_2 - 1) - \frac{8}{9},\\ \eta_2 &= \frac{\overline{C}_{11}\overline{C}_{23} - \overline{C}_{12}\overline{C}_{13}}{\Xi},\ \ \Xi = \overline{C}_{11}\overline{C}_{22} - \overline{C}_{12}^2\end{aligned} \tag{108}$$

Detailed Derivation for Tensor P in Eq. 58

$$\sigma_{ij}^{o} = \left[\mathrm{P}_{\mathrm{IK}}^{(1)}\delta_{ij}\delta_{kl} + \mathrm{P}_{\mathrm{IJ}}^{(2)}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right)\right] : \overline{\sigma}_{kl}, \quad i,j,k,l = 1,2 \tag{109}$$

where

$$\begin{aligned}\mathrm{P}_{\mathrm{IK}}^{(1)} &= -\frac{\Lambda_{\mathrm{IK}}}{2\hat{\mathrm{W}}_{\mathrm{II}}^{(2)}},\ \ \mathrm{P}_{\mathrm{IJ}}^{(2)} = \frac{1}{4\hat{\mathrm{W}}_{\mathrm{IJ}}^{(2)}},\\ \left\{\begin{matrix}\Lambda_{\mathrm{I1}}\\ \Lambda_{\mathrm{I2}}\end{matrix}\right\} &= \begin{bmatrix}\hat{\mathrm{W}}_{11}^{(1)} + 2\hat{\mathrm{W}}_{11}^{(2)} & \hat{\mathrm{W}}_{21}^{(1)}\\ \hat{\mathrm{W}}_{12}^{(1)} & \hat{\mathrm{W}}_{22}^{(1)} + 2\hat{\mathrm{W}}_{22}^{(2)}\end{bmatrix}^{-1}\left\{\begin{matrix}\hat{\mathrm{W}}_{\mathrm{I1}}^{(1)}\\ \hat{\mathrm{W}}_{\mathrm{I2}}^{(1)}\end{matrix}\right\}\end{aligned} \tag{110}$$

$$\hat{\mathrm{W}}_{\mathrm{IK}}^{(1)} = \sum_{m=1}^{2}\lambda_0\Gamma_{\mathrm{mK}}^{(1)} + 2\mu_0\Gamma_{\mathrm{IK}}^{(1)} + 2\lambda_0\Gamma_{\mathrm{KK}}^{(2)},\ \ \hat{\mathrm{W}}_{\mathrm{IJ}}^{(2)} = 2\mu_0\Gamma_{\mathrm{IJ}}^{(2)} \tag{111}$$

$$\Gamma_{\mathrm{IK}}^{(1)} = \sum_{m=1}^{2}\hat{\Gamma}_{\mathrm{Im}}^{(1)}\left(\frac{-\lambda_0}{4\mu_0(\lambda_0+\mu_0)}\right) + 2\hat{\Gamma}_{\mathrm{II}}^{(2)}\left(\frac{-\lambda_0}{4\mu_0(\lambda_0+\mu_0)}\right) + \frac{2\hat{\Gamma}_{\mathrm{IK}}^{(1)}}{4\mu_0},\ \ \Gamma_{\mathrm{IJ}}^{(2)} = \frac{2\hat{\Gamma}_{\mathrm{IJ}}^{(2)}}{4\mu_0} \tag{112}$$

$$\hat{\Gamma}_{\mathrm{IK}}^{(1)} = \sum_{m=1}^{2}\widetilde{\Gamma}_{\mathrm{IK}}^{(1)}\left(\frac{-(4\nu_0-1)}{8(1-\nu_0)}\right) + 2\widetilde{\Gamma}_{\mathrm{IK}}^{(1)}\left(\frac{1}{2} - \frac{3-4\nu_0}{8(1-\nu_0)}\right) + 2\widetilde{\Gamma}_{\mathrm{KK}}^{(2)}\left(-\frac{4\nu_0-1}{8(1-\nu_0)}\right) \tag{113}$$

$$\hat{\Gamma}_{\mathrm{IJ}}^{(2)} = \frac{1}{2} + 2\left(\frac{1}{2} - \frac{3-4\nu_0}{8(1-\nu_0)}\right)\left(\widetilde{\Gamma}_{\mathrm{IJ}}^{(2)}\right) \tag{114}$$

$$\begin{aligned}\widetilde{\Gamma}_{\mathrm{IK}}^{(1)} &= \phi^{(1)}\widetilde{\lambda}^{(1)\prime} + \phi^{(2)}\overline{\Gamma}_{\mathrm{IK},11}^{(1)} + \phi^{(3)}\overline{\Gamma}_{\mathrm{IK},12}^{(1)},\\ \widetilde{\Gamma}_{\mathrm{IJ}}^{(2)} &= \phi^{(1)}\widetilde{\mu}^{(1)\prime} + \phi^{(2)}\overline{\Gamma}_{\mathrm{IJ},11}^{(2)} + \phi^{(3)}\overline{\Gamma}_{\mathrm{IJ},12}^{(2)}\end{aligned} \tag{115}$$

$$\overline{\Gamma}_{\mathrm{IK},12}^{(1)} = -\frac{\widetilde{\Lambda}_{\mathrm{IK},12}^{(2)}}{2\left(\frac{3-4\nu_0}{8(1-\nu_0)} + \chi_{\mathrm{II},12}^{(2)}\right)},\ \ \overline{\Gamma}_{\mathrm{IJ},12}^{(2)} = \frac{1}{4\left(\frac{3-4\nu_0}{8(1-\nu_0)} + \chi_{\mathrm{IJ},12}^{(2)}\right)} \tag{116}$$

$$\left\{\begin{matrix}\widetilde{\Lambda}^{(2)}_{\mathrm{I1},12}\\ \widetilde{\Lambda}^{(2)}_{\mathrm{I2},12}\end{matrix}\right\}=\left[\begin{matrix}\left[\begin{matrix}\frac{4\nu_0-1}{8(1-\nu_0)}+\chi^{(1)}_{11,12}\\ +2\left(\frac{3-4\nu_0}{8(1-\nu_0)}+\chi^{(2)}_{11,12}\right)\end{matrix}\right] & \frac{4\nu_0-1}{8(1-\nu_0)}+\chi^{(1)}_{21,12}\\ \frac{4\nu_0-1}{8(1-\nu_0)}+\chi^{(1)}_{12,12} & \left[\begin{matrix}\frac{4\nu_0-1}{8(1-\nu_0)}+\chi^{(1)}_{22,12}\\ +2\left(\frac{3-4\nu_0}{8(1-\nu_0)}+\chi^{(2)}_{22,12}\right)\end{matrix}\right]\end{matrix}\right]^{-1}\left\{\begin{matrix}\frac{4\nu_0-1}{8(1-\nu_0)}+\chi^{(1)}_{\mathrm{I1},12}\\ \frac{4\nu_0-1}{8(1-\nu_0)}+\chi^{(1)}_{\mathrm{I2},12}\end{matrix}\right\} \tag{117}$$

$$\overline{\Gamma}^{(1)}_{\mathrm{IK},11}=-\frac{\widetilde{\Lambda}^{(2)}_{\mathrm{IK},11}}{2\left(\frac{3-4\nu_0}{8(1-\nu_0)}+\chi^{(2)}_{\mathrm{II},11}\right)},\ \overline{\Gamma}^{(2)}_{\mathrm{IJ},11}=\frac{1}{4\left(\frac{3-4\nu_0}{8(1-\nu_0)}+\chi^{(2)}_{\mathrm{IJ},11}\right)} \tag{118}$$

$$\left\{\begin{matrix}\widetilde{\Lambda}^{(2)}_{\mathrm{I1},11}\\ \widetilde{\Lambda}^{(2)}_{\mathrm{I2},11}\end{matrix}\right\}=\left[\begin{matrix}\left[\begin{matrix}\frac{4\nu_0-1}{8(1-\nu_0)}+\chi^{(1)}_{11,11}\\ +2\left(\frac{3-4\nu_0}{8(1-\nu_0)}+\chi^{(2)}_{11,11}\right)\end{matrix}\right] & \frac{4\nu_0-1}{8(1-\nu_0)}+\chi^{(1)}_{21,11}\\ \frac{4\nu_0-1}{8(1-\nu_0)}+\chi^{(1)}_{12,11} & \left[\begin{matrix}\frac{4\nu_0-1}{8(1-\nu_0)}+\chi^{(1)}_{22,11}\\ +2\left(\frac{3-4\nu_0}{8(1-\nu_0)}+\chi^{(2)}_{22,11}\right)\end{matrix}\right]\end{matrix}\right]^{-1}\left\{\begin{matrix}\frac{4\nu_0-1}{8(1-\nu_0)}+\chi^{(1)}_{\mathrm{I1},11}\\ \frac{4\nu_0-1}{8(1-\nu_0)}+\chi^{(1)}_{\mathrm{I2},11}\end{matrix}\right\} \tag{119}$$

$$\begin{aligned}&\widetilde{\lambda}^{(1)'}=-\frac{\overline{\lambda}^{(1)'}}{4\overline{\mu}^{(1)'}\left(\overline{\lambda}^{(1)'}+2\overline{\mu}^{(1)'}\right)},\ \widetilde{\mu}^{(1)'}=\frac{1}{4\overline{\mu}^{(1)'}},\\ &\overline{\lambda}^{(1)'}=\frac{4\nu_0-1}{8(1-\nu_0)}+\lambda^{(1)'},\ \overline{\mu}^{(1)'}=\frac{3-4\nu_0}{8(1-\nu_0)}+\mu^{(1)'}\end{aligned} \tag{120}$$

$$\lambda^{(1)'}=\frac{\lambda_0\mu_1-\lambda_1\mu_0}{(\mu_1-\mu_0)[3(\lambda_1-\lambda_0)+2(\mu_1-\mu_0)]},\ \mu^{(1)'}=\frac{\mu_0}{2(\mu_1-\mu_0)} \tag{121}$$

$$\left(\chi^{(2)}\right)_{ijkl}=\chi^{(1)}_{\mathrm{IK},11}\delta_{ij}\delta_{kl}+\chi^{(2)}_{\mathrm{IJ},11}\left(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}\right) \tag{122}$$

$$\begin{aligned}&\chi^{(1)}_{\mathrm{IK},11}=\left[\frac{\sum_{\mathrm{m}=1}^{2}-\overline{\Lambda}^{(2)}_{\mathrm{Im},11}\lambda_0}{2\left(\mu^{(2)}_{\mathrm{II},11}-\mu_0\right)}+\frac{2\lambda_0}{4\left(\mu^{(2)}_{\mathrm{II},11}-\mu_0\right)}+\frac{-2\overline{\Lambda}^{(2)}_{\mathrm{IK},11}\mu_0}{2\left(\mu^{(2)}_{\mathrm{II},11}-\mu_0\right)}\right],\\ &\chi^{(2)}_{\mathrm{IJ},11}=\frac{2\mu_0}{4\left(\mu^{(2)}_{\mathrm{IJ},11}-\mu_0\right)}\end{aligned} \tag{123}$$

$$\left(\chi^{(3)}\right)_{ijkl}=\chi^{(1)}_{\mathrm{IK},12}\delta_{ij}\delta_{kl}+\chi^{(2)}_{\mathrm{IJ},12}\left(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}\right) \tag{124}$$

$$\chi^{(1)}_{\mathrm{IK},12} = \left[\frac{\sum_{m=1}^{2} -\overline{\Lambda}^{(2)}_{\mathrm{Im},12}\lambda_0}{2\left(\mu^{(2)}_{\mathrm{II},12}-\mu_0\right)} + \frac{2\lambda_0}{4\left(\mu^{(2)}_{\mathrm{II},12}-\mu_0\right)} + \frac{-2\overline{\Lambda}^{(2)}_{\mathrm{IK},12}\mu_0}{2\left(\mu^{(2)}_{\mathrm{II},12}-\mu_0\right)}\right],$$
$$\chi^{(2)}_{\mathrm{IJ},12} = \frac{2\mu_0}{4\left(\mu^{(2)}_{\mathrm{IJ},12}-\mu_0\right)} \tag{125}$$

$$\begin{Bmatrix}\overline{\Lambda}^{(2)}_{\mathrm{I1},11}\\ \overline{\Lambda}^{(2)}_{\mathrm{I2},11}\end{Bmatrix} = \begin{bmatrix}\left(\lambda^{(2)}_{11,11}-\lambda_0\right)+2\left(\mu^{(2)}_{11,11}-\mu_0\right) & \left(\lambda^{(2)}_{21,11}-\lambda_0\right)\\ \left(\lambda^{(2)}_{12,11}-\lambda_0\right) & \left(\lambda^{(2)}_{22,11}-\lambda_0\right)+2\left(\mu^{(2)}_{22,11}-\mu_0\right)\end{bmatrix}^{-1}\begin{Bmatrix}\lambda^{(2)}_{\mathrm{I1},11}-\lambda_0\\ \lambda^{(2)}_{\mathrm{I2},11}-\lambda_0\end{Bmatrix} \tag{126}$$

$$\begin{Bmatrix}\overline{\Lambda}^{(2)}_{\mathrm{I1},12}\\ \overline{\Lambda}^{(2)}_{\mathrm{I2},12}\end{Bmatrix} = \begin{bmatrix}\left(\lambda^{(2)}_{11,12}-\lambda_0\right)+2\left(\mu^{(2)}_{11,12}-\mu_0\right) & \left(\lambda^{(2)}_{21,12}-\lambda_0\right)\\ \left(\lambda^{(2)}_{12,12}-\lambda_0\right) & \left(\lambda^{(2)}_{22,12}-\lambda_0\right)+2\left(\mu^{(2)}_{22,12}-\mu_0\right)\end{bmatrix}^{-1}\begin{Bmatrix}\lambda^{(2)}_{\mathrm{I1},12}-\lambda_0\\ \lambda^{(2)}_{\mathrm{I2},12}-\lambda_0\end{Bmatrix} \tag{127}$$

Conclusions

In this chapter, a probabilistic micromechanical evolutionary elastoplastic-damage framework is proposed for fiber-reinforced ductile composites to characterize the process of interfacial partial debonding between the fibers and the matrix and to estimate the effects of progressive interfacial debonding on the overall elastoplastic-damage responses of composites. The damage evolution is represented by the debonding arc angle that is dependent on the external loading condition. Three different types of partial debonding modes are considered and analyzed. For each type of debonded fibers, the elastic equivalency is constructed in terms of the equivalent orthotropic stiffness tensor. That is, the debonded isotropic fibers are replaced by the orthotropic, yet perfectly bonded fibers or voids. Moreover, the evolutionary volume fractions of partially debonded fibers are characterized by Weibull's probabilistic approach. The first-order micromechanical approximation method of Ju and Chen (1994a, b, c), Ju and Zhang (1998a), and Ko and Ju (2012, 2013b) is utilized to estimate the effective stiffness tensor of the resultant multiphase composites. The proposed constitutive framework is suitable for accommodating general 2-D loading conditions. Finally, numerical simulations of the effective elastoplastic-damage behavior of fiber-reinforced composites are presented to illustrate the potential of the proposed partial debonding damage model under the transverse uniaxial and biaxial loading.

Future research is warranted to extend the present framework to accommodate the effects of *fiber coating* on the overall responses of composites. For example, Baker et al. (1999) studied the responses of SiC fibers to vacuum plasma spraying

(VPS) and vacuum hot pressing (VHP) during fabrication of titanium metal matrix composites. It is indicated that from the secondary electron SEM (scanning electron microscope) image, a C-coating layer exists between the Sigma 1140+ SiC fibers and the Ti–6Al–4V metal matrix, while a thin reaction layer exists between the C-coating layer and the matrix. In addition, the composites under the normal or tangential shear loading may fail on the weakest plane within the interfacial region, e.g., within the coating or the reaction products or at any sub-interfaces (i.e., fiber coating, coating reaction product, or reaction product matrix; cf. Gundel and Miracle 1998). In the case of *interphase* failure within the composites, the effects of fiber coating and the reaction product (layer) on the overall elastoplastic-damage responses should be taken into account. The effects of fiber coating and the thin reaction layer on the overall elastic properties and on local fields in fiber-reinforced composites have been studied by many researchers (Mikata and Taya 1985; Benveniste et al. 1989; Yang and Mal 1995; Lee and Mal 1997). However, the studies on the effects of fiber coating and the thin reaction layer on the overall elastoplastic-damage responses of composites are still limited. Therefore, based on present framework, it is desirable to develop a probabilistic micromechanical damage model to accommodate the effects of fiber coating on the overall responses of composites by means of functionally graded materials (FGMs).

References

R.K. Abu Al-Rub, G.Z. Voyiadjis, A finite strain plastic-damage model for high velocity impact using combined viscosity and gradient localization limiters: Part I-Theoretical formulation. Int. J. Damage Mech. **15**(4), 293–334 (2006)

A.M. Baker, P.S. Grant, M.L. Jenkins, The response of SiC fibres to vacuum plasma spraying and vacuum hot pressing during the fabrication of titanium matrix composites. J. Microsc. **196** (Pt 2), 162–174 (1999)

Y. Benveniste, G.J. Dvorak, T. Chen, Stress fields in composites with coated inclusions. Mech. Mater. **7**, 305–317 (1989)

J.G. Berryman, Estimates and rigorous bounds on pore-fluid enhanced shear modulus in poroelastic media with hard and soft anisotropy. Int. J. Damage Mech. **15**(2), 133–167 (2006)

M. Berveiller, A. Zaoui, An extension of the self-consistent scheme to plastically flowing polycrystals. J. Mech. Phys. Solids **26**, 325–344 (1979)

S.W. Case, K.L. Reifsnider, Micromechanical analysis of fiber fracture in unidirectional composite materials. Int. J. Solids Struct. **33**(26), 3795–3812 (1996)

M. Cheng, W. Chen, Modeling transverse behavior of Kevlar KM2 single fibers with deformation-induced damage. Int. J. Damage Mech. **15**(2), 121–132 (2006)

J.D. Eshelby, The determination of the elastic field of an ellipsoidal inclusion, and related problems. Proc. R. Soc. **A241**, 376–396 (1957)

D.B. Gundel, D.B. Miracle, The influence of interface structure and composition on the response of single-fiber SiC/Ti–6Al–4V composites to transverse tension. Appl. Compos. Mater. **5**, 95–108 (1998)

Z. Hashin, S. Shtrikman, A variational approach to the theory of the elastic behavior of multiphase materials. J. Mech. Phys. Solids **11**, 127–140 (1962)

L.H. He, C.W. Lim, Time-dependent interfacial sliding in fiber composites under longitudinal shear. Compos. Sci. Technol. **61**(4), 579–584 (2001)

C.H. Hsueh, Interfacial debonding and fiber pull-out stresses of fiber-reinforced composites. Mater. Sci. Eng. A **123**(1), 1–11 (1990)
J.W. Hutchinson, H.M. Jensen, Models of fiber debonding and pullout in brittle composites with friction. Mech. Mater. **9**, 139–163 (1990)
I. Jasiuk, Y. Tong, The effect of interface on the elastic stiffness of composites, Mechanics of Composite Materials and Structures, in J.N. Reddy, J.L.Teply (eds) *Proceedings of the 3rd Joint ASCE/ASME Mechanics Conference,* University of California, San Diego, La Jolla, California, 49–54 (1989)
J.W. Ju, A micromechanical damage model for uniaxial reinforced composites weakened by interfacial arc microcracks. J. Appl. Mech. **58**, 923–930 (1991a)
J.W. Ju, On two-dimensional self-consistent micromechanical damage models for brittle solids. Int. J. Solids Struct. **27**(2), 227–258 (1991b)
J.W. Ju, On micromechanical evolutionary damage models for polycrystalline ceramics. Int. J. Damage Mech. **5**(2), 113–137 (1996)
J.W. Ju, T.M. Chen, Micromechanics and effective moduli of elastic composites containing randomly dispersed ellipsoidal inhomogeneities. Acta Mech. **103**, 103–121 (1994a)
J.W. Ju, T.M. Chen, Effective elastic moduli of two-phase composites containing randomly dispersed spherical inhomogeneities. Acta Mech. **103**, 123–144 (1994b)
J.W. Ju, T.M. Chen, Micromechanics and effective elastoplastic behavior of two-phase metal matrix composites. ASME J. Eng. Mater. Technol. **116**, 310–318 (1994c)
J.W. Ju, T.M. Chen, Effective elastic moduli of two-dimensional brittle solids with interacting microcracks. Part I: basic formulations. J. Appl. Mech. ASME **61**, 349–357 (1994d)
J.W. Ju, T.M. Chen, Effective elastic moduli of two-dimensional brittle solids with interacting microcracks. Part II: evolutionary damage models. J. Appl. Mech. ASME **61**, 358–366 (1994e)
J.W. Ju, Y.F. Ko, Multi-level elastoplastic damage mechanics for elliptical fiber reinforced composites with evolutionary complete fiber debonding. Int. J. Damage Mech. **18**, 419–460 (2009)
J.W. Ju, X. Lee, Micromechanical damage models for brittle solids. Part I: tensile loadings. J. Eng. Mech. ASCE **117**(7), 1495–1515 (1991)
J.W. Ju, H.K. Lee, A micromechanical damage model for effective elastoplastic behavior of ductile matrix composites considering evolutionary complete particle debonding. Comput. Methods Appl. Mech. Eng. **183**, 201–222 (2000)
J.W. Ju, H.K. Lee, A micromechanical damage model for effective elastoplastic behavior of partially debonded ductile matrix composites. Int. J. Solids Struct. **38**, 6307–6332 (2001)
J.W. Ju, L.Z. Sun, A novel formulation for the exterior-point Eshelby's tensor of an ellipsoidal inclusion. J. Appl. Mech. ASME **66**, 570–574 (1999)
J.W. Ju, L.Z. Sun, Effective elastoplastic behavior of metal matrix composites containing randomly located aligned spheroidal inhomogeneities. Part I: micromechanics-based formulation. Int. J. Solids Struct. **38**, 183–201 (2001)
J.W. Ju, K.H. Tseng, A three-dimensional statistical micromechanical theory for brittle solids with interacting microcracks. Int. J. Damage Mech. **1**(1), 102–131 (1992)
J.W. Ju, K.H. Tseng, An improved two-dimensional micromechanical theory for brittle solids with many randomly located interacting microcracks. Int. J. Damage Mech. **4**(1), 23–57 (1995)
J.W. Ju, K.H. Tseng, Effective elastoplastic behavior of two-phase ductile matrix composites: a micromechanical framework. Int. J. Solids Struct. **33**, 4267–4291 (1996)
J.W. Ju, K.H. Tseng, Effective elastoplastic algorithms for ductile matrix composites. J. Eng. Mech. ASCE **123**, 260–266 (1997)
J.W. Ju, K. Yanase, Micromechanical elastoplastic mechanics for elliptical fiber-reinforced composites with progressive partial fiber debonding. Int. J. Damage Mech. **18**(7), 639–668 (2009)
J.W. Ju, X.D. Zhang, Micromechanics and effective transverse elastic moduli of composites with randomly located aligned circular fibers. Int. J. Solids Struct. **35**(9–10), 941–960 (1998a)
J.W. Ju, Y. Zhang, A thermomechanical model for airfield concrete pavement under transient high temperature loadings. Int. J. Damage Mech. **7**(1), 24–46 (1998b)

J.W. Ju, Y. Zhang, Axisymmetric thermomechanical constitutive and damage modeling for airfield concrete pavement under transient high temperatures. Mech. Mater. **29**, 307–323 (1998c)
J.W. Ju, X.D. Zhang, Effective elastoplastic behavior of ductile matrix composites containing randomly located aligned circular fibers. Int. J. Solids Struct. **38**, 4045–4069 (2001)
J.W. Ju, Y.F. Ko, H.N. Ruan, Effective elastoplastic damage mechanics for fiber reinforced composites with evolutionary complete fiber debonding. Int. J. Damage Mech. **15**(3), 237–265 (2006)
J.W. Ju, Y.F. Ko, H.N. Ruan, Effective elastoplastic damage mechanics for fiber reinforced composites with evolutionary partial fiber debonding. Int. J. Damage Mech. **17**, 493–537 (2008)
B.W. Kim, J.A. Nairn, Observations of fiber fracture and interfacial debonding phenomena using the fragmentation test in single fiber composites. J. Compos. Mater. **36**, 1825–1858 (2002)
Y.F. Ko, J.W. Ju, New higher-order bounds on effective transverse elastic moduli of three-phase fiber reinforced composites with randomly located and interacting aligned circular fibers. Acta Mech. **223**(11), 2437–2458 (2012)
Y.F. Ko, J.W. Ju, Effect of fiber-cracking on elastoplastic damage behavior of fiber-reinforced metal matrix composites. Int. J. Damage Mech. **22**(1), 56–79 (2013a)
Y.F. Ko, J.W. Ju, Effective transverse elastic moduli of three-phase hybrid fiber reinforced composites with randomly located and interacting aligned circular fibers of distinct elastic properties and sizes. Acta Mech. **224**(1), 157–182 (2013b)
X. Lee, J.W. Ju, Micromechanical damage models for brittle solids. Part II: compressive loadings. J. Eng. Mech. ASCE **117**(7), 1516–1537 (1991)
H.K. Lee, J.W. Ju, A three-dimensional stress analysis of a penny-shaped crack interacting with a spherical inclusion. Int. J. Damage Mech. **16**, 331–359 (2007)
H.K. Lee, J.W. Ju, 3-D micromechanics and effective moduli for brittle composites with randomly located interacting microcracks and inclusions. Int. J. Damage Mech. **17**, 377–417 (2008)
J. Lee, A. Mal, A volume integral equation technique for multiple inclusion and crack interaction problems. J. Appl. Mech. **64**, 23–31 (1997)
J. Lee, A. Mal, Characterization of matrix damage in metal matrix composites under transverse loads. Comput. Mech. **21**, 339–346 (1998)
V.C. Li, Y. Wang, S. Backer, Effect of inclining angle, bundling and surface treatment on synthetic fibre pull-out from a cement matrix. Composites **21**(2), 132–140 (1990)
S.H. Li, S.P. Shan, Z. Li, T. Mura, Micromechanical analysis of multiple fracture and evaluation of debonding behavior for fiber-reinforced composites. Int. J. Solids Struct. **30**(11), 1429–1459 (1993)
H.T. Liu, L.Z. Sun, Effects of thermal residual stresses on effective elastoplastic behavior of metal matrix composites. Int. J. Solids Struct. **41**, 2189–2203 (2004)
H.T. Liu, L.Z. Sun, J.W. Ju, An interfacial debonding model for particle-reinforced composites. Int. J. Damage Mech. **13**, 163–185 (2004)
H.T. Liu, L.Z. Sun, J.W. Ju, Elastoplastic modeling of progressive interfacial debonding for particle-reinforced metal matrix composites. Acta Mech. **181**(1–2), 1–17 (2006)
J. Lu, X. Zhang, Y.-W. Mai, A preliminary study on damage wave in elastic-brittle materials. Int. J. Damage Mech. **14**(2), 127–147 (2005)
D.B. Marshall, W.L. Morris, B.N. Cox, J. Graves, J.R. Porter, D. Kouris, R.K. Everett, Transverse strengths and failure mechanisms in Ti_3Al matrix composites. Acta Metallurgica et Materialia **42**, 2657–2673 (1994)
Y. Mikata, M. Taya, Stress field in coated continuous fiber composite subjected to thermomechanical loadings. J. Compos. Mater. **19**, 554–579 (1985)
T. Mori, K. Tanaka, Average stress in matrix and average elastic energy of materials with misfitting inclusions. Acta Metall. **21**, 571–574 (1973)
T. Mura, *Micromechanics of Defects in Solids*, 2nd edn. (Kluwer, Dordrecht, 1987)
S. Nemat-Nasser, M. Hori, *Micromechanics: Overall Properties of Heterogeneous Materials* (Elsevier Science Publisher B. V, Dordrecht, 1993)
R.P. Nimmer, R.J. Bankert, E.S. Russell, G.A. Smith, P.K. Wright, Micromechanical modeling of fiber/matrix interface effects in transversely loaded SiC/Ti-6-4 metal matrix composites. J. Compos. Technol. Res. JCTRER **13**, 3–13 (1991)

N.J. Pagano, G.P. Tandon, Modeling of imperfect bonding in fiber reinforced brittle matrix composites. Mech. Mater. **9**, 49–64 (1990)
G.H. Paulino, H.M. Yin, L.Z. Sun, Micromechanics-based interfacial debonding model for damage of functionally graded materials with particle interactions. Int. J. Damage Mech. **15** (3), 267–288 (2006)
R. Pereyra, Y.-L. Shen, Characterization of indentation-induced 'particle crowding' in metal matrix composites. Int. J. Damage Mech. **14**(3), 197–213 (2005)
J. Qu, Effects of slightly weakened interfaces on the overall elastic properties of composite materials. Mech. Mater. **14**, 269–281 (1993)
A.S. Sangani, G. Mo, Elastic interactions in particulate composites with perfect as well as imperfect interfaces. J. Mech. Phys. Solids **45**, 2001–2031 (1997)
P.S. Steif, Stiffness reduction due to fiber breakage. J. Compos. Mater. **17**, 153–172 (1984)
P.S. Steif, A model for kinking in fiber composites-I. Fiber breakage via micro-buckling. Int. J Solids Struct. **26**(5–6), 549–561 (1990)
L.Z. Sun, J.W. Ju, Effective elastoplastic behavior of metal matrix composites containing randomly located aligned spheroidal inhomogeneities. Part II: applications. Int. J. Solids Struct. **38**, 203–225 (2001)
L.Z. Sun, J.W. Ju, Elastoplastic modeling of metal matrix composites containing randomly located and oriented spheroidal particles. J. Appl. Mech. ASME **71**, 774–785 (2004)
L.Z. Sun, J.W. Ju, H.T. Liu, Elastoplastic modeling of metal matrix composites with evolutionary particle debonding. Mech. Mater. **35**, 559–569 (2003a)
L.Z. Sun, H.T. Liu, J.W. Ju, Effect of particle cracking on elastoplastic behavior of metal matrix composites. Int. J. Numer. Methods Eng. **56**, 2183–2198 (2003b)
G.P. Tandon, G.J. Weng, A theory of particle-reinforced plasticity. J. Appl. Mech. ASME **55**, 126–135 (1988)
K. Tohgo, G.J. Weng, A progress damage mechanics in particle-reinforced metal-matrix composites under high triaxial tension. J. Eng. Mater. Technol. **116**, 414–420 (1994)
G.Z. Voyiadjis, R.K. Abu Al-Rub, A finite strain plastic-damage model for high velocity impact using combined viscosity and gradient localization limiters: Part II-Numerical aspects and simulation. Int. J. Damage Mech. **15**(4), 335–373 (2006)
G.Z. Voyiadjis, D.H. Allen, *Damage and Interfacial Debonding in Composites*. Studies in Applied Mechanics, vol. 44 (Elsevier, Amsterdam, 1996). 275 p
W. Weibull, A statistical distribution function of wide applicability. J. Appl. Mech. **18**, 293–297 (1951)
F.C. Wong, A. Ait-Kadi, Analysis of particulate composite behavior based on non-linear elasticity and modulus degradation theory. J. Mater. Sci. **32**, 5019–5034 (1997)
Z.C. Xia, J.W. Hutchinson, A.G. Evans, B. Budiansky, On large scale sliding in fiber-reinforced composites. J. Mech. Phys. Solids **42**(7), 1139–1158 (1994)
L.R. Xu, A.J. Rosakis, Impact damage visualization of heterogeneous two-layer materials subjected to low-speed impact. Int. J. Damage Mech. **14**(3), 215–233 (2005)
K. Yanase, J.W. Ju, Effective elastic moduli of spherical particle reinforced composites containing imperfect interfaces. Int. J. Damage Mech. **21**(1), 97–127 (2012)
K. Yanase, J.W. Ju, Toughening behavior of unidirectional fiber reinforced composites containing a crack-like flaw: matrix crack without fiber break. Int. J. Damage Mech. **21**(1), 97–127 (2013)
R.B. Yang, A.K. Mal, The effective transverse moduli of a composite with degraded fiber-matrix interfaces. Int. J. Eng. Sci. **33**(11), 1623–1632 (1995)
Y.H. Zhao, G.J. Weng, Transversely isotropic moduli of two partially debonded composites. Int. J. Solids Struct. **34**, 493–507 (1997)
Y.H. Zhao, G.J. Weng, The effect of debonding angle on the reduction of effective moduli of particle and fiber-reinforced composites. J. Appl. Mech. **69**, 292–302 (2002)
Y. Zhou, H. Mahfuz, S. Jeelani, Numerical simulation for high strain rate failure process of unidirectional SiC_f-Al composites. Int. J. Damage Mech. **14**(4), 321–341 (2005)

New Strain-Energy Based Coupled Elastoplastic Damage-Healing Mechanics Accounting for Matric Suction Effect for Geomaterials

31

K. Y. Yuan and Jiann-Wen Woody Ju

Contents

Abstract

Innovative energy-based coupled elastoplastic hybrid isotropic damage-healing models for partially saturated soils have been developed and implemented for numerical simulation of soil moving processes. A class of elastoplastic constitutive ***damage-healing*** models, based on a continuum thermodynamic framework, is proposed within an initial elastic strain energy-based formulation. In particular, change of effective stress due to ***matric suction*** is considered, and the governing incremental damage and healing evolutions are coupled and characterized through the effective stress in conjunction with the hypothesis of strain equivalence. Further, plastic flow is introduced by means of an additive split of the stress tensor. Two characteristic energy norms of the tensile and compressive

K.Y. Yuan • J.-W.W. Ju (✉)
Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, USA
e-mail: yuanky@ucla.edu; juj@ucla.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_14

strain tensors, respectively, are introduced for the corresponding damage and healing mechanisms.

By incorporating micromechanics-motivated damage and healing characterizations, the proposed model and computational algorithms have been implemented to demonstrate the significant flexibility on numerical simulation of earth-pushing processes. Completely new computational algorithms are systematically developed based on the two-step ***operator splitting*** methodology. The elastic-damage-healing predictor and the plastic corrector are implemented within the existing RKPM (Reproducing Kernel Particle Method) meshfree codes. A numerical example under soil pushing is presented to illustrate the effect of matric suction for partially saturated soils.

Introduction

In the early 1980s, the development of constitutive equations for saturated soils has involved three main concepts, including the concept of effective stress from soil mechanics, the existence of coupled elastoplastic damage models from damage mechanics, and the theory of two-phase mixtures for a solid skeleton and a fluid. Some engineering applications and numerical demonstrations by finite element analysis and meshfree method were reported in the literature; cf. Sanavia et al. (2006), Wu et al. (2001), and Murakami et al. (2005).

On the other hand, an unsaturated soil is commonly defined as having three phases: (1) solids, (2) water, and (3) air. However, it may be more correct to recognize the existence of a fourth phase, namely, the air–water interface or contractile skin (Fredlund and Morgenstern 1978). From the standpoint of the volume–mass relations for an unsaturated soil, it is acceptable to consider the soil as a three-phase system since the volume of the contractile skin is small and its mass can be considered as part of the mass of water. Some rigorous framework to define the constitutive behavior of unsaturated soils for three phases has been developed by Alonso et al. (1990), Wheeler and Sivakumar (1995), Wheeler et al. (2002, 2003), Bolzon et al. (1996), Borja (2004), and Georgiadis et al. (2005). For the stress analysis of a multiphase continuum, it should be realized that the air–water interface behaves as an independent phase (Fredlund and Rahardio 1993). A comprehensive framework to define the constitutive behavior of unsaturated soils for four phases was proposed by Loret and Khalili (2000).

It is widely accepted that theories for correctly describing the states of stress and failure in unsaturated soils require two fundamental considerations. Firstly, both the ***net stress*** and ***matric suction*** in Bishop's effective stress need to be considered independently; cf. Bishop (1959), Bishop et al. (1960), Fredlund et al. (1978), Fredlund (1979), Gallipoli et al. (2003a, b), Gens and Alonso (1992), and Escario and Saez (1986). Secondly, plasticity or failure models such as the Mohr–Coulomb (Fredlund et al. 1978), the Cam-Clay (Alonso et al. 1990), or the Cap models (Simo et al. 1988; Kohler and Hofstetter 2008) must be modified for unsaturated (partially

saturated) soils due to the matric suction. Further, to tackle different engineering problems, material variables (e.g., grain size and grain size distribution), state variables (e.g., degree of saturation), and the consequent interparticle forces (suction-induced effective stress or suction stress) need to be considered as well for more comprehensive soil models.

At variance with our previous work by Ju et al. (2012a, b) and Ju and Yuan (2012), the major contribution of this chapter is to develop the ***initial elastic strain energy***-based hybrid isotropic elastoplastic damage-healing formulations for partially saturated soils which can account for the effect of matric suction and be implemented for the 2D earth moving simulation. In particular, since main interests are focus in simulating rapid earth-pushing activities and the qualitative features of our approach, the following assumptions are made to simplify both modeling and computation efforts:

1. The air is assumed to remain at atmospheric pressure during the earth moving processes.
2. The suction refers here to the matric suction which is the difference between the air and water pressures induced by the capillary tension. The difference in pressures induced by osmotic effects, i.e., ionic disequilibrium, is not considered here.
3. The matric suction and the parameter χ in Bishop's effective stress are only functions of saturation which can be obtained by numerical equations (based on experimental observations) for different types of soils. The effects of grain size or grain size distribution on the matric suction and the parameter χ are not considered here.
4. For soils, the wetting (saturation changing from low to high) or drying (saturation changing from high to low) process will induce different soil–water characteristic curve. This soil–water hysteresis is not considered here.
5. During the rapid earth moving processes, the water is assumed undrained. Therefore, the gravimetric water content can be considered as a *constant*, but the degree of saturation varies due to void ratio change when soil particles undergo complex gliding and rolling motions. Unlike the typical geotechnical problems such as bearing capacity, seepage and flow net, or slope stability, Darcy's law or Navier–Stokes equations are not applicable here for our specific earth moving processes. Instead, a numerical equation based on experimental data is used to calculate the degree of saturation at different stages during the earth moving activities.
6. The mass transfer due to vaporization and condensation is neglected during the rapid earth moving processes, although vaporization certainly occurs in actual experiments where water and air are in contact over a long period of time.

The remaining part of this chapter is organized as follows. The change of effective stress due to the matric suction and the Bishop's effective stress for unsaturated soils are introduced. The experimental data and some numerical equations which are employed to evaluate the parameters of Bishop's effective stress

and matric suction from literature are systematically presented. Subsequently, the coupled elastoplastic hybrid isotropic damage-healing formulations featuring the effect of matric suction for unsaturated soils are developed. The corresponding efficient two-step operator splitting methodology is rendered. Finally, numerical simulation of a 2D earth-pushing process is demonstrated and discussed in detail. Conclusions and future work are addressed as well.

Strain Energy-Based Coupled Elastoplastic Hybrid Isotropic Damage-Healing Models with Matric Suction Effect

Change of Effective Stress Due to Matric Suction

The effective stress approach by Bishop (1959) for unsaturated soil expanded Terzaghi's classic effective stress equation as follows:

$$\boldsymbol{\sigma}' = (\boldsymbol{\sigma} - \mathbf{u_a}) + \chi(\mathbf{u_a} - \mathbf{u_w}) \tag{1}$$

The difference $\boldsymbol{\sigma} - \mathbf{u_a}$ is referred to as the *net normal stress*, the difference $\mathbf{u_a} - \mathbf{u_w}$ is the matric suction, and the effective stress parameter χ is a material variable that is generally considered to vary between zero and unity. For $\chi = 0$, it corresponds to completely dry soil; for $\chi = 1$, it corresponds to fully saturated soil, which leads to Terzaghi's classic effective stress equation for saturated soil. Notice that the pore water pressure, $\mathbf{u_w}$, in saturated soil is generally compressive and isotropic. On the other hand, the pore water pressure in unsaturated soil is generally tensile and matric suction is always positive.

Furthermore, cohesion is the component of shear strength of soil that is independent of interparticle friction. For partially saturated soil, in addition to the electrostatic forces and cementing by minerals (e.g., Fe_2O_3, $CaCO_3$, NaCl, etc.), the partial loss or recovery of cohesion between granular soil particles due to water menisci is given by the change of matric suction. For instance, for loose, uncemented sand that is either completely dry or completely saturated, the cohesion in the Mohr–Coulomb failure criterion may be considered essentially equal to zero. However, if some water is added to the completely dry sand, considerable cohesive strength may exist due to matric suction. This apparent cohesion unique to unsaturated soil arises from *negative* pore water pressure and surface tension effects occurring at the interface of the pore water, pore air, and soil solids among the unsaturated soil grains.

Matric Suction

The soil suction is commonly termed "total suction" which has two components, namely, the matric suction and the osmotic suction. However, it is primarily the matric suction component ($\mathbf{u_a} - \mathbf{u_w}$) that governs the engineering behavior of unsaturated soils in the lower suction range encountered in most field situations (Vanapalli et al. 1996). Moreover, laboratory data indicate that a change in total

suction is essentially equivalent to a change in matric suction where the water contents are less than the residual value (Krahn and Fredlund 1972). Some typical values of matric suction for different types of soils can be found in the literature; cf. Fredlund and Rahardio (1993), Lu and Likos (2004), and Mitchell and Soga (2005). In general, the latter values range from 0 to 1,000 KPa.

Disregarding pore air pressure, the contribution of pore water pressure to total stress depends on the degree of saturation and pore size distribution. Based on the studies on granular physics, four states of saturation can be characterized as follows (Mitarai and Nori 2006):

1. Pendular state: Soil particles are held together by water bridge at their contact points.
2. Funicular state: Some pores are fully saturated by water, but there still remain voids filled with air.
3. Capillary state: All voids between soil particles are filled with water, but the surface water is drawn back into the pore under capillary action.
4. Slurry state: Soil particles are fully immersed in water and the surface of water is convex, i.e., no capillary action at the surface.

Mitarai and Nori (2006) is referred for the schematic diagram and physical description of these four regimes. To macroscopically estimate the matric suction in different saturation states, it is necessary to introduce the soil–water characteristic curve (SWCC). The SWCC is a fundamental constitutive relationship in unsaturated soil mechanics, which can be employed to predict unsaturated soil properties such as the hydraulic conductivity, diffusivity, adsorption, shear strength, and volume change. In general terms, the SWCC describes the relationship between the matric suction and water content (or degree of saturation).

Numerous mathematical models were proposed for the soil–water characteristic curve, and various key equations were summarized by Zapata et al. (2000) and Nishimura et al. (2006). In particular, Fredlund and Xing (1994) proposed the following relationship between the volumetric water content and the matric suction:

$$\theta = \theta_s \left[\frac{1}{\ln[e + (s/a)^n]} \right]^m \tag{2}$$

where θ is the volumetric water content, θ_s represents the *saturated* volumetric water content, s is the matric suction, e is the exponential number (approximately 2.718281828), and (m, n, a) are material parameters. Figures 1, 2, and 3 exhibit the plots of Eq. 2 with different sets of parameters (m, n, a).

Parameter in Bishop's Effective Stress

Loret and Khalili (2000) stated that a properly defined effective stress is an efficient tool in the qualitative and quantitative descriptions of the behavior of unsaturated soils subject to the following two provisos: (1) that the parameter χ is evaluated with sufficient accuracy and (2) that the observed stiffening effect due to suction is

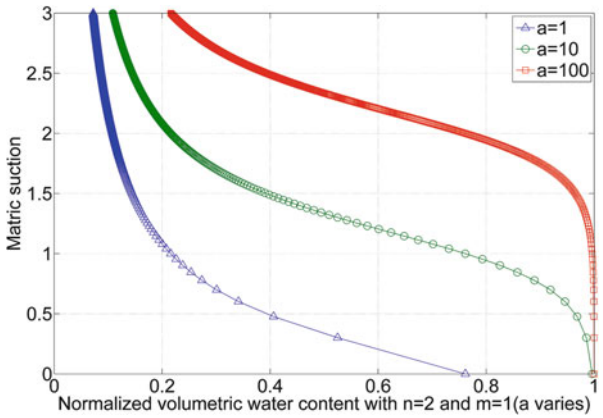


Fig. 1 The sample plots of Eq. 2 with $n = 2$ and $m = 1$ (with varying a)

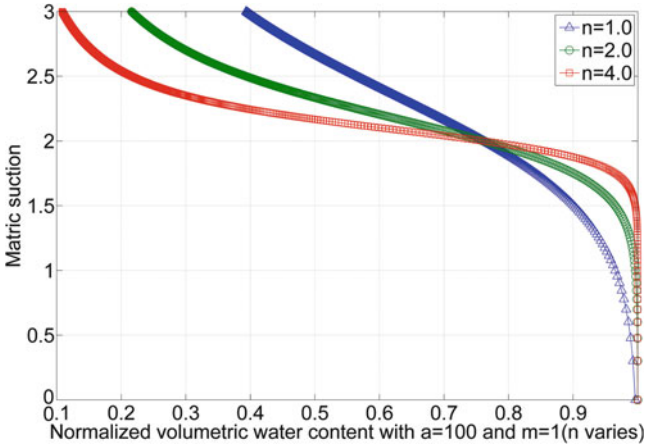


Fig. 2 The sample plots of Eq. 2 with $a = 100$ and $m = 1$ (with varying n)

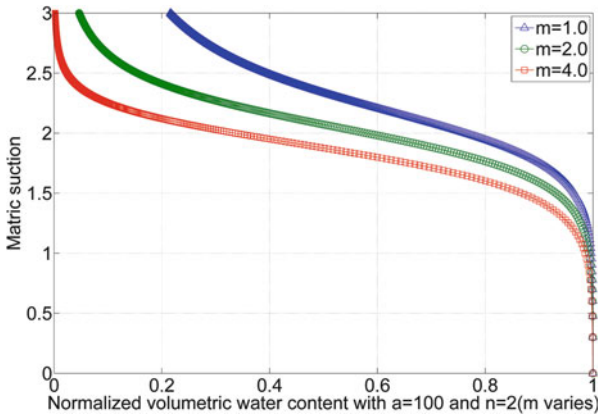


Fig. 3 The sample plots of Eq. 2 with $a = 100$ and $n = 2$ (with varying m)

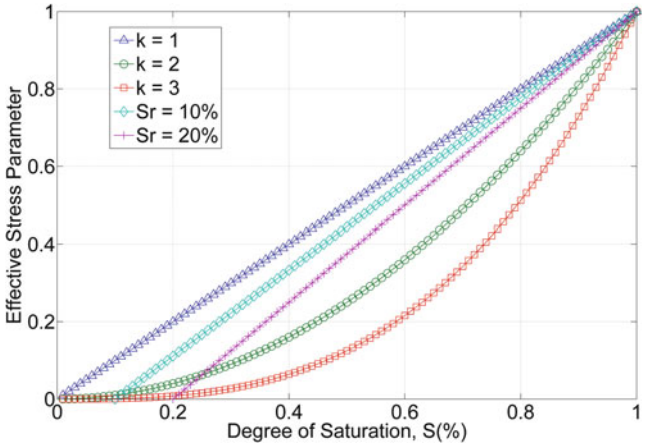


Fig. 4 Various forms for the effective stress parameter χ as functions of the degree of saturation

also accounted for. The material variable χ in Eq. 1 is captured by its strong dependency on the degree of pore water saturation. Moreover, the nature of χ can be described systematically from both the microscopic and macroscopic perspective; cf. Lu and Likos (2004).

The validity of several forms of χ as a function of the degree of saturation was also examined by Vanapalli and Fredlund (2000) using a series of shear strength test results for statically compacted mixtures of clay, silt, and sand from Escario et al. (1989). For the matric suction ranging between 0 and 1,500 KPa, the following two forms showed good fit to the experimental results. The first form reads

$$\chi = S^k = \left(\frac{\theta}{\theta_s}\right)^k \tag{3}$$

where S is the degree of saturation and k is a fitting parameter optimized to obtain the best fit between the measured and predicted values. The second form reads

$$\chi = \frac{S - S_r}{1 - S_r} = \frac{\theta - \theta_r}{\theta_s - \theta_r} \tag{4}$$

where θ_r is the residual volumetric water content and S_r is the residual degree of saturation. The nature of Eqs. 3 and 4 is illustrated in Fig. 4 for several values of k and S_r.

Determination of Saturation

From previous sections, it is shown that the matric suction and Bishop's effective stress parameter χ can be represented by the degree of saturation. However, as mentioned earlier, the continuity equation of pore water with Darcy's law or

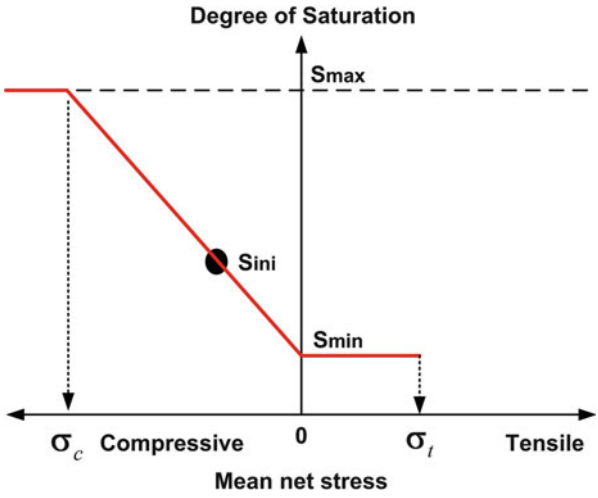


Fig. 5 A schematic plot of the relationship between the degree of saturation and the mean net stress

Navier–Stokes equations for unsaturated soils are not applicable here due to the instantaneous and complex rolling and gliding motions of soil particles during earth-pushing processes. In this section, some physically reasonable assumptions which are motivated by the literatures are made to calculate the degree of saturation at different time step for our specific earth-pushing problems.

Gallipoli et al. (2003b) suggested an improved form of expression for the variation of degree of saturation, accounting for the influence of changes in void ratio. The experimental data exhibits that for certain matric suction, the degree of saturation is ***monotonically increasing*** with the mean net stress. Motivated by these experimental observations, the following assumptions are proposed to simplify modeling and computation efforts:

1. The degree of saturation is a function of the mean net stress, gravimetric water content, and soil types.
2. For a certain gravimetric water content and soil type, the degree of saturation is monotonically increasing (linearly or nonlinearly depending on the soil types) with respect to the mean net stress.

For simplicity, it is assumed that the relationship between the mean net stress and the degree of saturation is linear as exhibited in Fig. 5. Here, in Fig. 5, S_{ini} is the assumed initial degree of saturation. When the mean net stress (compression) is increased, the degree of saturation increases linearly. Once the assumed critical compressive mean net stress $\boldsymbol{\sigma_c}$ is reached, it is assumed that the saturation will reach a maximum value, $S_{\max}$. On the other hand, if the mean net stress is zero or tensile, the soil will reach a minimum constant value of saturation, $S_{\min}$, and the effect of matric suction will be eliminated once the tensile mean net stress is greater than the maximum matric suction, $\boldsymbol{\sigma_t}$.

A Coupled Hybrid Isotropic Formulation for Partially Saturated Soils

The proposed ***initial elastic strain energy***-based hybrid isotropic damage-healing model for partially saturated soils is under the hypothesis that incremental damage and healing in the soils are directly linked to the history of total strains. The notion of effective stress along with the hypothesis of strain equivalence then follows from the assumed form of free energy. Attention is focused on the modification of effective stress due to the matric suction.

The Additive Stress Split

The plastic flow is introduced by means of an additive split of the stress tensor into the initial and inelastic parts:

$$\overline{\boldsymbol{\sigma}} = \frac{\partial \Psi^0(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}} - \overline{\boldsymbol{\sigma}}^{\mathbf{p}} - [u_a - \chi(S)(u_a - u_w)]\underset{\sim}{\mathbf{1}} \tag{5}$$

where $\overline{\boldsymbol{\sigma}}$ is the effective (undamaged) stress, $\boldsymbol{\varepsilon}$ is the total strain, $\Psi^0(\boldsymbol{\varepsilon})$ is the initial elastic stored energy function of the undamaged soil, $\overline{\boldsymbol{\sigma}}^{\mathbf{p}}$ is the effective (undamaged) plastic relaxation stress, u_a represents the pore air pressure, and u_w is the pore water pressure. Moreover, $\chi(S)$ is a function of the degree of saturation and represents the Bishop's effective stress parameter, and $\underset{\sim}{\mathbf{1}}$ denotes the second-order identity tensor. For the linear elasticity,

$$\Psi^0(\boldsymbol{\varepsilon}) = \frac{1}{2}\boldsymbol{\varepsilon} : \mathbf{C}^0 : \boldsymbol{\varepsilon} \tag{6}$$

where $\mathbf{C}^0$ denotes the linear elasticity tensor. Equation 5 then can be written as

$$\overline{\boldsymbol{\sigma}} = \mathbf{C}^0 : \boldsymbol{\varepsilon} - \overline{\boldsymbol{\sigma}}^{\mathbf{p}} - [u_a - \chi(S)(u_a - u_w)]\underset{\sim}{\mathbf{1}} \tag{7}$$

Thermodynamic Basis

In order to introduce both the net effect of damage-healing and plastic flow processes, a free energy potential of the following form is proposed:

$$\Psi(\boldsymbol{\varepsilon}, \boldsymbol{\sigma}^{\mathbf{p}}, \mathbf{q}, d^{\text{net}}) \equiv (1 - d^{\text{net}})\Psi^0(\boldsymbol{\varepsilon}) - \boldsymbol{\varepsilon} : \boldsymbol{\sigma}^{\mathbf{p}} + \Xi(\mathbf{q}, \boldsymbol{\sigma}^{\mathbf{p}}) \tag{8}$$

where $\boldsymbol{\varepsilon}$ denotes the total strain tensor, $\boldsymbol{\sigma}^{\mathbf{p}}$ is the plastic relaxation stress tensor, and $\mathbf{q}$ is a suitable set of internal (plastic) variables. Here, d^{net} is the isotropic scalar variable of net (combined) effect of damage-healing which depends on the damage variable d, the incremental damage variable Δd, and the incremental healing variable ΔR, all varying between 0 and 1 in numerical values. $\Psi^0(\boldsymbol{\varepsilon})$ is the initial elastic stored energy function of the undamaged material and $\Xi(\mathbf{q}, \boldsymbol{\sigma}^{\mathbf{p}})$ is the plastic potential function.

Confining our attention to the purely mechanical theory, the Clausius–Duhem (reduced dissipation) inequality (Coleman and Gurtin 1967) (for purely mechanical isothermal theory) takes the form for any admissible process

$$-\dot{\Psi} + \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} \geq 0 \tag{9}$$

By taking the time derivative of Eq. 8, substituting it into Eq. 9, and making use of standard arguments (the Gurtin-Coleman argument) along with the additional assumption that the net effect of damage-healing and plastic unloading are elastic processes, the stress–strain constitutive law can be obtained by

$$\begin{aligned} \boldsymbol{\sigma} &= \frac{\partial \Psi(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}} = (1 - d^{\text{net}}) \frac{\partial \Psi^0(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}} - \boldsymbol{\sigma}^{\mathbf{p}} \\ &= (1 - d^{\text{net}}) \left\{ \overline{\boldsymbol{\sigma}} + [u_a - \chi(S)(u_a - u_w)] \underline{\mathbf{1}} \right\} \end{aligned} \tag{10}$$

and the dissipative inequalities

$$-\frac{\partial \Xi}{\partial \mathbf{q}} \bullet \dot{\mathbf{q}} - \left(\frac{\partial \Xi}{\partial \boldsymbol{\sigma}^{\mathbf{p}}} - \boldsymbol{\varepsilon} \right) : \dot{\boldsymbol{\sigma}}^{\mathbf{p}} \geq 0 \tag{11}$$

$$\Psi^0(\boldsymbol{\varepsilon})\, \dot{d}^{\text{net}} - \frac{\partial \Xi}{\partial \mathbf{q}} \bullet \dot{\mathbf{q}} - \left(\frac{\partial \Xi}{\partial \boldsymbol{\sigma}^{\mathbf{p}}} - \boldsymbol{\varepsilon} \right) : \dot{\boldsymbol{\sigma}}^{\mathbf{p}} \geq 0 \tag{12}$$

It follows from Eq. 10 that within the present strain space formulation, the stress tensor is split into the elastic-damage-healing with suction effect and the plastic relaxation parts. From Eqs. 11 and 12, it shows that the dissipative energy by the plasticity itself is positive, and if damage-healing effect with suction effect is involved, the sum of them (the damage-healing effect and the plasticity) is also positive. It is also clear from Eqs. 10 to 12 that the present framework is capable of accommodating general (nonlinear) elastic response and general plastic response.

The potential $\Xi(\mathbf{q}, \boldsymbol{\sigma}^{\mathbf{p}})$ is linked to the plastic dissipation. Its role is such that inequality Eq. 11 is satisfied for arbitrary processes. Note that it is assumed $\Xi(\mathbf{q}, \boldsymbol{\sigma}^{\mathbf{p}})$ independent of d^{net}. From Eq. 8, it then follows that

$$-Y \equiv -\frac{\partial \Psi(\boldsymbol{\varepsilon}, \boldsymbol{\sigma}^{\mathbf{p}}, \mathbf{q}, d^{\text{net}})}{\partial d^{\text{net}}} = \Psi^0(\boldsymbol{\varepsilon}) \tag{13}$$

Hence, the initial (undamaged) elastic strain energy $\Psi^0(\boldsymbol{\varepsilon})$ is the thermodynamic force Y conjugate to the net damage-healing variable d^{net}.

Characterization of Initial Elastic Strain Energy-Based Hybrid Isotropic Damage Evolution

Firstly, the progressive degradation of mechanical properties of soils due to damage is characterized by means of a simple isotropic damage mechanism. To this effect,

the notion of equivalent tensile strain ξ^+ is ascribed as the (undamaged) energy norm of the tensile strain tensor. The isotropic damage mechanism is called ***hybrid***, since the computation of this equivalent tensile strain involves the principle tensile direction of total strain. Accordingly, motivated by Eq. 13, it is set that

$$\xi^+ \equiv \sqrt{\Psi^0(\boldsymbol{\varepsilon}^+)} = \sqrt{\frac{1}{2}\boldsymbol{\varepsilon}^+ : \mathbf{C}^0 : \boldsymbol{\varepsilon}^+} \tag{14}$$

where $\boldsymbol{\varepsilon}^+ \equiv \mathbf{P}^+ : \boldsymbol{\varepsilon}$. The $\boldsymbol{\varepsilon}$ denotes the total strain, and the fourth-order tensor $\mathbf{P}^+$ denotes the mode I positive (tensile) projection tensor with components

$$\mathbf{P}^+_{ijkl}(\boldsymbol{\varepsilon}) \equiv \frac{1}{2}\left(\mathbf{Q}^+_{ik}\ \mathbf{Q}^+_{jl} + \mathbf{Q}^+_{il}\ \mathbf{Q}^+_{jk}\right) \tag{15}$$

where $\mathbf{Q}^+ \equiv \sum_{i=1}^{2} \hat{H}(\boldsymbol{\varepsilon}_i)\mathbf{p}_i \otimes \mathbf{p}_i$; $\boldsymbol{\varepsilon} = \sum_{i=1}^{2} \boldsymbol{\varepsilon}_i \mathbf{p}_i \otimes \mathbf{p}_i$, $\|\mathbf{p}_i\| = 1$ (for 2D simulations in this chapter). Here, $\boldsymbol{\varepsilon}_i$ is the i-th principal strain, $\mathbf{p}_i$ is the i-th corresponding unit vector in the principal direction, and $\hat{H}(\boldsymbol{\varepsilon}_i)$ is the smoothed Heaviside function.

Then, the state of damage in soils by means of a damage criterion $\phi^d(\xi_t^+, g_t) \leq 0$ is characterized, formulated in the strain space, with the following functional form:

$$\phi^d\left(\xi_t^+, g_t\right) \equiv \xi_t^+ - g_t \leq 0 \quad t \in R_+ \tag{16}$$

Here, the subscript t refers to a value at the current time $t \in R_+$, and g_t is the damage threshold at the current time t. Note that g_t will be numerically lowered due to the potential incremental healing from the previous time step. If g_0 denotes the initial damage threshold before any loading is applied, it must have $g_t \geq g_0$. Here, g_0 is considered to be a characteristic material property. Condition (16) then states that damage in soils is initiated when the energy norm of the tensile strain tensor ξ_t^+ exceeds the initial damage threshold g_0. For the isotropic damage case, the evolutions of the damage variable d and the damage threshold g_t are defined by the rate equations

$$\dot{d}_t = \dot{\mu} H\left(\xi_t^+, d_t\right) \tag{17}$$

$$\dot{g} = \dot{\mu} \tag{18}$$

where $\dot{\mu} \geq 0$ is a damage consistency parameter that defines damage loading/unloading conditions according to the Kuhn–Tucker relations:

$$\dot{\mu} \geq 0, \quad \phi^d\left(\xi_t^+, g_t\right) \leq 0, \quad \dot{\mu}\, \phi^d\left(\xi_t^+, g_t\right) = 0 \tag{19}$$

Moreover, H in Eq. 17 signifies the damage hardening function. The conditions (19) are standard for problems involving unilateral constraint. If $\phi^d(\xi_t^+, g_t) < 0$, the

damage criterion is not satisfied and by conditions (19), $\dot{\mu} = 0$; hence, the damage rule (19) implies that $\dot{d} = 0$ and no further damage takes place. If, on the other hand, $\dot{\mu} > 0$, that is, further damage ("tensile loading") is taking place, conditions (19) implies that $\phi^d(\xi_t^+, g_t) = 0$. In this event, the value of $\dot{\mu}$ is determined by the damage consistency condition, i.e.,

$$\phi^d\left(\xi_t^+, g_t\right) = \dot{\phi}^d\left(\xi_t^+, g_t\right) = 0 \Rightarrow \dot{\mu} = \dot{\xi}^+ \tag{20}$$

so that g_t is defined by the expression

$$g_t = \max\left(g_0, \max_{s \in (-\infty, t)} \xi_s^+\right) - g_{h,(t-1)} \tag{21}$$

where $g_{h(t-1)}$ is the reduced value for the current damage threshold due to the incremental healing (if any) from the previous time step.

If $H(\xi_t^+, d_t)$ in condition (17) is independent of $\dot{d}_t$, the above formulation may be rephrased as follows: let $G : R \to R_+$ be such that $H(\xi_t^+) \equiv \partial G(\xi_t^+)/\partial \xi_t^+$. $G(\bullet)$ is assumed monotonic. A damage criterion entirely equivalent to condition (16) is given by $\overline{\phi}^d\left(\xi_t^+, g_t\right) \equiv G\left(\xi_t^+\right) - G(g_t) \leq 0$. The flow rule and loading/unloading conditions then become

$$\dot{d}_t = \dot{\mu} \frac{\partial \overline{\phi}^d\left(\xi_t^+, g_t\right)}{\partial \xi_t^+}, \quad g_t = \dot{\mu} \tag{22}$$

$$\dot{\mu} \geq 0, \quad \overline{\phi}^d\left(\xi_t^+, g_t\right) \leq 0, \quad \dot{\mu}\overline{\phi}^d\left(\xi_t^+, g_t\right) = 0 \tag{23}$$

Conditions (22) and (23) are simply the Kuhn–Tucker optimality conditions of a principle of maximum damage dissipation.

Characterization of Initial Elastic Strain Energy-Based Hybrid Isotropic Healing Evolution

Similar to the characterization of damage in previous section, the progressive recovery of mechanical properties of soils due to healing is characterized by means of a simple isotropic healing mechanism. The notion of equivalent compressive strain ξ^- is used as the energy norm of the compressive strain tensor. Accordingly, it is set that

$$\xi^- \equiv \sqrt{\Psi^0(\boldsymbol{\varepsilon}^-)} = \sqrt{\frac{1}{2}\boldsymbol{\varepsilon}^- : \mathbf{C}^0 : \boldsymbol{\varepsilon}^-} \tag{24}$$

where $\boldsymbol{\varepsilon}^- \equiv \mathbf{P}^- : \boldsymbol{\varepsilon}$. The fourth-order tensor $\mathbf{P}^-$ denotes the mode I negative (compressive) projection tensor with components

$$\mathbf{P}^-_{ijkl}(\boldsymbol{\varepsilon}) \equiv \frac{1}{2}\left(\mathbf{Q}^-_{ik}\ \mathbf{Q}^-_{jl} + \mathbf{Q}^-_{il}\ \mathbf{Q}^-_{jk}\right) \tag{25}$$

where $\mathbf{Q}^- = \underset{\sim}{\mathbf{1}} - \mathbf{Q}^+$.

The state of healing in soils by using a healing criterion $\phi^h(\xi_t^-, r_t) \leq 0$ is also characterized, formulated in the strain space, with the following functional form:

$$\phi^h\left(\xi_t^-, r_t\right) \equiv \xi_t^- - r_t \leq 0 \qquad t \in R_+ \tag{26}$$

Here, r_t is the healing threshold at the current time t. Note that r_t will be numerically lowered due to the incremental damage from the previous time step. If r_0 denotes the initial healing threshold before any loading is applied, it must have $r_t \geq r_0$. Further, r_0 is considered as a material property. Condition (26) then states that healing in the material is initiated when the energy norm of the compressive strain tensor ξ_t^- exceeds the initial healing threshold r_0. For the isotropic healing case, it is defined the evolution of the healing variable R and the healing threshold by the rate equations

$$\dot{R}_t = \dot{\zeta}\ Z\left(\xi_t^-, R_t\right) \tag{27}$$

$$\dot{r} = \dot{\zeta} \tag{28}$$

where $\dot{\zeta} \geq 0$ is a healing consistency parameter that defines healing loading/unloading conditions according to the Kuhn–Tucker relations

$$\dot{\zeta} \geq 0, \quad \phi^h\left(\xi_t^-, r_t\right) \leq 0, \quad \dot{\zeta}\, \phi^h\left(\xi_t^-, r_t\right) = 0 \tag{29}$$

Further, Z in Eq. 27 renders the healing hardening function. Conditions (29) are standard for problems involving unilateral constraint. If $\phi^h(\xi_t^-, r_t) < 0$, the healing criterion is not satisfied and by conditions (29), $\dot{\zeta} = 0$; hence, the healing rule (27) implies that $\dot{R} = 0$ and no further healing takes place. If, on the other hand, $\dot{\zeta} > 0$, that is, further healing ("compressive loading") is taking place, conditions (29) implies that $\phi^h(\xi_t^-, r_t) = 0$. In this event, the value of $\dot{\zeta}$ is determined by the healing consistency condition, i.e.,

$$\phi^h\left(\xi_t^-, r_t\right) = \dot{\phi}^h\left(\xi_t^-, r_t\right) = 0 \Rightarrow \quad \dot{\zeta} = \dot{\xi}^- \tag{30}$$

so that r_t is given by the expression

$$r_t = \max\left(r_0, \max_{s \in (-\infty, t)} \xi_s^-\right) - r_{d,(t-1)} \tag{31}$$

where $r_{d,(t-1)}$ is the reduced value for the current healing threshold due to the incremental damage (if any) from the previous time step.

If $Z(\xi_t^-, R_t)$ in condition (27) is independent of $\dot{R}_t$, the above formulation may be rephrased as follows: let $G^* : R \to R_+$ be such that $Z(\xi_t^-) \equiv \partial G^*(\xi_t^-)/\partial \xi_t^-$. $G^*(\bullet)$ is assumed monotonic. A healing criterion entirely equivalent to conditions (26) is given by $\overline{\phi}^h\left(\xi_t^-, r_t\right) \equiv G^*\left(\xi_t^-\right) - G^*(r_t) \leq 0$. The flow rule and loading/unloading conditions then become

$$\dot{R}_t = \dot{\zeta}\frac{\partial \overline{\phi}^h\left(\xi_t^-, r_t\right)}{\partial \xi_t^-}, \quad r_t = \dot{\zeta} \tag{32}$$

$$\dot{\zeta} \geq 0, \quad \overline{\phi}^h\left(\xi_t^-, r_t\right) \leq 0, \quad \dot{\zeta}\overline{\phi}^h\left(\xi_t^-, r_t\right) = 0 \tag{33}$$

Net (Combined) Effect of the Hybrid Isotropic Damage and Healing

In the previous work (Ju et al. 2012a, b; Ju and Yuan 2012), it has been shown that the following equation for the net effect of damage and healing is physically incorrect:

$$d^{\text{net}} = d \ (1 - R) \tag{34}$$

The net effect of damage and healing mechanism for new energy-based model is identical to the micromechanics-motivated scalar incremental form as expressed in Eq. 33 of previous work (Ju et al. 2012a, b).

Modification of the Drucker–Prager Model

For saturated soil, shear strength is commonly described by the Mohr–Coulomb failure criterion, which defines shear strength in terms of the material variables ϕ' and **c**' as

$$\boldsymbol{\tau}_{\mathbf{f}} = \mathbf{c}' + (\boldsymbol{\sigma} - \mathbf{u}_{\mathbf{w}})_f \tan\phi' \tag{35}$$

where $\boldsymbol{\tau}_{\mathbf{f}}$ is the shear stress on the failure plane, c' is the effective cohesion, $(\boldsymbol{\sigma} - \mathbf{u}_{\mathbf{w}})$ is the effective normal stress, and ϕ' is the effective angle of internal friction.

For unsaturated soil, modern experimental studies regarding the shear strength date back to 1950s and 1960s. Inspection of Blight's triaxial testing results (Blight 1967) and Escario's direct shear test results (Escario et al. 1980) demonstrates two general trends in the shear strength behavior of unsaturated soil. Firstly, the shear strength of unsaturated soil generally increases as net normal stress increase. Secondly, emerging from the triaxial and direct shear testing results is that shear strength increases as applied matric suction increases. Fredlund et al. (1978) formulated an extended Mohr–Coulomb criterion to describe the shear strength behavior of unsaturated soil. The failure envelope is a planar surface in the stress space of the stress state variables $\boldsymbol{\sigma} - \mathbf{u}_{\mathbf{a}}$ and $\mathbf{u}_{\mathbf{a}} - \mathbf{u}_{\mathbf{w}}$; the shear stress $\boldsymbol{\tau}$ may be written as (Lu and Likos 2004)

$$\boldsymbol{\tau}_{\mathbf{f}} = \mathbf{c}' + (\boldsymbol{\sigma} - \mathbf{u}_{\mathbf{a}})_f \tan\phi' + (\mathbf{u}_{\mathbf{a}} - \mathbf{u}_{\mathbf{w}}) \tan\phi^b \tag{36}$$

where **c**' is the cohesion at zero matric suction and zero net normal stress, $\mathbf{u}_{\mathbf{a}}$ is the pore air pressure, $\mathbf{u}_{\mathbf{w}}$ is the pore water pressure, ϕ ' is the angle of internal friction associated with the net normal stress variable, and ϕ^b is an internal friction angle associated with the matric suction that describes the rate of increase in shear strength relative to matric suction. Some experimental data of (**c** ', ϕ ', ϕ^b) for a wide variety of soil types can be found in the literature; cf. Fredlund and Rahardio (1993), and Lu and Likos (2004). Furthermore, the Drucker–Prager criterion can be written in the form

$$f(\mathrm{I}_1, J_2) = \sqrt{J_2} - \alpha \mathrm{I}_1 - k = 0 \tag{37}$$

where I_1 is the first invariant of the Cauchy stress tensor, J_2 is the second invariant of the deviatoric stress tensor, and the two parameters α and k are (positive) material constants. Motivated by the discussion of Mohr–Coulomb failure criterion for unsaturated soil proposed by Fredlund et al. (1978), Kohler and Hofstetter (2008) proposed the extension of a cap model to describe the material behavior of partially saturated soils. Ignoring the effect of the third invariant of the deviatoric stress tensor, the shear failure surface (Drucker–Prager model) can be described as follows:

$$f\left(\mathbf{I}_1^{'}, J_2\right) = \sqrt{J_2} - \alpha' \mathbf{I}_1^{'} - k' - \lambda(\mathbf{u_a} - \mathbf{u_w}) = 0 \tag{38}$$

where $\mathbf{I}_1^{'}$ is the first invariant of the net stress tensor, $\mathbf{u_a}$ is the pore air pressure, $\mathbf{u_w}$ is the pore water pressure, and (α', k', λ) are material constants.

Computational Algorithms: Two-Step Operator Splitting

In the previous section, initial elastic strain energy-based hybrid isotropic damage-healing formulations are developed based on the effective stress concept for partially saturated soil. In this section, computational aspects are focused in details on the proposed models within the context of numerical method. More precisely, the attention is focused on the following local elastoplastic damage-healing rate constitutive equations:

$$\begin{aligned}
&\dot{\varepsilon} = \nabla^s \dot{\mathrm{u}}(t) \\
&\begin{cases} \dot{d}_t = \dot{\mu}\ H\left(\xi_t^+, d_t\right) \\ \dot{g} = \dot{\mu} \end{cases} \\
&\dot{\mu} \geq 0, \quad \phi^d\left(\xi_t^+, g_t\right) \leq 0, \quad \dot{\mu}\, \phi^d\left(\xi_t^+, g_t\right) = 0 \\
&\begin{cases} \dot{R}_t = \dot{\zeta}\ Z\left(\xi_t^-, R_t\right) \\ \dot{r} = \dot{\zeta} \end{cases} \\
&\dot{\zeta} \geq 0, \quad \phi^h\left(\xi_t^-, r_t\right) \leq 0, \quad \dot{\zeta}\, \phi^h\left(\xi_t^-, r_t\right) = 0 \\
&\dot{d}^{\mathrm{net}} = \dot{d} - \dot{R}\ \hat{d} \\
&\dot{\overline{\sigma}} = \frac{d}{dt}\left[\frac{\partial \Psi^0(\varepsilon)}{\partial \varepsilon}\right] - \dot{\overline{\sigma}}^p - \frac{d}{dt}[u_a - \chi(S)(u_a - u_w)]\underset{\sim}{1} \\
&\begin{cases} \dot{\overline{\sigma}}^{\mathbf{p}} = \dot{\lambda}\dfrac{\partial f}{\partial \varepsilon}\left(\dfrac{\partial \Psi^0(\varepsilon)}{\partial \varepsilon} - \overline{\sigma}^p - [u_a - \chi(S)(u_a - u_w)]\underset{\sim}{1}, \mathbf{q}\right) & \text{(associative flow rule)} \\ \dot{q} = \dot{\lambda}\mathbf{h}\left(\dfrac{\partial \Psi^0(\varepsilon)}{\partial \varepsilon} - \overline{\sigma}^p - [u_a - \chi(S)(u_a - u_w)]\underset{\sim}{1}, \mathbf{q}\right) & \text{(plastic hardening law)} \\ f\left(\dfrac{\partial \Psi^0(\varepsilon)}{\partial \varepsilon} - \overline{\sigma}^p - [u_a - \chi(S)(u_a - u_w)]\underset{\sim}{1}, \mathbf{q}\right) \leq 0 & \text{(yield condition)} \end{cases}
\end{aligned} \tag{39}$$

From an algorithmic standpoint, the problem of integrating the evolution equations in Eq. 39 reduces to updating the basic variables $\{\boldsymbol{\sigma}, d^{\text{net}}, \overline{\boldsymbol{\sigma}}^{\mathbf{p}}, \mathbf{q}, \mathbf{u_a}, \mathbf{u_w}, \chi, S\}$ in a manner consistent with the constitutive model. It is essential to realize that in this computational process, the history of strains $t \to \varepsilon \equiv \nabla^s \mathrm{u}(t)$ is assumed to be given.

Equations of evolution are to be solved incrementally over a sequence of given time steps $[t_n, t_{n+1}] \subset R_+$, $n = 0, 1, 2 \ldots$. Thus, the initial conditions for equations are

$$\{\boldsymbol{\sigma}, d^{\text{net}}, \overline{\boldsymbol{\sigma}}^{\mathbf{p}}, \mathbf{q}, \mathbf{u_a}, \mathbf{u_w}, \chi, S\}|_{t=t_n} = \left\{\boldsymbol{\sigma}_n, d_n^{\text{net}}, \overline{\boldsymbol{\sigma}}_n^{\mathbf{p}}, \mathbf{q}_n, \mathbf{u}_{\mathbf{a}(n)}, \mathbf{u}_{\mathbf{w}(n)}, \chi_n, S_n\right\} \tag{40}$$

In accordance with the notion of operator split, the following additive decomposition of problem of evolution into elastic-damage-healing and plastic parts can be considered:

$$\begin{aligned}
&\mathbf{1.Elastic-damage-healing\ part:}\\
&\dot{\boldsymbol{\varepsilon}} = \nabla^s \dot{\mathbf{u}}(t)\\
&\dot{d} = \begin{cases} H(\xi^+)\dot{\xi}^+ & \text{iff} \quad \phi_t^d = \dot{\phi}_t^d = 0 \\ 0 & \text{otherwise} \end{cases}\\
&\dot{g} = \begin{cases} \dot{\xi}^+ & \text{iff} \quad \phi_t^d = \dot{\phi}_t^d = 0 \\ 0 & \text{otherwise} \end{cases}\\
&\dot{R} = \begin{cases} Z(\xi^-)\dot{\xi}^- & \text{iff} \quad \phi_t^h = \dot{\phi}_t^h = 0 \\ 0 & \text{otherwise} \end{cases}\\
&\dot{r} = \begin{cases} \dot{\xi}^- & \text{iff} \quad \phi_t^h = \dot{\phi}_t^h = 0 \\ 0 & \text{otherwise} \end{cases}\\
&\dot{d}^{\text{net}} = \dot{d} - \dot{R}\hat{d}\\
&\dot{\overline{\boldsymbol{\sigma}}} = \frac{d}{dt}\left[\frac{\partial \Psi^0(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}} - [u_a - \chi(S)(u_a - u_w)]\underset{\sim}{\mathbf{1}}\right]\\
&\dot{\overline{\boldsymbol{\sigma}}}^{\mathbf{p}} = \mathbf{0}\\
&\dot{\mathbf{q}} = \mathbf{0}
\end{aligned} \tag{41}$$

$$\begin{aligned}
&\mathbf{2.Plastic\ part:}\\
&\dot{\boldsymbol{\varepsilon}} = \mathbf{0}\\
&\dot{d}^{net} = 0\\
&\dot{r} = 0\\
&\dot{\overline{\boldsymbol{\sigma}}} = -\dot{\overline{\boldsymbol{\sigma}}}^{\mathbf{p}}\\
&\dot{\overline{\boldsymbol{\sigma}}}^{\mathbf{p}} = \dot{\lambda}\frac{\partial f}{\partial \boldsymbol{\varepsilon}}\left(\frac{\partial \Psi^0(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}} - \overline{\boldsymbol{\sigma}}^{\mathbf{p}} - \frac{d}{dt}[u_a - \chi(S)(u_a - u_w)]\underset{\sim}{\mathbf{1}},\ \mathbf{q}\right)\\
&\dot{\mathbf{q}} = \dot{\lambda}\mathbf{h}\left(\frac{\partial \Psi^0(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}} - \overline{\boldsymbol{\sigma}}^{\mathbf{p}} - \frac{d}{dt}[u_a - \chi(S)(u_a - u_w)]\underset{\sim}{\mathbf{1}},\ \mathbf{q}\right)
\end{aligned} \tag{42}$$

It is noted that Eqs. 41 and 42 do indeed add up to Eq. 39 in agreement with the notion of operator split. The formulation of an algorithm consistent with equations is based on the following fundamental result concerning operator or split method. Given the two algorithms, the first one is consistent with problem (41)

(elastic-damage-healing predictor) and the second one is consistent with problem (42) (return mapping corrector). In turn, the product algorithm obtained by successive application of these two algorithms is consistent with the original problem.

The Elastic-Damage-Healing Predictor

An algorithm consistent with problem (39), referred to as the elastic-damage-healing predictor in the sequel, is given by the following step-by-step procedure:

Step 1: Strain update: Given the incremental displacement field $\mathbf{u}_{n+1}$, the strain tensor is updated as

$$\boldsymbol{\varepsilon}_{n+1} = \boldsymbol{\varepsilon}_n + \nabla^s \mathbf{u}_{n+1} \tag{43}$$

Step 2: Compute the mode I positive 4th rank projection operator $\mathbf{P}^+_{n+1}$ based on the total strain

tensor $\boldsymbol{\varepsilon}_{n+1}$

$$\mathbf{P}^+_{ijkl}(\boldsymbol{\varepsilon}_{n+1}) = \frac{1}{2}\left(\mathbf{Q}^+_{ik}\ \mathbf{Q}^+_{jl} + \mathbf{Q}^+_{il}\ \mathbf{Q}^+_{jk}\right) \tag{44}$$

where $\mathbf{Q}^+_{n+1} = \sum_{i=1}^{2} \hat{H}(\boldsymbol{\varepsilon}_i)\mathbf{p}_i \otimes \mathbf{p}_i$; $\boldsymbol{\varepsilon}_{n+1} = \sum_{i=1}^{2} \boldsymbol{\varepsilon}_i \mathbf{p}_i \otimes \mathbf{p}_i$ (for 2D simulations in this chapter) and $\hat{H}(\boldsymbol{\varepsilon}_i)$ is the smoothed Heaviside function.

Step 3: Compute the negative 4th rank projection operator $\mathbf{P}^-_{n+1}$ based on the total strain tensor

$$\boldsymbol{\varepsilon}_{n+1}$$

$$\mathbf{P}^-_{ijkl}(\boldsymbol{\varepsilon}_{n+1}) = \frac{1}{2}\left(\mathbf{Q}^-_{ik}\ \mathbf{Q}^-_{jl} + \mathbf{Q}^-_{il}\ \mathbf{Q}^-_{jk}\right) \tag{45}$$

where $\mathbf{Q}^-_{n+1} = \underset{\sim}{\mathbf{1}} - \mathbf{Q}^+_{n+1}$

Step 4: Compute the initial (undamaged) elastic tensile and compressive strain energies, ξ^+ and ξ^-,

$$\xi^+_{n+1} \equiv \sqrt{\Psi^0\left(\boldsymbol{\varepsilon}^+_{n+1}\right)} = \sqrt{\frac{1}{2}\ \boldsymbol{\varepsilon}^+_{n+1} : \mathbf{C}^0 : \boldsymbol{\varepsilon}^+_{n+1}}\ , \quad \text{if linear elastcity} \tag{46}$$

$$\xi^-_{n+1} \equiv \sqrt{\Psi^0\left(\boldsymbol{\varepsilon}^-_{n+1}\right)}\ = \sqrt{\frac{1}{2}\ \boldsymbol{\varepsilon}^-_{n+1} : \mathbf{C}^0 : \boldsymbol{\varepsilon}^-_{n+1}}\ , \quad \text{if linear elastcity} \tag{47}$$

where $\boldsymbol{\varepsilon}^+ = \mathbf{P}^+ : \boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}^- = \mathbf{P}^- : \boldsymbol{\varepsilon}$

Step 5: Update the scalar damage parameter threshold $\tilde{d}_{n+1}$ (a history variable, initially 0).

$\tilde{d}_{n+1}$ denotes the net amount of scalar damage threshold after an incremental healing takes place from the previous time step (if any).

If $\Delta R_n > 0$ and $\tilde{d}_n > 0$, then update (lower) the scalar damage parameter threshold

$$\tilde{d}_{n+1} = \tilde{d}_n(1.0 - \Delta R_n). \tag{48}$$

Otherwise, set

$$\tilde{d}_{n+1} = \tilde{d}_n. \tag{49}$$

Step 6: Compute the scalar damage predictor d_{n+1} based on ξ^+_{n+1}.

If ξ^+_{n+1} at current time step is less than the initial damage threshold, no further damage is generated. Set $\Delta d_{n+1} = 0$. If ξ^+_{n+1} is larger than the given initial damage threshold, compute the scalar damage predictor d_{n+1} by using one of the following nonlinear damage functional evaluation:

$$d_{n+1}\left(\xi^+_{n+1}\right) = \frac{k_c\left[\xi^+_{n+1} - k_i\right]}{\xi^+_{n+1}\left[k_c - k_i\right]} \tag{50}$$

where k_c and k_i are material constants for damage evolution

or

$$d_{n+1}\left(\xi^+_{n+1}\right) = 1 - \frac{A(1-B)}{\xi^+_{n+1}} - B\ \exp\left[A - \xi^+_{n+1}\right] \tag{51}$$

where A and B are material constants for damage evolution.

Step 7: Check the incremental scalar damage criterion Δd_{n+1}.

Compute $\Delta d_{n+1} = d_{n+1} - \tilde{d}_{n+1}$.

$$\text{If } \Delta d_{n+1} \left\{ \begin{array}{l} \leq 0, \text{ no further damage, } \Delta d_{n+1} = 0 \Rightarrow \text{Go to } \textbf{Step 9} \\ > 0, \text{ further damage} \Rightarrow \text{Set } \tilde{d}_{n+1} = d_{n+1}; \text{ continue} \end{array} \right\}$$

Step 8: Compute the incremental hybrid isotropic damage predictor tensor $\Delta \mathrm{D}_{n+1}$ and the updated hybrid isotropic nonsymmetric *interim* damage tensor $\hat{\mathrm{D}}_{n+1}$:

$$\Delta \mathrm{D}_{n+1} = \Delta d_{n+1}\ \mathrm{I} \tag{52}$$

$$\hat{\mathrm{D}}_{n+1} = \mathrm{D}^{\text{net}}_n + \Delta \mathrm{D}_{n+1} \tag{53}$$

where **I** denotes the fourth-order identity tensor.

Step 9: Update the scalar healing parameter threshold $\tilde{R}_{n+1}$: (a history variable, initially 0)

$\tilde{R}_{n+1}$ denotes the **net** amount of healing sustained after damage takes place. If $\Delta d_{n+1} > 0$ and

$\tilde{R}_n > 0$, then update the (lower) healing parameter threshold for the next time step:

$$\tilde{R}_{n+1} = \tilde{R}_n(1.0 - \Delta d_{n+1}). \tag{54}$$

Otherwise, set

$$\tilde{R}_{n+1} = \tilde{R}_n. \tag{55}$$

Step 10: Compute the scalar healing predictor R_{n+1} based on ξ^-_{n+1}.

If ξ^-_{n+1} at current time step is less than the initial healing threshold, no further healing occurs. Set $\Delta R_{n+1} = 0$. If ξ^-_{n+1} is larger than the given initial healing threshold, compute the scalar healing predictor R_{n+1} by using one of the following nonlinear healing functional evaluation:

$$R_{n+1}\left(\xi^-_{n+1}\right) = \frac{h_c\left[\xi^-_{n+1} - h_i\right]}{\xi^-_{n+1}\left[h_c - h_i\right]} \tag{56}$$

where h_c and h_i are material constants for healing evolution

or

$$\frac{R_{n+1}\left(\xi^-_{n+1}\right) = 1 - \overline{A}\left(1 - \overline{B}\right)}{\xi^-_{n+1} - \overline{B}\ \exp\left[\overline{A} - \xi^-_{n+1}\right]} \tag{57}$$

where $\overline{A}$ and $\overline{B}$ are material constants for healing evolution.

Step 11: Check the incremental scalar healing criterion ΔR_{n+1}

Compute $\Delta R_{n+1} = R_{n+1} - \tilde{R}_{n+1}$.

$$\text{If } \Delta R_{n+1} \left\{ \begin{array}{l} \leq 0, \text{ no further healing}, \Delta R_{n+1} = 0 \Rightarrow \text{ Go to } \textbf{Step 14} \\ > 0, \text{ further healing} \Rightarrow \text{Set } \tilde{R}_{n+1} = \hat{R}_{n+1}; \text{ continue} \end{array} \right\}$$

Step 12: Compute the hybrid isotropic incremental healing tensor $\Delta\mathbf{R}_{n+1}$

$$\Delta\mathbf{R}_{n+1} = \Delta R_{n+1}\ \mathbf{I} \tag{58}$$

Step 13: Compute the incremental healing corrector, and update the hybrid isotropic net (combined) damage-healing tensor (true damage measure, a history variable):

$$\begin{array}{l} \Delta\mathbf{D}^H_{n+1} = -\hat{\mathbf{D}}_{n+1} \bullet \Delta\mathbf{R}_{n+1} \\ \mathbf{D}^{\text{net}}_{n+1} \ = \hat{\mathbf{D}}_{n+1} + \Delta\mathbf{D}^H_{n+1} \end{array} \tag{59}$$

Step 14: Calculate the mean net stress and compute the associated degree of saturation based on the previous discussions. Once the degree of saturation is

known, the corresponding Bishop's effective stress parameter χ and matric suction can be obtained.

Trial (predictor) stress: By mere substitution into the potential for the stress tensor, the following is obtained:

$$\boldsymbol{\sigma}_{n+1}^{0} = \frac{\partial \Psi^{0}(\boldsymbol{\varepsilon}_{n+1})}{\partial \boldsymbol{\varepsilon}} - [u_a - \chi(S)(u_a - u_w)]\,\underset{\sim}{\mathbf{1}} \quad \overline{\boldsymbol{\sigma}}_{n+1}^{trial} = \boldsymbol{\sigma}_{n+1}^{0} - \overline{\boldsymbol{\sigma}}_{n}^{\mathbf{p}} \quad \mathbf{q}_{n+1}^{trial} = \mathbf{q}_{n} \tag{60}$$

The Effective Plastic Return Mapping Corrector

To develop an algorithm consistent with the plastic part of the operator split, one first checks the loading/unloading conditions.

Step 15: Check for yielding: The algorithmic counterpart of the Kuhn–Tucker conditions is trivially implemented in terms of the elastic-damage trial stress. One simply checks

$$f\left(\overline{\boldsymbol{\sigma}}_{n+1}^{\text{trial}}, \mathbf{q}_{n+1}^{\text{trial}}\right) \quad \begin{cases} \leq 0 & \text{elastic} - \text{damage} - \text{healing} \Rightarrow \text{predictor} = \text{final state} \\ > 0 & \text{plastic} \Rightarrow \text{return mapping} \end{cases} \tag{61}$$

Multi-surface plasticity: In the case of plastic loading, it is necessary to determine the active plastic surface for Drucker–Prager criterion.

Step 16: Plastic return mapping corrector: In the case of plastic loading, the predictor stresses and internal variables are "returned back" to the yield surface along the algorithmic counterpart of the flow generated by Eq. 42. The algorithmic construction of this flow follows a proposed procedure which was inspired by a form of Kelley's convex cutting plane method for nonlinear optimization, with its basic structure inherited from Newton's method. Two fundamental advantages of this procedure are (a) the quadratic rate of convergence toward the yield surface and (b) the need for computing the gradient of the flow rule and hardening law that are entirely bypassed.

Step 17: Update the homogenized (nominal) stress $\boldsymbol{\sigma}_{n+1}$:

$$\boldsymbol{\sigma}_{n+1} = \left[\mathbf{I} - \mathbf{D}_{n+1}^{\text{net}}\right] : \left[\overline{\boldsymbol{\sigma}}_{n+1} + \{u_a - \chi(S)(u_a - u_w)\}\ \underset{\sim}{\mathbf{1}}\right] \tag{62}$$

Numerical Simulations

To demonstrate the effect of matric suction, a soil pushing process is performed. In the numerical simulation, the modified Drucker–Prager associative multi-surface plasticity formulation as expressed in Eq. 38 is employed to model the soil behavior for the sake of simplicity. The associated soil properties and assumed numerical

Table 1 The associated soil properties

Soil properties	Young's modulus	Poisson's ratio	Density	Cohesion	Lame constants
Values	24.7 (MPa)	0.35	$1.88 \times 10^3 (\text{kg/m}^3)$	0.19 (MPa)	$\lambda = 2.13457 \times 10^7 (\text{Pa})$ $\mu = 9.14815 \times 10^7 (\text{Pa})$

Table 2 The assumed numerical material constants

Numerical material constants	$\chi(S)$ in Eq. 1	(m, n, a) in Eq. 2	$\boldsymbol{\sigma}_c$, $\boldsymbol{\sigma}_t$, S_{ini}, $S_{\max}$, $S_{\min}$	λ in Eq. 38
Values	$k = 1$ in Eq. 3	$m = 1, n = 2, a$ varies so that the matric suction ranges from 0 to 440 KPa, 0 to 890 KPa, and 0 to 1,130 KPa	$\boldsymbol{\sigma}_c = 2,000 (\text{KPa})$ $\boldsymbol{\sigma}_t = 500 (\text{Pa})$ $S_{\text{ini}} = 0.2$ $S_{\max} = 1.0$ $S_{\min} = 0.1$	0.5

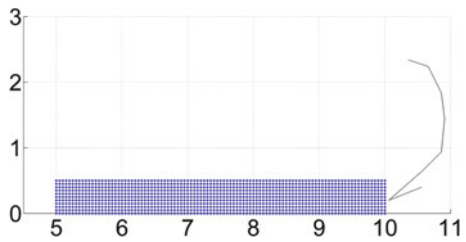


Fig. 6 The initial configuration and discretization of the earth-pushing processes

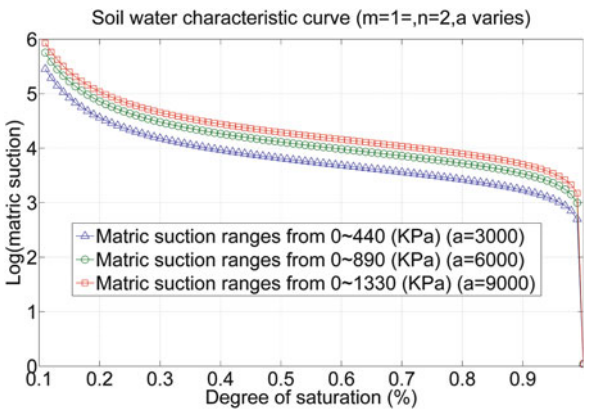


Fig. 7 Three soil–water characteristic curves for the earth-pushing simulations

material constants are listed in Tables 1 and 2. The NMAP (Nonlinear Meshfree Analysis Program) meshfree codes were provided by Prof. J.S. Chen's group at UCLA; cf. Chen and Wang (2000), Chen (2001), Chen et al. (2001), Wu et al. (2001), Chen and Wu (2007), and Chen et al. (2009).

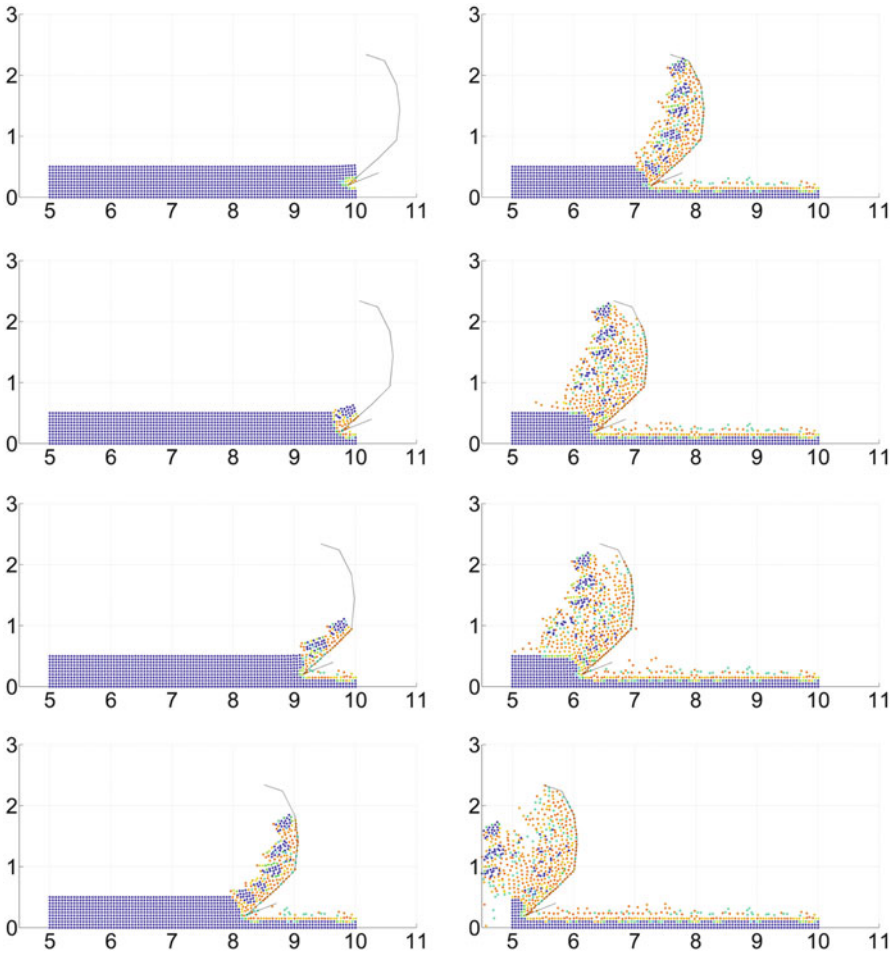


Fig. 8 The damage contours and soil deformations for earth-pushing simulations

The initial configuration and discretization of the earth-pushing process is rendered in Fig. 6. The blade of the bulldozer is treated as a rigid body and thus represented by two contact surfaces in black color. To model the contact between soil particles and blade surfaces, the traditional penalty formulation is employed. A layer of soil with dimension 5×0.5 (meter) is discretized into $101 \times 11 = 1111$ uniformly distributed particles. The blade as exhibited in Fig. 6 is controlled to move horizontally from right to left for 5 m.

The initial degree of saturation is assumed 20 % and varies linearly when the mean net stress of soil particles is changed as discussed. The corresponding matric suction then follows the soil–water characteristic curves as presented in Fig. 7 which are obtained from Eq. 2. By setting $m = 1$, $n = 2$ and allowing varying a,

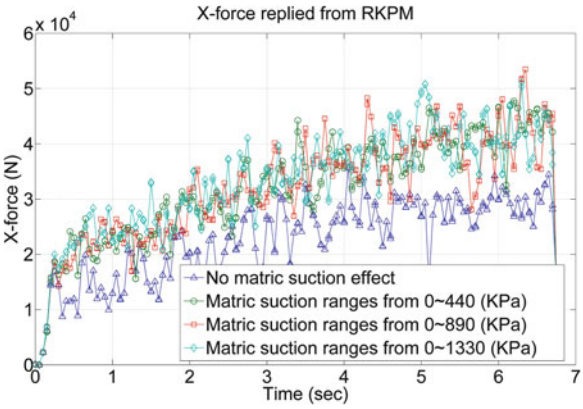


Fig. 9 The horizontal resulting forces on the bulldozer blade for different ranges of the matric suction

Table 3 The maximum horizontal resulting force on bulldozer blade

Initial elastic energy-based hybrid isotropic damage-healing models	No matric suction effect	Matric suction ranges from 0 to 440 (KPa)	Matric suction ranges from 0 to 890 (KPa)	Matric suction ranges from 0 to 1,130 (KPa)
Max. horizontal resulting force on bulldozer blade (N)	3.5681e +004	4.7764e+004	5.3514e+004	5.1145e+004
Percentage	100 %	133.86 %	149.98 %	143.34 %

three curves are generated with the matric suction ranging from 0 $\sim$ 440 (KPa), 0 $\sim$ 890 (KPa), to 0 $\sim$ 1,330 (KPa).

The initial elastic strain energy-based couple elastoplastic hybrid isotropic damage-healing model with three different ranges of matric suction is then applied to generate the earth-pushing simulations as demonstrated in Fig. 8. In these figures, the blue solid circles represent the soil particles with no damage ($d^{\text{net}} = 0$), the red solid circles symbolize the soil particles with full damage ($d^{\text{net}} = 0.95$, the preset upper bound damage in these simulations), and the solid circles in other colors (such as green, yellow, and orange) signify partial damage between 0 and 0.95.

The histories of horizontal resultant force on bulldozer blade by considering three different ranges of matric suction are plotted in Fig. 9, and their corresponding maximum values are listed in Table **3**.

From Fig. 9 and Table 3, the numerical results exhibit the trend of the effect of matric suction. Namely, for partially saturated soils, matric suction is an important factor to influence its mechanical behavior. More specifically, the effective stress is increased and soils become more stiffened when the matric suction increases.

In the absence of associated experimental data, the proposed initial elastic strain energy-based coupled elastoplastic hybrid isotropic damage-healing formulations with the consideration of matric suction for partially saturated soil models only demonstrate the salient feature of the effect of matric suction. Additional future

efforts are required to obtain the validations to the numerical material constants and parameters in the proposed formulations.

Conclusions

In this study, new ***initial elastic strain energy-***based coupled elastoplastic hybrid isotropic damage-healing formulations have been presented for partially saturated soils in the specific simulation of earth-pushing processes. The equivalent strain based on $\mathbf{P}^{+}$ and $\mathbf{P}^{-}$ for the energy criteria of damage and healing is used for the proposed models. New two-step operator split algorithms (featuring the elastic-damage-healing predictors and the effective plastic return mapping corrector) are presented. In addition, the effect of matric suction and the coupling of damage and healing are proposed in an incremental form by using the ***predictor formula*** and ***corrector formula*** as introduced. In the absence of associate experimental data, the proposed innovative formulations and algorithms for partially saturated soils only demonstrate the effect of matric suction. Further validations to the material parameters in each damage and healing evolutions will be performed once the associated experiment data become available in the near future. In addition, much gratitude is expressed to Prof. J.S. Chen and Dr. Pai-Chen Guan for their assistance in implementing the proposed models into the NMAP (Nonlinear Meshfree Analysis Program) meshfree code by Chen (2001).

References

E.E. Alonso, A. Gens, A. Josa, A constitutive model for partially saturated soils. Geotechnique **40**(3), 405–430 (1990)

A.W. Bishop, The principle of effective stress. Technisk Ukeflad. **106**(39), 859–863 (1959)

A.W. Bishop, I. Alpan, G.E. Blight, I.B. Donald, Factors controlling the shear strength of partly saturated cohesive soils, in *ASCE Research Conference on Shear Strength of Cohesive Soils*, University of Colorado, Boulder, 1960

G.E. Blight, Effective stress evaluation for unsaturated soils. ASCE J. Soil Mech. Found. Div. **93**(SM2), 125–148 (1967)

G. Bolzon, B.A. Schrefler, O.C. Zienkiewicz, Elastoplastic soil constitutive laws generalized to partially saturated states. Geotechnique **46**(2), 279–289 (1996)

R.I. Borja, Cam-clay plasticity. Part V: a mathematical framework for three-phase deformation and strain localization analyses of partially saturated porous media. Comput. Methods Appl. Mech. Eng. **193**(48–51), 5301–5338 (2004)

J.S. Chen, *Nonlinear Meshfree Analysis Program (NMAP)* (University of California, Los Angeles, 2001)

J.S. Chen, H.P. Wang, New boundary condition treatments for meshless computation of contact problems. Comput. Methods Appl. Mech. Eng. **187**, 441–468 (2000)

J.S. Chen, Y. Wu, Stability in lagrangian and semi-Lagrangian reproducing Kernel discretizations using nodal integration in nonlinear solid mechanics, in *Computational Methods in Applied Sciences*, ed. by V.M.A. Leitao, C.J.S. Alves, C.A. Duarte (Springer, Dordrecht, 2007), pp. 55–77

J.S. Chen, C.T. Wu, S. Yoon, Y. You, A stabilized conforming nodal integration for Galerkin meshfree methods. Int. J. Numer. Methods Eng. **50**, 435–466 (2001)

J.S. Chen, Y. Wu, P. Guan, H. Teng, J. Gaidos, K. Hofstetter, M. Alsaleh, A semi-Lagrangian reproducing Kernel formulation for modeling earth moving operations. Mech. Mater. **41**, 670–683 (2009)

B.D. Coleman, M.E. Gurtin, Thermodynamics with internal state variables. J. Chem. Phys. **47**(2), 597–613 (1967)

V. Escario, Suction-controlled penetration and shear tests, in *Proceedings of the 4th International Conference on Expansive Soils*, Denver, 1980, pp. 781–787

V. Escario, J. Saez, The shear-strength of partly saturated soils. Geotechnique **36**(3), 453–456 (1986)

V. Escario, J. Juca, M.S. Coppe, Strength and deformation of partly saturated soils, in *Proceedings of the 12th International Conference on Soil Mechanics and Foundation Engineering*, vol. 3, Rio de Janeiro, 1989, pp. 43–46

D.G. Fredlund, 2nd Canadian geotechnical colloquium – appropriate concepts and technology for unsaturated soils. Can. Geotech. J. **16**(1), 121–139 (1979)

D.G. Fredlund, N.R. Morgenstern, Stress state variables for unsaturated soils. J. Geotech. Eng. Div. ASCE **104**(11), 1415–1416 (1978)

D.G. Fredlund, H. Rahardio, *Soil Mechanics for Unsaturated Soils* (Wiley, New York, 1993)

D.G. Fredlund, A.Q. Xing, Equations for the soil-water characteristic curve. Can. Geotech. J. **31**(4), 521–532 (1994)

D.G. Fredlund, N.R. Morgenstern, R.A. Widger et al., Shear-strength of unsaturated soils. Can. Geotech. J. **15**(3), 313–321 (1978)

D. Gallipoli, A. Gens, R. Sharma, J. Vaunat, An elasto-plastic model for unsaturated soil incorporating the effects of suction and degree of saturation on mechanical behaviour. Geotechnique **53**(1), 123–135 (2003a)

D. Gallipoli, S.J. Wheeler, M. Karstunen, Modelling the variation of degree of saturation in a deformable unsaturated soil. Geotechnique **53**(1), 105–112 (2003b)

A. Gens, E.E. Alonso, A framework for the behavior of unsaturated expansive clays. Can. Geotech. J. **29**(6), 1013–1032 (1992)

K. Georgiadis, D.M. Potts, L. Zdravkovic, Three-dimensional constitutive model for partially and fully saturated soils. Int. J. Geomech. **5**(3), 244–255 (2005)

J.W. Ju, K.Y. Yuan, New strain energy based coupled elastoplastic two-parameter damage and healing models for earth moving processes. Int. J. Damage Mech. **21**(7), 989–1019 (2012)

J.W. Ju, K.Y. Yuan, A.W. Kuo, Novel strain energy based coupled elastoplastic damage and healing models for geomaterials – part I: formulations. Int. J. Damage Mech. **21**(4), 525–549 (2012a)

J.W. Ju, K.Y. Yuan, A.W. Kuo, J.S. Chen, Novel strain energy based coupled elastoplastic damage and healing models for geomaterials – part II: computational aspects. Int. J. Damage Mech. **21**(4), 551–576 (2012b)

R. Kohler, G. Hofstetter, A cap model for partially saturated soils. Int. J. Numer. Anal. Methods Geomech. **32**(8), 981–1004 (2008)

J. Krahn, D.G. Fredlund, Total, matric and osmotic suction. Soil Sci. **114**(5), 339–348 (1972)

B. Loret, N. Khalili, A three-phase model for unsaturated soils. Int. J. Numer. Anal. Methods Geomech. **24**(11), 893–927 (2000)

N. Lu, W. Likos, *Unsaturated Soil Mechanics* (Wiley, Hoboken, 2004)

N. Mitarai, F. Nori, Wet granular materials. Adv. Phys. **55**(1–2), 1–45 (2006)

J.K. Mitchell, K. Soga, *Fundamentals of Soil Behavior*, 3rd edn. (Wiley, New York, 2005)

A. Murakami, T. Setsuyasu, S. Arimoto, Mesh-free method for soil-water coupled problem within finite strain and its numerical validity. Soils Found. **45**(2), 145–154 (2005)

T. Nishimura, Y. Murasawa, T. Okami, Estimating air-water hydraulic conductivity using soil-water characteristic curve. Proc. Fourth Int. Conf. Unsatur. Soils Geotech. Spec. Publ. **147**(2), 1595–1606 (2006)

L. Sanavia, F. Pesavento, B.A. Schrefler, Finite element analysis of non-isothermal multiphase geomaterials with application to strain localization simulation. Comput. Mech. **37**(4), 331–348 (2006)

J.C. Simo, J.W. Ju, K.S. Pister, R.L. Taylor, An assessment of the cap model: consistent return algorithms and rate-dependent extension. J. Eng. Mech. ASCE **114**(2), 191–218 (1988)

S.K. Vanapalli, D.G. Fredlund, Comparison of empirical procedures to predict the shear strength of unsaturated soils using the soil-water characteristic curve, in *Advances in Unsaturated Geotechnics*, ed. by C.D. Shackelford, N.Y. Chang. GSP No. 99 (ASCE, Reston, 2000), pp. 195–209

S.K. Vanapalli, D.G. Fredlund, D.E. Pufahl, A.W. Clifton, Model for the prediction of shear strength with respect to soil suction. Can. Geotech. J. **33**(3), 379–392 (1996)

S.J. Wheeler, V. Sivakumar, An elasto-plastic critical state framework for unsaturated soil. Geotechnique **45**(1), 35–53 (1995)

S.J. Wheeler, D. Gallipoli, M. Karstunen, Comments on use of the Barcelona basic model for unsaturated soils. Int. J. Numer. Anal. Methods Geomech. **26**(15), 1561–1571 (2002)

S.J. Wheeler, R.S. Sharma, M.S.R. Buisson, Coupling of hydraulic hysteresis and stress-strain behaviour in unsaturated soils. Geotechnique **53**(1), 41–54 (2003)

C.T. Wu, J.S. Chen, L. Chi, F. Huck, Lagrangian meshfree formulation for analysis of geotechnical materials. J. Eng. Mech. **127**, 440–449 (2001)

C.E. Zapata, W.N. Houston, S.L. Houston, K.D. Walsh, Soil-water characteristic curve variability, in C. D. Shackelford, S. L. Houston, and N. Y. Chang, (eds) *Advanced in Unsaturated Geotechnics*. ASCE, Reston, VA, 84–124

New Strain-Energy Based Coupled Elastoplastic Two-Parameter Damage and Healing Models for Geomaterials

32

Jiann-Wen Woody Ju and K. Y. Yuan

Contents

Abstract

Novel strain-energy-based coupled elastoplastic ***two-parameter*** damage and healing formulations for geomaterials are developed and implemented for a numerical simulation of 2D earth-moving processes. A new class of elastoplastic damage-healing models is proposed within an ***initial-elastic strain-energy***-based framework. The governing incremental damage and healing evolutions are *coupled* in ***volumetric*** and ***deviatoric*** parts and characterized through the effective stress concept. The plastic flow is established via an additive split of the

J.-W.W. Ju (✉) • K.Y. Yuan
Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, USA
e-mail: juj@ucla.edu; yuanky@ucla.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_13

stress tensor. Specifically, four characteristic energy norms of the tensile volumetric, tensile deviatoric, compressive volumetric, and compressive deviatoric strain tensors are introduced, respectively, for the corresponding volumetric and deviatoric damage and healing mechanisms.

By adopting micromechanics-motivated damage and healing characterizations in the volumetric and deviatoric parts, the proposed two-parameter damage-healing models are implemented to demonstrate considerable versatility on numerical simulations of earth-moving processes. New computational algorithms are systematically developed based on the two-step operator splitting methodology. The volumetric and deviatoric elastic-damage-healing predictor and the effective plastic corrector are implemented within the RKPM meshfree codes. Numerical examples under earth excavation, transport, and compaction are presented to illustrate salient features of soils such as shear band and partial recovery of soil stiffness due to compaction by the new two-parameter damage-healing models.

Introduction

Certain soil motions (such as rolling and gliding) can cause considerable amount of soil spillage during earth-moving processes, which decrease the soil-carrying capacity of earth-moving equipment and thus increase the costs of earthwork operations. These rolling and gliding motions also generate higher frictional contacts between the surfaces of soil particles and the equipment blades, leading to faster blade wear and higher energy consumption by the tractor (Xia 2008). To enhance the operational efficiency of equipment, extensive laboratory experiments and field testing of competing designs for earth-moving equipment must be performed. To enable more economic and powerful design, a versatile computational framework that can be used to simulate soil motions during earth-moving activities is highly desired. Further, comprehensive physics-based material models of soils are needed in order to establish a rigorous computational framework to accommodate complex and instantaneous motions of cohesive soils.

Since the introduction of the scalar damage concept by Kachanov (1958) and Rabotnov et al. (1963) for creep of metals, the continuum damage mechanics has become an emerging field of active research. Extensive phenomenological damage models were proposed in the literature. In particular, scalar damage variables were widely employed for isotropic or one-dimensional phenomenological damage models (Lemaitre 1985), while vectorial, second-order, and fourth-order tensorial damage variables were often used for anisotropic phenomenological damage models (Kachanov 1980; Simo and Ju 1987a, b; Simo et al. 1988; Ju 1989a, b, 1990; Simo and Ju 1989; Ju et al. 2011a, b). However, there are few publications in coupled plasticity and sound damage models for cohesive soils. In particular, damage models are rarely built on two parts, i.e., the volumetric and deviatoric parts. As pointed out by Ju (1990), the *scalar* damage model implies that the Poisson's ratio always remains a *constant* and is therefore only a special case of

the *isotropic* damage. Moreover, the general form of the fourth-order isotropic damage tensor consists of two parts: the volumetric and deviatoric components.

From different fields, some materials experimentally show that they can be repaired or healed in various manners such as chemical, physical, or biological phenomena, leading to progressive recovery of internal material defects. Barbero et al. (2005) are referred for damage and irreversible deformation processes for a self-healing fiber-reinforced lamina. Further literature review of self-healing materials can be found in Herbst and Luding (2008). However, there has been no published literature on rigorous formulations of damage and healing (recovery of elastic stiffness) models for granular cohesive soils, except the very recent work by Ju et al. (2011a, b). Nevertheless, the latter did not invoke the decomposition of damage tensors into the volumetric and deviatoric parts. Therefore, there is a fundamental need for physical and reliable modeling of the progressive coupled elastoplastic *two-parameter* damage and healing in cohesive granular soils under complex and cyclic loading during earth-moving processes.

The remainder of this chapter is organized as follows. New initial-elastic strain-energy-based coupled elastoplastic two-parameter damage-healing formulations are presented in the next section, following by brand new computational algorithms which are systematically developed based on the two-step operator splitting methodology. Numerical simulations of earth-moving processes are subsequently presented to illustrate the versatility and salient features of the proposed new two-parameter soil damage-healing models. The governing volumetric and deviatoric damage and healing evolutions are coupled in an incremental form and characterized through the effective stress concept in conjunction with the hypothesis of strain equivalence.

New Initial-Elastic-Strain-Energy-Based Coupled Elastoplastic Two-Parameter Damage and Healing Models

A Coupled Two-Parameter Formulation

A crucial idea underlining the *two-parameter* coupled elastoplastic damage-healing models presented here concerns the criteria to determine when and how the individual volumetric or deviatoric, damage or healing mechanisms take place. In order to introduce the ***net effect*** of the volumetric and deviatoric damage-healing and plastic flow processes, a free energy potential of the following form is proposed:

$$\Psi\left(\boldsymbol{\varepsilon}_v, \mathbf{e}, \boldsymbol{\sigma}^{\mathrm{p}}, \mathbf{q}, d_v^{\mathrm{net}}, d_d^{\mathrm{net}}\right) \equiv \left(1 - d_v^{\mathrm{net}}\right)\Psi_v^0(\boldsymbol{\varepsilon}_v) + \left(1 - d_d^{\mathrm{net}}\right)\Psi_d^0(\mathbf{e}) - \boldsymbol{\varepsilon} : \boldsymbol{\sigma}^{\mathrm{p}} + \Xi(\mathbf{q}, \boldsymbol{\sigma}^{\mathrm{p}}) \tag{1}$$

where $\boldsymbol{\varepsilon}_v \equiv \frac{1}{3}\boldsymbol{\varepsilon}_{kk}\ \underset{\sim}{\mathbf{1}}$ denotes the volumetric strain tensor, $\mathbf{e}$ is the deviatoric strain tensor; $\boldsymbol{\sigma}^{\mathbf{p}}$ is the plastic relaxation stress tensor, and $\mathbf{q}$ denotes a suitable set of

internal (plastic) hardening variables. Here, d_v^{net} signifies the scalar variable of *net* (combined) effect of volumetric damage and healing, which depends on the volumetric damage variable d_v, the incremental volumetric damage variable Δd_v, and the incremental volumetric healing variable ΔR_v, all between 0 and 1 in numerical values. Similarly, d_d^{net} is the scalar variable of *net* (combined) effect of deviatoric damage and healing, which depends on the deviatoric damage variable d_d, the incremental deviatoric damage variable Δd_d, and the incremental deviatoric healing variable ΔR_d, all between 0 and 1 in numerical values. Moreover, $\Psi_v^0(\boldsymbol{\varepsilon}_v)$ defines the initial-elastic volumetric stored energy function of the undamaged material, $\Psi_d^0(\mathbf{e})$ is the initial-elastic deviatoric stored energy function of the undamaged material, and $\Xi(\mathbf{q}, \boldsymbol{\sigma}^{\mathbf{p}})$ is the plastic potential function. For the linear elasticity case, one has $\Psi_v^0(\boldsymbol{\varepsilon}_v) = \frac{1}{2}\boldsymbol{\varepsilon}_v : \left(K \underset{\sim}{\mathbf{1}} \otimes \underset{\sim}{\mathbf{1}}\right) : \boldsymbol{\varepsilon}_v = \frac{1}{2}K(\boldsymbol{\varepsilon}_{kk})^2$, where K denotes the bulk modulus; $\Psi_d^0(\mathbf{e}) = G\mathbf{e} : \mathbf{e}$, where G represents the shear modulus.

Confining our focus to the purely mechanical theory, the Clausius-Duhem (reduced dissipation) inequality (for purely mechanical isothermal theory) takes the form for any admissible process:

$$-\dot{\Psi} + \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} \geq 0 \tag{2}$$

By taking the time derivative of Eq. 1, substituting it into Eq. 2, and making use of standard arguments (Coleman and Gurtin 1967) along with the additional assumption that the net effect of damage-healing and plastic unloading are elastic processes, the stress-strain constitutive law (Ju 1989a; Ju et al. 2011a) can be obtained:

$$\boldsymbol{\sigma} = \left(1 - d_v^{\text{net}}\right) \frac{\partial \Psi_v^0(\boldsymbol{\varepsilon}_v)}{\partial \boldsymbol{\varepsilon}_v} + \left(1 - d_d^{\text{net}}\right) \frac{\partial \Psi_d^0(\mathbf{e})}{\partial \mathbf{e}} - \boldsymbol{\sigma}^{\text{p}} \tag{3}$$

and the dissipative inequalities

$$-\frac{\partial \Xi(\mathbf{q}, \boldsymbol{\sigma}^{\mathbf{p}})}{\partial \mathbf{q}} \cdot \dot{\mathbf{q}} - \left(\frac{\partial \Xi(\mathbf{q}, \boldsymbol{\sigma}^{\text{p}})}{\partial \boldsymbol{\sigma}^{p}} - \boldsymbol{\varepsilon}\right) : \dot{\boldsymbol{\sigma}}^{\mathbf{p}} \geq 0 \tag{4}$$

$$\dot{d}_v^{\text{net}} \ \Psi_v^0(\boldsymbol{\varepsilon}_v) + \dot{d}_d^{\text{net}} \ \Psi_d^0(\mathrm{e}) - \frac{\partial \Xi(\mathrm{q}, \boldsymbol{\sigma}^{\text{p}})}{\partial \mathrm{q}} \cdot \dot{\mathrm{q}} - \left(\frac{\partial \Xi(\mathrm{q}, \boldsymbol{\sigma}^{\text{p}})}{\partial \boldsymbol{\sigma}^{\text{p}}} - \varepsilon\right) : \dot{\boldsymbol{\sigma}}^{\text{p}} \geq 0 \tag{5}$$

It follows from Eq. 3 that within the present strain-space formulation, the stress tensor is split into the volumetric and deviatoric elastic-damage-healing and the plastic relaxation parts. From Eqs. 4 and 5, it is observed that the dissipative energy by plasticity itself is positive; if the damage-healing effect is involved, the sum of dissipative energy by the damage-healing and plasticity effect is also positive. It is clear from Eqs. 3, 4, and 5 that the present framework is capable of accommodating general (nonlinear) elastic response and general plastic response.

The potential $\Xi(\mathbf{q}, \boldsymbol{\sigma}^{\mathbf{p}})$ is linked to plastic dissipation. Its role is such that inequality Eq. 5 is satisfied for arbitrary processes. Note that assumption is made that $\Xi(\mathbf{q}, \boldsymbol{\sigma}^{\mathbf{p}})$ independent of d_v^{net} and d_d^{net}. From Eq. 1, it then follows that

$$-Y_v \equiv -\frac{\partial \Psi\left(\boldsymbol{\varepsilon}_v, \mathbf{e}, \boldsymbol{\sigma}^{\mathbf{p}}, \mathbf{q}, d_v^{\text{net}}, d_d^{\text{net}}\right)}{\partial d_v^{\text{net}}} = \Psi_v^0(\boldsymbol{\varepsilon}_v) \tag{6}$$

$$-Y_d \equiv -\frac{\partial \Psi\left(\boldsymbol{\varepsilon}_v, \mathbf{e}, \boldsymbol{\sigma}^{\mathbf{p}}, \mathbf{q}, d_v^{\text{net}}, d_d^{\text{net}}\right)}{\partial d_d^{\text{net}}} = \Psi_d^0(\mathbf{e}) \tag{7}$$

Hence, the initial (undamaged) *volumetric* elastic strain energy $\Psi_v^0(\boldsymbol{\varepsilon}_v)$ is the thermodynamic force Y_v conjugate to the net volumetric damage-healing variable d_v^{net}; the initial (undamaged) deviatoric elastic strain energy $\Psi_d^0(\mathbf{e})$ is the thermodynamic force Y_d conjugate to the net deviatoric damage-healing variable d_d^{net}. Attention is then focused on the volumetric and deviatoric damage, the volumetric and deviatoric healing, and their interactions.

Characterization of Initial-Elastic-Strain-Energy-Based Two-Parameter Damage Model

The proposed damage mechanism is called ***two parameter*** since the damage effects or evolutions are divided into the volumetric and deviatoric parts. Motivated by Eq. 6, the notion of equivalent tensile volumetric strain ξ_v^+ is first ascribed as the (undamaged) energy norm of the tensile volumetric strain tensor:

$$\xi_v^+ \equiv \sqrt{\Psi_v^0\left(\boldsymbol{\varepsilon}_v^+\right)} = \sqrt{\frac{1}{2}K\left(\boldsymbol{\varepsilon}_{kk}^+\right)^2} \tag{8}$$

where $\boldsymbol{\varepsilon}_{kk}^+ = \boldsymbol{\varepsilon}_{11}^+ + \boldsymbol{\varepsilon}_{22}^+ + \boldsymbol{\varepsilon}_{33}^+$ and $\boldsymbol{\varepsilon}^+ \equiv \mathbf{P}^+ : \boldsymbol{\varepsilon}$. The $\boldsymbol{\varepsilon}$ tensor denotes the total strain, and the fourth-order tensor $\mathbf{P}^+$ denotes the "mode I" positive (tensile) projection tensor with components

$$\mathbf{P}_{ijkl}^+(\boldsymbol{\varepsilon}) = \frac{1}{2}\left(\mathbf{Q}_{ik}^+ \ \mathbf{Q}_{jl}^+ + \mathbf{Q}_{il}^+ \ \mathbf{Q}_{jk}^+\right) \tag{9}$$

where $\mathbf{Q}^+ \equiv \sum_{i=1}^{2} \hat{H}(\boldsymbol{\varepsilon}_i)\mathbf{p}_i \otimes \mathbf{p}_i$; $\boldsymbol{\varepsilon} = \sum_{i=1}^{2} \boldsymbol{\varepsilon}_i \ \mathbf{p}_i \otimes \mathbf{p}_i$, $\|\mathbf{p}_i\| = 1$ for the 2D simulations in this chapter. Here, $\boldsymbol{\varepsilon}_i$ is the ith principal strain, $\mathbf{p}_i$ is the ith corresponding unit vector in the principal direction, and $\hat{H}(\boldsymbol{\varepsilon}_i)$ is the smoothed Heaviside function.

Then, the state of volumetric damage in soils is characterized by means of a volumetric damage criterion $\phi_v^d(\xi_{v,t}^+, g_{v,t}) \leq 0$, formulated in the strain space, with the following form:

$$\phi_v^d\left(\xi_{v,t}^+, g_{v,t}\right) \equiv \xi_{v,t}^+ - g_{v,t} \leq 0, \qquad t \in \mathbf{R}_+ \tag{10}$$

Here, the subscript t refers to a value at current time $t \in \mathbf{R}_+$, and $g_{v,t}$ is the volumetric damage threshold at the current time t. The numerical value of $g_{v,t}$ would be lowered due to the incremental volumetric healing (if any) from the previous time step. Let $g_{v,0}$ denote the initial volumetric damage threshold before any loading is applied; it must have $g_{v,t} \geq g_{v,0}$. Here, $g_{v,0}$ is considered to be a characteristic material property. Condition Eq. 10 then states that volumetric damage in soils is initiated when the energy norm of the tensile volumetric strain $\xi_{v,t}^+$ exceeds the initial volumetric damage threshold $g_{v,0}$. The evolutions of the volumetric damage variable $d_{v,t}$ and the volumetric damage threshold $g_{v,t}$ are defined, respectively, by the rate equations

$$\dot{d}_{v,t} = \dot{\mu}_v H_v\left(\xi_{v,t}^+, d_{v,t}\right) \tag{11}$$

$$\dot{g}_v = \dot{\mu}_v \tag{12}$$

where $\dot{\mu}_v \geq 0$ is the volumetric damage consistency parameter that defines the volumetric damage loading/unloading conditions according to the Kuhn-Tucker relations:

$$\dot{\mu}_v \geq 0, \;\; \phi_v^d\left(\xi_{v,t}^+, g_{v,t}\right) \leq 0, \;\; \dot{\mu}_v \phi_v^d\left(\xi_{v,t}^+, g_{v,t}\right) = 0 \tag{13}$$

Moreover, H_v in Eq. 11 denotes the volumetric damage hardening function. Conditions Eq. 13 are standard for problems involving unilateral constraint. If $\phi_v^d(\xi_{v,t}^+, g_{v,t}) < 0$, the volumetric damage criterion is not satisfied and by conditions Eq. 13, $\dot{\mu}_v = 0$; hence, the volumetric damage rule Eq. 11 implies that $\dot{d}_v = 0$ and no further volumetric damage takes place. If, on the other hand, $\dot{\mu}_v > 0$, that is, further volumetric damage ("tensile loading") is taking place, conditions Eq. 13 imply that $\phi_v^d(\xi_{v,t}^+, g_{v,t}) = 0$. In this event, the value of $\dot{\mu}_v$ is determined by the volumetric damage consistency condition; i.e.,

$$\phi_v^d\left(\xi_{v,t}^+, g_{v,t}\right) = \dot{\phi}_v^d\left(\xi_{v,t}^+, g_{v,t}\right) = 0 \quad \Rightarrow \quad \dot{\mu}_v = \dot{\xi}_v^+ \tag{14}$$

and $g_{v,t}$ is defined by the expression

$$g_{v,t} = \max\left(g_{v,0}, \max_{s \in (-\infty,t)} \xi_{v,s}^+\right) - \quad g_{vh,(t-1)} \tag{15}$$

where $g_{vh,(t-1)}$ defines the reduced value for the current volumetric damage threshold due to incremental volumetric healing (if any) from the previous time step.

If $H_v(\xi^+_{v,t},\, d_{v,t})$ in Eq. 11 is independent of $\dot{d}_{v,t}$, the above formulation may be rephrased as follows. Let $G_v : \mathbf{R} \rightarrow \mathbf{R}_+$ be such that $H_v(\xi^+_{v,t}) \equiv \partial G_v(\xi^+_{v,t})/\partial \xi^+_{v,t}$. $G_v(\bullet)$ is assumed monotonic. A volumetric damage criterion entirely equivalent to condition Eq. 10 is given by $\overline{\phi}^d_v\left(\xi^+_{v,t}, g_{v,t}\right) \equiv G_v\left(\xi^+_{v,t}\right) - G_v\left(g_{v,t}\right) \leq 0$. The flow rule and loading/unloading conditions become

$$\dot{d}_{v,t} = \dot{\mu}_v \frac{\partial \overline{\phi}^d_v\left(\xi^+_{v,t}, g_{v,t}\right)}{\partial \xi^+_{v,t}}, \qquad g_{v,t} = \dot{\mu}_v \tag{16}$$

$$\dot{\mu}_v \geq 0, \quad \overline{\phi}^d_v\left(\xi^+_{v,t}, g_{v,t}\right) \leq 0, \quad \dot{\mu}_v \overline{\phi}^d_v\left(\xi^+_{v,t}, g_{v,t}\right) = 0 \tag{17}$$

Conditions Eqs. 16 and 17 are simply the Kuhn-Tucker optimality conditions of a principle of maximum damage dissipation.

Likewise, the notion of equivalent *tensile* deviatoric strain ξ^+_d is ascribed as the (undamaged) energy norm of the tensile deviatoric strain tensor. Motivated by Eq. 7, it can be defined that

$$\xi^+_d \equiv \sqrt{\Psi^0_d(\mathbf{e}^+)} = \sqrt{\frac{1}{2} G\, \mathbf{e}^+ : \mathbf{e}^+} \tag{18}$$

where $\mathbf{e}^+ = \boldsymbol{\varepsilon}^+ - \frac{1}{3}\boldsymbol{\varepsilon}^+_{kk}\ \underset{\sim}{\mathbf{1}}$ and $\boldsymbol{\varepsilon}^+ \equiv \mathbf{P}^+ : \boldsymbol{\varepsilon}$.

Similar to the characterization of volumetric damage, the state of deviatoric damage can be characterized by means of a deviatoric damage criterion $\phi^d_d\ (\xi^+_{d,t},\, g_{d,t}) \leq 0$, formulated in the strain space, with the following functional form:

$$\phi^d_d\left(\xi^+_{d,t}, g_{d,t}\right) \equiv \xi^+_{d,t} - g_{d,t} \leq 0, \qquad t \in \mathbf{R}_+ \tag{19}$$

Here, the subscript t refers to a value at the current time $t \in \mathbf{R}_+$, and $g_{d,t}$ is the deviatoric damage threshold at the current time t. It is noted that $g_{d,t}$ will be lowered in value due to the incremental deviatoric healing from the previous time step. If $g_{d,0}$ denotes the initial deviatoric damage threshold before any loading is applied, it must have $g_{d,t} \geq g_{d,0}$. Note that $g_{d,0}$ is considered as a characteristic material property. Condition Eq. 19 states that deviatoric damage in soils is initiated when the energy norm of the tensile deviatoric strain tensor $\xi^+_{d,t}$ exceeds the initial deviatoric damage threshold $g_{d,0}$. The evolutions of the deviatoric damage variable $d_{d,t}$ and the deviatoric damage threshold $g_{d,t}$ are defined by the rate equations:

$$\dot{d}_{d,t} = \dot{\mu}_d\ H_d\left(\xi^+_{d,t}, d_{d,t}\right) \tag{20}$$

$$\dot{g}_d = \dot{\mu}_d \tag{21}$$

where $\dot{\mu}_d \geq 0$ is the deviatoric damage consistency parameter that defines deviatoric damage loading/unloading conditions according to the Kuhn-Tucker relations:

$$\dot{\mu}_d \geq 0, \quad \phi_d^d\left(\xi_{d,t}^+, g_{d,t}\right) \leq 0, \quad \dot{\mu}_d \ \phi_d^d\left(\xi_{d,t}^+, g_{d,t}\right) = 0 \tag{22}$$

Moreover, H_d in Eq. 20 signifies the deviatoric damage hardening function. Conditions Eq. 22 are standard for problems involving unilateral constraint. If $\phi_d^d(\xi_{d,t}^+, g_{d,t}) < 0$, the deviatoric damage criterion is not satisfied and by conditions Eq. 22, $\dot{\mu}_d = 0$; hence, the deviatoric damage rule Eq. 20 implies that $\dot{d}_d = 0$ and no further deviatoric damage takes place. If, on the other hand, $\dot{\mu}_d > 0$; that is, further deviatoric damage ("tensile loading") is taking place, conditions Eq. 22 imply that $\phi_d^d\ (\xi_{d,t}^+, g_{d,t}) = 0$. In this case, the value of $\dot{\mu}_d$ is determined by the deviatoric damage consistency condition; i.e.,

$$\phi_d^d\left(\xi_{d,t}^+, g_{d,t}\right) = \dot{\phi}_d^d\left(\xi_{d,t}^+, g_{d,t}\right) = 0 \quad \Rightarrow \quad \dot{\mu}_d = \dot{\xi}_d^+ \tag{23}$$

and $g_{d,t}$ is defined by the expression

$$g_{d,t} = \max\left(g_{d,0}, \max_{s\in(-\infty,t)} \xi_{d,s}^+\right) - g_{dh,(t-1)} \tag{24}$$

where $g_{dh,(t-1)}$ is the value for the current deviatoric damage threshold to be numerically reduced due to incremental deviatoric healing (if any) from the previous time step.

If $H_d\ (\xi_{d,t}^+, d_{d,t})$ in Eq. 20 is independent of $\dot{d}_{d,t}$, the above formulation may be rephrased as follows. Let $G_d : \mathbf{R} \to \mathbf{R}_+$ be such that $H_d(\xi_{d,t}^+) \equiv \partial G_d(\xi_{d,t}^+)/\partial \xi_{d,t}^+$. $G_d(\bullet)$ is assumed monotonic. A deviatoric damage criterion entirely equivalent to condition Eq. 19 is given by $\overline{\phi}_d^d\left(\xi_{d,t}^+, g_{d,t}\right) \equiv G_d\left(\xi_{d,t}^+\right) - G_d\left(g_{d,t}\right) \leq 0$. The flow rule and loading/unloading conditions read

$$\dot{d}_{d,t} = \dot{\mu}_d \ \frac{\partial \overline{\phi}_d^d\left(\xi_{d,t}^+, g_{d,t}\right)}{\partial \xi_{d,t}^+}, \qquad g_{d,t} = \dot{\mu}_d \tag{25}$$

$$\dot{\mu}_d \geq 0, \quad \overline{\phi}_d^d\left(\xi_{d,t}^+, g_{d,t}\right) \leq 0, \quad \dot{\mu}_d \ \overline{\phi}_d^d\left(\xi_{d,t}^+, g_{d,t}\right) = 0. \tag{26}$$

Characterization of Initial-Elastic-Strain-Energy-Based Two-Parameter Healing Model

Similar to the characterization of volumetric and deviatoric damage in the previous section, the progressive recovery of mechanical properties of soils due to healing is

characterized by means of a two-parameter healing mechanism. Here, the notion of equivalent compressive volumetric strain ξ_v^- is used as the energy norm of the compressive volumetric strain tensor. Accordingly, it can be set

$$\xi_v^- \equiv \sqrt{\Psi_v^0\left(\boldsymbol{\varepsilon}_v^-\right)} = \sqrt{\frac{1}{2}K\left(\boldsymbol{\varepsilon}_{kk}^-\right)^2} \tag{27}$$

where $\boldsymbol{\varepsilon}_{kk}^- = \boldsymbol{\varepsilon}_{11}^- + \boldsymbol{\varepsilon}_{22}^- + \boldsymbol{\varepsilon}_{33}^-$ and $\boldsymbol{\varepsilon}^- \equiv \mathbf{P}^- : \boldsymbol{\varepsilon}$. The fourth-order tensor $\mathbf{P}^-$ denotes the "mode I" negative (compressive) projection tensor with components

$$\mathbf{P}_{ijkl}^-(\boldsymbol{\varepsilon}) = \frac{1}{2}\left(\mathbf{Q}_{ik}^- \ \mathbf{Q}_{jl}^- + \mathbf{Q}_{il}^- \ \mathbf{Q}_{jk}^-\right) \tag{28}$$

where $\mathbf{Q}^- = \underset{\sim}{\mathbf{1}} - \mathbf{Q}^+$.

A volumetric criterion $\phi_v^h\left(\xi_{v,t}^-, r_{v,t}\right) \leq 0$ can be defined, formulated in the strain space, with the following functional form:

$$\phi_v^h\left(\xi_{v,t}^-, r_{v,t}\right) \equiv \xi_{v,t}^- - r_{v,t} \leq 0, \qquad t \in \mathbf{R}_+ \tag{29}$$

Here, $r_{v,t}$ defines the volumetric healing threshold at the current time t. If $r_{v,0}$ denotes the initial volumetric healing threshold before any loading is applied, it must have $r_{v,t} \geq r_{v,0}$. Note that $r_{v,0}$ is considered as a characteristic material property. Condition Eq. 28 then states that volumetric healing in the material is initiated when the energy norm of the compressive volumetric strain tensor $\xi_{v,t}^-$ exceeds the initial volumetric healing threshold $r_{v,0}$. The evolutions of the volumetric healing variable R_v and the volumetric healing threshold $r_{v,t}$ are defined by the rate equations

$$\dot{R}_{v,t} = \dot{\zeta}_v \ Z_v\left(\xi_{v,t}^-, R_{v,t}\right) \tag{30}$$

$$\dot{r}_v = \dot{\zeta}_v \tag{31}$$

where $\dot{\zeta}_v \geq 0$ is a volumetric healing consistency parameter that defines the volumetric healing loading/unloading conditions according to the Kuhn-Tucker relations:

$$\dot{\zeta}_v \geq 0, \quad \phi_v^h\left(\xi_{v,t}^-, r_{v,t}\right) \leq 0, \quad \dot{\zeta}_v \ \phi_v^h\left(\xi_{v,t}^-, r_{v,t}\right) = 0 \tag{32}$$

Moreover, Z_v in Eq. 29 stands for the volumetric healing hardening function. Conditions Eq. 31 are standard for problems involving unilateral constraint. If $\phi_v^h(\xi_{v,t}^-, r_{v,t}) < 0$, the volumetric healing criterion is not satisfied and conditions Eq. 31, $\dot{\zeta}_v = 0$; hence, the volumetric healing rule Eq. 29 implies that $\dot{R}_v = 0$ and no further volumetric healing takes place. If, on the other hand, $\dot{\zeta}_v > 0$; that is, further

volumetric healing ("compressive loading") is taking place, conditions Eq. 31 imply that $\phi_v^h(\xi_{v,t}^-, r_{v,t}) = 0$. In this event, the value of $\dot{\zeta}_v$ is determined by the volumetric healing consistency condition; i.e.,

$$\phi_v^h\left(\xi_{v,t}^-, r_{v,t}\right) = \dot{\phi}_v^h\left(\xi_{v,t}^-, r_{v,t}\right) = 0 \quad \Rightarrow \quad \dot{\zeta}_v = \dot{\xi}^- \tag{33}$$

and $r_{v,t}$ is given by the expression

$$r_{v,t} = \max\left(r_{v,0}, \max_{s\in(-\infty,t)} \xi_{v,s}^-\right) - r_{vd,(t-1)} \tag{34}$$

where $r_{vd,(t-1)}$ is the value for the current volumetric healing threshold to be reduced due to the incremental volumetric damage (if any) from the previous time step.

If $Z_v(\xi_{v,t}^-, R_{v,t})$ in Eq. 29 is independent of $\dot{R}_{v,t}$, the above formulation may be rephrased as follows. Let $G_v^* : \mathbf{R} \to \mathbf{R}_+$ be such that $Z_v(\xi_{v,t}^-) \equiv \partial G_v^*(\xi_{v,t}^-)/\partial \xi_{v,t}^-$. $G_v^*(\bullet)$ is assumed monotonic. A volumetric healing criterion entirely equivalent to conditions Eq. 28 is given by $\overline{\phi}_v^h\left(\xi_{v,t}^-, r_{v,t}\right) \equiv G_v^*\left(\xi_{v,t}^-\right) - G^*\left(r_{v,t}\right) \le 0$. The flow rule and loading/unloading conditions become

$$\dot{R}_{v,t} = \dot{\zeta}_v \frac{\partial \overline{\phi}_v^h\left(\xi_{v,t}^-, r_{v,t}\right)}{\partial \xi_{v,t}^-}, \qquad r_{v,t} = \dot{\zeta}_v \tag{35}$$

$$\dot{\zeta}_v \ge 0, \quad \overline{\phi}_v^h\left(\xi_{tv}^-, r_{tv}\right) \le 0, \quad \dot{\zeta}_v \ \overline{\phi}_v^h\left(\xi_{tv}^-, r_{tv}\right) = 0 \tag{36}$$

Next, the notion of equivalent compressive deviatoric strain ξ_d^- is ascribed as the (undamaged) energy norm of the compressive deviatoric strain tensor. It can be defined that

$$\xi_d^- \equiv \sqrt{\Psi_d^0(\mathbf{e}^-)} = \sqrt{\frac{1}{2} G\, \mathbf{e}^- : \mathbf{e}^-} \tag{37}$$

where $\mathbf{e}^- = \boldsymbol{\varepsilon}^- - \frac{1}{3}\boldsymbol{\varepsilon}_{kk}^- \underline{\mathbf{1}}$ and $\boldsymbol{\varepsilon}^- \equiv \mathbf{P}^- : \boldsymbol{\varepsilon}$. Then, the state of deviatoric healing by using a deviatoric healing criterion $\phi_d^h(\xi_{d,t}^-, r_{d,t}) \le 0$ is characterized, formulated in the strain space, with the following functional form:

$$\phi_d^h\left(\xi_{d,t}^-, r_{v,t}\right) \equiv \xi_{d,t}^- - r_{d,t} \le 0, \qquad t \in \mathbf{R}_+ \tag{38}$$

Here, $r_{d,t}$ is the deviatoric healing threshold at the current time t. If $r_{d,0}$ denotes the initial deviatoric healing threshold before any loading is applied,

it must have $r_{d,t} \geq r_{d,0}$. Note that $r_{d,0}$ is also considered as a characteristic material property. Condition Eq. 37 then states that deviatoric healing in the material is initiated when the energy norm of the compressive deviatoric strain tensor $\xi^{-}_{d,t}$ exceeds the initial deviatoric healing threshold $r_{d,0}$. For the case, the evolution of the deviatoric healing variable R_d and the deviatoric healing threshold $r_{d,t}$ can be defined by the rate equations

$$\dot{R}_{d,t} = \dot{\zeta}_d \; Z_d\left(\xi^{-}_{d,t}, R_{d,t}\right) \tag{39}$$

$$\dot{r}_d = \dot{\zeta}_d \tag{40}$$

where $\dot{\zeta}_d \geq 0$ is a deviatoric healing consistency parameter that defines deviatoric healing loading/unloading conditions according to the Kuhn-Tucker relations:

$$\dot{\zeta}_d \geq 0, \quad \phi^h_d\left(\xi^{-}_{d,t}, r_{d,t}\right) \leq 0, \;\; \dot{\zeta}_d \, \phi^h_d\left(\xi^{-}_{d,t}, r_{d,t}\right) = 0 \tag{41}$$

Moreover, Z_d in Eq. 38 represents the deviatoric healing hardening function. Conditions Eq. 40 are standard for problems involving unilateral constraint. If $\phi^h_d\,(\xi^{-}_{d,t},\, r_{d,t}) < 0$, the deviatoric healing criterion is not satisfied and by conditions Eq. 40, $\dot{\zeta}_d = 0$; hence, the deviatoric healing rule Eq. 38 implies that $\dot{R}_d = 0$ and no further deviatoric healing takes place. By contrast, if $\dot{\zeta}_d > 0$, further deviatoric healing ("compressive loading") takes place and conditions Eq. 40 imply that $\phi^h_d\,(\xi^{-}_{d,t},\, r_{d,t}) = 0$. In this event, the value of $\dot{\zeta}_d$ is determined by the deviatoric healing consistency condition; i.e.,

$$\phi^h_d\left(\xi^{-}_{d,t}, r_{d,t}\right) = \dot{\phi}^h_d\left(\xi^{-}_{d,t}, r_{d,t}\right) = 0 \quad \Rightarrow \quad \dot{\zeta}_d = \dot{\xi}^{-}_d \tag{42}$$

and $r_{d,t}$ is given by the expression

$$r_{d,t} = \max\left(r_{d,0}, \max_{s\in(-\infty,t)} \xi^{-}_{d,s}\right) - r_{dd,(t-1)} \tag{43}$$

where $r_{dd,(t-1)}$ is the numerically reduced value for the current deviatoric healing threshold due to the incremental deviatoric damage (if any) from the previous time step.

If $Z_d(\xi^{-}_{d,t},\, R_{d,t})$ in Eq. 38 is independent of $\dot{R}_{d,t}$, the above formulation may be rephrased as follows. Let $G^*_d : \mathbf{R} \rightarrow \mathbf{R}_+$ be such that $Z_d(\xi^{-}_{d,t}) \equiv \partial G^*_d(\xi^{-}_{d,t})/\partial \xi^{-}_{d,t}$. $G^*_d(\bullet)$ is assumed monotonic. A deviatoric healing criterion entirely equivalent to conditions Eq. 37 is given by $\overline{\phi}^h_d\left(\xi^{-}_{d,t}, r_{d,t}\right) \equiv G^*_d\left(\xi^{-}_{d,t}\right) - G^*\left(r_{d,t}\right) \leq 0$. The flow rule and loading/unloading conditions become

$$\dot{R}_{d,t} = \dot{\zeta}_d \frac{\partial \overline{\Phi}_d^h\left(\xi_{d,t}^-, r_{d,t}\right)}{\partial \xi_{d,t}^-}, \qquad r_{d,t} = \dot{\zeta}_d \tag{44}$$

$$\dot{\zeta}_d \geq 0, \quad \overline{\Phi}_d^h\left(\xi_{d,t}^-, r_{d,t}\right) \leq 0, \quad \dot{\zeta}_d \ \overline{\Phi}_d^h\left(\xi_{d,t}^-, r_{d,t}\right) = 0. \tag{45}$$

Net (Combined) Effects of the Two-Parameter Damage and Healing Models

In the recent work (Ju et al. 2011a, b), the following characterization is clearly illustrated for the net effects of damage and healing is physically incorrect and misleading:

$$d^{net} = d(1 - R) \tag{46}$$

Then, a micromechanics-motivated scalar incremental form is proposed to compute the net effects of isotropic damage healing, d^{net}, as expressed in Eq. 32 of Ju et al. (2011a).

As for the net effects of ***two-parameter*** damage-healing mechanisms, two independent variables are employed (the volumetric variable d_v^{net} and the deviatoric variable d_d^{net}). The following interacting mechanisms are proposed:

Interim volumetric and deviatoric damage :

$$\hat{d}_{v,n+1} = d_{v,n}^{net} + \Delta d_{v,n+1}, \qquad \hat{d}_{d,n+1} = d_{d,n}^{net} + \Delta d_{d,n+1}$$

With volumetric and deviatoric healing :

$\left(\Delta R_{v,n+1} = 0 \text{ and } \Delta R_{d,n+1} = 0, \text{if there is no healing}\right)$:

$$\Delta d_{v,n+1}^h = -\hat{d}_{v,n+1} \ \Delta R_{v,n+1}, \qquad \Delta d_{d,n+1}^h = -\hat{d}_{d,n+1} \ \Delta R_{d,n+1}$$
$$d_{v,n+1}^{net} = \hat{d}_{v,n+1} + \Delta d_{v,n+1}^h, \qquad d_{d,n+1}^{net} = \hat{d}_{d,n+1} + \Delta d_{d,n+1}^h$$

Net incremental volumetric and deviatoric damage :

$$\Delta d_{v,n+1}^{net} = \Delta d_{v,n+1} - \hat{d}_{v,n+1} \ \Delta R_{v,n+1}, \qquad \Delta d_{d,n+1}^{net} = \Delta d_{d,n+1} - \hat{d}_{d,n+1} \ \Delta R_{d,n+1} \tag{47}$$

where $\Delta d_{v,n+1}$ is the incremental volumetric damage variable, $\Delta d_{d,n+1}$ means the incremental deviatoric damage variable, $\hat{d}_{v,n+1}$ is defined as the interim volumetric damage variable, $\hat{d}_{d,n+1}$ is the interim deviatoric damage variable, $\Delta d_{v,n+1}^h$ represents the incremental volumetric recovery due to the incremental volumetric

healing, and $\Delta d^{h}_{d,n+1}$ signifies the incremental deviatoric recovery due to the incremental deviatoric healing. Take the volumetric part, for instance, if there is no volumetric healing, the interim volumetric damage variable $\hat{d}_{v,n+1}$ is equal to $d_{v,n+1}{}^{net}$. If the volumetric healing exists, the net (combined) volumetric damage-healing variable $d^{net}_{v,n+1}$ must be modified by $\Delta d^{h}_{v,n+1}$.

With Δd^{net}_{v} and Δd^{net}_{d} known, the incremental fourth-rank damage tensor can be obtained by the following equation:

$$\Delta \mathbf{D} = \Delta d^{net}_{v}\ \underset{\sim}{\mathbf{1}} \otimes \underset{\sim}{\mathbf{1}} + \Delta d^{net}_{d}\ \mathbf{I}_{dev}. \tag{48}$$

Computational Algorithms: Two-Step Operator Splitting

In the previous sections, the new initial-elastic strain-energy-based two-parameter damage-healing models are proposed based on the effective stress concept. In this section, the focus is placed in detail on the computational aspects of the proposed new damage-healing models within the context of two-step operator splitting methodology. More precisely, the attention is placed on the following local two-parameter elastoplastic-damage-healing rate constitutive equations:

$$\begin{aligned}
&\dot{\varepsilon} = \nabla^{s}\dot{\mathbf{u}}(t)\\
&\begin{cases} \dot{d}_{v,t} = \dot{\mu}_{v} H_{v}\left(\xi^{+}_{v,t}, d_{v,t}\right)\\ \dot{g}_{v} = \dot{\mu}_{v}\\ \dot{\mu}_{v} \geq 0, \quad \phi^{d}_{v}\left(\xi^{+}_{v,t}, g_{v,t}\right) \leq 0, \quad \dot{\mu}_{v}\ \phi^{d}_{v}\left(\xi^{+}_{v,t}, g_{v,t}\right) = 0 \end{cases}\\
&\begin{cases} \dot{d}_{d,t} = \dot{\mu}_{d} H_{d}\left(\xi^{+}_{d,t}, d_{d,t}\right)\\ \dot{g}_{d} = \dot{\mu}_{d}\\ \dot{\mu}_{d} \geq 0, \quad \phi^{d}_{d}\left(\xi^{+}_{d,t}, g_{d,t}\right) \leq 0, \quad \dot{\mu}_{d}\ \phi^{d}_{d}\left(\xi^{+}_{d,t}, g_{d,t}\right) = 0 \end{cases}\\
&\begin{cases} \dot{R}_{v,t} = \dot{\zeta}_{v} Z_{v}\left(\xi^{-}_{v,t}, R_{v,t}\right)\\ \dot{r}_{v} = \dot{\zeta}_{v}\\ \dot{\zeta}_{v} \geq 0, \quad \phi^{h}_{v}\left(\xi^{-}_{v,t}, r_{v,t}\right) \leq 0, \quad \dot{\zeta}_{v}\ \phi^{h}_{v}\left(\xi^{-}_{v,t}, r_{v,t}\right) = 0 \end{cases}\\
&\begin{cases} \dot{R}_{d,t} = \dot{\zeta}_{d} Z_{d}\left(\xi^{-}_{d,t}, R_{d,t}\right)\\ \dot{r}_{d} = \dot{\zeta}_{d}\\ \dot{\zeta}_{d} \geq 0, \quad \phi^{h}_{d}\left(\xi^{-}_{d,t}, r_{d,t}\right) \leq 0, \quad \dot{\zeta}_{d}\ \phi^{h}_{d}\left(\xi^{-}_{d,t}, r_{d,t}\right) = 0 \end{cases}\\
&\Delta d^{net}_{v,n+1} = \Delta d_{v,n+1} - \hat{d}_{v,n+1}\ \Delta R_{v,n+1}; \quad \Delta d^{net}_{d,n+1} = \Delta d_{d,n+1} - \hat{d}_{d,n+1}\ \Delta R_{d,n+1}\\
&\Delta \mathbf{D} = \Delta d^{net}_{v}\ \underset{\sim}{\mathbf{1}} \otimes \underset{\sim}{\mathbf{1}} + \Delta d^{net}_{d}\ \mathbf{I}_{dev}
\end{aligned} \tag{49}$$

$$
\begin{aligned}
&\dot{\overline{\boldsymbol{\sigma}}} = \frac{d}{dt}\left[\frac{\partial \Psi^0(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}}\right] - \dot{\overline{\boldsymbol{\sigma}}}^{\mathbf{p}} \\
&\begin{cases}
\dot{\overline{\boldsymbol{\sigma}}}^{\mathbf{p}} = \dot{\lambda}\dfrac{\partial f}{\partial \boldsymbol{\varepsilon}}\left(\dfrac{\partial \Psi^0(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}} - \overline{\boldsymbol{\sigma}}^{\mathbf{p}}, \mathbf{q}\right) & (\textit{associative flow rule}) \\
\dot{\mathbf{q}} = \dot{\lambda}\mathbf{h}\left(\dfrac{\partial \Psi^0(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}} - \overline{\boldsymbol{\sigma}}^{\mathbf{p}}, \mathbf{q}\right) & (\textit{plastic hardening law}) \\
f\left(\dfrac{\partial \Psi^0(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}} - \overline{\boldsymbol{\sigma}}^{\mathbf{p}}, \mathbf{q}\right) \le 0 & (\textit{yield condition})
\end{cases}
\end{aligned}
\tag{50}
$$

From an algorithmic standpoint, the problem of integrating the evolution in Eqs. 46 and 47 reduces to updating the basic variables $\left\{\boldsymbol{\sigma}, d_v^{net}, d_d^{net}, \overline{\boldsymbol{\sigma}}^{\mathbf{p}}, \mathbf{q}\right\}$ in a manner consistent with the constitutive models. It is essential to realize that in this computational process, the history of strains $t \to \boldsymbol{\varepsilon} \equiv \nabla^s \mathbf{u}(t)$ is assumed to be given.

Equations of evolution are to be solved incrementally over a sequence of given time steps $[t_n, t_{n+1}] \subset R_+$, $n = 0, 1, 2 \ldots$. Therefore, the initial conditions for equations are

$$
\left\{\boldsymbol{\sigma}, d_v^{net}, d_d^{net}, \overline{\boldsymbol{\sigma}}^{\mathbf{p}}, \mathbf{q}\right\}|_{t=t_n} = \left\{\boldsymbol{\sigma}_n, d_{v,n}^{net}, d_{d,n}^{net}, \overline{\boldsymbol{\sigma}}_n^{\mathbf{p}}, \mathbf{q}_n\right\}
$$

In accordance with the notion of operator split, the following additive decomposition of problem of evolution is considered into the elastic-damage-healing predictor and the plastic corrector parts.

1.Elastic−damage−healing part:

$$
\begin{aligned}
&\dot{\boldsymbol{\varepsilon}} = \nabla^s \dot{\mathbf{u}}(t) \\
&\dot{d}_v = \begin{cases} H_v(\xi_v^+)\dot{\xi}_v^+, \ \textit{iff} \quad \phi_{v,t}^d = \dot{\phi}_{v,t}^d = 0 \\ 0, \ \textit{otherwise} \end{cases} ; \quad \dot{d}_d = \begin{cases} H_d(\xi_d^+)\dot{\xi}_d^+, \ \textit{iff} \quad \phi_{d,t}^d = \dot{\phi}_{d,t}^d = 0 \\ 0, \ \textit{otherwise} \end{cases} \\
&\dot{g}_v = \begin{cases} \dot{\xi}_v^+, \ \textit{iff} \quad \phi_{v,t}^d = \dot{\phi}_{v,t}^d = 0 \\ 0, \ \textit{otherwise} \end{cases} ; \quad \dot{g}_d = \begin{cases} \dot{\xi}_d^+, \ \textit{iff} \quad \phi_{d,t}^d = \dot{\phi}_{d,t}^d = 0 \\ 0, \ \textit{otherwise} \end{cases} \\
&\dot{R}_v = \begin{cases} Z_v(\xi_v^-)\dot{\xi}_v^-, \ \textit{iff} \quad \phi_{v,t}^h = \dot{\phi}_{v,t}^h = 0 \\ 0, \ \textit{otherwise} \end{cases} ; \quad \dot{R}_d = \begin{cases} Z_d(\xi_d^-)\dot{\xi}_d^-, \ \textit{iff} \quad \phi_{d,t}^h = \dot{\phi}_{d,t}^h = 0 \\ 0, \ \textit{otherwise} \end{cases} \\
&\dot{r}_v = \begin{cases} \dot{\xi}_v^-, \ \textit{iff} \quad \phi_{v,t}^h = \dot{\phi}_{v,t}^h = 0 \\ 0, \ \textit{otherwise} \end{cases} ; \quad \dot{r}_d = \begin{cases} \dot{\xi}_d^-, \ \textit{iff} \quad \phi_{d,t}^h = \dot{\phi}_{d,t}^h = 0 \\ 0, \ \textit{otherwise} \end{cases} \\
&\Delta d_{v,n+1}^{net} = \Delta d_{v,n+1} - \hat{d}_{v,n+1}\Delta R_{v,n+1} \quad ; \quad \Delta d_{d,n+1}^{net} = \Delta d_{d,n+1} - \hat{d}_{d,n+1}\Delta R_{d,n+1} \\
&\Delta D = \Delta d_v^{net}\, \underset{\sim}{\mathbf{1}} \otimes \underset{\sim}{\mathbf{1}} + \Delta d_d^{net}\, \mathbf{I}_{dev} \\
&\dot{\overline{\boldsymbol{\sigma}}} = \frac{d}{dt}\left[\frac{\partial \Psi^0(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}}\right] \\
&\dot{\overline{\boldsymbol{\sigma}}}^{\mathbf{p}} = \mathbf{0} \\
&\dot{\mathbf{q}} = \mathbf{0}
\end{aligned}
\tag{51}
$$

2.Plastic part:

$$\begin{aligned}
&\dot{\boldsymbol{\varepsilon}} = \mathbf{0}\\
&\dot{d}_v^{net} = 0\\
&\dot{d}_d^{net} = 0\\
&\dot{r}_v = 0\\
&\dot{r}_d = 0\\
&\dot{\overline{\boldsymbol{\sigma}}} = -\dot{\overline{\boldsymbol{\sigma}}}^{\mathbf{p}}\\
&\dot{\overline{\boldsymbol{\sigma}}}^{\mathbf{p}} = \dot{\lambda}\frac{\partial f}{\partial \boldsymbol{\varepsilon}}\left(\frac{\partial \Psi^0(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}} - \overline{\boldsymbol{\sigma}}^{\mathbf{p}}, \mathbf{q}\right)\\
&\dot{\mathbf{q}} = \dot{\lambda}\mathbf{h}\left(\frac{\partial \Psi^0(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}} - \overline{\boldsymbol{\sigma}}^{\mathbf{p}}, \mathbf{q}\right)
\end{aligned} \tag{52}$$

It is noted that Eqs. 48 and Eq. 49 do indeed add up to Eqs. 46 and 47, in agreement with the notion of operator split. The formulation of an algorithm consistent with equations is based on the following fundamental result concerning operator split method. Given the two algorithms, the first one is consistent with problem Eq. 48 (the elastic-damage-healing predictor) and the second one is consistent with problem Eq. 49 (the plastic return mapping corrector). In turn, the product algorithm obtained by successive application of these two algorithms is consistent with the original problem.

The Two-Parameter Elastic-Damage-Healing Predictor

An algorithm consistent with problems Eq. 48 and Eq. 49, referred to as the two-parameter elastic-damage-healing predictor, is rendered by the following step-by-step procedure.

Step 1: Strain update: Given the incremental displacement field $\mathbf{u}_{n+1}$, the strain tensor is updated as

$$\boldsymbol{\varepsilon}_{n+1} = \boldsymbol{\varepsilon}_n + \nabla^s \mathbf{u}_{n+1}.$$

Step 2: Compute the "mode I positive 4th rank projection operator" $\mathbf{P}^+_{n+1}$ (different from Simo and Ju 1987a, b, 1989 and Ju 1989a), based on the **total** strain tensor $\boldsymbol{\varepsilon}_{n+1}$:

$$\mathbf{P}^+_{ijkl}(\boldsymbol{\varepsilon}_{n+1}) = \frac{1}{2}\left(\mathbf{Q}^+_{ik}\ \mathbf{Q}^+_{jl} + \mathbf{Q}^+_{il}\ \mathbf{Q}^+_{jk}\right) \tag{53}$$

where $\mathbf{Q}^+_{n+1} = \sum_{i=1}^{2} \hat{H}(\boldsymbol{\varepsilon}_i)\mathbf{p}_i \otimes \mathbf{p}_i; \varepsilon_{n+1} = \sum_{i=1}^{2} \boldsymbol{\varepsilon}_i\ \mathbf{p}_i \otimes \mathbf{p}_i$ (for 2D simulations in this chapter) and $\hat{H}(\boldsymbol{\varepsilon}_i)$ is the smoothed Heaviside function.

Step 3: Compute the "negative 4th rank projection operator" $\mathbf{P}^-_{n+1}$ based on the **total** strain tensor $\boldsymbol{\varepsilon}_{n+1}$:

$$\mathbf{P}^{-}_{ijkl}(\boldsymbol{\varepsilon}_{n+1}) = \frac{1}{2}\left(\mathbf{Q}^{-}_{ik}\ \mathbf{Q}^{-}_{jl} + \mathbf{Q}^{-}_{il}\ \mathbf{Q}^{-}_{jk}\right) \tag{54}$$

where $\mathbf{Q}^{-}_{n+1} = \underset{\sim}{\mathbf{1}} - \mathbf{Q}^{+}_{n+1}$.

Step 4: Compute the initial (undamaged) elastic tensile and compressive volumetric and deviatoric strain energies, $\xi^{+}_{v,n+1}$, $\xi^{+}_{d,n+1}$, $\xi^{-}_{v,n+1}$ and $\xi^{-}_{d,n+1}$.

$$\xi^{+}_{v,n+1} \equiv \sqrt{\Psi^{0}_{v}\left(\boldsymbol{\varepsilon}^{+}_{v,n+1}\right)} = \sqrt{\frac{1}{2}K\left(\boldsymbol{\varepsilon}^{+}_{kk,n+1}\right)^{2}}, \quad \text{if linear elasticity} \tag{55}$$

$$\xi^{+}_{d,n+1} \equiv \sqrt{\Psi^{0}_{d}\left(\mathbf{e}^{+}_{n+1}\right)} = \sqrt{\frac{1}{2}G\,\mathbf{e}^{+}_{n+1} : \mathbf{e}^{+}_{n+1}} \tag{56}$$

$$\xi^{-}_{v,n+1} \equiv \sqrt{\Psi^{0}_{v}\left(\boldsymbol{\varepsilon}^{-}_{v,n+1}\right)} = \sqrt{\frac{1}{2}K\left(\boldsymbol{\varepsilon}^{-}_{kk,n+1}\right)^{2}}, \quad \text{if linear elasticity} \tag{57}$$

$$\xi^{-}_{d,n+1} \equiv \sqrt{\Psi^{0}_{d}\left(\mathbf{e}^{-}_{n+1}\right)} = \sqrt{\frac{1}{2}G\,\mathbf{e}^{-}_{n+1} : \mathbf{e}^{-}_{n+1}} \tag{58}$$

where $\boldsymbol{\varepsilon}^{+}_{kk} = \boldsymbol{\varepsilon}^{+}_{11} + \boldsymbol{\varepsilon}^{+}_{22} + \boldsymbol{\varepsilon}^{+}_{33}$, $\mathbf{e}^{+} = \boldsymbol{\varepsilon}^{+} - \frac{1}{3}\boldsymbol{\varepsilon}^{+}_{kk}\ \underset{\sim}{\mathbf{1}}$, $\boldsymbol{\varepsilon}^{+} = \mathbf{P}^{+} : \boldsymbol{\varepsilon}$, and $\boldsymbol{\varepsilon}^{-} = \mathbf{P}^{-} : \boldsymbol{\varepsilon}$.

Step 5: Update the two scalar damage parameter thresholds $\left(\tilde{d}_{v}\right)_{n+1}$ and $\left(\tilde{d}_{d}\right)_{n+1}$. For the **volumetric** part:

(a) If $(\Delta R_{v})_{n} > 0$ and $\left(\tilde{d}_{v}\right)_{n} > 0$, update the volumetric damage parameter threshold:

$$\left(\tilde{d}_{v}\right)_{n+1} = \left(\tilde{d}_{v}\right)_{n}\ \left(1.0 - (\Delta R_{v})_{n}\right) \tag{59}$$

(b) Otherwise, set

$$\left(\tilde{d}_{v}\right)_{n+1} = \left(\tilde{d}_{v}\right)_{n} \tag{60}$$

For the **deviatoric** part:

(a) If $(\Delta R_{d})_{n} > 0$ and $\left(\tilde{d}_{d}\right)_{n} > 0$, update the deviatoric damage parameter threshold:

$$\left(\tilde{d}_{d}\right)_{n+1} = \left(\tilde{d}_{d}\right)_{n}\left(1.0 - (\Delta R_{d})_{n}\right) \tag{61}$$

(b) Otherwise, set

$$\left(\tilde{d}_{d}\right)_{n+1} = \left(\tilde{d}_{d}\right)_{n} \tag{62}$$

Step 6: Compute the two scalar damage predictors $(d_v)_{n+1}$ and $(d_d)_{n+1}$. Define:

UBVD = the upper bound of the volumetric damage threshold ($0 <$ UBVD < 1).

ξ^+_{UBVD} = the upper bound of tensile volumetric strain energy obtained from the volumetric damage functional evaluation for specific UBVD.

UBDD = the upper bound of deviatoric damage threshold ($0 <$ UBDD < 1).

ξ^+_{UBVD} = the upper bound of tensile deviatoric strain energy.

For the **volumetric** part:

(a) If $\xi^+_{v,n+1} \leq \xi^{0+}_{vol}$, no further volumetric damage, set $\Delta(d_v)_{n+1} = 0$, where ξ^{0+}_{vol} is the initial volumetric damage threshold.

(b) If $\xi^{0+}_{vol} < \xi^+_{v,n+1} < \xi^+_{UBVD}$, compute $(d_v)_{n+1}$ using a nonlinear damage functional evaluation, such as one of the following two functions (Wu et al. 2005):

$$(d_v)_{n+1}\left(\xi^+_{v,n+1}\right) = \frac{(k_v)_c \left[\xi^+_{v,n+1} - (k_v)_i\right]}{\xi^+_{v,n+1}\left[(k_v)_c - (k_v)_i\right]} \tag{63}$$

where $(k_v)_c$ and $(k_v)_i$ are material constants for the volumetric damage evolution.

$$(d_v)_{n+1}\left(\xi^+_{v,n+1}\right) = 1 - \frac{A_v\ (1 - B_v)}{\xi^+_{v,n+1}} - B_v\ \exp\left[A_v - \xi^+_{v,n+1}\right] \tag{64}$$

where A_v and B_v are material constants for the volumetric damage evolution.

(c) If $\xi^+_{v,n+1} \geq \xi^+_{UBVD}$, set $(d_v)_{n+1} =$ UBVD and a counter N_v. If $N_v > 1$, no further volumetric damage,
set $\Delta(d_v)_{n+1} = 0$.

For the **deviatoric** part:

(a) If $\xi^+_{d,n+1} \leq \xi^{0+}_{dev}$, no further deviatoric damage, set $\Delta(d_d)_{n+1} = 0$, where ξ^{0+}_{dev} is the initial deviatoric damage threshold.

(b) $\xi^{0+}_{dev} < \xi^+_{d,n+1} < \xi^+_{UBDD}$, compute $(d_d)_{n+1}$ using a nonlinear damage functional evaluation, such as one of the following two functions (Wu et al. 2005):

$$(d_d)_{n+1}\left(\xi^+_{d,n+1}\right) = \frac{(k_d)_c \left[\xi^+_{d,n+1} - (k_d)_i\right]}{\xi^+_{d,n+1}\left[(k_d)_c - (k_d)_i\right]} \tag{65}$$

$$(d_d)_{n+1}\left(\xi^+_{d,n+1}\right) = 1 - \frac{A_d\ (1 - B_d)}{\xi^+_{d,n+1}} - B_d\ \exp\left[A_d - \xi^+_{d,n+1}\right] \tag{66}$$

(a) If $\xi^+_{d,n+1} \geq \xi^+_{UBDD}$, set $(d_d)_{n+1} =$ UBDD and a counter N_d. If $N_d > 1$, no further deviatoric damage, set $\Delta(d_d)_{n+1} = 0$.

If both $\Delta(d_v)_{n+1} = 0$ and $\Delta(d_d)_{n+1} = 0$, go to **Step 10**.

Step 7: Check the two incremental scalar damage criteria $\Delta(d_v)_{n+1}$ and $\Delta(d_d)_{n+1}$:

$$\Delta(d_v)_{n+1} = (d_v)_{n+1} - \left(\tilde{d}_v\right)_{n+1} \tag{67}$$

$$\Delta(d_d)_{n+1} = (d_d)_{n+1} - \left(\tilde{d}_d\right)_{n+1} \tag{68}$$

If $\Delta\xi^{+}_{n+1} \leq 0$, then set $\Delta(d_v)_{n+1} = 0$ and $\Delta(d_d)_{n+1} = 0$.

Only when both $\Delta(d_v)_{n+1} \leq 0$ and $\Delta(d_d)_{n+1} \leq 0$, go to **Step 10**.

Otherwise, if $\Delta(d_v)_{n+1} > 0$, set $\left(\tilde{d}_v\right)_{n+1} = (d_v)_{n+1}$; continue.

If $\Delta(d_d)_{n+1} > 0$, set $\left(\tilde{d}_d\right)_{n+1} = (d_d)_{n+1}$; continue

where $\left(\tilde{d}_v\right)_n$ and $\left(\tilde{d}_d\right)_n$ denote the scalar volumetric and deviatoric damage parameter thresholds from the previous time step.

Step 8: Compute the incremental two-parameter damage predictor tensor $\Delta\mathbf{D}_{n+1}$ and the updated nonsymmetric interim two-parameter damage tensor $\hat{\mathbf{D}}_{n+1}$:

$$\Delta\mathbf{D}_{n+1} = (\Delta d_v)_{n+1}\underset{\sim}{\mathbf{1}} \otimes \underset{\sim}{\mathbf{1}} + (\Delta d_d)_{n+1}\mathbf{I}_{dev} \tag{69}$$

$$\hat{\mathbf{D}}_{n+1} = \mathbf{D}_n^{net} + \Delta\mathbf{D}_{n+1} \tag{70}$$

Step 9: Update the two scalar healing parameter thresholds $\left(\tilde{R}_v\right)_{n+1}$ and $\left(\tilde{R}_d\right)_{n+1}$:

$\left(\tilde{R}_v\right)_{n+1}$ and $\left(\tilde{R}_d\right)_{n+1}$ denote the net amount of volumetric and deviatoric healing sustained after damage takes place.

For the **volumetric part**:

(a) If $(\Delta d_v)_{n+1} > 0$ and $\left(\tilde{R}_v\right)_n > 0$, update the volumetric healing threshold for next step:

$$\left(\tilde{R}_v\right)_{n+1} = \left(\tilde{R}_v\right)_n \left(1.0 - (\Delta d_v)_{n+1}\right) \tag{71}$$

(b) Otherwise, set:

$$\left(\tilde{R}_v\right)_{n+1} = \left(\tilde{R}_v\right)_n \tag{72}$$

For the **deviatoric part**:

(a) If $(\Delta d_d)_{n+1} > 0$ and $\left(\tilde{R}_d\right)_n > 0$, update the deviatoric healing threshold for next step:

$$\left(\tilde{R}_d\right)_{n+1} = \left(\tilde{R}_d\right)_n \left(1.0 - (\Delta d_d)_{n+1}\right) \tag{73}$$

(b) Otherwise, set:

$$\left(\tilde{R}_d\right)_{n+1} = \left(\tilde{R}_d\right)_n \tag{74}$$

Step 10: Compute the two scalar healing predictors $(R_v)_{n+1}$ and $(R_d)_{n+1}$. Define:

UBVH = the upper bound of the volumetric healing threshold ($0 <$ UBVH < 1).

$\bar{\xi}_{UBVH}$ = the upper bound of compressive volumetric strain energy obtained from the volumetric healing functional evaluation for specific UBVH (set as 0.6 for example).

UBDH = the upper bound of the deviatoric healing threshold ($0 <$ UBDH < 1).

$\bar{\xi}_{UBVH}$ = the upper bound of compressive deviatoric strain energy obtained from the deviatoric damage functional evaluation for specific UBDH (set as 0.6 for example).

For the **volumetric part**:

(a) If $\bar{\xi}_{v,n+1} \leq \bar{\xi}^0_{vol}$, no further volumetric healing, set $\Delta(R_v)_{n+1} = 0$

where $\bar{\xi}^0_{vol}$ is the initial volumetric healing threshold.

(b) If $\bar{\xi}^0_{vol} < \bar{\xi}_{v,n+1} < \bar{\xi}_{UBVH}$, compute $(R_v)_{n+1}$ using a nonlinear healing functional evaluation, such as the following function (Wu et al. 2005):

$$(R_v)_{n+1}\left(\bar{\xi}_{v,n+1}\right) = \frac{(h_v)_c \left[\bar{\xi}_{v,n+1} - (h_v)_i\right]}{\bar{\xi}_{v,n+1}\left[(h_v)_c - (h_v)_i\right]} \tag{75}$$

where $(h_v)_c$ and $(h_v)_i$ are material constants for the volumetric healing evolution.

Or, use the following function:

$$(R_v)_{n+1}\left(\bar{\xi}_{v,n+1}\right) = 1 - \frac{\overline{A}_v\ (1\ -\ \overline{B}_v)}{\bar{\xi}_{v,n+1}} - \overline{B}_v\ \exp\left[\overline{A}_v - \bar{\xi}_{v,n+1}\right] \tag{76}$$

where $\overline{A}_v$ and $\overline{B}_v$ are material constants for volumetric healing evolution.

(c) If $\bar{\xi}_{v,n+1} \geq \bar{\xi}_{UBVH}$, set $(R_v)_{n+1} =$ UBVH and a counter M_v.

If $M_v > 1$, no further volumetric healing, set $\Delta(R_v)_{n+1} = 0$

For the **deviatoric part**:

(a) If $\bar{\xi}_{d,n+1} \leq \bar{\xi}^0_{dev}$, no further deviatoric healing, set $\Delta(R_d)_{n+1} = 0$

where $\bar{\xi}^0_{dev}$ is the initial deviatoric healing threshold.

(b) If $\bar{\xi}^0_{dev} < \bar{\xi}_{d,n+1} < \bar{\xi}_{UBDH}$, compute $(R_d)_{n+1}$ using a nonlinear healing functional evaluation, such as the following function (Wu et al. 2005):

$$(R_d)_{n+1}\left(\bar{\xi}_{d,n+1}\right) = \frac{(h_d)_c \quad \left[\bar{\xi}_{d,n+1} \; - \; (h_d)_i\right]}{\bar{\xi}_{d,n+1} \quad \left[(h_d)_c \; - \; (h_d)_i\right]} \tag{77}$$

where $(h_d)_c$ and $(h_d)_i$ are material constants for the deviatoric healing evolution.

$$(R_d)_{n+1}\left(\bar{\xi}_{d,n+1}\right) \;\; = \;\; 1 - \frac{\overline{A}_d \;\; (1 \; - \; \overline{B}_d)}{\bar{\xi}_{d,n+1}} - \overline{B}_d \;\; \exp\left[\overline{A}_d - \bar{\xi}_{d,n+1}\right] \tag{78}$$

(c) where $\overline{A}_d$ and $\overline{B}_d$ are material constants for deviatoric healing evolution.

If $\bar{\xi}_{d,n+1} \geq \bar{\xi}_{UBDH}$, set $(R_d)_{n+1} =$ UBDH and a counter M_d.
If $M_d > 1$, no further deviatoric healing, set $\Delta(R_d)_{n+1} = 0$.
Only when both $\Delta(R_v)_{n+1} = 0$ and $\Delta(R_d)_{n+1} = 0$, go to **Step 14**.
Otherwise, go to **Step 11**.

Step 11: Check the two incremental scalar healing criteria $\Delta(R_v)_{n+1}$ and $\Delta(R_d)_{n+1}$:

$$\Delta(R_v)_{n+1} = (R_v)_{n+1} - \left(\tilde{R}_v\right)_{n+1} \tag{79}$$

$$\Delta(R_d)_{n+1} = (R_d)_{n+1} - \left(\tilde{R}_d\right)_{n+1} \tag{80}$$

If $\Delta\bar{\xi}_{n+1} \leq 0$, then set $\Delta(R_v)_{n+1} = 0$ and $\Delta(R_d)_{n+1} = 0$.
Only when both $\Delta(R_v)_{n+1} \leq 0$ and $\Delta(R_d)_{n+1} \leq 0$, go to **Step 14**.
Otherwise, if $\Delta(R_v)_{n+1} > 0$, set $\left(\tilde{R}_v\right)_{n+1} = (R_v)_{n+1}$.
If $\Delta(R_d)_{n+1} > 0$, set $\left(\tilde{R}_d\right)_{n+1} = (R_d)_{n+1}$.
Where $\left(\tilde{R}_v\right)_n$ and $\left(\tilde{R}_d\right)_n$ denote the scalar volumetric and deviatoric healing parameter thresholds from the previous time step.

Step 12: Compute the two-parameter incremental healing tensor $\Delta\mathbf{R}_{n+1}$:

$$\Delta\mathbf{R}_{n+1} = (\Delta R_v)_{n+1} \;\; \underset{\sim}{\mathbf{1}} \otimes \underset{\sim}{\mathbf{1}} + (\Delta R_d)_{n+1} \;\; \mathbf{I}_{dev} \tag{81}$$

Step 13: Compute the incremental healing corrector, and update the two-parameter net (combined) damage-healing tensor:

$$\begin{aligned} \Delta\mathbf{D}^H_{n+1} \;\; &= -\hat{\mathbf{D}}_{n+1} \bullet \Delta\mathbf{R}_{n+1} \\ \mathbf{D}^{net}_{n+1} \;\; &= \hat{\mathbf{D}}_{n+1} + \Delta\mathbf{D}^H_{n+1} \end{aligned} \tag{82}$$

Trial (predictor) stress: By substitution into the potential for the stress tensor, it can be obtained:

$$\begin{aligned} \boldsymbol{\sigma}^0_{n+1} &= \frac{\partial \Psi^0(\boldsymbol{\varepsilon}_{n+1})}{\partial \boldsymbol{\varepsilon}} \\ \overline{\boldsymbol{\sigma}}^{trial}_{n+1} &= \boldsymbol{\sigma}^0_{n+1} - \overline{\boldsymbol{\sigma}}^{\mathrm{p}}_n \\ \mathbf{q}^{trial}_{n+1} &= \mathbf{q}_n \end{aligned} \tag{83}$$

The Effective Plastic Return Mapping Corrector

To develop an algorithm consistent with the plastic part of the operator split, one first checks the loading/unloading conditions.

Step 14: Check plastic yielding and active mode. The algorithmic counterpart of the Kuhn-Tucker conditions is trivially implemented in terms of the elastic-damage trial stress. One simply checks

$$f\left(\overline{\boldsymbol{\sigma}}_{n+1}^{trial}, \mathbf{q}_{n+1}^{trial}\right)\begin{cases} \leq 0 & elastic-damage-healing\ predictor = final\ state \\ > 0 & plastic \Rightarrow return\ mapping \end{cases} \tag{84}$$

Multi-surface plasticity: In the case of plastic loading, it is necessary to determine the active plastic surface, such as the CAP model.

Step 15: Effective plastic return mapping corrector. In the event of plastic loading, the predictor stresses and internal variables are "returned back" to the yield surface along the algorithmic counterpart of the flow generated by Eq. 49. The algorithmic construction of this flow follows a proposed procedure inspired by a form of Kelley's convex cutting plane method for nonlinear optimization, with its basic structure inherited from the Newton's method. Two fundamental advantages of this procedure are (a) the quadratic rate of convergence toward the yield surface and (b) the need for computing the gradient of the flow rule and hardening law are bypassed.

Step 16: Update the homogenized (nominal) stress $\boldsymbol{\sigma}_{n+1}$:

$$\boldsymbol{\sigma}_{n+1} = \left[\mathbf{I} - \mathbf{D}_{n+1}^{net}\right] : \overline{\boldsymbol{\sigma}}_{n+1}. \tag{85}$$

Numerical Simulations of Soil Compression, Excavation, and Compaction Motions

In order to demonstrate that the new two-parameter elastoplastic damage-healing models can generate more versatile simulations than the hybrid isotropic elastoplastic damage-healing models proposed in our previous work (Ju et al. 2011a, b), the same numerical examples of soil compression, excavation, and compaction are performed to render a fair comparison. The associated soil properties are listed in Table 1; cf. Wu et al. (2005). The Drucker-Prager associative multi-surface plasticity formulation is employed to model the soil behavior for simplicity; cf. Simo et al. (1988), Wu et al. (2001), and Chen et al. (2009).

The initial configuration and discretization of the earth-moving process is exhibited in Fig. 1. The blade of the bulldozer is treated as a rigid body and thus represented by two contact surfaces in black color. To model the contact between soil particles and blade surfaces, the traditional penalty formulation is adopted. A layer of soil with dimension 4 m $\times$ 2 m is discretized into $41 \times 21 = 861$ uniformly distributed particles. As an important extension of

Table 1 The associated soil properties

Young's modulus	24.7 (MPa)
Poisson's ratio	0.35
Density	$1.88 \times 10^3 (\text{kg/m}^3)$
Cohesion	0.19 (MPa)
Lame constants	$\lambda = 2.13457 \times 10^7 (\text{Pa})$ $\mu = 9.14815 \times 10^7 (\text{Pa})$
Yield stress	$1.88312 \times 10^5 (\text{Pa})$

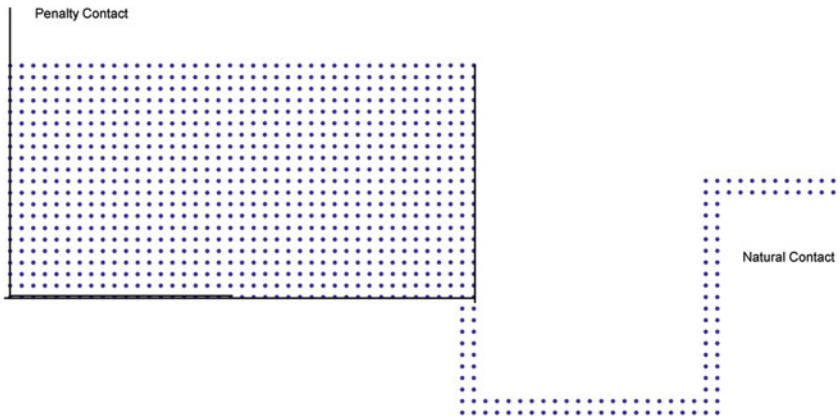


Fig. 1 The initial configuration and discretization of the earth-moving process

semi-Lagrangian discretization, a new particle-based "partition of unity" contact algorithm was developed (Wu et al. 2005). The natural contact algorithm (Wu et al. 2005) is applied to model the contact between soil particles and ground surfaces, which is found on the right-hand side in Fig. 1. The blade in Fig. 1 is controlled to move horizontally to the right for 0.3 m, lift the soil vertically for 3 m, move horizontally to the right for 1.8 m, rotate 45° to dump the soil over the wall, and then rotate back to move forward and downward to compact the soil in the ground pit.

For the new two-parameter elastoplastic damage-healing models, in addition to the material parameters in damage and healing evolutions, there are four more material parameters to control the soil behavior under complex numerical simulations. These include the initial volumetric damage threshold ξ_{vol}^{0+}, initial deviatoric damage threshold ξ_{dev}^{0+}, initial volumetric healing threshold ξ_{vol}^{0-}, and initial deviatoric healing threshold ξ_{dev}^{0-}. In the numerical simulations, the following two assumptions are made for simplicity. (i) The soil complies with the volumetric and deviatoric damage-healing evolutions as described in Eqs. 60, 62, 72, and 74.

Table 2 A parametric study on ξ_{vol}^{0+}, ξ_{dev}^{0+}, ξ_{vol}^{0-}, and ξ_{dev}^{0-}

	Initial volumetric damage threshold ξ_{vol}^{0+}	Initial deviatoric damage threshold ξ_{dev}^{0+}	Initial volumetric healing threshold ξ_{vol}^{0-}	Initial deviatoric healing threshold ξ_{dev}^{0-}
Hybrid isotropic elastoplastic damage-healing model				
	0.050	0.050	0.080	0.080
Two-parameter elastoplastic damage-healing models				
Effect of volumetric damage	**0.040**	0.050	Infinite	Infinite
	0.045			
	0.055			
	0.060			
Effect of deviatoric damage	0.050	**0.040**	Infinite	Infinite
		0.045		
		0.055		
		0.060		
Effect of volumetric healing	0.050	0.050	**0.070**	0.080
			0.075	
			0.085	
			0.090	
Effect of deviatoric healing	0.050	0.050	0.080	**0.070**
				0.075
				0.085
				0.090

(ii) The volumetric and deviatoric damage-healing evolutions use the same material parameters, i.e., $(k_v)_c = (k_d)_c = (h_v)_c = (h_d)_c = 0.05$ and $(k_v)_i = (k_d)_i = (h_v)_i = (h_d)_i = 50$ for ductile soils.

Subsequently, a parametric study is performed to compare the new two-parameter damage-healing models with the previous hybrid isotropic damage-healing model, as rendered in Table 2, on the initial volumetric damage threshold, the initial deviatoric damage threshold, the initial volumetric healing threshold, and the initial deviatoric healing threshold.

Table 2 clearly shows that while the hybrid isotropic elastoplastic damage and healing models can generate only one numerical simulation in our previous work (Ju et al. 2011a, b), the new two-parameter elastoplastic damage and healing models can have multiple versatile combinations of numerical simulations by slightly adjusting ξ_{vol}^{0+}, ξ_{del}^{0+}, ξ_{vol}^{0-}, and ξ_{dev}^{0-} for the underlying soil compression, excavation, and compaction operations.

In this section, selected numerical results are presented to illustrate the salient features and versatile simulation capability of the proposed damage-healing models. Furthermore, the comparisons of progressive deformations at various time steps for the effects of initial deviatoric damage threshold are displayed in

Fig. 2 and for the effects of initial volumetric and deviatoric healing thresholds are rendered in Fig. 3. In these figures, the blue solid circles represent the soil particles with no volumetric (deviatoric) damage, the red solid circles symbolize the soil particles with full volumetric (deviatoric) damage (the upper bound damage UBVD and UBDD in these simulations), and the solid circles in other colors (such as green, yellow, and orange) signify partial volumetric (deviatoric) damage between 0 and the upper bound damage.

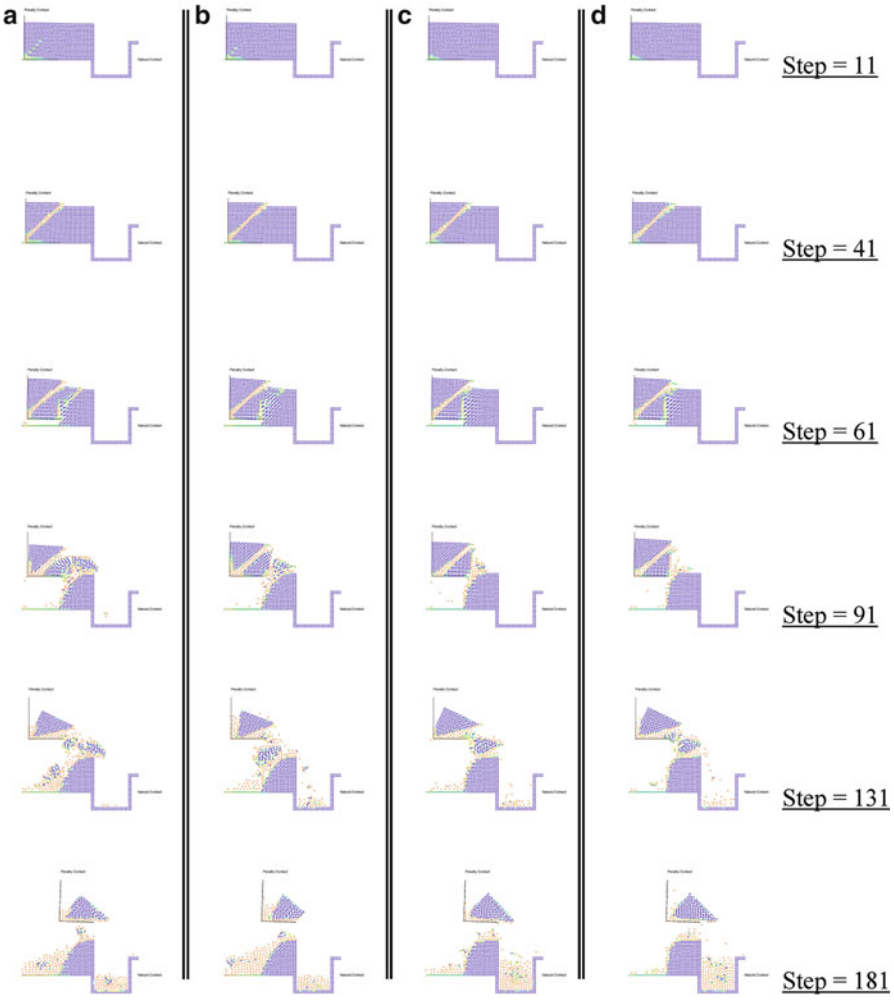


Fig. 2 (continued)

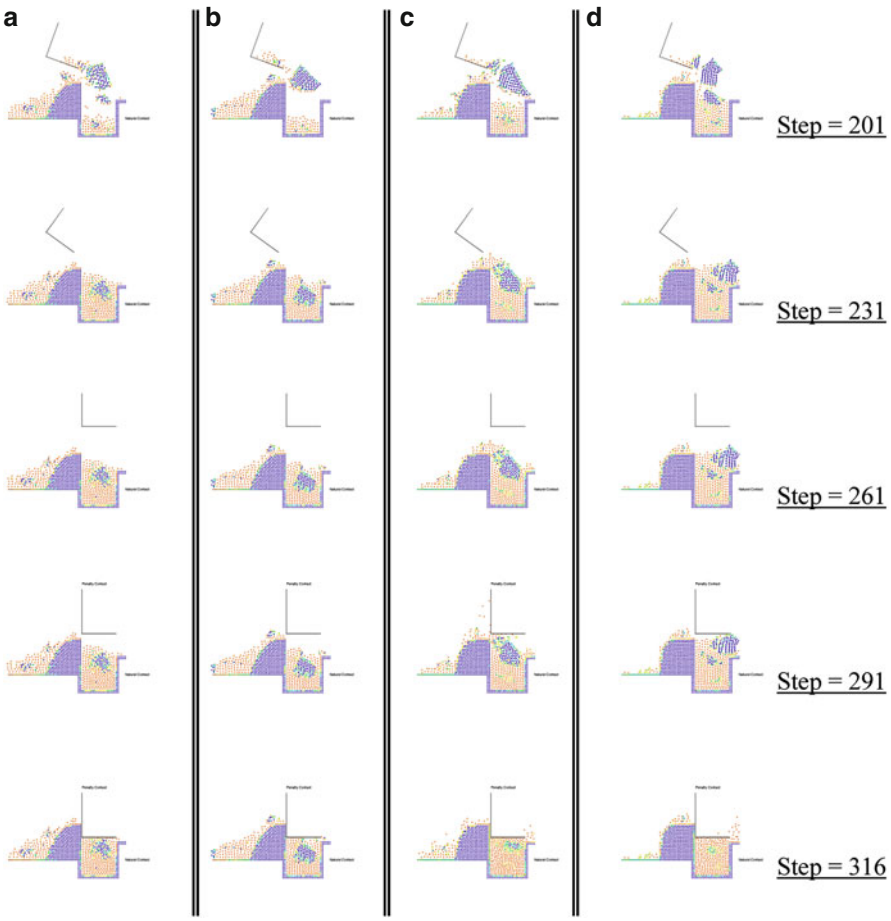


Fig. 2 A parametric study on the initial deviatoric damage threshold ξ_{dev}^{0+} ($\xi_{vol}^{0+} = 0.05$, $\xi_{vol}^{0-} = \xi_{dev}^{0-} =$ infinite) of deformation of soils for various stages of the earth-moving simulations by the two-parameter damage-healing models

Conclusions

In this chapter, new ***initial-elastic-strain-energy-based*** coupled elastoplastic ***two-parameter*** damage-healing formulations have been presented for the simulations of complex earth-moving processes. The equivalent *strain* based on $\mathbf{P}^{+}$ and $\mathbf{P}^{-}$ in volumetric and deviatoric parts for energy criteria of the volumetric and deviatoric damage and healing are employed for the proposed new models. Innovative two-step operator split algorithms are presented, featuring the volumetric and

deviatoric elastic-damage-healing predictors and the effective plastic return mapping corrector. In addition, the coupling of the volumetric and deviatoric damage and healing mechanisms is characterized in an incremental form by using the ***predictor formula*** and ***corrector formula***. In the absence of associated experimental data, the proposed two-parameter innovative formulations and algorithms for soils only illustrate their significant versatility and the salient damage-healing features. Experimental validations for the key parameters in damage and healing evolutions will be performed once the associated experiment data becomes available. Fully saturated soils are considered in this chapter. The effects of moisture and matric suction for partially saturated soils will be considered in the next chapter.

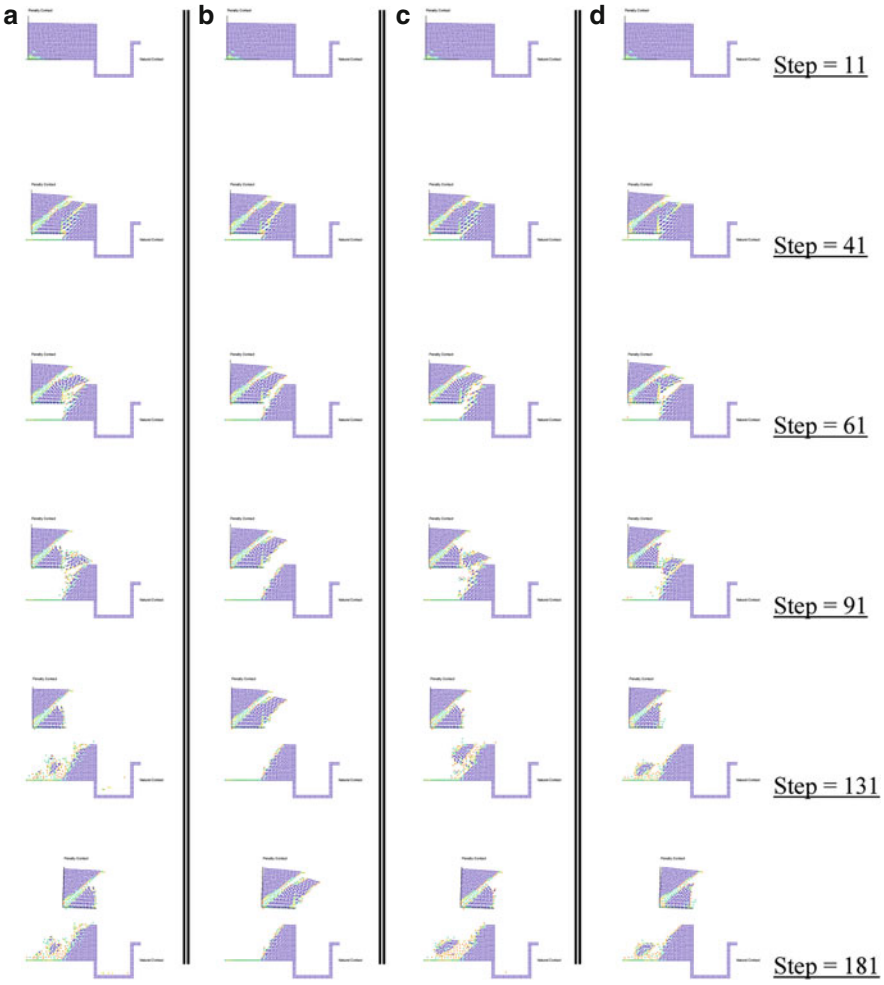


Fig. 3 (continued)

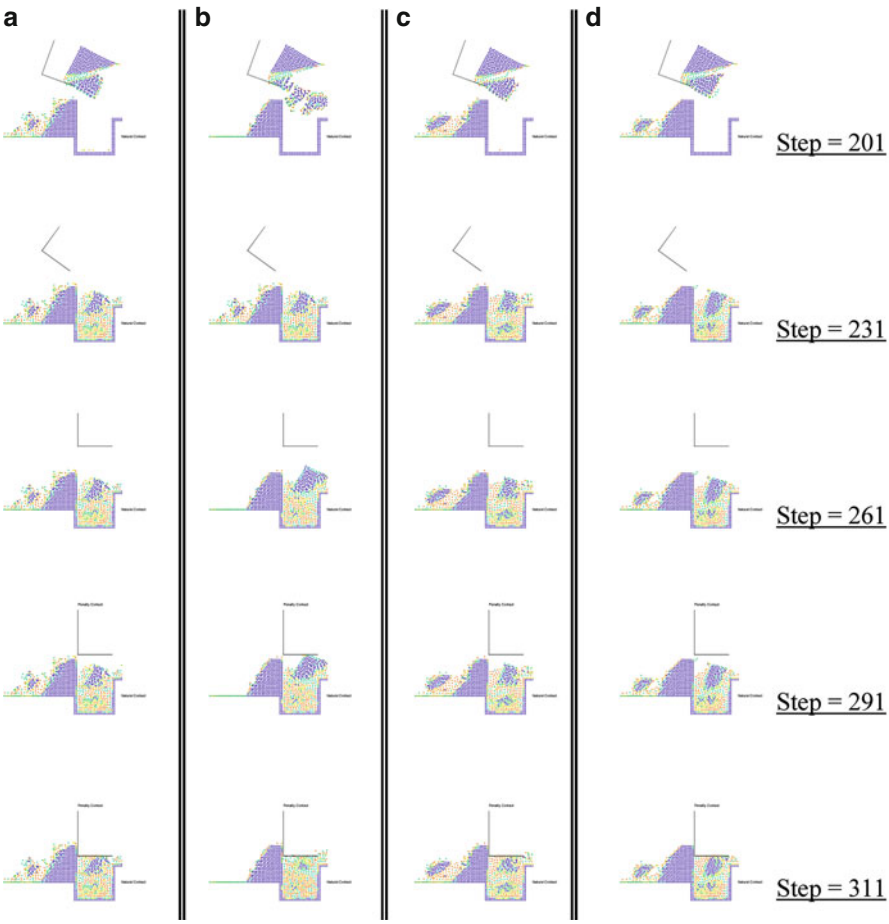


Fig. 3 A parametric study on the initial volumetric and deviatoric healing thresholds (ξ^{0-}_{vol}, ξ^{0-}_{dev}) ($\xi^{0+}_{vol} = \xi^{0+}_{dev} = 0.05$) of deformation of soils for various stages of the earth-moving simulations by the two-parameter damage-healing models

Further, extension of the current work can be made to accommodate the 4th-rank anisotropic damage-healing formulations proposed recently by Ju et al. (2011a, b). In addition, much gratitude is expressed to Prof. J.S. Chen and Dr. Pai-Chen Guan for their assistance in implementing the proposed models into the NMAP meshfree code.

References

E.J. Barbero, F. Greco, P. Lonetti, Continuum damage-healing mechanics with application to self-healing composites. Int. J. Damage Mech. **14**(1), 51–81 (2005)

J.S. Chen, Y. Wu, P. Guan, H. Teng, J. Gaidos, K. Hofstetter, M. Alsaleh, A semi-Lagrangian reproducing Kernel formulation for modeling earth moving operations. Mech. Mater. **41**, 670–683 (2009)

B.D. Coleman, M.E. Gurtin, Thermodynamics with internal state variables. J. Chem. Phys. **47**(2), 597–613 (1967)

O. Herbst, S. Luding, Modeling particulate self-healing materials and application to uni-axial compression. Int. J. Fract. **154**(1–2), 87–103 (2008)

J.W. Ju, On energy-based coupled elastoplastic damage theories – constitutive modeling and computational aspects. Int. J. Solids Struct. **25**(7), 803–833 (1989a)

J.W. Ju, On energy-based coupled elastoplastic damage models at finite strains. J. Eng. Mech. ASCE **115**(11), 2507–2525 (1989b)

J.W. Ju, Isotropic and anisotropic damage variables in continuum damage mechanics. J. Eng. Mech. ASCE **116**(12), 2764–2770 (1990)

J.W. Ju, K.Y. Yuan, A.W. Kuo, Novel strain energy based coupled elastoplastic damage and healing models for geomaterials – part I: formulations. Int. J. Damage Mech. **21**(4), 525–549 (2011a)

J.W. Ju, K.Y. Yuan, A.W. Kuo, J.S. Chen, Novel strain energy based coupled elastoplastic damage and healing models for geomaterials – part II: computational aspects. Int. J. Damage Mech. **21** (4), 551–576 (2011b)

L.M. Kachanov, Rupture time under creep conditions. Izvestia Akademii Nauk SSSR Otdelenie Tekhnicheskich Nauk **8**, 26–31 (1958)

M. Kachanov, Continuum model of medium with cracks. J. Eng. Mech. Div. ASCE **106**(5), 1039–1051 (1980)

J. Lemaitre, A continuous damage mechanics model for ductile fracture. J. Eng. Mater. Technol. ASME **107**(1), 83–89 (1985)

I.N. Rabotnov, On the equations of state for creep, in *Progress in Applied Mechanics-the Prager Anniversary Volume*, (MacMillan, New York, 1963), pp. 307–315

J.C. Simo, J.W. Ju, Strain-based and stress-based continuum damage models.1. Formulation. Int. J. Solids Struct. **23**(7), 821–840 (1987a)

J.C. Simo, J.W. Ju, Strain-based and stress-based continuum damage models.2. Computational aspects. Int. J. Solids Struct. **23**(7), 841–869 (1987b)

J.C. Simo, J.W. Ju, On continuum damage-elastoplasticity at finite strains: a computational framework. Comput. Mech. **5**(5), 375–400 (1989)

J.C. Simo, J.W. Ju, K.S. Pister, R.L. Taylor, An assessment of the cap model: consistent return algorithms and rate-dependent extension. J. Eng. Mech. ASCE **114**(2), 191–218 (1988)

Y. Wu, A stabilized semi-Lagrangian Galerkin meshfree formulation for extremely large deformation analysis. PhD Dissertation, University of California, Los Angeles, 2005

C.T. Wu, J.S. Chen, L. Chi, F. Huck, Lagrangian meshfree formulation for analysis of geotechnical materials. J. Eng. Mech. **127**, 440–449 (2001)

K. Xia, A framework for earthmoving blade/soil model development. J. Terramech. **45**(5), 147–165 (2008)

Particle-Cracking Modeling of Metal Matrix Composites

33

L. Z. Sun, H. T. Liu, and Jiann-Wen Woody Ju

Contents

Abstract

This chapter aims to model the mechanical behavior of particle-reinforced metal matrix composites with particle cracking. Specifically a micromechanics-based elastoplastic constitutive model is coupled with damage mechanics due to particle cracking to predict the overall mechanical behavior of particle-reinforced metal matrix composites. Unidirectionally aligned spheroidal elastic particles, some of which contain penny-shaped cracks, are randomly distributed

L.Z. Sun (✉)
Civil and Environmental Engineering, University of California, Irvine, CA, USA
e-mail: lsun@uci.edu

H.T. Liu
American Bureau of Shipping, Houston, TX, USA
e-mail: dr.haitao.liu@gmail.com

J.-W.W. Ju
Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, USA
e-mail: juj@ucla.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_9

in the elastoplastic metal matrix. These imperfect particles, attributed to progressive particle cracking, are modeled by using the double-inclusion concept. The ensemble-volume averaged homogenization procedure is employed to estimate the effective yield function of the damaged composites. The elastoplastic mechanical behavior of particulate composites under uniaxial loading condition is simulated and compared with available experimental results.

Introduction

Particle-reinforced metal matrix composites (PRMMCs) have received an increasing amount of attention due to their good formability and machinability as well as highly specific mechanical properties. However, the potential of widespread use of PRMMCs is somewhat restricted due to their low ductility and fracture toughness related to microstructural damage evolution. Minimizing these limitations through microstructural design requires a thorough understanding of the micromechanisms of PRMMC's intrinsic damage processes.

Experimental investigations (Clyne and Withers 1993; Suresh et al. 1993) show that normally there are three prevalent damage micromechanisms in PRMMCs, including the brittle cracking of reinforcements, the debonding along the interface between the matrix and reinforcements, and the ductile plastic localization in the matrix. Accordingly, various models exist to estimate the effects of these microstructural damage processes on the overall mechanical behavior of PRMMCs. It should be noted that the dominant damage mechanism is intimately related to the strengths of reinforcement, the composite matrix and its interface, the reinforcement shape and concentration, and the loading mode. This chapter is confined to the reinforcing particle-cracking mechanism. In a companion chapter, Sun et al. (▶ Chap. 34, "Particle-Debonding Modeling of Metal-Matrix Composites") propose the modeling of the evolutionary interfacial particle debonding of metal matrix composites based on the micromechanics concept for effective elastoplastic behavior of partially debonded composites (Liu et al. 2006).

A topic of practical importance regarding the applications of PRMMCs is the prediction of their overall elastoplastic responses due to microstructural particle-cracking damage. For example, a number of finite element micromechanical studies have been conducted to address the effects of particle cracking on the elastic and elastoplastic behavior of PRMMCs (Bao 1992; Finot et al. 1994; Brockenbrough and Zok 1995; Llorca et al. 1997; Wilkinson et al. 1997; Ghosh and Moorthy 1998; Steglich et al. 1999). Within the unit-cell micromechanical framework of two-dimensional axisymmetric finite element analyses, Bao (1992) examined the effect of both intact and cracked reinforcing particles on the flow response of PRMMCs with an elastic-perfectly plastic matrix, while Finot et al. (1994) and Brockenbrough and Zok (1995) further considered the matrix as a plastically hardening solid. Llorca et al. (1997) and Wilkinson et al. (1997) introduced the Weibull statistics to simulate the damage evolution of particle cracking with their finite element models. To consider composites with random reinforcement

distributions, Ghosh and Moorthy (1998) developed a Voronoi cell finite element approach to analyze the particle-cracking effect. Finally, Steglich et al. (1999) considered the particle-cracking by means of a cohesive zone model and crack propagation using the Gurson-Tvergaard-Needleman damage model (Gurson 1977; Tvergaard and Needleman 1984).

Theoretical investigation of effects of damaged particles on the overall mechanical behavior of PRMMCs appears to originate with Mochida et al. (1991), who estimated the elastic stiffness of the composites based on a combination of the Eshelby's equivalent inclusion method (Eshelby 1957) and the Mori–Tanaka's back stress analysis (Mori and Tanaka 1973). Bourgeois (Bourgeois et al. 1994) and Derrien et al. (2000) further extended Mochida et al.'s approach (Mochida et al. 1991) to the elastoplastic study using the concept of secant moduli (Berveiller and Zaoui 1979; Tandon and Weng 1988). More recently, Gonzalez and Llorca (2000) used the incremental self-consistent scheme (Hill 1965) to compute the effective elastoplastic responses of PRMMCs and the stress redistribution in particles induced by reinforcement fracture.

The objective of this chapter is to employ the micromechanics and homogenization (ensemble-volume average) approach to study the effective elastoplastic and damage behavior of metal matrix composites containing unidirectionally aligned particles. Contrary to most unit-cell methods with periodic microstructures, the present model is based on the random distribution of particles embedded in the matrix. Furthermore, the model emphasizes the effect of the fractured reinforcement phase during the progressive evolution of the particle-cracking process. The authors have developed a micromechanical framework to predict the effective elastoplastic behavior of PRMMCs with random microstructures of spheroidal particles under three-dimensional loading histories (Ju and Sun 2001; Sun and Ju 2001). A novel formulation has been derived for the local stress distribution in the matrix with embedded spheroidal particles. The ensemble averaging procedure has been employed to micromechanically derive the effective plastic yielding behavior of PRMMCs based on the probabilistic spatial distribution of spheroidal particles and particle–matrix influences. Extending the foregoing research, a micromechanics-based elastoplastic constitutive model is proposed in this paper to characterize the effect of particle cracking on PRMMCs.

The remaining part of this chapter is organized as follows: First of all, a micromechanics principle is recapitulated to quantify the local stress field in particles and the matrix in section "Micromechanics of Inclusions." The double-inclusion theory (Hori and Nemat-Nasser 1993; Shodja and Sarvestani 2001) is introduced to simulate the cracking process in particles. In section "Homogenization Procedure," the ensemble-volume average procedure is employed to derive the overall yield function of PRMMCs. Particle-cracking damage evolution and the effective elastoplastic constitutive formulation are proposed in sections "Damage Evolution" and "Composite Constitutive Modeling," respectively. Finally, efficient computational algorithms and numerical results are presented in sections "Computational Algorithm" and "Numerical Simulation" to demonstrate the proposed method.

Micromechanics of Inclusions

Consider a composite of an isotropically elastic matrix (phase 0) and unidirectionally aligned yet randomly distributed elastic spheroidal particles (phase 1) with distinct material properties. It is assumed that the two phases are perfectly bonded at interfaces. The isotropic stiffness of matrix and particles can be written as

$$C_{ijkl}^{(\beta)} = \lambda^{(\beta)}\delta_{ij}\delta_{kl} + \mu^{(\beta)}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right), \quad \beta = 0, 1 \tag{1}$$

where $(\lambda^{(0)}, \mu^{(0)})$ and $(\lambda^{(1)}, \mu^{(1)})$ are the Lame constants of the matrix and the particle, respectively.

Upon loading, the local stress $\boldsymbol{\sigma}(\boldsymbol{x})$ at any point $\boldsymbol{x}$ in the matrix is the sum of the far-field stress $\boldsymbol{\sigma}_0$ and the perturbed stress $\boldsymbol{\sigma}'(\boldsymbol{x})$ due to the presence of the inhomogeneous particles. Specifically, the stress perturbation due to one single particle centered at $\boldsymbol{x}'$ is derived as (Ju and Sun 2001)

$$\boldsymbol{\sigma}'(\boldsymbol{x}|\boldsymbol{x}') = \left[\boldsymbol{C}^{(0)} \cdot \overline{\boldsymbol{G}}^{(1)}(\boldsymbol{x} - \boldsymbol{x}')\right] : \boldsymbol{\varepsilon}_*^{(1)} \tag{2}$$

where $\boldsymbol{\varepsilon}_*^{(1)}$ is the equivalent eigenstrain which reads

$$\boldsymbol{\varepsilon}_*^{(1)} = -\left[\boldsymbol{S}^{(1)} + \left(\boldsymbol{C}^{(1)} - \boldsymbol{C}^{(0)}\right)^{-1} \cdot \boldsymbol{C}^{(0)}\right]^{-1} : \boldsymbol{\varepsilon}_0 \tag{3}$$

with $\boldsymbol{\varepsilon}_0 = \boldsymbol{C}^{(0)-1} : \boldsymbol{\sigma}_0$. Furthermore, $\overline{\boldsymbol{G}}^{(1)}(\boldsymbol{x} - \boldsymbol{x}')$ and $\boldsymbol{S}^{(1)}$ are the fourth-rank exterior- and interior-point Eshelby's tensors for spheroidal inclusions, explicit expressions of which can be found in Ju and Sun (1999). It is also noted that, throughout the paper, the double-dot symbol ":"" denotes the tensor contraction between a fourth-rank tensor and a second-rank tensor, while the symbol "·" represents the tensor multiplication between two fourth-rank tensors.

During the external loading process, particle cracking may take place inside the phase of reinforcement once the local stress intensity reaches the critical value (Lee et al. 1999). The micromechanisms for crack initiation, interaction, and growth are complicated. Since the major objective is to model the overall mechanical properties of PRMMCs through homogenization procedures rather than the locally accurate stress/strain fields, a penny-shaped crack (phase 2) is embedded in the perfect particle to simulate the actual particle-cracking damage, as shown in Fig. 1. From the double-inclusion theory (Hori and Nemat-Nasser 1993; Shodja and Sarvestani 2001), the local perturbed stress $\boldsymbol{\sigma}'(\boldsymbol{x})$ due to a damaged particle centered at $\boldsymbol{x}'$ can be estimated as

$$\boldsymbol{\sigma}'(\boldsymbol{x}|\boldsymbol{x}') = \left[\boldsymbol{C}^{(0)} \cdot \overline{\boldsymbol{G}}^{(1)}(\boldsymbol{x} - \boldsymbol{x}')\right] : \boldsymbol{\varepsilon}_*^{(1)} + \left[\boldsymbol{C}^{(0)} \cdot \overline{\boldsymbol{G}}^{(2)}(\boldsymbol{x} - \boldsymbol{x}')\right] : \left(\boldsymbol{\varepsilon}_*^{(2)} - \boldsymbol{\varepsilon}_*^{(1)}\right) \tag{4}$$

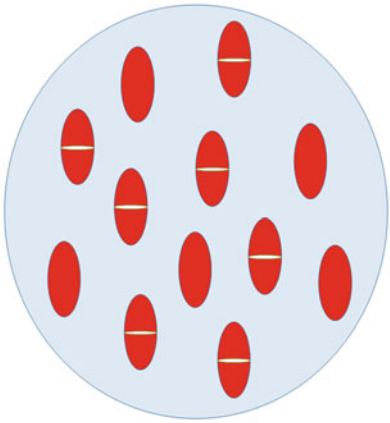


Fig. 1 Schematic diagram of particle-cracking microstructure of PRMMCs

where $\overline{\boldsymbol{G}}^{(2)}(\boldsymbol{x}-\boldsymbol{x}')$ is the exterior-point Eshelby's tensor for penny-shaped cracks. It can be calculated as the special case of $\overline{\boldsymbol{G}}^{(1)}(\boldsymbol{x}-\boldsymbol{x}')$ by setting the aspect ratio of spheroidal particles equal to zero. Moreover, it is noted that the equivalent eigenstrains $\boldsymbol{\varepsilon}_*^{(1)}$ and $\boldsymbol{\varepsilon}_*^{(2)}$ of spheroidal inclusions and penny-shaped cracks are *not* uniform inside their respective domains (Shodja and Sarvestani 2001). For ease of calculation, Hori and Nemat-Nasser's (1993) volume-averaged expressions are used in this paper:

$$\begin{aligned}
\boldsymbol{\varepsilon}_*^{(1)} &= -\left[\boldsymbol{S}^{(1)} + \left(\boldsymbol{C}^{(1)} - \boldsymbol{C}^{(0)}\right)^{-1} \cdot \boldsymbol{C}^{(0)}\right]^{-1} : \boldsymbol{\varepsilon}_0 \\
\boldsymbol{\varepsilon}_*^{(2)} &= -\left\{\boldsymbol{S}^{(2)} - \boldsymbol{I} + \left(\boldsymbol{S}^{(1)} - \boldsymbol{S}^{(2)}\right) \cdot \left[\boldsymbol{S}^{(2)} + \left(\boldsymbol{C}^{(1)} - \boldsymbol{C}^{(0)}\right)^{-1} \cdot \boldsymbol{C}^{(0)}\right]^{-1} \cdot \left(\boldsymbol{S}^{(2)} - \boldsymbol{I}\right)\right\}^{-1} : \boldsymbol{\varepsilon}_0 \\
&\quad + \left[\boldsymbol{S}^{(1)} + \left(\boldsymbol{C}^{(1)} - \boldsymbol{C}^{(0)}\right)^{-1} \cdot \boldsymbol{C}^{(0)}\right]^{-1} : \boldsymbol{\varepsilon}_0
\end{aligned} \tag{5}$$

where $\boldsymbol{S}^{(2)}$ is the interior-point Eshelby's tensor for penny-shaped cracks and $\boldsymbol{I}$ is the fourth-rank identity tensor.

Homogenization Procedure

To obtain the effective (overall) constitutive laws of PRMMCs, a homogenization procedure is usually performed within a mesoscopic representative volume element (RVE) (Nemat-Nasser and Hori 1999). At any matrix material local point $\boldsymbol{x}$, the microscopic stress $\boldsymbol{\sigma}(\boldsymbol{x})$ is assumed to satisfy the von Mises J_2-yield criterion:

$$F\left(\boldsymbol{\sigma}, \overline{e}_m^p\right) = \sqrt{\boldsymbol{\sigma} : \boldsymbol{I}_d : \boldsymbol{\sigma}} - K\left(\overline{e}_m^p\right) \leq 0 \tag{6}$$

where $\overline{e}_m^p$ and $K\left(\overline{e}_m^p\right)$ are the equivalent plastic strain and the isotropic hardening function of the matrix-only material, respectively. Moreover $\boldsymbol{I}_d$, denotes the deviatoric part of the fourth-rank identity tensor $\boldsymbol{I}$.

The square of the "current stress norm," denoted by $H(\boldsymbol{x}|g) = \boldsymbol{\sigma}(\boldsymbol{x}|g) : \boldsymbol{I}_d : \boldsymbol{\sigma}(\boldsymbol{x}|g)$, at a local point $\boldsymbol{x}$, contributes to the initial yielding criterion of composites for a given particle configuration g (assembly). Furthermore, $\langle H\rangle_m(\boldsymbol{x})$ is defined as the ensemble average of $H(\boldsymbol{x}|g)$ over all possible realizations for a matrix point $\boldsymbol{x}$:

$$\langle H\rangle_m(\boldsymbol{x}) = H^0 + \oint_g \left[H(\boldsymbol{x}|g) - H^0\right] P(g)\, dg \tag{7}$$

Here, $P(g)$ is the probability density function for finding a particle configuration g in the composite, and $H^0 = \boldsymbol{\sigma}_0 : \boldsymbol{I}_d : \boldsymbol{\sigma}_0$ denotes the square of the far-field stress norm applied to the composite.

The expression of $\langle H\rangle_m$ can be approximated by neglecting the *interaction* among neighboring particles. That is, a matrix point $\boldsymbol{x}$ simply *collects* the perturbation from all randomly located, non-interacting particles:

$$\begin{aligned}\langle H\rangle_m(\boldsymbol{x}) \cong H^0 &+ \oint_{\boldsymbol{x}'\notin\Xi(\boldsymbol{x})} \left[H^{(1)}(\boldsymbol{x}|\boldsymbol{x}') - H^0\right] P^{(1)}(\boldsymbol{x}')\, d\boldsymbol{x}' \\ &+ \oint_{\boldsymbol{x}'\notin\Xi(\boldsymbol{x})} \left[H^{(2)}(\boldsymbol{x}|\boldsymbol{x}') - H^0\right] P^{(2)}(\boldsymbol{x}')\, d\boldsymbol{x}'\end{aligned} \tag{8}$$

where $\Xi(\boldsymbol{x})$ is the exclusion zone of $\boldsymbol{x}$ for the center location $\boldsymbol{x}'$ of a particle in the probability space, which is identical to the shape and size of a spheroidal particle. In addition, $H^{(1)}$ and $H^{(2)}$ are the stress-norm collection contributions from perfect particles and cracked particles, respectively. They can be computed from Eqs. 2 and 4 in section "Micromechanics of Inclusions," respectively.

For simplicity, it is assumed that all particles are uniformly distributed in the composites. Thus, $P^{(\beta)}(\boldsymbol{x}')$ can be assumed to be $N^{(\beta)}/V$ ($\beta = 1, 2$), where $N^{(\beta)}$ is the total number of β-phase particles uniformly dispersed in volume V of RVE. After a series of lengthy but straightforward derivations, the ensemble-averaged $\langle H\rangle_m$ can be evaluated as

$$\langle H\rangle_m = \boldsymbol{\sigma}_0 : \boldsymbol{T} : \boldsymbol{\sigma}_0 \tag{9}$$

where the components of the fourth-rank tensor $\boldsymbol{T}$ take the form

$$T_{ijkl} = T_{IK}^{(1)}\delta_{ij}\delta_{kl} + T_{IJ}^{(2)}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) \tag{10}$$

with

$$T_{IK}^{(1)}=-\frac{1}{3}+\frac{2}{4725(1-\nu_0)^2}\left\{\begin{array}{l}3\left(35\nu_0^2-70\nu_0+36\right)\times\left(4A_{II}^{(2)}A_{KK}^{(2)}\Delta_{IK}^{p}\dfrac{\phi_2}{\alpha}+\Delta_{IK}\dfrac{\phi}{B_{II}B_{KK}}\right)\\+7\left(50\nu_0^2-59\nu_0+8\right)\times\left[4A_{II}^{(2)}A_{KK}^{(2)}(\Delta_I^p+\Delta_K^p)\dfrac{\phi_2}{\alpha}+(\Delta_I+\Delta_K)\dfrac{\phi}{B_{II}B_{KK}}\right]\\-2\left(175\nu_0^2-343\nu_0+103\right)\times\left(4A_{II}^{(2)}A_{KK}^{(2)}\Delta_{IK}^{p}\dfrac{\phi_2}{\alpha}+\Delta_{IK}\dfrac{\phi}{B_{II}B_{KK}}\right)\end{array}\right\}$$
$$+\frac{2(25\nu_0-2)(1-2\nu_0)}{225(1-\nu_0)^2}\times\left\{\frac{\phi(\Gamma_{II}+\Gamma_{KK})}{B_{II}B_{KK}}-2\frac{\phi_2}{\alpha}\left(A_{II}^{(1)}A_{KK}^{(2)}+A_{II}^{(2)}A_{kk}^{(1)}\right)\right\}$$
$$+\frac{2(25\nu_0-23)(1-2\nu_0)}{225(1-\nu_0)^2}\times\left\{\frac{\phi(\Gamma_{II}\Delta_K+\Gamma_{KK}\Delta_I)}{B_{II}B_{KK}}-2\frac{\phi_2}{\alpha}\left(A_{II}^{(1)}A_{KK}^{(2)}\Delta_K^p+A_{II}^{(2)}A_{kk}^{(1)}\Delta_I^p\right)\right\}$$
$$+\frac{2(1-2\nu_0)^2}{3(1-\nu_0)^2}\times\left\{A_{II}^{(1)}A_{KK}^{(1)}\frac{\phi_2}{\alpha}+\frac{\phi\Gamma_{II}\Gamma_{KK}}{B_{II}B_{KK}}\right\}$$

$$T_{IJ}^{(2)}=\frac{1}{2}+\frac{1}{1575(1-\nu_0)^2}\left\{\begin{array}{l}\left(70\nu_0^2-140\nu_0+72\right)\times\left(4A_{IJ}^{(2)}A_{IJ}^{(2)}\Delta_{IJ}^{p}\dfrac{\phi_2}{\alpha}+\Delta_{IJ}\dfrac{\phi}{B_{IJ}B_{IJ}}\right)\\-7\left(175\nu_0^2-266\nu_0+75\right)\times\left[2A_{IJ}^{(2)}A_{IJ}^{(2)}\left(\Delta_I^p+\Delta_J^p\right)\dfrac{\phi_2}{\alpha}+\dfrac{(\Delta_I+\Delta_J)}{2}\dfrac{\phi}{B_{IJ}B_{IJ}}\right]\\+\left(350\nu_0^2-476\nu_0+164\right)\times\left(4A_{IJ}^{(2)}A_{IJ}^{(2)}\dfrac{\phi_2}{\alpha}+\dfrac{\phi}{B_{IJ}B_{IJ}}\right)\end{array}\right\} \tag{11}$$

Here, ν_0 is Poisson's ratio of the matrix material, α is the aspect ratio of spheroidal particles, and ϕ, ϕ_1, and ϕ_2 are the volume fractions of the total particles, perfect particles, and the cracked particles, respectively. It is noted that Mura's (1987) tensorial indicial notation is followed; i.e., repeated lower-case indices are summed up from 1 to 3, whereas upper-case indices take on the same numbers as the corresponding lower-case ones but are not summed up. Other parameters in the above equation are given in (Ju and Tseng 1997).

The general relationship between the applied far-field stress $\boldsymbol{\sigma}_0$ and macroscopic (ensemble-volume averaged) stress $\overline{\boldsymbol{\sigma}}$ is given by Ju and Sun (2001)

$$\boldsymbol{\sigma}_0=\boldsymbol{P}:\overline{\boldsymbol{\sigma}} \tag{12}$$

where the fourth-rank tensor $\boldsymbol{P}$ reads

$$\boldsymbol{P}=\left\{\boldsymbol{C}^{(0)}\cdot\left[\boldsymbol{I}+\phi\left(\boldsymbol{I}-\boldsymbol{S}^{(1)}\right)\cdot\left(\boldsymbol{S}^{(1)}+\left(\boldsymbol{C}^{(1)}-\boldsymbol{C}^{(0)}\right)^{-1}\cdot\boldsymbol{C}^{(0)}\right)^{-1}\right]\cdot\boldsymbol{C}^{(0)^{-1}}\right\}^{-1} \tag{13}$$

The combination of Eqs. 9 and 12 then leads to an alternative expression of the ensemble-averaged square of the current stress norm as

$$\langle H\rangle_m=\overline{\boldsymbol{\sigma}}:\overline{\boldsymbol{T}}:\overline{\boldsymbol{\sigma}} \tag{14}$$

where the fourth-rank tensor $\overline{\boldsymbol{T}}=\boldsymbol{P}^T\cdot\boldsymbol{T}\cdot\boldsymbol{P}$. It is observed in the above equation that $\langle H\rangle_m$ can be reduced to that in Ju and Sun (2001) if no particle cracking occurs in the composites. Furthermore, Eq. 14 will reduce to the classical J_2-invariant for $\phi=0$ (matrix material only).

Damage Evolution

Although the micromechanisms for particle interaction with neighboring materials are complicated in composites, it is clear that the stress field inside the particle domain is crucial for particle cracking. During the modeling of the elastoplastic behavior of PRMMCs, the averaged normal stress on particles is the controlling factor for particle cracking. This implies that particle cracking will occur in a certain plane if the particle normal stress perpendicular to that plane reaches a critical value (Lee et al. 1999). On the other hand, all particles are assumed to be uniformly dispersed with identical geometry but not to initiate particle cracking simultaneously. The cracking process is best described by Weibull's (Sun et al. 2003) probability approach (Llorca et al. 1997; Derrien et al. 2000; Weibull 1951; Li et al. 1999).

Assuming that Weibull's statistics governs the initiation of particle cracking in the symmetric equatorial plane (2–3 plane), the total volume fraction of cracked particles ϕ_2 can be expressed as

$$\frac{\phi_2}{\phi} = 1 - \exp\left[-\left(\frac{\overline{\sigma}_{11}^{(1)}}{s}\right)\right]^m \tag{15}$$

where the Weibull parameters s and m are related to the critical fracture strength of particles and cracking evolution rate, respectively. For example, when an intermediate rate of $m = 5$ is selected, s will be determined as $1.09\,\sigma_{\mathrm{cri}}$. In addition, $\overline{\sigma}_{11}^{(1)}$ is the averaged normal stress of particle in the symmetric direction. From the micromechanics theory of section "Micromechanics of Inclusions," the following equation holds:

$$\overline{\boldsymbol{\sigma}}^{(1)} = \boldsymbol{C}^{(0)} \cdot \left\{ \boldsymbol{I} + \left(\boldsymbol{I} - \boldsymbol{S}^{(1)}\right) \cdot \left[\boldsymbol{S}^{(1)} + \left(\boldsymbol{C}^{(1)} - \boldsymbol{C}^{(0)}\right)^{-1} \cdot \boldsymbol{C}^{(0)}\right]^{-1} \right\} : \boldsymbol{\varepsilon}_0 \tag{16}$$

Composite Constitutive Modeling

When a small deformation is considered, the total macroscopic strain $\overline{\boldsymbol{\varepsilon}}$ consists of two parts:

$$\overline{\boldsymbol{\varepsilon}} = \overline{\boldsymbol{\varepsilon}}^{\mathrm{e}} + \overline{\boldsymbol{\varepsilon}}^{p} \tag{17}$$

where $\overline{\boldsymbol{\varepsilon}}^{\mathrm{e}}$ denotes the overall elastic strain and $\overline{\boldsymbol{\varepsilon}}^{p}$ represents the overall plastic strain of composites. The relationship between the macroscopic stress $\overline{\boldsymbol{\sigma}}$ and macroscopic elastic strain $\overline{\boldsymbol{\varepsilon}}^{\mathrm{e}}$ reads

$$\overline{\boldsymbol{\sigma}} = \overline{\boldsymbol{C}} : \overline{\boldsymbol{\varepsilon}}^{\mathrm{e}} \tag{18}$$

where the effective elastic stiffness of composites can be determined as (Wilkinson et al. 2001)

$$\begin{aligned}\overline{\boldsymbol{C}} = \boldsymbol{C}^{(0)} + \phi\ \boldsymbol{C}^{(0)} \cdot \left[\boldsymbol{K} - \phi\ \boldsymbol{S}^{(1)} + \tfrac{\phi_2}{\alpha}\boldsymbol{S}^{(2)} \cdot \boldsymbol{J}^{-1} \cdot \boldsymbol{K}\right]^{-1} \\ + \frac{\phi_2}{\alpha}\boldsymbol{C}^{(0)} \cdot \left[\frac{\phi_2}{\alpha}\boldsymbol{S}^{(2)} + \left(\boldsymbol{I} - \phi\ \boldsymbol{S}^{(1)} \cdot \boldsymbol{K}^{-1}\right) \cdot \boldsymbol{J}\right]^{-1}\end{aligned} \tag{19}$$

in which

$$\begin{aligned}\boldsymbol{K} &= \boldsymbol{S}^{(1)} + \left(\boldsymbol{C}^{(1)} - \boldsymbol{C}^{(0)}\right)^{-1} \cdot \boldsymbol{C}^{(0)} \\ \boldsymbol{J} &= \left\{\boldsymbol{I} + \left(\boldsymbol{S}^{(1)} - \boldsymbol{S}^{(2)}\right) \cdot \left[\boldsymbol{S}^{(2)} + \left(\boldsymbol{C}^{(1)} - \boldsymbol{C}^{(0)}\right)^{-1} \cdot \boldsymbol{C}^{(0)}\right]^{-1}\right\} \cdot \boldsymbol{E}\end{aligned} \tag{20}$$

and $\boldsymbol{E} = \lim_{\rho \to 0} \frac{\boldsymbol{S}^{(2)} - \boldsymbol{I}}{\rho}$ with ρ as the limiting aspect ratio of penny-shaped cracks.

The overall plastic flow of composites is postulated to be *associative*. The macroscopic plastic strain rate for PRMMCs takes the form

$$\dot{\overline{\boldsymbol{\varepsilon}}}^{p} = \dot{\lambda}\frac{\partial \overline{F}}{\partial \overline{\boldsymbol{\sigma}}} \tag{21}$$

where $\dot{\lambda}$ is the plastic consistency parameter. Moreover, $\overline{F}$ is the overall yield function of composites which, based on ensemble homogenization derivations in section "Homogenization Procedure," can be micromechanically determined as

$$\overline{F} = (1 - \phi_1)\sqrt{\overline{\boldsymbol{\sigma}} : \overline{\boldsymbol{T}} : \overline{\boldsymbol{\sigma}}} - K(\overline{e}^{p}) \leq 0 \tag{22}$$

where $\overline{e}^{p}$ represents the effective equivalent plastic strain, while the simple isotropic hardening function $K(\overline{e}^{p})$ is proposed as

$$K(\overline{e}^{p}) = \sqrt{\frac{2}{3}}\left[\sigma_y + h(\overline{e}^{p})^{q}\right] \tag{23}$$

Here, σ_y denotes the initial yield stress of matrix material, and h and q signify the linear and exponential isotropic hardening parameters.

Therefore, an effective elastoplastic constitutive model has been developed for particle reinforced metal matrix composites with a progressive particle-cracking process. The model is formulated in Eqs. 15, 16, 17, 18, 19, 20, 21, and–22 based on a micromechanics approach, ensemble homogenization procedures, and the statistical distribution method. The proposed model allows one to estimate the overall elastoplastic stress–strain responses of the composites.

Computational Algorithm

Consistent with the computational plasticity algorithms, this chapter employs the strain-driven algorithm in which the overall stress history is to be uniquely determined by the given overall strain history. Similar to the previous procedures (Ju and Chen 1994), the unknown state of a point $\left(\overline{\boldsymbol{\sigma}}_{n+1}, \overline{\boldsymbol{\varepsilon}}^{p}_{n+1}, \overline{e}^{p}_{n+1}\right)$ is to be determined at the end of the time step $t = t_{n+1}$, given the known state from the previous time step $\left(\overline{\boldsymbol{\varepsilon}}_{n}, \overline{\boldsymbol{\varepsilon}}^{p}_{n}, \overline{e}^{p}_{n}\right)$ at $t = t_n$. The total stress $\overline{\boldsymbol{\sigma}}_n$ at t_n can be evaluated as

$$\overline{\boldsymbol{\sigma}}_{n} = \overline{\boldsymbol{C}} : \left(\overline{\boldsymbol{\varepsilon}}_{n} - \overline{\boldsymbol{\varepsilon}}^{p}_{n}\right) \tag{24}$$

The two-step operator splitting methodology is adopted to split the elastoplastic loading process into the elastic predictor and plastic corrector. First, a trial elastic stress can be computed as

$$\overline{\boldsymbol{\sigma}}^{tr}_{n+1} = \overline{\boldsymbol{\sigma}}_{n} + \overline{\boldsymbol{C}} : \Delta\overline{\boldsymbol{\varepsilon}}_{n+1} \tag{25}$$

where $\Delta\overline{\boldsymbol{\varepsilon}}_{n+1}$ is the given strain increment at the (n + 1)th time step. The trial $\overline{\boldsymbol{\sigma}}^{tr}_{n+1}$ and the previous $\overline{e}^{p}_{n}$ are then plugged into the yield function as

$$\overline{F}^{tr}_{n+1} = (1-\phi_1)\sqrt{\overline{\boldsymbol{\sigma}}^{tr}_{n+1} : \overline{\boldsymbol{T}} : \overline{\boldsymbol{\sigma}}^{tr}_{n+1}} - K\left(\overline{e}^{p}_{n}\right) \tag{26}$$

If $\overline{F}^{tr}_{n+1} \leq 0$, then the incremental response is entirely elastic. It is therefore set that $\overline{\boldsymbol{\sigma}}_{n+1} = \overline{\boldsymbol{\sigma}}^{tr}_{n+1}$ and $\overline{e}^{p}_{n+1} = \overline{e}^{p}_{n}$.

In the event of $\overline{F}^{tr}_{n+1} > 0$, the trial elastic stress is located outside the yield surface, which is not allowed. The Kuhn–Tucker loading/unloading conditions must be enforced to bring the elastic predictor back to the yield surface. By employing the implicit backward Euler method which is unconditionally stable, it reads

$$\overline{\boldsymbol{\varepsilon}}^{p}_{n+1} = \overline{\boldsymbol{\varepsilon}}^{p}_{n} + \xi_{n+1}\overline{\boldsymbol{T}} : \overline{\boldsymbol{\sigma}}_{n+1}\overline{e}^{p}_{n+1} = \overline{e}^{p}_{n} + \xi_{n+1}\sqrt{\frac{2}{3}\overline{\boldsymbol{\sigma}}_{n+1} : \overline{\boldsymbol{T}} : \overline{\boldsymbol{\sigma}}_{n+1}} \tag{27}$$

with the following definition:

$$\xi_{n+1} = 2(1-\phi_1)^2\dot{\lambda}\,\Delta t_{n+1} \tag{28}$$

Consequently, the overall stress $\overline{\boldsymbol{\sigma}}_{n+1}$ can be updated as

$$\overline{\boldsymbol{\sigma}}_{n+1} = \overline{\boldsymbol{C}} : \left(\overline{\boldsymbol{\varepsilon}}_{n+1} - \overline{\boldsymbol{\varepsilon}}^{p}_{n} - \xi_{n+1}\overline{\boldsymbol{T}} : \overline{\boldsymbol{\sigma}}_{n+1}\right) \tag{29}$$

It is known that the effective yield function should always be zero during the plastic loading process. Therefore, the parameter ξ_{n+1} can be determined by enforcing the plastic consistency condition at $t = t_{n+1}$:

$$\overline{F}_{n+1}(\xi_{n+1}) = (1-\phi_1)\sqrt{\overline{\boldsymbol{\sigma}}_{n+1} : \overline{\boldsymbol{T}} : \overline{\boldsymbol{\sigma}}_{n+1}} - K\left(\overline{e}^{p}_{n+1}\right) = 0 \tag{30}$$

Substitution of Eqs. 27 and 29 into Eq. 30 renders a nonlinear scalar equation in terms of ξ_{n+1}. Once the parameter ξ_{n+1} is numerically solved from Eq. 30, the total current stress, the plastic strain, and the hardening parameter can be updated in terms of Eqs. 27 and 29.

For convenience, Table 1 summarizes the above micromechanical iterative computational algorithm for the overall elastoplastic responses of PRMMCs with damage.

Numerical Simulation

Uniaxial stress–strain curves are often referred to as important indicators of the mechanical behavior of materials. In order to illustrate the proposed micromechanics-based model, the uniaxial stress tension is considered as an example. In this case, the macroscopic stress $\overline{\boldsymbol{\sigma}}$ can be expressed as $\overline{\sigma}_{11} > 0$, and all other stress components can be expressed as $\overline{\sigma}_{ij} = 0$. The aligned, axisymmetric axis of spheroidal particles is denoted as the x_1-axis. Furthermore, unless noted otherwise during subsequent numerical simulations, the matrix material is taken as an aluminum alloy with Young's modulus $E_m = 70$ GPa, Poisson's ratio $\nu_m = 0.3$, the uniaxial yield strength $\sigma_y = 100$ MPa, and the strain-hardening parameters $h = 500$ MPa and $q = 0.35$. For the elastic reinforcement material, the Young's modulus $E_p = 450$ GPa and the Poisson's ratio $\nu_p = 0.2$ (similar to the elastic properties of the SiC particles). The critical strength for particle cracking is $\sigma_{\text{cri}} = 3\sigma_y$ and the particle-cracking evolution parameter is $m = 5$.

Simulation results on uniaxial stress–strain–damage behavior of PRMMCs are illustrated in Fig. 2. Specifically, Fig. 2a shows the effective stress–strain curves for a 15 % total volume fraction of prolate spheroidal particles with the aspect ratio $\alpha = 3$. It is demonstrated that the overall stress–strain responses with progressive particle cracking lie between those of the composite with all particles fractured at the outset (the lower bound) and the composite without particle cracking at all (the upper bound). Less strain hardening can be observed for composites with the lower critical fracture strength. Fig. 2b shows the volume fraction evolutions of cracked particles during the external loading process. The damage parameter indicated by ϕ_2 initiates somewhere before the composite reaches its yield strength. The parameter ϕ_2 keeps increasing until all particles crack. Weaker particle fracture strength leads to faster damage evolution process.

The effect of the *remaining particle* phase (double inclusion), in which penny-shaped cracks are embedded, is shown in Fig. 3. The model prediction demonstrates a significant effect when compared with the mechanical response of materials featuring penny-shaped cracks *directly embedded* in the matrix, even though the damage evolutions are set the same. Both the elastic and plastic properties will be underestimated if the remaining particle part is not taken into account in the calculations.

Table 1 Computational algorithm for overall elastoplastic-damage responses of PRMMCs

Given: $\{\overline{\sigma}_n, \overline{\varepsilon}_n, \overline{\varepsilon}_n^p, \overline{e}_n^p\}$at time step n and strain increment $\{\Delta\overline{\varepsilon}_{n+1}\}$	
Solve: $\{\overline{\sigma}_{n+1}, \overline{\varepsilon}_{n+1}, \overline{\varepsilon}_{n+1}^p, \overline{e}_{n+1}^p\}$ the value of time step (n + 1)	
(i)	Initialize: $\{\overline{\sigma}_0 = 0, \overline{\varepsilon}_0 = 0, \overline{\varepsilon}_0^p = 0, \overline{e}_0^p = 0\}, \{\phi_1|_0 = \phi, \phi_2|_0 = 0\}$
(ii)	Compute: $\overline{C}_{n+1}, \overline{T}_{n+1}$
(iii)	Compute: $\overline{\sigma}_{n+1}^{tr} = \overline{\sigma}_n + \overline{C} : \Delta\overline{\varepsilon}_{n+1}, \overline{\varepsilon}_{n+1} = \overline{\varepsilon}_n + \Delta\overline{\varepsilon}_{n+1}$
(vi)	Check: $\overline{F}_{n+1}^{tr} = (1-\phi_1)\sqrt{\sigma_{n+1}^{tr} : \overline{T} : \sigma_{n+1}^{tr}} - K(\overline{e}_n^p) \le TOL(?)$
	If $F_{n+1}^{tr} \le TOL$ (elastic), $\xi_{n=1} = 0$; Set $\overline{\sigma}_{n+1} = \overline{\sigma}_{n+1}^{tr}; \overline{e}_{n+1}^p = \overline{e}_n^p$; Go to (vi)
	Else (plastic), do the plastic correction (v)
(v)	Perform plastic correction: (return mapping algorithm)
	$\overline{\varepsilon}_{n+1}^p = \overline{\varepsilon}_n^p + \xi_{n+1}\overline{T} : \overline{\sigma}_{n+1}; \overline{e}_{n+1}^p = \overline{e}_n^p + \xi_{n+1}\sqrt{\frac{2}{3}\overline{\sigma}_{n+1} : \overline{T} : \overline{\sigma}_{n+1}}$ $\overline{\sigma}_{n+1} = \overline{C} : \left(\overline{\varepsilon}_{n+1} - \overline{\varepsilon}_n^p - \xi_{n+1}\overline{T} : \overline{\sigma}_{n+1}\right); \xi_{n+1} = 2(1-\phi_1|_n)^2\dot{\lambda}\Delta t_{n+1}$
	Use local Newton iteration to solve the nonlinear equation: $\overline{F}_{n+1}(\xi_{n+1}) = (1-\phi_1|_n)\sqrt{\overline{\sigma}_{n+1} : \overline{T} : \overline{\sigma}_{n+1}} - K(\overline{e}_{n+1}^p) = 0$
	Calculate plastic correction: $\{\overline{\sigma}_{n+1}, \overline{\varepsilon}_{n+1}, \overline{\varepsilon}_{n+1}^p \overline{e}_{n+1}^p\}$
(vi)	Computer the evolution of volume fraction of particles:
	(a) Compute the normal stresses inside particles: $\overline{\sigma}_{n+1}^{(1)} = C^{(0)}.\left\{I + \left(I - S^{(1)}\right).\left[S^{(1)} + \left(C^{(1)} + C^{(0)}\right)^{-1}.C^{(0)}\right]^{-1}\right\} : \varepsilon_0|_{n+1}$
	(b) Calculate the volume fraction of cracked particles: $\phi_2|_{n+1} = \phi\left\{1 - \exp\left[\frac{\sigma_{n+1}^{(1)}}{s}\right]^{n+1}\right\}$
	(c) Calculate the volume fraction of undamaged particles: $\phi_1|_{n+1} = \phi - \phi_2|_{n+1}$
	(d) GO TO (ii) (the next time step computation)

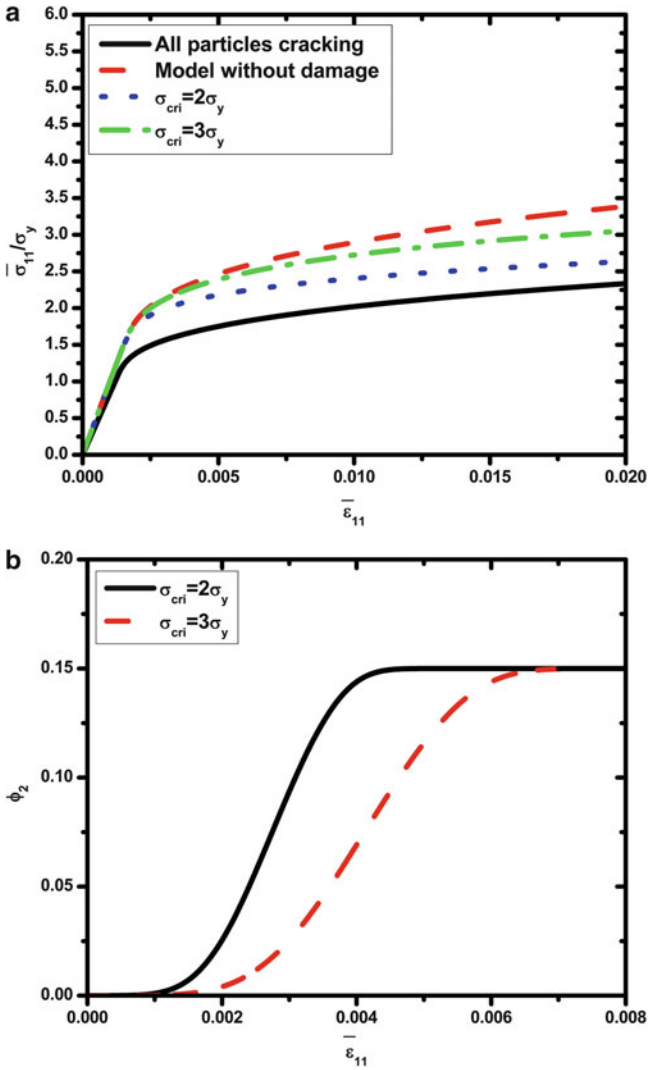


Fig. 2 Simulations of overall uniaxial stress–strain curves (**a**) and microstructural damage evolutions (**b**) of PRMMCs

Furthermore, Fig. 4 shows the comparison between the simulation and the experimental results of Wilkinson et al. (1997). In their experiment, uniaxial stress–strain behavior was recorded for SiC particulate-reinforced A356 aluminum alloy composites. The following material properties are chosen: $E_m = 70$ GPa, $\nu_m = 0.3$, $\sigma_y = 65$ MPa, $h = 460$ MPa, $q = 0.36$, $E_p = 450$ GPa, $\nu_p = 0.2$, $\phi = 10$ % and 20 %, and $\alpha = 3$. In addition, the particle fracture strength σ_{cri} is intentionally chosen as $4\sigma_y$, and the damage evolution parameter is taken as 5.

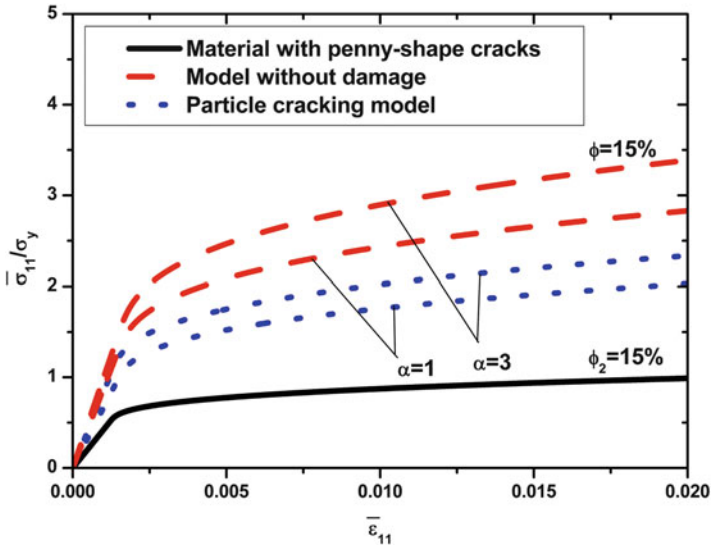


Fig. 3 Effect of the remaining particle phase on overall elastoplastic behavior of PRMMCs

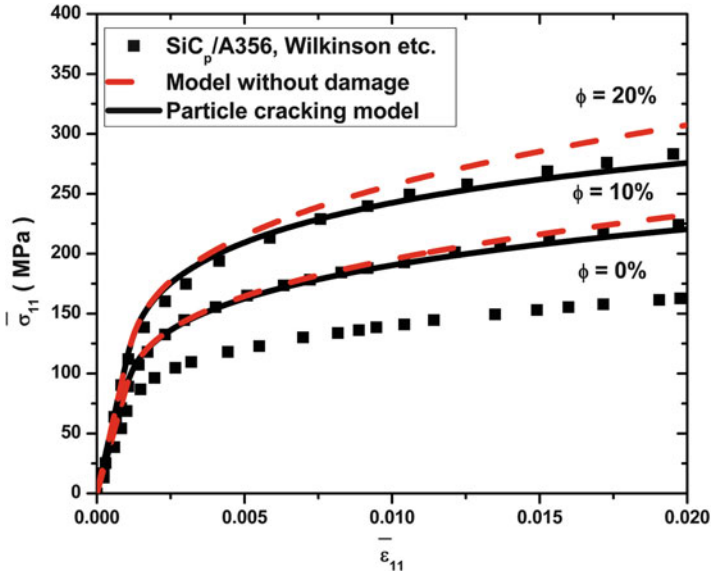


Fig. 4 Comparisons of uniaxial stress–strain responses with experimental data (Wilkinson et al. 1997) for SiCp/Al composites

It is shown in Fig. 4 that the present predictions with particle cracking compare well with the experimental data in the entire range of elastic and plastic stages. In particular, the particle damage effect is clearly illustrated within the plastic behavior.

Conclusions

Emanating from the eigenstrain concept of micromechanics and macroscopic homogenization, ensemble-averaged elastoplastic constitutive equations are developed for particle-reinforced metal matrix composites with particle-cracking damage evolution. Those cracked particles are modeled using the *double-inclusion* concept. The damage evolution of particles is considered in accordance with Weibull's statistical function to characterize the varying probability of reinforcement cracking. Numerical simulations are conducted to capture the elastoplastic mechanical behavior of particulate composites under uniaxial loading conditions. Comparison with available experimental data shows good agreement and illustrates the potential applicability of the proposed method for PRMMCs with evolutionary particle-cracking damage.

References

G. Bao, Damage due to fracture of brittle reinforcements in a ductile matrix. Acta Metall. Mater. **40**(10), 2547–2555 (1992)

M. Berveiller, A. Zaoui, An extension of the self-consistent scheme to plastically-flowing polycrystals. J. Mech. Phys. Solids **26**, 325–344 (1979)

N. Bourgeois, Caracterisation et modelisation micromecanique du comportement et de lendommagement dun composite a matrice metallique, Al/SiCp. Doctoral Thesis, Ecole Centrale des Arts et Manufactures: Chatenay-Malabry, France, 1994

J.R. Brockenbrough, F.W. Zok, On the role of particle cracking in flow and fracture of metal-matrix composites. Acta Metall. Mater. **43**(1), 11–20 (1995)

T.W. Clyne, P.J. Withers, *An Introduction to Metal Matrix Composites* (Cambridge University Press, Cambridge, 1993)

K. Derrien, J. Fitoussi, G. Guo, D. Baptiste, Prediction of the effective damage properties and failure properties of nonlinear anisotropic discontinuous reinforced composites. Comput. Methods Appl. Mech. Eng. **185**, 93–107 (2000)

J.D. Eshelby, The determination of the elastic field of an ellipsoidal inclusion and related problem. Proc. R. Soc. Lond. A. **241**, 376–396 (1957)

M. Finot, Y.L. Shen, A. Needleman, S. Suresh, Micromechanical modeling of reinforcement fracture in particle-reinforced metal-matrix composites. Metall. Mater. Trans. A **25**(11), 2403–2420 (1994)

S. Ghosh, S. Moorthy, Particle fracture simulation in non-uniform microstructures of metal-matrix composites. Acta Mater. **46**(3), 965–982 (1998)

C. Gonzalez, J. Llorca, A self-consistent approach to the elasto-plastic behaviour of two-phase materials including damage. J. Mech. Phys. Solids **48**, 675–692 (2000)

A.L. Gurson, Continuum theory of ductile rupture by void nucleation and growth, part I – yield criterion and flow rules for porous ductile media. ASME J. Eng. Mater. Technol. **99**(1), 2–15 (1977)

R. Hill, Continuum micro-mechanics of elastoplastic polycrystals. J. Mech. Phys. Solids **13**, 89–101 (1965)

M. Hori, S. Nemat-Nasser, Double-inclusion model and overall moduli of multi-phase composites. Mech. Mater. **14**, 189–206 (1993)

J.W. Ju, T.M. Chen, Micromechanics and effective moduli of elastic composites containing randomly dispersed ellipsoidal inhomogeneities. Acta Mech. **103**, 103–121 (1994)

J.W. Ju, L.Z. Sun, A novel formulation for exterior-point Eshelby's tensor of an ellipsoidal inclusion. ASME J. Appl. Mech. **66**, 570–574 (1999)

J.W. Ju, L.Z. Sun, Effective elastoplastic behavior of metal matrix composites containing randomly located aligned spheroidal inhomogeneities, Part I: micromechanics-based formulation. Int. J. Solids Struct. **38**, 183–201 (2001)
J.W. Ju, K.H. Tseng, Effective elastoplastic algorithms for ductile matrix composites. ASCE J. Eng. Mech. **123**(3), 260–266 (1997)
K. Lee, S. Moorthy, S. Ghosh, Multiple scale computational model for damage in composite materials. Comput. Methods Appl. Mech. Eng. **172**, 175–201 (1999)
M. Li, S. Ghosh, O. Richmond, H. Weiland, T.N. Rouns, Three dimensional characterization and modeling of particle reinforced metal matrix composites part II: damage characterization. Mater. Sci. Eng. A **266**, 221–240 (1999)
H.T. Liu, L.Z. Sun, J.W. Ju, Elastoplastic modeling of progressive interfacial debonding for particle-reinforced metal matrix composites. Acta Mech. **181**, 1–17 (2006)
J. Llorca, J.L. Martinez, M. Elices, Reinforcement fracture and tensile ductility in sphere-reinforced metal-matrix composites. Fatigue Fract. Eng. Mater. Struct. **20**(5), 689–702 (1997)
T. Mochida, M. Taya, M. Obata, Effect of damaged particles on the stiffness of a particle/metal matrix composite. JSME Int. J. Ser. I. Solid Mech. Strength Mater. **34**(2), 187–193 (1991)
T. Mori, K. Tanaka, Average stress in matrix and average elastic energy of materials with misfitting inclusions. Acta Metall. **21**, 571–574 (1973)
T. Mura, *Micromechanics of Defects in Solids*, 2nd edn. (Kluwer, Dordrecht, 1987)
S. Nemat-Nasser, M. Hori, *Micromechanics: Overall Properties of Heterogeneous Materials*, 2nd edn. (North-Holland, Amsterdam, 1999)
H.M. Shodja, A.S. Sarvestani, Elastic fields in double inhomogeneity by the equivalent inclusion method. ASME J. Appl. Mech. **68**(1), 3–10 (2001)
D. Steglich, T. Siegmund, W. Brocks, Micromechanical modeling of damage due to particle cracking in reinforced metals. Comput. Mater. Sci. **16**(1–4), 404–413 (1999)
L.Z. Sun, J.W. Ju, Effective elastoplastic behavior of metal matrix composites containing randomly located aligned spheroidal inhomogeneities, Part II: applications. Int. J. Solids Struct. **38**, 203–225 (2001)
L.Z. Sun, H.T. Liu, J.W. Ju, Effect of particle cracking on elastoplastic behavior of metal matrix composites. Int. J. Numer. Method Eng. **56**, 2183–2198 (2003)
S. Suresh, A. Mortensen, A. Needleman, *Fundamentals of Metal-Matrix Composites* (Butterworth-Heinemann Publisher, Boston, 1993)
G.P. Tandon, G.J. Weng, A theory of particle-reinforced plasticity. ASME J. Appl. Mech. **55**, 126–135 (1988)
V. Tvergaard, A. Needleman, Analysis of the cup-cone fracture in a round tensile bar. Acta Metall. **32**(1), 157–169 (1984)
W. Weibull, A statistical distribution function of wide applicability. ASME J. Appl. Mech. **18**, 293–297 (1951)
D.S. Wilkinson, E. Maire, J.D. Embury, The role of heterogeneity on the flow of two-phase materials. Mater. Sci. Eng. A **233**(1–2), 145–154 (1997)
D.S. Wilkinson, W. Pompe, M. Oeschner, Modeling the mechanical behaviour of heterogeneous multi-phase materials. Prog. Mater. Sci. **46**, 379–405 (2001)

Particle-Debonding Modeling of Metal-Matrix Composites

34

L. Z. Sun, H. T. Liu, and Jiann-Wen Woody Ju

Contents

Abstract

This chapter aims to model the mechanical behavior of particle-reinforced metal-matrix composites with progressive particle–matrix interfacial debonding. The partial-debonding process is represented by the debonding angles in composites. The equivalent orthotropic elasticity tensor is constructed for the

L.Z. Sun (✉)
Civil and Environmental Engineering, University of California, Irvine, CA, USA
e-mail: lsun@uci.edu

H.T. Liu
American Bureau of Shipping, Houston, TX, USA
e-mail: dr.haitao.liu@gmail.com

J.-W.W. Ju
Department of Civil and Environmental Engineering, University of California, Los Angeles CA, USA
e-mail: juj@ucla.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_10

debonded yet isotropic particles to characterize the reduction of the load-transfer ability in the debonded directions. Micromechanical homogenization procedures are utilized to estimate the effective moduli and the overall yield function of the resultant multiphase composites. The associative plastic flow rule and isotropic hardening law are postulated based on the continuum plasticity theory. The effect of partially interfacial debonding on the overall stress–strain relations of the composites is investigated and illustrated via numerical examples.

Introduction

In the last few decades, particle-reinforced metal-matrix composites (PRMMCs) have been rapidly developed to meet the need for better materials and structures. While the enhanced mechanical properties mostly come from the reinforcing particle phase, the latter also leads to new damage mechanisms that restrict the potential for widespread use of composites. Among the three prevalent damage micromechanisms in PRMMCs (i.e., debonding at the matrix–particle interface, internal particle cracking, and ductile plastic localization in the matrix (Clyne and Withers 1993; Suresh et al. 1993)), matrix–particle interfacial debonding is the predominant damage mode when the interfacial strength is relatively weak and the composite system is under high triaxial loading. To model the interfacial debonding, Jasiuk and Tong (1989), Pagano and Tandon (1990), Qu (1993), and Sangani and Mo (1997) introduced either a linear spring layer with vanishing thickness or an interlayer with constant thickness. In their models, a different elastic constant of the spring-/interlayer from the matrix and reinforcements is used to simulate the loss of load-transfer ability through the interface due to debonding. Since the spring-/interlayer elastic properties in their models are not position dependent, such models may not be applicable for partial-debonding mechanisms. Another simple yet physically meaningful means used to model the interfacial debonding is the equivalent stiffness method (Zhao and Weng 1997; Wong and Ait-Kadi 1997; Sun et al. 2003) in which the isotropic debonded particles are replaced by the perfectly bonded particles with constructed equivalent anisotropic stiffness to characterize the reduction of the load-transfer ability of the debonded interface. Thus, the conventional Eshelby's inclusion theory and the micromechanics method (Eshelby 1957, 1959; Mura 1987) can be applied to deal with the multiphase composites.

To capture the progressive process of the interfacial debonding, the evolution of the damage, and the transition between various debonding modes, a micromechanical framework has been developed to simulate the interfacial debonding between the matrix and reinforcing particles and to estimate its effect on the overall elastic behavior of particle-reinforced composites (Liu et al. 2004). In this model, the progressive damage process is represented by the debonding angles that are governed by external loads. To extend the study, this chapter focuses on the elastoplastic model to investigate the effect of the progressive partial interfacial debonding on the overall nonlinear response of ductile composites containing randomly dispersed particles. The effective elastoplastic-damage constitutive responses of the composites under various loading conditions are numerically simulated and compared with available experimental results.

Progressive Debonding Model

In real situations, interfacial debonding between particle and matrix is a progressive process, in which the debonding area evolves with the change of applied loads. The damage parameters should be dependent on the magnitude and mode of external loads. Following this consideration, interfacial progressive debonding model is developed in this chapter. For the particle-reinforced composites considered, the reinforcement consists of identical spherical particles, which are randomly distributed in the matrix. Both particles and matrix are assumed of isotropic materials. With applied external loading, some of the particles inside the material begin to debond, while the debonding areas and the number of debonded particles (volume fraction) will also proportionately change in the whole loading process. On the microstructural level, the debonding mechanism is treated as a fracture process, in which the local stresses play a key role. The elastic equivalency is constructed in terms of stiffness tensor for those debonded particles. Namely, the isotropic yet debonded particles are replaced by the orthotropic but perfect particles. The volume fraction evolution of debonded particles is phenomenologically characterized by the Weibull statistical approach. Incorporating these considerations into the general micromechanics frame shown in Fig. 1, a damage model of the PRMMCs with progressive interfacial debonding is established in this section.

Debonding Modes

In what follows, a local *Cartesian* coordinate system is selected as the same as the three principal directions of the local stress field inside a particle. The three local principal stresses (σ_1, σ_2, σ_3) follow the conventional order of $\sigma_1 \geq \sigma_2 \geq \sigma_3$. Under this local coordinates, the normal direction at a certain surface point P of a spherical particle can be expressed as $\mathbf{n} = \{\sin\,\phi\,\cos\,\theta,\ \sin\,\phi\,\sin\,\theta,\ \cos\,\phi\}$ where θ and ϕ are the two Eulerian angles. The normal stress at point P can be derived as

$$\sigma^{normal} = \sigma_1(\sin\phi\cos\theta)^2 + \sigma_2(\sin\phi\sin\theta)^2 + \sigma_3(\cos\phi)^2. \tag{1}$$

Emanating from the assumption that the normal stress controls the debonding initiation on the interface, as the (tensile) normal stress reaches a critical interfacial debonding strength σ_{cri}, the debonding process would initiate at the point P. The debonding criterion can be conveniently written as

$$\sigma_1(\sin\phi\cos\theta)^2 + \sigma_2(\sin\phi\sin\theta)^2 + \sigma_3(\cos\phi)^2 \geq \sigma_{cri}. \tag{2}$$

Therefore, the relationship between the local principal stresses and the critical debonding strength results in the following four different types of interfacial debonding modes:

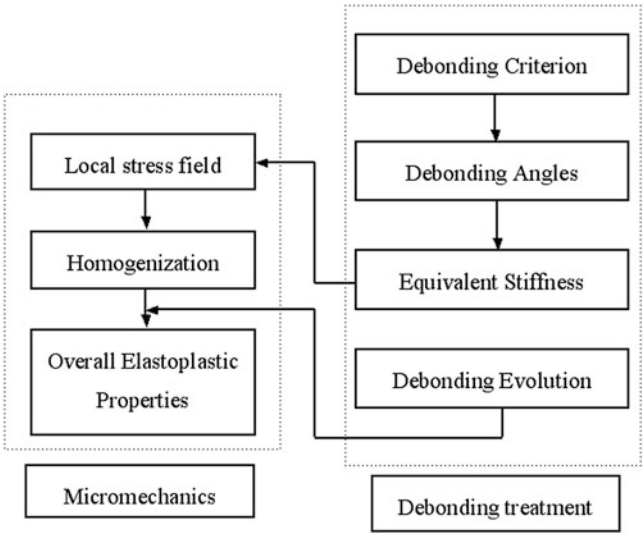


Fig. 1 Micromechanics framework of incorporating progressive debonding models into elastoplastic constitutive relations of PRMMCs

Mode 1: $\sigma_{cri} \geq \sigma_1 \geq \sigma_2 \geq \sigma_3$. This is actually the original case – all the particles are perfectly bonded since none of the principal stresses exceeds the critical debonding strength. No debonding process has been activated.

Mode 2: $\sigma_1 \geq \sigma_{cri} \geq \sigma_2 \geq \sigma_3$. Only one principal stress is greater than the critical interfacial debonding strength σ_{cri}. In this case, the interfacial debonding initiates from the local x_1 direction and propagates progressively towards the other two principal directions. The debonding area can be characterized by two debonding angles $\alpha_{12}^{(2)}$ and $\alpha_{13}^{(2)}$ (Fig. 2), which are determined as

$$\alpha_{1\gamma}^{(2)} = \sin^{-1}\sqrt{\frac{\sigma_1 - \sigma_{cri}}{\sigma_1 - \sigma_\gamma}}, \quad \gamma = 2, 3. \tag{3}$$

The incremental form of the debonding angles is consequently given by

$$d\alpha_{1\gamma}^{(2)} = \frac{(\sigma_{cri} - \sigma_\gamma)d\sigma_1 + (\sigma_1 - \sigma_{cri})d\sigma_\gamma}{\sqrt{(\sigma_1 - \sigma_\gamma)^4 - (\sigma_1 - \sigma_\gamma)^2(\sigma_{cri} - \sigma_1)^2}}, \quad \gamma = 2, 3. \tag{4}$$

Mode 3: $\sigma_1 \geq \sigma_2 \geq \sigma_{cri} > \sigma_3$. In this case, both the principal stresses σ_1 and σ_2 in the x_1 and x_2 directions exceed the critical debonding strength. The normal stress must be greater than the critical strength, which in turn indicates that the debonding is complete in the entire $x_1 - x_2$ plane. As a result, only two debonding angles $\alpha_{13}^{(3)}$ and $\alpha_{23}^{(3)}$ need to be introduced, representing the debonding processes in the planes of $x_1 - x_3$ and $x_2 - x_3$, respectively, as shown in Fig. 3 for the one-eighth part of the particle-debonding process. Following the same argument as in mode 2, these mode 3 debonding angles are determined as

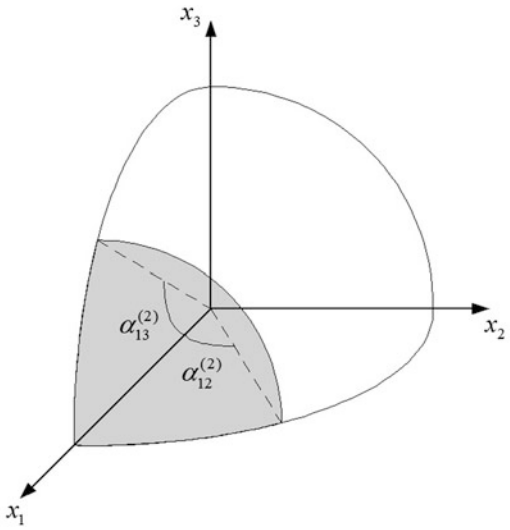


Fig. 2 Mode 2 debonding process initiating from the x_1 direction. Two debonding angles are used to represent the debonding extents in x_2 and x_3 directions, respectively

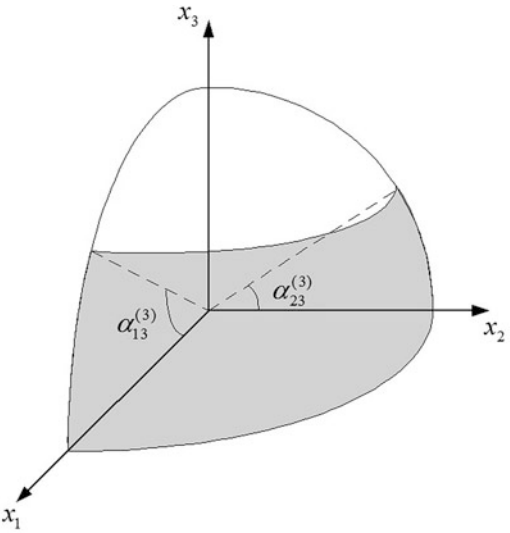


Fig. 3 Mode 3 debonding process in two directions. Two debonding angles are used to represent the debonding extents in x_3 direction from x_1 and x_2 directions, respectively

$$\alpha_{\gamma 3}^{(3)} = \sin^{-1}\sqrt{\frac{\sigma_\gamma - \sigma_{cri}}{\sigma_\gamma - \sigma_3}}, \quad \gamma = 1, 2. \tag{5}$$

The corresponding incremental form of the debonding angles reads

$$d\alpha_{\gamma 3}^{(3)} = \frac{(\sigma_{cri} - \sigma_3)d\sigma_\gamma + (\sigma_\gamma - \sigma_{cri})d\sigma_3}{\sqrt{(\sigma_\gamma - \sigma_3)^4 - (\sigma_\gamma - \sigma_3)^2(\sigma_{cri} - \sigma_\gamma)^2}}, \quad \gamma = 1, 2. \tag{6}$$

Mode 4: $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \sigma_{cri}$. All three principal stresses exceed the critical debonding strength, which indicates that the normal stress at any point on a particle surface is greater than the critical strength. Therefore, the entire interface is debonded.

It is noted that for all of the above four damage modes, the range of debonding angles is between 0 and $\pi/2$. The lower and upper bounds of the debonding angles in a certain principal direction correspond to the perfect bonding and total debonding, respectively, in that direction.

Equivalent Inclusion Treatment

The partially interfacial debonding results in the *partial* loss of load-transfer capacity in the debonding directions, which is simulated by the reduction of the corresponding elastic stiffness of the debonded particles in that direction (Liu et al. 2004; Zhao and Weng 2002). Therefore, the partially debonded isotropic particles are replaced by perfectly bonded particles with constructed equivalent orthotropic stiffness tensors. To establish the relationship between the debonding angles and the loss of tensile load-transfer capacity, which is manifested by a reduction in elastic stiffness, three distinct interfacial damage parameters $D_i^{(\beta)}$ serving as the measures of elastic stiffness reduction in certain directions are defined using the ratio between the projected damage area in a certain direction and the original interface area. Here, the superscript $\beta = 1, 2, 3, 4$ refers to the four distinct debonding modes and the subscript $i = 1, 2, 3$ represents the damage effect on the three principal directions, respectively. In accordance with the four different debonding modes, the interfacial damage parameters can be developed as follows for various debonding modes, respectively:

$$D_1^{(1)} = D_2^{(1)} = D_3^{(1)} = 0. \tag{7}$$

$$\begin{aligned}
D_1^{(2)} &= \sin\alpha_{12}^{(2)} \sin\alpha_{13}^{(2)} \\
D_2^{(2)} &= \frac{2}{\pi}\left[\alpha_{13}^{(2)} - \frac{\sin\alpha_{12}^{(2)}\cos^2\alpha_{12}^{(2)}\sinh^{-1}\left(\sqrt{\sin^2\alpha_{12}^{(2)} - \sin^2\alpha_{13}^{(2)}}/\cos\alpha_{12}^{(2)}\right)}{\sqrt{\sin^2\alpha_{12}^{(2)} - \sin^2\alpha_{13}^{(2)}}}\right] \\
D_3^{(2)} &= \frac{2}{\pi}\left[\alpha_{12}^{(2)} - \frac{\sin\alpha_{12}^{(2)}\cos^2\alpha_{13}^{(2)}\sin^{-1}\left(\sqrt{\sin^2\alpha_{12}^{(2)} - \sin^2\alpha_{13}^{(2)}}/\cos\alpha_{13}^{(2)}\right)}{\sqrt{\sin^2\alpha_{12}^{(2)} - \sin^2\alpha_{13}^{(2)}}}\right].
\end{aligned} \tag{8}$$

$$
\begin{aligned}
D_1^{(3)} &= \frac{2}{\pi}\left[\alpha_{23}^{(3)} + \frac{\sin^2\alpha_{13}^{(3)}\cos\alpha_{23}^{(3)}\sin^{-1}\left(\sqrt{\sin^2\alpha_{13}^{(3)} - \sin^2\alpha_{23}^{(3)}}/\sin\alpha_{13}^{(3)}\right)}{\sqrt{\sin^2\alpha_{13}^{(3)} - \sin^2\alpha_{23}^{(3)}}}\right] \\
D_2^{(3)} &= \frac{2}{\pi}\left[\alpha_{13}^{(3)} + \frac{\sin^2\alpha_{23}^{(3)}\cos\alpha_{13}^{(3)}\sinh^{-1}\left(\sqrt{\sin^2\alpha_{13}^{(3)} - \sin^2\alpha_{23}^{(3)}}/\sin\alpha_{23}^{(3)}\right)}{\sqrt{\sin^2\alpha_{13}^{(3)} - \sin^2\alpha_{23}^{(3)}}}\right] \\
D_3^{(3)} &= 1 - \cos\alpha_{13}^{(3)}\cos\alpha_{23}^{(3)}.
\end{aligned}
\tag{9}
$$

$$
D_1^{(4)} = D_2^{(4)} = D_3^{(4)} = 1. \tag{10}
$$

The debonding damage parameters vary from 0 to 1. A greater value of the damage parameter signifies a more significant level of *reduction* in tensile load-transfer capability.

With the help of the damage parameters defined in Eqs. 7, 8, 9, and 10, for the β-mode ($\beta = 1, 2, 3, 4$) particles, the equivalent orthotropic elastic compliance matrix can be formulated as

$$
\mathbf{M}^{(\beta)} = \begin{pmatrix}
\frac{1}{E^{(1)}\xi_{11}^{(\beta)}} & -\frac{\nu^{(1)}}{E^{(1)}\xi_{12}^{(\beta)}} & -\frac{\nu^{(1)}}{E^{(1)}\xi_{13}^{(\beta)}} & 0 & 0 & 0 \\
-\frac{\nu^{(1)}}{E^{(1)}\xi_{12}^{(\beta)}} & \frac{1}{E^{(1)}\xi_{22}^{(\beta)}} & -\frac{\nu^{(1)}}{E^{(1)}\xi_{23}^{(\beta)}} & 0 & 0 & 0 \\
-\frac{\nu^{(1)}}{E^{(1)}\xi_{13}^{(\beta)}} & -\frac{\nu^{(1)}}{E^{(1)}\xi_{23}^{(\beta)}} & \frac{1}{E^{(1)}\xi_{33}^{(\beta)}} & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1+\nu^{(1)}}{E^{(1)}\xi_{23}^{(\beta)}} & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1+\nu^{(1)}}{E^{(1)}\xi_{12}^{(\beta)}} & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1+\nu^{(1)}}{E^{(1)}\xi_{13}^{(\beta)}}
\end{pmatrix}
\tag{11}
$$

where $E^{(1)}$ and $\nu^{(1)}$ are the Young's modulus and Poisson's ratio of the original (perfect) particle (with corresponding elastic stiffness tensor $C_{ijkl}^{(1)}$), respectively, and the parameters $\xi_{ij}^{(\beta)}$ are defined as

$$
\xi_{ij}^{(\beta)} = \left(1 - D_i^{(\beta)}\right)\left(1 - D_j^{(\beta)}\right), \quad i, j = 1, 2, 3. \tag{12}
$$

The equivalent elastic compliance $\mathbf{M}^{(\beta)}$ ($\beta = 1, 2, 3, 4$) shows that the equivalent orthotropic Young's modulus and shear modulus in a certain direction decrease as

the corresponding damage parameters increase. When all damage parameters are equal to 0, the compliance matrix reduces to the original (perfectly bonded) elastic compliance. By contrast, when a certain damage parameter reaches 1, the load-transfer capability in *that* direction becomes totally lost. In general, the compliance matrix of particles is orthotropic and reflects the reduction in the elastic moduli of particles as the applied far-field loads increase.

The equivalent stiffness matrix of particles is further obtained through the inversion procedure of the foregoing compliance matrix. The corresponding fourth-rank equivalent elastic stiffness tensor of particles follows the form

$$C_{ijkl}^{(\beta)} = \lambda_{IK}^{(\beta)} \delta_{ij}\delta_{kl} + \mu_{IJ}^{(\beta)} \left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) \tag{13}$$

where

$$\begin{aligned} \lambda_{IK}^{(\beta)} &= \lambda^{(1)} \left(1 - D_I^{(\beta)}\right)\left(1 - D_K^{(\beta)}\right) \\ \mu_{IJ}^{(\beta)} &= \mu^{(1)} \left(1 - D_I^{(\beta)}\right)\left(1 - D_J^{(\beta)}\right) \end{aligned} \tag{14}$$

with $\lambda^{(1)}$, $\mu^{(1)}$ denoting the isotropic Lame constants of the original particles (perfectly bonded particles). Here, Mura's tensorial indicial notation (Mura 1987) is followed, i.e., the repeated lowercase indices are summed up from 1 to 3, whereas the uppercase indices take on the same numbers as the corresponding lowercase ones but are not summed up. This indicial expression facilitates the subsequent derivations and computations. By using the elastic equivalency treatment, all partially debonded particles are replaced by perfectly bonded particles with the aforementioned equivalent orthotropic elastic stiffness. Therefore, multiphase micromechanical approaches can be established to characterize the progressive interfacial debonding processes of metal-matrix composites.

Damage Evolution

The evolution of the volume fractions of the damaged particles can be simulated using a two-parameter Weibull distribution function as (Tohgo and Weng 1994)

$$P_i = \begin{cases} 1 - \exp\left[-\left(\frac{\sigma_i - \sigma_{cri}}{S}\right)^M\right], & \sigma_i \geq \sigma_{cri} \\ 0, & \sigma_i < \sigma_{cri} \end{cases} \quad (i = 1, 2, 3). \tag{15}$$

Here, the Weibull parameter M signifies the evolution rate of the volume fraction of debonded particles. The parameter S is not independent when the relationship between the mean value of the Weibull distribution function and the critical bonding strength σ_{cri} is established. For example, when an intermediate debonding evolution rate $M = 5$ is chosen and the mean of the Weibull function is equal to the critical bonding strength, S can be calculated as $S = 1.09\sigma_{cri}$.

The volume fractions of the four various damage modes $\phi^{(\beta)}$ ($\beta = 1, 2, 3, 4$) are expressed as follows to characterize the evolution of interfacial partial debonding and the transition between the four debonding modes:

$$\begin{aligned}\phi^{(4)} &= \phi^{Total} P_3 \\ \phi^{(3)} &= \phi^{Total}[P_2 - P_3] \\ \phi^{(2)} &= \phi^{Total}[P_1 - P_2] \\ \phi^{(1)} &= \phi^{Total}[1 - P_1]\end{aligned} \tag{16}$$

where ϕ^{Total} is the total volume fraction of all particles in the composite. The probabilistic function $P_i (i = 1, 2, 3)$ (Eq. 15) represents a normal stress-controlled debonding process and can be treated as the debonding probability in the ith principal direction. When the smallest principal stress (the third principal stress) reaches the critical stress, the other two principal stresses are greater or equal to the critical stress, and thus the total debonding (the fourth mode) occurs. Therefore, the evolution of the fourth debonding mode is characterized by P_3. Similarly, the third debonding mode (two-dimensional debonding) and the second debonding mode (one-dimensional debonding) are controlled by P_2 and P_1, respectively. With the increase of external loading, part of the two-dimensional debonding converts to the total debonding, and part of the one-dimensional debonding evolves into the two-dimensional debonding. Hence, the volume fractions of the four debonding modes are constructed as Eq. 16 to reflect the transition of debonding modes.

Elastoplastic and Damage Modeling

Although the plastic deformation in composites is highly localized, the mean-field principles (homogenization procedures) can be directly applied to estimate the effective yield strength of composites since the initial yielding and plastic hardening of composites should be attributed to the collective responses of particle–matrix interactions (Mura 1987). To obtain the effective yield function of PRMMCs, the averaging homogenization is generally performed within a mesoscopic representative volume element (RVE) (Nemat-Nasser and Hori 1999). At any local matrix material point $\mathbf{x}$, the microscopic stress $\boldsymbol{\sigma}(\mathbf{x})$ is assumed to satisfy the von Mises J_2-yield criterion and the local matrix yield function takes the form

$$F\left(\boldsymbol{\sigma}, \bar{e}_m^p\right) = \sqrt{\boldsymbol{\sigma} : \mathbf{I}_d : \boldsymbol{\sigma}} - K\left(\bar{e}_m^p\right) \leq 0 \tag{17}$$

where $\bar{e}_m^p$ and $K\left(\bar{e}_m^p\right)$ are the equivalent plastic strain and the isotropic hardening function of the matrix-only material, respectively. Moreover, $\mathbf{I}_d$ denotes the deviatoric part of the fourth-rank identity tensor $\mathbf{I}$.

Following the ensemble-volume averaging processes (Ju and Sun 2001), the overall yield function for the composites can be expressed as

$$\overline{F} = \left(1 - \phi^{(1)}\right)\sqrt{\langle H\rangle_m} - K(\overline{e}^p) \le 0 \tag{18}$$

where $\overline{e}^p$ represents the effective equivalent plastic strain. It should be noted that the effect of the various debonding modes on the overall yield function is reflected through the change of stress field. This is shown in the expression of $\langle H\rangle_m$, which is a function of the volume fraction $\phi^{(i)}$, $(i = 1, 2, 3, 4)$. The expression of $\langle H\rangle_m$ for the composite with four phases of particles (corresponding to the four debonding modes) can be approximately obtained by neglecting the interaction among neighboring particles as

$$\langle H\rangle_m(\boldsymbol{x}) \cong H^0 + \sum_{\beta=1}^{4} \oint_{\mathbf{x}' \notin \Xi(\mathbf{x})} \left[H^{(\beta)}(\mathbf{x}|\mathbf{x}') - H^0\right] P^{(\beta)}(\mathbf{x}')\, d\mathbf{x}' \tag{19}$$

where $H^0 = \boldsymbol{\sigma}_0 : \mathbf{I}_d : \boldsymbol{\sigma}_0$ is the square of the far-field stress norm applied on the composite and $\Xi(\mathbf{x})$ is the exclusion zone of $\mathbf{x}$ for the center location $\mathbf{x}'$ of a particle in the probability space, which is identical to the shape and size of the particle. Further, $P^{(\beta)}(\mathbf{x}')$ is the probability density function for finding a β-phase particle in the exclusion zone of the particle center located at $\mathbf{x}'$. In addition, $H^{(\beta)}$ is the stress-norm collection contribution from the β-phase particles, i.e.,

$$H^{(\beta)} = \boldsymbol{\sigma}^{(\beta)}(\mathbf{x}) : \mathbf{I}_d : \boldsymbol{\sigma}^{(\beta)}(\mathbf{x}), \quad \beta = 1, 2, 3, 4 \tag{20}$$

where the local stress tensor in the matrix due to a β-phase particle centered at $\mathbf{x}'$ can be written as

$$\boldsymbol{\sigma}(\mathbf{x}) = \boldsymbol{\sigma}_0 + \mathbf{C}^0 : \overline{\mathbf{G}}(\mathbf{x} - \mathbf{x}') : \boldsymbol{\varepsilon}_*^{(\beta)} \tag{21}$$

in which $\boldsymbol{\varepsilon}_*^{(\beta)}$ is the eigenstrain tensor in the β-phase particles that can be expressed explicitly for spherical particles (Liu et al. 2004) and $\overline{\mathbf{G}}$ is the exterior-point Eshelby's tensor (Ju and Sun 1999) and has the following simple form for spherical particles:

$$\begin{aligned}\overline{G}_{ijkl}(\mathbf{x}) = {} & \frac{\rho^3}{30(1-\nu_0)} \\ & \times \begin{bmatrix} (3\rho^2 + 10\nu_0 - 5)\delta_{ij}\delta_{kl} + 15(1-\rho^2)\delta_{ij}n_k n_l \\ +(3\rho^2 - 10\nu_0 + 5)\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) \\ +15(1 - 2\nu_0 - \rho^2)\delta_{kl}n_i n_j + 15(7\rho^2 - 5)n_i n_j n_k n_l \\ +15(\nu_0 - \rho^2)\left(\delta_{ik}n_j n_l + \delta_{il}n_j n_k + \delta_{jk}n_i n_l + \delta_{jl}n_i n_k\right) \end{bmatrix}\end{aligned} \tag{22}$$

where ν_0 is the Poisson's ratio of the matrix and $\rho = a/r$, in which a is the radius of the sphere, $r = \sqrt{x_i x_i}$, and $n_i = x_i/r$.

For simplicity, let us consider that all particles are uniformly randomly distributed in the composites. Therefore, $P^{(\beta)}(\mathbf{x}')$ can be assumed to be $N^{(\beta)}/V$

($\beta = 1, 2, 3, 4$), where $N^{(\beta)}$ is the total number of β-phase particles uniformly dispersed in volume V of RVE. After a series of lengthy but straightforward derivations, the following ensemble-averaged $\langle H \rangle_m$ expression is derived:

$$\langle H \rangle_m = \boldsymbol{\sigma}_0 : \mathbf{T} : \boldsymbol{\sigma}_0 \tag{23}$$

where the components of the fourth-rank tensor $\mathbf{T}$ for spherical particles take the form

$$T_{ijkl} = T_{IK}^{(1)} \delta_{ij}\delta_{kl} + T_{IJ}^{(2)} \left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) \tag{24}$$

with

$$\begin{aligned} T_{IK}^{(1)} &= -\frac{1}{3} + \frac{2}{675(1-v_0)^2} \times \left[\begin{array}{l} \left(65v_0^2 - 50v_0 + 2\right) \sum_{\beta=1}^{4} \dfrac{\phi^{(\beta)}}{B_{II}^{(\beta)} B_{KK}^{(\beta)}} \\ -75(1-2v_0)^2 \sum_{\beta=1}^{4} \dfrac{\phi^{(\beta)} \left(\Gamma_{II}^{(\beta)} + \Gamma_{KK}^{(\beta)}\right)}{B_{II}^{(\beta)} B_{KK}^{(\beta)}} \\ +225(1-2v_0)^2 \sum_{\beta=1}^{4} \dfrac{\phi^{(\beta)} \Gamma_{II}^{(\beta)} \Gamma_{KK}^{(\beta)}}{B_{II}^{(\beta)} B_{KK}^{(\beta)}} \end{array} \right] \\ T_{IJ}^{(2)} &= \frac{1}{2} + \frac{\left(35v_0^2 - 50v_0 + 23\right)}{225(1-v_0)^2} \sum_{\beta=1}^{4} \frac{\phi^{(\beta)}}{B_{IJ}^{(\beta)} B_{IJ}^{(\beta)}}. \end{aligned} \tag{25}$$

Here, $\phi^{(\beta)}$ denotes the volume fraction of the β-phase particles. Other parameters in the above equation can be found in Liu et al. (2006).

The general relationship between the applied far-field stress $\boldsymbol{\sigma}_0$ and the macroscopic (ensemble-volume averaged) stress $\overline{\boldsymbol{\sigma}}$ takes the form (Ju and Chen 1994)

$$\boldsymbol{\sigma}_0 = \mathbf{P} : \overline{\boldsymbol{\sigma}} \tag{26}$$

where the fourth-rank tensor $\mathbf{P}$ reads

$$\mathbf{P} = \left\{ \mathbf{C}^{(0)} \cdot [\mathbf{I} + (\mathbf{I} - \mathbf{S}) \cdot \mathbf{Y}] \cdot \mathbf{C}^{(0)^{-1}} \right\}^{-1} \tag{27}$$

with $\mathbf{Y} = \sum_{\beta=1}^{4} \phi^{(\beta)} \left[\mathbf{S} + \left(\mathbf{C}^{(\beta)} - \mathbf{C}^{(0)} \right)^{-1} \cdot \mathbf{C}^{(0)} \right]^{-1}$ and $\mathbf{S}$ being the Eshelby's tensor for spherical particles. The combination of Eqs. 23 and 26 leads to an alternative expression of the ensemble-averaged square of the current stress norm as

$$\langle H \rangle_m = \overline{\boldsymbol{\sigma}} : \overline{\mathbf{T}} : \overline{\boldsymbol{\sigma}} \tag{28}$$

where the fourth-rank tensor reads $\overline{\mathbf{T}} = \mathbf{P}^{\mathbf{T}} \cdot \mathbf{T} \cdot \mathbf{P}$. It is observed from the above equation that $\langle H \rangle_m$ can be reduced to the corresponding equation in Ju and Sun

(Ju and Sun 2001) if no particle debonding occurs in the composites (i.e., $\phi^{(1)} > 0$ and $\phi^{(\beta)} = 0, \beta = 2, 3, 4$). Furthermore, Eq. 28 will recover the classical J_2-invariant for the matrix-only material (i.e., $\phi^{(\beta)} = 0, \beta = 1, 2, 3, 4$).

The total macroscopic strain $\overline{\boldsymbol{\varepsilon}}$ consists of two parts:

$$\overline{\boldsymbol{\varepsilon}} = \overline{\boldsymbol{\varepsilon}}^{\mathrm{e}} + \overline{\boldsymbol{\varepsilon}}^{p} \tag{29}$$

where $\overline{\boldsymbol{\varepsilon}}^{\mathrm{e}}$ denotes the overall elastic strain and $\overline{\boldsymbol{\varepsilon}}^{p}$ represents the overall plastic strain of the composites. The relationship between the macroscopic stress $\overline{\boldsymbol{\sigma}}$ and macroscopic elastic strain $\overline{\boldsymbol{\varepsilon}}^{\mathrm{e}}$ reads

$$\overline{\boldsymbol{\sigma}} = \overline{\mathbf{C}} : \overline{\boldsymbol{\varepsilon}}^{\mathrm{e}} \tag{30}$$

in which the effective elastic stiffness of composites can be determined as (Ju and Chen 1994)

$$\overline{\mathbf{C}} = \mathbf{C}^{(0)} \cdot \left[\mathbf{I} + \left(\mathbf{Y}^{-1} - \mathbf{S}\right)^{-1}\right]. \tag{31}$$

The plastic flow of composites is postulated to be associative for simplicity. The macroscopic plastic strain rate for PRMMCs thus takes the form

$$\dot{\overline{\boldsymbol{\varepsilon}}}^{p} = \dot{\lambda}\frac{\partial \overline{F}}{\partial \overline{\boldsymbol{\sigma}}} \tag{32}$$

where $\dot{\lambda}$ is the plastic consistency parameter. Moreover, $\overline{F}$ is the overall yield function of composites. The simplified isotropic power-law hardening function $K(\overline{\varepsilon}^{p})$ is proposed as

$$K(\overline{\varepsilon}^{p}) = \sqrt{\frac{2}{3}}\left[\sigma_y + h(\overline{\varepsilon}^{p})^{q}\right]. \tag{33}$$

Here, σ_y denotes the initial yield stress of matrix material, and h and q signify the linear and exponential isotropic hardening parameters.

The foregoing characterization together with the Kuhn–Tucker conditions

$$\dot{\lambda} \geq 0, \quad \overline{F} \leq 0, \quad \dot{\lambda}\overline{F} = 0, \quad \dot{\lambda}\dot{\overline{F}} = 0 \tag{34}$$

constitutes an effective elastoplastic-damage constitutive formulation for particle-reinforced metal-matrix composites with a progressive, partial interfacial debonding process. The proposed composite framework as formulated in Eqs. 29, 30, 31, 32, 33, and 34 is based on the micromechanics approach, the ensemble-averaging homogenization procedures, and the statistical distribution method. The proposed formulation offers a potentially viable framework to estimate the overall elastoplastic-damage stress–strain responses of the metal-matrix composites.

Numerical Examples

Uniaxial Loading

The uniaxial stress–strain curves are often referred to as important indicators of mechanical responses of composite materials. In order to illustrate the proposed micromechanics-based model, the uniaxial tension loading is considered. The components of the macroscopic stress $\overline{\boldsymbol{\sigma}}$ can be expressed as $\overline{\sigma}_{11} > 0$ and $\overline{\sigma}_{ij} = 0$ for all other stress components. The selected composite system signifies the SiC particle-reinforced aluminum matrix composite. The Young's moduli and Poisson's ratios of SiC particles and the aluminum matrix are taken as $E_p =$ 450 GPa, $E_m = 70$ GPa, $v_p = 0.2$ and $v_m = 0.3$, where the subscripts p and m represent the particles and the matrix, respectively. The Weibull parameter is selected as $M = 5$. The yield strength is taken as $\sigma_y = 300$ MPa and the hardening parameters are assumed to be $h = 1.0$ GPa and $q = 0.5$, respectively. Numerical results on the uniaxial elastoplastic-damage stress–strain behaviors of PRMMCs are displayed in Fig. 4. Clearly, the effective stress–strain responses for a 50 % volume fraction of spherical particles with progressive interfacial debonding lie between those of the porous material (the lower bound) and the composite material without debonding (the upper bound). The interfacial bonding strength plays an important role in the interfacial debonding process and has significant effects on the overall elastoplastic-damage behaviors of the composites. Specifically, less strain hardening can be clearly observed for composites with lower interfacial bonding strengths. For example, with a very low bonding strength (i.e., $\sigma_{cri} = 0.4\ \sigma_y$), the interfacial debonding occurs even before the composite reaches its overall yield point, and therefore a softening portion can be observed.

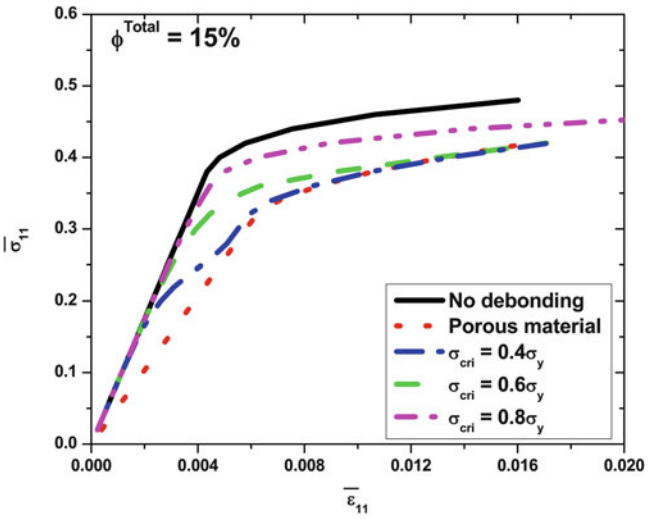


Fig. 4 The overall stress–strain curves of PRMMCs with progressive interfacial debonding

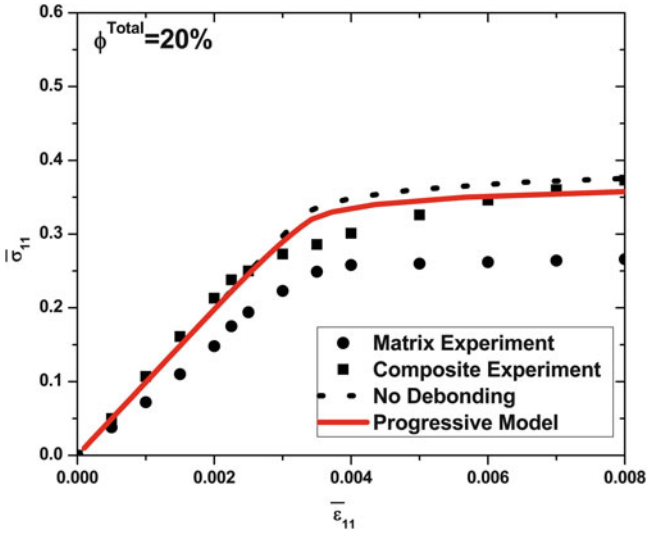


Fig. 5 The comparison between the prediction of current model and the experimental result (Papazian and Adler 1990)

Under the uniaxial loading condition, the first local principal stress is tensile and the other two local principal stresses are compressive with the same magnitudes. Therefore, only one debonding mode – debonding along the loading direction – will occur. Because of the equality of the second and third local principal stresses, the debonding angles $\alpha_{12}^{(2)}$ and $\alpha_{13}^{(2)}$ are equal to each other. It is observed that the debonding angle increases rapidly at the beginning stage and then becomes saturated as the overall deformation increases, implying that it is increasingly more difficult to further debond due to the compressive principal stresses in the second and third principal directions. Furthermore, Fig. 5 exhibits the comparison between the proposed model predictions and the experimental results of particulate-reinforced SiC/Al5456 composites reported by Papazian and Adler (Papazian and Adler 1990). With the introduction of progressive interfacial debonding, the current model obtains a better prediction than the non-debonding model. The effects of debonding on the overall initial yield strengths of the composites can be demonstrated. When the particle volume fraction is zero, there are no reinforcements in the composite, and therefore it becomes a matrix-only material. The overall yield strengths of the composites increase with increasing volume fractions of particles without considering debonding, which reflects the enhancing hardening influence of the existence of reinforcing particles. By contrast, for porous materials with voids, a higher volume fraction of voids leads to a decrease of the overall initial yield strength. With higher particle concentrations in the composites, the interfacial debonding mechanism creates more pronounced effects on the overall initial yield strengths of the composites. Moreover, as the bonding strength becomes weaker, particle debonding occurs in early loading stage; and as the overall initial yield

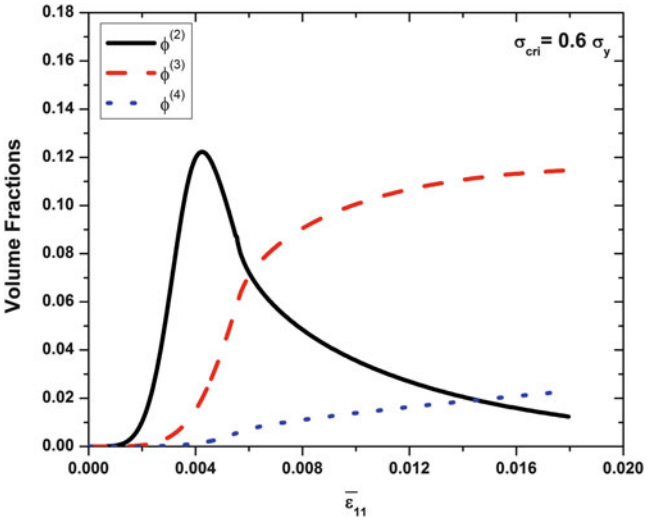


Fig. 6 The evolution of volume fractions with the increased overall strain under triaxial loading condition

takes place, there are already many void-like particles in the composites. As a consequence, the overall initial yield strengths of the weak-interface composites go even lower than the matrix-only material and the voids-prone composites behave like porous materials.

Triaxial Loading

Under the uniaxial loading condition, only one debonding mode is active since only the first local principal stress is tensile. To investigate the transformation between different debonding modes, the composite is subjected to the triaxial loading case, in which $\overline{\sigma}_{11} > 0, \overline{\sigma}_{22} = 0.6\overline{\sigma}_{11}$ and $\overline{\sigma}_{33} = 0.4\overline{\sigma}_{11}$. Figure 6 shows the evolution and transformation of the damage volume fractions. With the increase of the overall external loading, the first local principal stress reaches the critical strength first and activates the second debonding mode (one-dimensional debonding). Further increasing the external loading, the third debonding mode (two-dimensional debonding) becomes active when the second local principal stress reaches the critical strength. At the beginning stage, the number of newly formed mode 2 particles from the perfectly bonded (mode 1) particles is larger than the amount of mode 2 particles that evolve into mode 3 and mode 4 ones. Therefore, both $\phi^{(2)}$ and $\phi^{(3)}$ increase with the increasing external loading until a specific point where more mode 2 particles become mode 3 and mode 4 while $\phi^{(2)}$ begins to decrease. Finally, all the local principal stresses go beyond the critical strength, and all four modes of particles exist in the matrix simultaneously. The progression of debonding angles is shown in Fig. 7. Comparing with the uniaxial loading case, the debonding

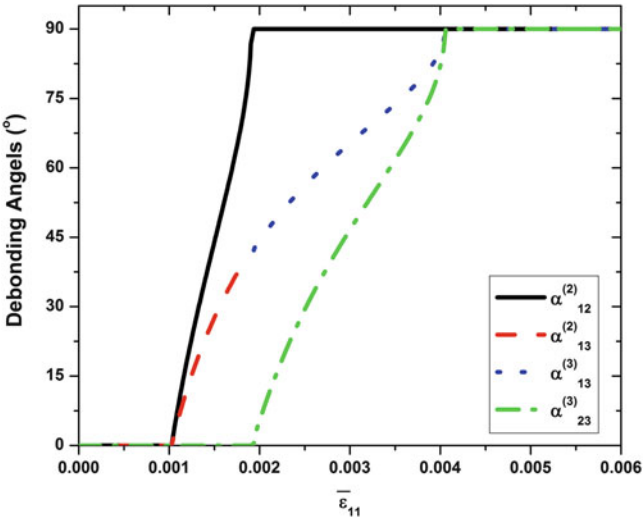


Fig. 7 The progress of debonding angles with the increased overall strain under triaxial loading condition

process develops more rapidly and the debonding angles reach the maximum value (90°) finally due to the tensile state of the principal stresses in all directions. When the second principal stress reaches the critical strength, the debonding angle $\alpha_{12}^{(2)}$ becomes 90° indicting a total debonding between the local x_1 and x_2 directions. Therefore, the two-dimensional debonding mode is activated (referring to $\phi^{(3)}$ in Fig. 6). The debonding angle in $x_1 - x_3$ direction ($\alpha_{13}^{(2)}$) keeps developing under the two-dimensional debonding mode ($\alpha_{13}^{(3)}$). Once all the debonding angles progress to 90°, the total debonding mode (mode 4) is achieved (referring to $\phi^{(4)}$ in Fig. 6).

To investigate the effects of interfacial particle debonding on the overall initial yield surfaces, axisymmetric (biaxial) loading cases are considered here. In this loading case, $\overline{\sigma}_{11} > 0, \overline{\sigma}_{22} = \overline{\sigma}_{33} > 0$ and $\overline{\sigma}_{12} = \overline{\sigma}_{13} = \overline{\sigma}_{23} = 0$. The initial effective yield surfaces are presented in terms of the normalized volumetric and effective stresses in Figs. 8 and 9. Specifically, the volumetric and effective stresses can be easily obtained from their basic definitions under the axisymmetric condition as $\overline{\sigma}_v = (\overline{\sigma}_{11} + 2\overline{\sigma}_{22})/3$ and $\overline{\sigma}_e = \overline{\sigma}_{11} - \overline{\sigma}_{22}$, respectively. Figure 8 shows that the overall initial yielding of the composites is not of the von Mises type, even though the particles are spherical in shape and randomly distributed. The effect of interfacial bonding strength is significant on the normalized volumetric yield stress, whereas considerably less influence is observed on the normalized effective yield stress. With decreasing bonding strength, early debonding occurs in the composite and causes the degradation of the initial plastic yielding point for the volumetric stress. On the other hand, for a constant (specified) bonding strength, the effects of particle volume fraction on the overall initial yield surface in the volumetric and effective stress space are exhibited in Fig. 9. In particular, decreasing particle volume fraction leads to an increase in the normalized volumetric yield stress yet

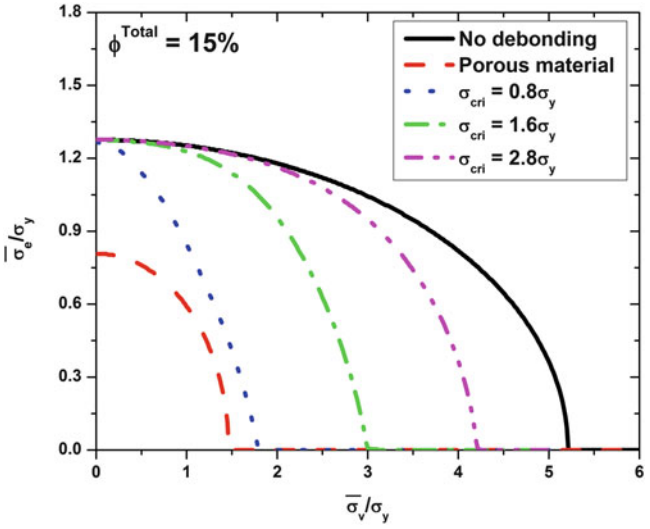


Fig. 8 Effects of bonding strengths on the overall yield surfaces for PRMMCs

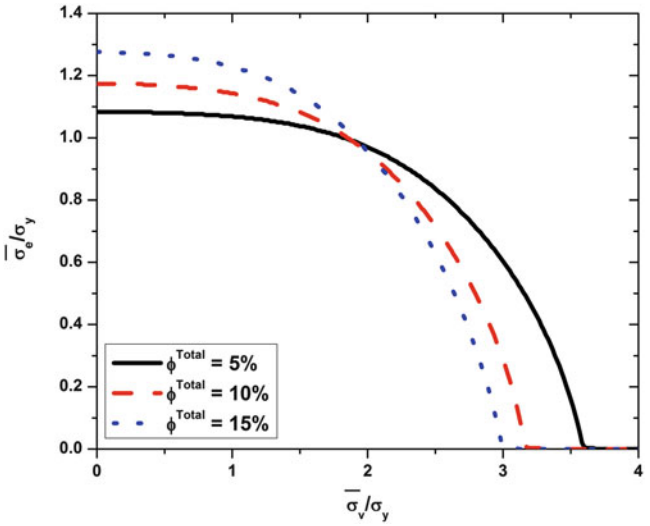


Fig. 9 Effects of volume fractions on the overall yield surfaces for PRMMCs

a decrease in the normalized effective yield stress. This unique feature clearly illustrates that the non-von Mises type of composite yielding is mainly caused by the existence of particles. As the volume fraction of particles reduces to zero, the matrix-only material recovers the von Mises plastic yielding, as expected.

Conclusions

Emanating from the eigenstrain concept of micromechanics and homogenization, the ensemble-averaged elastoplastic-damage constitutive equations are derived for PRMMCs with progressive partial interfacial debonding evolution. The debonding areas are represented by the corresponding debonding angles. Four different debonding modes are considered and the corresponding equivalent orthotropic stiffness tensors are constructed systematically. The proposed formulation is subsequently applied to the uniaxial loading conditions to illustrate the potential capabilities of the present framework. Comparisons between the model predictions and the available experimental data are also conducted. It is noted that the proposed elastoplastic-damage formulation is amenable to generalization to handle multiaxial loading conditions. Further research is warranted to extend the current framework to accommodate the spheroidal and other particle shapes. In addition to the stress-based particle-debonding criterion, the strain-based, the energy-based, or the mixed matrix–particle interfacial debonding criteria can be taken into consideration for different composite materials.

References

T.W. Clyne, P.J. Withers, *An Introduction to Metal Matrix Composites* (Cambridge University Press, Cambridge, 1993)

J.D. Eshelby, The determination of the elastic field of an ellipsoidal inclusion and related problems. Proc. R. Soc. Lond. **A241**, 376–396 (1957)

J.D. Eshelby, The elastic field outside an ellipsoidal inclusion. Proc. R. Soc. Lond. **A252**, 561–569 (1959)

I. Jasiuk, Y. Tong, The effect of interface on the elastic stiffness of composites, in *Mechanics of Composite Materials and Structures. Proceedings of the 3rd Joint ASCE/ASME Conference*, ed. by J.N. Reddy, J.L. Teply, American Society of Mechanical Engineers, New York, New York. 1989, pp. 49–54

J.W. Ju, T.M. Chen, Micromechanics and effective moduli of elastic composites containing randomly dispersed ellipsoidal inhomogeneities. Acta Mech. **103**, 103–121 (1994)

J.W. Ju, L.Z. Sun, A novel formulation for exterior-point Eshelby's tensor of an ellipsoidal inclusion. J. Appl. Mech. **66**, 570–574 (1999)

J.W. Ju, L.Z. Sun, Effective elastoplastic behavior of metal matrix composites containing randomly located aligned spheroidal inhomogeneities. Part I: micromechanics-based formulation. Int. J. Solids Struct. **38**, 183–201 (2001)

H.T. Liu, L.Z. Sun, J.W. Ju, An interfacial debonding model for particle-reinforced composites. Int. J. Damage Mech. **14**, 163–185 (2004)

H.T. Liu, L.Z. Sun, J.W. Ju, Elastoplastic modeling of progressive interfacial debonding for particle-reinforced metal-matrix composites. Acta Mech. **181**, 1–17 (2006)

T. Mura, *Micromechanics of Defects in Solids*, 2nd edn. (Kluwer, Boston, 1987)

S. Nemat-Nasser, M. Hori, *Micromechanics: Overall Properties of Heterogeneous Materials*, 2nd edn. (North-Holland, Amsterdam, 1999)

N.J. Pagano, G.P. Tandon, Modeling of imperfect bonding in fiber reinforced brittle matrix composites. Mech. Mater. **9**, 49–64 (1990)

J.M. Papazian, P.N. Adler, Tensile properties of short fiber-reinforced SiC/Al composites. Part I: effect of matrix precipitates. Metall. Trans. **A21**, 401–410 (1990)

J. Qu, The effect of slightly weakened interfaces on the overall elastic properties of composite materials. Mech. Mater. **14**, 269–281 (1993)
A.S. Sangani, G. Mo, Elastic interactions in particulate composites with perfect as well as imperfect interfaces. J. Mech. Phys. Solids **45**, 2001–2031 (1997)
L.Z. Sun, J.W. Ju, H.T. Liu, Elastoplastic modeling of metal matrix composites with evolutionary particle debonding. Mech. Mater. **35**, 559–569 (2003)
S. Suresh, A. Mortensen, A. Needleman, *Fundamentals of Metal-Matrix Composites* (Butterworth-Heinemann, Boston, 1993)
K. Tohgo, G.J. Weng, A progressive damage mechanics in particle-reinforced metal-matrix composites under high triaxial tension. J. Eng. Mater-T ASME **116**, 414–420 (1994)
F.C. Wong, A. Ait-Kadi, Analysis of particulate composite behavior based on non-linear elasticity and modulus degradation theory. J. Mater. Sci. **32**, 5019–5034 (1997)
Y.H. Zhao, G.J. Weng, Transversely isotropic moduli of two partially debonded composites. Int. J. Solids Struct. **34**, 493–507 (1997)
Y.H. Zhao, G.J. Weng, The effect of debonding angle on the reduction of effective moduli of particle and fiber-reinforced composites. J. Appl. Mech. **69**, 292–302 (2002)

Section VIII

Damage Under Dynamic Loading

Anisotropic Damage for Extreme Dynamics 35

Tomasz Łodygowski and Wojciech Sumelka

Contents

Abstract

Within this document, recent achievements of the Perzyna-type viscoplasticity theory for metallic materials subjected to extreme loading are considered. The complexity of the subject of matter lays in fact that for robust modelling the number of important phenomena included in the description is high. In consequence, the number of material parameters in the model is considerable. Thus,

T. Łodygowski (✉) • W. Sumelka
Division of Computer Aided Design, Institute of Structural Engineering, Poznan University of Technology, Poznań, Poland
e-mail: tomasz.lodygowski@put.poznan.pl; wojciech.sumelka@put.poznan.pl

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_32

very detailed experimental examination of a particular material, under vast range of strain rates, temperatures, and scales of observations, is needed for applications.

The important features of the presented theory, which assure reliability of the extreme dynamics modelling, in its qualitative and quantitative meaning, can be summarized as follows: (i) the description is invariant with respect to any diffeomorphism (covariant material model); (ii) the obtained evolution problem is well posed; (iii) sensitivity to the rate of deformation; (iv) finite elasto-viscoplastic deformations; (v) plastic non-normality; (vi) dissipation effects (anisotropic description of damage); (vii) thermomechanical couplings; and (viii) length-scale sensitivity are included.

Introduction

Since its invention by Perzyna in 1963 (cf. classical paper (Perzyna 1963)), viscoplasticity theory has come long way and nowadays is one of the most often choice when dealing with rate-dependent materials. Although Perzyna model is commonly associated with famous definition of rate of viscoplastic strains in terms of overstress function only, its present form belongs to most general and elegant formulations in mechanics (Perzyna 2005, 2008) and includes (i) invariance with respect to any diffeomorphism (covariant material model), (ii) well-posedness of evolution problem, (iii) sensitivity to the rate of deformation, (iv) finite elasto-viscoplastic deformations, (v) plastic non-normality, (vi) dissipation effects (anisotropic description of damage), (vii) thermomechanical couplings, and (viii) length-scale sensitivity. On the other hand, it is important that the viscoplasticity theory, being a physical one, has a deep physical interpretation derived from the analysis of a single crystal and a polycrystal behavior.

Recently, the Perzyna theory was extended to include the effects of anisotropy of microdamage (Perzyna 2008; Glema et al. 2009; Sumelka 2009) as a necessary condition for proper/robust description of damage. It generalizes previously used scalar damage parameter (Dornowski 1999; Dornowski and Perzyna 2002a, b, c). The introduced anisotropy allows to distinguish two levels of damage approximation: global and local (Sumelka and Łodygowski 2011). Good *global damage approximation* (GDA) is obtained if (global) strain-stress curves from experiment and mathematical model are close to each other. Good *local damage approximation* (LDA) is obtained if apart from the global, particularly good coincidence in macrodamage initiation time, velocity of macrodamage evolution, and the geometry of macrodamage pattern is observed. Under continuum damage mechanics, it can be shown that the first class of problems (GDA) can be covered by scalar damage models, while for the second class of problems (LDA) one needs higher-order tensor in the model describing damage (Glema et al. 2010a; Łodygowski et al. 2012).

It should be emphasized that the problem of modelling extreme dynamic events for metallic materials including strain rates over $10^7\ s^{-1}$ and temperatures reaching

melting point is still open in theoretical, applied, and computational mechanics. The embarrassing situation is found that the experiment cannot still give a detailed/unique answer for extremely fast thermomechanical processes about evolution of crucial phenomena (e.g., strain or temperature evolution). Hence, a situation is faced in which the theory must extrapolate/forecast experimental results. The reason is that the thermomechanical processes are highly influenced by elasto-viscoplastic wave effects (their propagation and interaction (Glema et al. 2004; Łodygowski and Sumelka 2012)) and varying initial anisotropy caused by existing defects in metals structure like microcracks, microvoids, and mobile and immobile dislocations densities being together a cause of overall induced anisotropy during deformation.

The discussed material model for metallic materials is now well established in literature (Perzyna 2008; Glema et al. 2009; Sumelka 2009; Glema et al. 2010a, b; Sumelka and Łodygowski 2011, 2013; Łodygowski et al. 2012; Sumelka 2013; Sumelka and Łodygowski 2013). Nevertheless, within this document, apart classical part needed to understand the overall concept, the attention is focused on the evolution of fracture porosity (Cochran and Banner 1988; Meyers and Aimone 1983). This parameter controls the level of load carrying capacity by the material in the model. Such concept was suggested first in the viscoplasticity theory in (Sumelka 2009) and herein will be exemplified by full spatial modelling of spalling phenomena. Thus, the model can approximate additional experimental observation that under certain dynamic conditions, damage in metallic materials changes from ductile (for lower rates of deformation) to almost brittle (for extreme rates of deformation).

The remaining part of this chapter is organized as follows.

In section "Experimental Motivation," the experimental observations concerning metals anisotropy, evolution of fracture porosity, and their influence on behavior of metallic materials under dynamic conditions are discussed.

Section "Mathematical Modelling" governs the fundamental results of Perzyna-type viscoplasticity accounting for anisotropic damage description and identification of assumed material functions and parameters (Sumelka 2009).

In section "Material Model Identification Adiabatic Process," computer implementation of the model in Abaqus/Explicit program utilizing VUMAT subroutine is explained.

Finally, in section "Model Validation Numerical Examples," verification of the model based on numerical results of spalling phenomena including evolving fracture porosity is presented.

Section "Conclusions" concludes the document.

Experimental Motivation

The attention is focused on experimental observations of metals behaviors concerning especially microdamages anisotropy in metals being a source of the overall metal anisotropy. Other sources of anisotropy like different sizes and shapes

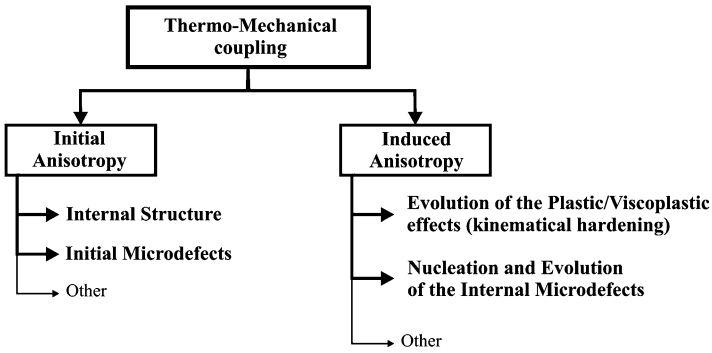


Fig. 1 Anisotropy in thermomechanics of metals

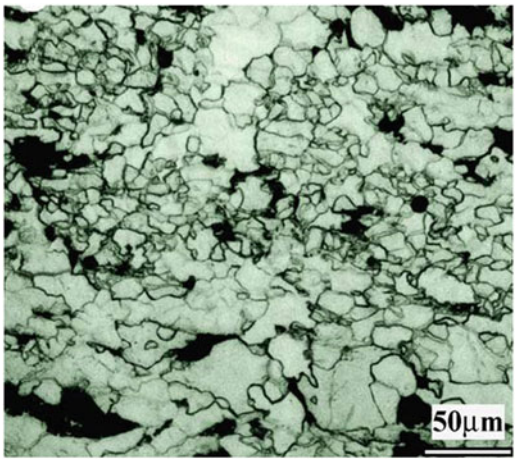


Fig. 2 The anisotropy of the HSLA-65 steel microstructure (Narayanasamy et al. 2009)

of adjacent grains (Narayanasamy et al. 2009) and presence of different phases like pearlite or ferrite (Pęcherski et al. 2009) are not discussed (cf. Fig. 1). The discussed experimental evidence considers also rate dependence of fracture porosity, being the measure of load carrying capacity by the material in the model.

On purpose the description of well recognized and described in literature phenomena, like e.g., kinematical hardening, rate sensitivity, length-scale sensitivity, or plastic non-normality (they are included into the model of course), is omitted. For detailed information, please see the review reports, e.g., Łodygowski (1996), Perzyna (1998, 2005), Glema et al. (2004).

So, the always existing defects in metals structure like microcracks, microvoids, and mobile and immobile dislocations densities (Abu Al-Rub and Voyiadjis 2006; Voyiadjis and Abu Al-Rub 2006; cf. Fig. 2) cause anisotropy of metals. It is then clear that for a proper mathematical modelling of metal behavior, one should include this type of anisotropy into the formal description. The frequently used isotropic simplification for metals should be thought of as a first approximation

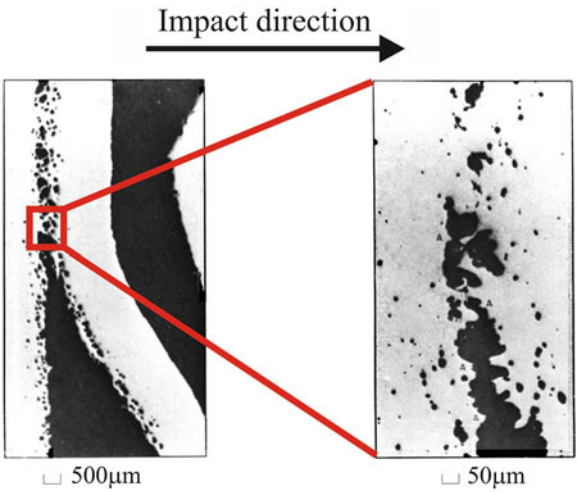


Fig. 3 Cracks anisotropy in 1145 aluminum after flat plate impact experiment (Seaman et al. 1976)

which carries not enough information for modern applications (Glema et al. 2010b) (though it certainly does not disavow such an approach in many applications cf. Klepaczko et al. (2007), Rusinek and Klepaczko (2009)).

Coming back to the experimental results for metallic materials, being the crux of the matter of this section, the following three statements are proposed (Sumelka 2009):

(i) Intrinsic microdefects are anisotropic.
(ii) Evolution of microdamages is directional.
(iii) Damage changes from ductile to almost brittle under dynamic loading.

Statement (i) confirms the experimental results that metal anisotropy, caused by the intrinsic defects, comes not only from its existence but especially from its inhomogeneous structure.

As an example, the effects of flat plate impact experiment in 1145 aluminum are considered (Seaman et al. 1976). The separation observed is preceded by the evolution of microdamages (microvoids), consisting for undamaged material of three stages: nucleation, growth, and coalescence. Notice in Fig. 3 that all of the microdefects are elongated perpendicularly to impact direction, thus to maximal tensile stresses. In this experiment, they have approximately an ellipsoidal shape. So, intrinsic defects have directional nature. Their anisotropy influences the whole deformation process, having a considerable impact on it.

Statement (ii) expresses explicitly experimental fact that the anisotropic properties of the continuum body evolve directionally during the deformation process (cf. experimental results presented in Grebe et al. 1985). Notice that it is a consequence of structure rearrangement itself but especially by directional evolution of intrinsic defects. As an example in Fig. 4, the evolution of microvoids in the region of forming shear band is presented. It is clearly seen that the evolution is

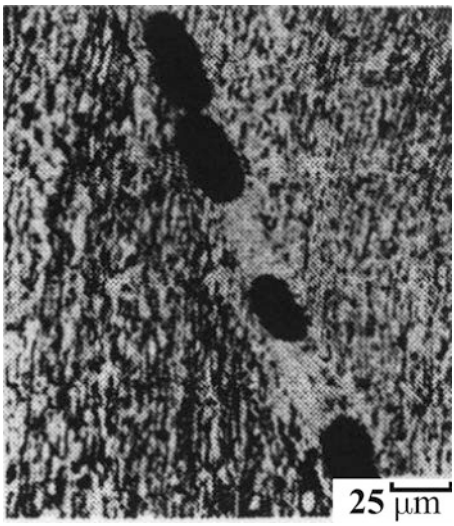


Fig. 4 Anisotropic microcracks in the shear band region in Ti-6 pct Al-4 pct V alloy (After Grebe et al. 1985)

directional; microvoids are being elongated through the shear band. So, the existing or nucleating microdamages growth is directional according to the imposed deformation process, inducing the anisotropic evolution of material properties.

Statement (iii) shows the experimental results that damage in metallic materials changes from ductile (for lower rates of deformation) to almost brittle (for extreme rates of deformation). In Fig. 5 the evolution of the ratio of the spall pick to the shock peak (denoted by R) in uranium versus flayer plate velocity during flat plate impact experiment (Cochran and Banner 1988) is presented. It is seen that when velocity of flayer plate increases, the ratio $\boldsymbol{R}$ increases. It confirms that by increasing rate of deformation, there is no time for microdamage growth. Equivalently, for extreme loading, once damage nucleates, it causes brittle cracking, while for lowest velocities after nucleation there is a possibility for damage growth, giving ductile damage.

As a concluding remark of this section, recall that the microdamage evolution mechanism in metals generally has three stages: defects nucleation, their growth and coalescence, and the evolution limits of microdamage that depend on rate of deformation. These observations will be transferred into the model utilizing the concept of microdamage tensor.

Mathematical Modelling

The Concept of Microdamage Tensor

The microdamage tensor plays an important role in the presented formulation. Notice that although the prefix “micro” is used, it should not be confused with microscale observations. Namely, the continuum description is used, so the

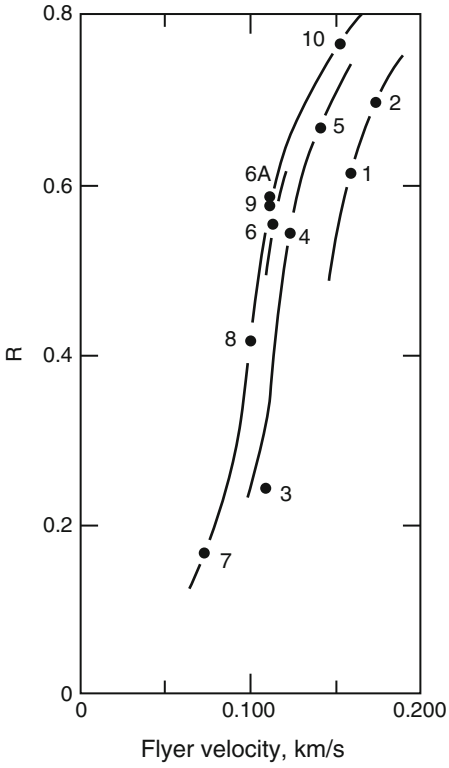


Fig. 5 The evolution of R in uranium versus flayer plate velocity during flat plate impact experiment (After Cochran and Banner 1988)

Table 1 Representation of materials defects

Type of defects	Mathematical description	Science
Microcracks, microvoids	Stochastical distributions, overall variables, and fractal mechanics	Micromechanics
Microcracks, microvoids	**Damage variables**	**Continuum damage mechanics**
Macrocracks	Embedded displacement discontinuities	Fracture mechanics

observation scale between macro- and mesoscales ($10^{-4} \div 10^{-3}$ m) is considered, but the nomenclature from microscale "transferred" (Longere et al. 2005). An analogous problem was discussed in Tikhomirov et al. (2001) according to the notion *microcracks* (cf. Table 1).

The introduced concept of microdamage tensor has clear physical interpretation giving a straight answer concerning how to conduct real measurements. To understand the whole concept, one should follow the results discussed in Sumelka (2009) and Glema et al. (2010a).

The Physical Interpretation of Microdamage Tensor

Suppose that for the selected points P_i in the material body $\mathcal{B}$, on three perpendicular planes, the ratio of the damaged area to the assumed characteristic area of the representative volume element (RVE) can be measured, i.e.,

$$\frac{A_i^p}{A}, \tag{1}$$

where A_i^p is a damaged area and A denotes assumed characteristic area of the RVE – cf. Fig. 6. Based on the calculated ratios ($\frac{A_i^p}{A}$), three vectors are obtained. Theirs modules are equal to those ratios and are normal to RVE's planes (see Fig. 6).

Measurements presented above can be repeated in any different configuration by rotating these three planes about point O – Fig. 6. So, for every measurement configuration, from those three vectors, the resultant is composed. Afterward, the configuration is chosen, in which the resultant module is largest. Such resultant is called the *main microdamage vector* and is denoted by $\hat{\boldsymbol{\xi}}^{(m)}$, i.e.,

$$\hat{\boldsymbol{\xi}}^{(m)} = \frac{A_1^p}{A}\hat{\mathbf{e}}_1 + \frac{A_2^p}{A}\hat{\mathbf{e}}_2 + \frac{A_3^p}{A}\hat{\mathbf{e}}_3, \tag{2}$$

where $\hat{(\cdot)}$ denotes the principal directions of microdamage with $A_1^p \geq A_2^p \geq A_3^p$.

In the following step, based on the main microdamage vector, the *microdamage vector* is built, denoted by $\hat{\boldsymbol{\xi}}^{(n)}$ (Sumelka et al. 2007)

$$\hat{\boldsymbol{\xi}}^{(n)} = \frac{1}{\left\|\hat{\boldsymbol{\xi}}^{(n)}\right\|}\left(\left(\frac{A_1^p}{A}\right)^2\hat{\mathbf{e}}_1 + \left(\frac{A_2^p}{A}\right)^2\hat{\mathbf{e}}_2 + \left(\frac{A_3^p}{A}\right)^2\hat{\mathbf{e}}_3\right). \tag{3}$$

Finally, the existence of microdamage tensorial field $\boldsymbol{\xi}$ is postulated

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_{11} & \xi_{12} & \xi_{13} \\ \xi_{21} & \xi_{22} & \xi_{23} \\ \xi_{31} & \xi_{32} & \xi_{33} \end{bmatrix}, \tag{4}$$

which is defined in its principal directions by applying the formula combining the microdamage vector and microdamage tensor (Sumelka et al. 2007)

$$\hat{\boldsymbol{\xi}}^{(n)} = \hat{\boldsymbol{\xi}}_{\mathbf{n}}, \tag{5}$$

where

$$\mathbf{n} = \sqrt{3}\left\|\hat{\xi}^{(m)}\right\|^{-1}\left(\hat{\xi}_1^{(m)}\hat{\mathbf{e}}_1 + \hat{\xi}_2^{(m)}\hat{\mathbf{e}}_2 + \hat{\xi}_3^{(m)}\hat{\mathbf{e}}_3\right), \tag{6}$$

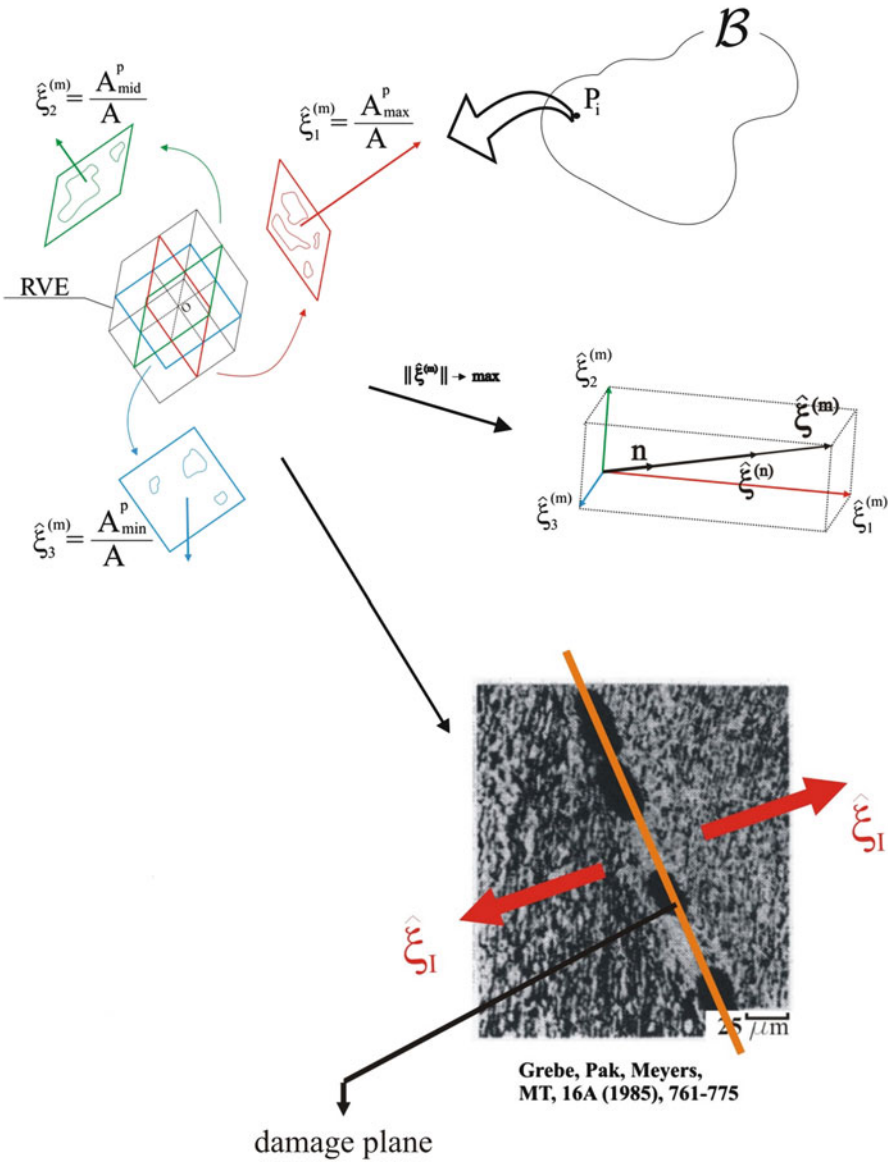


Fig. 6 The concept of microdamage tensor

leading to the fundamental result

$$\hat{\boldsymbol{\xi}} = \frac{\sqrt{3}}{3}\begin{bmatrix} \hat{\xi}_1^{(m)} & 0 & 0 \\ 0 & \hat{\xi}_2^{(m)} & 0 \\ 0 & 0 & \hat{\xi}_3^{(m)} \end{bmatrix}. \tag{7}$$

Hence, the following physical interpretation of the microdamage tensor components is obtained: *The diagonal components ξ_{ii} of the microdamage tensor $\boldsymbol{\xi}$, in its principal directions, are proportional to the components of the main microdamage vector $\xi_i^{(m)}$ which defines the ratio of the damaged area to the assumed characteristic area of the RVE, on the plane perpendicular to the i direction.*

Such interpretation clearly states that the damage plane is the one perpendicular to the maximal principal value of $\boldsymbol{\xi}$. Simultaneously, it gives a tool for a graphical presentation of the anisotropy evolution during post processing of the numerical results. Namely, by tracing the principal directions of $\boldsymbol{\xi}$, the softening directions are traced, and macrodamage path(s) can be predicted (Glema et al. 2010a, b).

Moreover, taking the Euclidean norm from the microdamage field $\hat{\boldsymbol{\xi}}$, the following relation is obtained

$$\sqrt{\boldsymbol{\xi} : \boldsymbol{\xi}} = \frac{\sqrt{3}}{3}\sqrt{\left(\frac{A_1^p}{A}\right)^2 + \left(\frac{A_2^p}{A}\right)^2 + \left(\frac{A_3^p}{A}\right)^2}. \tag{8}$$

Now, assuming that the characteristic length of RVE cube is l, one can rewrite Eq. 8 as

$$\sqrt{\boldsymbol{\xi} : \boldsymbol{\xi}} = \frac{\frac{\sqrt{3}\,l}{3}\sqrt{\left(\frac{A_1^p}{A}\right)^2 + \left(\frac{A_2^p}{A}\right)^2 + \left(\frac{A_3^p}{A}\right)^2}}{l^3}. \tag{9}$$

Equation 8 results in additional physical interpretation for microdamage tensor appears, namely, the Euclidean norm of the microdamage field defines the scalar quantity called the *volume fraction porosity* or simply *porosity* (Perzyna 2008):

$$\sqrt{\boldsymbol{\xi} : \boldsymbol{\xi}} = \frac{V - V_s}{V} = \frac{V_p}{V}, \tag{10}$$

where ξ denotes porosity (scalar damage parameter), V is the volume of a material element, V_s is the volume of the solid constituent of that material element, and V_p denotes void volume:

$$V_p = \frac{\sqrt{3}l}{3}\sqrt{\left(A_1^p\right)^2 + \left(A_2^p\right)^2 + \left(A_3^p\right)^2}. \tag{11}$$

The Limit Values of Microdamage Evolution

The interpretations of the microdamage tensorial field impose the mathematical bounds for a microdamage evolution, as

$$\xi \in < 0, 1 >, \quad \text{and} \quad \hat{\xi}_{ii} \in < 0, 1 >. \tag{12}$$

However, the physical bounds are different and are rate dependent (Cochran and Banner 1988; Meyers and Aimone 1983). This fact is crucial for further numerical analyses.

There is unwavering experimental evidence showing the existence of the initial porosity in metals (denoted by ξ_0) which is of order $\xi_0 \cong 10^{-4} \div 10^{-3}$ (Nemes and Eftis 1991). That porosity, however, cannot reach the theoretical full saturation, i.e., $\xi = 1$, during the deformation process and more is rate dependent causing that in limit even its close to zero value can cause fracture (transition from ductile to brittle fracture mode). Real maximal fracture porosity in metals depends on the tested material and the process characteristics (strain rates) and under extreme loading is of the order $0.09 \div 0.35$ (Dornowski and Perzyna 2002a, 2006).

It should be mentioned that it is common that the only experimental information regarding the initial microdamage tensorial field is its norm (porosity). Thus, particular numerical calculations must be preceded by assumption existing microdamage directions and its spatial distribution (keeping its norm of course). The meaningful role of such assumption in the deformation process was discussed in Sumelka and Łodygowski (2011). It was proved that different initial microdamage states (note that the porosity must increase its initial value by several orders to reach the fracture porosity) influence, in particular LDA, macrodamage initiation time, the velocity of macrodamage evolution, and the geometry of macrodamage, but have limited influence on GDA (it is in agreement with experiment, e.g., for tensile test of a particular material strain-stress curves are reproducible, but damage patterns differ in details).

Theory of Perzyna-Type Viscoplasticity Accounting for Anisotropic Damage

The description starts with analysis of the kinematics of the body, then fundamental constitutive axioms are postulated, and finally by fulfilling standard set of balance principles, constitutive model is obtained. Notice that more general

description including anisotropic elastic range and in consequence anisotropic yield criterion based on the extension of Lord Kelvin's formulation (Thomson 1856) was discussed in Sumelka (2009).

Kinematics

Two descriptions of the material body motion are considered, namely, Lagrangian (material, referential) and Eulerian (spatial, current). Those descriptions span two manifolds denoted by $\mathcal{B}$ and $\mathcal{S}$ respectively (Marsden and Hughes 1983).

Points in $\mathcal{B}$ are denoted by $\mathbf{X}$ while in $\mathcal{S}$ by $\mathbf{x}$. Coordinate system for $\mathcal{B}$ is denoted by $\{X^A\}$ with base $\mathbf{E}_A$ and for $\mathcal{S}$ by $\{x^a\}$ with base $\mathbf{e}_a$. Dual bases in those coordinate systems are denoted by $\mathbf{E}^A$ and $\mathbf{e}^a$, respectively.

The tangent spaces in $\mathcal{B}$ and $\mathcal{S}$ are written as $T_{\mathbf{X}}\mathcal{B} = \{\mathbf{X}\} \times V^3$ and $T_{\mathbf{x}}\mathcal{S} = \{\mathbf{x}\} \times V^3$. It is understood as Euclidean vector space V^3, regarded as vectors emanating from points $\mathbf{X}$ and $\mathbf{x}$, respectively (Marsden and Hughes 1983).

Taking Riemannian space on manifolds $\mathcal{B}$ and $\mathcal{S}$, i.e., $\{\mathcal{B}, \mathbf{G}\}$ and $\{\mathcal{S}, \mathbf{g}\}$, the metric tensors are defined, namely, $\mathbf{G} : T\mathcal{B} \rightarrow T^*\mathcal{B}$ and $\mathbf{g} : T\mathcal{B} \rightarrow T^*\mathcal{S}$, where $T\mathcal{B}$ and $T\mathcal{S}$ denotes the tangent bundles of $\mathcal{B}$ and $\mathcal{S}$, respectively, while $T^*\mathcal{B}$ and $T^*\mathcal{S}$ denotes their dual tangent bundles. Explicit definitions for metric tensors are then $G_{AB}(\mathbf{X}) = (\mathbf{E}_A, \mathbf{E}_B)_{\mathrm{x}}$ and $g_{ab}(\mathbf{x}) = (\mathbf{e}_a, \mathbf{e}_b)_{\mathrm{x}}$ where $(,)_{\mathbf{X}}$ and $(,)_{\mathrm{x}}$ denote inner product in $\mathcal{B}$ and $\mathcal{S}$, respectively.

The regular motion of the material body, treated as a series of the immersing of the abstract body $\mathcal{B}$ in the Euclidean point space E^3 (Rymarz 1993), can be written as

$$\mathbf{x} = \phi(\boldsymbol{X}, t), \tag{13}$$

thus ϕ_t: $\mathcal{B} \rightarrow \mathcal{S}$ is a $\boldsymbol{C^1}$ actual configuration of $\mathcal{B}$ in $\mathcal{S}$, at time t. The tangent of ϕ defines the two-point tensor field $\mathbf{F}$, called deformation gradient, which describes all local deformation properties and is the primary measure of deformation (Perzyna 1978; Holzapfel 2000). Thus,

$$\boldsymbol{F}(\boldsymbol{X}, t) = T\phi = \frac{\partial \phi(\boldsymbol{X}, t)}{\partial \mathbf{X}}, \tag{14}$$

and using the notion of tangent space

$$\boldsymbol{F}(\boldsymbol{X}, t) = T_{\mathbf{X}}\mathcal{B} \rightarrow T_{\mathbf{x}=\phi(\mathbf{X}, t)}\mathcal{S}, \tag{15}$$

so $\mathbf{F}$ is a linear transformation for each $\mathbf{X} \in \mathcal{B}$ and $t \in \boldsymbol{I} \subset \mathbb{R}^1$.

The map ϕ is assumed to be uniquely invertible (smooth homeomorphism) ($\mathbf{X} = \phi^{-1}(\mathbf{x}, t)$); hence, there exists the inverse of deformation gradient

$$\mathbf{F}^{-1}(\mathbf{x}, t) = \frac{\partial \phi^{-1}(\mathbf{x}, t)}{\partial \mathbf{x}}, \tag{16}$$

and the tensor field $\mathbf{F}$ is nonsingular ($\det(\mathbf{F}) \neq 0$), and because of the impenetrability of matter $\det(\mathbf{F}) > 0$. The following important decomposition, called polar decomposition, holds

$$\mathbf{F} = \mathbf{RU} = \mathbf{vR}, \tag{17}$$

where the $\mathbf{R}$ is rotation tensor (unique, proper orthogonal) which measures local orientation and U and v define unique, positive define, symmetric tensors called the right (or material) stretch tensor and the left (or spatial) stretch tensor, respectively (stretch tensors measure the local shape). Using the notion of tangent space, the result is obtained that for each $\mathbf{X} \in \mathcal{B}$, $\mathbf{U}(\mathbf{X})$: $T_{\mathbf{X}}\mathcal{B} \rightarrow T_{\mathbf{X}}\mathcal{B}$, and for each $\mathbf{x} \in \mathcal{S}$, $\mathbf{v}(\mathbf{x})$: $T_{\mathrm{x}}\mathcal{S} \rightarrow T_{\mathrm{x}}\mathcal{S}$. Thus, the local motion characterized by F can be decomposed into pure stretch and pure rotation.

The general class of Lagrangian and Eulerian strain measures can be defined through one single scale function given by (cf. Hill 1978; Xiao et al. 1998),

$$\mathbf{E} = g(\boldsymbol{C}) = \sum_{i=1}^{3} g(\breve{\chi}_i)\boldsymbol{C}_i,$$

and

$$\mathbf{e} = g(\mathbf{B}) = \sum_{i=1}^{3} g(\breve{\chi}_i)\mathbf{B}_i,$$

where the scale function $g(\cdot)$ is a smooth increasing function with the normalized property $g(1) = g'(1) - 1 = 0$; $\breve{\chi}_i$ are used to denote distinct eigenvalues of the right and left Cauchy-Green tensors $\mathbf{C}$ and $\mathbf{B}$, respectively; and $\mathbf{C}_i$ and $\mathbf{B}_i$ are the corresponding subordinate eigenprojections.

The Green-Lagrange strain tensor has been chosen (Perzyna 2005) (E: $T_{\mathrm{X}}\mathcal{B} \rightarrow T_{\mathrm{X}}\mathcal{B}$):

$$2\mathbf{E} = \mathbf{C} - \mathbf{I}, \tag{18}$$

where $\mathbf{E}$ stands for the Green-Lagrange strain tensor, $\mathbf{I}$ denotes the identity on $T_{\mathrm{X}}\mathcal{B}$, and

$$\mathbf{C} = \mathbf{F}^T \cdot \mathbf{F} = \mathbf{U}^2 = \mathbf{B}^{-1}. \tag{19}$$

By analogy for the spatial (Eulerian) strain measure, the Euler-Almansi strain tensor has been accepted (e: $T_{\mathrm{x}}\mathcal{S} \rightarrow T_{\mathrm{x}}\mathcal{S}$):

$$2\mathbf{e} = \mathbf{i} - \mathbf{c}, \tag{20}$$

where **e** stands for the Euler-Almansi strain tensor and i denotes the identity on $T_x\mathcal{S}$. One has also

$$\mathbf{c} = \mathbf{b}^{-1} \quad \text{and} \quad \mathbf{b} = \mathbf{F} \cdot \mathbf{F}^T = \mathrm{v}^2, \tag{21}$$

where tensor b is sometimes referred to as the Finger deformation tensor.

Using pushforward and pullback operation, the following relations are obtained:

$$\mathbf{e}^\flat = \phi_*\left(\mathbf{E}^\flat\right) = \mathbf{F}^{-T}\mathbf{E}^\flat\mathbf{F}^{-1}, \tag{22}$$

and

$$\mathbf{E}^\flat = \phi^*\left(\mathbf{e}^\flat\right) = \mathbf{F}^T\mathbf{e}^\flat\mathbf{F}, \tag{23}$$

where

$$\phi_*\left((\cdot)^\flat\right) = \mathbf{F}^{-T}(\cdot)^\flat\mathbf{F}^{-1} \tag{24}$$

stands for pushforward and

$$\phi^*\left((\cdot)^\flat\right) = \mathbf{F}^T(\cdot)^\flat\mathbf{F} \tag{25}$$

for pullback.

To describe the finite elasto-viscoplastic deformation, the multiplicative decomposition of the total deformation gradient has been accepted:

$$\mathbf{F}(\mathbf{X}, t) = \mathbf{F}^e(\mathbf{X}, t) \cdot \mathbf{F}^p(\mathbf{X}, t). \tag{26}$$

This decomposition justified by the micromechanics of single crystal plasticity (Perzyna 1998) states that the component $\mathbf{F}^e$ is a lattice contribution to **F**, while $\mathbf{F}^p$ describes the deformation solely due to plastic shearing on crystallographic slip systems.

The inverse of the local elastic deformation $\mathbf{F}^{e^{-1}}$ releases from the stress state in every surrounding ($\mathcal{N}(\mathbf{x}) \subset \phi(\mathcal{B})$) in an actual configuration. The configuration obtained by the linear map $\mathbf{F}^{e^{-1}}$ from actual configuration $\mathcal{S}$ is called unstressed configuration and is denoted by $\mathcal{S}'$. Thus, one can write (see Fig. 7)

$$\mathbf{F}^e : T_{\mathrm{y}}\mathcal{S}' \to T_{\mathrm{X}}\mathcal{S}, \quad \mathbf{F}^p : T_{\mathrm{X}} \to T_{\mathrm{y}}\mathcal{S}', \tag{27}$$

where the material point in the configuration $\mathcal{S}'$ is characterized by y.

The introduced decomposition of **F** enables to define the fundamental strain tensors in both formulations considered.

The viscoplastic strain tensor $\mathbf{E}^p : T_{\mathrm{X}}\mathcal{B} \to \boldsymbol{T}_{\mathbf{X}}\mathcal{B}$ can be written as

$$2\mathbf{E}^p = \mathbf{C}^p - \mathbf{I}, \tag{28}$$

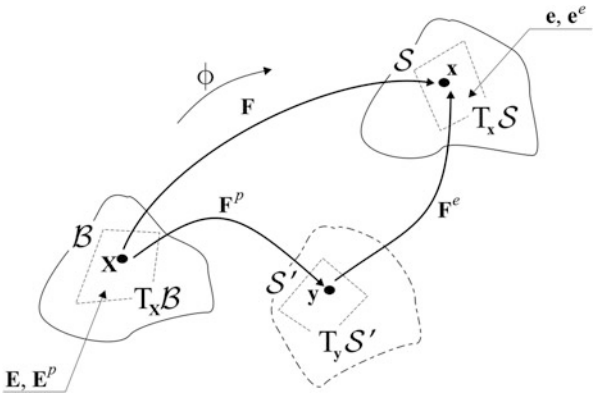


Fig. 7 The interpretation of the multiplicative decomposition of **F**

where

$$\mathbf{C}^p = \mathbf{F}^{pT} \cdot \mathbf{F}^p = \mathbf{U}^{p2} = \mathbf{B}^{-1} \quad \text{and} \quad \mathbf{E}^e = \mathbf{E} - \mathbf{E}^p, \tag{29}$$

while the elastic strain tensor $\mathbf{e}^e : T_x\mathcal{S} \rightarrow T_x\mathcal{S}$ is

$$2\mathbf{e}^e = \mathbf{i} - \mathbf{c}^e, \tag{30}$$

where

$$\mathbf{c}^e = \mathbf{b}^{e-1} \quad \text{and} \quad \mathbf{b}^e = \mathbf{F}^e \cdot \mathbf{F}^{eT} = \mathbf{v}^{e2} \quad \text{and} \quad \mathbf{e}^p = \mathbf{e} - \mathbf{e}^e. \tag{31}$$

At the end of material body kinematic description, it is fundamental for further discussion of the rate-type constitutive relations to define the rate of fields that describe the change of shape, position, and orientation of a continuum body.

Using motion defined by Eq. 13, the spatial velocity υ is simply

$$\boldsymbol{\upsilon}(\boldsymbol{x}, t) = \dot{\boldsymbol{x}} = \frac{\partial \boldsymbol{\phi}}{\partial t}. \tag{32}$$

Taking the gradient of $\boldsymbol{\upsilon}$, the tensor field (nonsymmetric, second order), called spatial velocity gradient, is obtained (Holzapfel 2000):

$$\mathbf{l}(\mathbf{x}, t) = \frac{\partial \boldsymbol{\upsilon}(\mathbf{x}, t)}{\partial \mathbf{x}}, \tag{33}$$

where $\mathbf{l}$ stands for spatial velocity gradient. Using the notion of deformation gradient and multiplicative decomposition of deformation gradient, the Eq. 33 is rewritten to the form (Perzyna 2005)

$$\mathbf{l} = \dot{\mathbf{F}} \cdot \mathbf{F}^{-1} = \dot{\mathbf{F}}^e \cdot \mathbf{F}^{e-1} + \mathbf{F}^e \cdot \left(\dot{\mathbf{F}}^p \cdot \mathbf{F}^{p-1}\right) \cdot \mathbf{F}^{e-1} = \mathbf{l}^e + \mathbf{l}^p, \tag{34}$$

which introduces the elastic $\mathbf{l}^e$ and plastic $\mathbf{l}^p$ parts of spatial velocity gradient. On the other hand, the additive decomposition of spatial velocity gradient to symmetric

and antisymmetric parts generates covariant tensor field d called rate of deformation tensor and also covariant tensor field w called spin tensor, with the definitions:

$$\mathbf{l} = \mathbf{d} + \mathbf{w} = \mathbf{d}^e + \mathbf{w}^e + \mathbf{d}^p + \mathbf{w}^p, \tag{35}$$

$$\mathbf{d} = \frac{1}{2}\left(\mathbf{l} + \mathbf{l}^T\right), \tag{36}$$

$$\mathrm{w} = \frac{1}{2}\left(\mathbf{l} - \mathbf{l}^T\right). \tag{37}$$

First the notion of objective Lie derivative (assuring diffeomorphisms) is introduced. The Lie derivative of an arbitrary spatial tensor field φ is obtained using the following concept:

(i) Compute the pullback operation of φ; the material field Φ is obtained.
(ii) Take the material time derivative of Φ.
(iii) Carry out the pushforward operation of the result field from (ii).
(iv) The scheme can be summarized as

$$\mathrm{L}_v(\varphi) = \phi_*\left(\frac{\mathrm{D}}{\mathrm{D}t}\phi^*(\varphi)\right), \tag{38}$$

where L_v stands for Lie derivative.

So, applying Lie derivative to strain measure, the following fundamental result is obtained:

$$\mathbf{d}^\flat = \mathrm{L}_v\left(\mathbf{e}^\flat\right). \tag{39}$$

Thus, Lie derivative states a direct relationship between the stretching d, which is a direct natural measure of the rate of length of any line element and the rate of change of the angle between any two intersecting line elements in a deforming body, and the Eulerian strain e which measures a change of the length of any line element and change of the angle between any two intersecting line elements.

By analogy to Eq. 39, one can write

$$\mathbf{d}^{e^\flat} = \mathrm{L}_v\left(\mathbf{e}^{e^\flat}\right), \quad \mathbf{d}^{p^\flat} = \mathrm{L}_v\left(\mathbf{e}^{p^\flat}\right). \tag{40}$$

Constitutive Postulates

Assuming that the balance principles hold, namely, conservation of mass, balance of momentum, balance of moment of momentum, and balance of energy and entropy production, four constitutive postulates are defined (Perzyna 1986b, 2005) (below e depends on formulation of course):

(i) Existence of the free energy function ψ. Thus, formally

$$\psi = \hat{\psi}(\boldsymbol{e}, \mathbf{F}, \vartheta; \boldsymbol{\mu}), \tag{41}$$

where $\boldsymbol{\mu}$ denotes a set of internal state variables governing the description of dissipation effects and ϑ represents temperature. It is important to notice that the semicolon used to separate the last variable due to its different nature (it introduces a dissipation to the model), without $\boldsymbol{\mu}$ the presented model, describes thermoelasticity.

(ii) Axiom of objectivity. The material model should be invariant with respect to diffeomorphism (any superposed motion).

(iii) The axiom of the entropy production. For every regular process, the constitutive functions should satisfy the second law of thermodynamics.

(iv) The evolution equation for the internal state variables vector μ should be of the form

$$\mathrm{L}_{v}\boldsymbol{\mu} = \hat{\mathbf{m}}(\boldsymbol{e}, \mathbf{F}, \vartheta; \boldsymbol{\mu}). \tag{42}$$

where evolution function $\hat{\mathrm{m}}$ has to be determined based on the experimental observations.

Constitutive Relations: General Form

The reduced dissipation inequality states that (Marsden and Hughes 1983; Sumelka 2009)

$$\frac{1}{\rho_{\mathrm{Ref}}}\boldsymbol{\tau} : \mathbf{d} - \left(\eta\dot{\vartheta} + \dot{\psi}\right) - \frac{1}{\rho\vartheta}\mathbf{q} \cdot \mathrm{grad}\vartheta \geq 0, \tag{43}$$

where ρ denotes actual and ρ_{Ref} reference densities, τ denotes Kirchhoff stress, ψ is the free energy function, ϑ is absolute temperature, η denotes the specific (per unit mass) entropy, and $\mathbf{q}$ is the heat flux. Using postulate (i), Eq. 43 can be rewritten as

$$\left(\frac{1}{\rho_{\mathrm{Ref}}}\tau - \frac{\partial\hat{\psi}}{\partial \mathrm{e}}\right) : \mathbf{d} - \left(\eta + \frac{\partial\hat{\psi}}{\partial\vartheta}\right)\dot{\vartheta} - \frac{\partial\hat{\psi}}{\partial\boldsymbol{\mu}}\mathrm{L}_{v}\boldsymbol{\mu} - \frac{1}{\rho\vartheta}\mathbf{q} \cdot \mathrm{grad}\vartheta \geq 0. \tag{44}$$

Because of arbitrariness,

$$\boldsymbol{\tau} = \rho_{\mathrm{Ref}}\frac{\partial\hat{\psi}}{\partial\mathbf{e}}, \tag{45}$$

$$\eta = -\frac{\partial\hat{\psi}}{\partial\vartheta}. \tag{46}$$

Hence, Eq. 44 reduces to

$$-\frac{\partial\hat{\psi}}{\partial\boldsymbol{\mu}}\mathrm{L}_{v}\mu - \frac{1}{\rho\vartheta}\mathbf{q} \cdot \mathrm{grad}\vartheta \geq 0. \tag{47}$$

Assuming now that internal state vector consists of two variables (discussed in details in next section), namely (Perzyna 2008; Glema et al. 2009; Sumelka and Łodygowski 2011),

$$\boldsymbol{\mu} = (\in^{p}, \boldsymbol{\xi}), \tag{48}$$

where $\in^p$ is the equivalent plastic deformation $\dot{\in}^p = \left(\frac{2}{3}\mathbf{d}^p : \mathbf{d}^p\right)^{\frac{1}{2}}$, which describes the dissipation effects generated by viscoplastic deformation, and $\boldsymbol{\xi}$ is microdamage tensor which takes into account the anisotropic microdamage effects one can write general form of rate-type constitutive equations for thermomechanical process under consideration.

Applying Lie derivative to formula Eq. 45, with internal state vector constant, or in other words keeping the history constant (thermoelastic process), the evolution equation for Kirchhoff stress tensor is obtained in the form (Duszek–Perzyna et al. 1994)

$$(\mathrm{L}_v\boldsymbol{\tau})^e = \mathcal{L}^e : \mathbf{d}^e - \mathcal{L}^{th}\dot{\vartheta}, \tag{49}$$

where

$$\mathcal{L}^e = \rho_{\mathrm{Ref}}\frac{\partial^2\hat{\psi}}{\partial\mathbf{e}^2}, \tag{50}$$

$$\mathcal{L}^{th} = -\rho_{\mathrm{Ref}}\frac{\partial^2\hat{\psi}}{\partial\mathbf{e}\partial\vartheta}, \tag{51}$$

in the above $\mathcal{L}^e$ denotes elastic constitutive tensor and $\mathcal{L}^{th}$ is thermal operator. Using the relation

$$(\mathrm{L}_{\mathbf{v}}\boldsymbol{\tau})^e = \dot{\boldsymbol{\tau}} - \boldsymbol{\tau}\cdot\mathbf{d}^e - \mathbf{d}^e\cdot\boldsymbol{\tau}, \tag{52}$$

$$\mathbf{d} = \mathbf{d}^e + \mathbf{d}^p, \tag{53}$$

the final form of rate of Kirchhoff stresses is

$$\mathrm{L}_v\boldsymbol{\tau} = \mathcal{L}^e : \mathbf{d} - \mathcal{L}^{th}\dot{\vartheta} - (\mathcal{L}^e + \mathbf{g}\boldsymbol{\tau} + \boldsymbol{\tau}\mathbf{g}) : \boldsymbol{d}^p. \tag{54}$$

Using the energy balance in the form (Perzyna 2005; Sumelka 2009)

$$\rho\vartheta\dot{\eta} = -\mathrm{div}\mathbf{q} - \rho\frac{\partial\hat{\psi}}{\partial\boldsymbol{\mu}}\cdot\mathrm{L}_v\boldsymbol{\mu} \tag{55}$$

and taking the rate of entropy, defined by the Eq. 46, the evolution equation for temperature is obtained:

$$\rho c_p\dot{\vartheta} = -\mathrm{div}\mathbf{q} + \vartheta\frac{\rho}{\rho_{\mathrm{Ref}}}\frac{\partial\boldsymbol{\tau}}{\partial\vartheta} : \mathbf{d} + \chi^*\boldsymbol{\tau} : \boldsymbol{d}^p + \chi^*\mathbf{k} : L_v\boldsymbol{\xi}, \tag{56}$$

where the specific heat

$$c_p = -\vartheta \frac{\partial^2 \hat{\psi}}{\partial \vartheta^2}, \tag{57}$$

and the irreversibility coefficients χ^* and χ^{**} are determined by (p defines viscoplastic flow direction – discussed in details in next section)

$$\begin{aligned} \chi^* &= -\rho \left(\frac{\partial \hat{\psi}}{\partial \in^p} - \vartheta \frac{\partial^2 \hat{\psi}}{\partial \vartheta \partial \in^p} \right) \sqrt{\frac{2}{3}} \frac{1}{\boldsymbol{\tau} : \mathbf{p}}, \\ \chi^{**} &= -\rho \left(\frac{\partial \hat{\psi}}{\partial \boldsymbol{\xi}} - \vartheta \frac{\partial^2 \hat{\psi}}{\partial \vartheta \partial \boldsymbol{\xi}} \right) : \frac{1}{\mathbf{k}}. \end{aligned} \tag{58}$$

Next one needs to specify explicit definitions for assumed internal state variables ($\in^p$, $\boldsymbol{\xi}$), define materials functions, and finally identify material parameters.

Material Model Identification Adiabatic Process

Evolution Equations for Internal State Variables

The evolution equations for internal state variables are postulated as follows:

$$\mathbf{d}^p = \Lambda \mathbf{p}, \tag{59}$$

$$\mathrm{L}_v \boldsymbol{\xi} = \Lambda^h \frac{\partial h*}{\partial \boldsymbol{\tau}} + \Lambda^g \frac{\partial g*}{\partial \boldsymbol{\tau}}, \tag{60}$$

where Λ, Λ^h, Λ^g define the intensity of viscoplastic flow, microdamage nucleation, and microdamage growth, while $\mathbf{p}$, $\frac{\partial h*}{\partial \boldsymbol{\tau}}$, $\frac{\partial g*}{\partial \boldsymbol{\tau}}$, define viscoplastic flow direction, microdamage nucleation direction, and microdamage growth direction, respectively. It should be emphasized that there exists fracture microdamage state (ξ^F) for which catastrophe takes place, namely,

$$\kappa = \hat{\kappa}(\in^p, \vartheta, \boldsymbol{\xi})|_{\xi = \xi^F} = 0. \tag{61}$$

Having defined evolution of $\in^p$ and $\boldsymbol{\xi}$, one can state the initial boundary value problem (IBVP) as follows.

Initial Boundary Value Problem

Find $\phi, \boldsymbol{v}, \rho, \boldsymbol{\tau}, \boldsymbol{\xi}, \vartheta$ as functions of t and position $\mathbf{x}$ such that the following equations are satisfied (Perzyna 1994; Łodygowski 1996; Łodygowski and Perzyna 1997a, b):

(i) The field equations

$$\begin{aligned}
\dot{\phi} &= \boldsymbol{v}, \\
\dot{\boldsymbol{v}} &= \frac{1}{\rho_{Ref}}\left(\mathrm{div}\boldsymbol{\tau} + \frac{\boldsymbol{\tau}}{\rho}\cdot \mathrm{grad}\rho - \frac{\boldsymbol{\tau}}{1-(\boldsymbol{\xi}:\boldsymbol{\xi})^{\frac{1}{2}}}\mathrm{grad}(\boldsymbol{\xi}:\boldsymbol{\xi})^{\frac{1}{2}}\right), \\
\dot{\rho} &= -\rho \mathrm{div}\boldsymbol{v} + \frac{\rho}{1-(\boldsymbol{\xi}:\boldsymbol{\xi})^{\frac{1}{2}}}(L_v\boldsymbol{\xi}:L_v\boldsymbol{\xi})^{\frac{1}{2}}, \\
\dot{\boldsymbol{\tau}} &= \mathcal{L}^e : \mathbf{d} + 2\boldsymbol{\tau}\cdot \mathrm{d} - \mathcal{L}^{th}\dot{\vartheta} - (\mathcal{L}^e + \boldsymbol{g}\boldsymbol{\tau} + \boldsymbol{\tau}\boldsymbol{g}) : \mathbf{d}^p, \\
\dot{\boldsymbol{\xi}} &= 2\boldsymbol{\xi}\cdot\mathbf{d} + \frac{\partial g*}{\partial \boldsymbol{\tau}}\frac{1}{T_m}\left\langle \Phi^g\left[\frac{I_g}{\tau_{eq}(\boldsymbol{\xi},\vartheta,\in^p)} - 1\right]\right\rangle, \\
\dot{\vartheta} &= \frac{\chi*}{\rho c_p}\boldsymbol{\tau}:\mathbf{d}^p + \frac{\chi**}{\rho c_p}\mathbf{k}:\mathrm{L}_v\boldsymbol{\xi}.
\end{aligned} \tag{62}$$

(ii) The boundary conditions
  (a) Displacement ϕ is prescribed on a part Γ_ϕ of $\Gamma(\mathcal{B})$, and tractions $(\boldsymbol{\tau}\cdot\mathbf{n})^a$ are prescribed on a part Γ_τ of $\Gamma(\mathcal{B})$, where $\Gamma_\phi \cap \Gamma_\tau = 0$ and $\Gamma_\phi \cup \Gamma_\tau = \Gamma(\mathcal{B})$.
  (b) Heat flux $\mathbf{q}\cdot\mathbf{n} = 0$ is prescribed on $\Gamma(\mathcal{B})$.

(iii) The initial conditions $\phi, \boldsymbol{v}, \rho, \boldsymbol{\tau}, \boldsymbol{\xi}, \vartheta$ are given for each particle $\mathbf{X} \in \mathcal{B}$ at $t = 0$ and are satisfied.

In Eq. $(62)_5$ it was assumed that in the evolution of microdamage, nucleation term is omitted; thus, appropriate initial microdamage state in computations must be assumed (cf. Sumelka and Łodygowski 2011 to detailed discussion of such consequences), and because of adiabatic regime assumption, first two terms in temperature evolution in Eq. 56 are omitted. It should be emphasized that adiabatic condition assumption weakens the robustness of modelling due to the fact that first term in Eq. 56 introduces in a natural way non-locality. Nevertheless, recall that in viscoplasticity non-locality comes implicitly from relaxation time parameter (T_m).

Material Functions

For the evolution problem Eq. 62, the following assumptions are stated.

The elastic range is assumed isotropic and independent from microdamage state; thus, elastic constitutive tensor $\mathcal{L}^e$ is (for more general setup cf. Sumelka 2009)

$$\mathcal{L}^e = 2\mu\boldsymbol{\mathcal{I}} + \lambda(\mathbf{g}\otimes\mathbf{g}), \tag{63}$$

where μ, λ are Lam*é* constants.

Similarity for thermal expansion effects, its isotropy is postulated; thus, thermal operator $\mathcal{L}^{th}$ is

$$\mathcal{L}^{th} = (2\mu + 3\lambda)\theta\mathbf{g}, \tag{64}$$

where θ is thermal expansion coefficient.

The rate of viscoplastic strains $\mathbf{d}^p$ is assumed as common in Perzyna's theory (Perzyna 1963, 1966), namely,

$$\mathbf{d}^p = \Lambda^{vp}\mathbf{p}. \tag{65}$$

In flow rule Eq. (65) Λ^{vp} defines flow intensity, thus

$$\Lambda^{vp} = \frac{1}{T_m}\left\langle \Phi^{vp}\left(\frac{f}{\kappa} - 1\right)\right\rangle = \frac{1}{T_m}\left\langle \left(\frac{f}{\kappa} - 1\right)^{m_{pl}}\right\rangle, \tag{66}$$

where yield surface f is considered in the form (Shima and Oyane 1976; Perzyna 1986a, b; Glema et al. 2009)

$$f = \left\{J_2' + \left[n_1(\vartheta) + n_2(\vartheta)(\boldsymbol{\xi} : \boldsymbol{\xi})^{\frac{1}{2}}\right]J_1^2\right\}^{\frac{1}{2}}, \tag{67}$$

with defining parameters dependent on temperature

$$n_1(\vartheta) = 0, \quad n_2(\vartheta) = n = \text{const.}, \tag{68}$$

and work hardening-softening function κ is postulated as (Perzyna 1986b; Nemes and Eftis 1993)

$$\kappa = \{\kappa_s(\vartheta) - [\kappa_s(\vartheta) - \kappa_0(\vartheta)]\exp[-\delta(\vartheta) \in^p]\}\left[1 - \left(\frac{(\boldsymbol{\xi} : \boldsymbol{\xi})^{\frac{1}{2}}}{\xi^F}\right)^{\beta(\vartheta)}\right], \tag{69}$$

$$\begin{aligned} \overline{\vartheta} = \frac{\vartheta - \vartheta_0}{\vartheta_0}, \quad \kappa_s(\vartheta) &= \kappa_s^* - \kappa_s^{**}\overline{\vartheta}, \quad \kappa_0(\vartheta) = \kappa_0^* - \kappa_0^{**}\overline{\vartheta}, \\ \delta(\vartheta) &= \delta^{**} - \delta^{**}\overline{\vartheta}, \quad \beta(\vartheta) = \beta^{**} - \beta^{**}\overline{\vartheta}. \end{aligned} \tag{70}$$

Rate dependence of fracture porosity ξ^F is postulated as (Sumelka 2009)

$$\xi^F = \xi^{F*} - \xi^{F**}\left\langle \left(\frac{\|\mathbf{L}_v\boldsymbol{\xi}\| - \|\mathbf{L}_v\boldsymbol{\xi}_c\|}{\|\mathbf{L}_v\boldsymbol{\xi}_c\|}\right)^{m_F}\right\rangle, \tag{71}$$

where ξ^{F*} can be thought as a quasi-static fracture porosity and $\|\mathrm{L}_v\boldsymbol{\xi}_c\|$ denotes equivalent critical velocity of microdamage. Notice that the definition of

Eq. 71 introduces the analogy to cumulative fracture criterion (Campbell 1953; Klepaczko 1990a), namely, by writing

$$t_c = \int_0^{t_d} \left\langle \frac{\|\mathbf{L}_v \boldsymbol{\xi}\|}{\|\mathbf{L}_v \boldsymbol{\xi}_c\|} - 1 \right\rangle dt, \tag{72}$$

one can say that there exists critical time t_c (possibly a function of thermodynamic process) needed to saturate the microdamage process up to its fracture limit t_d that stands for damage time.

Finally, the normalized directions of viscoplastic flow are defined as

$$\mathbf{p} = \left.\frac{\partial f}{\partial \boldsymbol{\tau}}\right|_{\boldsymbol{\xi}=\text{const}} \left(\left\| \frac{\partial f}{\partial \boldsymbol{\tau}} \right\| \right)^{-1} = \frac{1}{\left[2J_2' + 3A^2 (\text{tr}\boldsymbol{\tau})^2 \right]^{\frac{1}{2}}} [\boldsymbol{\tau}' + A\text{tr}\boldsymbol{\tau}\delta]. \tag{73}$$

Other denotes: $\boldsymbol{\tau}'$ represents the stress deviator and J_1, J_2' are the first and the second invariants of Kirchhoff stress tensor and deviatoric part of the Kirchhoff stress tensor, respectively: $A = 2\left(n_1 + n_2 (\boldsymbol{\xi} : \boldsymbol{\xi})^{\frac{1}{2}} \right)$.

For microdamage mechanism, it is assumed that growth term is considered only, while nucleation is replaced by initial microdamage distribution assumption. So, taking the additional assumptions (Dornowski 1999; Glema et al. 2009):

- Velocity of the microdamage growth is coaxial with the principal directions of stress state.
- Only positive (tension) principal stresses induce the growth of the microdamage.

One has

$$\frac{\partial g^*}{\partial \boldsymbol{\tau}} = \left\langle \frac{\partial \hat{g}}{\partial \boldsymbol{\tau}} \right\rangle \left\| \left\langle \frac{\partial \hat{g}}{\partial \boldsymbol{\tau}} \right\rangle \right\|^{-1}, \qquad \text{and} \qquad \hat{g} = \frac{1}{2} \boldsymbol{\tau} : \boldsymbol{\mathcal{G}} : \boldsymbol{\tau}, \tag{74}$$

$$\Phi^g \left(\frac{I_g}{\tau_{eq}(\boldsymbol{\xi}, \vartheta, \in^p)} - 1 \right) = \left(\frac{I_g}{\tau_{eq}} - 1 \right)^{m_g}, \tag{75}$$

where void growth threshold stress function τ_{eq} is in the form

$$\begin{aligned} \tau_{eq} &= c(\vartheta) \left(1 - (\boldsymbol{\xi} : \boldsymbol{\xi})^{\frac{1}{2}} \right) \ln \frac{1}{(\boldsymbol{\xi} : \boldsymbol{\xi})^{\frac{1}{2}}} \{ 2\kappa_s(\vartheta) - [\kappa_s(\vartheta) - \kappa_0(\vartheta)] F(\xi_0, \boldsymbol{\xi}, \vartheta) \}, \\ c(\vartheta) &= \text{const.}, \end{aligned} \tag{76}$$

$$F = \left(\frac{\xi_0}{1 - \xi_0} \frac{1 - (\boldsymbol{\xi} : \boldsymbol{\xi})^{\frac{1}{2}}}{(\boldsymbol{\xi} : \boldsymbol{\xi})^{\frac{1}{2}}} \right)^{\frac{2}{3}\delta} + \left(\frac{1 - (\boldsymbol{\xi} : \boldsymbol{\xi})^{\frac{1}{2}}}{1 - \xi_0} \right)^{\frac{2}{3}\delta}, \tag{77}$$

and

$$I_g = \overline{b}_1 J_1 + \overline{b}_2 \left(J_2'\right)^{\frac{1}{2}} + \overline{b}_3 \left(J_3'\right)^{\frac{1}{3}}, \tag{78}$$

$\overline{b}_i (i = 1, 2, 3)$ are the material parameters, and J_3' is the third invariant of deviatoric part of the Kirchhoff stress tensor.

Now, taking into account the postulates for microdamage evolution and assuming that tensor $\mathcal{G}$ can be written as a symmetric part of the fourth-order unity tensor $\mathcal{I}$ (Łodygowski et al. 2008)

$$\mathcal{G} = \mathcal{I}^s \quad \mathcal{G}_{ijkl} = \frac{1}{2}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right), \tag{79}$$

one can write the explicit form of the growth function $\hat{g}$ as

$$\hat{g} = \frac{1}{2}\left(\tau_I^2 + \tau_{II}^2 + \tau_{III}^2\right). \tag{80}$$

The gradient of $\hat{g}$ with respect to the stress field gives the following matrix representation of a tensor describing the anisotropic evolution of microdamage:

$$\frac{\partial \hat{g}}{\partial \boldsymbol{\tau}} = \begin{bmatrix} g_{11}\tau_I & 0 & 0 \\ 0 & g_{22}\tau_{II} & 0 \\ 0 & 0 & g_{33}\tau_{III} \end{bmatrix}. \tag{81}$$

In Eq. 81 τ_I, τ_{II}, τ_{III} are the principal values of Kirchhoff stress tensor.

For temperature evolution, the following relation is considered:

$$\mathbf{k} = \boldsymbol{\tau}.$$

Material Parameters Identification for HSLA-65 Steel

For identification, experimental data for HSLA-65 steel presented in Nemat-Nasser and Guo (2005) are used. The HSLA-65 steel belongs to the class of HSLA steels (high strength low alloy) which were developed in 1960s. The high strength of this steel (flow stresses are in the rage 400 ÷ 1,200 *MPa* dependently on temperature), connected with good weldability, formability, toughness, elevated service life, and less weight to the traditional high-strength steel, causes the broad range of its nowadays applications, e.g., cars, trucks, cranes, bridges, naval surface vessels, submarines, and other structures, that are designed to handle large amounts of material efforts, frequently subjected to wide range of temperatures. The steel has the characteristics of the bcc structure, hence belongs to so-called ferritic steels. As a consequence, this metal has high temperature and strain rate sensitivity and displays good ductility and plasticity (true strain $>$ 60 %). The mechanical properties of the HSLA-65 steel are strongly affected by impurities in its internal structure. It is important that the processing (rolling) of the HSLA-65 steel can induce the anisotropy of its structure.

Table 2 Material parameters for HSLA-65 steel

$\lambda = 121.15$ GPa	$\mu = 80.769$ GPa	$T_m = 2.5\ \mu$s	$m_{pl} = 0.14$
$n_1 = 0$	$n_2 = 0.25$	$\overset{*}{\chi} = 0.8$	$\overset{**}{\chi} = 0.1$
$\overset{*}{\kappa_s} = 570$ MPa	$\overset{**}{\kappa_s} = 129$ MPa	$\overset{*}{\kappa_0} = 457$ MPa	$\overset{**}{\kappa_0} = 103$ MPa
$\overset{*}{\delta} = 6.0$	$\overset{**}{\delta} = 1.4$	$\overset{*}{\beta} = 11.0$	$\overset{**}{\beta} = 2.5$
$c = 0.067$	$\theta = 10^{-6}\ \mathrm{K}^{-1}$	$m_{md} = 1$	$m_F = 0.5$
$b_1 = 0.02$	$b_2 = 0.5$	$b_3 = 0$	$\|\|\mathrm{L}_v\boldsymbol{\xi}_c\|\| = 10^{-5}\ \mathrm{s}^{-1}$
$\overset{F^*}{\xi} = 0.36$	$\overset{F^{**}}{\xi} = 0.03$	$c_p = 470\frac{\mathrm{J}}{\mathrm{kgK}}$	$\rho_{Ref} = 7800\frac{\mathrm{kg}}{\mathrm{m}^3}$

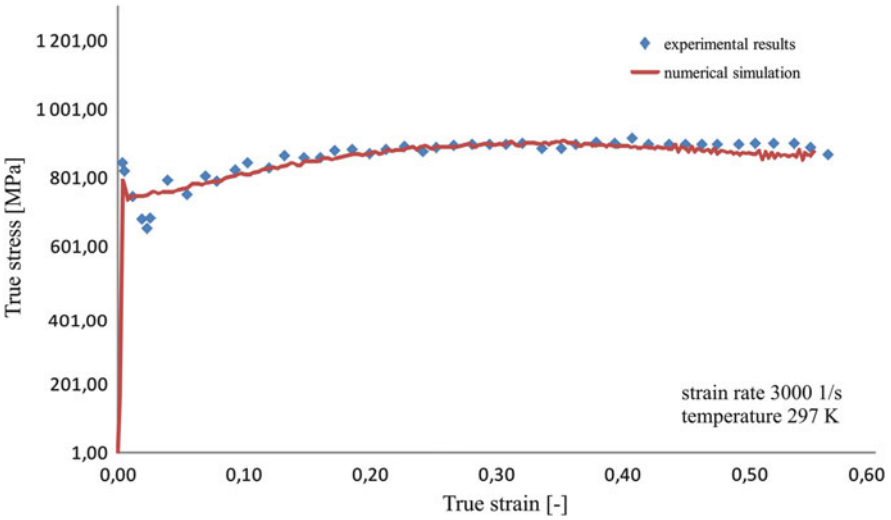


Fig. 8 The comparison of strain-stress curve from experiment and numerical simulation

The set of material parameters is presented in Table 2.

It should be emphasized however that the identification of such great number of material parameters is awkward. From one point of view, the results presented in Nemat-Nasser and Guo (2005) are not enough to calibrate the presented material model in which all variables (e.g., temperature, viscoplastic strain, microdamage) are coupled. On the other hand, recall that current experimental techniques cannot still give a detailed/unique answer for extremely fast thermomechanical processes. Thus, parameters from Table 2 should be thought as a compromise; hence, small fluctuations of them are possible (dependently on detailed experimental results showing competition of fundamental variables, e.g., temperature, viscoplastic strain, microdamage). For proposition of reduction of the number of material parameters using soft computing methods, cf. Sumelka and Łodygowski (2013).

In Fig. 8 the comparison of experimental and numerical results is presented. Notice that the numerical solution is obtained from full 3D thermomechanical analysis accounting for the previously mentioned anisotropic intrinsic

microdamage process (microdamage state was initially isotropic during calibration procedure; however, the obtained parameters are valid for other initial microdamage states since its influence on GDA scatter is in experimentally acceptable bounds (Sumelka and Łodygowski 2011)). In other words, the presented numerical results take into account the whole local processes. The curve fitting shows that using presented material model, one can obtain the numerical simulations in a very good agreement with the experimental observations.

Model Validation Numerical Examples

Introductory Remarks

Up to this point, a constitutive structure of Perzyna type including anisotropic description for damage was defined, and all material functions were postulated. Recall that the important features of this formulation are as follows: (i) the description is invariant with respect to any diffeomorphism (covariant material model); (ii) the obtained evolution problem is well posed; (iii) sensitivity to the rate of deformation; (iv) finite elasto-viscoplastic deformations; (v) plastic non-normality; (vi) dissipation effects; (vii) thermomechanical couplings; and (viii) length-scale sensitivity.

For the moment let the attention be focused to the feature (i) which is a result of consequent use of Lie derivative and is connected with objectivity of the description. The importance of the subject of matter lays in fact that it is common to define spatial tensor rates in terms of Zaremba-Jaumann (1903) or Green-Naghdi (1965) rates in most of popular constitutive models (and commercial programs like Abaqus (2012)) although in some situations they can lead to nonphysical solutions (Dienes 1979; Lehmann 1972; Nagtegaal and Jong 1982; Xiao et al. 1997a). The situation described is most strongly vivid in extreme dynamics. It is then crucial to choose appropriately the objective rate definition from the set of all objective rates which is infinite, unfortunately. This problem was discussed in Sumelka (2013) where the importance of using Lie derivative was pointed out. Summarizing using different objective rate, one can observe differences in global strain-stress space response, geometry and intensity of localized deformation zones, and initiation time of macrodamage its direction and final fracture pattern. One can conclude that covariant material model, being one of the most general in continuum mechanics, presents most robust and stable solution.

The proposed theory has been tested for many different types of IBVPs to indicate different aspects of the formulation. So, dynamic tension and twisting were discussed in Sumelka (2009) and Glema et al. (2010b); dynamic shearing was shown in Glema et al. (2010a) and Sumelka and Łodygowski (2013), while machining was a subject of Łodygowski et al. (2012; cf. Fig. 9). In every process mentioned, different types of combined loading were captured (e.g., different triaxiality, local temperatures, or strain rates), and robustness of formulation was proved.

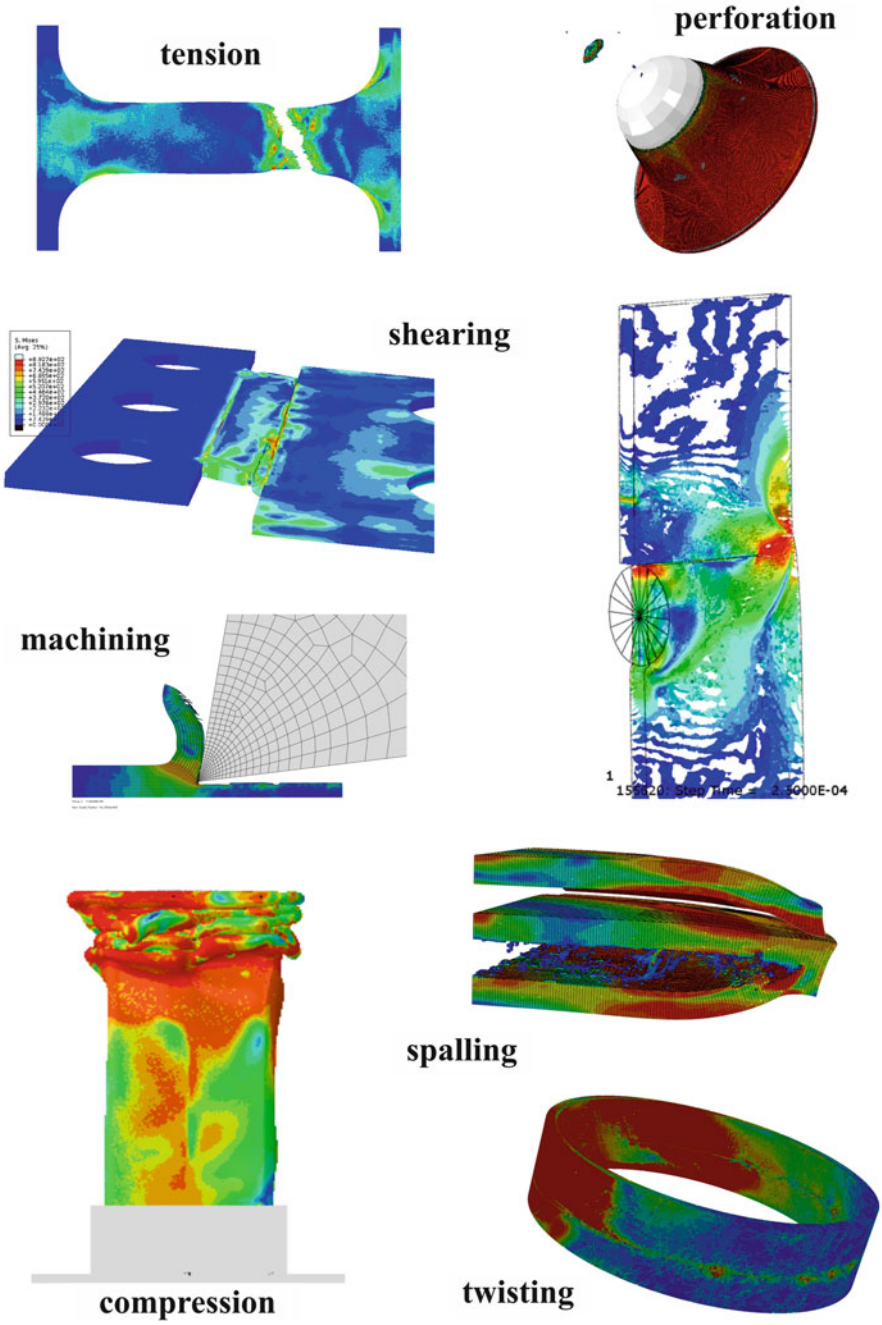


Fig. 9 The processes solved under Perzyna-type viscoplasticity accounting for anisotropic damage

Model Validation

In the reaming part of this section, the modelling of spalling phenomenon with analysis of the evolution of fracture porosity will be considered. However, first computer implementation of the theory in Abaqus/Explicit finite element program utilizing Lie derivative is discussed.

Computer Implementation in Abaqus/Explicit

The solution of the IBVP defined by Eq. 62 has been obtained with the use of finite element method. The Abaqus/Explicit commercial finite element code has been adapted as a solver. The Abaqus/Explicit utilizes central-difference time integration rule along with the diagonal ("lumped") element mass matrices. To remove damaged elements from the mesh (elements in which for every integration point, fracture porosity was reached, or equivalently load carrying capacity is zero Eq. 61), the so-called element deletion method is applied (Song et al. 2008). The model has been implemented in the software by taking advantage of a user subroutine VUMAT, which is coupled with Abaqus system (2012).

The stress update in VUMAT user subroutine needs some additional comments. During computations, the user subroutine VUMAT controls the evolution of stresses, viscoplastic deformation, temperature, and microdamage fields. Recall that in the presented material model, derivative has been taken into account for all rates, including stress rate.

Hence, for the stress rate, the Lie rate is enforced,

$$\mathbf{L}_{\mathbf{v}}\boldsymbol{\tau} = \dot{\boldsymbol{\tau}} - \mathbf{l}^{\mathrm{T}} \cdot \boldsymbol{\tau} - \boldsymbol{\tau} \cdot \mathbf{l}, \tag{83}$$

in opposition to the Green-Naghdi rate calculated by default in Abaqus/Explicit VUMAT user subroutine according to the following formula (Abaqus 2012)

$$\boldsymbol{\tau}^{(G-N)^{\circ}} = \dot{\boldsymbol{\tau}} + \boldsymbol{\tau} \cdot \boldsymbol{\Omega} - \boldsymbol{\Omega} \cdot \boldsymbol{\tau}, \tag{84}$$

where $\boldsymbol{\Omega} = \boldsymbol{\Omega}^{(G-N)} = \dot{\mathbf{R}} \cdot \mathbf{R}^{T}$ represents the angular velocity of the material (Dienes 1979) (or spin tensor (Xiao et al. 1997b)) and R denotes the rotation tensor. It is also worth mentioning that the material model in Abaqus/Explicit VUMAT user subroutine is defined in corotational coordinate system, being described by the spin tensor $\boldsymbol{\Omega}$ (see Fig. 10).

To keep the VUMAT algorithm objective in Lie sense, the following approach is applied. In the iterative procedure, the forward difference scheme is taken as the material derivative of the second-order tensor. Thus, for the material derivative of the Kirchhoff stress tensor, one has

$$\dot{\boldsymbol{\tau}}|_{i} = \frac{\boldsymbol{\tau}|_{i+1} - \boldsymbol{\tau}|_{i}}{\Delta t}. \tag{85}$$

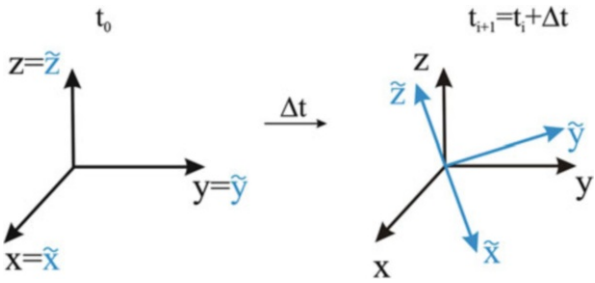


Fig. 10 Initial (XYZ), current (xyz), corotational $\tilde{x}\tilde{y}\tilde{z}$), and "damage" ($\hat{x}\hat{y}\hat{z}$) coordinate systems

Using Eqs. 83 and 84, one can write in the corotational coordinate system, respectively:

$$\widetilde{\boldsymbol{\tau}}|_{i+1} = \mathrm{R}^T\big|_{i+1}\Big[\boldsymbol{\tau}|_i + \Delta t \mathrm{L}_v\boldsymbol{\tau}|_i + \Delta t\Big(\mathbf{I}^T\big|_i \cdot \mathbf{I}|_i + \boldsymbol{\tau}|_i \cdot \mathbf{I}|_i\Big)\Big] + \mathbf{R}|_{i+1}, \tag{86}$$

and

$$\widetilde{\boldsymbol{\tau}}|_{i+1} = \mathbf{R}^T\big|_{i+1}\Big[\boldsymbol{\tau}|_i + \Delta t\boldsymbol{\tau}^{(G-N)^\circ}\big|_i + \Delta t\big(\boldsymbol{\Omega}|_i \cdot \boldsymbol{\tau}|_i - \boldsymbol{\tau}|_i \cdot \boldsymbol{\Omega}|_i\big)\Big]\mathbf{R}|_{i+1}. \tag{87}$$

Thus, it is clear that the Green-Naghdi rate produces an additional term

$$\Delta t\big(\boldsymbol{\Omega}|_i \cdot \boldsymbol{\tau}|_i - \boldsymbol{\tau}|_i \cdot \boldsymbol{\Omega}|_i\big). \tag{88}$$

That is why one has to subtract this term since different is proposed. Hence, in the presented formulation for the stress update in VUMAT, the following holds (Sumelka 2009):

$$\widetilde{\boldsymbol{\tau}}\big|_{i+1} = \mathbf{R}^T\big|_{i+1}\Big[\boldsymbol{\tau}\big|_i + \Delta t\big(2\boldsymbol{\tau}|_i \cdot \mathbf{d}|_i + \mathbf{L}_v\boldsymbol{\tau}|_i\big) + \boldsymbol{\Upsilon}\big|_i\Big]\mathbf{R}|_{i+1}, \tag{89}$$

where $\boldsymbol{\Upsilon}|_i = -\Delta t(\boldsymbol{\Omega}|_i \cdot \boldsymbol{\tau}|_i - \boldsymbol{\tau}|_i \cdot \boldsymbol{\Omega}|_i)$ and $\boldsymbol{\tau}|_i = \mathbf{R}|_i\widetilde{\boldsymbol{\tau}}|_i\mathbf{R}^T\big|_i$.

Such approach is necessary only for stresses, since other variables are kept as scalars. The detailed algorithm for the whole process can be found in Sumelka (2009).

Spall Fracture Phenomenon Modelling

Spall Fracture Phenomenon

Spallation is a particular kind of dynamic fracture (can lead to fragmentation) (Hanim and Klepaczko 1999). Due to its variety of applications, particularly in terminal ballistics and detonics, many studies have been reported for spall so far in literature (Klepaczko 1990b; Curran et al. 1987; Meyers and Aimone 1983).

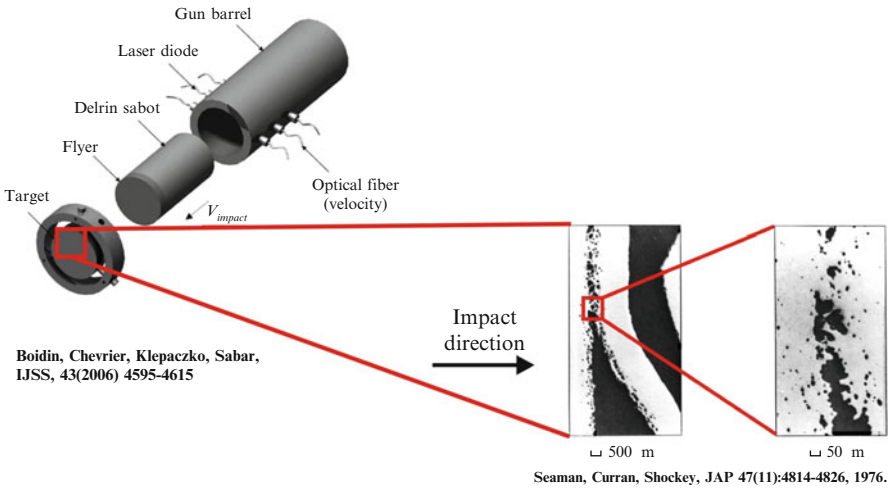


Fig. 11 Experimental setup for spall fracture investigation

Experimentally, spall fracture can be analyzed considering flat plate impact test (Boidin et al. 2006; cf. Fig. 11). In the experiment, two plates with high velocities impact each other. As result of impact (caused by the flyer plate), a complete or partial separation of the material can appear in the target plate (Fig. 11). This is due to tension in the target plate, induced by the interaction of two waves, incident and reflected. As mentioned, the separation is preceded by the evolution of microdamage (microvoids), consisting for undamaged material of three stages: nucleation, growth, and coalescence. It is experimentally observed that the level of porosity (called fracture porosity in the model) in spall zone changes dependently on velocity of flayer plate (Cochran and Banner 1988; Curran et al. 1987). For sufficiently high impact velocities, target plate breaks in brittle mode (porosity is very limited in spall zone in that case).

In following the numerical analysis of flat plate, impact test is considered.

Computational Model

The geometry for spall modelling is presented in Fig. 12. The dimensions of flayer plate are diameter $\phi_{\text{fla}} = 114$ mm and thickness $t_{\text{fla}} = 5$ mm, while for target plate diameter is $\phi_{\text{tar}} = 114$ mm and thickness is $t_{\text{tar}} = 10$ mm

Because of anisotropy, full spatial modelling was enforced for computational model. Thus, for the spall model, continuum elements were used – the C3D8R finite elements (8-node linear brick, reduced integration element). Totally $3M$ of finite elements were applied.

Six different initial velocities were applied $v_0 = 50, 150, 250, 350, 450,$ and $500\frac{\text{m}}{\text{s}}$, cf. Fig. 13 and the initial temperature 296 K was assumed. Because of the lack of the experimental data concerning initial microdamage distribution in the specimen, it was assumed equal in every material point in the body and isotropic. This simplification is crucial concerning the fact that the way of mapping of initial microdamage

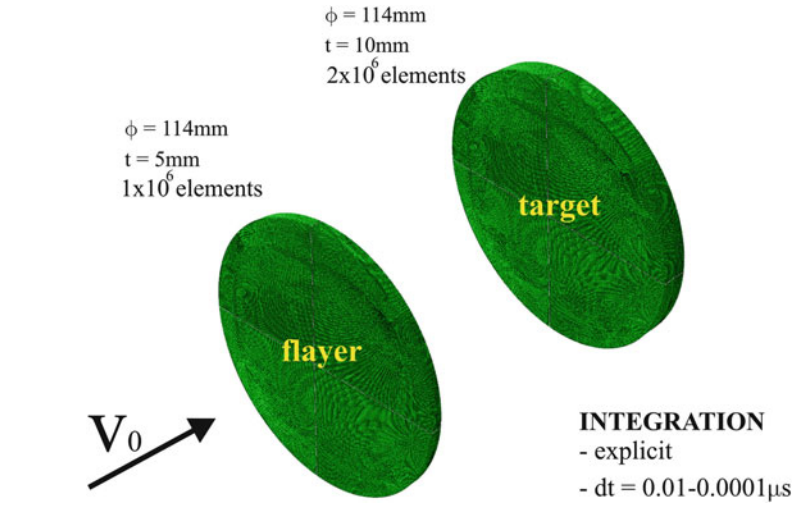


Fig. 12 Numerical model for modelling spall phenomenon

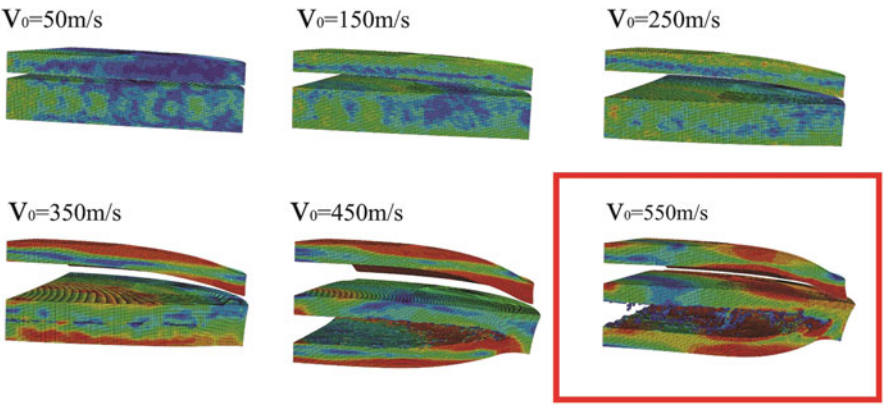


Fig. 13 The intensity of spalling in target plate versus velocity of flayer plate (the box indicates the case chosen for further detailed analysis)

state has a strong impact on a final failure mode and the global answer from the specimen (Sumelka and Łodygowski 2011). The components of microdamage tensor were chosen in a way to obtain initial porosity equal to $6 \cdot 10^{-4}$, namely,

$$\boldsymbol{\xi}_0 = \begin{bmatrix} 34.64 \cdot 10^{-5} & 0 & 0 \\ 0 & 34.64 \cdot 10^{-5} & 0 \\ 0 & 0 & 34.64 \cdot 10^{-5} \end{bmatrix}.$$

To assure proper contact conditions, the general contact in Abaqus/Explicit was applied, which includes the self-contact conditions. The properties of a

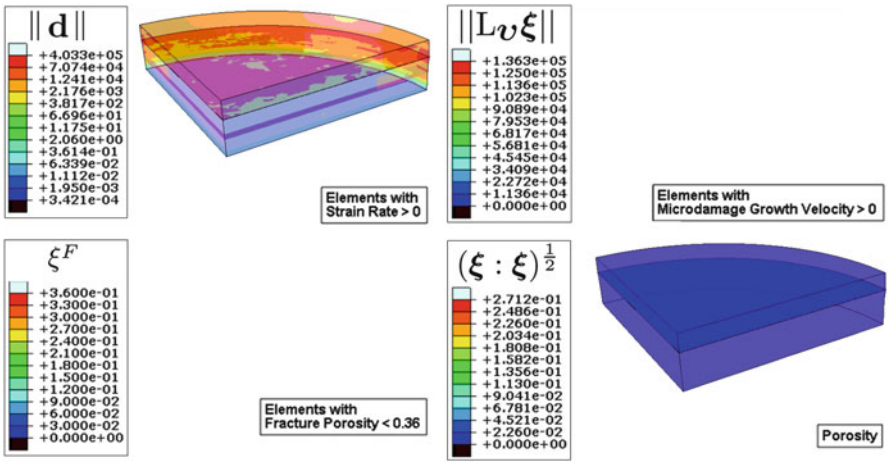


Fig. 14 The plot of strain rates, microdamage growth velocity, fracture porosity, and porosity for time process $1 \cdot 10^{-6}$ s (flayer velocity $v_0 = 550\frac{\text{m}}{\text{s}}$)

contact were the hard normal contact (without penetration and unlimited contact stresses) and the tangential contact with Coulomb friction model (friction coefficient equal 0.05).

Material being tested was HLSA-65 steel.

Results Discussion

In experiment, it is observed that the refection of the compressive incident wave from the free surface of the target produces a tensile stress wave, and at distance t_{fla} from the free surface, a high tensile stress occurs before the arrival of release waves from the edges of the plate. Thus, the central portion of the target is in a confined state of one-dimensional strain. The compressive wave in the flyer plate is reflected by the free surface as a tensile wave and returns to the impact surface. If the magnitude and duration of this tensile stress wave are high enough, spallation occurs (Hanim and Klepaczko 1999).

In Fig. 13 the numerical results of spalling for six different initial velocities of flayer plate are presented. It is seen that for the material tested, the critical velocity for which spall occurs is between $250 \div 350\frac{\text{m}}{\text{s}}$. For higher flayer plate velocities > 350 m, intensive spalling is observed. For further detailed analysis, the case with initial velocity $550\frac{\text{m}}{\text{s}}$ is chosen.

The evolution of strain rates, microdamage growth velocity, fracture porosity, and porosity during the process (the analysis of other variables like stresses, thermal stresses, strains, temperatures, etc., is not included) is presented in Figs. 14–17.

The local strain rates as high as ca. $7 \cdot 10^5$ s^{-1} are observed. The velocity of microdamage growth reaches ca. $4.5 \cdot 10^6\frac{\text{m}}{\text{s}}$ for begging of spall

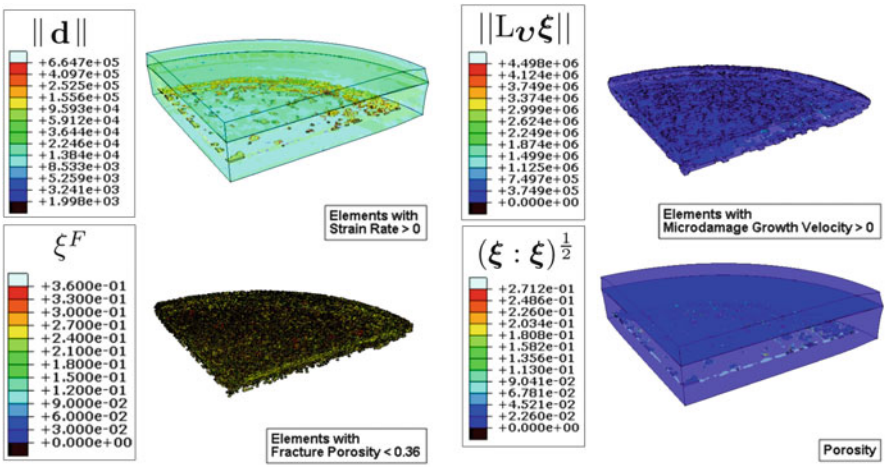


Fig. 15 The plot of strain rates, microdamage growth velocity, fracture porosity, and porosity for time process $4 \cdot 10^{-6}$ s (flayer velocity $v_0 = 550\frac{\text{m}}{\text{s}}$)

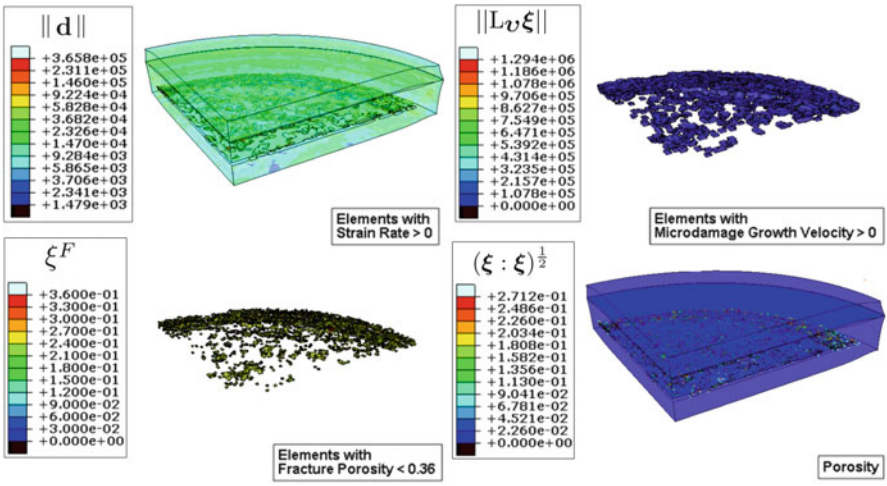


Fig. 16 The plot of strain rates, microdamage growth velocity, fracture porosity, and porosity for time process $5 \cdot 10^{-}6$ s (flayer velocity $v_0 = 550\frac{\text{m}}{\text{s}}$)

($t = 4 \cdot 10^{-6}$ s Fig. 15). On the other side, the evolution of fracture porosity, which is very intensive in spall zone (cf. Figs. 15 and 16), shows its reduction from static value $\xi^{F*} = 0.36$ to dynamic one which is ca. $\xi^F = 0.27$. The accumulation of porosity is proved to be in spall zone as in real experiment.

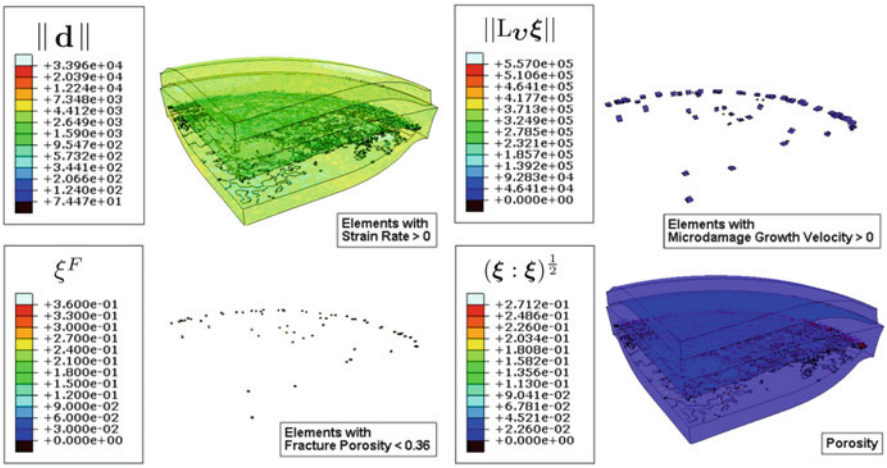


Fig. 17 The plot of strain rates, microdamage growth velocity, fracture porosity, and porosity for time process $3.7 \cdot 10^{-5}$ s (flayer velocity $v_0 = 550\frac{\text{m}}{\text{s}}$)

Conclusions

In this document recent developments of Perzyna-type viscoplasticity for extreme dynamic processes in metallic materials, including the effects of anisotropic damage, are presented. It is shown that to assure reliability of the extreme dynamics modelling results, in its qualitative and quantitative meaning, the phenomena included in the description is high. In consequence, the number of material parameters in the model is also considerable.

Within the document, all crucial stages of modelling are revised. The discussion is started with experimental motivations, and then the fundamental results of Perzyna-type viscoplasticity accounting for anisotropic damage are described and also its computer implementation, identification, and validation. The illustrative example considers numerical analysis of spall phenomenon.

Finally, one can state that utilizing the described model, one is able not only to trace stresses, strains, and temperatures during particular process but also directional evolution of microdamage initiation and growth. In other words, one is able to predict time initiation of damage, its evolution directions (even before material separation occurs – by tracing microdamage tensor principal directions), and velocities. On the other hand, seemingly, the drawback is that model identification needs very detailed and expensive experimental tests, due to the number of material parameters. However, if one considers extreme dynamics processes, it will be always the case.

References

Abaqus, Abaqus Version 6.12 Theory Manual, 2012, SIMULIA Worldwide, Headquarters, Providence, RI.

R.K. Abu Al-Rub, G.Z. Voyiadjis, A finite strain plastic-damage model for high velocity impact using combined viscosity and gradient localization limiters: part I – theoretical formulation. Int. J. Damage Mech. **15**(4), 293–334 (2006)

X. Boidin, P. Chevrier, J.R. Klepaczko, H. Sabar, Identification of damage mechanism and validation of a fracture model based on mesoscale approach in spalling of titanium alloy. Int. J. Solids Struct. **43**(14–15), 4029–4630 (2006)

J.D. Campbell, The dynamic yielding of mild steel. Acta Metall. **1**(6), 706–710 (1953)

S. Cochran, D. Banner, Spall studies in uranium. J. Appl. Phys. **48**(7), 2729–2737 (1988)

D.R. Curran, L. Seaman, D.A. Shockey, Dynamic failure of solids. Phys. Rep. **147**(5-6), 253–388 (1987)

J.K. Dienes, On the analysis of rotation and stress rate in deforming bodies. Acta Mech. **32**, 217–232 (1979)

W. Dornowski, Influence of finite deformations on the growth mechanism of microvoids contained in structural metals. Arch. Mech. **51**(1), 71–86 (1999)

W. Dornowski, P. Perzyna, Analysis of the influence of various effects on cycle fatigue damage in dynamic process. Arch. Appl. Mech. **72**, 418–438 (2002a)

W. Dornowski, P. Perzyna, Localized fracture phenomena in thermo-viscoplastic flow process under cyclic dynamic loadings. Acta Mech. **155**, 233–255 (2002b)

W. Dornowski, P. Perzyna, Numerical analysis of macrocrack propagation along a bimaterial interface under dynamic loading processes. Int. J. Solids Struct. **39**, 4949–4977 (2002c)

W. Dornowski, P. Perzyna, Numerical investigation of localized fracture phenomena in inelastic solids. Found. Civil Environ. Eng. **7**, 79–116 (2006)

M.K. Duszek–Perzyna, P. Perzyna, Material instabilities: theory and applications, in *Analysis of the Influence of Different Effects on Criteria for Adiabatic Shear Band Localization in Inelastic Solids*, eds. by R.C. Batra and H.M. Zbib, vol. 50 (ASME, New York, 1994)

A. Glema, Analysis of wave nature in plastic strain localization in solids, in *Rozprawy*, vol 379 (Publishing House of Poznan University of Technology, Poznan, Poland, 2004, in Polish)

A. Glema, T. Łodygowski, W. Sumelka, P. Perzyna, The numerical analysis of the intrinsic anisotropic microdamage evolution in elasto-viscoplastic solids. Int. J. Damage Mech. **18**(3), 205–231 (2009)

A. Glema, T. Łodygowski, W. Sumelka, Nowacki's double shear test in the framework of the anisotropic thermo-elasto-vicsoplastic material model. J. Theor. Appl. Mech. **48**(4), 973–1001 (2010a)

A. Glema, T. Łodygowski, W. Sumelka, Towards the modelling of an anisotropic solids. Comput. Methods Sci. Technol. **16**(1), 73–84 (2010b)

H.A. Grebe, H.-R. Pak, M.A. Meyers, Adiabatic shear localization in titanium and Ti-6 pct Al-4 pct V alloy. Metall. Mater. Trans. A **16**(5), 761–775 (1985)

A.E. Green, P.M. Naghdi, A general theory of an elastic–plastic continuum. Arch. Ration. Mech. Anal. **18**, 251–281 (1965)

S. Hanim, J.R. Klepaczko, Numerical study of spalling in an aluminum alloy 7020–T6. Int. J. Impact Eng. **22**, 649–673 (1999)

R. Hill, Aspects of invariance in solid mechanics. Adv. Appl. Mech. **18**, 1–75 (1978)

G.A. Holzapfel, *Nonlinear Solid Mechanics – A Continuum Approach for Engineering* (Wiley, New York, 2000)

J.R. Klepaczko, Behavior of rock like materials at high strain rates in compression. Int. J. Plast. **6**, 415–432 (1990a)

J.R. Klepaczko, Dynamic crack initiation, some experimental methods and modelling, in *Crack dynamics in metallic materials*, ed. by J.R. Klepaczko (Springer, Vienna, 1990b), pp. 255–453

J.R. Klepaczko, Constitutive relations in dynamic plasticity, pure metals and alloys, in *Advances in Constitutive Relations Applied in Computer Codes* (CISM, Udine, 2007). 23–27 July
T. Lehmann, Anisotrope plastische Formänderungen. Rom. J. Tech. Sci. Appl. Mech. **17**, 1077–1086 (1972)
T. Łodygowski, P. Perzyna, Localized fracture of inelastic polycrystalline solids under dynamic loading process. Int. J. Damage Mech. **6**, 364–407 (1997a)
T. Łodygowski, P. Perzyna, Numerical modelling of localized fracture of inelastic solids in dynamic loading process. Int. J. Numer. Methods Eng. **40**, 4137–4158 (1997b)
T. Łodygowski, W. Sumelka, Damage induced by viscoplastic waves interaction. Vib. Phys. Syst. **25**, 23–32 (2012)
T. Łodygowski, A. Rusinek, T. Jankowiak, W. Sumelka, Selected topics of high speed machining analysis. Eng. Trans. **60**(1), 69–96 (2012)
P. Longere, A. Dragon, H. Trumel, X. Deprince, Adiabatic shear banding-induced degradation in thermo-elastic/viscoplastic material under dynamic loading. Int. J. Impact Eng. **32**, 285–320 (2005)
J.E. Marsden, T.J.H. Hughes, *Mathematical Foundations of Elasticity* (Prentice-Hall, New Jersey, 1983)
M.A. Meyers, C.T. Aimone, *Dynamic Fracture (Spalling) of Materials*. Progress in Material Science, vol. 28 (Pergamon, New York, 1983)
J.C. Nagtegaal, J.E. de Jong, Some aspects of non-isotropic work-hardening in finite strain plasticity, in *Proceedings of the workshop on plasticity of metals at finite strain: theory, experiment and computation*, ed. by E.H. Lee, R.L. Mallet (Stanford University, New York, 1982), pp. 65–102
R. Narayanasamy, N.L. Parthasarathi, C.S. Narayanan, Effect of microstructure on void nucleation and coalescence during forming of three different HSLA steel sheets under different stress conditions. Mater. Des. **30**, 1310–1324 (2009)
S. Nemat-Nasser, W.-G. Guo, Thermomechanical response of HSLA-65 steel plates: experiments and modeling. Mech. Mater. **37**, 379–405 (2005)
J.A. Nemes, J. Eftis, Several features of a viscoplastic study of plate-impact spallation with multidimensional strain. Comput. Struct. **38**(3), 317–328 (1991)
J.A. Nemes, J. Eftis, Constitutive modelling of the dynamic fracture of smooth tensile bars. Int. J. Plast. **9**(2), 243–270 (1993)
R.B. Pęcherski, W.K. Nowacki, Z. Nowak, P Perzyna, Effect of strain rate on ductile fracture. a new methodology, in *Workshop, Dynamic Behaviour of Materials, In memory of our Friend and Colleague Prof. J.R. Klepaczko*, Metz, 13–15 May 2009, pp. 65–73
P. Perzyna, The constitutive equations for rate sensitive plastic materials. Quart. Appl. Math. **20**, 321–332 (1963)
P. Perzyna, Fundamental problems in viscoplasticity. Adv. Appl. Mech. **9**, 243–377 (1966)
P. Perzyna, *Termodynamika materiałów niesprężystych* (PWN, Warszawa, 1978) (in Polish)
P. Perzyna, Constitutive modelling for brittle dynamic fracture in dissipative solids. Arch. Mech. **38**, 725–738 (1986a)
P. Perzyna, Internal state variable description of dynamic fracture of ductile solids. Int. J. Solids Struct. **22**, 797–818 (1986b)
P. Perzyna, Instability phenomena and adiabatic shear band localization in thermoplastic flow process. Acta Mech. **106**, 173–205 (1994)
P. Perzyna, Constitutive modelling of dissipative solids for localization and fracture, chapter 3, in *Localization and Fracture Phenomena in Inelastic Solids*, ed. by P. Perzyna. CISM Course and Lectures, vol. 386 (Springer, New York, 1998), pp. 99–241
P. Perzyna, The thermodynamical theory of elasto-viscoplasticity. Eng. Trans. **53**, 235–316 (2005)
P. Perzyna, The thermodynamical theory of elasto-viscoplasticity accounting for microshear banding and induced anisotropy effects. Mech. Mater. **27**(1), 25–42 (2008)
A. Rusinek, J.R. Klepaczko, Experiments on heat generated during plastic deformation and stored energy for trip steels. Mater. Des. **30**(1), 35–48 (2009)

C. Rymarz, *Mechanika ośrodków ciągłych* (PWN, Warszawa, 1993) (in Polish)
L. Seaman, D.R. Curran, D.A. Shockey, Computational models for ductile and brittle fracture. J. Appl. Phys. **47**(11), 4814–4826 (1976)
S. Shima, M. Oyane, Plasticity for porous solids. Int. J. Mech. Sci. **18**, 285–291 (1976)
J.-H. Song, H. Wang, T. Belytschko, A comparative study on finite element methods for dynamic fracture. Comput. Mech. **42**, 239–250 (2008)
W. Sumelka, *The Constitutive Model of the Anisotropy Evolution for Metals with Microstructural Defects* (Publishing House of Poznan University of Technology, Poznań, 2009)
W. Sumelka, The role of the covariance in continuum damage mechanics. ASCE J. Eng. Mech. (2013). doi:10.1061/(ASCE)EM.1943-7889.0000600
W. Sumelka, T. Łodygowski, The influence of the initial microdamage anisotropy on macrodamage mode during extremely fast thermomechanical processes. Arch. Appl. Mech. **81**(12), 1973–1992 (2011)
W. Sumelka, T. Łodygowski, Reduction of the number of material parameters by an approximation. Comput. Mech. **52**, 287–300 (2013)
W. Sumelka, T. Łodygowski, Thermal stresses in metallic materials due to extreme loading conditions. ASME J. Eng. Mater. Technol. **135**, 021009–1–8 (2013)
W. Sumelka, A. Glema, The evolution of microvoids in elastic solids, in *17th International Conference on Computer Methods in Mechanics CMM-2007*, Łódź-Spała, 19–22 June 2007, pp. 347–348
T. Łodygowski, *Theoretical and Numerical Aspects of Plastic Strain Localization*, vol 312, D.Sc. Thesis, Publishing House of Poznan University of Technology, 1996
T. Łodygowski, A. Glema, W. Sumelka, Anisotropy induced by evolution of microstructure in ductile material, in *The 8th World Congress on Computational Mechanics (WCCM8), 5th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2008)*, Venice, 30 June–5 July 2008
W. Thomson, On six principal strains of an elastic solid. Philos. Trans. R. Soc. **166**, 495–498 (1856)
D. Tikhomirov, R. Niekamp, E. Stein, On three-dimensional microcrack density distribution. ZAMM J. Appl. Math. Mech. **81**(1), 3–16 (2001)
G.Z. Voyiadjis, R.K. Abu Al-Rub, A finite strain plastic-damage model for high velocity impacts using combined viscosity and gradient localization limiters: part II – numerical aspects and simulations. Int. J. Damage Mech. **15**(4), 335–373 (2006)
H. Xiao, O.T. Bruhns, A. Meyers, Hypo-elasticity model based upon the logarithmic stress rate. J. Elast. **47**, 51–68 (1997a)
H. Xiao, O.T. Bruhns, A. Meyers, Logarithmic strain, logarithmic spin and logarithmic rare. Acta Mech. **124**, 89–105 (1997b)
H. Xiao, O.T. Bruhns, A. Meyers, Strain rates and material spin. J. Elast. **52**, 1–41 (1998)
S. Zaremba, Sur une forme perfectiońée de la théorie de la relaxation. Bull. Int. Acad. Sci. Crac. **8**, 594–614 (1903)

Plasticity Conditions and Failure Criteria for Quasi-brittle Materials

36

Tomasz Jankowiak and Tomasz Łodygowski

Contents

Abstract

In this chapter, selected problems connected with failure of quasi-brittle materials in static and dynamic conditions are presented. The starting points of the

T. Jankowiak (✉)
Division of Computer Aided Design, Institute of Structural Engineering, Poznan University of Technology, Faculty of Civil and Environmental Engineering, Poznań, Poland
e-mail: tomasz.jankowiak@put.poznan.pl

T. Łodygowski
Division of Computer Aided Design, Institute of Structural Engineering, Poznan University of Technology, Poznań, Poland
e-mail: tomasz.lodygowski@put.poznan.pl

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_48

discussion are quasi-static and dynamic mechanical properties of concrete-like strength in uniaxial compression, tension, plane stress conditions, and meridian plane, which are confronted with dynamic strength in tension and in compression. There are at least few constitutive relations known from literature which reflect the slightly different ideas of authors. The important part of this chapter is the presentation of the selected plasticity conditions and failure criteria due to von Mises, Drucker-Prager, Bresler-Pister, Mróz, Willam-Warnke, Podgórski, and Burzyński. The interpretation of the specified criteria that are based on energy description is introduced to perform easily comparisons between them. The second important set of problems considers dynamic failure of concrete. The ideas of multiplicative failure criterion (MFC) and cumulative failure criterion (CFD) are shown. The CFD describes influence of constitutive parameters on the definition of time to failure. The method of generalization of this criterion to three-dimensional cases is proposed. The verification of the elastic material model with the strain rate-dependent failure criterion that is based on spalling test of concrete in the Hopkinson pressure bar is discussed. The last application describes numerically the dynamic behavior of concrete and reinforced concrete slab perforated by projectile which hits the obstacle (plate) with different incidence angles.

Introduction

The damage and failure of quasi-brittle materials appear under both quasi-static and dynamic conditions. These types of behavior exhibit, for example, concretes, ceramics, or glasses. This behavior is a consequence of an internal structure, in which, for example, concrete is composed of aggregate, mortar, and water. In the case of concrete – after hydration process – the mixture creates the artificial rock in which behavior is quasi-brittle.

For quasi-static behavior, the most important and useful property of concrete is its compressive strength particularly when designing the concrete and reinforced concrete structures (Popovics 1998 and Jankowiak 2011). The other important and measurable mechanical properties of concrete are tensile and shear strengths. The behavior of this material under multiaxial loadings should also be described as for plane stress conditions. Different tests exist to describe the above quantities, and they are repeated in Neville (2000). The other tests are used to describe the dynamic strength of concrete and other quasi-brittle materials; some results will be discussed in this chapter. It is observed that the dynamic strength of concrete significantly increases for higher strain rates (Abrams 1917) for both tension and compression.

The classical uniaxial static behavior of quasi-brittle material is presented in Fig. 1. For this material, much smaller strength in tension than in compression is typical. Damage for concrete is through crushing in compression but cracking in tension (Chen 1982). The normalized test is performed to describe the compressive behavior of concrete, using the cube or cylinder specimens (Neville 2000).

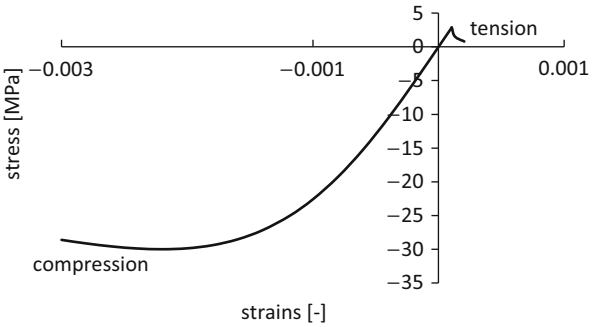


Fig. 1 The uniaxial compression and tensile behavior (concrete C30) under quasi-static loading

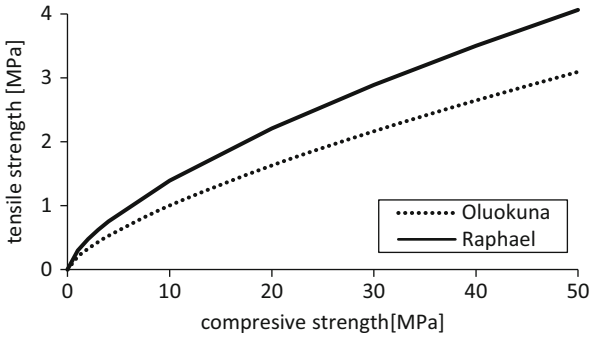


Fig. 2 The relationship between the tensile and the compressive strengths

In experimental practice, the description of tensile behavior and strength is complicated and may be performed using different methods (Neville 2000).

The tensile properties of concrete can be evaluated using three-point bending test, Brazilian test, or direct tension test. In practice, it can be also described by the compressive strength using the following formula:

$$f_t = k(f_c)^n, \tag{1}$$

where f_c and f_t are compressive and tensile strength and k and n are the parameters. The parameters $k = 0.3$ and $n = 2/3$ (Raphael 1984) give the best fitted results with experiments. The other possible values $k = 0.3$ and $n = 0.7$ (Oluokun 1991) present smaller tensile strength and give more brittle tensile behavior. Both curves that map the compressive strength onto tensile strength are presented in Fig. 2.

The important knowledge on the behavior of quasi-brittle materials as concrete or rocks comes from plane stress or three-directional state of stress tests (Chen 1982; Jankowiak 2011). The results of experiments for plain stress conditions are presented in Fig. 3 for typical concrete C30. One can notice the dominated zone of compression for the biaxial case (Kupfer et al. 1969).

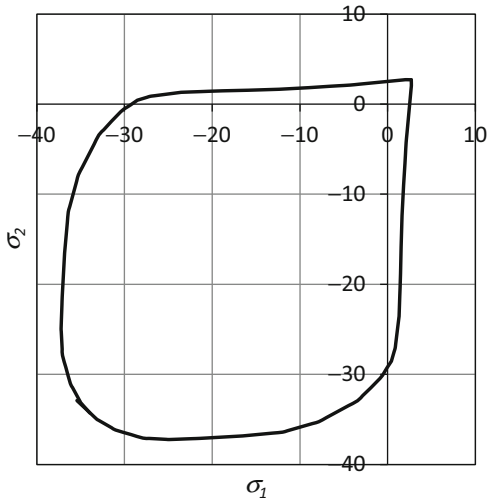


Fig. 3 Failure curve for plane quasi-static stress conditions (concrete C30)

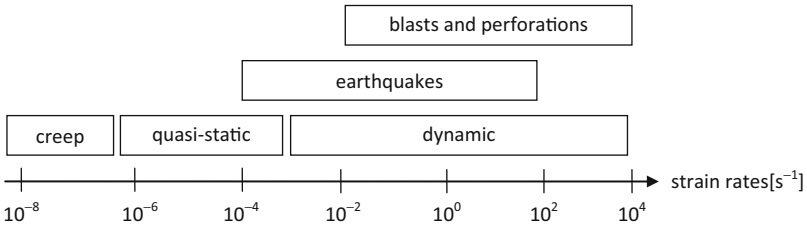


Fig. 4 Strain rates for different unique loadings

When talking on dynamic loadings, the selected range of strain rates is in mind. Different strain rates accompany the phenomena indicated in Fig. 4. The blasts, perforations, rigid hitting, and some earthquakes generate the deformation rates of order 10^2 or higher. These deformation rates are the focus of our interests.

There are different tests for tension and for compression (Klepaczko and Brara 2001; Ross et al. 1992 and Bischoff and Perry 1991) when wishing to describe the dynamic strength of quasi-brittle materials. Mainly the wave theory is adopted to explain the phenomena which appear when testing the specimen in the Hopkinson bar (Klepaczko and Brara 2001; Ross et al. 1992 and Bischoff and Perry 1991; Jankowiak et al. 2011). If the dynamic strength is reached, it may cause the local damage and as a consequence the failure of the whole structure. The effect can be often unpredictable.

The detailed analysis of the compression dynamic tests is presented by Bischoff and Perry (1991). The summary of the experiments is presented in Fig. 5. It shows how the compressive dynamic strength depends on the strain rates. On the ordinate axis is CDIF, i.e., compressive dynamic increase factor. It is clearly seen that CDIF reaches 2.0 for strain rates equal to 100 s^{-1}.

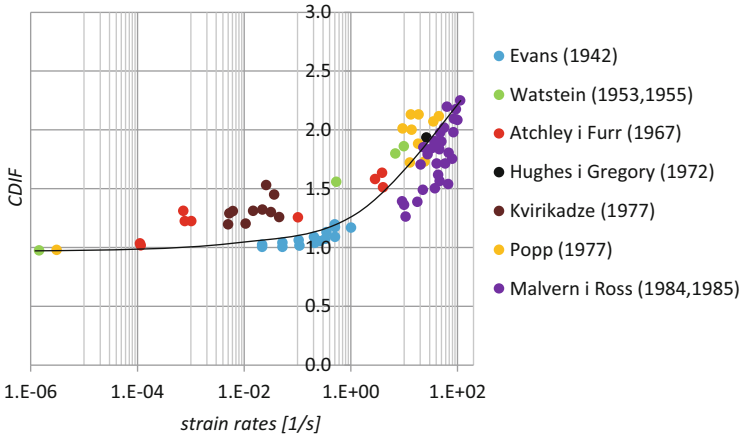


Fig. 5 Influence of strain rates on compressive dynamic increase factor

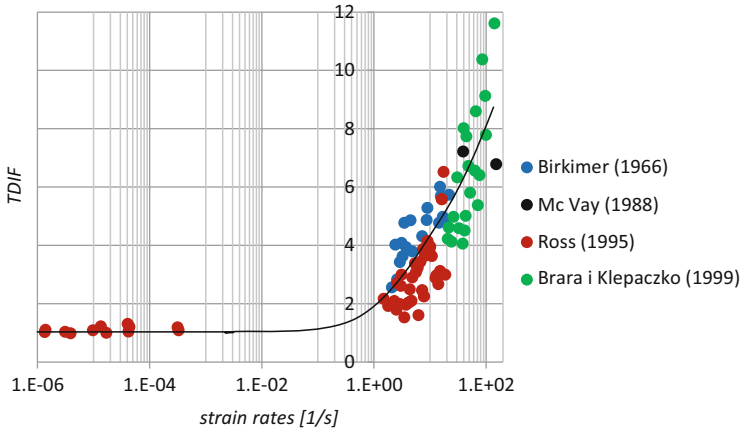


Fig. 6 Influence of strain rates on tensile dynamic increase factor

The experiments which describe the tensile dynamic strength of quasi-brittle material are more complicated to perform. The number of results is small in comparison with compression tests. The important tests of Klepaczko cover the strain rates in tension between 20 s^{-1} and 120 s^{-1} (Klepaczko and Brara 2001). The results from different authors are collected in Fig. 6. It shows the influence of strain rates on TDIF (tensile dynamic increase factor). The increase of dynamic tensile strength observed in experiments, which could be estimated, reaches even 10 for strain rates 120 s^{-1} (see Fig. 6).

The European Standards as CEB (Comite Euro-International du Beton) suggests using the dynamic increase factor in both tension and compression (CEB 1987).

These factors describe how many times the dynamic strength is higher than the static one. The empirical equations proposed by CEB for CDIF are as follows:

$$\mathrm{CDIF} = \frac{f_{cd}}{f_{cs}} = \begin{cases} \left(\frac{\dot{\varepsilon}_d}{\dot{\varepsilon}_{cs}}\right)^{1.026\alpha} & \text{if } \dot{\varepsilon}_d \le 30\mathrm{s}^{-1} \\ \gamma\left(\frac{\dot{\varepsilon}_d}{\dot{\varepsilon}_{cs}}\right)^{1/3} & \text{if } \dot{\varepsilon}_d > 30\mathrm{s}^{-1} \end{cases} \tag{2}$$

where f_{cd} is dynamic compressive strength and f_{cs} compressive quasi-static strength obtained for strain rate $\dot{\varepsilon}_{cs} = 0.00003\ \mathrm{s}^{-1}$. Parameters γ, α are defined by CEB as $\log \gamma = 6.156\alpha - 0.49$ and $\alpha = (5 + 3f_{cu}/4)^{-1}$, where f_{cu} is the tested strength based on cylindrical specimens. The equation proposed for TDIF is the following:

$$\mathrm{TDIF} = \frac{f_{td}}{f_{ts}} = \begin{cases} 1.0 & \text{if} \quad \dot{\varepsilon}_d \le 10s^{-4} \\ 2.06 + 0.26\log\dot{\varepsilon}_d & \text{if} \quad 10\mathrm{s}^{-4} < \dot{\varepsilon}_d \le 1\mathrm{s}^{-1} \\ 2.06 + 2\log\dot{\varepsilon}_d & \text{if} \quad 1\mathrm{s}^{-1} < \dot{\varepsilon}_d \end{cases}, \tag{3}$$

where f_{td} is dynamic tensile strength of concrete for strain rates $\dot{\varepsilon}_d$. The equation can be extended to strain rates up to 1,000 s^{-1} (Klepaczko 1990). Over this limit, the TDIF is assumed to be constant.

Plasticity Conditions and Failure Criteria

The general form of the plasticity/failure criteria expressed in the space of stresses and strain rates for quasi-brittle materials is the following:

$$f\left(\sigma_{ij}, \dot{\varepsilon}_{ij}\right) = 0. \tag{4}$$

It can be used in both cases as plasticity condition and also as a failure criterion. In the condition, function f depends on stress tensor σ_{ij} and strain rate tensor $\dot{\varepsilon}_{ij}$. In general, the function $f\left(\sigma_{ij}, \dot{\varepsilon}_{ij}\right) = 0$ can be spread on two parts and used in additive or multiplicative forms (Litoński 1977; Rusinek 2000). It is valid mainly for ductile metals and can be written in the following form:

$$g\left(\sigma_{ij}\right) \cdot h\left(\dot{\varepsilon}_{ij}\right) = 0. \tag{5}$$

In the case of function g, it is possible to use any quasi-static criterion, while for h it is necessary to increase the strength in tension TDIF and in compression CDIF according to the growing strain rates (Fig. 7).

Next, some criteria which can be used for brittle material in quasi-static loading are presented. Later, the strain rate sensitivity in the failure criterion or plasticity condition is discussed.

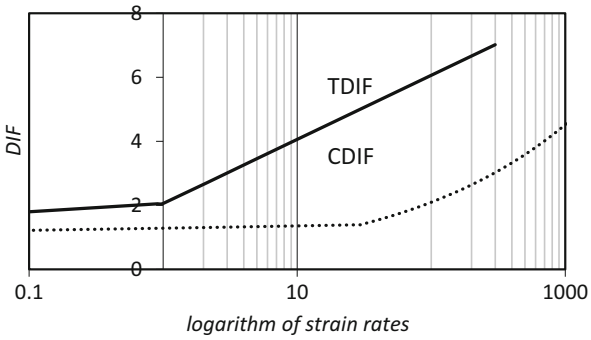


Fig. 7 DIF as a function of the strain rates in tension TDIF and in compression CDIF

Quasi-static Case

The failure criteria for quasi-static loading can be expressed in the following form:

$$f(I_1, J_2, J_3) = 0, \tag{6}$$

where I_1 is the first invariant of stress tensor σ_{ij} and J_2 and J_3 are the second and the third invariants of stress deviator s_{ij}, respectively. The rate of strain tensor $\dot{\varepsilon}_{ij}$ in Eqs. 4 and 5 is assumed to be zero.

The whole elastic energy density over the volume material unit W can be divided into two parts according to $W = W_1 + W_2$ (Gawęcki 1998). The first W_1 is connected with the change of volume (volumetric part), while the second one W_2 with the change of a shape (deviatoric part). Both elastic energy densities depend on stress and stress deviator invariants using the following relations:

$$W_1 = \frac{1 - 2v}{6E} I_1^2, \tag{7a}$$

$$W_2 = \frac{1 + v}{E} J_2. \tag{7b}$$

E i v are Young's modulus and Poisson ratio. Invariants I_1, J_2 are connected with the strain energy density of volume change W_1 and with strain energy density of distortion W_2, respectively.

The Burzyński (B) Energy Failure Criterion

This brilliant hypothesis which is almost a century old assumes that failure is conducted by the strain energy density of distortion as in the Huber-Mises condition increased by a part of volume change strain energy density (Burzyński 1928; Pęcherski 2008). The mathematical form of this assumption is expressed by the following equation:

$$\phi_f + \eta \phi_v = K, \tag{8}$$

where

$$\phi_f = W_2 = \frac{1+v}{E} J_2,$$

$$\phi_v = W_1 = \frac{1-2v}{6E} I_1^2, \tag{9}$$

are the strain energy densities of distortion and volume change, respectively. The last element is a function η, which is the matching function of the first invariant of stress tensor, and according to that proposed by Burzyński, the form is

$$\eta = \eta(W_1) = \omega + \frac{\delta}{3p} = \omega + \frac{\delta}{I_1} = \omega + \frac{\delta}{\pm\sqrt{\frac{6E}{1-2v} W_1}}, \tag{10}$$

where hydrostatic pressure p is

$$p = \frac{I_1}{3} \tag{11}$$

Equation 8 after taking into consideration dependences in Eqs. 9, 10, and 11 reaches the form

$$W_2 = K - \omega W_1 \mp \delta \sqrt{\frac{1-2v}{6E} W_1}, \tag{12}$$

where K, ω, δ are constants of the criterion. The sign "+" is applied if $I_1 \geq 0$ and "–" if $I_1 < 0$. Three constitutive parameters ω, δ, and K are identified by the three laboratory tests: uniaxial compression $\left(-f_c, 0, 0\right)$ and tension $\left(f_t, 0, 0\right)$ and biaxial uniform compression $\left(-f_{bc}, -f_{bc}, 0\right)$. Table 1 shows all the necessary data required to create the following system of equations:

$$\begin{cases} \dfrac{1+v}{E}\dfrac{f_t^2}{3} = K - \omega\dfrac{1-2v}{6E} f_t^2 + \delta\sqrt{\left(\dfrac{1-2v}{6E}\right)^2 f_t^2} \\ \dfrac{1+v}{E}\dfrac{f_c^2}{3} = K - \omega\dfrac{1-2v}{6E} f_c^2 - \delta\sqrt{\left(\dfrac{1-2v}{6E}\right)^2 f_c^2} \\ \dfrac{1+v}{E}\dfrac{f_{bc}^2}{3} = K - \omega\dfrac{1-2v}{6E} 4f_{bc}^2 - \delta\sqrt{\left(\dfrac{1-2v}{6E}\right)^2 4f_{bc}^2} \end{cases}. \tag{13}$$

By solving Eq. 13, the three necessary parameters K, ω, δ of failure surface are calculated. The last two columns in Table 1 are collected to construct the system of equation based on Eq. 12.

Table 1 Data for identification of the Burzyński criterion

Point	I_1	J_2	W_1	W_2
$(\sigma_1, \sigma_2, \sigma_3) = (f_t, 0, 0)$	f_t	$\frac{f_t^2}{3}$	$\frac{1-2v}{6E} f_t^2$	$\frac{1+v}{E} \frac{f_t^2}{3}$
$(\sigma_1, \sigma_2, \sigma_3) = (-f_c, 0, 0)$	$-f_c$	$\frac{f_c^2}{3}$	$\frac{1-2v}{6E} f_c^2$	$\frac{1+v}{E} \frac{f_c^2}{3}$
$(\sigma_1, \sigma_2, \sigma_3) = (-f_{bc}, -f_{bc}, 0)$	$-2f_{bc}$	$\frac{f_{bc}^2}{3}$	$\frac{1-2v}{6E} 4f_{bc}^2$	$\frac{1+v}{E} \frac{f_{bc}^2}{3}$

The solution of the system, Eq. 13 gives three constitutive parameters in the following form:

$$\begin{cases} \omega = -\dfrac{1}{3} \dfrac{a\left(-f_{bc}^2 - 2f_t f_{bc} + f_t f_c + 2f_c f_{bc}\right)}{b\left(-4f_{bc}^2 + 2f_c f_{bc} - 2f_t f_{bc} + f_t f_c\right)} \\ \delta = -\dfrac{a f_{bc}^2 (f_c - f_t)}{b\left(-4f_{bc}^2 + 2f_c f_{bc} - 2f_t f_{bc} + f_t f_c\right)} \\ K = \dfrac{a f_c f_t f_{bc}^2}{-4f_{bc}^2 + 2f_c f_{bc} - 2f_t f_{bc} + f_t f_c} \end{cases} \tag{14}$$

where

$$a = \frac{1+v}{E}, b = \frac{1-2v}{6E} \tag{15}$$

For example, assuming that $f_c = 30\,\text{MPa}$, $f_t = 3\,\text{MPa}$, and $f_{bc} = 33.6\,\text{MPa}$, the three constitutive parameters are obtained as $\omega = 1.1878$, $\delta = 140.069$, and $K = 0.001556$.

The Burzyński energy failure criterion (Pęcherski 2008; Jankowiak and Łodygowski 2010; Jankowiak 2011) in graphical form is presented in Fig. 8 (black continuous line). There are two visible lines, which consider invariant I_1 for opposite signs (in compressive and tensile zones). The curve in the Burzyński criterion passes through identification points, and the strain energy of distortion W_2 from the specifed place (maximum) decreases together with the increase of the volume change strain energy W_1. In addition, in Fig. 9, the shape of the Burzyński failure surface under plane stress conditions is presented (black continuous line). Figure 10 shows the shape of the failure surface in meridian plane. The following function describes the shape in meridian plane:

$$r = \sqrt{\frac{2}{a}\left(K - 3b\omega\xi^2 - \sqrt{3}b\delta\xi\right)}. \tag{16}$$

It is important to notice that the Burzyński failure criterion could be reduced to the other criteria. In fact, it is more general and includes some other criteria. Failure criteria are the specific limit functions of stresses and strains or the invariants of stresses and strains. These functions can be used as the plastic potential functions or

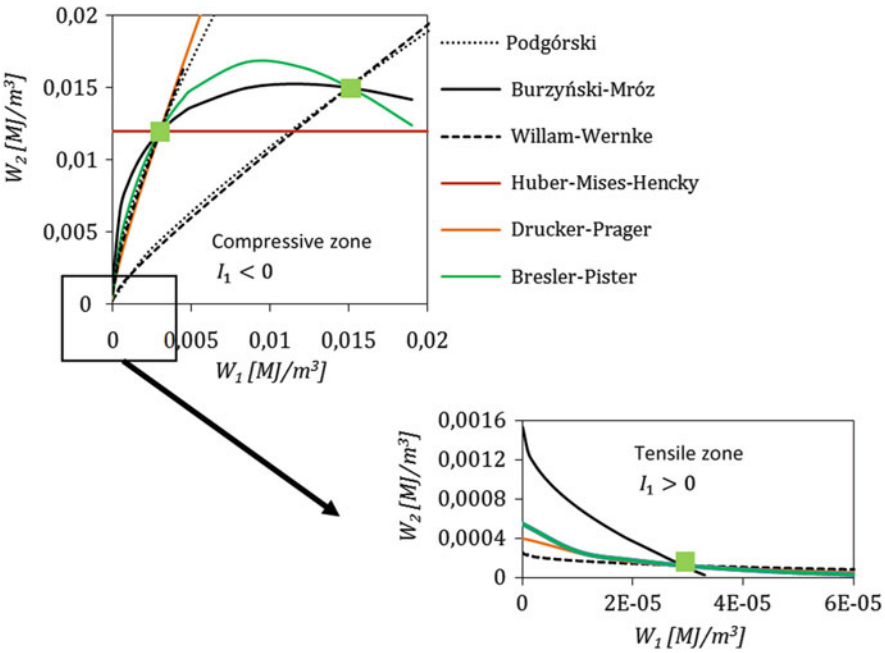


Fig. 8 Comparison of failure curves in space W_1–W_2

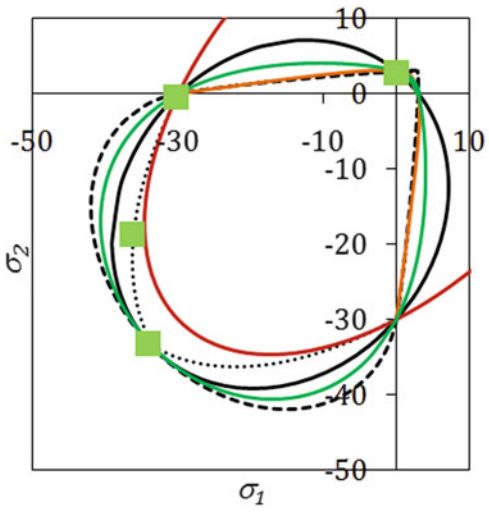


Fig. 9 Different failure criteria in plane stress conditions $\sigma_1 - \sigma_2$

the loading functions to describe the inelastic deformation of material. It is possible to describe the material effort criterion in energy form (Burzyński 1928). The collection of some representative criteria that depend on various numbers of parameters is presented below.

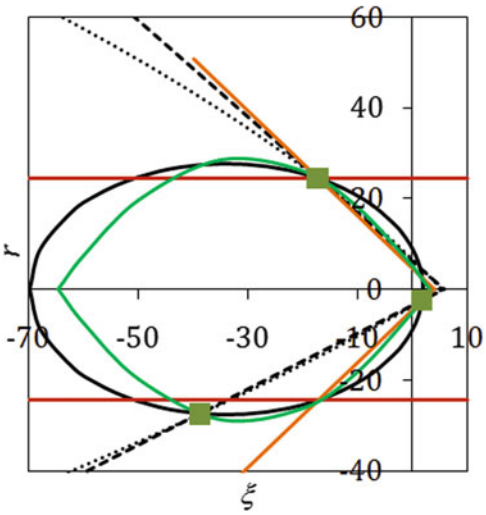


Fig. 10 Different failure criteria in meridian plane

The Huber-Mises-Hencky (HMH) Failure Criterion

The criterion assumes that the only responsible for failure of material is the strain energy of distortion (energy of deviatoric state) (Huber 1904; von Mises 1913). It is expressed in the following form:

$$f(J_2) = J_2 - k^2 = 0. \tag{17}$$

It assumes only one constitutive parameter k, which can be computed from the compressive strength of concrete f_c:

$$k = \frac{f_c}{\sqrt{3}}. \tag{18}$$

Taking into account Eq. 7, the strain energy of distortion, which decides on the material failure, is as follows:

$$W_2 = \frac{(1+v)}{E}\frac{f_c^2}{3} = 0.012\,\mathrm{MJ/m^3}. \tag{19}$$

In the space defined by the strain energy of distortion W_2 and the volume change of strain energy W_1, this criterion for the data, compressive strength $f_c = 30\,\mathrm{MPa}$, Young's modulus $E = 30e3\,\mathrm{MPa}$, and Poisson ratio $v = 0.2$, is presented in Fig. 8 as red line (parallel to W_1 axis). Additionally, the shapes of the curves which represent this condition in plane stress case and in meridian plane are shown in Figs. 9 and 10 as red lines.

The Drucker-Prager (DP) Failure Criterion

The criterion takes into consideration additionally the influence of the first invariant of stress state I_1 and the volume change strain energy density (Drucker 1959; Prager 1952). The criterion is written in the form

$$f(I_1, J_2) = mI_1 + \sqrt{J_2} - k = 0. \tag{20}$$

To identify the two constitutive parameters m and k, using the two points $(f_t, 0)$ and $(-f_c, 0)$ is necessary. It is required that the failure surface passes through both points. This condition is satisfied for the pair of numbers (m, k) in the form

$$\begin{cases} m = \dfrac{(f_c - f_t)}{\sqrt{3}(f_c + f_t)} \\ k = \dfrac{2f_c f_t}{\sqrt{3}(f_c + f_t)} \end{cases}. \tag{21}$$

Assuming $f_c = 30\,\text{MPa}$ and $f_t = 3\,\text{MPa}$, two parameters $m = 0.47$ and $k = 3.15$ are obtained. The strain energy of distortion which decides on the failure in DP criterion is dependent on the volume change strain energy according to the following expression:

$$W_2 = \begin{cases} \dfrac{1+v}{E}k^2 + 2mk\dfrac{1+v}{E}\sqrt{\dfrac{6E}{1-2v}W_1} + 6m^2\dfrac{1+v}{1-2v}W_1 & I_1 < 0 \\ \dfrac{1+v}{E}k^2 - 2mk\dfrac{1+v}{E}\sqrt{\dfrac{6E}{1-2v}W_1} + 6m^2\dfrac{1+v}{1-2v}W_1 & I_1 \geq 0 \end{cases}. \tag{22}$$

In Figs. 8, 9, and 10, the failure criterion in space $W_2 - W_1$, in plane stress conditions, and in meridian plane, for $f_c = 30\,\text{MPa}, f_t = 3\,\text{MPa}, E = 30e3\,\text{MPa}$, and $v = 0.2$, is shown as orange line.

The Bresler-Pister (BP) Failure Criterion

The BP criterion describes the strength in advanced state of stress and takes into account both the strain energy of distortion and the volume change strain energy density (Bresler and Pister 1958). It is presented in the following form:

$$f(I_1, J_2) = A + BI_1 + C(I_1)^2 - \sqrt{J_2} = 0. \tag{23}$$

This criterion depends on three independent parameters. For identification of these constitutive parameters $A, B,$ and C, it is necessary to use three

points defined by experimental tests. There are $(f_t, 0)$, $(-f_c, 0)$, and $(-f_{bc}, -f_{bc})$, where f_{bc} is the strength of concrete for biaxial uniform compression. The criterion passes through three identification points and defines the three parameters (A, B, C) in the form

$$\begin{cases} A = \dfrac{(f_c f_t f_{bc})(f_t + 3f_c + 8f_{bc})}{\sqrt{3}(f_c + f_t)(2f_{bc} - f_c)(2f_{bc} + f_t)} \\ B = \dfrac{(f_c - f_t)(f_{bc} f_c + f_{bc} f_t - f_t f_c - 4f_{bc}^2)}{\sqrt{3}(f_c + f_t)(2f_{bc} - f_c)(2f_{bc} + f_t)} \\ C = \dfrac{(3f_{bc} f_t - f_{bc} f_c - 2f_t f_c)}{\sqrt{3}(f_c + f_t)(2f_{bc} - f_c)(2f_{bc} + f_t)} \end{cases}. \tag{24}$$

Assuming $f_c = 30\,\text{MPa}$, $f_t = 3\,\text{MPa}$, and $f_{bc} = 33.6\,\text{MPa}$, the three constitutive parameters are of values $A = 3.6832$, $B = -0.6326$, and $C = -0.0059$. The strain energy of distortion as a function of the volume change strain energy is presented by the equation:

$$W_2 = \begin{cases} a\left(\begin{array}{r} A^2 - 2AB\sqrt{bW_1} - 2ACbW_1 + \\ 2BC\sqrt{(bW_1)^3} + B^2 bW_1 + C^2(bW_1)^2 \end{array}\right) I_1 < 0 \\ a\left(\begin{array}{r} A^2 + 2AB\sqrt{bW_1} + 2ACbW_1 + \\ 2BC\sqrt{(bW_1)^3} + B^2 bW_1 + C^2(bW_1)^2 \end{array}\right) I_1 \geq 0 \end{cases} \tag{25}$$

where

$$a = \frac{1+v}{E}, b = \frac{6E}{1-2v}. \tag{26}$$

In Figs. 8, 9, and 10, the curves of the Bresler-Pister failure criterion are presented in space $W_2 - W_1$, in plane stress conditions, and in meridian plane by green lines for comparison with the earlier failure criteria.

The Mróz (M) Failure Criterion

The three parameters failure criterion with ellipse shape in meridian plane were introduced by Mróz (Klisinski and Mróz 1988) in the form

$$f(I_1, J_2) = (I_1 - A)^2 + BJ_2 - C = 0. \tag{27}$$

The same identification points like in the cases of the Burzyński and the Bresler-Pister failure criteria were used to identify three constitutive parameters A, B, and C. The conditions are satisfied for the following three numbers (A, B, C):

$$\begin{cases} A = \dfrac{3}{2}\dfrac{(f_c - f_t)f_{bc}^2}{\Omega} \\ B = -3\dfrac{3f_{bc}^2 + \Omega}{\Omega} \\ C = \dfrac{\Omega\left[4\Omega f_t - 12f_t f_{bc}^2(f_c - 2f_t) - 4\Omega f_t^2\right] + 9(f_c - f_t)^2 f_{bc}^4}{4\Omega^2} \end{cases} \tag{28}$$

where

$$\Omega = (f_{bc} - f_c)(f_{bc} - f_t) - (f_c - f_t)f_{bc}. \tag{29}$$

Assuming as before $f_c = 30$ MPa, $f_t = 3$ MPa, and $f_{bc} = 33.6$ MPa, the three constitutive parameters are $A = -58.96$, $B = 10.10$, and $C = 3863.82$. The strain energy of distortion as a function of the volume change strain energy is presented by the equation:

$$W_2 = \begin{cases} \dfrac{a}{B}\left[C - \left(\sqrt{bW_1} - A\right)^2\right] & I_1 \geq 0 \\ \dfrac{a}{B}\left[C - \left(-\sqrt{bW_1} - A\right)^2\right] & I_1 < 0 \end{cases} \tag{30}$$

where a and b are identical as in Eq. 26.

The Mróz failure criterion in space $W_2 - W_1$, in plane stress conditions, and in meridian plane is presented in Figs. 8, 9, and 10 by black continuous lines. The Mróz criterion is identical with the Burzyński failure criterion for the same identification points, so in the future it will be called the Burzyński-Mróz failure criterion.

The Willam-Warnke (WW) Failure Criterion

The criterion introduced for concrete by Willam and Warnke (1975) in three-dimensional stress state is proposed in the form

$$f(I_1, J_2, J_3) = \frac{1}{3z}\frac{I_1}{f_c} + \sqrt{\frac{2}{5}}\frac{1}{r(\theta)}\frac{\sqrt{J_2}}{f_c} - 1, \tag{31}$$

where

$$r(\theta) = \frac{2r_c\left(r_c^2 - r_t^2\right)\cos\theta + r_c(2r_t - r_c)\sqrt{4\left(r_c^2 - r_t^2\right)\cos^2\theta + 5r_t^2 - 4r_c r_t}}{4\left(r_c^2 - r_t^2\right)\cos^2\theta + (r_c - 2r_t)^2}, \tag{32}$$

Table 2 Data for identification of the Willam-Warnke criterion parameters

Points	I_1	J_2	θ	$r(\theta)$
$(\sigma_1, \sigma_2, \sigma_3) = (f_t, 0, 0)$	f_t	$\frac{f_t^2}{3}$	0^0	r_t
$(\sigma_1, \sigma_2, \sigma_3) = (-f_c, 0, 0)$	$-f_c$	$\frac{f_c^2}{3}$	60^0	r_c
$(\sigma_1, \sigma_2, \sigma_3) = (-f_{bc}, -f_{bc}, 0)$	$-2f_{bc}$	$\frac{f_{bc}^2}{3}$	0^0	r_t

for

$$\theta = \frac{1}{3}\cos^{-1}\left(\frac{3\sqrt{3}}{2}\frac{J_3}{\sqrt[2]{J_2^3}}\right),$$

$$\cos\theta = \frac{2\sigma_1 - \sigma_2 - \sigma_3}{\sqrt{2\left[(\sigma_1-\sigma_2)^2 + (\sigma_2-\sigma_3)^2 + (\sigma_1-\sigma_3)^2\right]}}. \tag{33}$$

The identification of the three parameters r_c, r_t, and z can be performed based on the same points as before. The values of invariants I_1, J_2, and θ, which correspond to the three identification points $(f_t, 0, 0)$, $(-f_c, 0, 0)$, and $(-f_{bc}, -f_{bc}, 0)$, are collected in Table 2.

The three parameters are found in the form

$$\begin{cases} r_t = \sqrt{\frac{6}{5}}\left[\frac{f_{bc} f_t}{f_c(2f_{bc} + f_t)}\right] \\ r_c = \sqrt{\frac{6}{5}}\left[\frac{f_{bc} f_t}{3f_{bc} f_t + f_c(f_{bc} - f_t)}\right] \\ z = \frac{f_{bc} f_t}{f_c(f_{bc} - f_t)} \end{cases} \tag{34}$$

The failure surface passes through the three identification points. Assuming as before $f_c = 30\,\text{MPa}$, $f_t = 3\,\text{MPa}$, and $f_{bc} = 33.6\,\text{MPa}$, the three parameters have the values $r_t = 0.0524316$, $r_c = 0.090479$, and $z = 0.1098039$. The strain energy density of distortion W_2 as a function of the volume change strain energy density W_1 for the Willam-Warnke failure criterion is presented by the following equation:

$$W_2 = \begin{cases} a\left[\frac{5}{2}f_c^2 r^2(\theta) - \frac{5}{3}\frac{f_c r^2(\theta)}{z}\sqrt{bW_1} + \frac{5}{18}\frac{r^2(\theta)}{z^2}\right] & I_1 \geq 0 \\ a\left[\frac{5}{2}f_c^2 r^2(\theta) + \frac{5}{3}\frac{f_c r^2(\theta)}{z}\sqrt{bW_1} + \frac{5}{18}\frac{r^2(\theta)}{z^2}\right] & I_1 < 0 \end{cases}, \tag{35}$$

where a and b are identical as in Eq. 26.

In Figs. 8, 9, and 10, the curves of the Willam-Warnke failure criterion are presented by black dotted lines (in space $W_2 - W_1$, in plane stress conditions, and in

meridian plane). In this case, in the deviatoric plane, it is a curve that consists of parts of ellipses. It is necessary to use in Eq. 35 the proper value of $r(\theta)$, as in Table 2, according to Eq. 32, $r(\theta)$ changes between r_t and r_c.

The Podgórski (P) Failure Criterion

There exist many other multiparameter criteria. The final failure criterion which will be discussed here has five parameters and was proposed for advanced state of stresses by Podgórski (1984) in the form

$$f(I_1,J_2,J_3) = \frac{1}{3}I_1 - A + BP(J)\sqrt{\frac{2}{3}J_2} + \frac{2}{3}CJ_2 = 0, \tag{36}$$

for

$$P(J) = \cos\left(\frac{1}{3}\cos^{-1}DJ - E\right),$$

$$J = \cos(3\theta) = \frac{3\sqrt{3}}{2}\frac{J_3}{J_2^{3/2}}. \tag{37}$$

The identification of five parameters $A, B, C, D,$ and E is performed based on five identification points $(f_t, 0, 0)$, $(-f_c, 0, 0)$, $(-f_{bc}, -f_{bc}, 0)$, $(-f_{cc}, -1/2f_{cc}, 0)$, and (f_{tt}, f_{tt}, f_{tt}). The values of invariants for those points are collected in Table 3.

The system of equations, which should be solved to compute five material parameters, is as follows:

$$\begin{cases} \frac{1}{3}f_t - A + B\cos\left(\frac{1}{3}\cos^{-1}D - E\right)\frac{\sqrt{2}}{3}f_t + \frac{2}{3}C\frac{f_t^2}{3} = 0 \\ -\frac{1}{3}f_c - A + B\cos\left(\frac{1}{3}\cos^{-1} - D - E\right)\frac{\sqrt{2}}{3}f_c + \frac{2}{3}C\frac{f_c^2}{3} = 0 \\ -\frac{2}{3}f_{bc} - A + B\cos\left(\frac{1}{3}\cos^{-1}D - E\right)\frac{\sqrt{2}}{3}f_{bc} + \frac{2}{3}C\frac{f_{bc}^2}{3} = 0 \\ -\frac{1}{2}f_{cc} - A + B\cos\left(\frac{1}{3}\cos^{-1}0 - E\right)\frac{1}{\sqrt{6}}f_{cc} + \frac{2}{3}C\frac{f_{cc}^2}{4} = 0 \\ f_{tt} - A = 0 \end{cases} \tag{38}$$

The above system of equations (38) is solved iteratively by the Newton method. Assuming data $f_c = 30\,\text{MPa}$, $f_t = 3\,\text{MPa}$, $f_{bc} = 33.6\,\text{MPa}$, $f_{cc} = 36\,\text{MPa}$, and $f_{tt} = 3\,\text{MPa}$, the five constitutive parameters $A = 3$, $B = 1.4276$, $C = 0.0112$, $D = 1$, and $E = 0.03902$ are specified. The criterion has four identification points in plane stress conditions; that is why the shape in this plane (Fig. 9) fits the experimental data (Kupfer et al. 1969).

Table 3 Data for identification of the Podgórski failure criterion parameters

Points	I_1	J_2	J	$P(J)$
$(\sigma_1, \sigma_2, \sigma_3) = (f_t, 0, 0)$	f_t	$\frac{f_t^2}{3}$	1	$\cos\left(\frac{1}{3}\cos^{-1}D - E\right)$
$(\sigma_1, \sigma_2, \sigma_3) = (-f_c, 0, 0)$	$-f_c$	$\frac{f_c^2}{3}$	-1	$\cos\left(\frac{1}{3}\cos^{-1} - D - E\right)$
$(\sigma_1, \sigma_2, \sigma_3) = (-f_{bc}, -f_{bc}, 0)$	$-2f_{bc}$	$\frac{f_{bc}^2}{3}$	1	$\cos\left(\frac{1}{3}\cos^{-1}D - E\right)$
$(\sigma_1, \sigma_2, \sigma_3) = (-f_{cc}, -1/2f_{cc}, 0)$	$-\frac{3}{2}f_{cc}$	$\frac{1}{4}f_{cc}^2$	0	$\cos\left(\frac{1}{3}\cos^{-1}0 - E\right)$
$(\sigma_1, \sigma_2, \sigma_3) = (f_{tt}, f_{tt}, f_{tt})$	$3f_{tt}$	0	–	–

The strain energy of distortion which decides on the failure criterion is the following function of the volume change strain energy:

$$W_2 = \left(\frac{-BP(J)\sqrt{\frac{2}{3}a} + \sqrt{\Delta}}{\frac{4}{3}Ca}\right)^2, \tag{39}$$

where a and b are identical as in Eq. 26, while Δ is defined in the following way:

$$\Delta = \begin{cases} \frac{2}{3}aB^2P^2(J) - 4\left(\frac{2}{3}Ca\right)\left(\frac{1}{3}\sqrt{bW_1} - A\right) & I_1 \geq 0 \\ \frac{2}{3}aB^2P^2(J) - 4\left(\frac{2}{3}Ca\right)\left(-\frac{1}{3}\sqrt{bW_1} - A\right) & I_1 < 0 \end{cases}. \tag{40}$$

Dynamic Failure Criteria

The fast dynamic loadings lead (blasts, hitting) to damage or failure of a concrete structure in a very short time. When formulating the limit condition for plasticity, damage, or failure, it is necessary to consider the dependence on stress state (volumetric and deviatoric parts) as well as on strain rates according to experiments (Klepaczko and Brara 2001). It is also possible to use Eq. 5 with one of the quasi-static conditions (presented in section "Quasi-static Case" and in Figs. 8, 9, and 10) as a function $g(\sigma_{ij})$ and strain rate sensitivity according to the recommendations of CEB for TDIF and CDIF.

In this chapter, other types of criteria are presented. All of them have cumulative form and can be successfully used to describe the failure of the material under extreme loading (Tuler and Butcher 1968; Campbell 1953). Tuler, Butcher, and Campbell found the good correlation between the strength and the length of the load impulse. The failure criterion is presented in the following integral form:

$$\int_0^{t_c} \left(\sigma^{eq} - \sigma_0^{eq}\right)^{\lambda} dt = C, \tag{41}$$

where λ, σ_0^{eq} and C are the material constants and t_c is the time up to failure described by the criterion. The criterion describes t_c only if the intensity of stress σ^{eq} is higher than quasi-static strength σ_0^{eq}.

The other criterion but also in cumulative form was presented by Freund (1993) in the following form:

$$\int_0^{t_c} \left[\frac{\sigma^{eq}}{\sigma_0^{eq}} - 1 \right]^{\beta} dt = D, \tag{42}$$

where σ_0^{eq}, β and D are material parameters.

Here all the mentioned cumulative-type criteria describe the accumulation of the energy which has to be stored in material to destroy it. These criteria estimate the time for how long in the fast dynamic case the material can carry the state of stresses that is higher than its quasi-static strength. The next criterion of the same type was proposed by Campbell for ductile materials as metals (Campbell 1953). The cumulative failure criterion was used for brittle materials, for example, concrete (Stolarski 2004; Jankowiak 2011) and also glass (Jankowiak et al. 2013). The cumulative failure criterion was generalized to represent the behavior in three-axial state of stresses (Jankowiak and Łodygowski 2007). The criterion describes the time to failure for impulsive loadings and has the following integral form:

$$t_{c0} = \int_0^{t_c} \left(\frac{\sigma_F^{eq}(t)}{\sigma_{F0}^{eq}} \right)^{\alpha(T)} dt \text{ if } \sigma_F^{eq}(t) > \sigma_{F0}^{eq}. \tag{43}$$

In Eq. 43, t_{c0} is the critical failure time, $\alpha(T)$ is a parameter (can be a function of temperature) connected with energy activated during separation process (the breaking of the cohesion), and σ_{F0}^{eq} is the equivalent strength of material for quasi-static loading. It is possible to use any function presented for quasi-static cases. However, the following equivalent measure of stress σ_F^{eq} is used (Geers et al. 2000):

$$\sigma_F^{eq} = \frac{k-1}{2k(1-2v)} I_1 + \frac{1}{2k} \sqrt{\left(\frac{k-1}{1-2v} I_1 \right)^2 + \frac{6k}{(1-v)^2} J_2}, \tag{44}$$

where I_1 and J_2 are as before the first invariant of stress tensor and the second invariant of its deviator. The parameter k influences the shape of the function in the space of principal stresses, and it decides on the shifting of the limit condition in the space. The influence of the two parameters σ_{F0}^{eq} and k on the shape and the position of the limit surface in the space of principal stresses is presented in Fig. 11. Now, the limit surface can be expressed as follows:

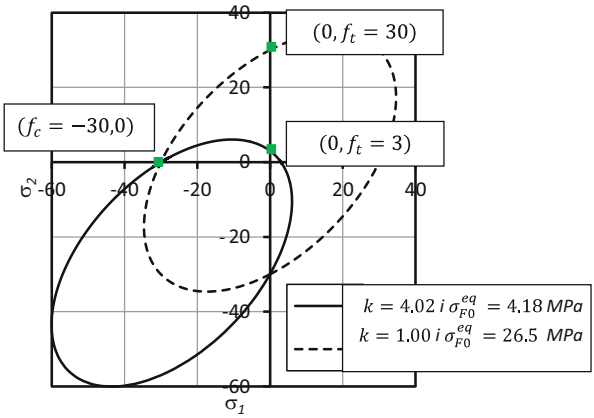


Fig. 11 Limit condition in plane stress state for different parameters k and σ_{F0}^{eq}

$$f(I_1,J_2) = \frac{k-1}{2k(1-2v)}I_1 + \frac{1}{2k}\sqrt{\left(\frac{k-1}{1-2v}I_1\right)^2 + \frac{6k}{(1-v)^2}J_2} - \sigma_F^{eq}. \tag{45}$$

For example, if the parameter $k = 1$, the classical Huber-Mises-Hencky condition is arrived (dotted line), while for parameter $k = 4.02$, the surface is shifted into the compression zone (continuous line).

In the meridian plane, the limit condition is presented in Fig. 12. The coordinates are defined by $r = \pm\sqrt{2J_2}$ and $\xi = \frac{1}{\sqrt{3}}I_1$. ξ is the distance along the hydrostatic axis, and r is the distance in the direction perpendicular to hydrostatic axis. Taking into account these formulae $r - J_2$ and $\xi - I_1$, Eq. 44 can be rewritten in the following form:

$$\sigma_F^{eq} = \frac{(k-1)\sqrt{3}}{2k(1-2v)}\xi + \frac{1}{2k}\sqrt{\left(\frac{(k-1)\sqrt{3}}{1-2v}\xi\right)^2 + \frac{6k}{(1-v)^2}\frac{r^2}{2}}. \tag{46}$$

Or it could be transformed to the form

$$r = \sqrt{\frac{4}{3}k\sigma_F^{eq}\frac{2(1-v)^2}{6k} - \frac{4}{3}\sqrt{3}\sigma_F^{eq}\frac{(k-1)(1-v)^2}{(1-2v)}\xi}, \tag{47}$$

for

$$\xi \le \xi_{\max} = \frac{k\sigma_F^{eq}(1-2v)}{(k-1)\sqrt{3}}. \tag{48}$$

The right side of the curves, Fig. 12, is limited by $\xi_{\max}$. It is the value for which r equals 0. If $k = 1$, the limit surface is reduced to HMH condition and could be drawn as the straight line parallel to the axis ξ. The other limit conditions are acceptable.

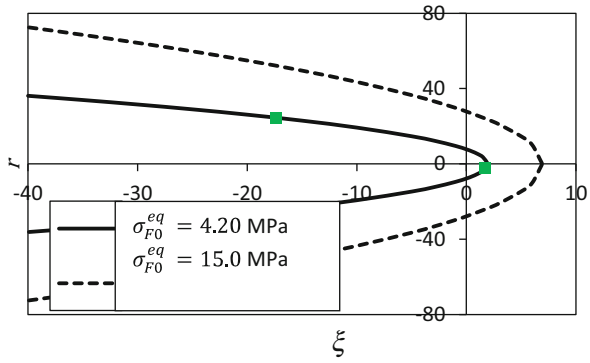
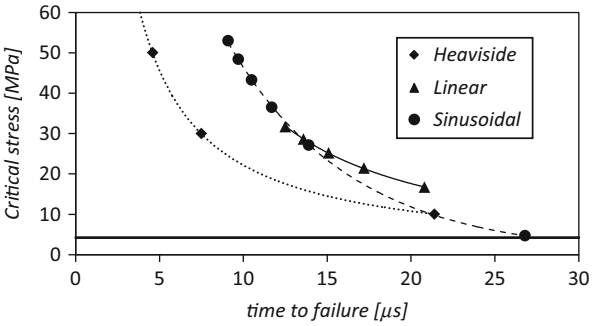


Fig. 12 The size of the limit condition for $k = 4.02$ in space $r - \xi$ (meridian plane)



t_{c0}	0.000049s
$\alpha(T)$	0.95
σ_F^{eq}	4.2e6Pa

Fig. 13 The comparison of the critical stress curves $\sigma_F^{eq} - t_c$ for different loading history

In the cumulative failure criterion, the stress level for different loading histories σ_F^{eq}(t) is described. The failure stress level depends on time, and generally for short impulse, the dynamic strength is higher. In Fig. 13, the influence of the loading history is presented. Three different shapes of the impulse are considered: heaviside, linear, and sinusoidal with different amplitudes (different rates of stress and strain). The comparison of the results obtained by cumulative failure criterion for linear strain rates and experimental results – black dots (Brara and Klepaczko 2006) – is presented in Fig. 14 for the strain rates between 20 s^{-1} and 120 s^{-1}. The agreement is acceptable for material parameters identified in Table 4.

The last plots present the influence of the material parameters used in the criteria σ_{F0}^{eq}, α, and t_{c0} on the estimation of the critical stress and time to failure (the curves are in the space $\sigma_F^{eq} - t_c$).

The influence of the quasi-static strength σ_{F0}^{eq} is presented in Fig. 15. The increasing of σ_{F0}^{eq} describes the higher level of the dynamic strength. Next, Fig. 16 shows the influence of α parameter which is connected with energy activated

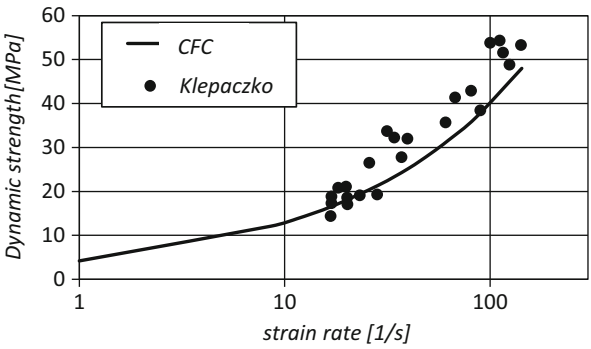


Fig. 14 Comparison of the strength taken from experiments and predicted by cumulative failure criterion

Table 4 Constitutive parameters of concrete

E	35 10^9[Pa]
ν	0.2
ρ	2,395 [kg/m^3]
k	4.03
t_{c0}	0.000049 [s]
α	0.95 [−]
σ_{F0}^{eq}	4.2e6 [Pa]

in separation process. It must be concluded that increasing of α decreases the dynamic strength. The influence of the parameter t_{c0} on the critical curve is presented in Fig. 17. The increase in critical failure time t_{c0} leads to the higher dynamic strength.

Examples

Two examples are presented in this chapter, and both describe the dynamic failure of concrete structures. The first one presents the verification of the failure criterion which is used to simulate the experiment performed by the spalling Hopkinson pressure bar (SHPB) test. The test was proposed by Klepaczko (Klepaczko 1990; Brara and Klepaczko 2006). The second example shows the perforation analysis of the pure and reinforced concrete slab.

Testing Concrete Specimen in the Hopkinson Bar

The brittle materials and also concrete show the increase of strength in tension and in compression for high strain rate loadings. In tension, the increment of the strength under deformation rates of order 100 (TDIF = 10) is higher than in compression (CDIF = 2.7). This fact is included into codes (CEB 1987). In everyday engineering practice, the fact that the concrete is able to carry the tensile

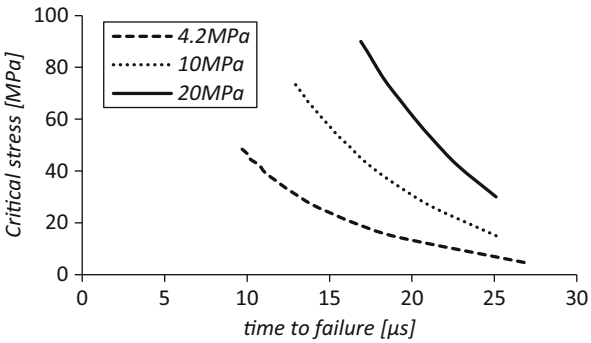


Fig. 15 The influence of σ_{F0}^{eq} on the shape of the critical curve

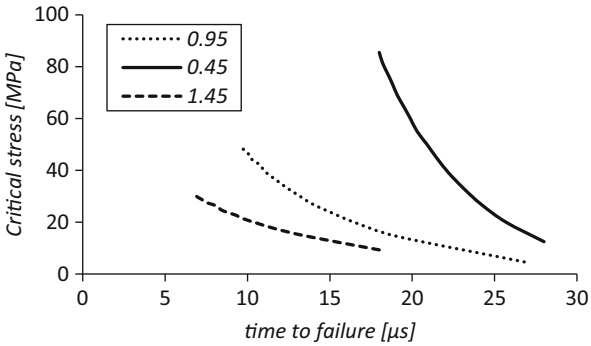


Fig. 16 The influence of α on the shape of the critical curve

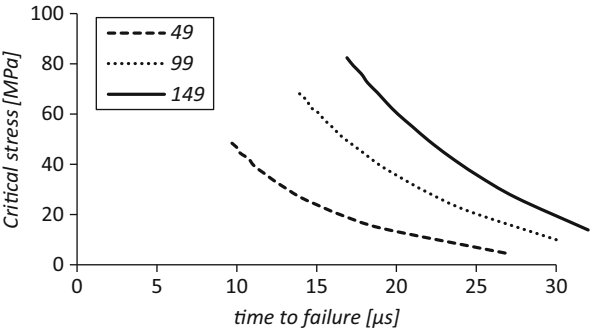


Fig. 17 The influence of t_{c0} on the shape of the critical curve

load is neglected, and in structures reinforcement is used to replace this gap. However, during fast impulses (impacts, blasts) in structures, the stress waves propagate and interact, and the spalling appears in tension zones.

The configuration of the SPHB experiment which is finally under analysis is shown in Fig. 18. The test predicts the tensile strength of the concrete using the wave analysis. In the experiment, it is necessary to describe the failure time of the concrete specimen. A similar test is also used for other brittle materials, for example, glass. The most important part is the aluminum bar (length 1 m and

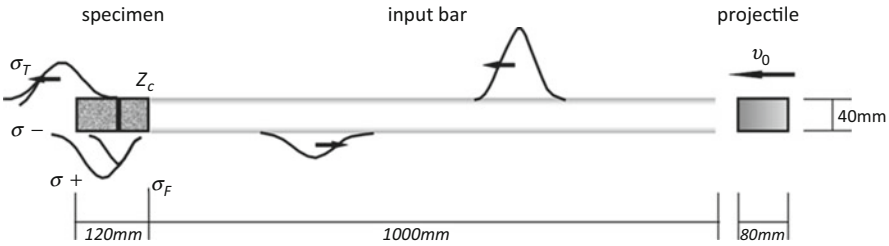


Fig. 18 Spalling Hopkinson pressure bar configuration

diameter 0.04 m). The bar is impacted by aluminum projectile (length 0.08 m and diameter 0.04 m). The length of the concrete specimen is 0.12 m, and its diameter is the same as the bar and projectile. All parts are fitted, and the concrete specimen is glued to the input bar. The aluminum is used because it has similar impedance to concrete.

After the impact of the projectile into the input bar, the compressive wave is transferred. The wave is propagated along the bar. Next, part of the wave is transmitted to the concrete specimen, and a part is reflected as tensile wave. Transmitted to the specimen wave is compressive, and it is important that its intensity must be smaller than the compressive strength of the concrete. After reflection from the free end, it propagates back as tensile wave throught the concrete specimen. If the dynamic tensile strength is reached, the spalling appears in the concrete specimen. In the experiment, the time when the spalling appears and the pattern of the failure can be measured and defined.

Abaqus/Explicit finite element code is used to simulate the process of failure. The axisymmetric model was analyzed (CAX4R finite element with characteristic length of 0.001 m) with contact between all three parts. The aluminum parts have Young's modulus of $E = 70 \times 10^9$ Pa, the Poisson ratio $v = 0.28$, and density $\rho = 2850$ kg/m^3. The material model of aluminum parts is elastic because both parts are working in this zone. In case of concrete, the elastic behavior was assumed with rate-dependent brittle failure. The constitutive parameters are the following: Young's modulus $E = 35 \times 10^9$ Pa, the Poisson ratio $v = 0.2$, density $\rho = 2395$ kg/m^3, asymmetry factor compression-tension $k = 4.03$, critical time $t_{c0} = 49$ μs, factor of the energy activated during the failure process $\alpha = 0.95$, and equivalent quasi-static tensile strength $\sigma_{F0}^{eq} = 4.2 \times 10^6$ Pa (see Table 4).

The results obtained in the laboratory experiments are presented in Fig. 19. There are selected frames from high-speed camera for two velocities of the projectile 7 and 12 m/s. For higher velocity, double cracking appears. The numerical simulation was performed to obtain the failure pattern. The corresponding numerical results are shown in Fig. 20 together with two frames of the longitudinal stress component before the failure (above) and after the failure (below) by tension. For the impact velocity 7 m/s, one crack appears, but for higher impact velocity 12 m/s, two cracks are predicted similarly as in experiments. In numerical model in the case of 7 m/s, the crack appears 55 mm from the impacted surface. In experiments, the

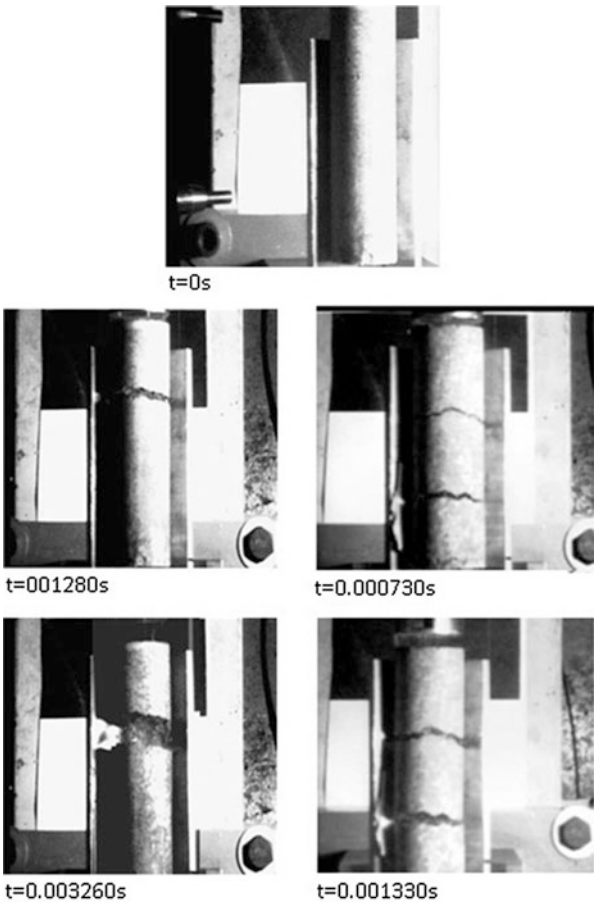


Fig. 19 The failure patterns for two impact velocities 7 and 12 m/s

cracking was observed in similar distance. For velocity 12 m/s, the position and the order of the crack appearance in computations are similar to the experiments. The first crack appears in the distance about 80 mm from the impacted side, and the second is in the distance of 55 mm from the impacted side. The predicted dynamic strength in the case of velocity 7 m/s is about 18 MPa but in the case of 12 m/s is 28 MPa. The numerical results fit to the experiments, so the cumulative failure criterion is confirmed.

Perforation Analysis of the RC Slab

The second numerical example describes the dynamic behavior of a pure concrete and reinforced concrete slabs subjected to impact, penetration, and perforation by the rigid projectile which hits them with different angles of incidence. The analyses are performed for pure concrete structure and reinforced by steel meshes

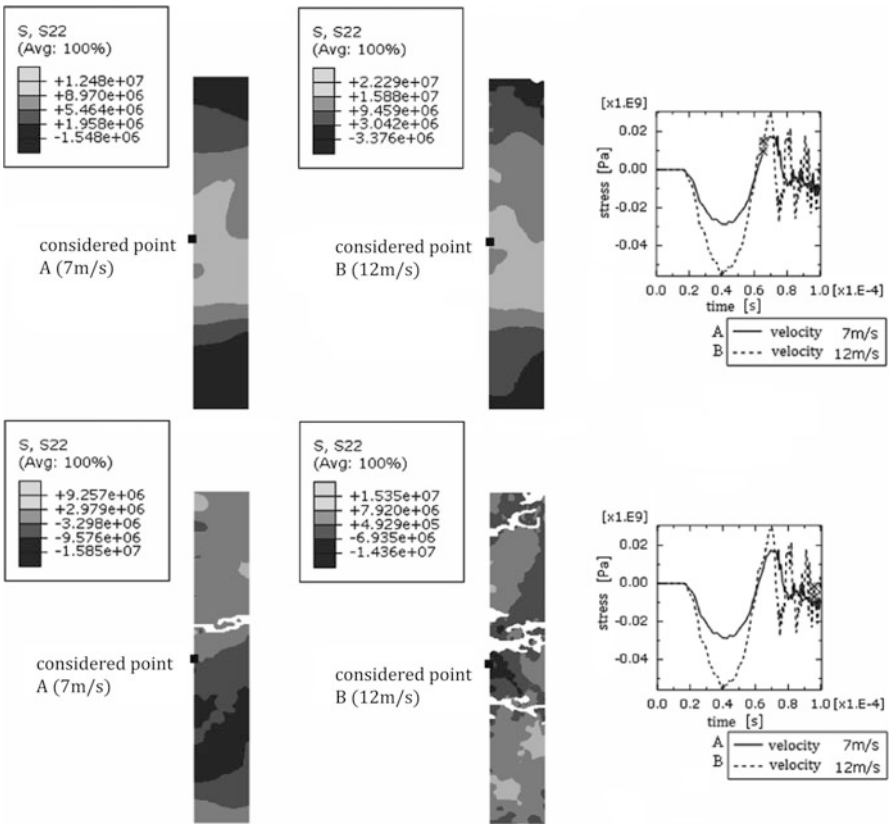


Fig. 20 Numerically obtained failure patterns for two impact velocities 7 and 12 m/s (half of symmetry due to axisymmetry; impact from the top side)

(see Figs. 21 and 22), The slab is square, and the reinforcement mesh is modeled as T3D2 finite elements (linear approximation of displacement). The concrete is discretized by solid element C3D8R (8 node bricks with reduced integration and hourglass control). The problem was modeled and solved using the 336400 elastic finite elements which are connected by 217635 elastic elements with cumulative failure criterion. In case of reinforced concrete slab, the reinforcement is glued into the solid concrete mesh. The projectile is assumed to be rigid with dimensions presented in Fig. 22.

The square slab has dimension 0.87 m and thickness 0.15 m (see Fig. 22). The distance between the bars in reinforcement mesh is 0.0434 m, and the cross-sectional area of a bar is 0.0001 m^2. Two layers of the reinforcement mesh are used in the distance of 0.1 m. The slab is fixed on perimeter. The direction of the impact is analyzed with angle of incidence (between 0° and 60°) and is presented in Fig. 22. The angle 0° means the perpendicular impact. In all presented numerical examples, the rigid projectile (mass=1.88 kg, length=0.22 m, and diameter=0.04 m)

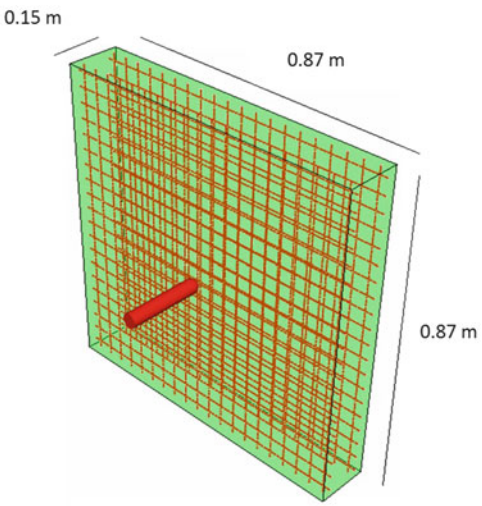
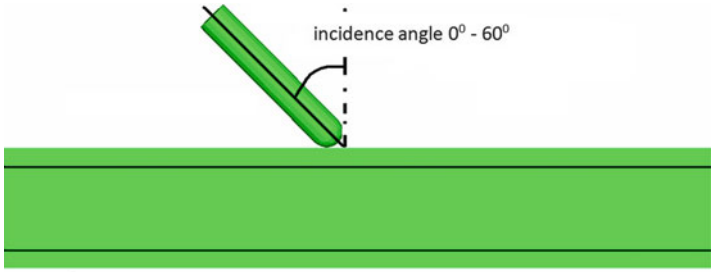


Fig. 21 The view of the reinforced concrete slab and of the projectile



Reinforcement:	
Bar area	0.0001 m²
Bar space	0.0434 m
Mesh distance (upper to lower)	0.1 m
Projectile:	
Diameter	0.04 m
Length	0.22 m
mass	1.88 kg
Velocity	200 m/s

Fig. 22 Geometry of the slab, reinforcement, and projectile

penetrates both the pure concrete and reinforced concrete slabs with initial velocity of 200 m/s. The modeling of the contact between the projectile, the slab, and the reinforcement was also important. Also the interior surfaces should be taken into account – these are created during concrete failure. The friction coefficient assumed in all analyzed cases equals 0.2.

All the simulations for different angles of incidence 0°, 15°, 30°, 45°, and 60° and for both types of structures were performed. In Figs. 23 and 24, the distribution of velocity in two time instances after the impact for 0° angle is presented. The

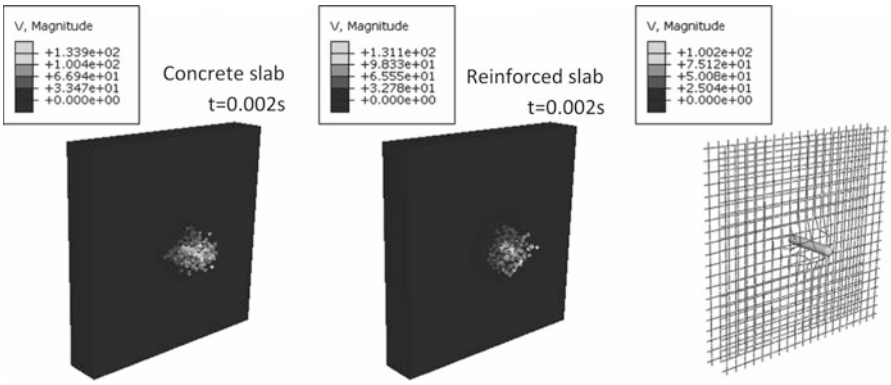


Fig. 23 The distribution of velocity at 0.002 s for 0° angle of incidence

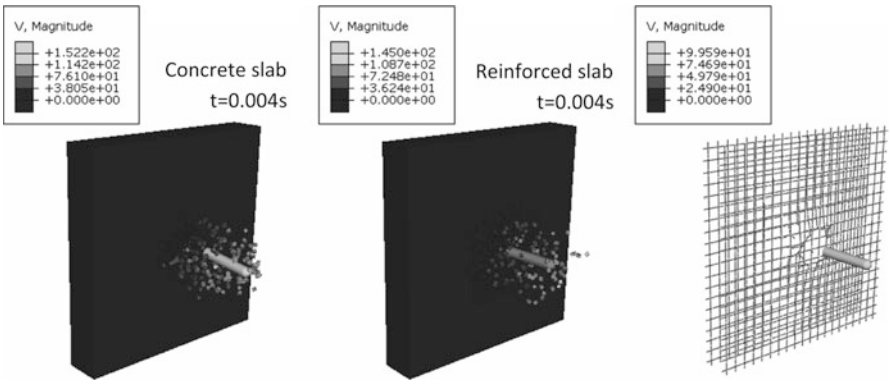


Fig. 24 The distribution of velocity at 0.004 s for 0° angle of incidence

reinforcement deformation is also presented. In case of pure concrete slab, the residual velocity of the projectile is 118 m/s and in reinforced slab is a bit slower 99 m/s.

The next presented results exhibit the influence of projectile in 45° angle of incidence (Fig. 25). This angle is measured between the line perpendicular to the slab and the trajectory of the projectile. The distribution of velocities for both types of structures is presented. For these two cases, full perforation is predicted, and the trajectory curvature is observed.

For the largest angle of incidence 60°, the perforation does not appear. The interaction between the projectile, concrete, and reinforcement induces the high trajectory curvature. The projectile reflects from the slab. In Fig. 26, the plot of the velocity time history for both structures is presented. In spite of the residual velocity, the curves are similar, but the failure of the concrete slab is much higher than the reinforced slab. However, there appears only penetration – not full

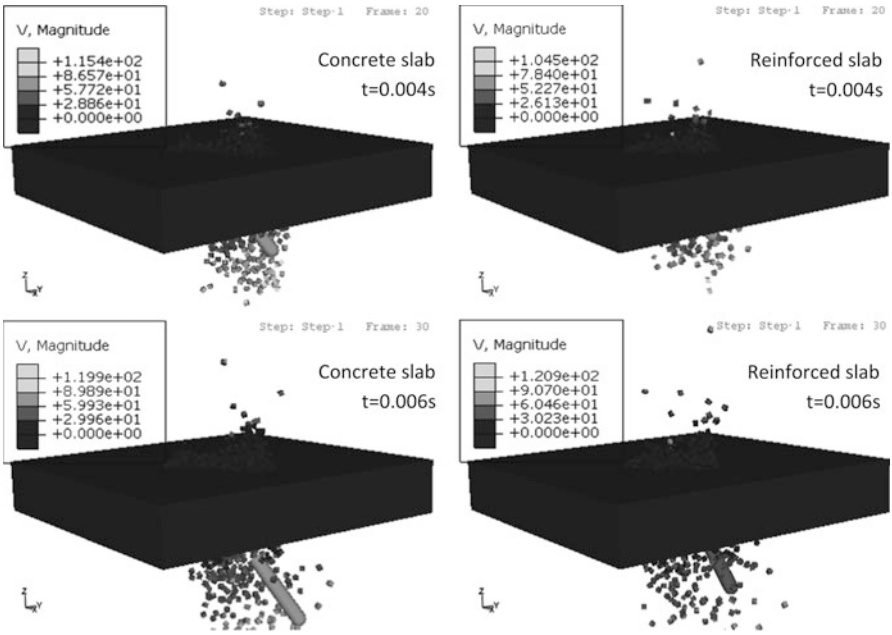


Fig. 25 The distribution of velocities for 45° angle of incidence

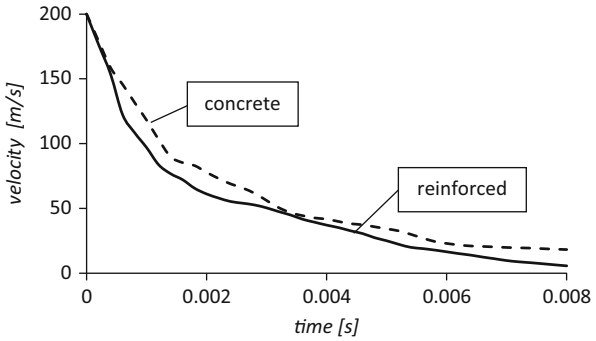
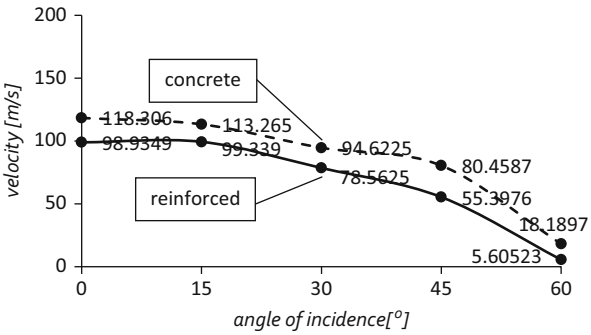


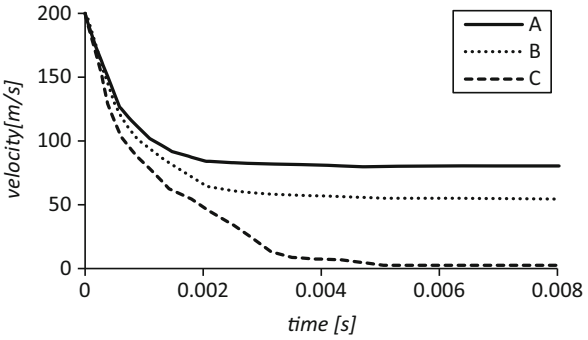
Fig. 26 The velocity time history for 60° angle of incidence

perforation. Figure 27 presents the residual velocity for all angles of incidence and for both structures. It is observed that the gap between the two curves is almost constant, and it represents the energy absorbed by the reinforcement.

The last case shows how it is possible to improve the structure strength by refinement of the reinforcement mesh (double) for 45° angle of incidence. The double reinforcement causes the stopping of the projectile inside the slab.

In Fig. 28, the deformation of the slab together with the velocity time history for three considered cases is shown. In case C, the projectile is stuck inside the slab. The other methods as smoothed particle hydrodynamics may be also used for this kind of application (Jankowiak and Lodygowski 2013).





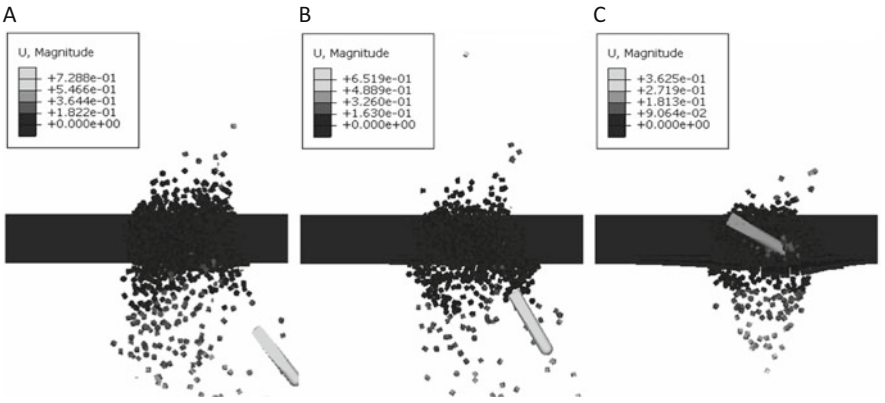


Fig. 27 The residual velocity of the projectile

Fig. 28 The deformation of the slabs for different structures; A, pure concrete; B, single reinforcement; and C, double reinforcement

Conclusions

The definition of the plasticity conditions and the failure criteria are important, in particular for modeling of static and dynamic behaviors of brittle materials and structures. The important aspect is to describe the strength of the material not only

in compression but also in tension and further definition of the strain rate sensitivity if it appears. Any numerical calculations, in particular those that describe fast dynamic processes, must be verified and confirmed by experiments which can prove the quality of the numerical models. The problem is very important because nowadays existing structures and of course future designed structures can be exposed to unique loading such as blast, impacts, and terrorist attacks.

References

A. Abrams, Effect of rate application of load on the compressive strength of concrete. Proc. Amer. Soc. Test. Mater. **17**, 364–377 (1917)

P. Bischoff, S. Perry, Compressive behaviour of concrete at high strain rate. Mater. Struct. **24**, 425–450 (1991)

A. Brara, J. Klepaczko, Experimental characterization of concrete in dynamic tension. Mech. Mater. **38**, 253–267 (2006)

B. Bresler, K. Pister, Strength of concrete under combined stresses. J. Am. Conc. Inst. **55**, 321–345 (1958)

W.T. Burzyński, *Study About the Strength Criteria (in Polish)* (Academy of Technical Sciences in Lwow, Lwów, 1928)

J. Campbell, The dynamic yielding of mild steel, Acta Metall. **1**, 706–710 (1953)

CEB, *Concrete Structure Under Impact and Impulsive Loading* (CEB, Lausanne, 1987)

W. Chen, *Plasticity in Reinforced Concrete* (McGraw-Hill, New York, 1982)

D.C. Drucker, A definition of stable inelastic materials. J. Appl. Mech. , 101–106 (1959)

L. Freund, Dynamic fracture mechanics (Cambridge University Press, Cambridge, 1993)

A. Gawęcki, *Mechanic of Material and Bar Structures (in Polish)* (Publishing House of Poznan University of Technology, Poznan, 1998)

M. Geers, R. De Borst, R. Peerlings, Damage and crack modeling in single-edge and double-edge notched concrete beams. Eng. Fract. Mech. **65**, 247–261 (2000)

M.T. Huber, Właściwa praca odkształcenia jako miara wytężenia materiału (in Polish) (Politechnique Association, Lwów, 1904)

T. Jankowiak, T. Łodygowski, Cumulative fracture criterion for concrete, in: *17th Int. Conference on Computer Methods in Mechanics*, Częstochowa. (2007)

T. Jankowiak, *Quasi-Static and Dynamic Failure Criteria of Concrete (in polish)* (Publishing House of Poznan University of Technology, Poznan, 2011)

T. Jankowiak, T. Lodygowski, Quasi-static failure criteria for concrete. Arch. Civil Eng. **LVI**, 2 (2010)

T. Jankowiak, T. Lodygowski, Smoothed particle hydrodynamics versus finite element method for blast impact. Bull. Pol. Acad. Sci. **61**, 111–121 (2013)

T. Jankowiak, A. Rusinek, T. Lodygowski, Validation of the Klepaczko–Malinowski model for friction correction and recommendations on split Hopkinson pressure bar. Finite Elem. Anal. Des. **47**, 1191–1208 (2011)

T. Jankowiak, A. Rusinek, P. Wood, Comments on paper: "Glass damage by impact spallation" by A. Nyoungue et al., Materials Science and Engineering A (2005) 407, 256-264". Mater Sci Eng A **564**, 206–212 (2013)

J. Klepaczko, Behavior of rock like materials at high strain rates in compression. Int. J. Plast. **6**, 415–432 (1990)

J. Klepaczko, A. Brara, An experimental method for dynamic tensile testing of concrete by spalling. Int. J. Impact Eng. **25**, 387–409 (2001)

M. Klisinski, Z. Mróz, *Description of Inelastic Deformation and Concrete Damage (in Polish)* (Publishing House of Poznan University of Technology, Poznan, 1988)

H. Kupfer, H.K. Hilsdo, H. Rusch, Behavior of concrete under biaxial stresses. ACI J. , 656–666 (1969)
J. Litoński, Plastic flow of a tube under adiabatic torsion. Biuletyn Polskiej Akademii Nauk **25**(1), 1–8 (1977)
A. Neville, *Własciwosci betonu* (Polski Cement, Kraków, 2000)
F. Oluokun, Prediction of concrete tensile strength from compressive strength: evaluation of existing relations for normal weight concrete. ACI Mater. J. **3**, 302–309 (1991)
R.B. Pęcherski, Burzyński yield condition vis-A-vis the related studies reported in the literature. Eng. Trans. **56**(4), 311–324 (2008)
J. Podgórski, Limit state condition and the dissipation function for isotropic materials. Arch. Mech. Mater. **36**(3), 323–342 (1984)
S. Popovics, *Strength and Related Properties of Concrete* (Wiley, New York, 1998)
W. Prager, Soil mechanics and plastic analysis or limit design. Q. Appl. Math. **10**(2), 157–165 (1952)
J. Raphael, Tensile strength of concrete. ACI Mater. J. **2**, 158–165 (1984)
C. Ross, J. Tedesco, S. Kuenen, Effects of strain rate on concrete strength. ACI Maer. J. **1**, 37–47 (1992)
A. Rusinek, Modelisation thermoviscoplastique d'une nuance de tole d'acier aux grandes vitesses de deformation. Etude experimentale et numerique du cisaillement, de la traction et de la perforation, Ph.D. thesis, University of Metz, Metz, 2000
A. Stolarski, Dynamic strength criterion for concrete, J. Eng. Mech. 1428–1435 (2004)
F. Tuler, B. Butcher (1968), A criterion for the time dependence of dynamic fracture, Int. J. Fract. Mech. **4**, 431–437 (1968)
R. von Mises, Mechanik der Festen Korper im plastisch deformablen Zustand. Nachr. Math. Phys **I**, 582–592 (1913)
K. Willam, W. Warnke, Constitutive models for the triaxial behavior of concrete. Int. Assoc. Bridge Struct. Eng. Proc. **19**, 1–30 (1975)

Section IX

Experimental Damage Characterization

Evaluating Damage with Digital Image Correlation: A. Introductory Remarks and Detection of Physical Damage

37

François Hild and Stéphane Roux

Contents

Abstract

The present chapter is devoted to the evaluation of damage with digital image correlation (DIC). This measurement technique provides 2D or 3D displacement fields that can be used to study damage mechanisms through the correlation residuals or even tune material parameters of damage models. Different aspects related to the use of full-field measurements will be illustrated in the context of damage mechanics. The main issues associated with damage measurements (i.e., definition of damage, difficulty of measuring damage, and image correlation) are introduced. Then, applications related to damage detection via image correlation are discussed. Two-dimensional surface measurements and even three-dimensional fields in the bulk of loaded materials are considered.

F. Hild (✉) • S. Roux
LMT-Cachan, ENS Cachan/CNRS/PRES UniverSud Paris, Cachan Cedex, France
e-mail: hild@lmt.ens-cachan.fr; stephane.roux@lmt.ens-cachan.fr

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_24

Damage Measurements

The reason for measuring damage is threefold. First, as will be illustrated in section "Detection and Evaluation of Physical Damage," the knowledge of the damage mechanisms is a priori information that is very useful when modeling the material response. Second, there is a need for the identification of parameters associated with damage laws for prediction purposes (see "► Chap. 38, Evaluating Damage with Digital Image Correlation: B. From Physical to Mechanical Damage" and "► Chap. 39, Evaluating Damage with Digital Image Correlation: C. Applications to Composite Materials"). Third, the evaluation of the damage state of a structure in use and its evolution are more and more frequently required in the field of civil and mechanical engineering.

The title of the present introduction refers to the seminal paper of Profs. Lemaitre and Dufailly (1987). Eight different methods were discussed to "measure" damage:

- *"Micrography"* or fractography performed in particular by using scanning electron microscopy (SEM). This technique is still extensively used to determine damage mechanisms (Mills 1991) but also as a quantitative tool (Chermant et al. 2001; Coster and Chermant 2001). SEM pictures can also be used to study damage via digital image correlation (Tasan et al. 2012).
- *Density measurements* are one of the early techniques to assess the porosity in damaged materials (Maire 2003). Nowadays, *porosity distributions* are also assessed by resorting to X-ray tomography (Baruchel et al. 2000; Babout et al. 2001). Similarly, SEM observations are utilized to estimate porosity distributions and/or locations. One key aspect is then to infer the *volume* of pores from 2D cuts. In all cases, image processing tools are needed to distinguish the solid phase(s) from the voids (Coster and Chermant 2001) and stereology to relate 2D (or 1D) observations to 3D characteristics.
- *Young's modulus* variations are one of the most popular damage indicators used to describe the gradual or sudden degradation in various classes of materials (Lemaitre and Chaboche 1978; Lemaitre et al. 1979). It is based on the coupling between elasticity and damage, and as discussed below, it needs very large-scale observations.
- *Ultrasonic waves* and acoustic emission are used as nondestructive techniques to assess the damage state of a material or a structure (Hellier 2001). One of the key aspects is related to the detection and the type of damage that is observed. It is to be stressed that depending on the wavelength of ultrasound as compared to the size of the microscale features responsible for damage, either the individual micromechanisms or the map of loss of stiffness is accessible.
- *Cyclic stress amplitudes* are another way of characterizing the onset and development of damage in fatigue studies (Lemaitre 1992). In that case, the coupling between plasticity and damage is considered.
- *Tertiary creep*, the first mechanism to be described by a damage variable (Kachanov 1958; Rabotnov 1963), is very frequently modeled within the

framework of continuum damage mechanics (CDM). With the years, the current damage models include more and more mechanism-based knowledge and microscopic features that call for new experimental ways of investigation to fine-tune material-dependent parameters.

- *Microhardness* measurements are less popular, even though they have recently received some renewed interest (Tasan et al. 2010). It was shown that the change of the hardness can be interpreted in terms of damage (Lemaitre et al. 1987) by using the coupling between plasticity and damage.
- *Electrical resistance* and its variation with damage is another way of characterizing the damage state of materials such as composites (Prabhakaran 1990; Abry et al. 1999). It is the counterpart of the so-called potential drop technique used to determine the crack length (Beevers 1982).

Many, if not all, of the previous techniques are still utilized today. However, one key change (some even talk about a revolution) is related to the fact that, since the turn of the century, full-field measurements are more and more developed and have now reached a degree of maturity that allows the users to measure displacement, temperature, lattice strain, phase, and grain orientation *fields* on the surface or even in the volume of samples (Sharpe 2008). The analysis of the measured fields usually leads to inverse problems in which the relevant properties are determined by inversion or identification. These procedures need strong coupling between experimental and simulation tools.

Among the various full-field measurement techniques (Rastogi 2000; Rastogi and Hack 2012), digital image (and volume) correlation (DIC and DVC) will be considered herein. The interested reader will find additional details and illustrations in the recent book collectively written by a French group of scientists dealing with full-field measurements and identification in solid mechanics (Grédiac and Hild 2012).

In all the introductory remarks, the term "damage" was used in a loose sense. It actually covers very different meanings and realities depending on the type of measurement used to assess the degree of degradation of a very small volume (or surface) element or that of a whole structure. The next section aims at defining damage as considered herein and in this chapter and in "▶ Chap. 38, Evaluating Damage with Digital Image Correlation: B. From Physical to Mechanical Damage" and "▶ Chap. 39, Evaluating Damage with Digital Image Correlation: C. Applications to Composite Materials."

Damage

To avoid any ambiguities, it is proposed in this chapter to use consistently a convention such that "mechanical damage" or in short "damage" often denoted as $D(\mathbf{x})$ will refer to the macroscopic relative loss of stiffness treated as a continuum field. It is worth remembering that this is one among many different choices, some of them listed above. With this meaning, damage can only be characterized through

the evaluation of local elastic properties. This description is limited to the largest scales of description and does not inform on the mechanisms by which the tangent stiffness is reduced, namely, microcracking or/and microvoidage.

Those mechanisms, related to the formation of microcavities or microcracks, are globally referred to as "physical damage," since they are distinctive features that can be detected, observed, and measured. However, their quantitative relationship with the (macroscopic) mechanical damage is difficult to establish and usually involves a modeling step.

The connection between physical and mechanical damage relies on a procedure called homogenization (Bornert et al. 2008). The latter reduces rich and detailed information (i.e., physical damage) to a single average effect (i.e., mechanical damage). CDM has shown over the past a remarkable ability to describe trustfully the behavior of a very broad range of materials, showing that indeed the most important feature was retained and that finer information (describing, for instance, the fluctuations of stress or strain within a representative element of volume) was irrelevant in most cases (Lemaitre and Desmorat 2005). The propensity of local defects to nucleate, grow, arrest, or coalesce does not preclude the validity of macroscopic descriptions in general.

However, homogenization rests on the key ingredient, namely, the representative volume element, and thus, at least implicitly, a length scale, ξ, is attached to the very definition of damage. Mechanical damage (with the restrictive meaning proposed above) only exists at scales of the order of and larger than ξ. In contrast, physical damage occurs at much smaller length scales since it designates the micromechanisms associated with mechanical damage. However, as no length scale enters the formulation of a local damage constitutive law, it is generally not emphasized. As macro-stress or macro-strain gradients become appreciable over the length scale, ξ, an enriched description may be called for, such as nonlocal (Pijaudier-Cabot and Bažant 1987) or gradient-based (Peerlings et al. 1996) damage models.

One difficulty (among others) is that the length scale, ξ, is naturally expected to grow with the damage magnitude (Hild et al. 1994; Forquin and Hild 2010), a point seldom considered from a theoretical standpoint. A second difficulty is that very few studies aim at identifying internal length scales for nonlocal or gradient-based models with full-field measurements (Geers et al. 1999). The identification of internal lengths is still in its infancy. Experimentally, it is also difficult to properly control an experiment under such circumstances. Furthermore, as strain and damage fields localize, "point" measurements become meaningless since an average value over a given gauge length (of an extensometer or strain gauge) is only a crude indication of the spatiotemporal features of displacement and strain fields when localized. The study of localized phenomena does call for spatially and temporally resolved kinematic fields.

Finally, even within the framework of CDM, i.e., without resorting to a microscopic representation of the origin of damage, the mechanical behavior

may display an instability, strain or damage localization, such that a spatially distributed damage over large length scales is no longer the unique solution, and therefore, small distances have to be accounted for properly. At a given stage, it is shown that there is a loss of uniqueness associated with the mechanical problem (Hill 1978), and the possibility of occurrence of localized modes in ductile (Billardon and Doghri 1989) and quasi-brittle composites (Hild et al. 1992). In this case, the unavoidable existence of a length scale below which damage is no longer an appropriate description (it is rather an illegitimate extrapolation from the large wavelength to the small ones) is recalled through apparent paradoxes that emerge if such length scales are forgotten. This issue, where different (and legitimate!) choices such as fracture mechanics (Kanninen and Popelar 1985) are available, will be discussed in "▶ Chap. 39, Evaluating Damage with Digital Image Correlation: C. Applications to Composite Materials."

Challenges for Measuring Damage

In the eight methods listed by Lemaitre and Dufailly (1987) and briefly summarized above, if postmortem analyses are excluded, two classes can be distinguished, namely, "physical" or "mechanical" descriptions of damage:

- First, nondestructive techniques such as ultrasonic waves (e.g., measurement of time of flight or time of flight diffraction) are used to detect cracks provided their size is larger than the wavelength of the probing ultrasound. One key aspect is related to the detection itself and in particular, to the corresponding resolution (Hellier 2001). Computed microtomography (μCT) also allows for the direct detection of physical damage (i.e., microcracks or microvoids). For instance, it is possible to monitor damage by resorting to ex situ or even in situ mechanical tests (Buffière et al. 2010). Further, the number of microcracks or microvoids, their size distribution and their center-to-center distance distribution and the crack opening distribution are now accessible in the bulk of many different materials (Baruchel et al. 2000; Salvo et al. 2003), allowing for a full statistical characterization. Potentially, quantities such as the size of the representative element of volume or the direct evaluation of internal length scales at the onset of localization should be accessible with such techniques.
- Second, indirect measurements are conducted by evaluating changes associated with physical and mechanical properties (e.g., Young's modulus, electrical or heat conductivity, yield stress, cyclic hardening parameters). The link to damage is obtained via couplings (e.g., stiffness loss is related to the coupling between elasticity and damage).

The first class corresponds to physical characterizations of damage where the connection to mechanical properties of the volume element formulation relies on various modeling strategies, for instance, using homogenization techniques. Conversely, the last class deals directly with the sought material properties. However, its physical meaning can be remote, and a quantitative link is often very uncertain. These facts constitute one of the difficulties associated with damage characterization and measurements. The complete equivalence between physical and mechanical description of damage still remains a challenge.

By keeping these remarks in mind, it may be concluded that there is no unique way of characterizing damage. This is true. However, measuring damage is directly related to modeling damage. Again, there is no unique modeling of damage (see other chapters in the present book). What should be aimed for is to have measuring techniques that allow the damage variable and parameters of its growth law to be identified and possibly validated. This is not always an easy task since it depends on the scale of modeling, on the type of information that needs to be extracted from the experiments, and on the robustness of the numerical tools used in simulations.

Digital Image Correlation

Digital image correlation (DIC) is a photomechanical technique that uses image registration algorithms (e.g., the correlation product) to measure 2D and 3D displacements in pictures (Sutton et al. 2009; Hild and Roux 2012). The raw outputs of correlation codes are 2D or 3D displacement *fields*. For the 3D displacements, either surface shape and deformations are obtained by resorting to stereocorrelation or kinematic data in the bulk by using digital volume correlation (DVC) applied to volumes obtained by computed tomography (e.g., X-ray μCT and optical coherence tomography) or magnetic resonance imaging. Strain fields are subsequently obtained by evaluating mean displacement gradients.

Among the different full-field measurement techniques (Rastogi 2000; Rastogi and Hack 2012), digital image correlation occupies a special place as it deals with pictures that can be obtained by different imaging devices (e.g., standard cameras, SEM, transmission electron microscopes (TEM), atomic force microscopes (AFM), infrared cameras). It makes DIC a tool of choice for multiscale measurements for which the physical size of one pixel ranges from values of the order of 0.1 nm (by using an AFM) to more than 10 m (for satellite images).

In the field of DIC, there are two different types of algorithms. First *local* approaches consist of maximizing the correlation between two interrogation windows (i.e., a small zone of interest or ZOI) taken one in the picture f of the reference configuration and another one g of the deformed configuration. The output of the correlation procedure is the mean displacement of each considered ZOI assigned to its center and a correlation score to evaluate the level of success associated with the registration (Sutton et al. 2009). More recently, *global* approaches were introduced in which the registration is performed at the level of the region of interest (ROI) by

minimizing the global correlation residual Φ_c (e.g., the sum of squared differences between f and g corrected by the sought displacement $\mathbf{u}$)

$$\Phi_c(\mathbf{u}) = \int_{\text{ROI}} \rho(\mathbf{x})^2 d\mathbf{x} \quad \text{with} \quad \rho(\mathbf{x}) = f(\mathbf{x}) - g(\mathbf{x} + \mathbf{u}(\mathbf{x})) \tag{1}$$

with respect to the set of unknown degrees of freedom u_i when the displacement field is decomposed over a given kinematic basis

$$\mathbf{u}(\mathbf{x}) = \sum_i u_i \boldsymbol{\mu}_i(\mathbf{x}) \tag{2}$$

where $\boldsymbol{\mu}_i$ are the trial displacement *fields* that can be defined over the whole ROI (Rayleigh-Ritz approach) or as in finite element procedures (Galerkin approach). The quality of the registration is characterized globally by Φ_c, which is the sum of squared correlation residuals ρ computed for each pixel belonging to the ROI (Hild and Roux 2012). The residuals can therefore be used at the *pixel* level as will be exemplified hereafter.

Only global approaches to DIC will be used herein. The main reason is that identification techniques, which will also be discussed, are generally based upon finite element formulations and simulations. Consequently, the same kinematic hypotheses are made at the measurement and at the identification levels. No reprojection (or filtering) will be needed. In the present chapter, different types of elements are used, namely, 4-noded (Q4) elements with bilinear displacement interpolations (Besnard et al. 2006), 3-noded (T3) elements with linear displacement interpolations for 2D analyses (Leclerc et al. 2009), and 8-noded (C8) elements for 3D analyses (Roux et al. 2008). In all these approaches, ℓ will refer to the size of an edge of the used element.

It is also possible to incorporate a mechanical regularization to the DIC functional so as to measure spatially resolved displacement fields with lower measurement uncertainties (Roux et al. 2012). In this case, the correlation functional Φ_c is supplemented by the equilibrium gap functional Φ_m (Claire et al. 2004) written for an elastic medium

$$\Phi_m(\mathbf{u}) = \{\mathbf{u}\}[\mathbf{K}]^t[\mathbf{K}]\{\mathbf{u}\} \tag{3}$$

where $[\mathbf{K}]$ denotes the stiffness matrix associated with inner nodes, which should have zero nodal forces (i.e., $[\mathbf{K}]\{\mathbf{u}\} = \{\mathbf{0}\}$), and $\{\mathbf{u}\}$ the displacement vector collecting all nodal displacements. In "▶ Chap. 39, Evaluating Damage with Digital Image Correlation: C. Applications to Composite Materials," the equilibrium gap method will be used to determine a stiffness distribution and parameters of damage growth laws. This regularization only concerns inner nodes and traction-free boundaries. For the other boundaries, another functional, Φ_b, is introduced (Tomičević et al. 2013).

The three functionals introduced previously are considered jointly as a weighted sum. The weights however define length scales. To make them apparent, a plane wave $\mathbf{v}$ of wavelength λ is used to normalize the different functionals, and the total functional to be minimized is written as

$$\Phi_t(\mathbf{u}) = \frac{\Phi_c(\mathbf{u})}{\Phi_c(\mathbf{v})} + \left(\frac{\ell_m}{\lambda}\right)^4 \frac{\Phi_m(\mathbf{u})}{\Phi_m(\mathbf{v})} + \left(\frac{\ell_b}{\lambda}\right)^4 \frac{\Phi_b(\mathbf{u})}{\Phi_b(\mathbf{v})} \tag{4}$$

where ℓ_m and ℓ_b are (resp. bulk and edge) regularization lengths. They act in a way comparable to the element size ℓ when FE discretization is the only regularization ingredient.

Detection and Evaluation of Physical Damage

With the above definition of physical damage, measurements of porosity and microcrack distributions are *direct* connections to the degradation of materials. In that case, procedures close to those utilized in nondestructive evaluation or structural health monitoring are considered. The first question is the detection of a given defect, and the second is its quantification in terms of size and location.

For microvoids, the easiest way of characterization is to resort to image processing techniques for diameters exceeding the physical size of one voxel (or one pixel). The voxel size is therefore a natural cutoff in the evaluation of the distribution of porosity. For sub-voxel sizes, mechanical-based analyses are the only hope to tackle this problem. However, measuring displacements in the sub-voxel range is possible but with element sizes that are at least one order of magnitude larger than the voxel (Réthoré et al. 2011). Voxel-scale DVC (or pixel-scale DIC) is possible provided a mechanical regularization is added to the correlation procedure (Leclerc et al. 2011, 2012).

Cracks (and microcracks in particular) are a feature that can be detected and quantified by DIC or DVC (Roux et al. 2009). This is due to the fact that they induce displacement jumps across the crack length or surface. For mode I cracks, the pictures themselves can be used for openings greater than the pixel or voxel size as they induce a change of gray levels. Conversely, mode II and mode III cracks are more difficult to detect since they are not necessarily accompanied by gray-level changes. When using displacement measurements, any mode of cracking can be analyzed since displacements can be quantified in all three cases. Different aspects can be addressed for microcracks, namely, their volume density, their size, their orientation, and their openings. Depending on the type of modeling, either or all of the above quantities need to be assessed. Another aspect is related to damage mechanisms associated with the detection of cracks (e.g., matrix-cracking in brittle matrix composites, aggregate/cement paste debonding in concrete, mesodelamination in layered composites).

The following examples shed some light on microstructural features (i.e., physical damage) responsible for mechanical damage at a macroscopic scale.

The detection and measurement of microcracks, their location, orientation, extension, or opening are typical questions to address where the resolution of DIC is challenged. The next section describes different strategies that can be considered within DIC to reach this objective.

Strategies

Kinematic Analyses

A microcrack is a crack, and hence in mechanical terms, it is the support of a localized discontinuity in the displacement field. The most natural way to detect it is to seek within the measured displacement field for a discontinuity. However, the measured displacement field is generally discretized, and hence, a discontinuity has not a clear-cut definition. Such a discontinuity will reveal through a large apparent strain within the element crossed by the crack. Consequently, the spatial resolution of the displacement field should be good enough to detect a strain concentration from its surrounding. Thus the region of interest should be finely meshed, but this fineness has a cost in terms of displacement (and hence strain) uncertainty. A very dense distribution of microcracks that cannot be resolved will result in an enhanced strain for the same level of stress or equivalently, it will appear as a reduction in local stiffness.

Correlation Residuals

A second strategy consists of the study, not of the displacement field itself, but rather of the correlation residuals. If the very nature of the displacement field is not well treated because of a priori assumptions, then the quality of image registration should reveal it clearly. More precisely, if the displacement field was assumed to be continuous, then it would not account for the displacement discontinuity occurring along the crack, and hence in the correlation residuals, a large error should become visible especially on the crack mouth. This second way of addressing the problem is quite illuminating for finding the crack position and detailed geometry, as the residual field can be computed at the pixel level. However, little information is provided on, say, the crack opening as the correlation residual is evaluated in terms of gray-level difference and not in terms of displacements. This may constitute the first step of a more sophisticated algorithm where, once cracks are identified, their kinematics (i.e., the displacement discontinuity) can be incorporated in the DIC analysis. In this spirit, an eXtended-DIC scheme, replicating the enrichment technique of the extended finite element method (XFEM), can be devised (Réthoré et al. 2008).

Sampling of Local Enrichment

A third strategy goes one step further along this direction. A microcrack has generally a localized influence as compared to the surrounding medium. Typically, the displacement perturbation caused by a microcrack of length a will affect a volume of extension a in all space directions. Generally, an anticipation of this

localized change in the displacement is available (it may be only schematic). Thus after a first (coarse) DIC analysis where microcracks are ignored, a possible enrichment may be considered where such a single microcrack displacement field would be proposed to a DIC algorithm to check whether or not it would allow for a better description of the actual kinematics. Thus, it is extremely appealing to resort to such a prospective enrichment as it would consist of a single degree of freedom DIC, performed for one single step, and hence involving only a minute additional computational task. As this degree of freedom would involve a discontinuity of the displacement across the crack, the estimated amplitude, A, of this degree of freedom could automatically be converted into a microcrack opening. Moreover, the decrease in residuals, $\Delta\Phi_c$, is also a natural output of this elementary computation, which is to be repeated for all crack locations and sizes (Rupil et al. 2011).

The following examples discuss practical implementations of these methodologies for various materials and scales.

Cracking in a Layered Composite

Composite materials form a wide class of materials where damage is an essential concept, as these materials are by design susceptible to different forms of microcracks or debonding, but in such a way that the phase arrangement leads to crack arrest, and thus at a macroscopic scale a sustainable damage. "► Chap. 39, Evaluating Damage with Digital Image Correlation: C. Applications to Composite Materials" will come back in details to model damage in various types of composite materials.

To illustrate the present approach of crack detection, 0/90 composites provide a good example of the wealth of information that can be obtained on the individual events responsible for damage. The material considered hereafter consists of a layered composite made of 3 plies of long unidirectional carbon fibers in an epoxy matrix. The plies are oriented along the (0°, 90°, 0°) direction. A uniaxial tensile test along the 0 direction is analyzed. Different scales of observation are chosen, namely, a zoom to capture the details of a single event and a larger field of view, where the ambition is to resolve and count the microcrack density. The former set of images will be discussed in section "Types of Damage and Constitutive Laws" of "► Chap. 39, Evaluating Damage with Digital Image Correlation: C. Applications to Composite Materials," while the larger scale is discussed here.

The ability of this composite material to sustain the presence of transverse cracks in different plies implies that a dense distribution of such cracks will take place as the material is loaded in tension. Thus at a more macroscopic scale, the material can be described within damage theory. This requires however to have a good statistical sampling of the density of microcracks. Capturing this phenomenon with DIC now requires a large field of view and hence fewer pixels to reveal each of these microcracks. Moreover, the crack opening is also smaller and hence more difficult to capture. Figure 1 shows such a large field of view.

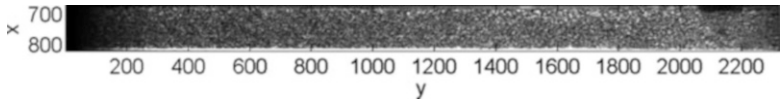


Fig. 1 Side view of a layered composite under tension (horizontally) considered at a macroscopic scale. The physical size of the pixel is 15.1 μm. The central ply is only about 500 μm (or 33-pixel) thick (Courtesy of P. Aimedieu and B. Leboime)

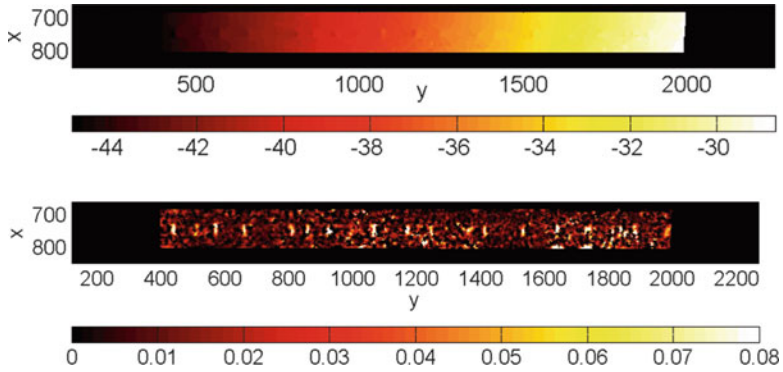


Fig. 2 (*top*) Horizontal displacement (in pixel) and (*bottom*) volumetric strain showing a quasi-periodic cracking pattern. The physical size of the pixel is 15.1 μm. The area of the ROI is 1.5×24.2 mm^2

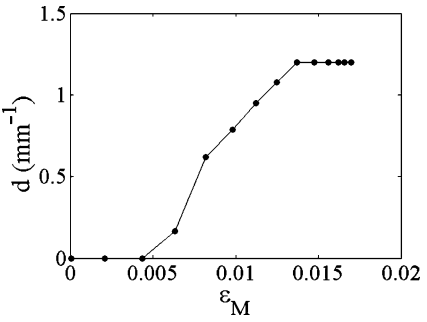


Fig. 3 Microcrack density (per unit length) in the central ply as a function of the mean longitudinal strain

DIC was performed with a very fine T3 mesh ($\ell \approx 5$ pixels) and a small regularization length ($\ell_m = 20$ pixels; $\ell_b = \ell_m/2$). Figure 2 shows the horizontal displacement and volumetric strain fields that allow meeting the previous requirements. Note that the large displacement is an additional difficulty to achieve a good convergence of the DIC algorithm.

The crack density (number of transverse cracks per unit length) can thus be estimated as shown in Fig. 3 as a function of the mean extension of the specimen, and their spatial separation is observed to be quite regular as could be anticipated from shear-lag-type models (Curtin 1993).

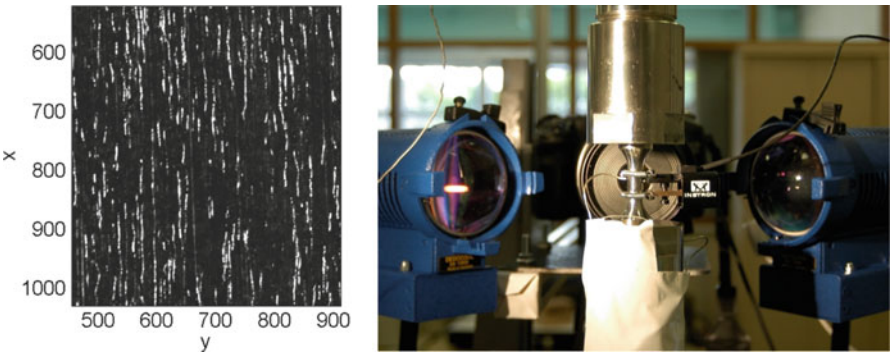


Fig. 4 Image of the microstructure of the 304L stainless steel surface observed during the test (*left*). Note that no paint has been used, and the revealed raw microstructure is used for the DIC analysis. The physical size of one pixel is 3.2 μm, and the area of the picture is 1.4×1.6 mm^2. Experimental setup (*right*) (Courtesy of J. Rupil and L. Vincent)

Fatigue Test on Stainless Steel

Fatigue of a 304L stainless steel is studied in this section. A flat gauge zone is machined and polished, on a cylindrical. The shallow notch creates local stress heterogeneities on the surface to localize microcrack initiations. The notched area is mechanically polished. A final electropolishing is performed resulting in a good quality of the surface finish, revealing the ferritic phase (elongated inclusions) in an austenitic matrix consisting of grains of an average diameter of 40 μm. No specific marking is applied onto the surface. Figure 4 shows the raw microstructure, which is not the best suited for analyzing transverse (i.e., horizontal) cracks. The specimen is cycled by sine-wave loading (5 Hz frequency) in a servohydraulic fatigue testing machine. The stress amplitude is controlled with a load ratio $R = -1$. The total strain is measured by an extensometer (gauge length: 10 mm). The images are captured with a digital single-lens reflex camera (Canon EOS 5D) and a macrolens (Canon MPE65) with a magnification of $\times 5$. Two lighting sources were needed to make the microstructure visible (Fig. 4).

DIC analysis was performed here using a regularized approach with a regularization length of $\ell_m = 40$ pixels ($\ell_b = \ell_m/2$) and a very fine mesh (T3 elements with $\ell \approx 4$ pixels). The displacement and strain maps shown in Fig. 5 reveal quite clearly a number of microcracks in spite of their small opening (1 μm or less) and limited spatial extension (roughly 150 μm). However, the DIC analysis here assumes continuity of the displacement, and the regularization smears out the displacement discontinuity.

The second strategy of considering correlation residuals, as shown in Fig. 6, allows for a finer spatial resolution as these residuals are computed at the pixel level. Although the above procedure can be considered as satisfactory, it requires a high-resolution DIC analysis, which is time-consuming. All along the fatigue test, the number of microcracks that initiated was quite large, and their characteristics

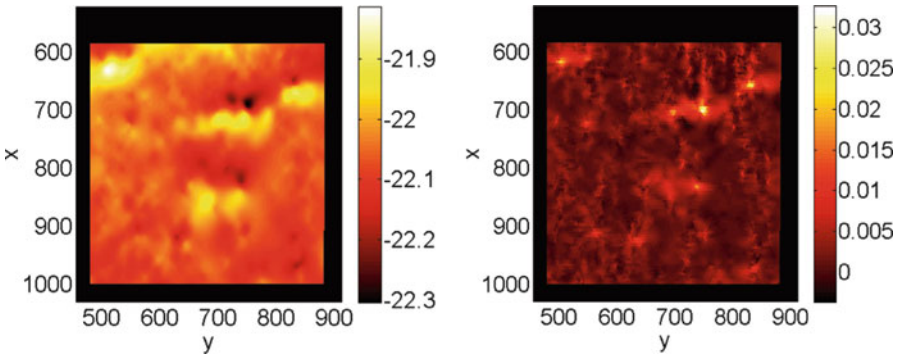


Fig. 5 (*left*) Vertical component of the displacement field (note the small overall dynamic range of 0.5 pixel or about 1.6 μm) and (*right*) corresponding maximum (apparent) strain map. The brightest spots correspond to microcracks. Their opening is of the order of 1 μm. The physical size of the ROI is $\approx 1.4 \times 1.4$ mm^2

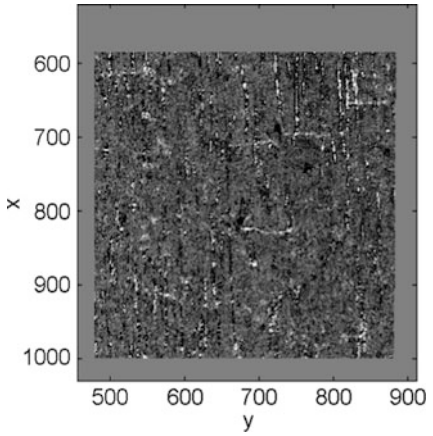


Fig. 6 Residual map after image registration (*white* is null; *black* is the largest *gray*-level difference). Note that the zones of highest strains in the previous figure can be seen as horizontal marks that are the traces of the microcracks here resolved at the pixel level. The physical size of one pixel is 3.2 μm

are to be analyzed in a statistical framework to validate the modeling robustness with respect to the particular sampling of the observation zone. Thus, the repeated task of analyzing a large number of large images, with such a small resolution becomes soon limiting. It is for addressing such a difficulty that the third strategy of sampling local enrichment of the displacement basis within DIC was designed.

Without discussing the details of the DIC procedure (Rupil et al. 2011), Fig. 7 shows a one-to-one comparison of the displacement field as it can be estimated first from a classical global DIC procedure with a very small Q4 element size (4 pixels) and second with the sampling of enrichment technique. The overall agreement is good (note that the overall range of displacement is only one pixel for a 1,500 pixel long image). Figure 8 compares the change of the crack density with the number of cycles with a manual analysis and with the fully automated procedure. A good agreement is obtained to be within the detection uncertainties.

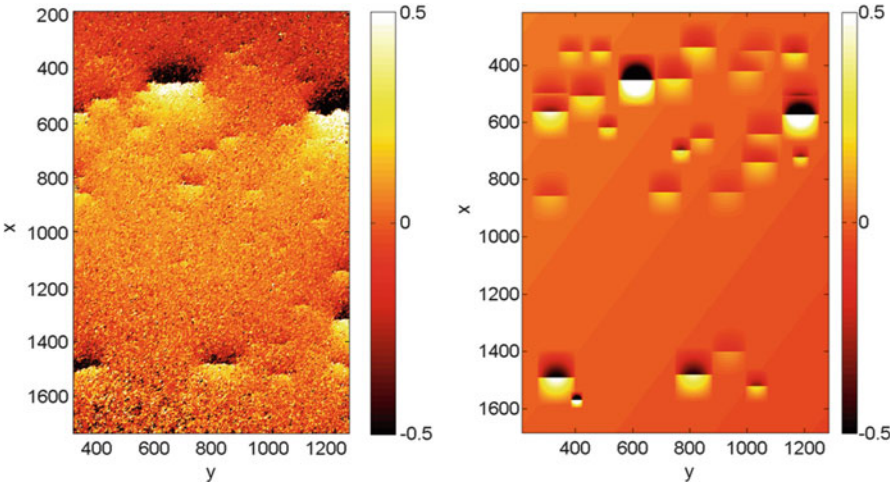


Fig. 7 Comparison between the classical global DIC analysis of the displacement field (*left*) with a very fine regular square mesh of quadrangular elements (4 pixel wide) and (*right*) the procedure of individual probing of a microcrack test field. The physical size of one pixel is 3.2 μm, and the area of the ROI is $\approx 4.8 \times 3.2$ mm^2 (Courtesy of J. Rupil)

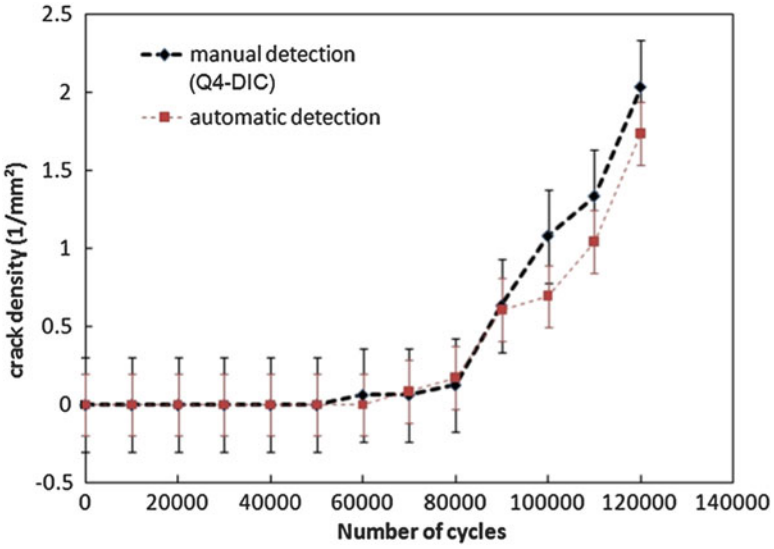


Fig. 8 Quantitative evaluation of the crack density based on a manual detection based on Q4-DIC displacement field (♦) or automated enrichment sensitivity (■) (Courtesy of J. Rupil)

Whatever the strategy used, the important message is that the DIC toolbox can provide a very fine information in terms of microcrack density, which in the present case, has been validated against replica measurements, a very accurate and reliable technique that is however so work demanding that it limits its use for a statistical analysis.

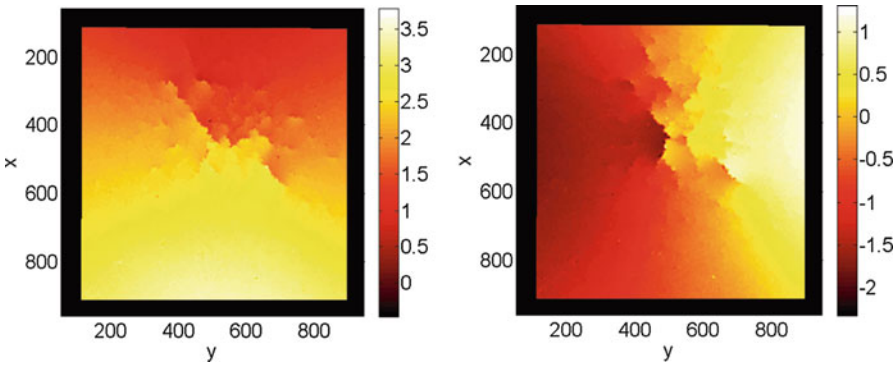


Fig. 9 Vertical (*left*) and horizontal (*right*) displacement fields (in pixels) resolved after 7,000 cycles of a biaxial test. The physical size of one pixel is 6 μm, and the area of the ROI is $\approx 4.8 \times 4.8\ \text{mm}^2$

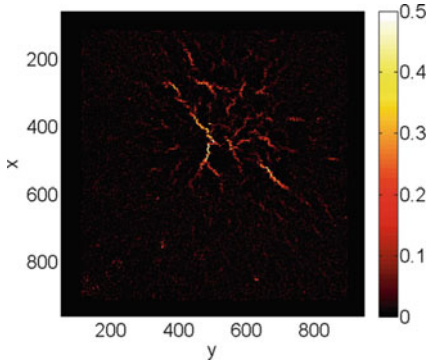


Fig. 10 Volumetric strain at the same stage as in the previous figure

2D Crack Networks

In the previous examples, the type of cracks was expected in terms of orientation (uniaxial tensile tests) and of size for the first example (constrained by the ply thickness). However, there are situations where the problem is more difficult. The same material as in the previous example is considered, 304L stainless steel, but subjected to a cyclic biaxial loading. The specimen is cross-shaped and its central part is thinned so as to concentrate the stress and control the region in which damage develops.

Again, DIC was performed using a very fine triangular mesh ($\ell \approx 5$ pixels) regularized with a small length of 10 pixels. As can be seen in Figs. 9 and 10, displacement and strain fields are successfully retrieved. Although the displacement field reveals a complex pattern of discontinuities, it becomes very difficult to appreciate cracks whose opening is very small. The same conclusion holds from the reading of the strain maps, or the residual field. Although resolving the cracks as individual objects with all their detailed geometrical characteristics (e.g., size, shape, roughness, branching) is a challenging and interesting question, a large

part of this information cannot be exploited or plays no role in a coarse-grained picture. In the example of Fig. 9, when the overall behavior of the sample is considered, only a mean loss of stiffness in the central region will matter. A microcrack that does not open to the level of detectability is unlikely to play any significant part in the present stress/strain response. It may grow and at a later stage become the initiation site of a macrocrack that will ultimately lead to failure, but this escapes from the realm of (standard) CDM.

3D Analyses

Up to now, only the pictures acquired by a single camera have been used to assess cracking under various conditions thanks to 2D-DIC (i.e., 2D displacements of the monitored surface are measured). With the development of 3rd generation synchrotrons and even more recently lab tomographs, it is possible to have access to 3D-reconstructed volumes during ex situ or even in situ mechanical tests (Buffière et al. 2010) by resorting to X-ray tomography. These volumes can subsequently be correlated through DVC to determine 3D displacements in the bulk of various materials. The two main damage mechanisms, namely, void nucleation and growth (Babout et al. 2001; Salvo et al. 2003; Maire et al. 2012), and microcrack initiation and propagation (Hild et al. 2011) can be quantified. Coalescence can also be analyzed (Maire et al. 2012), but it will not be discussed herein.

Analysis of Damage in High Strength Steel

Dual phase (DP) steels are commonly used in the automotive industry. Because of the presence of the (hard) martensitic phase in the ferritic matrix, high strength levels are observed. However, martensite can also promote damage because of its brittleness. The understanding and modeling of damage is needed to further improve the performance of DP steels. In situ tests are one way of analyzing the three steps of damage development (Maire et al. 2008). Figure 11 shows the distribution of porosities by thresholding the reconstructed scan. It is possible to follow the change of the number of porosities with the mean longitudinal strain in the neck. The latter was determined by assuming a uniform displacement per cross section and plastic incompressibility, so that tracking the boundary could provide an approximate strain level.

Analysis of Damage Development in Propellant by Using Digital Volume Correlation

The formulation of energetic materials (e.g., propellants) represents a challenge for material scientists mainly because of the difficulty/danger of tests performed with the real materials. It is thus of uttermost importance to be able to anticipate any difficulty that may occur at the processing stage using a phantom material. The present example is based on such a phantom, called PBX. It consists of a brittle phase forming smooth grains of typical size 500 μm, which are dispersed in a polymer matrix with a good bonding to the particles (Fig. 12). The mechanical behavior of this material is also essential to understand its functionality.

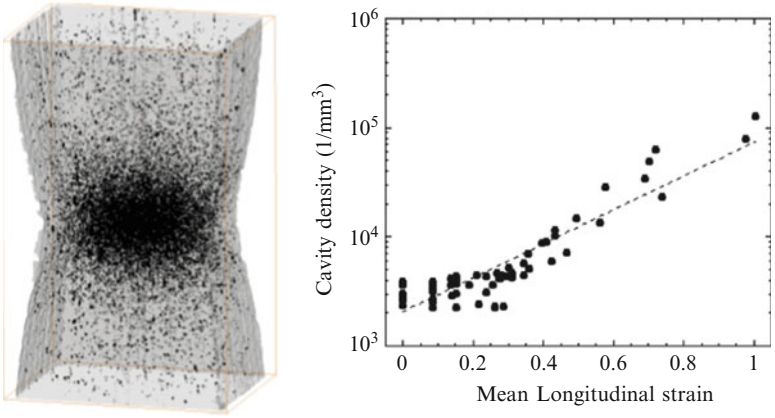


Fig. 11 3D rendering of the distribution of porosities within the deformed sample prior to fracture (*left*). Density of cavities as a function of the mean longitudinal strain (*right*). The solid symbols are experimental data, and the dashed line is an exponential fit (Courtesy of E. Maire)

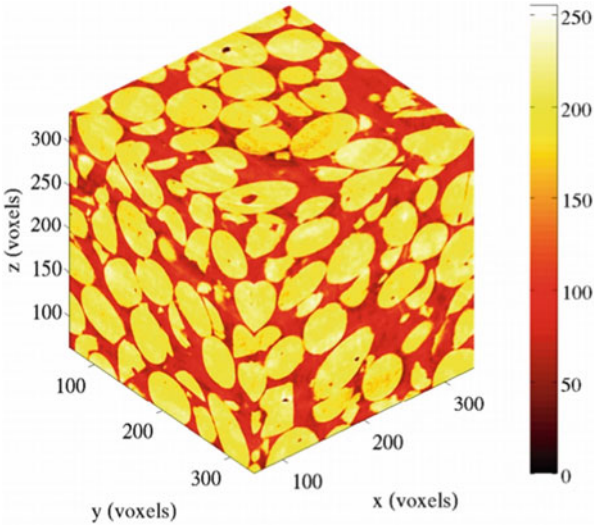


Fig. 12 Microstructure of the propellant phantom material (PBX) as revealed by tomography. The round inclusions (shown in *light gray/yellow*) are brittle, while the matrix phase (shown in *dark gray/red*) is a polymer binder. The physical size of a voxel is 7.4 μm, and that of the ROI is $\approx 2.1 \times 2.1 \times 2.1$ mm^3 (Courtesy of J. Adrien, A. Fanget and E. Maire)

In a uniaxial tensile test, failure occurs through the development of a major transverse crack, normal to the tensile axis. Tomography and DVC appear as unique tools to get some insight into the nature of the microcracks and the mechanism for their coalescence. Inclusion breakage and interface debonding are two possible mechanisms that can hardly be distinguished from a macroscopic characterization, and they are affected by the presence of a free edge, so that surface observation cannot be a priori considered trustworthy. Thus, a series of tomographic images was acquired during a mechanical test (uniaxial tension) at the European Synchrotron Radiation Facility (ESRF). Figure 12 shows the reference specimen.

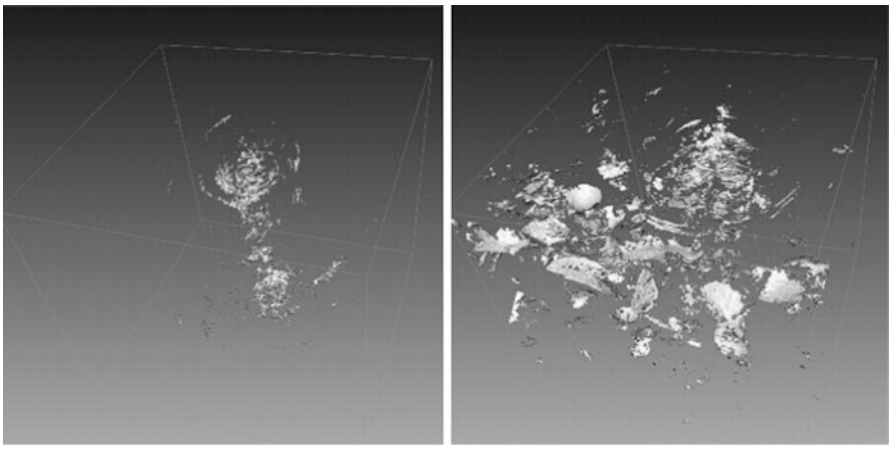


Fig. 13 Perspective view of the (thresholded) residual field after two different stages of deformation (*left*: 0.37 %; *right* 1.75 %). Whereas most of the residuals correspond to reconstruction artifacts at low strain (*left*), debonding cracks become active as the strain level increases

The displacement field at different stages of loading is determined from global DVC where the displacement field is decomposed over a finite element regular mesh consisting of 68-voxel C8 elements. Neither the displacement nor the strain fields by themselves could indicate the formation of microcracks. This is in part due to the coarse mesh chosen for this analysis. However, the residual field is very informative on the violation of the continuity assumption for the displacement field. Figure 13 displays perspective views of the thresholded residuals at two different levels of load.

The left residual field does not reveal any crack or defects other than those due to the way the images are obtained. The residual field presented in Fig. 13 (right) shows that numerous features appeared as the load increases, in particular smooth surfaces at the interface between particles and the matrix. Moreover, they tend to be oriented normal to the tensile axis, as could be anticipated for mode I opening. Thus, the residual analysis is able to bring this invaluable piece of information about the origin of the mechanical degradation. Moreover, as exemplified earlier, these patterns become apparent for a subvoxel crack opening, and hence, they could not be isolated from single images even with sophisticated image analysis techniques.

Summary

The present chapter has shown through a variety of examples that, from different outputs of image or volume correlations, it is possible to detect and quantify microstructural defects (i.e., physical damage) responsible for the degradation of the mechanical properties of materials. A particular emphasis has been put on

microcracks rather than microvoids as they are typically more challenging in terms of detectability. DIC is also a very powerful tool for identifying the constitutive parameters of mechanical damage understood as part of a constitutive law. This dimension will be explored in the two following chapters ("► Chap. 38, Evaluating Damage with Digital Image Correlation: B. From Physical to Mechanical Damage" and "► Chap. 39, Evaluating Damage with Digital Image Correlation: C. Applications to Composite Materials").

Acknowledgments Many results reported herein have been obtained thanks to the help of and discussions with J. Adrien, P. Aimedieu, A. Fanget, B. Leboime, E. Maire, B. Raka, J. Rupil, N. Swiergiel, and L. Vincent. Part of the work has been supported by Agence Nationale de la Recherche (VULCOMP phases 1 and 2), by Ile de France region (SESAME project entitled "*Plate-forme francilienne d'expérimentations mécaniques de 3e génération*," and DICCIT project).

This chapter has been reviewed by Profs. A. Dragon and J. Lemaitre. The authors would like to warmly thank them for their patience and help in improving the compuscript. Last, this chapter is dedicated to Prof. F.A. Leckie (1929–2013) with whom the authors discussed many issues presented herein.

References

J.C. Abry, S. Bochard, A. Chateauminois, M. Salvia, G. Giraud, In situ detection of damage in CFRP laminates by electrical resistance measurements. Comp. Sci. Technol. **59**, 925–935 (1999)

L. Babout, E. Maire, J.-Y. Buffière, R. Fougères, Characterisation by X-ray computed tomography of decohesion, porosity growth and coalescence in model metal matrix composites. Acta Mater. **49**(11), 2055–2063 (2001)

J. Baruchel, J.-Y. Buffière, E. Maire, P. Merle, G. Peix, *X-ray Tomography in Material Sciences* (Hermes Science, Paris, 2000)

C.J. Beevers (ed.), *Advances in Crack Length Measurement* (EMAS, West Midlands, 1982)

G. Besnard, F. Hild, S. Roux, "Finite-element" displacement fields analysis from digital images: application to Portevin-Le Châtelier bands. Exp. Mech. **46**, 789–803 (2006)

R. Billardon, I. Doghri, Prediction of macro-crack initiation by damage localization. C. R. Acad. Sci. Paris, II **308**(4), 347–352 (1989)

M. Bornert, T. Bretheau, P. Gilormini, *Homogenization in Mechanics of Materials* (Lavoisier, Paris, 2008)

J.-Y. Buffière, E. Maire, J. Adrien, J.-P. Masse, E. Boller, In situ experiments with X ray tomography: an attractive tool for experimental mechanics. Exp. Mech. **50**(3), 289–305 (2010)

J.L. Chermant, G. Boitier, S. Darzens, M. Coster, L. Chermant, Damage morphological parameters. Image Anal. Stereol. **20**, 207–211 (2001)

D. Claire, F. Hild, S. Roux, A finite element formulation to identify damage fields: the equilibrium gap method. Int. J. Numer. Meth. Eng. **61**(2), 189–208 (2004)

M. Coster, J.L. Chermant, Image analysis and mathematical morphology for civil engineering materials. Cement Concrete Comp. **23**(2–3), 133–151 (2001)

W.A. Curtin, Multiple matrix cracking in brittle matrix composites. Acta Metall. Mater. **41**(5), 1369–1377 (1993)

P. Forquin, F. Hild, A probabilistic damage model of the dynamic fragmentation process in brittle materials. Adv. Appl. Mech. **44**, 1–72 (2010)

M.G.D. Geers, R. De Borst, T. Peijs, Mixed numerical-experimental identification of non-local characteristics of random-fibre-reinforced composites. Comp. Sci. Tech. **59**, 1569–1578 (1999)

M. Grédiac, F. Hild (eds.), *Full-Field Measurements and Identification in Solid Mechanics* (ISTE/Wiley, London, 2012)
C.J. Hellier (ed.), *Handbook of Nondestructive Evaluation* (McGraw Hill, New York, 2001)
F. Hild, S. Roux, Digital image correlation, in *Optical Methods for Solid Mechanics. A Full-Field Approach*, ed. by P. Rastogi, E. Hack (Wiley-VCH, Weinheim, 2012), pp. 183–228
F. Hild, P.-L. Larsson, F.A. Leckie, Localization due to damage in fiber reinforced composites. Int. J. Solids Struct. **29**(24), 3221–3238 (1992)
F. Hild, A. Burr, F.A. Leckie, Fiber breakage and fiber pull-out of fiber-reinforced ceramic-matrix composites. Eur. J. Mech. A/Solids **13**(6), 731–749 (1994)
F. Hild, A. Fanget, J. Adrien, E. Maire, S. Roux, Three dimensional analysis of a tensile test on a propellant with digital volume correlation. Arch. Mech. **63**(5–6), 1–20 (2011)
R. Hill, Aspects of invariance in solid mechanics. Adv. Appl. Mech. **18**, 1–75 (1978)
L.M. Kachanov, Time of the rupture process under creep conditions. Bull. SSR Acad. Sci. Div. Tech. Sci. **8**, 26–31 (1958)
M.F. Kanninen, C.H. Popelar, *Advanced Fracture Mechanics* (Oxford University Press, Oxford, 1985)
H. Leclerc, J.-N. Périé, S. Roux, F. Hild, Integrated digital image correlation for the identification of mechanical properties, in *MIRAGE 2009*, ed. by A. Gagalowicz, W. Philips, vol. LNCS 5496 (Springer, Berlin, 2009), pp. 161–171
H. Leclerc, J.-N. Périé, S. Roux, F. Hild, Voxel-scale digital volume correlation. Exp. Mech. **51**(4), 479–490 (2011)
H. Leclerc, J.-N. Périé, F. Hild, S. Roux, Digital volume correlation: what are the limits to the spatial resolution? Mech. Indust. **13**, 361–371 (2012)
J. Lemaitre, *A Course on Damage Mechanics* (Springer, Berlin, 1992)
J. Lemaitre, J.-L. Chaboche, Phenomenological approach of damage rupture. J. Méc. Appl. **2**(3), 317–365 (1978)
J. Lemaitre, R. Desmorat, *Engineering Damage Mechanics* (Springer, Berlin, 2005)
J. Lemaitre, J. Dufailly, Damage measurements. Eng. Fract. Mech. **28**(5–6), 643–661 (1987)
J. Lemaitre, J.-P. Cordebois, J. Dufailly, Elasticity and damage coupling. C. R. Acad. Sci. Paris Ser. B **288**(23), 391–394 (1979)
J. Lemaitre, J. Duffailly, R. Billardon, Evaluation de l'endommagement par mesures de microdureté. C. R. Acad. Sci. Paris **304**(series II), 601–604 (1987)
E. Maire, Quantitative measurement of damage, in *Local Approach to Fracture*, ed. by J. Besson (Presses Ecole des Mines de Paris, Paris, 2003), pp. 79–108
E. Maire, O. Bouaziz, M. Di Michiel, C. Verdu, Initiation and growth of damage in a dual-phase steel observed by X-ray microtomography. Acta Mater. **56**(18), 4954–4964 (2008)
E. Maire, T. Morgeneyer, C. Landron, J. Adrien, L. Helfen, Bulk evaluation of ductile damage development using high resolution tomography and laminography. C. R. Phys. **13**(3), 328–336 (2012)
K. Mills (ed.), *Fractography*, vol. 12 (ASM International, Materials Park, 1991)
R.H.J. Peerlings, R. de Borst, W.A.M. Brekelmans, J.H.P. de Vree, Gradient-enhanced damage for quasi-brittle materials. Int. J. Numer. Method Eng. **39**, 3391–3403 (1996)
G. Pijaudier-Cabot, Z.P. Bažant, Nonlocal damage theory. ASCE J. Eng. Mech. **113**(10), 1512–1533 (1987)
R. Prabhakaran, Damage assessment through electrical resistance measurement in graphite fiber-reinforced composites. Exp. Tech. **14**(1), 16–20 (1990)
Y.N. Rabotnov, On the equations of state for creep, in *Progress in Applied Mechanics, Prager Anniversary Volume*, ed. by W.T. Koiter (McMillan, New York, 1963), pp. 307–315
P.K. Rastogi (ed.), *Photomechanics*, vol. 77 (Springer, Berlin, 2000)
P. Rastogi, E. Hack (eds.), *Optical Methods for Solid Mechanics. A Full-Field Approach* (Wiley-VCH, Berlin, 2012)
J. Réthoré, F. Hild, S. Roux, Extended digital image correlation with crack shape optimization. Int. J. Numer. Method Eng. **73**(2), 248–272 (2008)

J. Réthoré, N. Limodin, J.-Y. Buffière, F. Hild, W. Ludwig, S. Roux, Digital volume correlation analyses of synchrotron tomographic images. J. Strain Anal. **46**, 683–695 (2011)

S. Roux, F. Hild, P. Viot, D. Bernard, Three dimensional image correlation from X-ray computed tomography of solid foam. Comp. Part A **39**(8), 1253–1265 (2008)

S. Roux, J. Réthoré, F. Hild, Digital image correlation and fracture: an advanced technique for estimating stress intensity factors of 2D and 3D cracks. J. Phys. D Appl. Phys. **42**, 214004 (2009)

J. Rupil, S. Roux, F. Hild, L. Vincent, Fatigue microcrack detection with digital image correlation. J. Strain Anal. **46**(6), 492–509 (2011)

S. Roux, F. Hild, H. Leclerc, Mechanical assistance to DIC, in *Proceeding. Full Field Measurements and Identification in Solid Mechanics*, vol. IUTAM Procedia, 4, ed. by H. Espinosa, F. Hild (Elsevier, 2012), pp. 159–168

L. Salvo, P. Cloetens, E. Maire, S. Zabler, J.J. Blandin, J.Y. Buffiere, W. Ludwig, E. Boller, D. Bellet, C. Josserond, X-ray micro-tomography an attractive characterisation technique in materials science. Nucl. Instrum. Method B **200**, 273–286 (2003)

W.N. Sharpe Jr. (ed.), *Springer Handbook of Experimental Solid Mechanics* (Springer, New York, 2008)

M.A. Sutton, J.-J. Orteu, H. Schreier, *Image Correlation for Shape, Motion and Deformation Measurements: Basic Concepts, Theory and Applications* (Springer, New York, 2009)

C.C. Tasan, J.P.M. Hoefnagels, M.G.D. Geers, Indentation-based damage quantification revisited. Scr. Mater. **63**, 316–319 (2010)

C.C. Tasan, J.P.M. Hoefnagels, E.C.A. Dekkers, M.G.D. Geers, Multi-axial deformation setup for microscopic testing of sheet metal to fracture. Exp. Mech. **52**(7), 669–678 (2012)

Z. Tomičević, F. Hild, F. Roux, Mechanics-aided digital image correlation. J. Strain Anal. **48**, 330–343 (2013)

Evaluating Damage with Digital Image Correlation: B. From Physical to Mechanical Damage

38

François Hild and Stéphane Roux

Contents

Abstract

This chapter is devoted to the evaluation of damage with digital image correlation (DIC). This measurement technique provides displacement fields that can be used to tune material parameters of damage models. The potential of DIC for identification purposes is illustrated hereafter for simple 1D-type geometries in the context of quasi-brittle and ductile behaviors. Beams and frames as encountered in civil engineering application can be described through damage laws. The localization of damage at single points (lumped damage mechanics) and the

F. Hild (✉) • S. Roux
LMT-Cachan, ENS Cachan/CNRS/PRES UniverSud Paris, Cachan Cedex, France
e-mail: hild@lmt.ens-cachan.fr; stephane.roux@lmt.ens-cachan.fr

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_25

resulting damage law can be determined very precisely based on dedicated versions of DIC incorporating the descriptive language of beam theory. At smaller scale, plasterboards (plaster core lined with paper) constitute another example of a structure (rather than a material), which can be described as damageable. Those cases will be analyzed in details.

1D Geometry of Various Materials

The previous chapter (▶ Chap. 37, Evaluating Damage with Digital Image Correlation A. Introductory Remarks and Detection of Physical Damage) dealt with the detection of microcracks, as revealed from their mechanical effect. However, as discussed in the introduction, damage most often refers to a continuum level description where microcracks are not considered individually but rather through their effect, say, in stiffness reduction. This can be justified from homogenization, or simply considered as a phenomenological, though thermodynamically sound and well founded, description. Such a global and averaged view can also be tackled with DIC, either by providing the displacement field needed to identify a damage state or through a more intimate coupling between modeling and image registration.

To discuss this issue with the minimal degree of complexity, this part is based upon one-dimensional problems amenable to a beam-type description, where damage mechanics is also here (albeit less frequently used) a very appropriate and efficient formalism to account for the progressive degradation of mechanical properties. It is worth noting that plates and shells show in a different manner a tendency to concentrate cusps and ripples (Audoly and Pomeau 2010) leading to a rich coupling between geometrical and constitutive nonlinearities. This remark shows that besides the illustrative purpose of this topic, it contains specific ramifications that would require a research effort per se.

This part allows a number of challenges in the analysis of an experiment to be revisited. The very first question that will be addressed is the best determination of a large-scale deformation. Although, it may appear as a trivial question, it will raise interesting issues in terms of methodology. The exploitation of such large-scale deformation will open the way to a better knowledge of the actual conditions of a given test. In fact, there is no need to assume that a test is perfect or sticks to the ideal picture of what it should be, if full-field measurements can bring the actual information on the test geometry, on the load distribution, or on the actual location of an applied force. This better characterization is of help to the analysis, allowing for less variability, say, in the measurement of elastic stiffnesses. However, in damage mechanics, chances that an assumed symmetry of the test is maintained all along the test are dubious at best. Damage has a natural propensity to amplify any dissymmetry present in the test, and thus it is very important in the analysis to allow for such an evolution and capture it with the analysis tools (DIC in this chapter) in order not to bias further exploitation of the results.

Another challenge is to adjust the metrology to the choice for a particular description level. This is a key issue lying at the very heart of CDM

(see ▶ Chap. 37, Evaluating Damage with Digital Image Correlation A. Introductory Remarks and Detection of Physical Damage). Unless the "physical damage" cannot be resolved, the choice is often open to decide either for a description of say microcracks distributed within the analyzed domain or for an averaged vision. (It is recalled that throughout this chapter, the chosen convention for "physical damage" designates the microscopic mechanisms (microcracking or cavitation) responsible for the degradation of the mechanical properties, whereas "mechanical damage" refers to the averaged description of such degradation through a suited constitutive law.) If the second choice is made, the corresponding displacement field resulting from the most appropriate damage law will not be the actual displacement field. The challenge on DIC is thus to capture the long-wavelength component of the strain field, without being affected by the physical damage features. Finally, for the particular example of beams, it may be chosen to work in the opposite direction, namely, rather than spreading damage over a large portion of the beam, it may rather be localized (or lumped) at a single point, using "lumped damage mechanics," to concentrate nonlinearities and preserve the simplicity of the undamaged description in between these "damaged hinges" (Cipollina et al. 1995). Here again, DIC is a very convenient tool to identify such features.

Thus, the major emphasis of this part is to build an identification strategy to *measure* damage laws. The literature is quite abundant on different methods proposed to estimate the elastic properties from measured displacement fields (Grédiac and Hild 2012). The method presented hereafter is the most natural *updating method*. It is the most appropriate, but it is intrinsically nonlinear (in the generic case of significant damage). Other methods will be shown in "▶ Chap. 39, Evaluating Damage with Digital Image Correlation: C. Applications to Composite Materials" such as the "equilibrium gap method." However, the "reconditioned equilibrium gap method" is essentially equivalent to the updating method used hereafter. Hence, the terminology can be somewhat confusing. The "virtual field method" (Pierron and Grédiac 2012) consists of having a weak formulation of equilibrium and constitutive law, which is common to all the abovementioned methods, for arbitrary trial displacement fields. However, there is no specific prescription for the selection of these fields, even if some are more appropriate than others. An overview of these different methods is provided by Avril et al. (2008). It is to be emphasized that they are all correct, and their merits are to be evaluated from the more or less detrimental effect of noise on their identification capacity in particular for nonlinear constitutive laws such as those involving damage.

To illustrate these aspects of the problem, different examples dealing with beam geometries are commented. The beam problems are ubiquitous and are met at large scales for civil engineering concrete or steel beams, at centimeter/millimeter scale for materials such as plasterboard or laminated composite plates, or even at a microscale for microcantilevers encountered on lab-on-a-chip devices (Amiot et al. 2007). The case studies chosen to illustrate this part concern:

- A cantilever steel beam where buckling at its foot can be described as a localized or lumped damage.
- A plasterboard specimen subjected to a simple four-point bend test up to failure.

The parallel in the procedures to be followed for civil engineering-type structures and material science lab-scale tests is emphasized to underline the usefulness of such slender geometries.

Before addressing the identification of damage features, one key aspect is related to the extraction of the experimental boundary conditions. This question may appear as secondary, but if the boundary conditions are not properly captured, they may bias at best or even prevent the evaluation of the sought information, namely, the damage field and the corresponding growth law. In the first example, the determination of the boundary conditions (flexural moment, application point of the transverse force) can be uncoupled from the analysis of the damage field per se since an elastic part is still accessible. In the second example, boundary conditions and damage field are identified at the same time.

Case Studies

Steel Beam

A civil engineering setup consisting of a cantilever beam of length 1.50 m and square hollow cross section of outer (resp. inner) size of 120 mm (resp. 112 mm) is studied. The beam is made of conventional construction steel (ASTM-A-500). It is shown in Fig. 1. It is anchored in the massive block at its foot so that it can be considered as clamped. At the opposite end, a transverse force is applied with a jack. A series of high-definition (3,888 $\times$ 2,592 pixels) camera pictures are shot during the test, i.e., for a progressive increase of the transverse loading, for each 10 mm increment of the stroke. The physical size of 1 pixel is 0.39 mm. A random black-and-white painting is deposited onto the surface to allow for a more precise image registration.

Plasterboard

Plasterboard is extremely common as a building material, with good thermal and acoustic insulation properties, excellent fire resistance, ease of installation, nice surface finish suited for decoration, together with very low cost. The mechanical performance expected from such boards is generally modest, and hence, optimization of insulation properties is privileged over strength. To withstand the bending due to processing and handling of these boards, the paper lining of the outer faces of the plaster board plays an essential role. Even if the plaster core is brittle, paper will allow withstanding a flexural moment larger than the one needed to initiate the first crack. Even if debonding naturally occurs at the interface between plaster and paper close to a transverse crack, load transfer will allow for initiation of other cracks past a characteristic distance. Such flexural strength is generally qualified from three- or four-point bend tests.

A specimen 160 cm in length, 50 cm in width, and 13 mm in height is subjected to a four-point bend test. The outer support span is equal to 150 mm and the inner

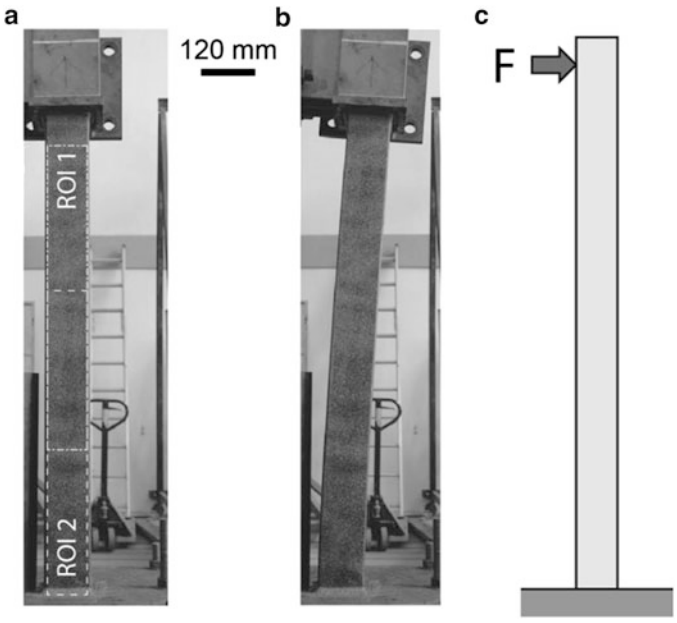


Fig. 1 (**a**) Reference configuration of a cantilever steel beam clamped at its foot and subjected to a transverse force. The two regions of interest that will be considered in the following analyses are depicted as white boxes. (**b**) Geometry of the beam at the end of the test, a local buckling can be guessed close to the base (Courtesy of N.J. Guerrero, M.E. Marante and R.A. Picon). (**c**) Schematic representation of the mechanical test

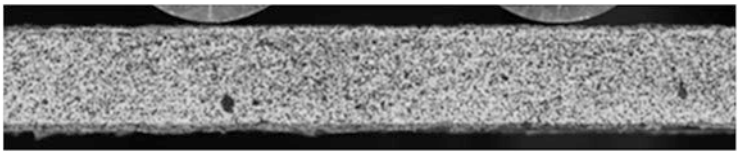


Fig. 2 Side view of a four-point bend test on a plasterboard specimen speckled with black paint over its observation face. The outer fixtures supporting the specimen are located outside the field of view. The inner load-bearing cylinders can be seen above the specimen (Courtesy of A. Bouterf)

yoke span is 40 mm. In order not to damage the specimen, the load-bearing contacting elements are cylinders 20 mm in diameter. Figure 2 shows a side view of the central part of the sample; only the central part is shown. Images of the speckled face are captured with a camera of definition of 3,888 × 2,592 pixels whose gray levels are stored with 8-bit digitization. The physical size of the pixel is 37 μm.

In the following, the displacement fields are obtained with DIC analyses based on a regular square mesh composed of 12 pixel wide Q4 elements. A priori

estimates of the displacement uncertainty are found to be equal to 0.04 pixel or 1.5 μm. Because of such a large radius of the cylinders and because of the crushing of the paper facing, the exact equivalent points of load application are imprecisely defined from the bare experimental setup.

Formalism

Within Euler-Bernoulli theory, in an elastic beam along the x-axis, the curvature, $\kappa(x)$, is proportional to the flexural moment, $M(x)$, with a flexural stiffness $k_0 = EI$, where E is the Young modulus and I the moment of geometric inertia. If $v(x)$ denotes the deflection of the beam, then $\kappa(x) = \mathrm{d}^2 v(x)/\mathrm{d}x^2$. In the presence of a degradation mechanism (such as microcrack formation but also buckling or other phenomena affecting either the elastic properties or the geometric cross section of the beam), damage, $D(x)$, is introduced to account for a reduction in the flexural stiffness, so that

$$k(x) = (1 - D(x))k_0 \tag{1}$$

Different presentations (either starting from the expression of the free energy as discussed in "▶ Chap. 39, Evaluating Damage with Digital Image Correlation: C. Applications to Composite Materials" or the effective stress (here flexural moment)) can be used. Under a monotonic loading history, damage is thus expected to be a function of the local curvature only

$$D(x) = \Phi(k(x)) \tag{2}$$

where Φ is a function that completely describes the damage growth law. In a test where the load is known, the curvature is a solution to the nonlinear problem where

$$(1 - \Phi(k(x)))k_0 k(x) = M(x) \tag{3}$$

Theoretically, DIC gives access to the deflection, $v(x)$, and hence combining this measurement, and the knowledge of the flexural moment $M(x)$, the function Φ can be estimated. Moreover, in simple cases such as the four-point bend test, the central region is studied where M (and thus κ) are uniform, and thus the damage law can directly be read from the test. This elegant idea has been exploited by Leplay et al. (2010) to evaluate the damage law of a porous ceramic material.

The practical difficulty of the previous pathway to the measurement of damage is the fact that the primary measurand of DIC is a displacement, such as the deflection, and hence the second-order derivative required for the computation of the curvature acts as a strong amplifier of noise. As a consequence, the proper procedure to evaluate Φ should be based on the displacement itself. For a uniform and uncorrelated Gaussian noise, the most secure determination of the displacement is obtained from a least squares minimization over a basis that contains the expected form.

The use of DIC to measure the beam kinematics can be performed with different strategies. The most obvious one is to use standard (global or local) DIC softwares and then projecting the two-dimensional displacement field onto the Euler-Bernoulli description in a least squares sense. It is worth mentioning that it is possible to incorporate the beam kinematics, already in the basis of function used in global DIC (Hild et al. 2009). If this is the chosen strategy, then the output of the DIC analysis is directly the deflection and rotation as a function of the coordinate along the beam axis, i.e., the most appropriate kinematic description. In the two discussed examples, a different choice is made for the measurement of the beam kinematics for reasons to be discussed below.

For both approaches, either the projection or the beam-DIC approach, the discretization of the deflection $v(x)$ is a key issue. It is natural to make sure that the deflection and its derivative (the rotation) are continuous. If several beam elements are considered in a partition of the beam, then it is easy enough to control the continuity of deflection and rotation, for instance, using Lagrange multipliers. If no localized torque is applied at the interface, then the flexural moment should be continuous, and hence if the flexural stiffness is also continuous, this corresponds to a continuity of the curvature field. Similarly, if no transverse force is applied at the interface, then the shear force should be continuous, and again, for a continuous flexural stiffness, the third derivative of the deflection should be continuous. If all continuity conditions are prescribed, this is equivalent to having a single beam element (with homogeneous elastic properties and no applied load along its length), over which $v(x)$ is a third-order polynomial. A further exploitation of the regularity of the deflection will be discussed below.

DIC Measurement of the Boundary Conditions

The beam is first assumed to be undamaged. The first goal to be aimed at is the identification of the proper boundary conditions. To take full advantage of DIC, a large-scale portion of the beam should be exploited. The discussion is first specialized to a segment (say $0 < x < L$) where no distributed load is applied. Integration of the above equations in the absence of damage leads to the deflection, $v(x)$, being a third-order polynomial of the coordinate x. Since $v(x)$ is assumed to be measured from DIC (and simultaneously corrupted by a random noise), the most robust way to determine the flexural moment or the shear force is not from a second- or third-order derivative of the displacement field. Rather, a much better way is to perform a least squares fit over a third-order polynomial. In the case where the noise affecting $v(x)$ is white (uncorrelated) and Gaussian, it can even be shown that this procedure is *the* most robust.

Moreover, as the measure is uniform over the interval $0 < x < \mathrm{L}$, it is convenient to choose as a polynomial basis, the Legendre polynomials, $Q_n(x) = \left(\sqrt{2/L}\right)P_n(2x/L - 1)$, since they form an orthogonal basis. Here P_n denotes the normalized

Legendre polynomials defined over the $[-1;1]$ interval. The least squares polynomial regression, $v(x) = \sum a_n Q_n(x)$, simply writes

$$a_n = \int_0^L Q_n(x) v(x) dx \tag{4}$$

and hence, working out a simple algebra, the least squares estimate for the curvature is given by

$$\kappa(x) = \frac{60}{L^2}\left[\int_0^1 \left(6y^2 - 6y + 1\right) v(yL) dy + 7\left(\frac{2x}{L} - 1\right)\int_0^1 \left(20y^3 - 30y^2 + 12y - 1\right) v(yL) dy\right] \tag{5}$$

and that of the third-order derivative of $v(x)$

$$\kappa' = \frac{840}{L^3}\left[\int_0^1 \left(20y^3 - 30y^2 + 12y - 1\right) v(yL) dy\right] \tag{6}$$

It is interesting to note the structure of the evaluation of such quantities of interest, generically denoted as q, is related to the measured displacement through a linear operator that can be written as a simple scalar product of $v(x)$ with an "extractor" field $E_q(x)$

$$q = \int_0^L E_q(x) v(x) dx \tag{7}$$

It is to be emphasized that the point, x_0, where the flexural moment vanishes can easily be read from the previous expression of $\kappa(x)$

$$\left(\frac{x_0}{L}\right) = \frac{\displaystyle\int_0^1 \left(70y^3 - 108y^2 + 45y - 4\right) v(yL) dy}{\displaystyle 7\int_0^1 \left(20y^3 - 30y^2 + 12y - 1\right) v(yL) dy} \tag{8}$$

The above procedure is the most appropriate way of dealing with a displacement field obtained from DIC. In this respect, there is no need to introduce a specific regularization. When beam-DIC is used, in fact such a procedure is already implicitly performed in the sense that the deflection is already a third-order polynomial that is adjusted thanks to image registration, which can be shown to be equivalent to a least squares regression for a parabolic pair correlation function of the surface texture (Hild and Roux 2012).

As a side remark, it is worth emphasizing that the existence of an "extractor" field is a very general property for all quantities that are linear in the displacement field. For instance, such a structure can be encountered for the measurement

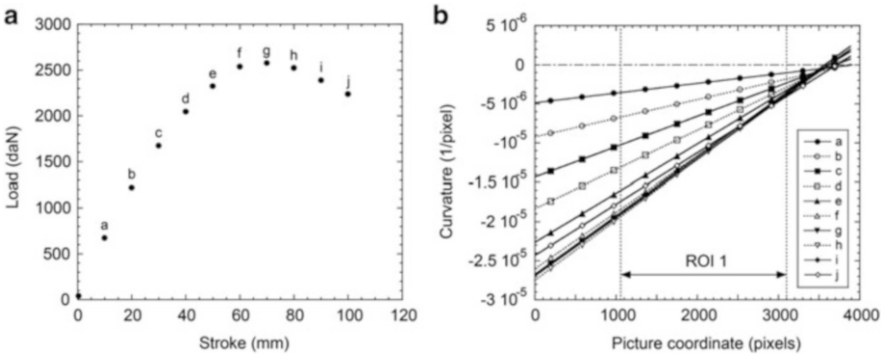


Fig. 3 (**a**) Load versus stroke of the steel beam at the points (*a–j*) where images are acquired. (**b**) Evaluation of the curvature of the beam versus abscissa along the beam for different stages of loading labeled with a letter. The region of interest (ROI 1; see Fig. 1a) over which the measurement is performed is indicated; its length is approximately 800 mm. Outside this region, the data are extrapolated. This procedure allows for a precise determination of the loading, and it is noted that the point of application of the transverse load (abscissa where the flexural moment, and hence curvature, vanishes) slowly drifts along the test (Hild et al. 2011)

of stress intensity factors, from a DIC measurement of a displacement field (Réthoré et al. 2011). Similarly, the crack-tip position can be evaluated as the above x_0 from the ratio of two quantities of interest (one primary quantity and its derivative; see Roux et al. (2009)).

One possible application of the above procedure is the analysis of the loading state for the steel beam. As mentioned above, the load is applied to the beam end through a jack. However, in order not to damage the beam at the contact point of application, a protecting device is introduced. Consequently, it is difficult to have a very accurate estimate of the position where the load can be considered as an equivalent point force. Yet, equipped with the above formulas or with a beam-DIC analysis, if the kinematics over a region of the cantilever beam corresponds to that given by an elastic medium (i.e., ROI 1 of Fig. 1a), such an estimation is quite straightforward.

Figure 3 shows the evaluation of the curvature as a function of position at different stages of loading, based on a portion of the beam where the behavior can be guaranteed to be elastic. Knowing the flexural stiffness of the beam (in its undamaged state) allows the transverse load to be evaluated from the measurement of the curvature $F = k_0\, \mathrm{d}\kappa(x)/\mathrm{d}x$ and the point of application where $\kappa(x)$ vanishes. Similarly, the rotation at the foot validates the relevance of the clamping condition (and thereby determines the most appropriate location of the beam end).

For the four-point bend test on plasterboard, a precise measurement of the loading (and geometry) of the setup may be addressed using the above methodology in different portions of the specimen (side or central parts) in between supporting or loading cylinders. However, such a partition in three independent parts is not

optimal as the continuities across the different parts are not exploited. Rather, a complete numerical description is an easy way to circumvent this difficulty and has the flexibility of addressing arbitrary complexity of the setup (and can easily be extended to frames). The procedure is a straightforward extension of the previous approach. First, a parametric modeling of the setup is to be chosen. This step is important since it relies on a good balance of a sufficiently general case to encompass all practical deviations from what the ideal test should be and constrained enough to allow for a precise determination of the parameters. The set of parameters defining the experiment are gathered in a vector, $\mathbf{a} = \{a_i\}$, for $i = 1, \ldots, n$, so that the expected deflection $v_{\text{comp}}(x;\mathbf{a})$ can be numerically determined. The sensitivity fields

$$w_i(x;\mathbf{a}) = \frac{\partial v_{\text{comp}}(x;\mathbf{a})}{\partial a_i} \tag{9}$$

are introduced and computed either formally, or through finite differences. In the framework of a linear theory, some of the sensitivity fields are constant (e.g., if the corresponding parameter is say a load at a particular point, so that the sensitivity is commonly termed an influence function), but the general case is that these sensitivity fields are not independent of the parameters $\mathbf{a}$. For instance, the position of a load-bearing point has a nonlinear influence on the deflection, even within linear elasticity. The ability to deal with nonlinearities (and these ones are gentle!) opens the way to the identification of more complex constitutive laws. Interestingly, the sensitivity fields often have a simple mechanical interpretation. When the parameter a_i is the location x_0 of an applied load F, the sensitivity field is the deflection induced by a load F at $x_0 + dx_0$ and $-F$ at x_0. This is precisely a local torque $F\,dx_0$ exerted at x_0.

To determine the best set of parameters, $\mathbf{a}$, to account for the measured displacement $v_{\text{DIC}}(x)$, it is convenient to minimize the quadratic norm of the differences

$$T(\mathbf{a}) = \int_0^L \left(v_{\text{comp}}(x;\mathbf{a}) - v_{\text{DIC}}(x)\right)^2 \, \mathrm{d}x \tag{10}$$

so that a current estimate of the parameters $\mathbf{a}^p$ at iteration p can be corrected to $\mathbf{a}^{p+1} = \mathbf{a}^p + \mathrm{d}\mathbf{a}$, where the argument of the objective function T is Taylor expanded

$$T(\mathbf{a}) = \int_0^L \left(v_{\text{comp}}(x;\mathbf{a}) + \sum\nolimits_i w_i(x;\mathbf{a})\,\mathrm{d}a_i - v_{\text{DIC}}(x)\right)^2 \, \mathrm{d}x \tag{11}$$

In other words, the correction $\mathrm{d}\mathbf{a}$ is determined as the least squares regression of the difference $(v_{\text{DIC}} - v_{\text{comp}})$ over the sensitivity fields. The answer is thus simply given by the pseudoinverse

$$\mathrm{d}\mathbf{a} = \mathbf{M}^{-1}\mathrm{d}\mathbf{b} \tag{12}$$

where

$$\begin{aligned} \mathrm{d}b_i &= \int_0^L \big(v_{\mathrm{DIC}}(x) - v_{\mathrm{comp}}(x;\mathbf{a})\big) w_i(x;\mathbf{a}) \ \mathrm{d}x \\ M_{ij} &= \int_0^L w_i(x;\mathbf{a}) w_j(x;\mathbf{a}) \ \mathrm{d}x \end{aligned} \tag{13}$$

It is to be noted that the above setting takes again the form of an extractor, or its natural extension to the nonlinear case, namely, with a tangent extractor and an additive term as expected from the proposed Newton scheme.

In the particular case of the four-point bend setup, all the above sensitivity fields are computed analytically so that the complete analysis can be made very efficient. However, the explicit expression of the different fields is lengthy and presents little practical interest. It is worth noting the complete analogy that can be drawn with the simpler case of a single beam element supporting no distributed load that was introduced earlier. A major advantage of the proposed procedure is that it is very tolerant to experimental imperfections. In particular, as earlier mentioned, the load-bearing cylinders have a large radius, and the paper siding of the plasterboard is compressed over a region of a few millimeters. Hence, the precise location of the equivalent point-like load does not appear to be an easy problem. Similarly, an equal left/right balance of the load is often expected by symmetry arguments, but it may be violated in practice.

The chosen parameters were the exact position of the contacting elements, the force distribution between the two loading cylinders (whereas the sum of the two forces is known from the load cell) and a rigid body motion superimposed on the solution where the supporting elements are fixed. A first benefit of using this procedure lies in the measurement of the elastic properties of the beam. For the measurement of the flexural stiffness (in the linear elastic regime) performed over 10 different samples, the standard deviation of the flexural stiffness is reduced by almost a factor of 2 (a Young's modulus of 2.52 ± 0.18 GPa was identified to be compared with 2.40 ± 0.28 GPa for the classical beam theory based upon symmetry considerations). Hence, half of the fluctuations originate from test imperfections and not from intrinsic variability of the material. This issue will be further commented later on (see results reported in Fig. 9).

Damage Identification

The determination of the damage law is carried out as a post-processing step of the DIC analysis. In fact, the previous framework rests on two assumptions, namely, the first one is that the flexural moment is constant, while the second is that under a similar flexural moment, the local curvature is identical. The latter assumption excludes regimes past the breakdown of uniqueness occurring at the onset of localization for a homogeneous constitutive law. However, it is to be stressed that assuming a uniform constitutive law is a very stringent assumption, especially by the fact that homogenization lies at the very root of the existence/relevance of the damage law as such.

Relaxing the previous framework to account for less restrictive assumptions opens two rather different perspectives:

- The first one consists of allowing for different constitutive laws in different regions along the beam. It opens a rather wide scope. The number of unknowns to be determined increases with the domain size as each additional volume element brings its own unknowns and thus does not help for the determination of previous ones. This increase of unknowns may still be acceptable in 1D situations, but in two or three dimensions, this route may soon reveal inadequate, unless a very coarse description is chosen (see "► Chap. 39, Evaluating Damage with Digital Image Correlation: C. Applications to Composite Materials").
- The second one assumes a homogeneous constitutive law, yet allowing for an inhomogeneous loading. This second option rests on the hypothesis that damage will amount to the same level for the same maximum historical (equivalent) strain. Thus, the actual unknown is the relationship between damage and maximum equivalent strain, or essentially a 1D unknown function, whatever the space dimensionality. If an additional volume element is taken into account, then the number of unknowns does not increase, and hence, larger volume reduces the difficulty and possible poor conditioning of the problem at hand. Moreover, even if the loading is homogeneous, say in the central span of a 4-point bend test, past a bifurcation point where the solution is not unique, the observed displacement field is still subjected to very stringent constraints, and hence, rather than being stopped at a bifurcation point, this strategy is expected to allow for the identification in post-bifurcation regimes, which can hardly be tackled with traditional means.

The second option is more favorable than the first in terms of accuracy, for distributed damage, because it involves a smaller number of unknowns for a similar amount of information. However, it rests on an assumption to be justified. Moreover, the situation is not always as clear cut as for these two pure cases. A partition of the domain into a few regions may be proposed where each may be regarded as homogeneous, and the question of an appropriate partitioning is by itself a very interesting point, which will not be further discussed here. Yet, for beams that are not subjected to distributed loading, the most severely loaded parts are the ends, and hence it may occur quite frequently for civil engineering applications that damage is essentially localized at the foot of a pillar, and hence a simple partition of the beam in a localized zone concentrating the nonlinear evolutions and the remainder of the beam being essentially ruled by a linear elastic constitutive law may be an appropriate way of capturing the degradation of a structure with a fixed level of complexity. Moreover in that case, the damaged region may possibly be pictured as reduced to a single point, the so-called "lumped damage" approach, with a minimal complexity, and hence constituting the basis of a very efficient modeling (Marante and Flórez-López 2003).

The two examples introduced earlier will be analyzed in this section as they are representatives of the two above cases, namely, first localized damage and second distributed damage.

Localized Damage

The steel cantilever beam (Fig. 1) is considered again. It has been shown previously how to characterize the loading state based on the analysis of an elastic part of the structure used as a "load" sensor. Figure 4 shows, for the very same analysis, the rotation field. It is observed that for the first levels of loading, the clamping condition at the foot of the beam is appropriate. However, for larger loads, extrapolation of the rotation data leads to a significant rotation at the foot. Since the level of correlation residuals in the region of interest remains very low, there is no reason to disqualify the assumption of an elastic behavior, at least in the region of interest *itself*. The flexural moment has a known evolution similar to the curvature shown in Fig. 3b, and where the extrapolation outside of ROI 1 (Fig. 1a) is legitimate. Thus, the only possibility to reach a vanishing rotation at the foot of the beam is to accept the fact that the curvature is much higher than the extrapolated value in the region $x < 1{,}000$ pixels. This is consistent with the known flexural moment only if the bending stiffness is reduced.

Note that in reality the phenomenon at play is buckling, taking place at the foot of the beam. This goes together with a drastic reduction in bending stiffness (analyzed here as an effective damage) and presumably a plastic hinge leading to a permanent rotation. The partition into plasticity and damage would require information about unloading, which is unfortunately unavailable in the present case. Thus, the analysis is restricted within the framework of damage theory (as if there were no plasticity). In the event of a more complete set of data, the incorporation of plasticity together with damage would not represent any particular difficulty, and most of the discussion would continue to hold.

A detailed analysis of the correlation residuals, and sensitivity to the beam element discretization (not reported herein; see Hild et al. (2011)) yields the rotation fields measured in ROI 2 (see Fig. 1a) and shown in Fig. 5a. Most of the nonlinear (damage) behavior is confined within a region $\boldsymbol{\Delta}_{\text{nl}}$, $0 < x < 500$ pixels. A uniform damage is assumed within such an element and an undamaged behavior elsewhere. The boundary conditions are the following:

- The beam is clamped at its foot $x = 0$.
- At the interface between the damaged and undamaged beams, deflection and rotation fields are both continuous, as well as shear force and bending moment.
- At the free end of the cantilever, a pure shear force is exerted at a point to be determined.

A discontinuity in the curvature takes place at the interface between damaged and undamaged elements, $x = x_1$, and hence writing the continuity of moment leads to

$$(1 - D)k_{dam}(x_1) = k_{undam}(x_1) \tag{14}$$

Thus, the damage state is directly evaluated from the ratio of curvatures in the undamaged to the damaged region. Rewriting the damage variable as a function of

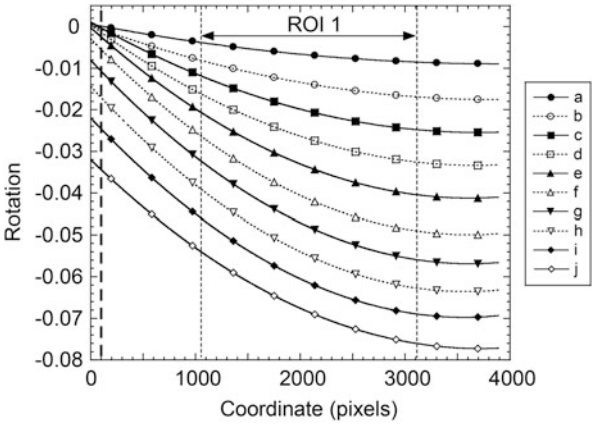


Fig. 4 Evaluation of the rotation field along the beam based on the same region of interest as previously (indicated as *ROI 1* in Fig. 1a) and for different levels of loading. Outside ROI 1, the data are extrapolated. The length of ROI 1 is 800 mm

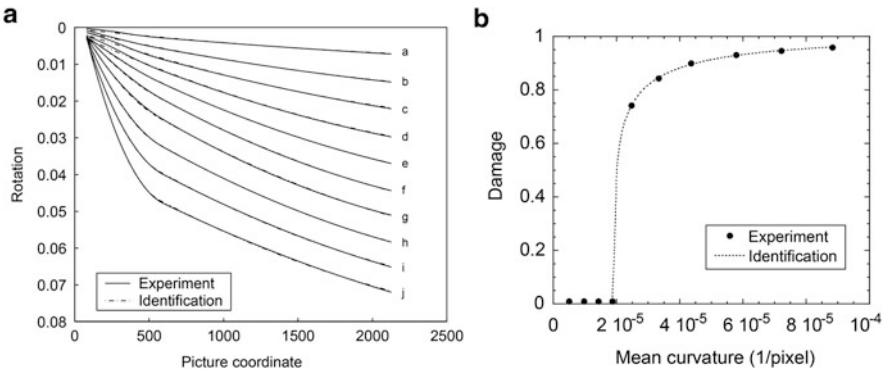


Fig. 5 (**a**) Rotation fields over a length of 800 mm for the ten analyzed load levels. (**b**) Damage as a function of the mean curvature in the damaged zone. The data points are obtained from the beam-DIC analysis, and the dotted curve is the fit to the proposed functional form. The physical size of one pixel is 0.39 mm

curvature $D = \Phi(\kappa)$, the analysis of the complete series of images leads to values of damage (one per image, representative for the entire "damaged area"), which can be interpolated by the following form

$$\Phi(\kappa) = 1 - \exp\left(-\left\langle \frac{\kappa - \kappa_{th}}{\kappa_0} \right\rangle^m\right) \tag{15}$$

where κ_{th} is a threshold curvature for the onset of damage (buckling), κ_0 a scaling factor, m an exponent (about 1/3), and $\langle \ldots \rangle$ designates the positive part of its argument. Figure 5b shows the above expression fitted to the measured data.

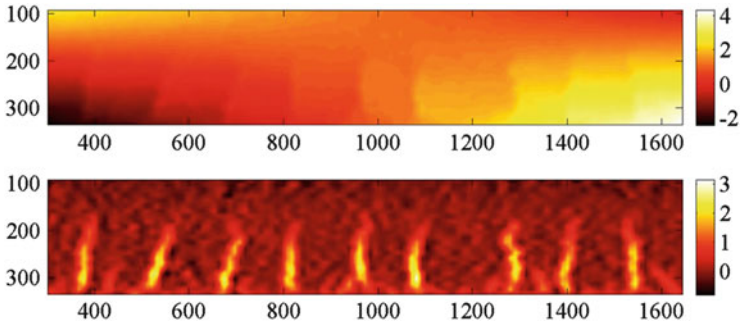


Fig. 6 (*top*) Horizontal component of the displacement field; (*bottom*) apparent strain (in %) field where cracks in plaster are clearly visible. The physical size of one pixel is 37 μm, and the area of the ROI is $\approx 9.3 \times 51.8$ mm^2

Distributed Damage

When Q4-DIC is applied to the four-point bend test of the plasterboard specimen (Fig. 2), at the later stages, it is quite clear that the strain field is not homogeneous anymore. Figure 6 shows the horizontal component of the displacement and of the strain fields over the same region as shown in Fig. 2.

To quantitatively evaluate the progressive degradation of the board up to failure, two routes can be followed. The first one consists of characterizing the cracks as they initiate, propagate, and potentially get arrested. Here the tools shown in Part A may reveal quite useful. However, the interface between the paper lining and the plaster core, which will suffer from delamination (i.e., mode II frictional cracks), is difficult to resolve. Moreover, such a description has to be based on a statistical description, and thus it requires a (very) large number of parameters to define all needed quantities. The second route ignores the presence of the cracks and substitutes to the localized kink due to the cracks a distributed smoothly varying curvature, through an equivalent damage field. The observation that the separation between cracks is very regular signals that although crack initiation is random by nature, the shielding or obscuration mechanism (well described by the shear-lag model (Kelly and Tyson 1965; Aveston et al. 1971)) is such that the stage captured in Fig. 6 is a "saturated" regime where the statistics of initiation is essentially hidden by the load transfer mechanism. Consequently, the behavior can be described by a deterministic process. A damage model is the minimum ingredient for a consistent description of the mechanical response of the board.

The observation of the displacement field sheds some light onto the discussion mentioned at the introduction of this section about the strategy to be used to capture the displacement field. Two strategies were mentioned. The first one used a classical DIC methodology, without any prior assumption, and a post-processing of the two-dimensional displacement field was applied to project the kinematics onto a basis of beam theory. The second one consisted of searching the kinematics

directly in the space spanned by beam theory. The second strategy did not involve any intermediate step, and thus it is consistent from the very beginning of the analysis and involves the minimum cost in information loss from one basis to another one. Hence, it may be considered as the strategy to be favored in all circumstances. Figure 6 however shows the limit of this argument. The actual displacement field that can be measured by DIC is not that with a homogenized vision underlying the damage theory. Hence, in such a case, the two objectives should be distinguished as distinct. On the one hand, the displacement field has to be captured accurately whatever its nature. On the other hand, the best approximation of the actual displacement field is sought within a certain framework (Euler-Bernoulli kinematics and slowly varying damage). Trying to do both at once may ruin both goals, unless it is expected that the representative element size is very small at the scale of the resolution of the imaging techniques (this condition is not met for the plasterboard). In such circumstances, a standard DIC analysis followed by a projection onto a beam kinematics is a recommended choice.

Identification of Distributed Damage

In the case of distributed damage, one strategy is to seek for the local stiffness written as a relative reduction of the flexural stiffness k_0 by a factor that may be written for obvious reason as $(1-D(x))$ although the "damage" $D(x)$ here introduced *does not* refer to any specific constitutive law. The proposed strategy is a direct extension of the previous determination of the geometrical parameters. It is straightforward to add the local damage variable $D(x)$ as additional parameters in the vector of parameters to be determined using the minimization of the quadratic difference of measured and computed deflections T. When referring to a function $D(x)$, what is really meant is a discretization of this function. In the previous (cantilever beam) example, only two regions were considered, and actually only one had a nonzero level D. In a more general framework, $D(x)$ can be considered to be piecewise constant over a partition of the beam into small size elements, and hence characterized by a vector of unknowns, $\boldsymbol{D} = \{D_i\}$. To simplify the discussion, it is assumed that the other geometrical parameters have been determined once for all in the elastic regime, so that the only unknowns are here gathered in $\boldsymbol{D}$.

To apply the previously discussed strategy, the associated sensitivity fields $w_i(x)$, which are equal to the displacement field that corresponds to a variation of damage, i.e., flexural stiffness, over the interval indexed by i, under fixed boundary conditions, need to be evaluated. They can be seen as Green's functions under a specific load that is two opposite torques being applied to the two ends of the considered element. These Green's functions depend on the entire damage state of the beam, but their computation is quite straightforward. For the sake of completeness, rigid body motions are to be added to the library of sought functions w_i.

The following procedure does not involve any new specific developments as compared to the previously discussed resolution. The only point to be noted is that the number of parameters may be large if the discretization of the damage field D is

fine. A large number of unknowns have the natural side effect of deteriorating the conditioning of the system, and high-frequency oscillations in $D(x)$ are to be feared, which have almost no overall effect and that will be activated by the noise inevitably present in the measured displacement fields. One common way to limit this influence is a regularization (Tikhonov and Arsenin 1977) where rapid variations of $D(x)$ are penalized, but this will artificially forbid the occurrence of localization, which may be a genuine mechanical instability possibly leading to a rapid failure. The following discussion indicates another pathway where a kind of regularization is introduced but with a clear mechanical justification.

Measuring a Damage Law

Considering now the strategy where a unique damage law is considered, the difference with the above treatment appears as modest in the general formalism. The constraint of having the same damage law, and hence the same curvature for a given flexural moment, simply restricts the initial degrees of freedom D_i. For instance, in a three-point bend test, the two parts of the beam with respect to the central point would be assumed to behave in the same way. Thus, as compared to the above, only half of the above modes would be considered, but the algebra would remain the same. Of course, the quality of the result is expected to be better because the number of degrees of freedom is reduced. Note that this conclusion however does not rest on the crucial question of the satisfaction of the hypothesized symmetry, as it was seen how taking into account a breakdown of symmetry (as a nonideality of the test) has a very small additional cost. The additional information of D being an increasing function of the curvature, which results from thermodynamic consistency, is much easier to implement.

The specificity of 1D geometries is that the analog of the stress is dictated by the boundary conditions alone (provided all loading forces or moments are known), irrespective of the constitutive law. Thus, the flexural moment $M(x)$ can generally be considered to be known. The problem then reduces to determining the damage $D(\kappa)$ as a function of curvature, for a known reduced moment $m(x) = M(x)/k_0$, which can be read as the curvature of the beam for the same flexural moment and no damage. Therefore, the problem can be written in the following form

$$\begin{aligned} &\big(1 - D\big(\kappa(x)\big)\big)\kappa\big(x\big) = m\big(x\big) \\ &\kappa(x) = \mathrm{d}^2 v\big(x\big)/\mathrm{d}x^2 \end{aligned} \tag{16}$$

where $v(x)$ is measured and $D(\kappa)$ to be determined, with $\psi(\kappa) = (1 - D(\kappa))\kappa$. The previous section dealt with the generic case where function ψ could be inverted, where

$$v(x) = a + bx + \int_0^x \int_0^\xi \psi^{-1}(m(\varsigma))\ \mathrm{d}\varsigma \mathrm{d}\xi \tag{17}$$

It suffices to decompose $\psi(\kappa)$ or $\psi^{-1}(m)$ over a basis of functions (typically with only few parameters) and to apply the previous strategy, through successive

linearizations. However, there is an obvious limit to the present approach, which is the issue of localization.

The Issue of Damage Localization

It may be of interest at this stage to address a rather common situation where $\psi(\kappa)$ cannot be inverted. In many practical cases, the flexural moment goes past a maximum and decreases. In such a case, localization sets in when the maximum is reached. For a three-point bend test, this event corresponds to the situation where under the central load, the flexural moment equals $k_0\psi_{\max}$. Past this point, an apparent brittle failure may take place. For the four-point bend, the discussion is less clear cut, as the entire mid-span reaches the maximum flexural moment at the same instant. Mathematically, a large number of solutions exist, although in practice, slight imperfections of the test will break the symmetry of the loading, and often the most "apparently brittle" response will emerge.

In those situations, in spite of the fact that many solutions may exist, only the ascending branch of ψ can be explored, and hence, ignoring the response past this point, the strategy mentioned in the previous subsection holds. Other interesting cases may be encountered, as illustrated in Fig. 7, where an algebraic form of $D(\kappa)$ was chosen here for illustration purpose. It can be seen that $\psi(\kappa)$ presents an absolute maximum, so that when the curvature exceeds the value for which the maximum is reached, global failure takes place (irrespective of the way the beam is loaded). This ultimate state is indicated with a star in Fig. 7. However, prior to this state, when damage first appears, because of the chosen very sudden rise, the relative flexural moment drops. The 1D geometry where the stress state is not controlled by the constitutive law (but dictated by the loading and geometry) makes the local descending part of the $\psi(\kappa)$ relationship inaccessible, namely, a discontinuity in the curvature is expected. In turn this leads to a discontinuity in the damage experienced along the beam. The condition to avoid instabilities is that $D(\kappa)$ rises more slowly than $1 - \beta/\kappa$ for any β value.

The occurrence of instabilities at intermediate regimes and final failure coupled to localization instabilities is a major source of difficulty for this identification, and thus it is important to test the proposed method with actual data.

Application to Plasterboard

The case of plasterboard is studied. Figure 8 shows two attempts to identify the damage law. In the first case (a), $D(\kappa)$ is chosen as $\langle 1 - \beta/\kappa \rangle$ to avoid any possible instability, but each level of loading of the four-point loading is treated independently. Although the data points obtained from one series can be seen quite clearly on the graph, the quasi-continuity of the data gives some confidence on the treatment. It is to be noted that the trend is to have a decreasing value of β from one load level to the next, so that if a unique algebraic form was chosen, then

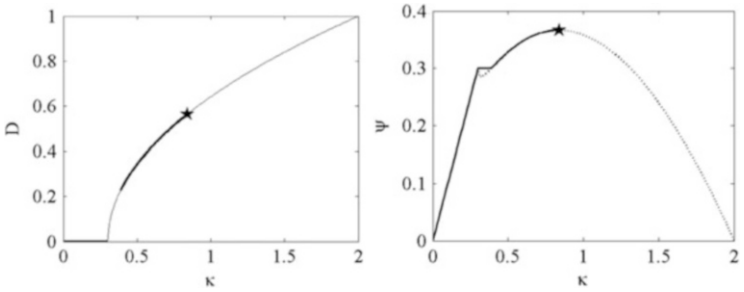


Fig. 7 (*left*) Schematic plot of a damage-curvature relationship (*thin black curve*). Although this function is defined for a large range of curvatures, only the bold segments are accessible to the measurement, and failure takes place at the point indicated by a star. (*right*) Reduced flexural moment, $\psi(\kappa)$, as resulting from the full expression of D (*dotted curve*) and accessible part of the curve (bold curve)

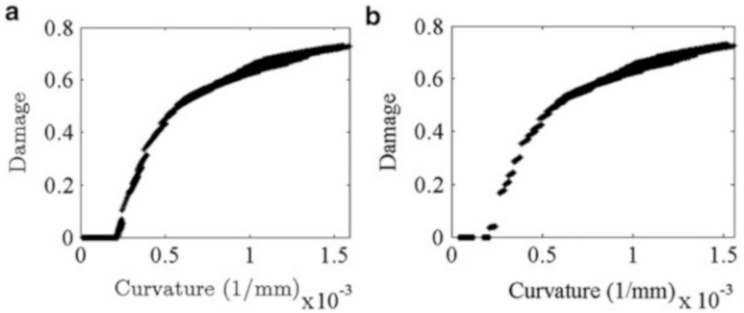


Fig. 8 Damage as a function of curvature as obtained from a series of 69 loading steps each treated independently. (**a**) The algebraic form $D = 1 - \beta/\kappa$ was chosen to limit possible instabilities. (**b**) An exponential form is used to identify the damage growth law

instabilities would be expected. Figure 8b corresponds to another attempt, which is based on a power-law form past a given threshold. Here again, a very good agreement is obtained, both with the previous attempt (signaling that plasterboards are not very far away from localization) and with the expected jumps.

Figure 9a shows that the quality of the deflection adjustment is excellent even if only a partial field of view is considered (the outer span is not visible). Figure 9b also shows that it is very important to allow for a breakdown of symmetry. Although the total force exerted on both inner loading contacts is well controlled, the distribution of the load among the two supports is a priori unknown. In the present case, the two forces were identified together with the damage law (as an additional unknown to vector). It is seen that as soon as the first crack appears, symmetry appears to be broken. The effect is so pronounced that an identification assuming the test to have kept the expected symmetry would certainly fail to

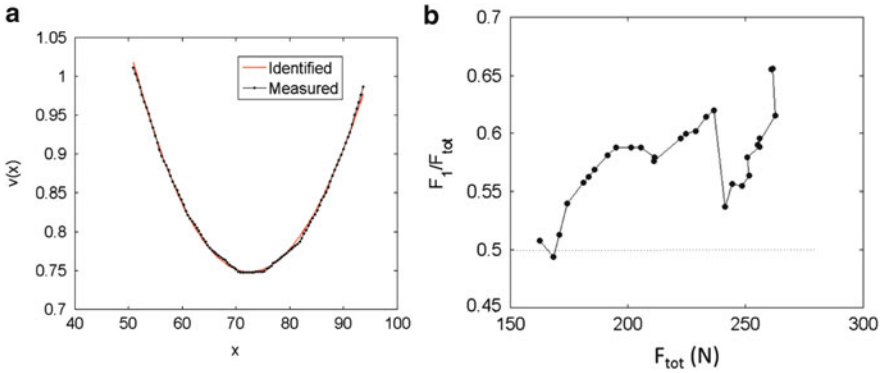


Fig. 9 (**a**) Example measured by DIC (*black*) and computed (*red*) deflection at a late stage of loading for the damage law identified in Fig. 8b. (**b**) Change of the ratio of left to total force as a function of image number. The departure from 0.5 signals a breakdown of the left-right symmetry due to damage

provide a consistent damage law. This test shows that a single test can bring about a considerable amount of information for the identification of the constitutive law and this redundant information may be used to validate the entire procedure. It also emphasizes the importance of staying as close as possible to the experiment. The breakdown of symmetry is not a flaw of the experiment, but the result of an instability that cannot be avoided.

The above treatment was considered within the framework of post-processing the displacement fields obtained from DIC. It is in fact possible to make progress along this direction, namely, as some limitations are considered in the definition of relevant degrees of freedom (for instance, discretization of the damage), this sets severe restrictions on the possible kinematics. Rather than projecting this deflection onto an appropriate subspace, the problem can be formulated with global DIC in such a way that the proposed kinematic basis for DIC actually spans the chosen subspace. This way of interrogating the actual images by a set of precomputed displacement fields obeying balance equations and some constitutive assumptions is called "integrated DIC" (Hild and Roux 2006). This procedure saves the intermediate step of choosing an intermediate basis, such as finite element shape functions, which are modestly suited to the properties assumed for the displacement field, and hence a net gain in uncertainty is to be expected (Roux and Hild 2006). The limit of this strategy, as earlier discussed for the beam-DIC approach, is when the actual displacement field is known to be different from that resulting from the chosen modeling (e.g., the presence of discrete cracks whereas a continuum damage model is chosen). In that case, the two displacement fields have to be distinguished, only the long-wavelength components of both displacements have to be secured, and a control on the amplitude of the difference between both deflections is to be introduced, so that the objective functional remains harmonic for all explored amplitudes.

Lumped Damage Mechanics

Slender geometries can be used to study materials (e.g., steel) or products (e.g., plasterboard), but the most common usage of such geometries is for building design, e.g., for the analysis of buildings from a frame and beam description (Cipollina et al. 1995; Marante and Flórez-López 2003; Guerrero et al. 2007). In this context, it is noteworthy that damage takes place essentially at those points where flexural moments are at their maximum value. The steel cantilever beam is a relevant example. Most of damage is expected to be initiated close to the clamp of the beam. For a hardening behavior, damage could then be thought as concerning larger and larger portions of the beam. However, the slender geometry is such that, in most cases, localization takes place and damage remains concentrated at the beam ends up to failure. This localization is a 1D simplified view for phenomena (e.g., cracking, debonding from reinforcements) that occur throughout the entire section of the beam. Yet, this observation leads to a drastic simplification of the description, where a frame or skeleton of a building can be described as a collection of elastic beams, which will remain elastic forever, in addition to nonlinear local elements located at those few places where flexural moments are the largest. This elegant simplification is called "lumped damage mechanics" (Marante and Flórez-López 2003).

When relevant, lumped damage mechanics has natural consequences on the way the nonlinear (i.e., damageable) point-wise elements should be characterized. From the previous description, it is now clear that an elastic beam, because of its simple kinematics (6 degrees of freedom in 2D situations), should be described as such for "integrated DIC" (Hild et al. 2009). Moreover, in their elastic state, the beams have generally well-known stiffnesses. Thus, a measurement of their kinematics can immediately be translated into corresponding static information. Therefore, a direct estimation of, say, the flexural moment is possible together with the rotation discontinuity (i.e., the appropriate limit of a concentrated curvature). Henceforth, the full set of generalized forces and displacements can be obtained from DIC, whereby the elastic beams actually play the role of force/torque sensors. Moreover, for a complete hyperstatic frame, kinematic continuity and balance equations can be exploited as additional constraints, easily implemented in DIC using Lagrange multipliers (Hild et al. 2011).

This lumped damage mechanics case is illustrative of how a known mechanical information can be used and exploited to take the best out of the DIC measurement as the kinematic descriptors can be tailored to fit the mechanical picture. In so doing, not only is the measurement uncertainty reduced to a minimum, but the amplitudes of the measured kinematic degrees of freedom have a direct mechanical interpretation. It also shows that damage has more than one face, and some freedom can be exerted on the way it is introduced, sometimes as smearing out the effect of localized cracks and, in other cases, as a way to lump altogether all nonlinearities at a single point ruled by an effective nonlinear behavior. In spite of those different faces of damage, DIC can be tuned to bridge the gap between modeling and experiment.

One dream in the long-term future is to have automated identifications based on such an approach to qualify the degradation of simple reinforced concrete structures after an earthquake to decide on possible rehabilitation, restoration, or destruction of buildings with a safe judgment.

Summary

Even though the geometry of the studied systems is very simple in the examples discussed in this part (i.e., beams), different aspects related to damage models have been illustrated when displacement fields are measured via DIC. Multiple cracking and local buckling are two mechanisms modeled in the framework of damage mechanics. In the first case, the damage description is based upon a field, whereas lumped variables are considered in the second example. In both cases, the use of displacement fields has allowed the growth law of damage variables to be constructed. Thus, in this chapter, damage was mostly regarded as mechanical damage, i.e., through its global manifestation in the mechanical behavior rather than through the responsible mechanisms that were considered in the previous chapter. The following chapter (▶ Chap. 39, Evaluating Damage with Digital Image Correlation: C. Applications to Composite Materials) will show that the same philosophy and formalism developed here for 1D elements can be applied to a class of materials designed and engineered to sustain damage. The higher dimensionality will simply require adapting the formulation of identification to a finite element framework.

Acknowledgments Many results reported in this chapter have been obtained thanks to the help of and discussions with A. Bouterf, R. Gras, N. Guerrero, M.-E. Marante, and R. Picon. Part of the work has been supported by ECOS-Nord/CNRS (project entitled "Detection and prediction of the damage state in frames subjected to earthquakes"), by Ile de France region (SESAME project entitled *Plate-forme francilienne d'expérimentations mécaniques de 3^e génération* and DICCIT project), and by Saint-Gobain Recherche.

This chapter has been reviewed by Profs. A. Dragon and J. Lemaitre. The authors would like to warmly thank them for their patience and help in improving the compuscript. Last, this chapter is dedicated to Prof. F. A. Leckie (1929–2013) with whom the authors discussed many issues presented herein.

References

F. Amiot, F. Hild, J.P. Roger, Identification of elastic property and loading fields from full-field displacement measurements. Int. J. Solids Struct. **44**, 2863–2887 (2007)

B. Audoly, Y. Pomeau, *Elasticity and Geometry* (Oxford University Press, Oxford, UK, 2010)

Aveston J, Cooper GA, and Kelly A (1971) Single and Multiple Fracture. In *Proceeing National Physical Laboratory: Properties of Fiber Composites* (IPC Science and Technology Press, Surrey), pp. 15–26.

S. Avril, M. Bonnet, A.-S. Bretelle, M. Grédiac, F. Hild, P. Ienny, F. Latourte, D. Lemosse, S. Pagano, E. Pagnacco, F. Pierron, Overview of identification methods of mechanical parameters based on full-field measurements. Exp. Mech. **48**(4), 381–402 (2008)

A. Cipollina, A. López-Inojosa, J. Flórez-López, A simplified damage mechanics approach to nonlinear analysis of frames. Comput. Struct. **54**(6), 1113–1126 (1995)
M. Grédiac, F. Hild (eds.), *Full-Field Measurements and Identification in Solid Mechanics* (ISTE/Wiley, London, 2012)
N. Guerrero, M.E. Marante, R. Picón, J. Flórez-López, Model of local buckling in steel hollow structural elements subjected to biaxial bending. J. Constr. Steel Res. **63**(6), 779–790 (2007)
F. Hild, S. Roux, Digital image correlation: From measurement to identification of elastic properties – A review. Strain **42**, 69–80 (2006)
F. Hild, S. Roux, Digital image correlation, in *Optical Methods for Solid Mechanics. A Full-Field Approach*, ed. by P. Rastogi, E. Hack (Wiley-VCH, Weinheim, 2012), pp. 183–228
F. Hild, S. Roux, R. Gras, N. Guerrero, M.E. Marante, J. Flórez-López, Displacement measurement technique for beam kinematics. Opt. Lasers Eng. **47**, 495–503 (2009)
F. Hild, S. Roux, N. Guerrero, M.E. Marante, J. Florez-Lopez, Calibration of constitutive models of steel beams subject to local buckling by using Digital Image Correlation. Eur. J. Mech. A/Solids **30**, 1–10 (2011)
A. Kelly, W.R. Tyson, Tensile properties of fibre-reinforced metals: copper/tungsten and copper/molybdenum. J. Mech. Phys. Solids **13**, 329–350 (1965)
P. Leplay, J. Réthoré, S. Meille, M.-C. Baietto, Damage law identification of a quasi brittle ceramic from a bending test using Digital Image Correlation. J. Eur. Ceram. Soc. **30**(13), 2715–2725 (2010)
M.E. Marante, J. Flórez-López, Three-dimensional analysis of reinforced concrete frames based on lumped damage mechanics. Int. J. Solids Struct. **40**(19), 5109–5123 (2003)
F. Pierron, M. Grédiac, *The Virtual Fields Method* (Springer, 2012)
J. Réthoré, S. Roux, F. Hild, Optimal and noise-robust extraction of fracture mechanics parameters from kinematic measurements. Eng. Fract. Mech. **78**(9), 1827–1845 (2011)
S. Roux, F. Hild, Stress intensity factor measurements from digital image correlation: post-processing and integrated approaches. Int. J. Fract. **140**(1–4), 141–157 (2006)
S. Roux, J. Réthoré, F. Hild, Digital image correlation and fracture: an advanced technique for estimating stress intensity factors of 2D and 3D Cracks. J. Phys. D Appl. Phys. **42**, 214004 (2009)
A.N. Tikhonov, V.Y. Arsenin, *Solutions of Ill-Posed Problems* (Wiley, New York, 1977)

Evaluating Damage with Digital Image Correlation: C. Applications to Composite Materials

39

François Hild, Jean-Noël Périé, and Stéphane Roux

Contents

Abstract

The present chapter is devoted to the evaluation of damage with digital image correlation (DIC). Applications will focus on composite materials. The latter ones are designed to accommodate microcracks through suited microstructures. As such, they constitute a natural class of materials for which damage (or rather damages) is an essential feature of their mechanical behavior. As discussed in a previous chapter (addressing detection of physical damage), DIC can reveal the elementary mechanisms (e.g., dense distribution of microcracks, crack branching along weak interfaces, progressive debonding of interfaces, and subsequent pullout or delamination). It will also be shown that damage laws can be identified with the help of DIC from mechanical tests imaged at different stages of loading. The followed strategy will be seen as reminiscent of the one that was used in the previous chapter dedicated to 1D (i.e., beam like) geometries (from physical to mechanical damage). Here, it will be necessary to couple DIC

F. Hild (✉) • J.-N. Périé • S. Roux
LMT-Cachan, ENS Cachan/CNRS/PRES UniverSud Paris, Cachan Cedex, France
e-mail: hild@lmt.ens-cachan.fr; jean-noel.perie@iut-tlse3.fr; stephane.roux@lmt.ens-cachan.fr

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_26

with finite element models. The benefit will be that in addition to the identified law, a full validation is naturally offered from the highly redundant piece of information contained in the measured displacement fields.

Identification of Mechanical Damage

This chapter follows the same footsteps as illustrated in the previous chapter based on beam geometry, but here it is dedicated to two- and three-dimensional systems. In spite of the fact that the examples concern composite materials, the methods are not limited to this class of materials.

When trying to identify damage from a mechanical test, the first step deals with the measurement of displacement fields. One of the measurement techniques is DIC, even though others can be used (e.g., grid methods, holography, and speckle interferometry). The outputs are displacement and strain fields that need to be post-processed. The damage variable(s) is(are) still hidden within these data. Inversion and identification techniques need to be used to extract mechanical damage fields. The problem thus consists of the evaluation of the local elastic properties, expressed as a relative loss of stiffness, $D(\mathbf{x})$ (where D can be a scalar or a tensor). This is a second step in which different metrics can be used to measure the distance between experimental measurements and numerical simulations (Avril et al. 2008; Grédiac and Hild 2012). This is a first route for identifying damage fields.

It is to be emphasized that the extraction of $D(\mathbf{x})$ from necessarily noisy kinematic fields is a difficult inverse problem (Tikhonov and Arsenin 1977). One way to help this determination is to assume that the same damage growth law applies everywhere within an analyzed material or structure. Thus, the unknowns are no longer a damage field but the parameters of a growth law. This represents a very drastic reduction in the number of unknowns, and hence it opens the way to a large reduction in the uncertainty of local damage. The assumption of a homogeneous constitutive law can be (or has to be) questioned. The invaluable advantage of dealing with field measurements, namely, the considerable amount of information, can be exploited to validate all assumptions proposed along the analysis. The pixel-wise comparison of measured and identified displacement fields or even images (e.g., computing the residual $\rho(\mathbf{x})$ fields (see Eq. (A.1)) from these different displacements) brings to light either the quality of the solution or its deficiencies. From the latter, it is generally clear to draw directions of progress or sophistication of, say, the constitutive law to reach a more satisfactory agreement.

A more elaborate strategy consists of coupling the measurement and identification steps. The unknowns to be measured associated with the trial displacement are no longer the standard degrees of freedom (e.g., nodal displacements in a Galerkin approach to DIC) but the parameters of the damage growth law in addition to boundary conditions to make the problem well posed. This type of global approach is referred to as *integrated* since the measured displacement field

is also mechanically admissible and yields mechanical parameters. An example of such an approach dedicated to beams was shown in "▶ Chap. 38, Evaluating Damage with Digital Image Correlation: B. From Physical to Mechanical Dam age" (see also Hild et al. 2011). The chosen displacement basis is either a closed-form solution (e.g., in a Brazilian test to identify elastic parameters (Hild and Roux 2006) or stress intensity factors in elastic media (Roux and Hild 2006)). The results of a numerical simulation can also be used to define the kinematic basis (e.g., in elasticity (Leclerc et al. 2009; Réthoré et al. 2009) or when dealing with damage (Réthoré 2010)).

Types of Damage and Constitutive Laws

When studying composite materials, one underlying difficulty is that the various damage mechanisms are strongly related to the material architecture, namely, particulate composites (Fig. 12 of "▶ Chap. 37, Evaluating Damage with Digital Image Correlation A. Introductory Remarks and Detection of Physical Damage"), composites with short fibers, and composites with continuous fibers (e.g., layered (Fig. 1 of "▶ Chap. 37, Evaluating Damage with Digital Image Correlation A. Introductory Remarks and Detection of Physical Damage"), woven, woven + stitched, interlock configurations) do not experience the same type of degradations. One of the consequences is that there are numerous models to describe their behavior (Orifici et al. 2008). Furthermore, there are also different possible choices for the smallest scale of modeling (i.e., microscopic (Burr et al. 1997), mesoscopic (Ladevèze 1992), or macroscopic (Périé et al. 2009)) to determine the behavior at the macroscopic scale, which is usually needed to run numerical simulations.

One modeling issue is to use either a discrete or continuum description of damage (Hild 2002). This choice will have consequences on the way damage is modeled and therefore experimentally characterized and identified. For example, in section "▶ Cracking in a Layered Composite" of "▶ Chap. 37, Evaluating Damage with Digital Image Correlation A. Introductory Remarks and Detection of Physical Damage," a continuum point of view was used to analyze the DIC results at the macroscopic scale (Figs. 2 and 3 of that chapter) of a 3-layer carbon-epoxy composite. However, mesoscopic observations are also possible, in particular to analyze the damage mechanisms.

Figure 1 shows the longitudinal displacement field where the individual damage mechanisms can be clearly read. It corresponds to an incremental approach where the reference configuration is chosen for a stress level of 1,120 MPa and 1,190 MPa in the deformed configuration. It becomes even clearer when analyzing the longitudinal strain field in which the two transverse cracks are observed. The lower crack, which initiated between these two stress levels, is more open than the upper one that initiated at an earlier stage. When analyzing the shear strain field, it is concluded that transverse cracking is accompanied by mesodelamination along the 0/90° interface for the lower crack. For the upper crack, mesodelamination has not evolved significantly.

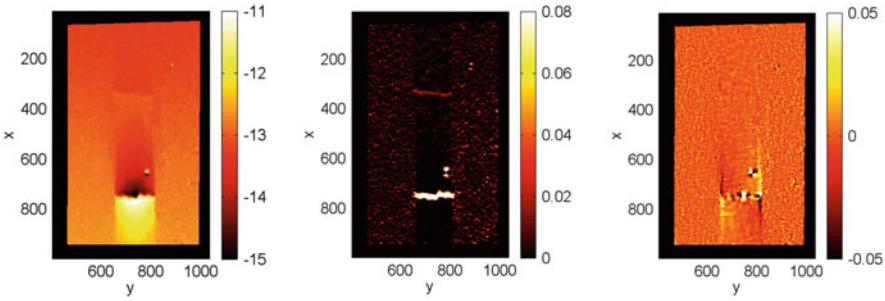


Fig. 1 (*Left*) Vertical displacement expressed in pixels (1 pixel $\leftrightarrow$ 3.5 µm), longitudinal strain (*center*), and in-plane shear strain (*right*) showing a transverse crack and mesodelamination. Note that another transverse crack ($x \approx 380$ pixels) that was created at a previous stage of loading is also visible. These results were obtained with regularized T3-DIC ($\ell = 5$ pixels, $\ell_m = 20$ pixels, $\ell_b = 10$ pixels). The region of interest has an area of $\approx 3 \times 2$ mm^2

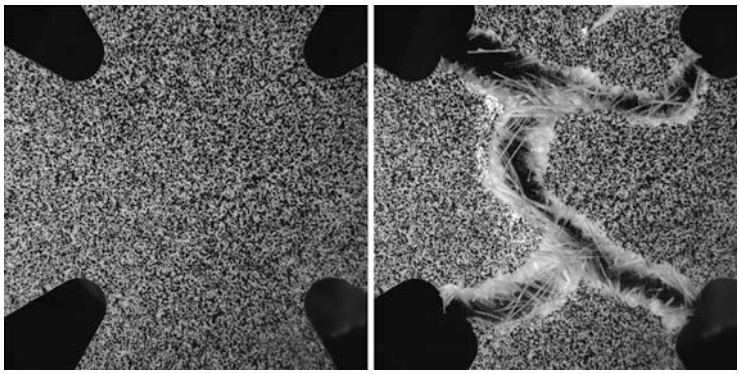


Fig. 2 Image of the sample prior to loading (*left*) and after failure (*right*). The observed region has an area of $\approx 68 \times 68$ mm^2

At this scale, an appropriate mechanical description would imply modeling the crack pattern in details, in particular, the combined effect of transverse cracking and mesodelamination, and the fact that the stresses in the vicinity of existing cracks are partially relaxed so that crack initiations are prevented very close to existing ones (this is a nonlocal effect). This hypothesis (Curtin 1991) allows the cracking pattern observed at the mesoscopic scale to be understood. It can be generalized to other situations dealing with multiple cracking (Hild 2002; Malésys et al. 2009; Forquin and Hild 2010; Guy et al. 2012).

When modeling the mechanical behavior of composite materials, the framework of continuum thermodynamics (Germain et al. 1983) turns out to be particularly powerful when applied to damage models. When cracks are described, it is more natural to express the state potential in terms of Gibbs' free enthalpy. However, in

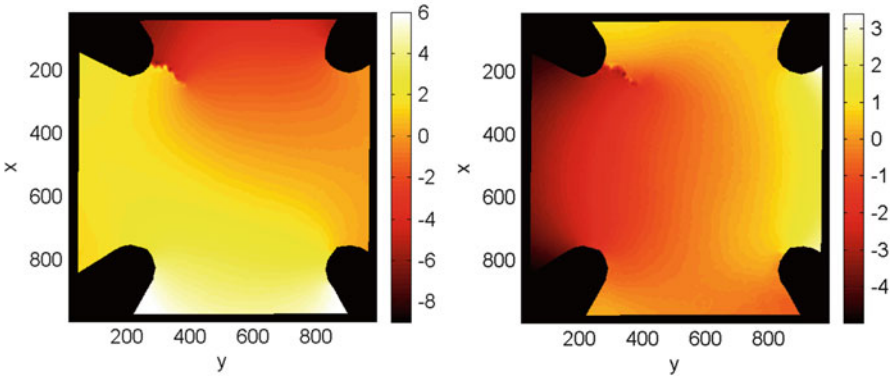


Fig. 3 Vertical (*left*) and horizontal (*right*) component of the displacement field (expressed in pixels) obtained from Q4-DIC ($\ell = 16$ pixels). The physical size of one pixel is 68 μm, and the region of interest has an area of $\approx 68 \times 68$ mm^2

numerical simulations, most of the formulations are displacement based, and it is therefore more convenient to consider Helmholtz' free energy density

$$\Psi = \frac{1}{2}\boldsymbol{\varepsilon}^e : \mathbf{C}(\mathbf{D}) : \boldsymbol{\varepsilon}^e + \Psi^{s-r}\left(\boldsymbol{\varepsilon}^i, \boldsymbol{\varepsilon}^c, \mathbf{d}, \mathbf{D}\right) \quad \text{with} \quad \boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^e + \boldsymbol{\varepsilon}^i + \boldsymbol{\varepsilon}^c \tag{1}$$

where $\boldsymbol{\varepsilon}$ is the infinitesimal strain tensor, $\boldsymbol{\varepsilon}^e$ the elastic strain tensor, $\boldsymbol{\varepsilon}^c$ the creep strain tensor (when needed (Begley et al. 1995; Du et al. 1997; Burr et al. 2001)), $\boldsymbol{\varepsilon}^i$ the inelastic strain tensor modeling all damage mechanisms leading to frictional sliding (e.g., fiber/matrix sliding, mesodelamination), $\mathbf{C}$ the macroscopic Hooke's tensor that contains various damage variables $\mathbf{D}$, and Ψ^{s-r} the stored and relaxed energy densities (Boudon-Cussac et al. 1998). In many models, inelasticity is described by resorting to isotropic hardening (e.g., Ladevèze and Le Dantec 1992). However, by using homogenization techniques, it is more natural to choose kinematic hardening (Andrieux et al. 1986; Hild et al. 1996) because it is associated with frictional sliding.

It can also be noted that when frictional sliding occurs, damage variables $\mathbf{d}$ are involved in the expression of the stored part of the state potential (Andrieux et al. 1986; Burr et al. 1997). Conversely, the relaxed part is caused by damage variables $\mathbf{D}$ that also lead to a stiffness loss (e.g., matrix-cracking in the presence of an initial residual stress field induced, say, by coefficient of thermal expansion mismatches (Budiansky et al. 1986; Boudon-Cussac et al. 1998) that is partially relaxed). Creep also involves a change in the self-balancing stress field, and therefore Ψ^{s-r} varies accordingly. Interfacial wear is a fatigue mechanism of many composite materials (Rouby and Reynaud 1993). It leads to a variation of the interfacial properties that influence the inelastic strain and the stored part of Ψ^{s-r} (Burr et al. 1998).

Furthermore, the state potential is generally postulated, and then the growth laws of the internal variables are written in terms of their associated forces obtained as partial derivatives of the state potential with respect to the former or combinations thereof. One critical issue to properly compute the dissipated energy is to evaluate

the part of the state potential Ψ^{s-r} that is stored or relaxed. The latter is the elastic energy density associated with the residual stresses induced by all dissipative mechanisms (Boudon-Cussac et al. 1998; Vivier et al. 2009, 2011).

Last, crack closure may occur, typically for mode I microcracks. This closure induces modeling issues that are not easy to tackle (Ladevèze et al. 1983; Chaboche et al. 1990; Desmorat 2000; Halm et al. 2002). Special care has to be exercised to avoid numerical and theoretical difficulties. This point will not be addressed herein. However, it can be noted that DIC (Hamam et al. 2007) and DVC (Limodin et al. 2009) can be used to analyze this phenomenon at different scales.

Damage in Low-Cost Composite

Two approaches are followed to study a vinylester matrix reinforced by an isotropic distribution of E-glass fibers. The first one consists of determining the contrast field of elastic stiffness, which can be interpreted as a signature of damage in a CDM context thanks to the coupling between elasticity and damage. The second one aims at identifying the growth law of an isotropic damage variable. Since only one internal (damage) variable is considered, any nonlinearity is attributed to damage. Consequently, the identification procedure does not need unloadings.

A thin plate made of this composite is prepared as a cross with wide arms, and subjected to biaxial loading (Fig. 2). The white surface of the test piece is sprayed with black paint so as to produce a fine random texture, which is needed for DIC. Digital images of the surface are shot for every 1-kN load increment in both directions up to complete failure, which occurred for 11.1 kN. Thus, 11 images are available for the analysis. The physical size of one pixel is 68 μm. In the sequel, this experiment is used to illustrate the results that are achieved when first seeking a stiffness field and second when identifying a damage law.

Inversion: Determination of Stiffness Field

In this first case, a field of elastic properties that account for the measured displacement field is sought. In the present case, Q4-DIC is considered (Fig. 3). The number of degrees of freedom in the inverse problem is of the same order of magnitude as the data (here the measured displacement field). The solution consists of the *inversion* of the mechanical problem.

It is assumed that the damage mechanism induces a heterogeneous contrast field such that the local Young's modulus is reduced to $\chi(\mathbf{x})E_0$ from its initial value E_0 while the Poisson's ratio, ν, remains unaltered. The equilibrium gap method (Claire et al. 2004) consists of exploiting the equilibrium equation

$$\mathrm{div}[\chi(\mathbf{x})\mathbf{C}_0\boldsymbol{\varepsilon}(\mathbf{u}(\mathbf{x}))] = \mathbf{0} \tag{2}$$

written here in the absence of body forces. The strain tensor $\boldsymbol{\varepsilon}$ is computed from the measured displacement field, and $\mathbf{C}_0$ is Hooke's tensor of the virgin material. Since

the displacement field is decomposed over a finite element basis (made of Q4 elements), the corresponding discretization of the "equilibrium gap" f_i is considered

$$f_i = \chi^m K_{ij}^m u_j = 0 \tag{3}$$

where K_{ij}^m is a component of the elementary rigidity matrix of the undamaged element m relating the displacement component u_j to the nodal force f_i^m (whose sum over all m elements should be equal to 0 for inner nodes). The rigidity matrix is linearly dependent upon the contrast χ. As can be seen from Eq. (3), since no static information is used (the right-hand side is vanishing at all inner nodes), the χ-field is determined up to an arbitrary scale factor. This is the consequence of limiting the analysis to the knowledge of kinematic data only. Only elastic *contrasts* can be determined in the present case. The formulation is complemented by arbitrarily prescribing that the average χ is equal to a constant. This is achieved by using a Lagrange multiplier or by eliminating one contrast value.

The inversion problem is ill-posed and some regularization is called for. The most straightforward approach is to search for the "best" solution for the χ-field in a subspace of smoothly varying fields in space. Q4 finite elements are used in the following, with however a mesh that is independent of (and coarser than) the measurement mesh. Shape functions N_i^m are introduced to provide the weight of the center of an element m for the i-th basis function

$$\chi^m = N_i^m b_i \tag{4}$$

where b_i are the unknown contrasts collected in a vector $\{\mathbf{b}\}$. This regularization limits the number of degrees of freedom $\{\mathbf{b}\}$, and hence the contrast field is obtained through the minimization of the global equilibrium gap

$$W(\{\mathbf{b}\}) = \sum_j \left(\sum_{i,m} L_j^m N_i^m b_i \right)^2 - \lambda \sum_{i,m} N_i^m b_i \tag{5}$$

where $L_j^m = K_{jk}^m u_k$ and λ is the Lagrange multiplier of the average contrast constraint. The minimization of functional W results in a linear problem that provides $\{\mathbf{b}\}$ amplitudes and, thus from Eq. (4), the contrast value in each element.

The final stage of loading is considered with a coarse mesh (10×10 elements) and a finer one (20×20 elements) for the contrast field (Fig. 4), which is to be compared with a 67×66 element mesh for the kinematic measurements (Fig. 3). In both cases, the stiffness reached a negative value at the top left corner (-0.02 and -0.3, respectively, for the coarse and fine meshes). This value has been artificially reset to 0.01. A satisfactory agreement between both results is obtained with a clear detection of crack initiation and propagation.

To evaluate the quality of the obtained contrast map, it is possible to resort to a standard elastic computation based on the determined stiffness contrast and using

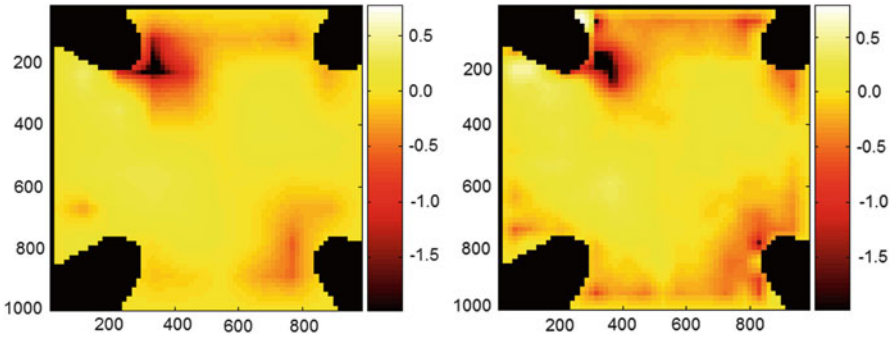


Fig. 4 Map of $\log_{10}\chi$ obtained for a coarse (10×10 elements, *left*) and fine (20×20 elements, *right*) mesh. The region of interest has an area of $\approx 68 \times 68$ mm^2

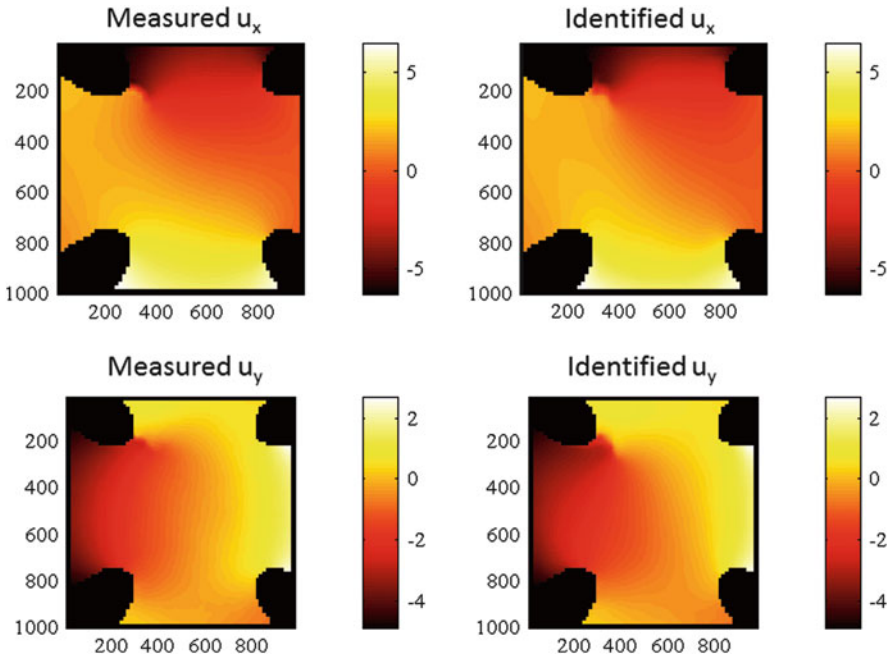


Fig. 5 Comparison between measured by Q4-DIC with $\ell = 16$ pixels (*left*) and recomputed (*right*) displacements expressed in pixels from the identified contrast field (fine mesh: 20×20 elements). The physical size of one pixel is 68 μm, and the region of interest has an area of $\approx 68 \times 68$ mm^2

Dirichlet boundary conditions, i.e., prescribing the displacement field on the boundary of the considered domain. The computed displacement field is then compared with the measured field in Fig. 5. A good agreement between both fields is observed.

To quantify the agreement, the following dimensionless "residual" ρ_u is defined: it is the standard deviation of the difference between identified and measured displacement

fields, normalized by the standard deviation of the measured displacement field. In the present case, for a fine mesh (i.e., 20×20 elements), this residual amounts to 14 %. The latter level is fair, and part of the discrepancy is related to the localized nature of damage, which cannot be fully captured with the chosen discretization.

Identification: Damage Growth Law

When a series of pictures is analyzed, there is no link between the contrast fields that can be determined by following the previous procedure. In particular, the choice of the additional condition to make the system invertible should be adapted to translate the contrast fields into damage fields. A growth law can be identified, but only by post-processing the previous results (Claire et al. 2007).

In the following, a regularized approach (Roux and Hild 2008) will be developed to move from an inversion to an identification (of a few material parameters). The spirit of the method is to require that elements with the same equivalent strain should also have the same damage level. This is a very strong requirement that allows the number of unknowns to be drastically reduced.

An isotropic description of damage is assumed to be valid. The behavior will be modeled by a unique scalar, $D(\mathbf{x})$, such that the Young's modulus is reduced to $(1-D)E_0$, from its initial value E_0, while the Poisson's ratio, ν, remains unaltered. The growth law of the damage variable D is described by a function of an equivalent strain ε_{eq} to be defined later on. The fact that D is related to ε_{eq} makes the identification easier since, if the damage parameter is assumed to be uniform over, say Q4 elements of the measurement mesh, it suffices to compute the equivalent strain per element before starting the identification procedure. The equilibrium gap norm is now written as

$$\mathfrak{R} = \sum_j \left(\sum_e L_j^e (1 - D^e) \right)^2 \tag{6}$$

and a decomposition of the damage growth law is chosen

$$D = \sum_i c_i \varphi_i\left(\widetilde{\varepsilon}_{\text{eq}}\right) \quad \text{with} \quad \widetilde{\varepsilon}_{\text{eq}}(t) = \max_{0 \le \tau \le t} \varepsilon_{\text{eq}}(\tau) \tag{7}$$

where φ_i are chosen functions (e.g., based on exponentials (Burr et al. 1997; Baptiste 2002)) and c_i the unknown parameters to identify (their number remains limited to a few units). The minimization of the equilibrium gap leads to the following linear system

$$\sum_p \sum_{e,f} \left(\sum_j L_j^e L_j^f \right) \varphi_p\left(\widetilde{\varepsilon}_{\text{eq}}\right) \varphi_q\left(\widetilde{\varepsilon}_{\text{eq}}\right) c_p = \sum_{e,f} \left(\sum_j L_j^e L_j^f \right) \varphi_q\left(\widetilde{\varepsilon}_{\text{eq}}\right) \tag{8}$$

The additional requirement prescribed to set the contrast scale can be dropped if the condition $\varphi_i(0) = 0$ is satisfied. It is worth noting that even though the initial problem is strongly nonlinear, the final formulation leads to a linear system.

As introduced above, the operator $\mathbf{L}$ is linear in terms of the measured displacements and involves the (uniform) elastic properties of the virgin material. The sum $\sum_e L_j^e = f_j$ can be interpreted as a nodal resultant force. The elastic problem with prescribed Dirichlet boundary conditions and known body forces is well posed and can be inverted to evaluate the nodal displacements

$$u_i = S_{ij} f_j \tag{9}$$

with $[\mathbf{S}] = [\mathbf{K}_0]^{-1}$, where $[\mathbf{K}_0]$ is the stiffness matrix of the virgin material. By noting that $\mathbf{L}$ is a second-order differential operator, the following form

$$\widetilde{\mathfrak{R}}(c_k) = \left\| \sum_j S_{ij} \sum_e L_j^e \left[1 - \sum_k c_k \varphi_k\left(\widetilde{\varepsilon}_{eq}\right) \right] \right\|^2 \tag{10}$$

allows the identification problem to be better conditioned when compared to Eq. (6). By noting that $[\mathbf{S}][\mathbf{L}] = \{\mathbf{u}^{\text{meas}}\}$, the reconditioned equilibrium gap becomes a distance written in terms of displacements and not its second-order derivatives

$$\widetilde{\mathfrak{R}}(c_k) = \left\| u_i^{meas} - \sum_j S_{ij} \sum_e L_j^e \sum_k c_k \varphi_k\left(\widetilde{\varepsilon}_{eq}\right) \right\|^2 \tag{11}$$

so that the identification will be less sensitive to measurement uncertainties.

The damage model tuned hereafter uses Helmholtz' free energy density Ψ when elasticity is coupled with damage ($\varepsilon = \varepsilon^e$ and $\Psi^{s-r} = 0$), which is expressed as (Marigo 1981)

$$\Psi = \frac{1}{2} \boldsymbol{\varepsilon} : \mathbf{C}_0 (1 - D) : \boldsymbol{\varepsilon} \tag{12}$$

so that the thermodynamic force Y associated with the damage variable D reads

$$Y = -\frac{\partial \Psi}{\partial D} = \frac{1}{2} \boldsymbol{\varepsilon} : \mathbf{C}_0 : \boldsymbol{\varepsilon} \tag{13}$$

Consequently, the equivalent strain ε_{eq} becomes, under a plane stress hypothesis,

$$\varepsilon_{\text{eq}}^2 = \frac{Y}{E_0} = \frac{\langle \varepsilon_1 \rangle^2 + 2\nu \langle \varepsilon_1 \rangle \langle \varepsilon_2 \rangle + \langle \varepsilon_2 \rangle^2}{2(1 - \nu^2)} \tag{14}$$

where ε_1, ε_2 are the two in-plane eigen strains and ν Poisson's ratio of the undamaged material. The functions φ defining the damage growth law (Eq. (7)) are assumed to be described by exponentials

$$\varphi_i\left(\widetilde{\varepsilon}_{\text{eq}}\right) = 1 - \exp\left(\frac{-\widetilde{\varepsilon}_{\text{eq}}}{\varepsilon_{ci}} \right) \tag{15}$$

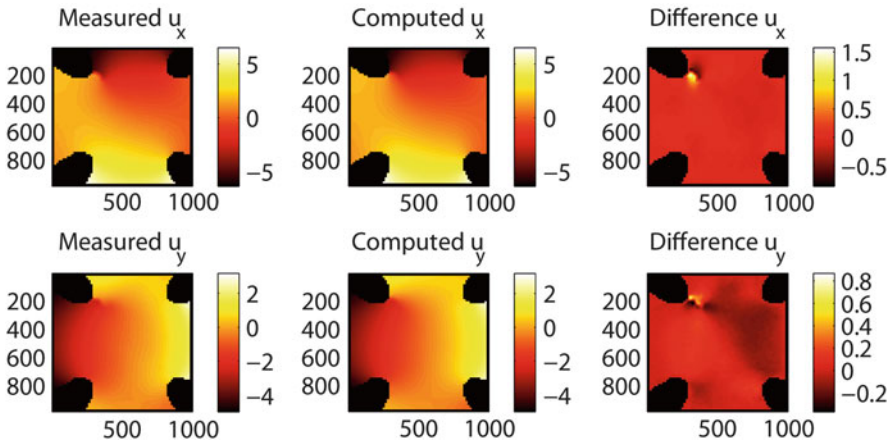


Fig. 6 Comparison between measured (*left*), identified (*center*) displacement fields at the final step of loading. The difference between the two displacements is also shown (*right*). The displacements are all expressed in pixels. The physical size of one pixel is 68 μm, and the region of interest has an area of $\approx 68 \times 68$ mm^2

where ε_{ci} are characteristic strains to be identified. Different characteristic strains in the damage law test functions were considered $\varepsilon_{ci} = 0.0067 \times (1, 2, 4, 8, 16)$ consistently with the range of equivalent strains encountered in the experiment. The five amplitudes were $\mathbf{c} = (0.87, 0, 0, 0, 0.13)$. The quality of the analysis appears to be good, namely, $\rho_u = 0.03, 0.03, 0.03$, and 0.05 for the last four load levels. It is worth noting that it is significantly lower than what was observed in the analysis of contrast fields.

Figure 6 shows a comparison between the measured and predicted displacement fields for the last load level. The fact that the quality deteriorates in this last level is due to a well-developed crack on the top left part of the sample. The crack is both crudely accounted for by the scalar damage model, but also presumably badly captured by the image correlation algorithm, which is designed for continuous displacement fields.

The damage map is very informative (Fig. 7a). Crack initiation on the left hand and top corner of the specimen is clearly captured, as also observed in the picture of step 11 (Fig. 2a) when magnified. Damage concentrations at other corners are also observed and correlate rather well with the final cracking pattern (Fig. 2b). The damage growth law consists of two regimes (Fig. 7b), possibly one prior to and one after damage localization. This problem will be addressed in section "Damage Localization Versus Cracking."

Anisotropic Damage Description for a Layered Composite

In the present section, the heterogeneity of the strain fields is enforced geometrically to ease the identification of the growth law. Consequently, the classical ±45° tensile test on coupons was modified by adding a lateral notch to the specimen.

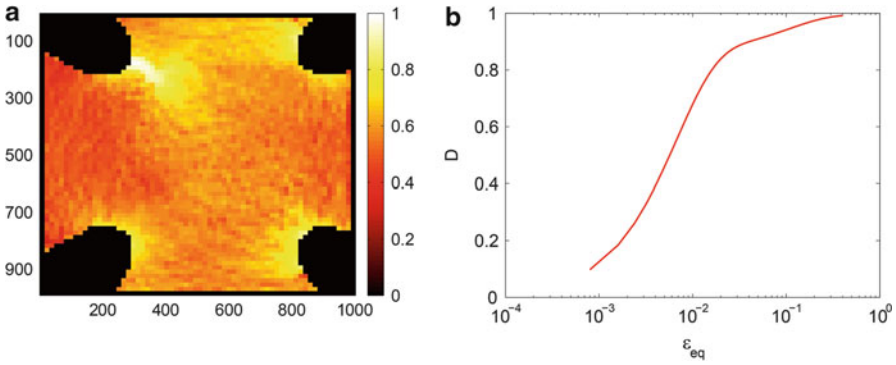


Fig. 7 (**a**) Maps of D for the last step of loading. One clearly sees in the left-hand top corner the initiation of a major crack that will lead to failure of the sample. Moreover secondary crack formations are also distinguished close to the other corners (see Fig. 2b for a detailed comparison with the final failure pattern). (**b**) Identified damage law

Each layer of the composite is made of unidirectionally aligned carbon fibers in a thermoset matrix. Contrary to the previous case, an anisotropic damage description is considered to describe the damage mechanism relative to shearing of the matrix. Gibbs' free enthalpy Φ of the composite then reads

$$\Phi = \frac{1}{2}\left[\frac{\sigma_{11}^2}{E_1} - \frac{\nu_{12}}{E_1}\sigma_{11}\sigma_{22} + \frac{\sigma_{22}^2}{E_2} + \frac{\sigma_{12}^2}{G_{12}(1-D)}\right] \tag{16}$$

where σ_{11}, σ_{22}, σ_{12} are the in-plane stress components expressed in the material frame (1, 2). The elastic properties are E_1 and E_2 (Young's moduli along the fiber directions), ν_{12} (Poisson's ratio), and G_{12} (in-plane shear modulus). The in-plane strain tensor $\boldsymbol{\varepsilon}$ is expressed as

$$\boldsymbol{\varepsilon} = \frac{\partial \Phi}{\partial \boldsymbol{\sigma}} \tag{17}$$

where $\boldsymbol{\sigma}$ is Cauchy's stress tensor and the thermodynamic force Y associated with the damage variable D becomes

$$Y = \frac{\partial \Phi}{\partial D} = \frac{\sigma_{12}^2}{2G_{12}(1-D)^2} = 2G_{12}\varepsilon_{12}^2 \tag{18}$$

Consequently, the equivalent strain becomes

$$\widetilde{\varepsilon}_{\text{eq}}(t) = \max_{0\le\tau\le t} |\varepsilon_{12}(\tau)| \tag{19}$$

The elementary stiffness matrix thus reads

$$[\mathbf{K}] = [\mathbf{K}_0] - D[\mathbf{K}_1] \tag{20}$$

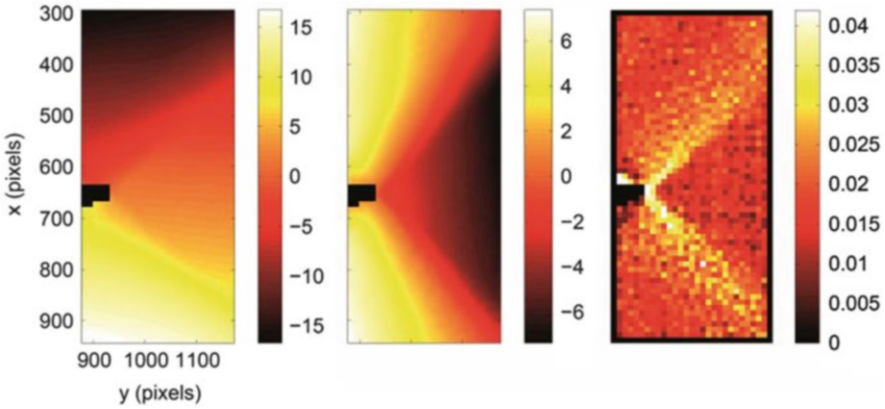


Fig. 8 Displacement maps u_x (*left*) and u_y (*center*) expressed in pixels (1 pixel $\leftrightarrow$ 36 µm) at maximum load (along the x-direction) of the notched specimen with a 12-pixel discretization. The rigid body motion was removed. Corresponding equivalent strain map ε_{eq} (*right*). Strain localization (along the fiber directions) is clearly visible. The region of interest has an area of $\approx 23.4 \times 10\ \text{mm}^2$

where $[\mathbf{K}_1]$ has only one nonzero (shear) term that is affected by damage. The reconditioned equilibrium gap becomes (Périé et al. 2009)

$$\breve{\mathfrak{R}}(c_k) = \left\| u_i^{meas} - \sum_j S_{ij} \sum_e \overline{L}_j^e \sum_k c_k \varphi_k\left(\widetilde{\varepsilon}_{eq}\right) \right\|^2 \tag{21}$$

with $\overline{L}_i^e = K_{ij}^{1e} u_j$.

In the following, 12-pixel Q4 element sizes are considered. The displacement maps are shown in Fig. 8 for the maximum load level (i.e., the last analyzed cycle). They are measured when the reference picture is the unloaded step after the maximum load level, which corresponds to the deformed configuration. The corresponding equivalent strain field is estimated by computing the average value in each finite element. The strain localization is clearly observed in the vicinity of the highly damaged (V-shaped) zones. Only the last of 14 loading/unloading cycles is shown herein. More results can be found in (Ben Azzouna et al. 2011).

By using the same expression as before for the damage law (15), the identification results yield the unknowns c_k for chosen characteristic strains ε_{ci}. The best solution is obtained when $\varepsilon_{ci} = 0.016 \times (1, 2, 4, 8)$ and $\mathbf{c} = (1.0, 0, 0, 0)$. The quality of the identification is first assessed globally by computing ρ_u. For the last load level, $\rho_u = 3.7$ %. This is a very low value giving confidence in the identification result. From the measurement of the equivalent strain map and the identified parameters of the damage law, it is possible to construct the damage maps for each analyzed cycle (Fig. 9 for the last five cycles).

For the last considered cycle (#14), a localized damage pattern is observed with a value of D approaching 1. This is in good accordance with the experimental failure

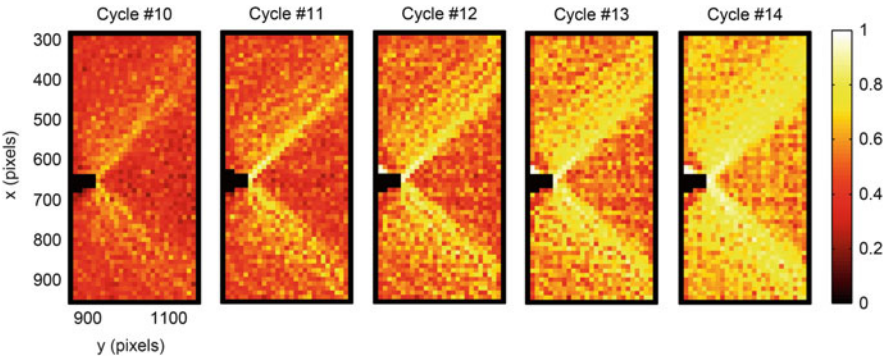


Fig. 9 Identified damage maps for the last five analyzed cycles

pattern (Ben Azzouna et al. 2011). The best solution is obtained with only one parameter ε_{c1} for which c_1 is therefore close to 1.

A first way of validating the identification results is to compare the computed displacement field using the damage law that was identified with the measured displacement field and then estimate the displacement residuals. Figure 10 shows the three maps corresponding to the displacement components for cycle #14. A very good agreement is observed.

Another way of validation is given by the damage law *itself*. The latter is compared with that obtained by following the classical identification procedure (Ladevèze and Le Dantec 1992). In the latter, only a single longitudinal strain level is available per unloading/loading cycle; from 10 to 15 points are generally accessible. For example, Fig. 11 shows the result from the analysis for which only five points lie above the damage threshold. With the present approach, the same results are shown in Fig. 11. Numerous identification points are available thanks to full-field data. It is also to be noted that the present approach allows for the identification of damage levels greater than 0.4, level at which the global method stops because of failure that occurs suddenly. In the range over which the two results can be compared, there is a good agreement, thereby validating the approach followed herein. It is to be remembered that the damage functions (15) do not incorporate a threshold parameter. This is the main difference between the two results. However, it is believed that this value is difficult to capture with fine meshes for which the measurement uncertainty is not sufficiently small.

Damage Localization Versus Cracking

The example of the cross-shaped composite sample discussed in section "Damage in Low-Cost Composite" is considered again. On the measurement side, a continuous displacement field was evaluated even at advanced stages of the experiment (Fig. 3). By analyzing the correlation residuals, it could have been concluded that the displacement continuity is questionable (Fig. 12), especially at the end of the experiment.

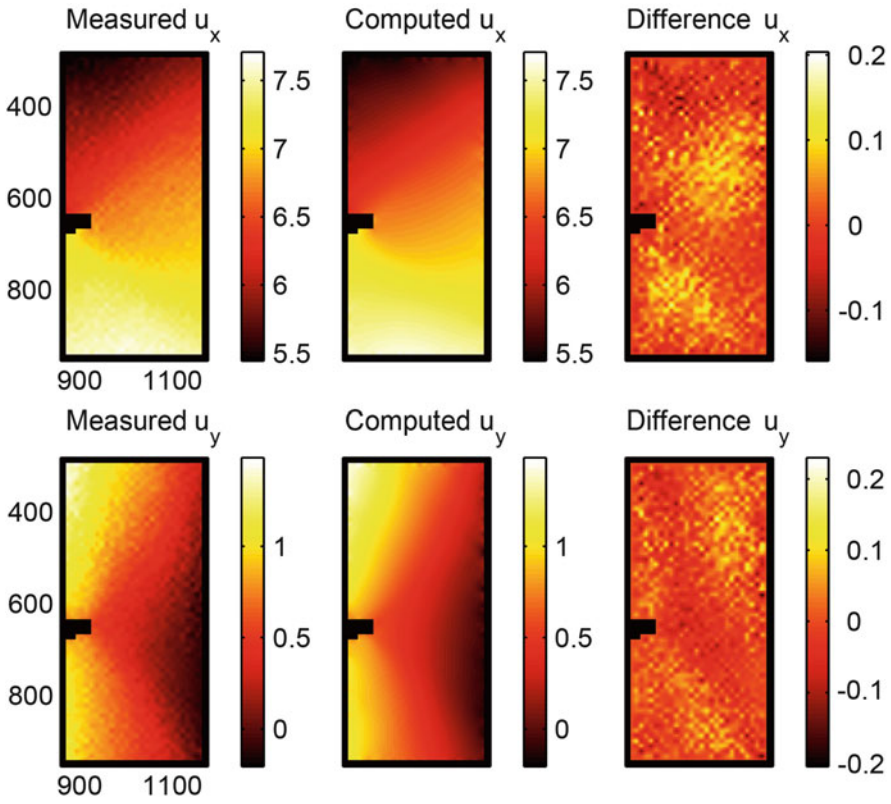


Fig. 10 Comparison of the measured and computed displacement fields for the last cycle (#14) for a 12-pixel discretization. The physical size of one pixel is 36 μm, and the region of interest has an area of $\approx 24.3 \times 10$ mm^2

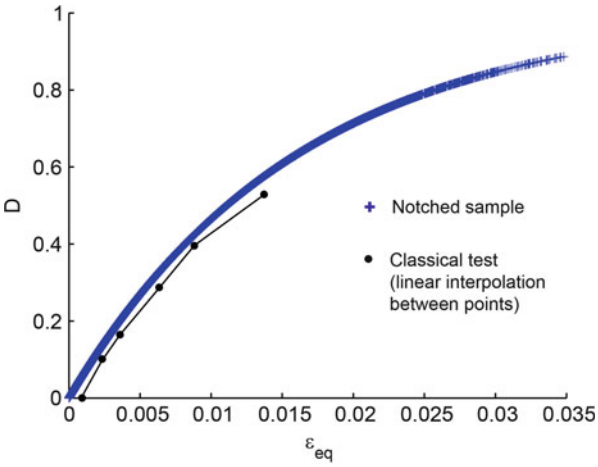


Fig. 11 Damage law identified by following the classical approach (few strain data are available) and by resorting to full-field measurements and the equilibrium gap method. Each cross corresponds to a measured strain

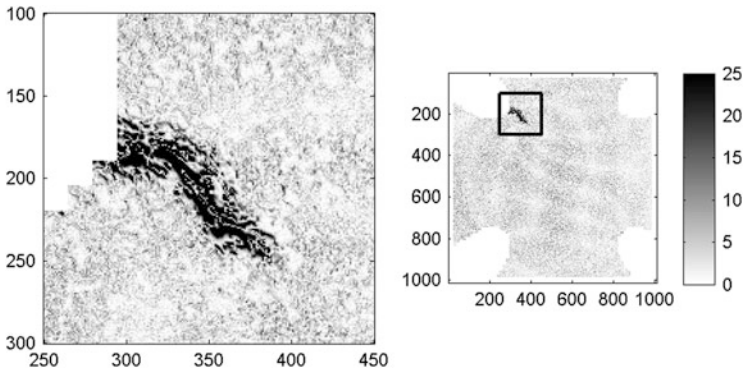


Fig. 12 Correlation residuals expressed in gray levels (the digitization of the analyzed pictures is 8 bits; see Fig. 2) for the displacement field shown in Fig. 3. The region of interest has an area of $\approx 68 \times 68$ mm^2. The crack is visible on the *top left corner* (see detail on the *left*)

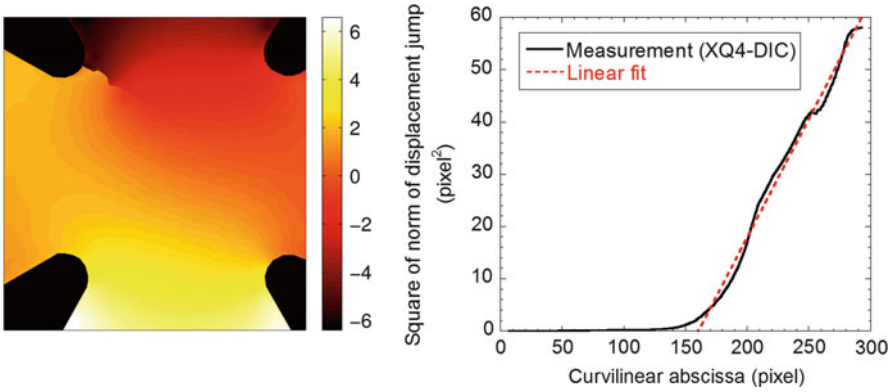


Fig. 13 Vertical (*left*) component of the displacement field (expressed in pixels) obtained from XQ4-DIC ($\ell = 32$ pixels). The region of interest has an area of $\approx 68 \times 68$ mm^2. Norm squared of displacement jump versus curvilinear abscissa (*right*). The physical size of one pixel is 68 μm (Courtesy of J. Réthoré)

To deal with displacement discontinuities in a global DIC context, two different routes can be followed. First, extended DIC (Réthoré et al. 2008) can be considered (e.g., XQ4-DIC). As in extended finite elements (Black and Belytschko 1999; Moës et al. 1999), it consists of enriching the displacement basis with discontinuous terms. Second, node splitting is also possible (Roux et al. 2012). Figure 13 shows the result obtained with XQ4-DIC. In particular, the crack opening displacement profile can be used to extract the stress intensity factor. In the plot of Fig. 13, the slope of the linear interpolation is equal to $8K/E_0\sqrt{2\pi}$, so that the value of the stress intensity factor K is 16 MPa $\sqrt{\text{m}}$.

On the modeling side, two regimes appear in the damage growth law shown in Fig. 7. The second one corresponds to strain levels that are greater than those

observed in tensile tests on the same material. They are related to the existence of a macrocrack whose description with CDM concepts may be questioned. To link linear elastic fracture mechanics and CDM, the dissipated energy (Lemaitre and Dufailly 1987) Δ is calculated by assuming a constant critical energy release rate G_c

$$\Delta = G_c a h \tag{22}$$

where a denotes the crack length and h the sample thickness. For the damage model, the dissipated energy density is first calculated for any broken element ($D = 1$)

$$\delta = \int_0^1 Y dD \tag{23}$$

With the chosen growth law, the dissipated energy density becomes

$$\delta = 2E_0 \sum_i c_i \varepsilon_{ci}^2 = 2Y_c \tag{24}$$

so that the dissipated energy of n broken elements reads

$$\Delta = n\ell^2 h\delta \tag{25}$$

where Y_c is the characteristic energy release rate density. By noting that $a \approx n\ell$, it follows

$$G_c = 2\ell Y_c \tag{26}$$

This result shows that the element size ℓ explicitly appears in the relationship between the critical energy release rate and the characteristic energy release rate density.

With the identified parameters of the damage growth law (see section "Identification: Damage Growth Law"), an estimate of the fracture toughness K_c is obtained

$$K_c = \sigma_c \sqrt{2\ell} \quad \text{with} \quad \varepsilon_c = \sqrt{\sum_i c_i \varepsilon_{ci}^2} \quad \text{and} \quad \sigma_c = E_0 \varepsilon_c \tag{27}$$

where σ_c denotes the characteristic strength and ε_c the characteristic strain. The fact that $\varepsilon_{c5} \gg \varepsilon_{c1}$ leads to the following approximation

$$K_c \approx E_0 \varepsilon_{c5} \sqrt{2\ell c_5} \tag{28}$$

This result proves that the second regime (i.e., for strain levels greater than 1 %) of the damage growth law (Fig. 7b) is associated with a localized mode (i.e., crack propagation) and not with a diffuse mechanism. The element size plays the role of a

nonlocal parameter. However, its physical meaning is lost since it was chosen at the measurement stage and not for any physical reason.

In the present case, it is found that $\sigma_c \approx 360$ MPa and $\varepsilon_c \approx 0.04$, with $\ell \approx 1.1$ mm and $E_0 = 9.3$ GPa. Consequently, the critical energy release rate becomes $G_c = 31$ kJ/m^2, and the corresponding fracture toughness $K_c = 17$ MPa$\sqrt{\text{m}}$. These values are rather high for composites. One of the reasons is due to the fact that a fiber mat allows cracks to be bridged, and therefore, brittle fracture is prevented from occurring. It was shown that this type of material is notch insensitive (Berthaud et al. 2000), which can be understood by the special architecture of the material. The level of fracture toughness is close to the value of the stress intensity factor estimated above. This constitutes a validation of the identification procedure.

Yet another way of modeling the presence of a crack and its process zone is to resort to cohesive zone models (CZMs). They consist of condensing all the nonlinearities along lines (in 2D simulations) and surfaces (in 3D simulations). However, contrary to standard fracture mechanics that only accounts for crack propagation (i.e., a crack is initially present in the considered structure), a CZM may account for initiation, propagation, and even coalescence when needed. One of the earlier models consists of writing the free energy density ψ of an elementary surface of the interface (Allix and Corigliano 1996) as

$$\psi = \frac{1}{2} k_n (1 - d) [u]^2 \tag{29}$$

where d denotes the interface damage variable, k_n the normal stiffness so that the normal traction t is related to the normal displacement jump $[u]$ by

$$t = \frac{\partial \psi}{\partial [u]} = k_n (1 - d) [u] \tag{30}$$

and the thermodynamic force y associated with the damage variable d reads

$$y = -\frac{\partial \psi}{\partial d} = \frac{1}{2} k_n [u]^2 \tag{31}$$

In the present case, only mode I is considered for the sake of simplicity. It can be generalized to account for the three fracture modes (Allix and Corigliano 1996).

When comparing Eqs. (12) and (13) with Eqs. (29), (30), and (31), the only difference is that the first ones are associated with energies per unit volume and the second ones by energies per unit surface. Consequently, if the free energies satisfy $\psi = \Psi \, \ell$, then the two damage variables are equal $d = D$ to achieve a total equivalence in terms of dissipated energy. It follows that the normal stiffness is such that $\ell \, k_n = E_0$, provided $\ell \, \varepsilon = [u]$, which is a good approximation in the present case.

Under these hypotheses, the parameters of the CZM are the normal stiffness $k_n = 4.3$ kN/mm^3, the characteristic strength $\sigma_c \approx 364$ MPa, and the characteristic crack opening displacement $\delta_c = \ell \, \varepsilon_c \approx 43$ μm (i.e., of the order of 0.6 pixel). In evaluating these parameters, the element size ℓ still appears.

It is of interest to draw an analogy with the lumped damage mechanics case that was discussed in "► Chap. 38, Evaluating Damage with Digital Image Correlation: B. From Physical to Mechanical Damage." The cantilever steel beam was shown to lead to a localized damage case. However, because the physical origin of this damage was a buckling of the beam walls, localization was distributed over a length scaling in proportion to the section. Yet, there is no obstacle to condense the nonlinear damage effect into a zone smaller than the actual one, provided the "damage" law is adjusted to the chosen size. To account for the same rotation, a curvature will have to scale as the inverse of the damaged zone. In the limit of a vanishing size, i.e., the "lumped" limit, the curvature diverges, but in such a way that the rotation undergoes a discontinuity. A quasi-perfect analogy exists with the crack versus localized damage discussion. Understanding what is a conventional choice and how physical quantities are to be scaled to comply with the chosen description shows the power of CDM and the sometimes heated debate around the notion of internal length scales, which used to be very active a few years back.

Conclusions and Perspectives

Among the various damage measurement methods discussed in the introduction (Part A), two of them were used. First, a detection technique was illustrated to analyze the development of physical microdamage (i.e., microcracks and microvoids) by resorting to 2D and 3D images, which are subsequently processed. Second, elasticity coupled with damage was utilized to invert damage fields and identify the parameters of a damage growth law. The choice of either approach is dictated by modeling strategies that mostly rely on the scale of observation and the taste of the scientist.

The identification of the (mechanical) damage law was first illustrated in the simple case of beam geometries (Part B) and then further applied herein to composite materials. The formulation of the damage law as one of the ingredients (together with equilibrium and compatibility) controlling the measured displacement field allows moving progressively to the analysis of a field of stiffness contrast and to robust determination of the damage growth law. Last, the question of the proper handling of a localized regime so that the chosen discretization remains compatible with the energy balance has been discussed both for lumped damage mechanics (for beams) and mesoscopic cracks (for composites). This very same path was followed both in the previous and present chapters.

To perform most of the analyses reported in this chapter, only one measurement technique (i.e., DIC) was used. Further, the identification procedures used the concept of equilibrium gap. It is worth noting that there are other full-field measurement and identification procedures (Grédiac and Hild 2012). The choice was made to link as strongly as possible both steps (i.e., measurement and identification). Global approaches to DIC are one way to achieve this goal and to seamlessly bridge the gap between experiments and numerical simulations.

In terms of damage models, only simple ones were used herein for illustration purposes and also because complexity has to be dealt with care. Inversion and identification belong to the class of inverse problems. Consequently, the more numerous the unknown parameters, the more measured data need to be collected to make the results trustworthy and robust. Further, only 2D displacement fields were used in this chapter. However, the identification procedures are generic and are currently being generalized to 3D surface and volume measurements.

As discussed in the introduction of Part A, various scales of measurement and modeling are possible when describing damage. For damage detection, different scales were considered. Smaller and larger ones can also be taken into consideration. For damage models, they were essentially written at the level of the volume element of continuum mechanics. Other choices could have been made.

All these developments are geared toward the emergence of simulation-based engineering sciences. Among the various challenges (Blue Ribbon Panel 2006), there are open problems associated with multiscale and multi-physics modeling. Continuum damage mechanics is one area of mechanics that needs further development to reach a level of confidence sufficiently high for engineers to use its models to design (damage-tolerant) structures. Real-time integration of simulation methods with measurement systems is another issue to be addressed. To achieve this goal, robust model identification and validation procedures need to be improved and made robust.

Acknowledgments Many results reported in this chapter have been obtained thanks to the help of and discussions with M. Ben Azzouna, S. Calloch, J.-M. Guimard, B. Leboime, J. Réthoré, and N. Swiergiel. Part of the work was supported by Agence Nationale de la Recherche (VULCOMP phases 1 and 2) and grants from Ile de France region (SESAME project entitled "*Plate-forme francilienne d' expérimentations mécaniques de 3^e génération*" and DICCIT project).

This chapter has been reviewed by Profs. A. Dragon and J. Lemaitre. The authors would like to warmly thank them for their patience and help in improving the compuscript. Last, this chapter is dedicated to Prof. F.A. Leckie (1929–2013) with whom the authors discussed many issues presented herein.

References

O. Allix, A. Corigliano, Modeling and simulation of crack propagation in mixed-modes interlaminar fracture specimens. Int. J. Fract. **77**, 111–140 (1996)

S. Andrieux, Y. Bamberger, J.-J. Marigo, Un modèle de matériau microfissuré pour les bétons et les roches. J. Méc. Th. Appl. **5**(3), 471–513 (1986)

S. Avril, M. Bonnet, A.-S. Bretelle, M. Grédiac, F. Hild, P. Ienny, F. Latourte, D. Lemosse, S. Pagano, E. Pagnacco, F. Pierron, Overview of identification methods of mechanical parameters based on full-field measurements. Exp. Mech. **48**(4), 381–402 (2008)

D. Baptiste, Damage micromechanics modelling of discontinuous reinforced composites, in *Continuum Damage Mechanics of Materials and Structures*, ed. by O. Allix, F. Hild (Elsevier, Amsterdam, 2002), pp. 115–163

M.R. Begley, A.G. Evans, R.M. McMeeking, Creep rupture in ceramic matrix composites with creeping fibers. J. Mech. Phys. Solids **43**(5), 727–740 (1995)

M. Ben Azzouna, J.-N. Périé, J.-M. Guimard, F. Hild, S. Roux, On the identification and validation of an anisotropic damage model by using full-field measurements. Int. J. Damage Mech. **20**(8), 1130–1150 (2011)

Y. Berthaud, S. Calloch, F. Collin, F. Hild, Y. Ricotti, Analysis of the degradation mechanisms in composite materials through a correlation technique in white light, in *Proceeding of the IUTAM Symposium on Advanced Optical Methods and Applications in Solid Mechanics,* ed. by A. Lagarde,(Kluwer, Dordrecht, 2000), pp. 627–634

T. Black, T. Belytschko, Elastic crack growth in finite elements with minimal remeshing. Int. J. Num. Meth. Eng. **45**, 601–620 (1999)

Blue Ribbon Panel, Simulation-Based Engineering Sciences. Final report, NFS (2006), www.nsf.gov/pubs/reports/sbes_final_report.pdf

D. Boudon-Cussac, A. Burr, F. Hild, On a continuum description of damage in fiber-reinforced composites, in *Proceedings of the McNU'97, Damage Mechanics in Engineering Materials,* eds. by G.Z. Voyadjis, J.-W.W. Ju, J.-L.Chaboche. Studies in applied mechanics, vol. 46 (Elsevier, Amsterdam, 1998), pp. 303–320

B. Budiansky, J.W. Hutchinson, A.G. Evans, Matrix fracture in fiber-reinforced ceramics. J. Mech. Phys. Solids **34**(2), 167–189 (1986)

A. Burr, F. Hild, F.A. Leckie, Continuum description of damage in ceramic-matrix composites. Eur. J. Mech. A/Solids **16**(1), 53–78 (1997)

A. Burr, F. Hild, F.A. Leckie, The mechanical behaviour under cyclic loading of ceramic-matrix composites. Mater. Sci. Eng. **A250**(2), 256–263 (1998)

A. Burr, F. Hild, F.A. Leckie, Isochronous analysis of ceramic-matrix composites under thermomechanical cyclic loading conditions. Comp. Sci. Tech. **61**(15), 2231–2238 (2001)

J.-L. Chaboche, On the description of damage induced anisotropy and active passive effect, in *Proceedings of the ASME Winter Annual Meeting,* 1990

D. Claire, F. Hild, S. Roux, A finite element formulation to identify damage fields: the equilibrium gap method. Int. J. Num. Meth. Eng. **61**(2), 189–208 (2004)

D. Claire, F. Hild, S. Roux, Identification of a damage law by using full-field displacement measurements. Int. J. Damage Mech. **16**(2), 179–197 (2007)

W.A. Curtin, Exact theory of fiber fragmentation in single-filament composite. J. Mater. Sci. **26**, 5239–5253 (1991)

R. Desmorat, Quasi-unilateral conditions in anisotropic elasticity. C. R. Acad. Sci. Paris IIB **328** (6), 445–450 (2000)

Z.Z. Du, A.C.F. Cocks, R.M. McMeeking, Power-law matrix creep in fiber composites due to transverse stress gradient. Eur. J. Mech. A/Solids **16**(3), 445 (1997)

P. Forquin, F. Hild, A probabilistic damage model of the dynamic fragmentation process in brittle materials. Adv. Appl. Mech. **44**, 1–72 (2010)

P. Germain, Q.S. Nguyen, P. Suquet, Continuum thermodynamics. ASME J. Appl. Mech. **50**, 1010–1020 (1983)

M. Grédiac, F. Hild (eds.), *Full-Field Measurements and Identification in Solid Mechanics* (ISTE/Wiley, London, 2012)

N. Guy, D.M. Seyedi, F. Hild, A probabilistic nonlocal model for crack initiation and propagation in heterogeneous brittle materials. Int. J. Num. Meth. Eng. **90**(8), 1053–1072 (2012)

D. Halm, A. Dragon, Y. Charles, A modular damage model for quasi-brittle solids – interaction between initial and induced anisotropy. Arch. Appl. Mech. **72**, 498–510 (2002)

R. Hamam, F. Hild, S. Roux, Stress intensity factor gauging by digital image correlation: application in cyclic fatigue. Strain **43**, 181–192 (2007)

F. Hild, Discrete versus continuum damage mechanics: a probabilistic perspective, in *Continuum Damage Mechanics of Materials and Structures*, ed. by O. Allix, F. Hild (Elsevier, Amsterdam, 2002), pp. 79–114

F. Hild, A. Burr, F.A. Leckie, Matrix cracking and debonding in ceramic-matrix composites. Int. J. Solids Struct. **33**(8), 1209–1220 (1996)

F. Hild, S. Roux, Digital image correlation: from measurement to identification of elastic properties – a review. Strain **42**, 69–80 (2006)
F. Hild, S. Roux, N. Guerrero, M.E. Marante, J. Florez-Lopez, Calibration of constitutive models of steel beams subject to local buckling by using digital image correlation. Eur. J. Mech. A/Solids **30**, 1–10 (2011)
P. Ladevèze, Sur une théorie de l'endommagement anisotrope. Internal report no. 34, LMT Cachan, 1983
P. Ladevèze, A damage computational method for composite structures. Comput. Struct. **44**(1–2), 79–87 (1992)
P. Ladevèze, E. Le Dantec, Damage modelling of the elementary ply for laminated composites. Comp. Sci. Tech. **43**(3), 257–267 (1992)
H. Leclerc, J.-N. Périé, S. Roux, F. Hild, Integrated digital image correlation for the identification of mechanical properties, in *Mirage*, ed. by A. Gagalowicz, W. Philips. LNCS, vol. 5496 (Springer, Berlin, 2009), pp. 161–171
J. Lemaitre, J. Dufailly, Damage measurements. Eng. Fract. Mech. **28**(5–6), 643–661 (1987)
N. Limodin, J. Réthoré, J.-Y. Buffière, A. Gravouil, F. Hild, S. Roux, Crack closure and stress intensity factor measurements in nodular graphite cast iron using 3D correlation of laboratory X ray microtomography images. Acta Mat. **57**(14), 4090–4101 (2009)
N. Malésys, L. Vincent, F. Hild, A probabilistic model to predict the formation and propagation of crack networks in thermal fatigue. Int. J. Fat. **31**(3), 565–574 (2009)
J.-J. Marigo, Formulation d'une loi d'endommagement d'un matériau élastique. C. R. Acad. Sci. Paris (série II) **292**, 1309–1312 (1981)
N. Moës, J. Dolbow, T. Belytschko, A finite element method for crack growth without remeshing. Int. J. Num. Meth. Eng. **46**(1), 133–150 (1999)
A.C. Orifici, I. Herszberg, R.S. Thomson, Review of methodologies for composite material modelling incorporating failure. Comp. Struct. **86**(1–3), 194–210 (2008)
J.N. Périé, H. Leclerc, S. Roux, F. Hild, Digital image correlation and biaxial test on composite material for anisotropic damage law identification. Int. J. Solids Struct. **46**, 2388–2396 (2009)
J. Réthoré, A fully integrated noise robust strategy for the identification of constitutive laws from digital images. Int. J. Num. Meth. Eng. **84**(6), 631–660 (2010)
J. Réthoré, F. Hild, S. Roux, Extended digital image correlation with crack shape optimization. Int. J. Num. Meth. Eng. **73**(2), 248–272 (2008)
J. Réthoré, S. Roux, F. Hild, An extended and integrated digital image correlation technique applied to the analysis fractured samples. Eur. J. Comput. Mech. **18**, 285–306 (2009)
D. Rouby, P. Reynaud, Fatigue behaviour related to interface modification during load cycling in ceramic-matrix fibre composites. Comp. Sci. Tech. **48**, 109–118 (1993)
S. Roux, F. Hild, Stress intensity factor measurements from digital image correlation: post-processing and integrated approaches. Int. J. Fract. **140**(1–4), 141–157 (2006)
S. Roux, F. Hild, Digital image mechanical identification (DIMI). Exp. Mech. **48**(4), 495–508 (2008)
S. Roux, F. Hild, H. Leclerc, Mechanical assistance to DIC, in *Proceedings of the Full Field Measurements and Identification in Solid Mechanics,* eds. by H. Espinosa, F. Hild. IUTAM Procedia, vol. 4 (Elesevier, 2012), pp. 159–168
A.N. Tikhonov, V.Y. Arsenin, *Solutions of Ill-Posed Problems* (Wiley, New York, 1977)
G. Vivier, H. Trumel, F. Hild, On the stored and dissipated energies in heterogeneous rate-independent systems: theory and simple examples. Continuum Mech. Thermodyn. **20**(7), 411–427 (2009)
G. Vivier, H. Trumel, F. Hild, On the stored and dissipated energies in heterogeneous rate-independent materials. Application to a quasi-brittle energetic material under tensile loading. Continuum Mech. Thermodyn **23**, 387–407 (2011)

Section X

Micromechanics of Damage for Laminated Composites

Experimental Methods to Quantify Microdamage and Microstructure Anomalies in Fiber-Reinforced Polymer Composites: Overview

40

Valeria La Saponara and Rani Elhajjar

Contents

Abstract

Composite structures are known to be susceptible to both manufacturing defects and in-service damage. Defects or damage can result in serviceability issues or a loss in the structural capability. Detection and characterization of defect and damage is thus of paramount importance in any successful deployment of fiber-reinforced polymer composites, particularly as they are used as primary load-carrying structures. Experimental methods for investigation of

V. La Saponara (✉)
Department of Mechanical and Aerospace Engineering, University of California, Davis, CA, USA
e-mail: vlasaponara@ucdavis.edu

R. Elhajjar
Department of Civil and Environmental Engineering, University of Wisconsin, Milwaukee, WI, USA
e-mail: elhajjar@uwm.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_15

microdamage or anomalies in composites present the challenge that one single method is not capable of identifying all damage mechanisms. This chapter presents a review of selected experimental methods aimed at providing a quantitative description of selected damage types of microscale dimensions (e.g., intra-ply cracks) and microstructure anomalies (waviness, porosity) in fiber-reinforced polymer composites. The chapter discusses in the first part microscale damage characterization using microscopy, radiography acoustics, and ultrasonic techniques, while the second part is focused on the characterization of microstructural anomalies that depend on manufacturing, namely, waviness and porosity.

Introduction

Fiber-reinforced polymer composites have been introduced since the 1970s for aerospace applications requiring high stiffness/weight and high strength/weight ratios. Since then, they have been adopted also for civil, marine, transportation, and wind energy applications. Examples include the new aircraft Boeing 787, Airbus A350, and F-35, with approximately 50 % of their structural weight made of composites; new vehicular and pedestrian bridges (starting with Miyun bridge in China back in 1982, with the first all-composite deck) and retrofitted bridges (seismic retrofit of the State Street Bridge on Interstate Highway 80 in Salt Lake City, UT, completed in 2001); the US M80 Stiletto launched in 2006 and the Swedish Visby Corvette commissioned in 2002; and the 37 m long composite blades for General Electric 1.5 MW wind turbines launched in 2004.

However, monolithic and sandwich fiber-reinforced polymer composites are plagued by complex damage types, such as microscale cracks, fiber/matrix interface debonding, delamination, barely visible impact damage, shear core failure, and micro-buckling. These damage modes may be triggered by manufacturing defects, compounded with damage incurred during service. Service damage may be caused by thermomechanical fatigue, impact, exposure to environmental factors (moisture, UV), and/or service fluids, overloads, and a nonlinear combination of all of the above. It has been estimated that the cost of inspection of a composite structure is approximately 30 % of the cost of its acquisition and operation (Bar-Cohen 2000). Therefore, structural safety relies on the ability to assess manufacturing defects and damage onset early enough for repairs while minimizing downtime for cost reasons.

This chapter presents a review of selected experimental methods aimed at providing a quantitative description of selected damage types of microscale dimensions (e.g., intra-ply cracks) and microstructure anomalies (waviness, porosity) in fiber-reinforced polymer composites. Part 1 focuses on microscale damage characterization using microscopy, radiography acoustics, and ultrasonic techniques. Part 2 addresses the characterization of manufacturing-dependent microstructural anomalies, waviness, and porosity.

Part 1: Microscale Damage

Overview

Many authors have investigated how damage evolves in fiber-reinforced polymer matrix composites and the respective roles of its constituents, matrix, and fibers on ultimate failure, since the 1970s (e.g., Aveston and Kelly 1973). In cross-ply composites manufactured with unidirectional 0° and 90° plies, damage caused by quasi-static and fatigue tensile loading follows a sequence that has been extensively studied in the past 40 years (see recent reviews by Berthelot 2003; McCartney et al. 2010; Talreja and Singh 2012): first, multiple matrix cracks appear in 90° plies (transverse plies with respect to the direction of loading), initiating from local defects (voids, fiber clusters, or resin-rich areas), with spacing that depends on the layer thickness. Matrix cracking may cause load redistribution in the composite, moisture ingress, and reduction of the stiffness of the cracked plies and thus loss of composite stiffness, strength, and toughness. With further load increase, these transverse cracks begin to interact with each other: the crack spacing decreases and eventually reaches saturation, a stage named the "characteristic damage state" (e.g., Highsmith and Reifsnider 1982), which depends on ply properties, ply thickness, and stacking sequence. A more complex damage progression has been documented for quasi-isotropic composites (e.g., Tong et al. 1997, Fig. 1; Adolfsson and Gudmundson 1999, Fig. 2), for example, with cracks at −45° orientation generally starting from 90° cracks.

The three-stage damage sequence discussed above has also been detected in cross-ply composites built by woven plies (e.g., Naik 2003). Woven composites are made of three major components: longitudinal strand (warp ply), transverse strand (fill or weft ply), and resin-only regions. In these composites, additional damage features are encountered: (a) the first stage of damage is associated with the formation of microscale damage and transverse cracks in the fill ply, and (b) in

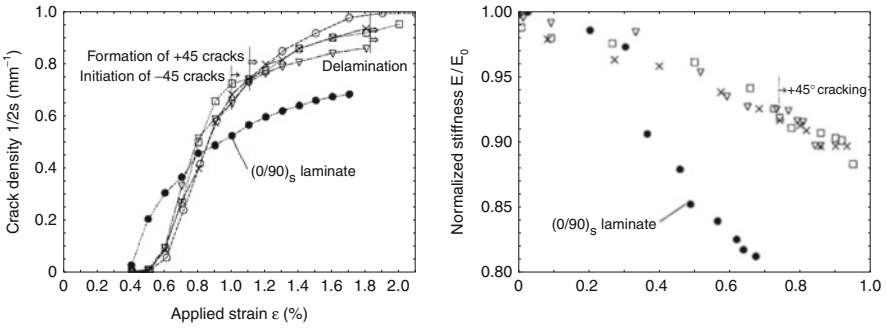


Fig. 1 Progressive transverse cracking for quasi-isotropic and cross-ply glass/epoxy laminates: (*left*) crack density versus applied strain; (*right*) stiffness loss versus crack density (Tong et al. 1997)

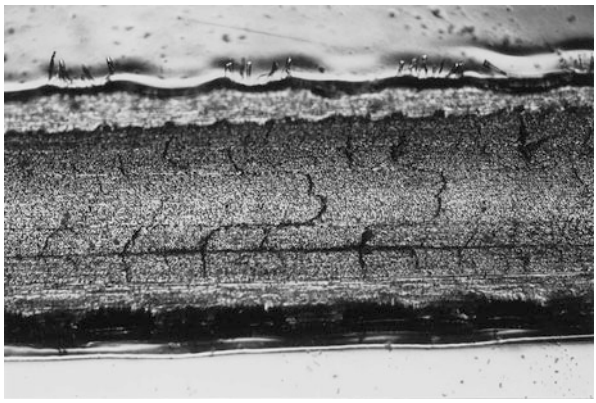


Fig. 2 Damage in glass/epoxy laminate with layup $[0/90/\pm45]_s$. Ply thickness ~0.125 mm (Adolfsson and Gudmundson 1999)

the middle stage of damage, macroscale damage such as shear failure in the warp plies appears. Cracks in the neat resin regions and delamination between warp and fill plies may also be found. Figure 3 shows damage in cross-ply glass/epoxy woven specimens under tensile fatigue, which demonstrates the increased complexity of failure modes in weaves. An example of fracture saturation may also be found in heterogeneous structures of much larger scales, such as rocks (Fig. 4; Schöpfer et al. 2011).

The lasting interest of the scientific community and the coexistence of numerous theories of damage initiation and progression in advanced composites and heterogeneous structures are shown by the very recent publication dates of papers and books on this topic (to name a few, Anderson et al. 2008; Cid Alfaro et al. 2010; Böhm and Hufenbach 2010; Schöpfer et al. 2011; Talreja and Singh 2012; Aggelis et al. 2012; Farge et al. 2012; Barbero et al. 2013).

Experimental validation should require in principle the use of robust techniques that minimize user bias and should also address the inherently stochastic nature of microscale damage features such as intra-ply cracks. This is particularly evident in woven composite plies, which are compacted differently during the manufacturing process. Thus, stress varies locally, intra-ply cracks are triggered at different times, and their spacing varies. These aspects of randomness of damage onset and stochastic distribution of damage were recognized early on by Manders et al. (1983) and are also discussed by Yurgartis et al. (1992), Bulsara et al. (1999), Wang and Yan (2005), Silberschmidt (2005), Anderson et al. (2008).

Part I of this chapter focuses on selected experimental methods for laboratory-scale testing to quantify microscale damage features. The practicality of the methods for field deployment is not addressed. Part I does not aim at advocating for one analytical/numerical approach or philosophy over another. The authors of this chapter have done their best effort to acknowledge the many authors who worked in this very populated but still exciting field of research.

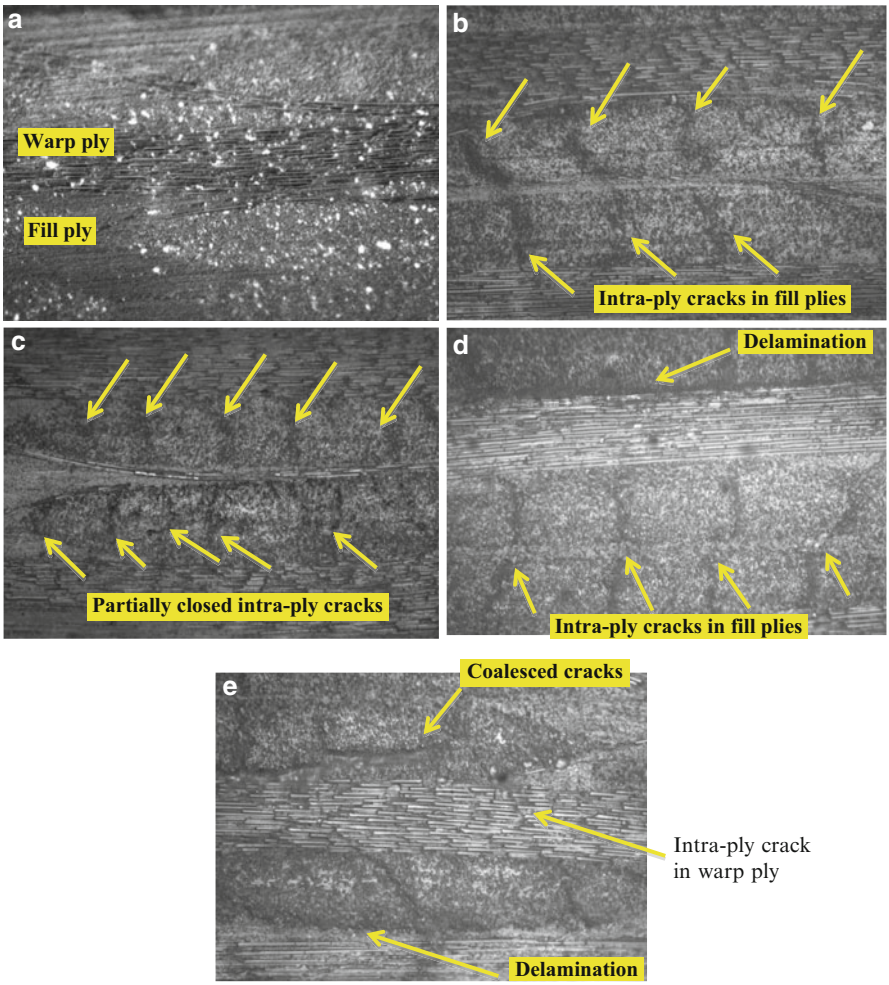


Fig. 3 Pictures of damage at several locations in woven glass/epoxy cross-ply specimen subject to tensile fatigue with stress ratio = 0.8 and frequency = 10 Hz. (**a**) pristine; (**b**) after 1,000 cycles, at the mean load of 9.12 kN; (**c**) unloaded, after 1,000 cycles; (**d**) after 14,000 cycles, at mean load; and (**e**) after 16,000 cycles, at mean load. Fill ply thickness ~200 μm. The experiments are described in La Saponara et al. 2011, and the pictures show specimen D, whose fatigue test were stopped after 18,000 cycles

Methods Based on Microscopy

Many researchers who published experimental data appear to have used optical microscopy and typically transmitted light microscopy applied to loaded and unloaded specimens, which were sufficiently transparent (e.g., glass/epoxy specimens). Some researchers directly examine samples or cross sections of samples (e.g., Manders et al. 1983; Yurgartis et al. 1992; Tong et al. 1997; Voyiadjis and

Fig. 4 Fracture saturation in rocks, with cracks developed within limestone beds located between mudrocks. The *arrow* indicates a 0.29 m long rock hammer, for scale (Schöpfer et al. 2011)

Almasri 2007, with scanning electron microscopy; Anderson et al. 2008; Thomas et al. 2008; Ogi et al. 2010; París et al. 2010). Other researchers use microscopy to view edge replications ("replicas") of the edges of polymer- and ceramic-matrix composites (e.g., Highsmith and Reifsnider 1982; Stinchcomb 1986; Yurgartis et al. 1992; Sørensen and Talreja 1993; Adolfsson and Gudmundson 1999, etc.). This method is used on the exposed specimen edges that have been cleaned and sanded prior to the test. Typically, the specimen is held in a testing machine. A strip of cellulose acetate film is softened by acetone just before being applied onto the specimens' edges, and it is then held in place by light pressure (e.g., by a thumb) for a short time interval (a couple of minutes). An impression of the surface is produced and may be viewed later by an optical microscope. Some challenges with this technique are reported below that are not explicitly discussed in the literature:

1. The film acquires curvature upon curing. The light of a microscope will thus scatter unevenly if the film is not held perfectly flat between microscope slides, thus complicating the matter of recognizing and possibly counting microscale damage features (see, e.g., shiny spots in Fig. 3a).
2. There is no access to the interior part of the specimen.
3. It is critical to have a statistically robust representation of the damage process, which is also not dependent on user bias. This may require tedious crack counting work, unless image processing software and stereology techniques are adopted. What is classified as an intra-ply crack should be clear-cut to different users (e.g., whether the crack needs to cross the entire width of a ply or not). Item 4 may hold in the case of use of image processing software.
4. There may be artifacts in the replica such as finger prints and drops of acetone "immortalized" by the replica, or film sections that were not cured by acetone and appear empty (Fig. 5). Should the image be analyzed by image processing software not supervised by a knowledgeable user, the software might log such artifacts as red-flag items related to damage.
5. Cellulose acetate films are viscoelastic materials. One manufacturer (TedPella) recommended storage at a low and constant temperature to minimize creep. Crack density measurements may be affected if the films are not stored properly and the replicas are not processed timely.

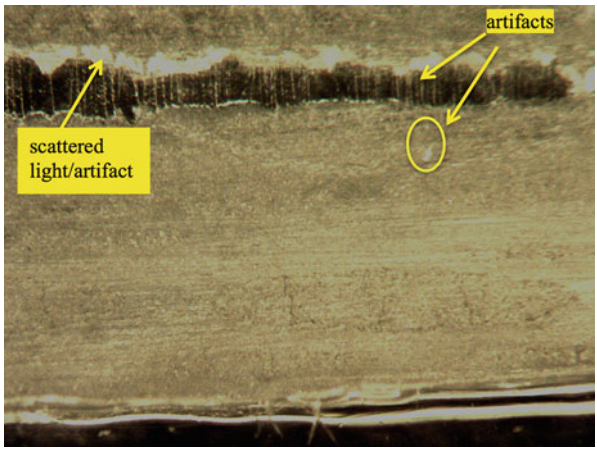


Fig. 5 Replica of woven cross-ply glass/epoxy specimen. The fill ply is ~200 μm thick

6. Although this technique is categorized as nondestructive, the systematic use of acetone on the surface of a polymeric specimen may actually contribute to damage. Exposure to acetone may also deteriorate the grips of the testing machine.

Yurgartis et al. (1992) devised an interesting image processing technique to classify intra-ply cracks (called "transverse bundle cracks") and delaminations. They consider ceramic-matrix composites, but the technique may be applied to other heterogeneous materials as well. The semiautomatic image processing involves the increase of the resolution of the cracks with respect to the background, producing a simplified image that can be used for a statistically sound measurement of damage (Fig. 6).

Wharmby et al. (2003) mention the use of software that was developed in-house, which processes images from a video camera and a digital frame grabber, but details of the software and of the decision-making process on what is labeled as a crack are not readily available (they are reported in a conference proceeding paper). By comparison, the work by Yurgartis et al. (1992) excels for its clarity and documentation.

Methods Based on Radiography

X-ray radiography is a technique that allows detection and quantification of micro- and macroscale damage features that are inside a specimen. It requires the use of a dye penetrant, for example, a di-iodobutane solution (Crossman et al. 1980), or zinc iodide in fiber-reinforced epoxy specimens (Stinchcomb 1986; Lafarie-Frenot and Hénaff-Gardin 1991). Stinchcomb (1986) lists two important conditions for successful radiography: (1) the penetrant solution should be able to flow easily inside the composite and thus should contain an additive to reduce the surface tension. The

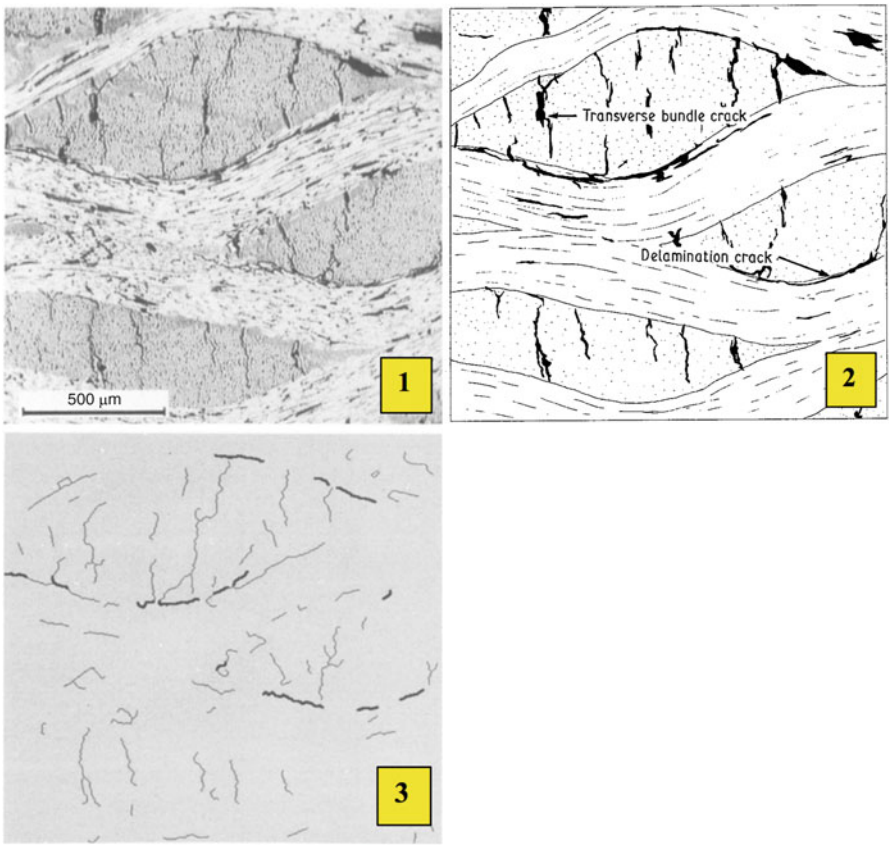


Fig. 6 Processing of microscope image of carbon-carbon specimens (Yurgartis et al. 1992). The original digital image (box *1*) is denoised, smoothed, and subjected to the so-called medial axis transformation, resulting in a simplified image highlighting the damage features of interest (boxes *2–3*). The discrimination of transverse bundle cracks from delamination cracks is carried out by an operator, who labels the cracks. The crack counting can then be automatic. The data was processed from 18 to 20 randomly selected and not overlapping images (Figures reproduced with kind permission from Springer Science+Business Media B.V)

penetrant may also not access some areas of the specimen, if those are not connected to an external surface where the penetrant is applied; (2) for damage such as large delaminations, capillary action will drive the penetrant to the delamination boundary, which will then appear in the radiograph, and could be ambiguous and misinterpreted. Moreover, radiography provides two-dimensional images with no depth information, and it may be necessary to take several radiographs of the specimen at different orientations with respect to the X-ray source (Talreja and Singh 2012).

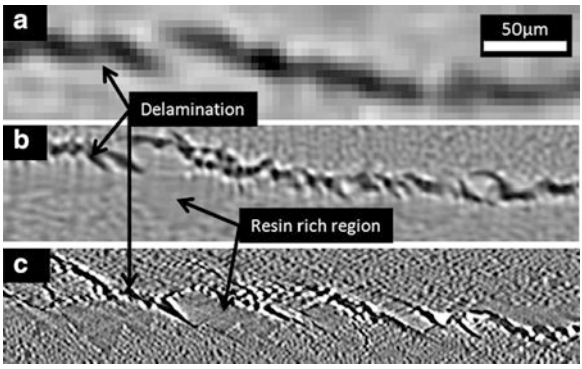


Fig. 7 Close-up of damaged regions using μCT, SRCT, and SRCL (Bull et al. 2013). (**a**) and (**b**) are of the same specimen at approximately the same location, while (**c**) is representative of similar damage on a separate specimen

Three-dimensional views of internal damage may be provided by X-ray computed tomography (CT) (Scott et al. 2011; Zhang et al. 2013; review by Withers and Preuss 2012) or its variations (laminography, see Moffat et al. 2010; synchrotron radiation tomography, see Wright et al. 2010). In CT, a series of high-spatial-resolution digital radiographs (projections) is collected as the sample is rotated by small increments between each image. A three-dimensional image is built mathematically after filtering/denoising the projections (usually with a Radon transform), followed by forwarding these filtered projections onto a grid and adding the contributions of the radiographs. The outcome is a three-dimensional representation of the object of interest.

Bull et al. (2013) reviewed the performance of micro X-ray CT (μCT), synchrotron radiation CT (SRCT), and synchrotron radiation computed laminography (SRCL) for the case of carbon/epoxy specimens subjected to impact (Fig. 7). The first technique allows identifying damage at the macro- and mesoscale levels, while the other two techniques are found to be superior for rapid scanning of three-dimensional microscale damage. The application of X-ray CT for identification of porosity in the microstructure will be further discussed in Part 2.

Methods Based on Acoustics/Ultrasonics

The resolution, sensitivity, and thus success of these nondestructive evaluation techniques depend on the ability to interpret the resulting signal and discriminate between as-manufactured defects (voids, distributed voids or porosity, inclusions, resin-rich and resin-poor areas, waviness) and damage caused by thermomechanical loading or other factors. Ultrasonic transducers generate wave forms in a frequency range between 20 kHz and few GHz, whose interaction with damage depends on several parameters: respective sizes of the wavelength and the defect/damage size, dispersion properties of the material, anisotropic nature of the material, interrogation distance, quality of the coupling between the transducer and the material, boundary and loading conditions (including those around the transducers,

especially if they are embedded in the material), and equipment-related nonlinearities deforming the transmitted waveform (e.g., amplifier above a given frequency range). In order for the composite structure to be "seen" by the transducers as a homogeneous medium, an appropriate operating frequency for the transducers needs to be selected: a frequency lower than 1 MHz satisfies the homogeneity condition for most composite materials (Castaings et al. 2000).

The transducers may transmit concurrent Lamb waves, whose behavior in a defect-free composite may be predicted by theoretical dispersion curves computed with the effective properties of a representative volume element of the material. Hence, theoretical dispersion curves are unable to capture the local variability of the material or the existence of manufacturing defects. Transducers embedded inside composite specimens may act as stress raisers and initiate damage, since they are typically of a size comparable with the ply thickness (Singh and Vizzini 1994; Schaaf et al. 2007; Tang et al. 2011), although some authors found this not to be always the case (Mall 2002; Qing et al. 2007).Tang et al. (2011) hypothesized a correlation between ultrasonic wave forms received by embedded transducers in the low ultrasonic range, processed with a Gabor wavelet transform, and the accumulation of intra-ply cracks in the host woven glass/epoxy specimens, when the specimens are subjected to tensile fatigue. This hypothesis is being currently tested through the very time-consuming and tedious process of analyzing over 100 replica films from loading/unloading tests (La Saponara et al. 2011; ongoing unpublished research). The transducers may also be tuned with a wedge or a rotating device to transmit a selected type of waves (in composites, quasi-longitudinal or fast quasi-transverse or slow quasi-transverse waves).

Wave propagation in composites and their ability to detect micro- and macroscale damage have been studied by numerous authors, this being a very active field of research. In most cases, the composites are analyzed as elastic materials. Few authors are cited here, who studied elastic composites and ultrasonics with contact transducers: Nayfeh (1991), Rose et al. (1993), Kessler et al. (2002), Seale et al. (1998), review by Su et al. (2006), review by Diamanti and Soutis (2010), and Aggelis et al. (2012). Few authors study wave propagation of bulk waves or Lamb waves in viscoelastic composite materials: Castaings et al. (2000) and Schubert and Herrmann (2011, 2012).

Another ultrasonic method uses C-scans, which may detect internal delaminations that are perpendicular to the transducer or may be used for assessments of manufacturing quality (e.g., Voyiadjis and Almasri 2007). In this ultrasonic method, one single transducer is used for the pulse-echo method, whereas two transducers (pulser and receiver) are needed for through-transmission. Accessibility issues sometimes require the use of one transducer in the pulse-echo form. Specimen thickness can also affect the wave propagation, which may necessitate the adoption of lower frequency sensors to investigate the composite. C-scans may not be able to detect microscale cracks, and their resolution is considered worse than that of radiography (Stinchcomb 1986). The method is also not capable of resolving many types of manufacturing anomalies. However, its benefit compared to

radiography is that it has easier deployment in manufacturing or in-service applications for detection of delaminations and porosity. More discussion about this method follows in Part 2.

Another approach based on wave propagation in solids is acoustic emission. Through acoustic emission, stress waves in the sound frequency range (20 Hz–20 kHz) are detected with passive transducers attached to the specimens, when there is a rapid redistribution of stress caused by damage such as fiber breaks or microcracks. Resulting stress wave bursts or acoustic emission (AE) events are captured by these passive transducers. The reliability of the acoustic emission method depends on the compatibility of the hardware used, the sensor and the preamplifier (Jemielniak 2001). The amplifiers typical of acoustic emission setups can have a high pass, low pass, or a band pass-type filters. They amplify and filter the acoustic emission signals coming directly from the sensors or from a preamplifier before they can be processed. Acoustic emission in composites has been investigated by Favre and Laizet (1989), Steiner et al. (1995), Tsamtsakis et al. (1998), and more recently by Böhm and Hufenbach (2010) and Aggelis et al. (2012). Early studies by Roy and Elghorba (1988) investigated the interlaminar failure in Mode II and used acoustic emission to investigate damage development of glass fiber/epoxy composite during monotonic and cyclic Mode II loadings. As mentioned above, drawbacks of the acoustic emission method include the difficulty in identifying damage location and the ability to discriminate between damage types. Qamhia et al. (2013) used acoustic emission to study the detection of porosity and waviness in carbon/epoxy composites. They manufactured specimens containing the defects separately and compounded. The porosity defects were found to produce higher counts and energy releases than the emissions from the waviness specimens. Frequency domain analysis shows the possibility to understand the time history of the AE signals and to differentiate the defect type. The coupling of the two defects under examination (porosity and waviness) was found to increase the difficulty in differentiating one source from the other. A detailed discussion of these defects continues in the second part of this chapter on manufacturing anomalies. To overcome some of the method limitations, some authors have combined acoustic emission with other techniques, e.g., Aggelis et al. (2012), who used ultrasonics in conjunction to acoustic emission to document microscale damage progression. Böhm and Hufenbach (2010) used a similar testing philosophy, adding video data and micrographs (Fig. 8). Part 2 of this chapter presents the use of ultrasonics (in particular, C-scans) for detection and quantification of microstructural anomalies. Source location of the events attributed to AE wave evolution can also be performed. The conventional methods rely on the presence of more than one sensor and are based on measuring the signal arrival time. This measurement of the arrival time can be accomplished either by the zonal method which uses the first sensor that detects the signals to specify the primary zone or by the triangulation method which measures the difference in arrival time of the signal to an array of sensors (Promboon 2000). Other methods based on waveforms or cross correlation techniques have also been reported (Ziola and Gorman 1991).

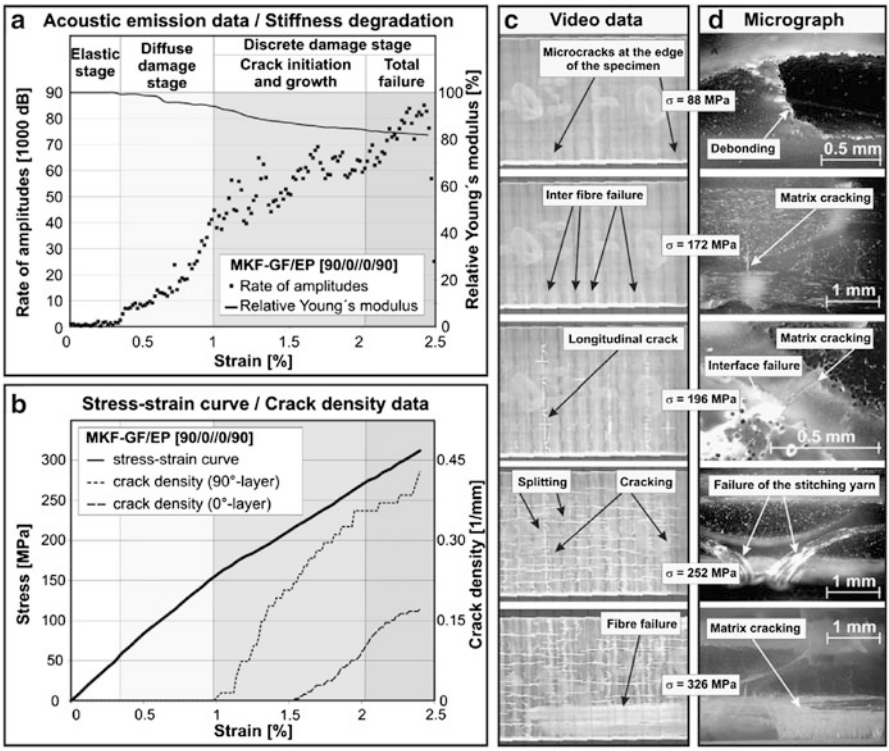


Fig. 8 Damage detected through acoustic emission, ultrasonics, video data, and micrographs (Böhm and Hufenbach 2010)

Part 2: Microstructure Anomalies

Fiber Waviness

The manufacturing of composite structures can inevitably lead to the introduction of defects or manufacturing anomalies into the product or structure being manufactured. The effect of the fiber waviness variability on the tension and compression properties has been previously discussed (Hsiao and Daniel 1996a, b). Other terms sometimes used to refer to fiber waviness include "tiger stripes, marcels, wrinkles, and zebra lines." The fiber waviness can significantly affect the tensile and compressive properties and nonlinear behavior (Chun et al. 2001; Hsiao and Daniel 1996c). Elhajjar and Lo Ricco (2012) have investigated the interaction of fiber waviness with stress concentrations like notches showing even higher reductions. There is no general consensus as to what level of fiber waviness can be considered a normal part of the build process. Figure 9 shows the fiber waviness in a carbon-fiber/epoxy bicycle fork near the head tube region (Sisneros et al. 2012).

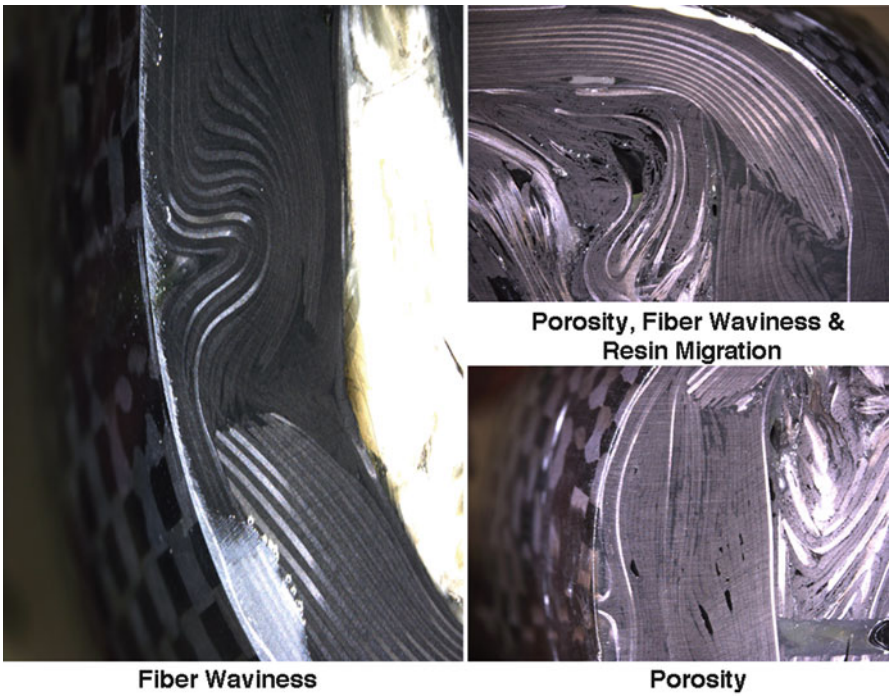


Fig. 9 A cross section from a bicycle fork containing fiber waviness and porosity defects (Sisneros et al. 2012)

In this study, the authors investigated "defective" bicycle forks using acoustic emission and fatigue loading. The acoustic method was capable of identifying the onset of failure. Large curvature changes or certain manufacturing methods result in large waviness zones such as those observed in the carbon-fiber/epoxy bicycle forks. Sometimes, the process of collating the plies, if not done with care, can cause the development of waviness. The move toward larger structures and larger scale layups increases the chances that such defects may occur. According to one article in the Seattle Times (Lovering 2009), during the ramp-up of production, Boeing (The Boeing Company; Chicago, Illinois, USA) has discovered unexpected fiber waviness in several of the one-piece fuselage sections of the 787 Dreamliner. The exact causes of fiber waviness may not always be evident, but research suggests that some of the causes include pressure from layers in a laminated structure, foreign objects, ply terminations, part curvature, co-curing, or residual stresses. One study has found larger incidence of this defect correlated to larger amounts of material being used (Lee and Soutis 2007). Several approaches have been proposed for the characterization and measurement of fiber waviness. Micrographic assessment methods are capable of identifying the waviness, but a major drawback is the destructive nature of this approach. The ability of predicting the location and severity of the waviness before the composite part is cured is theoretically of great interest, but limited success

has been achieved to date. Among the studies that showed some positive results, Wisnom and Atkinson (2000) artificially induced waviness in composite rings and were able to correlate the post-cure waviness levels using a displacement-based technique applied during the manufacturing stage. Most of the detection methodologies are focused on trying to characterize the waviness after the composite structure has cured. In certain waviness profiles, the presence of a resin pocket near the free surface can be used to indicate the presence of a more serious defect below.

Microstructural Analysis

Microstructure characterization methods emphasize the use of destructive cross-sectional scans extracted from the structure under investigation. In this approach, specimens are extracted from the structure, polished, and scanned using an optical microscope. Medium-quality results are also observed from using high-resolution desktop scanners. The general approach of using destructive examinations is limited in two ways: the first is the limited confidence that a given cross section relates to future parts produced; the second is the challenge in properly characterizing the waviness, this usually requires numerous specimens and cross sections per test. Despite these limitations, at this time this is the most accurate way to determine the morphology of fiber waviness defects. Improvements on this method through image analysis on polished cross section or X-ray scans have also been reported (Sutcliffe et al. 2012).

In the characterization of the fiber waviness, one can sometimes assume that the fiber waviness follows a sinusoidal function where the amplitude, angle, or length of the waviness zone can be related to a degradation of a particular mechanical property. The aspect ratio of the defect is sometimes used for its ease of measurement. The weakness of the aspect ratio approach is that it does not account for spatial variation, the thickness of the laminate, and the distribution of the waviness within the thickness. This results in inhibiting its connection to an analytical approach for mechanical properties and failure prediction. The challenge of using one sinusoidal wave or one aspect ratio for characterization of the fiber waviness can be seen in the fiber waviness morphologies shown in Fig. 10 (Chun et al. 2001). In the uniform waviness characterization (Fig. 10a), the aspect ratio of every ply is the same, whereas in the graded waviness (Fig. 10b), a function is used to reduce the amplitude from one face of the laminate to the other. The localized classification (Fig. 10c) refers to fiber waviness of uniform amplitude affecting only a limited number of plies. Whether the waviness follows a uniform, graded, or localized morphology will have a large impact on the type and degree of mechanical property degradation that is observed.

A Gaussian function can also be used to capture the bell curve response of the wavy plies (Elhajjar and Petersen 2011). The approach departs from the method by Hsiao and Daniel (1996a, c), who used a sinusoidal wave to characterize various waviness geometries. This approach can be used to analytically determine the elastic stiffness. The Gaussian function was found to better describe some forms of waviness geometry with respect to a sinusoidal wave. Since these analytic

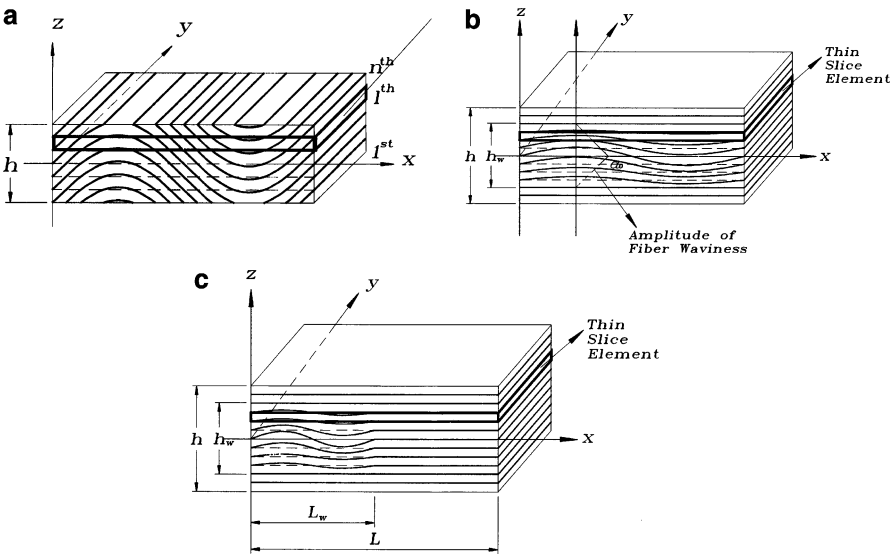


Fig. 10 Characterization of fiber waviness as (**a**) uniform, (**b**) graded, and (**c**) localized (Chun et al. 2001)

functions are differentiable, they can be used to determine the angle of the waviness and hence can be used in the transformation of elastic constants. The unit cell considered can be integrated over a length, L, and the contributions are then added through the thickness for each ply. In this approach, the waviness of each ply in a degrading waviness morphology can be expressed using the height, A_w, degraded from the maximum value at one surface to a diminished value at the other surface. This is accomplished by reducing the waviness height relative to the location of the ply within the stacking sequence. Thus, the z-coordinate along the x-axis, $v(x)$, of each ply can be represented as (Elhajjar and Petersen 2011)

$$v(x) = A_w \left[1 - \left(\frac{z_k - A_w}{h - A_w} \right)^2 \right] e^{-\frac{x^2}{2c^2}} \tag{1}$$

where h is the nominal thickness of the laminate and z_k is the z-coordinate of the upper surface of ply k. Note that the more severe the case of waviness, the more of the gage thickness is affected by the fiber waviness. One feature of this function is that it is independent of the length of the waviness, which is subjective in its determination.

Ultrasonic Methods

Active research in ultrasonics as a method to detect fiber waviness is underway, due to the popularity of this method in the engineering practice, as also presented in Part 1 of this chapter. Ultrasonic methods can be used to nondestructively provide

an indication of some aspects of the waviness morphology without destructive evaluation. However, full resolution of the waviness morphology is more challenging and is dependent on the material types and thicknesses inspected. For example, waviness may not always be uniform, as it maybe localized to a few layers of the laminate, or extend over large areas as described above. It can also be coupled with porosity, complicating the reflection or transmission characteristics of the ultrasonic wave. Dayal (1995) proposed using ultrasound to measure the through-the-thickness waviness and showed the periodicity of the waviness by correlation to C-scan measurements. Theoretical analysis was carried out for the longitudinal wave propagating through the material, with the reflection coefficients calculated at various locations. Wooh and Daniel (1995) proposed using a discrete ray-tracing model based on elastodynamic ray theory. The varying anisotropy of the material causes the ultrasonic waves to travel along curved paths. The theoretical model was found to match the period of the fiber waviness. Scattering due to the fibers and porosity in the matrix makes this approach challenging to implement. Chakrapani et al. (2012) used air-coupled ultrasonic transducers to determine the aspect ratios at different depths in composite specimens. The Rayleigh waves were generated using a 200 kHz frequency and were correlated to the aspect ratio of the waviness. The waviness with lower aspect ratios (producing larger degradation) are more clearly shown in the C-scans obtained using this approach (Figure 11).

Porosity

Porosity or distributed voids are a common manufacturing defect that can have detrimental effects on the static and fatigue performance of composites. It also can allow for moisture ingression and acceleration of environmental deterioration. The causes of porosity are usually related to entrapment of air or moisture between plies, unequal consolidation, or uneven cure pressure (Gauvin et al. 1987). Lower quality processes using low consolidation usually result in a larger porosity content (Liu et al. 2006). Sometimes the stacking sequence may also influence the porosity formation. Toscano and Vitiello (2011) showed how porosity levels were found to be greater when plies are stacked perpendicular to one another, or in a [45/−45] sequence, and were the lowest when stacked in the unidirectional orientation. Muller de Almeida and Nogueira Neto (1994) showed the effects of moisture on the likelihood of porosity. The negative effects of porosity have been investigated in numerous studies (Judd and Wright 1978; Koller et al. 2007).

The influence of porosity can greatly impact the loading capability of composites. In terms of axial loading, the literature seems to indicate more influence on compression strength compared to the tension response, thus pointing toward the matrix-dominated properties (Oliver et al. 1995). During tension, the fiber bears the majority of the load, whereas in compression the resin plays a far more dominant role. Other properties such as the interlaminar shear strength (ILSS) showed significant decrease in performance. Liu et al. (2006) concluded that the decrease in the ILSS is due to the reduction in interface contact between plies due to the

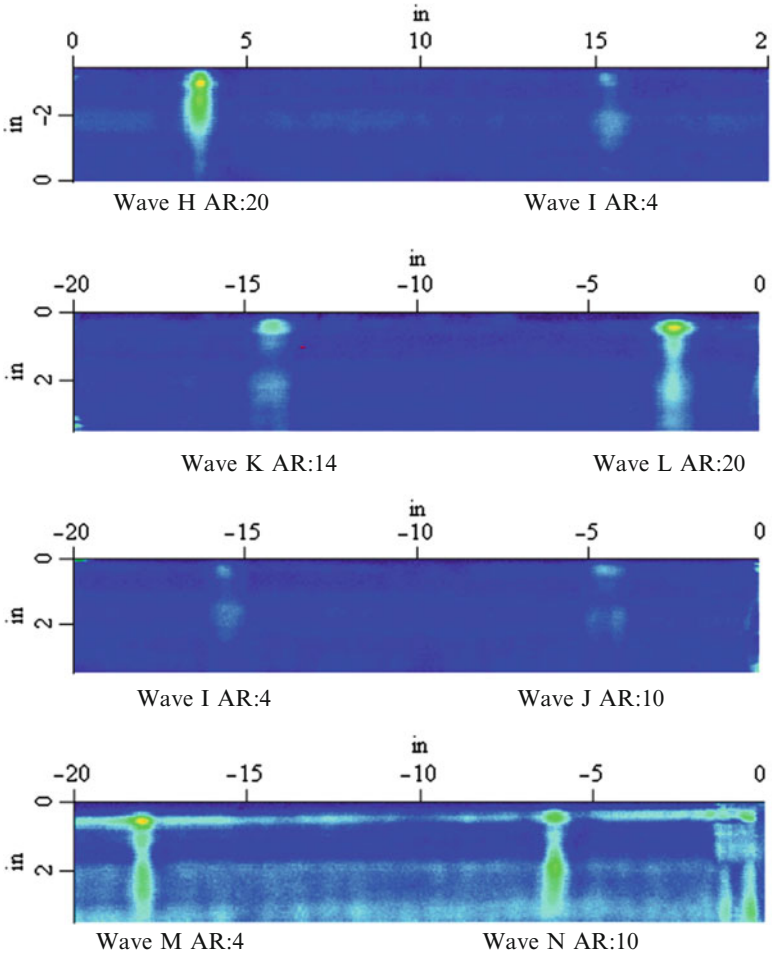


Fig. 11 C-scan correlation to waviness aspect ratio (Chakrapani et al. 2012)

voids. It is important to note that not all properties are impacted. For example, one study has shown that the flexural modulus is generally enhanced due to increases in thickness when porosity occurs (Yang and Elhajjar 2012).

Indirect Measurement Method

The void content in composites can be measured destructively using the indirect measurement method. The method does not give a spatial description of the porosity or voids, but can be helpful in determining the void content. If the density of the composite, fiber, resin and the weight of the fibers and resin are known, it is possible to estimate the void content. The void content in percent can be calculated from the following (ASTM D2734 – 09 2009):

$$V = 100 - M_d\left(\frac{r}{d_r} + \frac{f}{d_f}\right)$$

where M_d is the measured density, r and f are the resin and fiber weight percentage, and d_r and d_f are the density of the resin and fibers, respectively.

Image Analysis Methods

The image analysis method uses the grayscale histogram approach (Daniel et al. 1992). The technique consists of taking the cross-sectional microscopic image and producing its grayscale histogram graph (Mathworks Matlab 2011). This graph is then plotted with intensity of pixels (ranging from 1 to 256) against the frequency that the pixel occurs as shown in Fig. 12. From the examination of the frequency distribution, it is possible to determine the matrix, fiber, and porosity constituents. The fiber/matrix regions are a lighter gray, whereas the porosity regions are dark. The percentage of porosity can be estimated by the ratio of the threshold pixels to the total number of pixels in the image. The volume of voids V_v is estimated from Daniel et al. (1992), where P_t are the number of thresholded pixels and N is the total number of pixels:

$$V_v = \frac{P_t}{N}$$

Ultrasonic and Tomography Methods

Ultrasonic measurements can provide an estimate of the porosity content in a composite laminate and are one of the most commonly used approaches in industry. The method can detect delaminations, matrix cracking, and also voids. The attenuation of the signal and wave speed can also be used to quantify the void fraction. The voids create a change in the acoustic impedance that is directly related to the density allowing the differentiation of the two zones. As mentioned in Part 1, the application of ultrasonic methods in composites is complicated due to the scattering induced by the anisotropic nature of the composite. Ultrasonic methods can be applied in either pulse-echo or through-transmission methods. The attenuation drop in regions of porosity can be correlated to calibration standards obtained from cross-sectional scans. One disadvantage of these methods is that they offer either an overall estimate or a sectional response and provide little information about the overall morphology of the porosity defects. There is little information on whether the porosity manifests itself in a homogenous fashion or in some directional pattern. Some differentiation of stratified or distributed porosity can be made with the methods discussed. This has significant importance when considering the current reliance on 2D cross-sectional methods for analysis of porosity. Other techniques have also been proposed for measurement of porosity in composites, including infrared thermography, spectroscopy, laser ultrasonics, eddy currents, radiography, acoustic emission, and acousto-ultrasonics. In any of these methods, cross-sectional analysis or constituent level methods are used to develop calibration standards.

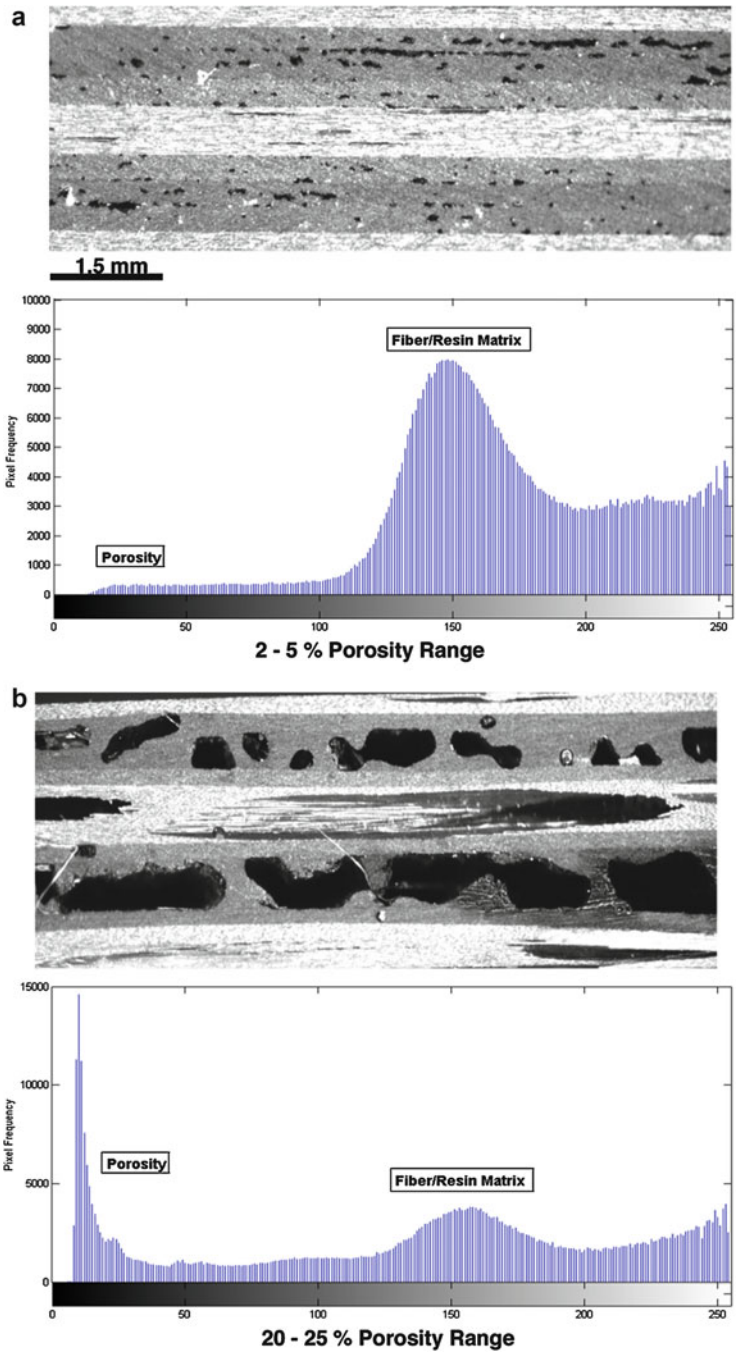


Fig. 12 Grayscale image histogram for estimation of porosity levels (Yang and Elhajjar 2012)

High-resolution X-ray computed tomography (CT) offers one approach for characterizing porosity in composites. In this method, highly charged X-ray particles penetrate from different angles to produce 3D microstructure reconstructions. X-ray CT has been used in studying the microstructure of porous silica–calcium phosphate nanocomposite in load bearing biomedical applications (Gupta et al. 2005). X-ray synchrotron tomography, which uses radiation within the electromagnetic spectrum, has shown promising results in identifying iron-rich areas of metal matrix composites and the detection of porosity as well as damage growth (de Andrade Silva et al. 2010; Williams et al. 2010), and also in self-healing composites (Ghezzo et al. 2010). Comparison of X-ray computed tomography with several conventional methods for void content analysis for fiber-reinforced polymer composites (FRP) shows that X-ray CT can yield more accurate results in determining void content, when compared to ultrasonic testing and acid digestion (Kastner et al. 2010). CT is able to detect void and porosity within 0.5 %, whereas ultrasonic and acid digestion as accurate to within ± 1 %. Damage in FRP such as microcracks, delamination, and fiber breakage can also be observed three-dimensionally and in high resolution (Schilling et al. 2005; Write et al. 2008). At this point, the X-ray CT methods are not portable and only small specimens can be assessed.

Conclusions

The susceptibility of composite structures to both manufacturing defects and in-service damage requires the use of advanced methodologies to characterize the type, location, and extent of these defects. Undetected or ignored, these defects or damage types can result in serviceability issues or a loss in the structural capability. As discussed in this chapter, although ultrasonic methods such as pulse-echo or through-transmission are common in industrial practice, they are not capable of fully quantifying and assessing these important types of damage. The identification of the complex fiber waviness profiles and their link to the structural performance is likely to be a continuing area of research for nondestructive methods. Among the experimental methods for investigation of microscale anomalies or manufacturing anomalies in composites, there is unfortunately not a single one capable of identifying all damage mechanisms. The fact that defects are sometimes coupled together creates further complexity in identifying them. With this chapter, the authors sought to offer an overview of experimental methods able to detect microscale damage and manufacturing anomalies, highlighting challenges and benefits of each method.

References

E. Adolfsson, P. Gudmundson, Matrix crack initiation and progression in composite laminates subjected to bending and extension. Int. J. Solids Struct. **36**, 3131–3169 (1999)

D.G. Aggelis, N.-M. Barkoula, T.E. Matikas, A.S. Paipetis, Acoustic structural health monitoring of composite materials: damage identification and evaluation in cross ply laminates using acoustic emission and ultrasonics. Compos. Sci. Technol. **72**, 1127–1133 (2012)

J. Andersons, R. Joffe, E. Spārniņš, Statistical model of the transverse ply cracking in cross-ply laminates by strength and fracture toughness based failure criteria. Eng. Fract. Mech. **75**, 2651–2665 (2008)

ASTM D2734 – 09, *Standard Test Methods for Void Content of Reinforced Plastics: ASTM D2734 – 09 Standard Test Methods for Void Content of Reinforced Plastics* (ASTM International, West Conshohocken, 2009)

J. Aveston, A. Kelly, Theory of multiple fracture of fibrous composites. J. Mater. Sci. **8**, 352–362 (1973)

E.J. Barbero, F.A. Cosso, F.A. Campo, Benchmark solution for degradation of elastic properties due to transverse cracking in laminated composites. Compos. Struct. **98**, 242–252 (2013)

Y. Bar-Cohen, Emerging NDE technologies and challenges at the beginning of the 3rd millennium, Part II, NDT.net. **5** (2000). Available on http://www.ndt.net/article/v05n02/barcohen/barcohen.htm. Downloaded Oct 2013

J.-M. Berthelot, Transverse cracking and delamination in cross-ply glass-fiber and carbon-fiber reinforced plastic laminates: static and fatigue loading. ASME Appl. Mech. Rev. **56**, 111–147 (2003)

R. Böhm, W. Hufenbach, Experimentally based strategy for damage analysis of textile-reinforced composites under static loading. Compos. Sci. Technol. **70**, 1330–1337 (2010)

D.J. Bull, L. Helfen, I. Sinclair, S.M. Spearing, T. Baumbach, A comparison of multi-scale 3D X-ray tomographic inspection techniques for assessing carbon fibre composite impact damage. Compos. Sci. Technol. **75**, 55–61 (2013)

V.N. Bulsara, R. Talreja, J. Qu, Damage initiation under transverse loading of unidirectional composites with arbitrarily distributed fibers. Compos. Sci. Technol. **59**, 673–682 (1999)

M. Castaings, B. Hosten, T. Kundu, Inversion of ultrasonic, plane-wave transmission data in composite plates to infer viscoelastic material properties. NDT&E Int. **33**, 377–392 (2000)

S. K. Chakrapani, V. Dayal, D.J. Barnard, A. Eldal, R. Krafka, Ultrasonic Rayleigh wave inspection of waviness in wind turbine blades: experimental and finite element method, in *Review of Progress in Quantitative Nondestructive Evaluation*, Burlington, July 2011. AIP Conf. Proc. **1430**, 1911–1917 (2012), http://proceedings.aip.org/resource/2/apcpcs/1430/1/1911_1. Accessed 23 Feb 2013

H.-J. Chun, J.-Y. Shin, I.M. Daniel, Effects of material and geometric nonlinearities on the tensile and compressive behavior of composite materials with fiber waviness. Compos. Sci. Technol. **61**(1), 125–134 (2001)

M.V. Cid Alfaro, A.S.J. Suiker, R. De Borst, Transverse failure behavior of fiber-epoxy systems. J. Compos. Mater. **44**, 1493–1516 (2010)

F.W. Crossman, W.J. Warren, A.S.D. Wang, G.E. Law Jr., Initiation and growth of transverse cracks and edge delamination in composite laminates Part 2. Experimental correlation. J. Compos. Mater. **14**, 88–108 (1980)

I.M. Daniel, S.C. Wooh, I. Komsky, Quantitative porosity characterization of composite materials by means of ultrasonic attenuation measurement. J. Nondestruct. Eval. **11**(1), 1–8 (1992)

V. Dayal, Wave propagation in a composite with a wavy sublamina. J. Nondestruct. Eval. **14**(1), 1–7 (1995). doi:10.1007/bf00735666

S.F. de Andrade Silva, J.W. Williams, B.R. Müller, M.P. Hentschel, P.D. Portella, N. Chawla, Three-dimensional microstructure visualization of porosity and Fe-rich inclusions in SiC particle-reinforced Al Ally matrix composites by X-ray synchrotron tomography. Metall. Mater. Trans. A **41**(8), 2121–2128 (2010)

K. Diamanti, C. Soutis, Structural health monitoring techniques for aircraft composite structures. Prog. Aerosp. Sci. **46**, 342–352 (2010)

R. F. Elhajjar, M. T. Lo Ricco, A modified average stress criterion for open-hole tension strength in the presence of localized wrinkling. Plast. Rubbers Compos. **41**(9), 396–406 (2012)

R.F. Elhajjar, D.R. Petersen, Gaussian function characterization of unnotched tension behavior in a carbon/epoxy composite containing localized fiber waviness. Compos. Struct. **93**(9), 2400–2408 (2011). doi:10.1016/J.Compstruct.2011.03.029

L. Farge, J. Varna, Z. Ayadi, Use of full-field measurements to evaluate analytical models for laminates with intralaminar cracks. J. Compos. Mater. **46**, 2739–2752 (2012)
J.P. Favre, J.C. Laizet, Amplitude and counts per event analysis of the acoustic emission generated by the transverse cracking of cross-ply CFRP. Compos. Sci. Technol. **36**, 27–43 (1989)
R. Gauvin, M. Chibani, P. Lafontaine, The modeling of pressure distribution in resin transfer molding. J. Reinf. Plast. Compos. **6**(4), 367–377 (1987)
F. Ghezzo, D.R. Smith, T.N. Starr, T. Perram, A.F. Starr, T.K. Darlington, R.K. Baldwin, S.J. Oldenburg, Development and characterization of healable carbon fiber composites with a reversibly cross linked polymer. J. Compos. Mater. **44**(13), 1587–1603 (2010)
G. Gupta, A. Zbib, A. El-Ghannam, M. Khraisheh, H. Zbib, Characterization of a novel bioactive composite using advanced X-ray computed tomography. Compos. Struct. **71**(3), 423–428 (2005)
A.L. Highsmith, K.L. Reifsnider, Stiffness reduction mechanisms in composite laminates, in *Damage in Composite Materials*, ed. by K.L. Reifsnider (ASTM STP 775, Philadelphia, 1982), pp. 103–117
H.M. Hsiao, I.M. Daniel, Effect of fiber waviness on stiffness and strength reduction of unidirectional composites under compressive loading. Compos. Sci. Technol. **56**(5), 581–593 (1996a)
H.M. Hsiao, I.M. Daniel, Nonlinear elastic behavior of unidirectional composites with fiber waviness under compressive loading. J. Eng. Mater. T ASME **118**(4), 561–570 (1996b)
H.M. Hsiao, I.M. Daniel, Elastic properties of composites with fiber waviness. Compos. Part A Appl. S **27**(10), 931–941 (1996c)
K. Jemielniak, Some aspects of acoustic emission signal pre-processing. J. Mater. Process Technol. **109**(3), 242–247 (2001)
N.C.W. Judd, W.W. Wright, Voids and their effects on the mechanical properties of composites-an appraisal. SAMPE J. **14**(1), 10–14 (1978)
J. Kastner, B. Plank, D. Salaberger, J. Sekelja. Defect and porosity determination of fibre reinforced polymers by x-ray computed tomography, in *2nd International Symposium on NDT in Aerospace 2010* (Hamburg, Germany 2010), http://www.ndt.net/article/aero2010/papers/we1a2.pdf. Accessed 23 Feb 2013
S.S. Kessler, S.M. Spearing, C. Soutis, Damage detection in composite materials using Lamb wave methods. Smart Mater. Struct. **11**, 269–278 (2002)
R. Koller, S. Chang, Y. Xi, Fiber-reinforced polymer bars under freeze-thaw cycles and different loading rates. J. Compos. Mater. **41**(1), 5–25 (2007)
M.C. Lafarie-Frenot, C. Hénaff-Gardin, Formation and growth of 90° ply fatigue cracks in carbon/epoxy laminates. Compos. Sci. Technol. **40**, 307–324 (1991)
V. La Saponara, W. Lestari, C. Winkelmann, L. Arronche, H-Y. Tang, in *Review of Progress in Quantitative Nondestructive Evaluation*, San Diego, July 2010. AIP Conf. Proc. **1335**, 927–934 (2011), http://proceedings.aip.org/resource/2/apcpcs/1335/1/927_1. Accessed 23 Feb 2013
J. Lee, C. Soutis, A study on the compressive strength of thick carbon fibre-epoxy laminates. Compos. Sci. Technol. **67**(10), 2015–2026 (2007)
L. Liu, B.-M. Zhang, D.-F. Wang, Z.-J. Wu, Effects of cure cycles on void content and mechanical properties of composite laminates. Compos. Struct. **73**(3), 303–309 (2006)
D. Lovering, Boeing finds new problem in 787, installing Patch. *Seattle Times* (2009), http://seattletimes.nwsource.com/html/localnews/2009664552_apusboeing7874thldwritethru.html. Accessed 23 Feb 2013
S. Mall, Integrity of graphite/epoxy laminate embedded with piezoelectric sensor/actuator under monotonic and fatigue loads. Smart Mater. Struct. **11**, 527–533 (2002)
P.W. Manders, T.-W. Chou, F.R. Jones, J.W. Rock, Statistical analysis of multiple fracture in 0°/90°/0° glass fibre/epoxy resin laminates. J. Mater. Sci. **18**, 2876–2889 (1983)
Mathworks, *Matlab*, R2011a edn. (Mathworks, Natick, 2011)
L.N. McCartney, G.A. Schoeppner, W. Becker, Comparison of models for transverse ply cracks in composite laminates. Compos. Sci. Technol. **60**, 2347–2359 (2010)

A.J. Moffat, P. Wright, L. Helfen, T. Baumbach, G. Johnson, S.M. Spearing, I. Sinclair, In situ synchrotron computed laminography of damage in carbon fibre-epoxy [90/0]s laminates. Scripta Mater. **62**, 97–100 (2010)

S.F. Muller de Almeida, Z.S. Nogueira Neto, Effect of void content on the strength of composite laminates. Compos. Struct. **28**(2), 139–148 (1994)

N.K. Naik, Woven-fibre thermoset composites, in *Fatigue in Composites: Science and Technology of the Fatigue Response of Fibre-Reinforced Plastics*, ed. by B. Harris (CRC Press, Boca Raton, 2003)

A.H. Nayfeh, The general problem of elastic wave propagation in multilayered anisotropic media. J. Acoust. Soc. Am. **89**(4), 1521–1531 (1991)

K. Ogi, S. Yashiro, K. Niimi, A probabilistic approach for transverse crack evolution in a composite laminate under variable amplitude cyclic loading. Compos. Part A Appl. S **41**, 383–390 (2010)

P. Oliver, J.P. Cottu, B. Ferret, Effects of cure cycle pressure and voids on some mechanical properties of carbon/epoxy laminates. Composites **26**(7), 509–515 (1995)

F. París, A. Blázquez, L.N. McCartney, A. Barroso, Characterization and evolution of matrix and interface related damage in [0/90]s laminates under tension. Part II: Experimental evidence. Compos. Sci. Technol. **70**, 1176–1183 (2010)

Y. Promboon, *Acoustic Emission Source Location* (The University of Texas, Austin, 2000), p. 343

I. I. Qamhia, E. M. Lauer-Hunt, R. Elhajjar, Identification of acoustic emissions from porosity and waviness defects in continuous fiber reinforced composites. ASTM J. Adv. Civ. Eng. Mater. **2**(1), 14 pp (2013)

X.P. Qing, S.J. Beard, A. Kumar, T.K. Ooi, F.-K. Chang, Built-in sensor network for structural health monitoring of composite structures. J. Intel. Mater. Syst. Struct. **18**, 39–49 (2007)

J.L. Rose, A. Pilarski, J.J. Ditri, An approach to guided wave mode selection for inspection of laminated plate. J. Reinf. Plast. Compos. **12**, 536–544 (1993)

C. Roy, M. Elghorba, Monitoring progression of mode-II delamination during fatigue loading through acoustic-emission in laminated glass-fiber composite. Polym. Compos. **9**(5), 345–351 (1988)

K. Schaaf, P. Rye, F. Ghezzo, A. Starr, S. Nemat-Nasser, Optimization of mechanical properties of composite materials with integrated embedded sensors networks, in *Proceedings of SPIE, Smart Structures and Materials: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, vol 6174, (San Diego, CA, 2007), p. 617443 (5 pp.)

P.J. Schilling, B.P.R. Karedia, A.K. Tatiparthi, M.A. Verges, P.D. Herrington, X-ray computed microtomography of internal damage in fiber reinforced polymer matrix composites. Compos. Sci. Technol. **65**, 2071–2078 (2005)

M.P.J. Schöpfer, A. Arslan, J.J. Walsh, C. Childs, Reconciliation of contrasting theories for fracture spacing in layered rocks. J. Struct. Geol. **33**, 551–565 (2011)

K.J. Schubert, A.S. Herrmann, On attenuation and measurement of Lamb waves in viscoelastic composites. Compos. Struct. **94**, 177–185 (2011)

K.J. Schubert, A.S. Herrmann, On the influence of moisture absorption on Lamb wave propagation and measurement in viscoelastic CFRP using surface applied piezoelectric sensors. Compos. Struct. **94**, 3635–3643 (2012)

A.E. Scott, M. Mavrogordato, P. Wright, I. Sinclair, S.M. Spearing, In situ fibre fracture measurement in carbon-epoxy laminates using high resolution computed tomography. Compos. Sci. Technol. **71**, 1471–1477 (2011)

M.D. Seale, B.T. Smith, W.H. Prosser, Lamb wave assessment of fatigue and thermal damage in composites. J. Acoust. Soc. Am. **103**, 2416–2424 (1998)

V.V. Silberschmidt, Matrix cracking in cross-ply composites: effect of randomness. Compos. Part A Appl. S **36**, 129–135 (2005)

D.A. Singh, A.J. Vizzini, Structural integrity of composite laminates with interlaced actuators. Smart Mater. Struct. **3**, 71–79 (1994)

P.M. Sisneros, P. Yang, R.F. Elhajjar, Fatigue and impact behaviour of carbon fibre composite bicycle forks. Fatigue Fract. Eng. M **35**(7), 672–682 (2012)
BF. Sørensen, R. Talreja, Effect of nonuniformity of fiber distribution on thermally-induced residual stresses and cracking in ceramic matrix composites. Mech. Mater. **16**, 351–363 (1993)
K.V. Steiner, R.F. Eduljee, X. Huang, J.W. Gillespie, Ultrasonic NDE techniques for the evaluation of matrix cracking in composite laminates. Compos. Sci. Technol. **53**(2), 193–198 (1995)
W.W. Stinchcomb, Nondestructive evaluation of damage accumulation processes in composite laminates. Compos. Sci. Technol. **25**, 103–118 (1986)
Z. Su, L. Ye, Y. Lu, Guided Lamb waves for identification of damage in composite structures: a review. J. Sound Vib. **295**, 753–780 (2006)
M.P.F. Sutcliffe, S.L. Lemanski, A.E. Scott, Measurement of fibre waviness in industrial composite components. Compos. Sci. Technol. **72**(16), 2016–2023 (2012)
R. Talreja, C.V. Singh, *Damage and Failure of Composite Materials* (Cambridge University Press, New York, 2012)
H.-Y. Tang, C. Winkelmann, W. Lestari, V. La Saponara, Composite structural health monitoring through use of embedded PZT sensors. J. Intel. Mater. Syst. Struct. **22**, 739–755 (2011)
M. Thomas, N. Boyard, L. Perez, Y. Jarny, D. Delaunay, Representative volume element of anisotropic unidirectional carbon-epoxy composite with high-fibre volume fraction. Compos. Sci. Technol. **68**, 3184–3192 (2008)
J. Tong, F.J. Guild, S.L. Ogin, P.A. Smith, On matrix crack growth in quasi-isotropic laminates −1. Experimental investigation. Compos. Sci. Technol. **57**, 1527–1535 (1997)
C. Toscano, C. Vitiello, Influence of the stacking sequence on the porosity in carbon fiber composites. J. Appl. Polym. Sci. **122**(6), 3583–3589 (2011)
D. Tsamtsakis, M. Wevers, P. de Meester, Acoustic emission from CFRP laminates during fatigue loading. J. Reinf. Plast. Compos. **17**, 1185–1201 (1998)
G.Z. Voyiadjis, A.H. Almasri, Experimental study and fabric tensor quantification of microcrack distribution in composite materials. J. Compos. Mater. **41**, 713–745 (2007)
A.S.D. Wang, K.C. Yan, On modeling matrix failure in composites. Compos. Part A Appl. S **36**, 1335–1346 (2005)
A.W. Wharmby, F. Ellyin, J.D. Wolodko, Observations on damage development in fibre reinforced polymer laminates under cyclic loading. Int. J. Fatigue. **25**, 437–446 (2003)
J.J. Williams, Z. Flom, A.A. Amell, N. Chawla, X. Xiao, F. De Carlo, Damage evolution in SiC particle reinforced Al alloy matrix composites by X-ray synchrotron tomography. Acta Mater. **58**, 6194–6205 (2010)
M.R. Wisnom, J.W. Atkinson, Fibre waviness generation and measurement and its effect on compressive strength. J. Reinf. Plast. Compos. **19**(2), 96–110 (2000)
P.J. Withers, M. Preuss, Fatigue and damage in structural materials studied by X-ray tomography. Annu. Rev. Mater. Res. **42**, 81–103 (2012)
S.-C. Wooh, I.M. Daniel, Wave propagation in composite materials with fibre waviness. Ultrasonics **33**(1), 3–10 (1995)
P. Wright, A. Moffat, I. Sinclair, S.M. Spearing, High resolution tomographic imaging and modelling of notch tip damage in a laminated composite. Compos. Sci. Technol. **70**, 1444–1452 (2010)
P. Wright, X. Fu, I. Sinclair, S.M. Spearing, Ultra-high resolution computed tomography of damage in notched carbon fiber-epoxy composites. J. Compos. Mater. **42**, 1993–2002 (2008)
P. Yang, R. Elhajjar, Porosity defect morphology effects in carbon fiber – epoxy composites. Polym. Plast. Technol. **51**(11), 1141–1148 (2012)
S.W. Yurgartis, B.S. MacGibbon, P. Mulvaney, Quantification of microcracking in brittle-matrix composites. J. Mater. Sci. **27**, 6679–6686 (1992)
C. Zhang, W.K. Binienda, G.N. Morscher, R.E. Martin, L.W. Kohlman, Experimental and FEM study of thermal cycling induced microcracking in carbon/epoxy triaxial braided composites. Compos. Part A Appl. S **46**, 34–44 (2013)
S.M. Ziola, M.R. Gorman, Source location in thin plates using cross-correlation. J. Acoust. Soc. Am. **90**(5), 2551–2556 (1991)

Modeling Deformation and Damage of Random Fiber Network (RFN) Materials

41

Rickard Hägglund and Per Isaksson

Contents

Abstract

Fiber materials are used in a wide range of products. Fibers can be directly bonded to each other to form 2D and 3D network materials, commonly referred to as random fiber network (RFN) materials. Examples of such materials are

R. Hägglund (✉)
SCA R&D CENTRE AB, SCA Group, Sundsvall, Sweden
e-mail: rickard.hagglund@sca.com

P. Isaksson
Department of Engineering Sciences, Division of Applied Mechanics, Uppsala Universitet, Uppsala, Sweden
e-mail: per.isaksson@angstrom.uu.se

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_17

paper, textiles, or nonwoven felts. An RFN represents a random structure on the microscale, which governs a complex deformation and fracture behavior. This is due to a size effect in the mechanical behavior of the network because of its heterogeneous nature: the fibers introduce long-range microstructural effects in the material that distribute forces and deformation in a complex manner that is very different compared to more continuous materials. A fiber network material used in a product must meet performance requirements through the entire use cycle including manufacturing and end-use situations. This chapter presents a framework for analyzing deformation, damage, and fracture in network materials using continuum damage mechanics. Material degradation, or damage, is described in a diffuse sense and the influence of damage on the mechanical properties is governed by an internal length variable. To correctly describe gradients in strain and damage, a nonlocal field theory (gradient theory) must be added to the framework. A mathematical framework for such theory is presented.

Introduction

A fiber is an elongated object that can be used to form different materials. There are many uses for fibers in society today. They can be spun into ropes and wires or used as a component of composite materials. Fibers can be directly bonded to each other to form 2D and 3D network materials, referred to as random fiber network (RFN) materials. 2D network materials, i.e., materials where the fibers are essentially organized in one plane, are seen in a large range of applications. Conventional paper is an important network material and is used on a daily basis for a variety of industry and household needs. Paper consists of an essentially planar network of cellulose fibers and is manufactured by dewatering a fiber suspension (Fig. 1). Bonds are created between the fibers when a fiber suspension is dried. The bonds can be tailored to have different properties using chemical additives depending on the requirements.

Nonwoven felt is a planar network material used across a wide range of applications and products. These fabrics are manufactured from a set of polymer fibers consolidated by bonds of different nature, such as entanglement, local thermal fusion, or chemical binders. These materials are found in products such as baby diapers, filters, and packaging. An emerging group of network materials are materials formed by fibers having cross-sectional dimensions on the nanometer scale. An example of that is nanocellulose (Henriksson et al. 2008; Eichhorn et al. 2010). By extracting cellulose at the nanometer scale, the majority of the defects associated with the hierarchical structure of wood fibers can be removed, and new small cellulose building blocks are formed that are very stiff and slender. 3D networks are found in various contexts such as living tissues like bones, muscles, and plant stems, all being examples of natural fiber networks. Manufactured fiber network materials are frequently used in the automotive and aerospace industry. Novel biomedical implants are being developed to mimic the network structure of human tissue.

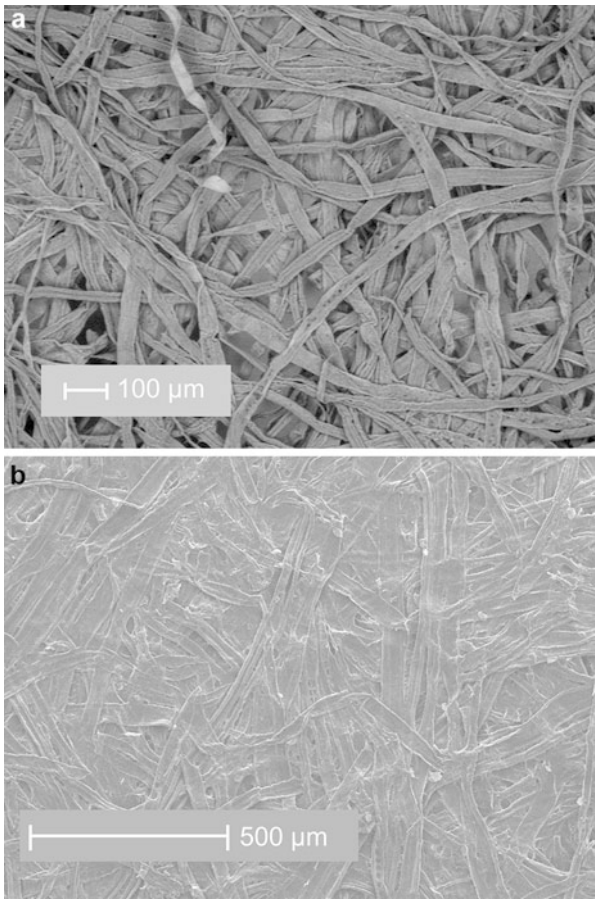


Fig. 1 (**a**) Open network (tissue paper). (**b**) Dense network (packaging paper)

Fig. 2 Fracture testing of tissue paper. Short (10 mm) and long (70 mm) initial cracks

An RFN represents a random structure on the microscale, which governs its complex deformation and fracture behavior. Figure 2 shows images from typical fracture tests on tissue paper containing central prefabricated straight cracks of 10 and 70 mm. A relatively large defect size is required to localize macroscopic fracture to the crack, i.e., that material is relatively insensitive to defects.

In a homogenous material, such as metals, glass, or polymers, the fracture process would localize to the initial crack tips and the crack would grow essentially along the original crack plane. In the open fiber network, the local stress intensity at a tip is reduced because of deformation process taking place on the microscale (Isaksson and Hägglund 2009). In present chapter the nature of deformation and damage in RFN materials is presented along with modeling concepts. During the last decades the fracture process in randomly distributed fiber networks has been extensively studied (cf. Delaplace et al. 1996; Herrmann and Roux 1990; Heyden and Gustafsson 1998; Niskanen 1993; Ramasubramanian and Perkins 1987 or Åström and Niskanen 1991, 1993). A fiber network material used in a product must meet performance requirements through the entire use cycle including manufacturing and end-use situations. To be able to design products in a cost-effective manner, it is necessary to know the coupling between the properties of the fibers and the mechanical response of a product. This coupling can be accomplished with network models where each fiber is described with coupled structural members that can be stretched and bent. Besides elastic properties, strength and damage can be analyzed with such an approach. However, for large structures such an approach becomes difficult to realize in common computer environments due to the high computational cost involved. In addition, during the manufacturing of network materials, the fibers sometimes change dimensions and mechanical properties and internal stresses are introduced, which complicates the coupling between structural levels. In contrast to network models, theories of continuum damage mechanics (CDM) offer a framework for describing material degradation at macroscopic continuum level. Special attention is given to planar network materials where the fibers extend predominantly in one plane. For such sheet materials assumptions of plane stress conditions can be made.

Deformations and Damage Mechanisms in Network Materials

Fibers form a material only if the fibers are bonded to each other in a coherent structure. Forces are transferred from fiber to fiber through the bonds upon loading. The mechanical properties of network materials are governed by three principal parameters: (i) fiber properties, (ii) properties of bonds and (iii) number of load-carrying bonds along the fibers.

Elasticity

The effective elastic modulus of network materials is less than the elastic modulus of the constituent fibers, which is partly explained by the presence of voids. The length distribution of free fiber segments is exponential in an RFN structure (Kallmes and Corte 1960). Cox (1952) developed a model that relates elastic constants of the sheet (Young's modulus E and Poisson's ratio υ) to fiber stiffness based on sheet and fiber densities and elastic modulus of the fibers (ρ_s, ρ_f, E_f)

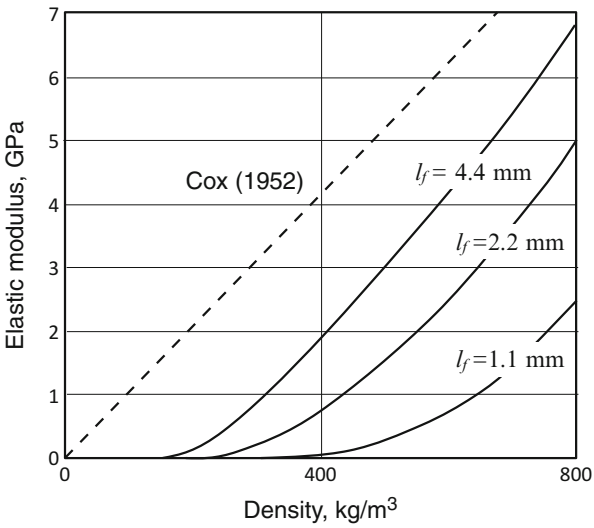


Fig. 3 Elastic modulus versus density for three different constant fiber lengths: $l_f = 1.1$, 2.2 and 4.4 mm (After Bronkhorst (2003))

according to $E = \frac{1}{3}\rho_f E_f/\rho_s$ and $\nu = \frac{1}{3}$. The model assumes that the sheet is in-plane isotropic and that the fibers are infinitely long and not allowed to bend or interact with other fibers. The theory overpredicts the sheet modulus. In true network materials, loads are transferred from fiber to fiber and the axial load on a fiber is zero at its ends, which creates a gradient in stress along the fiber. The effect of fiber length and network density must be included in order to properly analyze coupling between sheet modulus and fiber modulus. In a model developed by Van den Akker (1962), it was assumed that apart from axial strain the unbonded parts of the fibers could also sustain bending and shear. Network models have been used to establish detailed relations between fiber stiffness and network stiffness (Jangmalm and Östlund 1995; Heyden 2000; Åslund and Isaksson 2011). Figure 3 shows estimates of sheet elastic modulus obtained using a network model (Bronkhorst 2003). The graph shows the relation for networks relevant for paper materials for different fiber lengths and sheet densities. For very low densities the fibers are unable to form a coherent network, which makes the structure unable to carry and results in a stiffness of zero. In open fiber structures, the free fiber segments are in average long and therefore allowed to bend, which makes the material relatively compliant. As the number of bonds per fiber increases, the fiber segments become shorter and the network becomes stiffer.

The upper broken line in Fig. 3 shows the model suggested by Cox (1952). Bronkhorst's model is qualitatively in agreement with experiments on paper materials reported in literature (Luner et al. 1961).

Typically Poisson's ratio is close to 0.3 for in-plane isotropic planar networks (Heyden 2000). According to network models the in-plane Poisson's ratio is only marginally influenced by network density.

Damage

When a network material, such as paper or nonwoven felts, is strained into the nonlinear regime, the elastic modulus is reduced (Fig. 4) (Isaksson et al. 2004; Allaoui et al. 2008; Ridruejo et al. 2010; Coffin 2012). This is a macroscopic manifestation of the damage process taking place at microscale.

In the initial linear part of the load–displacement curve, the material deforms elastically. At continued loading the damage evolution in paper materials is initially random over the loaded volume of the material and exhibits stiffness reduction. Typically the dominating damage processes are combinations of bond failure or fiber breakage. For paper materials there is a loop in the load–displacement curve when the material is subjected to loading/unloading cycles, which is due to deformation mechanisms in the cellulose fiber (Jentzen 1964). As damage evolution progress, micro damage localizes to a narrow band. From that point the material undergoes a softening behavior. Finally formation and growth of a macroscopic crack occur.

Figure 5 shows typical damage behavior of paper materials in tension. In open networks of low degree of bonding, the failure process is dominated by debonding. In dense paper structures, discrete microcracks are formed during loading, oriented perpendicular to the loading direction (Allaoui et al. 2008). The microcracks appear to be a combination of debonding and fiber breakage. At final failure, microcracks coalesce to generate a macrocrack.

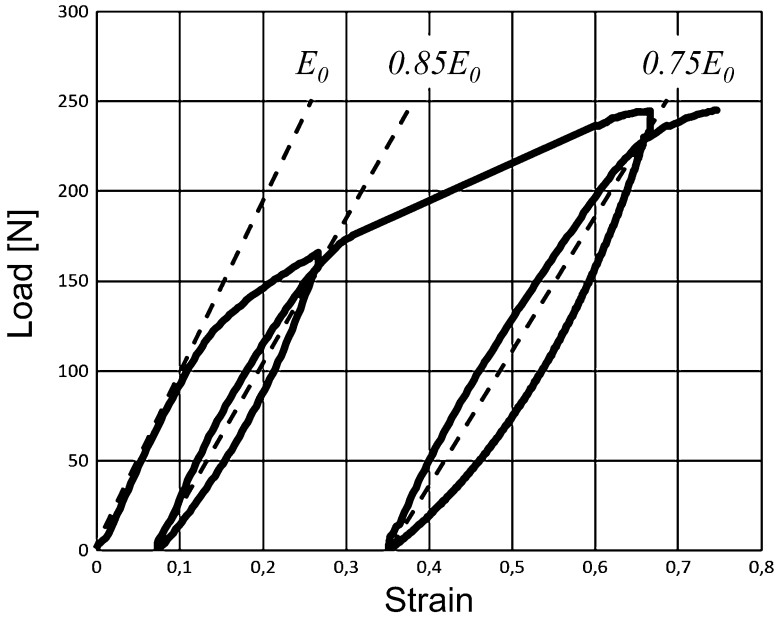


Fig. 4 Cyclic loading in tension of a slender specimen made of packaging paper (dense network). Here, the gauge length is 100 mm and the width is 15 mm

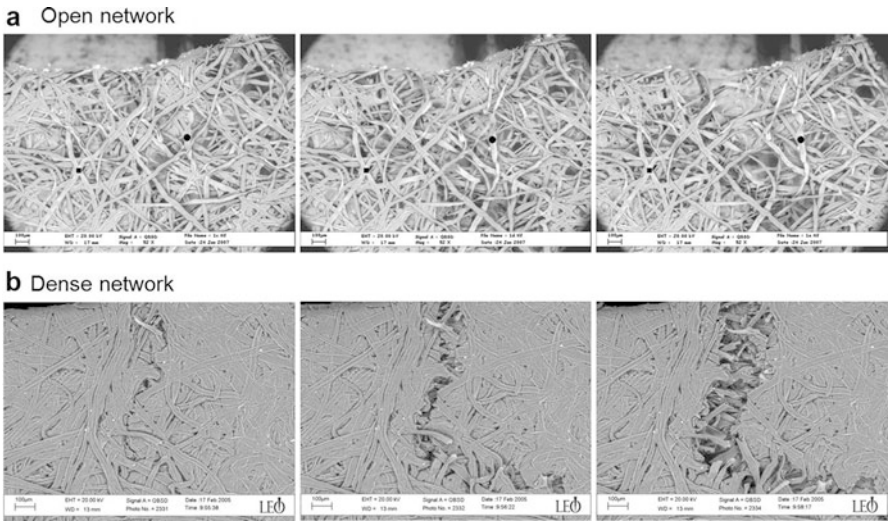


Fig. 5 Damage in paper under tensile loading. (**a**) Open network (tissue paper). (**b**) Dense network (packaging paper)

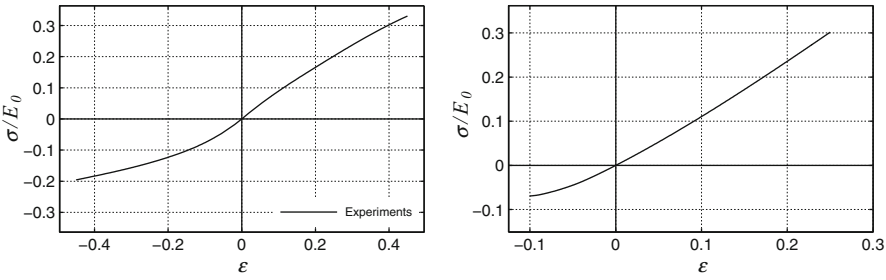


Fig. 6 *Left*: Experimental tensile and compression test. *Right*: Example of stress–strain curve estimated by a numerical model. The stress is normalized with the initial stiffness E_0 (Reprinted from J. Comp. Materials, Åslund PE, Isaksson P (2011). A note on the nonlinear mechanical behavior of planar random network structures subjected to in-plane compression. Comp. Mater. 45: 2697–2703, with permission from SAGE)

For paper materials, the damage caused by breaking of bonds can be monitored, e.g., using silicon-impregnated paper samples. When breakings of bonds occur, the light scattering coefficient of the materials is altered and the damage zone appears (Korteoja et al. 1996). For relatively dense paper materials, the stiffness of the silicone is low compared to that of the fiber. Niskanen et al. (2001) showed in an experimental investigation that the width of the damage band is governed by the fiber length.

On a macroscopic level, fiber networks often have different stress–strain response in tension and compression (Fig. 3). Typically, networks are more resistant in tension than in compression. There are practical difficulties in measuring the

edgewise compressive properties of thin sheets of network materials. The stress–strain response is dictated by the network structure, fiber, and bond properties. Sachs and Kuster (1980) suggest that dense paper fails in compression by a combination of enlargement of voids, tearing of fiber walls, and separation between the fiber layers. Fellers et al. (1980) suggest that compressive strength in dense paper is governed by the compressive strength of the fibers. For open network structures failure in compression is governed by buckling of fiber segments. Åslund and Isaksson (2011) showed in a numerical model that even though a linear elastic material model is used for the individual fibers, the network gives a nonlinear response in compression because individual fibers buckle and bend (Fig. 6).

Structural Effects

In materials where inherent heterogeneities such as pores, grains, or fibers are relatively large compared to the size of other relevant dimensions, such as the size of the body or defects, global-scale deformations given by classical elasticity theory are distorted due to deformation processes taking place at a smaller scale in the material. This dependence on the size of the microstructure is often referred to as *material length scales* or *length effects*. This behavior is seen for RFN materials and especially when the network density is low. The fibers introduce long-range microstructural effects in the material that distribute macroscopic stresses and strains in a complex manner. This structural phenomenon resembles what happens in a game of pick-up sticks (*Mikado*) when a player is trying to remove a stick without disturbing the remaining ones (Fig. 7).

If a macroscopic crack is present, the fibers distribute stresses near the crack tip and thus reduce the probability of failure immediately ahead of the macroscopic crack since forces are transferred to remote regions from the tip. For open fiber network materials, such as tissue paper and nonwoven materials, it is favorable to include length scale formulations to account for structural effects (cf. Coffin and

Fig. 7 A game of pick-up sticks

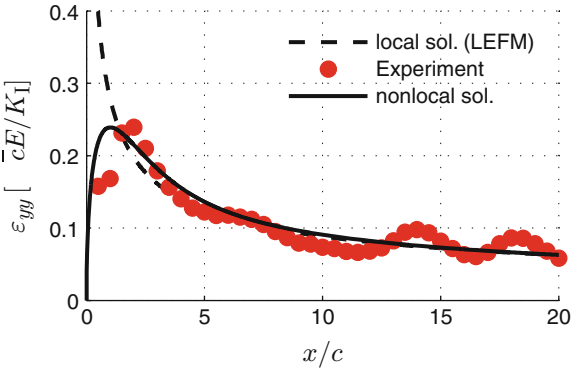


Fig. 8 Experimentally estimated strain field in front of a mode I crack in an open network (tissue paper). Also plotted are a corresponding nonlocal strain field and a classical singular LEFM strain field. For illustrative purposes, the fields are slightly differently normalized. K_{I} is the mode I stress intensity factor acting far away from the tip (Reprinted from Engineering Fracture Mechanics, Isaksson P, Hägglund R (2013). Crack-tip fields in gradient-enhanced elasticity. Engng Frac. Mech. 97:186–192, and from International Journal of Solids and Structures, Isaksson P, Hägglund R (2009) Structural effects on deformation and fracture of random fiber networks and consequences on continuum models. Int. J. Solids Struct. 46: 2320–2329, with permission from Elsevier)

Li 2011; Isaksson et al. 2004; Ridruejo et al. 2010). Nonlocal or gradient theories provide a framework that enables length parameters to be included in a continuum framework by introducing an internal length scale in the governing equations. The difference between a local and nonlocal model is illustrated in Fig. 8. The nonlocal model produces a finite strain field that is similar to the measured, while the local model tends to infinity at the crack tip.

Continuum Damage Model

The model discussed here is very much influenced by the work of, e.g., Peerlings et al. (1996) or Pijaudier-Cabot and Bazant (1987). A planar fiber material is here described by a two-dimensional continuum. The fundamental idea is that material degradation, or damage, can be described in a diffuse sense and the influence of damage on the mechanical properties is represented by an internal variable. Historically this idea dates back to a paper by Kachanov from (1958) where creep rupture in terms of a continuously growing damage parameter was discussed. In the simplest form of CDM, an isotropic scalar damage parameter D is introduced characterizing the degradation in a point of the material. A virgin undamaged material is characterized by $D = 0$ while $D = 1$ corresponds to a fully disintegrated material. Isotropic damage is referred to the situation where the degradation of the elastic stiffness tensor depends on a single parameter and the damage growth rate also depends on a single parameter.

General Equations for Continuum Damage Mechanics

The physical stress tensor σ_{ij} is replaced by an effective stress tensor $\hat{\sigma}_{ij}$, and the stresses are related by the scalar damage parameter according to

$$\sigma_{ij} = (1 - D)\hat{\sigma}_{ij}. \tag{1}$$

Apparently the physical stress $\sigma_{ij} = \hat{\sigma}_{ij}$ for an undamaged material and $\sigma_{ij} \to 0$ at the moment of rupture. There exist different theories regarding the effect of D on the constitutive parameters. According to the elastic strain equivalence postulate (cf. Lemaitre and Chaboche 1990), the relation between the present elastic stiffness tensor C_{ijkl} and the virgin stiffness tensor C^0_{ijkl} is given by

$$C_{ijkl} = (1 - D)C^0_{ijkl}. \tag{2}$$

The stress–strain relation can consequently be written as

$$\sigma_{ij} = (1 - D)C^0_{ijkl}\varepsilon_{kl}. \tag{3}$$

where ε_{ij} is the elastic macroscopic strain tensor. On incremental form, Eq. 3 results in $\mathrm{d}\sigma_{ij}(1 - D)\mathrm{d}\hat{\sigma}_{ij} - \mathrm{d}D\hat{\sigma}_{ij}$ or, the equivalent, $\mathrm{d}\sigma_{ij} = (1 - D)C^0_{ijkl}\mathrm{d}\varepsilon_{kl} - \mathrm{d}DC^0_{ijkl}\varepsilon_{kl}$. The driving force for damage evolution in a point is referred to as the damage energy release rate ψ and is the thermodynamic force conjugate to D. According to Clausius–Duhem inequality, $\psi = -\partial W/\partial D$ where W is the elastic strain energy density. Assuming that the material in its virgin state is linear elastic and that the damage is isotropic, the driving force for damage reads

$$\psi = \frac{1}{2}\varepsilon_{ij}C^0_{ijkl}\varepsilon_{kl}. \tag{4}$$

To study progress of damage, it is necessary to define a damage criterion and evolution law. According to some performed network simulation models, the development of macroscopic material degradation follows an exponential two-parameter law, consisting of an onset parameter and a fracture rate parameter (Hägglund and Isaksson 2007). Thus, for RFN it is physically motivated to use a damage evolution law that on incremental form reads

$$\mathrm{d}D = k(1 - D)\mathrm{d}\psi \quad \text{for} \quad \psi \geq \psi_0, \text{and on integrated form} \tag{5}$$

$$D = 1 - e^{-\beta(\psi/\psi_0 - 1)} \text{ for } \psi \geq \psi_0, \tag{6}$$

where the dimensionless parameter $\beta = k\psi_0$ is introduced. The parameter ψ_0 is the damage threshold, equal to the strain energy density at damage nucleation, and k governs the *rate* of damage evolution. In the context of damage in paper materials, it has been shown that the two damage parameters can be calibrated using data from acoustic emission monitored tensile tests. The theory can be extended to anisotropic network material where a damage parameter in each material direction is considered.

Gradient Formulation

Standard local continuum damage models as described by Eqs. 1, 2, 3, 4, 5, and 6 break down when the macroscopic strains become inhomogeneous as discussed in section "Structural Effects." Such theory does not take into account structural effects in the material. In network materials where damage develops in a diffuse manner upon localization and develops over a region governed by network and fiber dimensions, such as fiber length l_f and free fiber segment length l_s. Stresses depend on the strains not only at an individual point under consideration, but in all points in a specified volume surrounding the point (Fig. 9) (see section "Structural Effects").

There exist several techniques for incorporating structural effects into the governing equations. These methods control the scaling of damage localization and provide means for regularizing the governing equations. Physically it can be interpreted as taking long-range interaction at microscale into account. Examples of such theories include nonlocal integral formulations (cf. Kröner 1967; Eringen and Edelen 1972; Eringen 2002; Silling 2000) and gradient theories (cf. Peerlings et al. 2001; Aifantis 2011). Herein, a gradient formulation is considered. The theory is based on the assumption that a nonlocal counterpart $\overline{\xi}$ in a point (x_1, x_2) of a local state variable ξ over a surrounding two-dimensional infinite domain Ω is given by spatial averaging in a vicinity of the point according to

$$\overline{\xi}(x_1, x_2) = \chi^{-1} \int_{\Omega} \phi\left(x_1', x_2'; x_1, x_2\right) \xi\left(x_1', x_2'\right) \mathrm{d}\Omega, \tag{7}$$

where (x_1', x_2') is the position of the infinitesimal area $\mathrm{d}\Omega$, $\chi = \int_{\Omega} \phi\left(x_1', x_2'; x_1, x_2\right) \mathrm{d}\Omega$ a normalizing factor, and ϕ a symmetric weight function. The formulation (7) can for sufficiently smooth fields of ξ be rewritten into a gradient formulation around (x_1, x_2) according to the Taylor expansion:

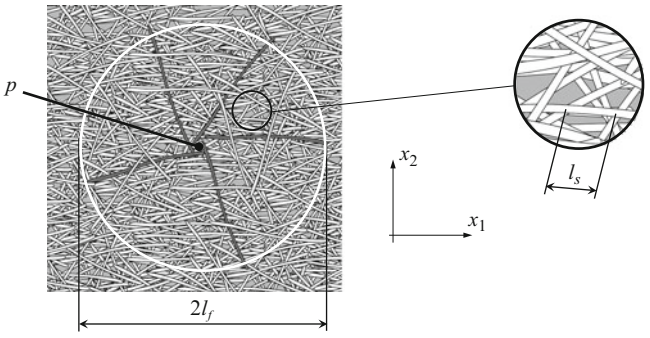


Fig. 9 Interaction between fibers around a point p

$$\xi\left(x_1', x_2'\right) = \sum_{n=0}^{j} \frac{1}{n!}(r\nabla)^n \xi(x_1, x_{23}), \tag{8}$$

where j denote the order of the series expansion, ∇ denote the differential operator, and r is the distance between (x_1, x_2) and (x_1', x_2'). Assuming a Gaussian distribution of the weight function ϕ and dropping unsymmetric terms in Eq. 8 yields after substitution of Eqs. 8 into 7

$$\overline{\xi}(x_1, x_2) = \xi(x_1, x_{23}) + \frac{1}{m!} c^{2m} \frac{\partial^2 \xi(x_1, x_2)}{\partial x_i^2} + \ldots, \tag{9}$$

where $m = 1, 2, \ldots$ and the range of nonlocal actions is controlled by a characteristic length c. Differentiating (9) twice and substituting back into Eq. 9 and truncating the series for derivatives of order four and higher, an implicit formulation is obtained in the form of a modified inhomogeneous Helmholtz equation:

$$\overline{\xi}(x_1, x_2) - c^2 \nabla^2 \overline{\xi}(x_1, x_2) = \xi(x_1, x_2) \text{ in } \Omega, \tag{10}$$

where ∇^2 is the Laplacian operator. In Eq. 10, derivatives of the forth order and higher have been neglected. One observes that an advantage of the implicit Eq. 10, compared to the explicit Eq. 9, is that Eq. 10 limits singularities in the local field, while Eq. 9 amplifies them, hence behaving like the experimental result in Fig. 2. It has been argued that in the context of damage models, a natural boundary condition to Eq. 10 according to vanishing normal derivatives on the boundaries, i.e., $n\nabla\overline{\xi} = 0$ (n is the normal to the boundary), is the only physically acceptable one. Moreover, in the context of continuum damage, there exist several variables that can be spatially averaged to obtain continuous development of damage. The model originally proposed by Pijaudier-Cabot and Bazant (1987) averages the damage energy release rate ψ according to

$$\left(1 - c^2 \nabla^2\right) = \psi / \overline{\psi} \text{ and the natural boundary condition } n\nabla\overline{\psi} = 0. \tag{11}$$

The implicit formulation (11) can fairly easily be implemented in a conventional finite element code. The theory provides good computational economy and can be used to study failure of complex macroscopic structures, given an accurate gradient length c is used.

Example A: Uniaxial Tension of a Rod

The fundamental properties of continuum damage mechanics combined with gradient enhancement are demonstrated in a uniaxial setting. The governing equations are solved numerically, e.g., Isaksson et al. (2004). To initiate failure, an imperfection in stiffness of 1 % is assigned the center point. The model has one elastic parameter (E), two parameters characterizing onset and evolution of damage

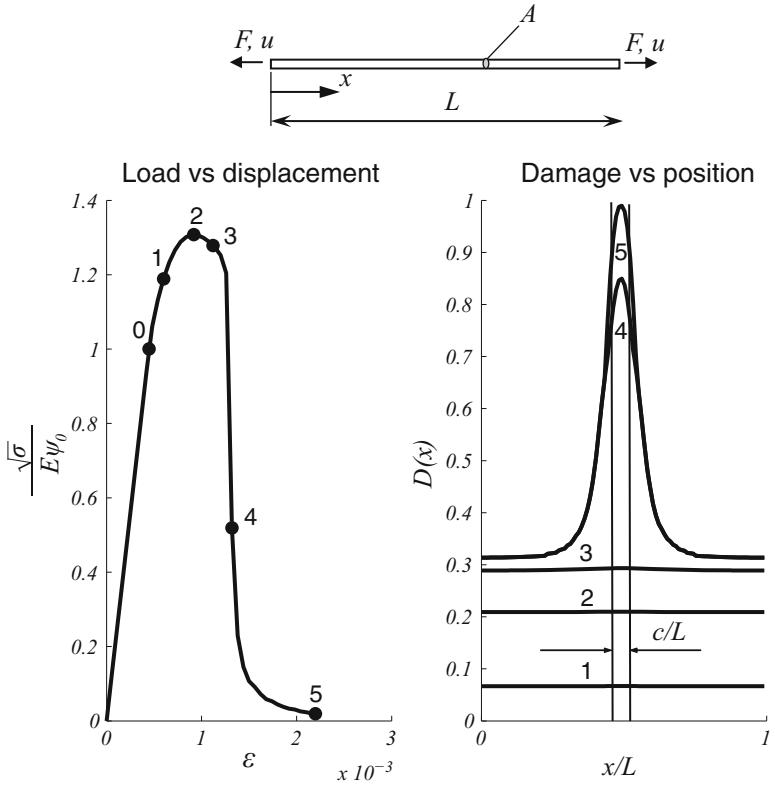


Fig. 10 Uniaxial example of nonlocal continuum damage ($\sigma = F/A$ and $\varepsilon = u/L$). Here, $E/\psi_0 = 5 \cdot 10^6$, $\beta = 0.1$, and $c/L = 1/28$

(ψ_0, β), and one gradient parameter (c) controlling the range of nonlocal actions. Figure 10 shows the computed stress–strain response and damage profile corresponding to six load levels during monotonic loading. The load cycle can be divided into three stages. Below load level 0, the body exhibits a purely linear elastic behavior. At a load level 0, the stress is given by $\sigma = [\psi_0 E]^{1/2}$ and there is onset of damage growth. Between load level 0 and load level 2, the damage in the body develops in an approximately homogenous manner and the damage growth is controlled by the parameter β. At load level 2, the maximum load is reached and the damage localizes to the center of the bar as softening begins, and the characteristic length c controls the width of the localized zone.

Example B: Analyzing Uniaxial Tension of Paper

A tissue-type paper with an open network structure is considered. The material is analyzed in the cross-machine direction, i.e., slender samples are cut perpendicular

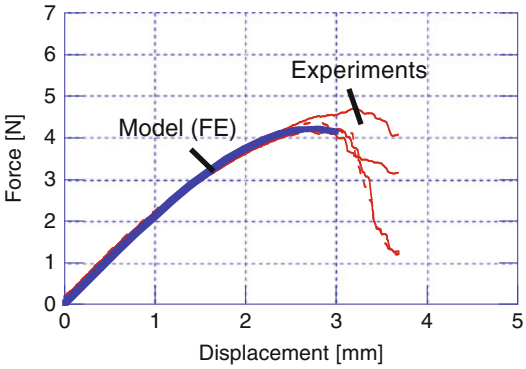


Fig. 11 Numerical model (FEM) fitted to uniaxial load–displacement curves for tissue paper. The width of the specimens is 30 mm and the gauge length 100 mm (Reprinted from Hägglund and Isaksson (2006), with permission from Elsevier)

to the manufacturing direction. The isotropic elastic modulus E of the material is determined from the initial slope of the force–displacement curves and was estimated to $E = 6.3$ kN/m. The damage evolution parameter β and the damage threshold ψ_0 were determined by fitting the constitutive model to the post-elastic regime of the load–displacement curve in Fig. 11. The calibration procedure gives $\psi_0/E = 3.2 \times 10^{-5}$ and $\beta = 4 \times 10^{-3}$.

Sample Applications: Mode I Fracture Paper Using CDM

In this section, the continuum damage model described in section "Continuum Damage Model" is used for fracture analysis of paper material. A tensile-loaded rectangular specimen consisting of low-basis-weight paper containing a prefabricated slit is examined (Fig. 12). The dimension of the specimen is given by the size of the crack ($2a$), length ($2h$), and width of the specimen ($2w$). A state of plane stress is assumed.

The principal directions of material symmetry of paper are defined as the machine direction (MD), which is the direction of manufacture; cross direction (CD), which is the transverse web direction; and thickness direction (Z). Specimens loaded in CD are considered. The material is assumed to be relatively brittle so that plastic strains are neglected. The effect of a notch size on fracture load is investigated. Because use is made of an orthotropic material description to describe the elastic response of the considered material, the model parameters are estimated from tensile tests in both CD and MD. Here index 1 refers to CD and index 2 to MD. The model requires five elastic parameters (E_1, E_2, G_{12}, ν_{12}, ν_{21}) that describe the elastic behavior, two parameters (ψ_0, β) that characterize the (isotropic) damage evolution, and a gradient parameter (c) that controls the range of nonlocal mechanisms. Macroscopic crack growth is assumed to occur when the damage variable D of a material point approaches one.

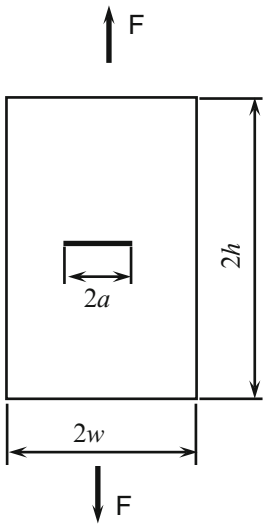


Fig. 12 Geometry of fracture specimen

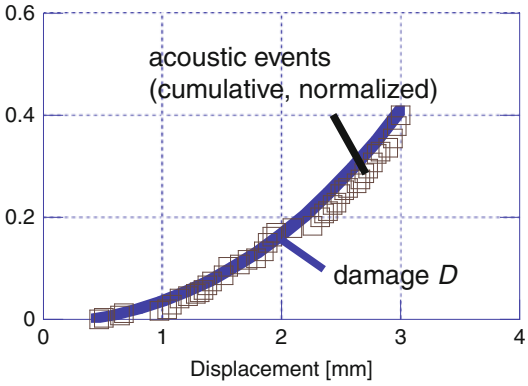


Fig. 13 Fitting of damage-related constitutive parameters. The size of the specimen was 100 × 30 mm (Fig. 11) (Reprinted from Hägglund and Isaksson (2006), with permission from Elsevier)

Calibration of the Constitutive Model

The constitutive model is calibrated using data from acoustic emission (AE) monitored uniaxial tensile tests (Fig. 13). Work carried out (among others, Salminen et al. 2003; Yamauchi 2004) has shown that the way paper fractures in tensile experiments can be followed by acoustic sensors. The basic principle in AE is that during the fracturing of a fiber-to-fiber bond (which will occur in the order of microseconds), stresses in the neighborhood of the fracture site will be redistributed and cause a rapid release of elastic energy, which can be recorded on the surface of the material using an appropriate acoustic emission sensor. In this way each microfracture is the source of one acoustic emission event. Eventually, the microfracture density becomes so severe as the material collapse is measurable in tensile tests as a reduction of the stiffness. The AE sensor was positioned in the

center point of each specimen and attached to the samples with a magnetic holder placed on the alternate side of the paper specimen.

The elastic modulus in the principal material directions of the material was determined from the initial slope of the load–displacement curves, where $E_1 = 6.3$ kN/m and $E_2 = \frac{1}{2}E_1$. The damage hardening parameter β is determined by fitting the constitutive model to the post-elastic regime of the load–displacement curve when loaded in CD. Displayed in Fig. 11 are four experimentally obtained load–displacement curves. The damage threshold ψ_0 was obtained from onset of acoustic emission (AE). The calibration procedure gives $\psi_0/E = 3.2 \times 10^{-5}$ and $\beta = 4 \times 10^{-3}$ (section "Example B: Analyzing Uniaxial Tension of Paper"). The in-plane Poisson's ratios were set to $\nu_{12} = 0.45$ and $\nu_{21} = 0.22$, while the in-plane shear modulus was assigned the value $G_{12} = 0.4\sqrt{E_1 E_2}$ following Baum et al. (1981). Notably is that the acoustic emission results support the use of an exponential damage evolution law for low-basis weight paper (Eqs. 5 and 6).

Results

The model is solved numerically in a nonlinear finite element procedure and is compared with data from fracture test (Hägglund and Isaksson 2006). In the fracture testing load–displacement is measured using specimens having a width and gauge length of 150 mm ($2w$) and 230 mm ($2h$), respectively, at a deformation rate of 1.6 mm/min. The specimens are tested in CD only. The prefabricated crack is manually cut in the center of the specimen oriented in MD to a length ($2a$) ranging from 5 to 70 mm.

Figure 14 shows a comparison between the model and experiments for notched specimens. The 95 % confidence limit is presented along with the average value of eight specimens for each notch length. The computed results of fracture load (i.e., maximum load) are presented for three different values of c, i.e., the characteristic length parameter. In Fig. 14, the length of the slit (a) has been normalized with respect to the width of the specimen and fracture load against the tensile strength of the material (i.e., the peak load of an unnotched specimen). Obviously, as the stress field is approximately homogenous for an unnotched specimen, the influence of

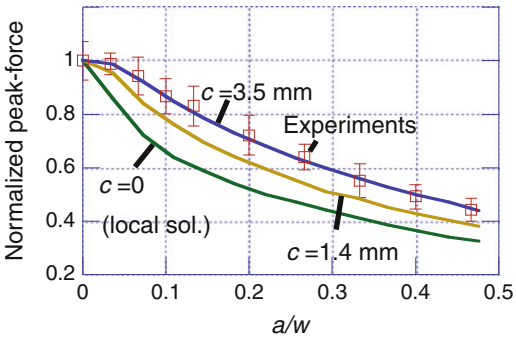


Fig. 14 Fracture test results (Reprinted from Hägglund and Isaksson (2006), with permission from Elsevier)

c vanishes. According to the results, slits smaller than 10–15 mm has marginal influence on the fracture load. Thus, for the case of small defects, failure may localize to other inherent weak spots in the sheet. In dense paper networks, a relatively small initial crack may localize fracture while an open network structure reduces the strain gradients ahead of the defect, and a large crack is required to localize fracture (Isaksson and Hägglund 2009; Coffin and Li 2011). Hence, the model is able to capture the phenomenon shown in Fig. 2.

Crack-Tip Fields

It may be of interest to analyze the strains and stresses close to a crack tip in an RFN material. In Isaksson and Hägglund (2013), crack-tip stress and strain fields in gradient-enhanced elastic materials are derived on closed form. The special case of an opening mode I crack is analyzed. Physical requirements of finite stresses and strains at infinity and at the tip are applied to remove singularities, following Altan and Aifantis (1997) and Aifantis (2011). A Cartesian (x_1, x_2) and a polar coordinate system ($r = [x_1^2 + x_2^2]^{1/2}$, $\theta = \tan^{-1}[x_2/x_1]$) are introduced with their origins coinciding with the crack tip. The crack occupies the negative part of the x_1-axis, i.e., $x_1 < 0$ and $x_2 = 0$, and Ω is the infinite domain containing the crack. Distant from the crack tip, a pure mode I opening field acts and the macroscopic stress tensor σ_{ij} is given by

$$\sigma_{ij} = \frac{K_\mathrm{I}}{\sqrt{2\pi r}} f_{ij}(\theta) \quad \text{as } r \to \infty, \tag{12}$$

where K_I is the LEFM mode I stress intensity factor and $f_{ij}(\theta)$ are angular functions. The gradient-enhanced stress tensor $\overline{\sigma}_{ij}$ is calculated using the LEFM stress tensors as source terms in the inhomogeneous Helmholtz Eq. 10, i.e.,

$$\overline{\sigma}_{ij} - c^2 \nabla^2 \overline{\sigma}_{ij} = \frac{K_\mathrm{I}}{\sqrt{2\pi r}} f_{ij}(\theta), \tag{13}$$

which has the solutions

$$\begin{aligned}
\overline{\sigma}_{11} &= \frac{K_\mathrm{I}}{\sqrt{2\pi r}} \left[\frac{3}{4} \cos\frac{\theta}{2} \left[1 - e^{-r/c} \right] + \frac{1}{4} \cos\frac{5\theta}{2} \left[1 - 6c^2/r^2 + 2e^{-r/c} \left(3c^2/r^2 + 3c/r + 1 \right) \right] \right] \\
\overline{\sigma}_{22} &= \frac{K_\mathrm{I}}{\sqrt{2\pi r}} \left[\frac{5}{4} \cos\frac{\theta}{2} \left[1 - e^{-r/c} \right] - \frac{1}{4} \cos\frac{5\theta}{2} \left[1 - 6c^2/r^2 + 2e^{-r/c} \left(3c^2/r^2 + 3c/r + 1 \right) \right] \right] \\
\overline{\sigma}_{12} &= \frac{K_\mathrm{I}}{\sqrt{2\pi r}} \left[\frac{-1}{4} \sin\frac{\theta}{2} \left[1 - e^{-r/c} \right] + \frac{1}{4} \sin\frac{5\theta}{2} \left[1 - 6c^2/r^2 + 2e^{-r/c} \left(3c^2/r^2 + 3c/r + 1 \right) \right] \right]
\end{aligned} \tag{14}$$

It is seen in Eq. 14 that the normal stresses are vanishing on the crack surfaces while the shear stress is nonzero in a small region behind the tip. Consequently, the

condition of vanishing surface shear stress is fulfilled only in a *strong* approximate sense. The gradient-enhanced elastic strain tensor $\overline{\varepsilon}_{ij}$ is readily given by Hooke's generalized law $\overline{\varepsilon}_{ij} = \left[\overline{\sigma}_{ij} - \overline{\sigma}_{kk}\delta_{ij}\nu/(1+\nu)\right]/E$, where E is Young's modulus, ν is Poisson's ratio, and δ_{ij} is Kronecker's delta. For the sake of completeness, they are written here:

$$\begin{aligned}
\overline{\varepsilon}_{11} &= \frac{K_{\mathrm{I}}}{E\sqrt{2\pi r}}\left[\frac{3-5\nu}{4}\cos\frac{\theta}{2}\left[1-e^{-r/c}\right]+\frac{1+\nu}{4}\cos\frac{5\theta}{2}\left[1-6c^2/r^2+2e^{-r/c}\left(3c^2/r^2+3c/r+1\right)\right]\right]\\
\overline{\varepsilon}_{22} &= \frac{K_{\mathrm{I}}}{E\sqrt{2\pi r}}\left[\frac{5-3\nu}{4}\cos\frac{\theta}{2}\left[1-e^{-r/c}\right]-\frac{1+\nu}{4}\cos\frac{5\theta}{2}\left[1-6c^2/r^2+2e^{-r/c}\left(3c^2/r^2+3c/r+1\right)\right]\right]\\
\overline{\varepsilon}_{12} &= \frac{K_{\mathrm{I}}}{E\sqrt{2\pi r}}\left[-\frac{1+\nu}{4}\sin\frac{\theta}{2}\left[1-e^{-r/c}\right]+\frac{1+\nu}{4}\sin\frac{5\theta}{2}\left[1-6c^2/r^2+2e^{-r/c}\left(3c^2/r^2+3c/r+1\right)\right]\right]
\end{aligned} \tag{15}$$

Expressions for stresses and strains in polar coordinates are obtained using conventional transformation rules. Interestingly, the solutions deviate slightly from Eringen's original nonlocal theory. The two normal stresses on the crack surfaces are vanishing in Eq. 14, while Eringen's approximation gives nonvanishing normal stress on the crack surfaces close behind the tip. Further, at distances $r/c > 2$ the gradient-enhanced stress is approximately equal to the classical LEFM stress in the whole body, a behavior that is the same for all stress components. The stresses (and strains) are finite and the position of maximum hoop stress is located close to the tip. Moreover, a numerical analysis reveals that the maximum hoop stress $\overline{\sigma}_{\theta\theta}$ is equal to $K_{\mathrm{I}}[5/2 - 19\mathrm{e}^{-1}/4]/\sqrt{2\pi c} \approx 0.30K_{\mathrm{I}}/\sqrt{c}$ and is situated on the crack plane at the distance $r/c \approx 1.1$. Notably is that the maximum hoop stress in the near-tip region is substantially lower than in Eringen's theory. Most importantly, the obtained strain field (Eq. 14) resembles experimental results obtained in fracture experiments on fiber materials (Fig. 8).

Concluding Remarks

The mechanical behavior of fiber networks is governed by complex multiple mechanisms. This chapter presents a framework for analyzing deformation, damage, and fracture in network materials using continuum damage mechanics. The fundamental idea is that material degradation, or damage, is described in a diffuse sense (continuum assumption) and the influence of damage on the mechanical properties is governed by an internal variable. A classical continuum description of this class of material is too simple to capture the essential mechanical behavior. To correctly describe gradients in strain and damage, a nonlocal field theory (gradient theory) must be added to the framework. The reason for this is the microstructure in the material. Firstly, there is defect-size dependence in the heterogeneous fiber network that is not captured by classical methods. Secondly, since the network consists of relatively stiff fibers connected to other fibers located at remote distances, the fibers introduce long-range microstructural effects in the

material. Thus, at any point in the network, mechanical nonlocal actions are experienced that have to be captured. A nonlocal field theory substantially improves estimates of deformation fields and fracture loads in fiber-based network materials as compared to classical mechanical theories. For ductile fiber networks (i.e., where fibers exhibit significant plastic straining upon loading), the theory may be extended by appropriate theories of plasticity.

References

E.C. Aifantis, A note on gradient elasticity and nonsingular crack fields. J. Mech. Behav. Mater. **20**, 103–105 (2011)

S. Allaoui, Z. Aboura, M.L. Benzeggagh, Phenomena governing uni-axial tensile behaviour of paperboard and corrugated cardboard. Comp. Struct. **87**, 80–92 (2008)

B.S. Altan, E.C. Aifantis, On some aspects in the special theory of gradient elasticity. J. Mech. Behav. Mater. **8**(3), 231–282 (1997)

P.E. Åslund, P. Isaksson, A note on the nonlinear mechanical behavior of planar random network structures subjected to in-plane compression. Comp. Mater. **45**, 2697–2703 (2011)

J. Åström, K. Niskanen, Simulation of network fracture, in *Proceedings of the 1991 International Paper Physics Conference* (TAPPI, Espoo Finland, 1991), pp. 31–47

J. Åström, K. Niskanen, Symmetry-breaking fracture in random fiber networks. Europhys. Lett. **21**, 557–562 (1993)

G.A. Baum, D.C. Brennan, C.C. Habeger, Orthotropic elastic constants of paper. Tappi **64**, 97–101 (1981)

C.A. Bronkhorst, Modeling paper as a two-dimensional elastic–plastic stochastic network. Int. J. Solids Struct. **40**(20), 5441–5454 (2003)

D.W. Coffin, Use of the efficiency factor to account for previous straining on the tensile behavior of paper. Nord. Pulp Pap. Res. J. **27**(2), 305–312 (2012)

D. W. Coffin, K. Li, On the fracture behavior of paper, in *Proceedings: Progress in Paper Physics*, 5–8 Sept, Graz, 2011

H.L. Cox, The elasticity and strength of paper and other fibrous materials. Br. J. Appl. Phys. **3**, 72–79 (1952)

A. Delaplace, G. Pijaudier-Cabot, S. Roux, Progressive damage in discrete models and consequences on continuum modelling. J. Mech. Phys. Solids **44**(1), 99–136 (1996)

S.J. Eichhorn, A. Dufresne, M. Aranguren, N.E. Marcovich, J.R. Capadona, S.J. Rowan, C. Weder, W. Thielemans, M. Roman, S. Renneckar, W. Gindl, S. Veigel, J. Keckes, H. Yano, K. Abe, M. Nogi, A.N. Nakagaito, A. Mangalam, J. Simonsen, A.S. Benight, A. Bismarck, L.A. Berglund, T. Peijs, Review: current international research into cellulose nanofibres and nanocomposites. J. Mater. Sci. **45**, 1–33 (2010)

A.C. Eringen, *Nonlocal Continuum Field Theories* (Springer, New York, 2002)

A.C. Eringen, D.G.B. Edelen, On nonlocal elasticity. Int. J. Eng. Sci. **10**, 233–248 (1972)

C. Fellers, A. de Ruvo, J. Elfström, M. Htun, Edgewise compression properties. A comparison of handsheets made from pulps of various yields. Tappi J. **63**(6), 109–112 (1980)

R. Hägglund, P. Isaksson, Analysis of localized failure in low-basis weight paper. Int. J. Solids Struct. **43**, 5581–5592 (2006)

R. Hägglund, P. Isaksson, On the coupling between macroscopic material degradation and interfiber bond fracture in an idealized fiber network. Int. J. Solids Struct. **45**, 868–878 (2007)

M. Henriksson, L.A. Berglund, P. Isaksson, T. Lindström, T. Nishino, Cellulose nanopaper structures of high toughness. Biomacromolecules **9**(6), 1579–1585 (2008)

H.J. Herrmann, S. Roux (eds.), *Statistical Models for the Fracture of Disordered Media* (North Holland, Amsterdam, 1990)

S. Heyden, Network modelling for the evaluation of mechanical properties of cellulose fiber fluff. Ph.D.-thesis, Lund University, Sweden, 2000
S. Heyden, P.J. Gustafsson, Simulation of fracture in a cellulose fiber network. J. Pulp Pap. Sci. **24**, 160–165 (1998)
P. Isaksson, R. Hägglund, Structural effects on deformation and fracture of random fiber networks and consequences on continuum models. Int. J. Solids Struct. **46**, 2320–2329 (2009)
P. Isaksson, R. Hägglund, Crack-tip fields in gradient enhanced elasticity. Eng. Fract. Mech. **97**, 186–192 (2013)
P. Isaksson, R. Hägglund, P. Gradin, Continuum damage mechanics applied to paper. Int. J. Solids Struct. **41**, 4731–4755 (2004)
A. Jangmalm, S. Östlund, Modelling of curled fibres in two-dimensional networks. Nord. Pulp Pap. Res. J. **10**, 156–161 (1995)
C.A. Jentzen, The effect of stress applied during drying on some of the properties of individual pulp fibers. Tappi **47**(7), 412–418 (1964)
L.M. Kachanov, Time of the rupture process under creep condition. Izv. Akad. Nauk SSSR, Otd. Tekhn. Nauk, 26–31, 1958 (in Russian)
O. Kallmes, H. Corte, The structure of paper. I. The statistical geometry of an ideal two-dimensional fibre network. Tappi **43**(9), 737–752 (1960)
M.J. Korteoja, A. Lukkarinen, K. Kaski, D.J. Gunderson, J.L. Dahlke, K.J. Niskanen, Local strain fields in paper. Tappi J. **79**(4), 217–223 (1996)
E. Kröner, Elasticity theory of materials with long range cohesive forces. Int. J. Solids Struct. **3**, 731–742 (1967)
J. Lemaitre, J.L. Chaboche, *Mechanics of Solid Materials* (Cambridge University Press, Cambridge, UK, 1990)
P. Luner, A.E.U. Karna, C.P. Donofrio, Studies in interfibre bonding of paper. The use of optical bonded area with high yield pulps. Tappi J. **44**(6), 409–414 (1961)
K.J. Niskanen, *Strength and fracture of paper* (KCL Paper Science Centre, Espoo, 1993)
K. Niskanen, H. Kettunen, Y. Yu, Damage width: a measure of the size of fracture process zone. In: *12th Fundamental Research Symposium*, Oxford, UK, 2001
R.H.J. Peerlings, R. de Borst, W.A.M. Brekelmans, J.H.P. de Vree, Gradient enhanced damage for quasi-brittle materials. Int. J. Numer. Met. Eng. **39**, 3391–3403 (1996)
R.H.J. Peerlings, M.G.D. Geers, R. de Borst, W.A.M. Brekelmans, A critical comparison of nonlocal and gradient-enhanced softening continua. Int. J. Solids Struct. **38**(44–45), 7723–7746 (2001)
G. Pijaudier-Cabot, Z.P. Bazant, Nonlocal damage theory. J. Eng. Mech. **113**, 1512–1533 (1987)
M.K. Ramasubramanian, R.W. Perkins, Computer simulation of the uniaxial elastic–plastic behavior of paper. ASME J. Eng. Mater. Tech. **110**(2), 117–123 (1987)
A. Ridruejo, C. González, J. LLorca, Damage micromechanisms and notch sensitivity of glass-fiber non-woven felts: an experimental and numerical study. J. Mech. Phys. Solids **58**, 1628–1645 (2010)
I.B. Sachs, T.A. Kuster, Edgewise compression failure mechanism of linerboard observed in a dynamic mode. Tappi J. **63**, 69 (1980)
L.I. Salminen, A.I. Tolvanen, M.J. Alava, Acoustic emission from paper fracture. Phys. Rev. Lett. **89**, 185503 (2003)
S.A. Silling, Reformulation of elasticity theory for discontinuities and long-range forces. J. Mech. Phys. Solids **48**, 175–209 (2000)
J.A. Van den Akker, Some theoretical considerations on the mechanical properties of fibrous structures, in *The Formation and Structure of Paper,* ed. by F. Bolam. Technical Section British Paper and Board Makers Association, London (1962), pp. 205–241
T. Yamauchi, Effect of notches on micro failures during tensile straining of paper. Jpn. Tappi J. **58**(11), 105–112 (2004)

Predicting Damage Evolution in Composites with Explicit Representation of Discrete Damage Modes

42

Q. D. Yang and B. C. Do

Contents

Abstract

Polymer matrix composites (PMCs) are playing rapidly increasing roles in future military and civilian industries. Damage tolerance analysis is an integral part of PMC structural design. Considerable research efforts have been invested to establish predictive capabilities, but thus far high-fidelity strength and durability prediction capabilities are yet to be established. Advanced numerical methods that can explicitly resolve the multiple-damage processes and their nonlinear coupling at various scales are highly desired. This paper first reviews the recent development of advanced numerical methods, including eXtended Finite

Q.D. Yang (✉) • B.C. Do
Department of Mechanical and Aerospace Engineering, University of Miami, Coral Gables, FL, USA
e-mail: qdyang@miami.edu; b.do1@umiami.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_16

Element Method (X-FEM), phantom node methods (PNM), and the Augmented Finite Element Method (A-FEM), in handling the multiple-damage coupling in composites. The capability of these methods in representing various composite damage modes explicitly with embedded nonlinear fracture models (such as cohesive zone models) makes them excellent candidates for high-fidelity failure analyses of composites. The detailed formulation of A-FEM and its implementation to a popular commercial software package (ABAQUS) as a user-defined element has been given. Successful simulations of composites at various scales using the framework of A-FEM are presented and the numerical and material issues associated with these high-fidelity analyses are discussed. Through the numerical predictions and the direct comparisons to experimental results, it has been demonstrated that high-fidelity failure analyses can be achieved with the A-FEM through careful calibration of nonlinear material properties and cohesive fracture parameters and with proper considerations of the different length scales within which these damage processes operate.

Introduction

The past decades have witnessed an exponential increase in composite usage and the trend is still continuing. Composite materials have been successfully inserted into new commercial and military airplanes (e.g., Boeing 787 Dreamliners, Airbus A380, and Joint Strike Fighters of US Air Force) and to the rapidly expanding renewable energy industries (e.g., wind turbine blades). Compared to structural metals, composites offer several key advantages such as significant weight savings, better reliability and durability, and lower manufacturing and maintenance costs. In particular, composites offer the exciting advantage of designing hierarchical features at micro- or mesoscale to achieve the desired structural performance through the optimization process (Cox and Yang 2006; LLorca and González 2011; Yang et al. 2011).

Unlike structural metals which are homogeneous and isotropic, composites are inherently inhomogeneous and anisotropic. Microscopic flaws and imperfections are inevitably present in such structures due to manufacturing processes. Therefore, damage tolerance design has to be an integral part of composite structural design. Currently such design relies heavily on large experimental test programs that are extremely costly and time consuming. The analytical capability for composite design currently consists of empirical fits to large experimental databases from which "knockdown" factors for static design and stress–life curves or similar aids for "lifing" are determined. The burden of testing to establish design allowables is immense: a typical large airframe, for example, currently requires $\sim 10^4$ tests of material specimens, along with tests of components and structures up to the entire tails, wing boxes, and fuselages, to achieve safety certification (Fawcett et al. 1997), and a single bearing fatigue test can cost up to $50,000. Such lengthy and costly procedures have been increasingly cited as the bottleneck to fast insertion of new composites in engineering applications.

There is a high demand in developing high-fidelity simulation capabilities that can offer quick and accurate strength and durability predictions based on constituent materials properties, so that many (but not all) of the composite tests can be partially replaced with high-fidelity simulations, as has been the case in automotive industry. There have been several calls for simulation-based engineering science (SBES) from various government agencies recently (Oden et al. 2006; Dowlbow et al. 2004). However, although computational stress analyses have been routine exercises in composite structural design, predicting the progressive damage process prior to final failure remains a very difficult task, owing to the highly heterogeneous nature of composites. Many of the challenging issues are related to the difficulty of developing practicable formulations for dealing with materials containing complex material heterogeneity at various scales. Heterogeneity in composites poses special problems for accurate prediction of local stress and strain fields, which can vary strongly with local material features, and with predicting cracks and localized damage bands, which can appear during damage evolution not only on the material boundaries but also on other surfaces that cannot be specified a priori.

Such highly heterogeneous materials present unique challenges that cannot be resolved by mainstream formulations of conventional material/structural modeling which are based on hierarchical homogenization through representative material volumes (Oskay and Fish 2007; Ramanathan et al. 1997). One challenge is that the scale of material heterogeneity is on the same scale as that of the features of the structures, which negates the common strategy of homogenizing the material properties in simulations. In the past, many strength criteria based on various homogenization theories have been proposed, including Hashin's mechanism-based criterion and the widely used Tsai–Wu criterion. The success of these phenomenological criteria in assessing composite strength has been limited. A recent report on blind failure prediction of unidirectional polymer matrix composites (PMCs) revealed rather disturbing predictive capabilities of such strength criteria: virtually none of these homogenized composite theories can predict laminate strength satisfactorily under general loading conditions. Discrepancies among different theoretical predictions are more than 100 % and up to 1,500 % in extreme cases (Kaddorur et al. 2004).

Continuum damage mechanics (CDM) has also been used to account for the progressive damage in composites (McCartney 2003; Chang et al. 1991; Shokrieh and Lessard 2000; Chaboche et al. 1995). This approach considers the gradual reduction of load-bearing capability due to damage by introducing a damage parameter (or a tensor) into the material constitutive relation, which leads to irreversible damage evolution (Chang et al. 1991; Shokrieh and Lessard 2000; Chaboche et al. 1995, 1997; de Borst et al. 1995). The CDM-based theories can be conveniently formulated into traditional finite element programs. But there are two inherent difficulties associated with this approach: (1) it cannot deal with highly concentrated crack-like damage, which is universal in composites, and (2) damage parameters calibrated from individual modes do not account for multiple-damage mode coupling (Talreja 2006; Van de Meer and Sluys 2009a). Linear elastic fracture mechanics (LEFM) and its computational form virtual crack closure

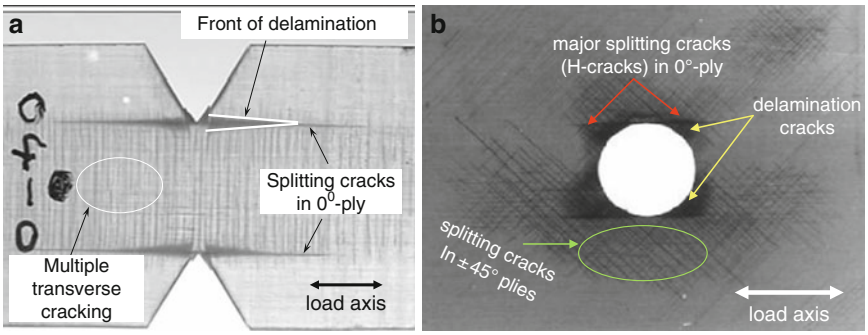


Fig. 1 X-ray radiography reveals damage mechanisms viewed through the ply stack in (**a**) a double-notch tension specimen with symmetric orthogonal ply stack $[0/90]_s$ and (**b**) a quasi-isotropic laminate $[-45/+45/90/0]_s$ with a circular open hole under tensile loading (Reprinted from Journal of the Mechanics and Physics of Solids, vol 59, X.J. Fang, Z.Q. Zhou, B.N. Cox, Q.D. Yang, *High-fidelity simulations of multiple fracture processes in a laminated composite in tension*, Pages No. 1355–1373, Copyright 2011, with permission from Elsevier)

technique (VCCT) have been used to study delamination cracks in laminated composites. But this method requires preexisting cracks of finite size and therefore cannot deal with damage initiation from pristine materials or from traction-free composite edges (Tay 2003).

Multiscale hierarchical modeling approach based on unit cell (UC) or representative volume element (RVE) analyses and followed by homogenization from one scale to next higher level scale (hierarchical model) has also been pursued to account for the progressive damage evolution on macroscopic composite properties (Oden et al. 1999; Tang et al. 2006; Inglis et al. 2007; Fish and Ghouali 2001; Reddy 2005; Gonzalez and LLorca 2006; Ladeveze 2004). But without explicit inclusion of the multiple-damage coupling, the homogenization process will have to rely on either theoretical hypothesis or costly experimental programs to calibrate key parameters (and in composites there are many of them). This point is illustrated through the progressive damage evolution recorded for two composites shown in Fig. 1.

Figure 1a shows multiple-damage modes in a double-notched tension specimen with symmetric $[0/90]_s$ ply stack. Dominant splitting cracks in the 0° ply appear as sharply defined horizontal lines (in an H configuration) and eventually span the specimen. Many transverse cracks in the 90° ply occur during the load increase. In addition, the major splitting cracks are accompanied by wedge-shaped delaminations between the plies (areas of shadow around the splitting cracks). Figure 1b shows the multiple-cracking features in a quasi-isotropic laminate $[-45/+45/90/0]_s$ with a central open hole (Case and Reifsnider, 1999). The splitting cracks are shorter and the delaminations are lobe shaped. Transverse cracking occurs predominantly in the ±45° plies. The intraply cracks, i.e., the splitting cracks in 90° ply and off-axial cracks in +/−45° plies, are located in different plies and are coupled

through local inter-ply delamination (dark lobed regions) at inter-ply interfaces. How to account for these multiple-damage mechanisms with strong interactions in an accurate and computationally efficient scheme remains a difficult task.

Traditionally, the intraply and inter-ply damage processes were treated separately with different theories: the delamination problems were extensively analyzed using either LEFM (Tay 2003; Shahwan and Waas 1997) or cohesive zone models (CZMs) (Yang and Cox 2005; Turon et al. 2006; Hallett and Wisnom 2006a; Wisnom and Chang 2000), while strength-based criteria coupled with CDM for strength degradation have been used for the intraply damage modes (Lapczyk and Hurtado 2007; Laurin et al. 2007; Maimi et al. 2007; Matzenmiller et al. 1995; Pinho et al. 2006). Direct coupling of CDM-based in-plane damage modes with various fracture mechanics models for delamination has also been attempted by several research groups but with only limited success (Cox and Yang 2006; Choi and Chang 1992; Yang and Cox 2005; Van de Meer and Sluys 2009a; Carpinteri and Ferro 2003). Several recent studies have shown that the homogenization process at the mesoscale in CDMs leads to loss of key information concerning multiple-damage coupling at the macroscopic scale and may result in inaccurate predictions of the crack path (Talreja 2006; Van de Meer and Sluys 2009a). Unexpected severe stress locking in numerical simulations may also occur (Iarve et al. 2005).

The inadequacy of traditional CDMs has led to a recent trend to integrate explicit representations of all major cracking events into global composite structure models to achieve direct coupling (Gonzalez and LLorca 2006; Van de Meer and Sluys 2009a, b; Iarve et al. 2005; Hallett and Wisnom 2006a; Rudraraju et al. 2010, 2011; Van de Meer et al. 2010; Fang et al. 2010, 2011a; Ling et al. 2009, 2011). The explicit consideration of multiple-damage modes and their nonlinear coupling during their evolution processes are enabled by two critical elements: (1) improved nonlinear fracture models such as the CZMs for bulk and interface crack problems to achieve unification of crack initiation and propagation and (2) improved numerical methods that allow for arbitrary locations of crack initiation and propagation in heterogeneous media.

The nonlinear fracture model of CZMs pioneered by Dugdale and Barenblatt (Barenblatt 1959, 1962; Dugdale 1960) and extended by Hillerborg et al (1976) for numerical applications has been widely used in the damage and fracture analyses of composites (Cox and Yang 2006; Turon et al. 2006; Wisnom and Chang 2000; Yang and Cox 2005; Hallett and Wisnom 2006a; Xie et al. 2006; Yang et al. 2006a, b; Song and Waas 1995; Thouless and Yang 2008; Yang and Thouless 2001a; Yang et al. 1999, 2001). In its original formulation, however, a CZM requires the potential crack path to be known *a priori*, so that CZM elements can be directly implanted along the path. This greatly limits the application of CZM for problems with evolving arbitrary discontinuities.

Recently several novel numerical methods have been developed to allow for arbitrary cohesive crack initiation and propagation without the need to specify crack paths *a priori*. The nonlinear cohesive laws are typically embedded in the numerical formulation in such a way that if under a given initiation criterion an

element is deemed to initiate a cohesive crack, the gradual separation of the severed subdomains is coupled by the cohesive stresses acting on the cracked surfaces. For example, cohesive models have been integrated into the extended finite element (X-FEM) framework and used successfully to model fracture in homogeneous quasi-brittle materials (Moës and Belytschko 2002; de Borst et al. 2006). However, when this method is applied to composites, it remains unclear how the nonlinear crack-tip stress field, which is necessary for nodal enrichment in the ply elements, is transferred to the interlaminar cohesive elements in a physically consistent way.

Another promising numerical method for handling arbitrary cracking problems is the augmented finite element method (A-FEM). This line of development follows the original work of Hansbo and Hansbo (Hansbo and Hansbo 2004), who first established that an arbitrary discontinuity can be introduced within an element by adding an extra element on top of the existing element, with each element accounting for the stiffness and force contribution from one part of the bisected physical domain. The two discontinuous domains are connected by linear or nonlinear springs (Hansbo and Hansbo 2004) or cohesive failure descriptions (Mergheim et al. 2005). The addition of elements is typically realized by introducing additional nodes that are geometrically identical to the original corner nodes (hence this method is also named the phantom node method; see Van de Meer and Sluys 2009b; Song et al. 2006). One major advantage of this method is that it uses only standard FE shape functions and thus avoids the use of the partition-of-unity method, as in X-FEM, which results in loss of elemental locality in the element that hosts a singular crack tip (Moes et al. 1999). It has been shown (Song et al. 2006) that this method is equivalent to the extended finite element method (XFEM) in which a discontinuity in the displacement field is introduced by enrichment of the shape functions with the Heaviside step function. This advantage renders the A-FEM completely compatible with standard FE programs. More recently, this method has been extended to account for the inevitable and important issue of material heterogeneity in composite materials (Ling et al. 2009).

This paper will review and discuss the recent development of advanced numerical methods that can potentially meet the challenge of faithfully simulating the coupled multiple damages at various scales in composite materials. The detailed formulation of the A-FEM and its implementation to commercial software packages such as ABAQUS as a user-defined element will be given. The enabling capabilities of the A-FEM in faithfully predicting the progressive damage evolution involving many damage modes will be demonstrated through numerical examples.

Nonlinear Fracture Models for Composite Materials

Most damage modes in composites are in the form of crack-like entities, and they interact with each other at various scales as evidenced in Fig. 1. When dealing with multiple-crack interactions, the nonlinear CZMs are essential because they unify

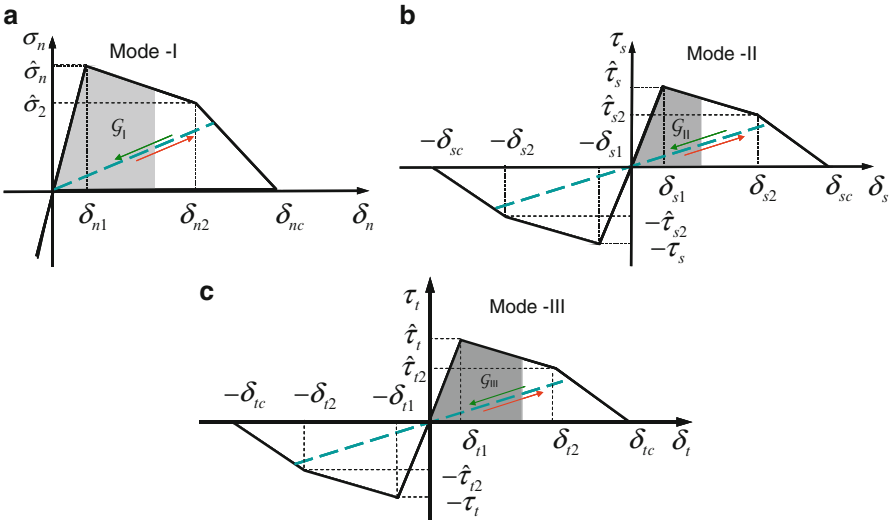


Fig. 2 A cohesive model with piece-wise linear cohesive laws for individual fracture modes

the processes of crack initiation and growth within a single physically consistent model (Barenblatt 1962; Dugdale 1960). A cohesive model is usually a phenomenological description of the mechanical behavior between the surface traction and separation (or displacement jump) across the crack surfaces. The key feature of such relation is that as a pair of cohesive surfaces separate due to the local stress environment, the cohesive stress increases initially, reaches a maximum (cohesive strength), and then drops gradually to zero at a critical separation displacement. This description captures the essence of the progressive failure in many engineering materials including PMCs. Furthermore, when such a CZM is embedded in a structural model, a fracture process zone (or cohesive zone) of finite size will develop which relates directly the microscopic material failure process to the macroscopic structural behaviors (Yang and Cox 2005; Camanho et al. 2003; Turon et al. 2007) (e.g., most fracture process zones in PMCs are of size 0.1–1.0 mm). This multiscale feature is the key to many successful applications of CZMs in composite analyses because it offers a clear guideline for many numerical issues such as mesh resolution.

This study will focus on the mode-independent cohesive law developed initially by Yang and Thouless (2001a) and later extended to 3D problems by Yang and Cox (2005). This cohesive law utilizes independent traction separations for opening (mode I) and shear (mode II and mode III) cracking, which are illustrated in Fig. 2.

The traction–separation law for each fracture mode can be written as

$$\sigma_{\mathrm{n}} = f_n(\delta_{\mathrm{n}}) = \begin{cases} \hat{\sigma}_n \cdot (\delta_{\mathrm{n}}/\delta_{\mathrm{n1}}) & (\text{if } \delta_{\mathrm{n}} \le \delta_{\mathrm{n1}}) \\ \hat{\sigma}_n - (\hat{\sigma}_n - \hat{\sigma}_2)(\delta_{\mathrm{n2}} - \delta_{\mathrm{n}})/(\delta_{\mathrm{n2}} - \delta_{\mathrm{n1}}) & (\text{if } \delta_{\mathrm{n1}} < \delta_{\mathrm{n}} \le \delta_{\mathrm{n2}}) \\ \hat{\sigma}_2 \cdot (\delta_{\mathrm{nc}} - \delta_{\mathrm{n}})/(\delta_{\mathrm{nc}} - \delta_{\mathrm{n2}}) & (\text{if } \delta_{\mathrm{n2}} < \delta_{\mathrm{n}} \le \delta_{\mathrm{nc}}) \\ 0 & (\text{if } \delta_{\mathrm{n}} > \delta_{\mathrm{nc}}) \end{cases}$$

$$\tau_{\mathrm{s}} = f_{\mathrm{s}}(\delta_{\mathrm{s}}) = \begin{cases} \hat{\tau}_s \cdot (\delta_{\mathrm{s}}/\delta_{\mathrm{s1}}) & (\text{if } |\delta_{\mathrm{s}}| \le \delta_{\mathrm{s1}}) \\ \mathrm{sgn}(\delta_{\mathrm{s}}) \cdot [\hat{\tau}_s - (\hat{\tau}_s - \hat{\tau}_{s2})(\delta_{\mathrm{s2}} - |\delta_{\mathrm{s}}|)/(\delta_{\mathrm{s2}} - \delta_{\mathrm{s1}})] & (\text{if } \delta_{\mathrm{s1}} < |\delta_{\mathrm{s}}| \le \delta_{\mathrm{s2}}) \\ \mathrm{sgn}(\delta_{\mathrm{s}}) \cdot \hat{\tau}_{s2} \cdot (\delta_{\mathrm{sc}} - |\delta_{\mathrm{s}}|)/(\delta_{\mathrm{sc}} - \delta_{\mathrm{s2}}) & (\text{if } \delta_{\mathrm{s2}} < |\delta_{\mathrm{s}}| \le \delta_{\mathrm{sc}}) \\ 0 & (\text{if } |\delta_{\mathrm{s}}| > \delta_{\mathrm{sc}}) \end{cases}$$

$$\tau_{\mathrm{t}} = f_{\mathrm{t}}(\delta_{\mathrm{t}}) = \begin{cases} \hat{\tau}_t \cdot (\delta_{\mathrm{t}}/\delta_{t1}) & (\text{if } |\delta_{\mathrm{t}}| \le \delta_{\mathrm{t1}}) \\ \mathrm{sgn}(\delta_{\mathrm{t}}) \cdot [\hat{\tau}_t - (\hat{\tau}_t - \hat{\tau}_{t2})(\delta_{\mathrm{t2}} - |\delta_{\mathrm{t}}|)/(\delta_{\mathrm{t2}} - \delta_{\mathrm{t1}})] & (\text{if } \delta_{\mathrm{t1}} < |\delta_{\mathrm{t}}| \le \delta_{\mathrm{t2}}) \\ \mathrm{sgn}(\delta_{\mathrm{t}}) \cdot \hat{\tau}_{t2} \cdot (\delta_{\mathrm{tc}} - |\delta_{\mathrm{t}}|)/(\delta_{\mathrm{tc}} - \delta_{\mathrm{t2}}) & (\text{if } \delta_{\mathrm{t2}} < |\delta_{\mathrm{t}}| \le \delta_{\mathrm{tc}}) \\ 0 & (\text{if } |\delta_{\mathrm{t}}| > \delta_{\mathrm{tc}}) \end{cases} \tag{1}$$

where $\mathrm{sgn}(\cdot)$ is a sign function and $|\cdot|$ denotes absolute value; σ_{n}, τ_{s}, and τ_{t} are normal and tangential tractions along the cohesive crack; δ_{n}, δ_{s}, and δ_{t} are normal and tangential displacement jumps across the crack line measured in the local coordinates as shown in Fig. 2; δ_{nc}, δ_{sc}, and δ_{tc} are critical normal and tangential displacements under pure modes beyond which the cohesive stresses become zero, indicating complete fracture; and $\hat{\sigma}_n$, $\hat{\tau}_s$, and $\hat{\tau}_{\mathrm{t}}$ are cohesive strengths for mode-I, mode-II, and mode-III fracture, respectively. $\hat{\sigma}_{n2}$, $\hat{\tau}_{s2}$, and $\hat{\tau}_{\mathrm{t2}}$ are secondary strengths for mode-I, mode-II, and mode-III fracture, respectively, which, together with the two intermedium displacements of each mode (δ_{n1} and δ_{n2} for mode I, δ_{s1} and δ_{s2} for mode II, and δ_{t1} and δ_{t2} for mode III), can be accommodated to achieve different shapes of the cohesive laws (e.g., with softening, hardening, or the trapezoidal shapes). But they are in general of secondary importance (Yang and Thouless 2001a; Yang et al. 1999; Yang et al. 2001; Kafkalidis et al. 2000) as compared to the cohesive strengths and toughnesses.

A mixed-mode cohesive model can be constructed by recognizing that the total traction–separation work absorbed during fracture, $\mathcal{G}$, can be separated into the opening (mode I) and shear (mode II and mode III) components, $\mathcal{G}_{\mathrm{I}}$, $\mathcal{G}_{\mathrm{II}}$, and $\mathcal{G}_{\mathrm{III}}$, so that,

$$\mathcal{G} = \mathcal{G}_{\mathrm{I}} + \mathcal{G}_{\mathrm{II}} + \mathcal{G}_{\mathrm{III}} \tag{2}$$

The three separate components can be calculated by integration of the mode-I, mode-II, and mode-III traction–separation curves (Fig. 2):

$$\begin{aligned} \mathcal{G}_{\mathrm{I}} &= \int_0^{\delta_n} \sigma_{\mathrm{n}}(\delta)\,\mathrm{d}\delta \\ \mathcal{G}_{\mathrm{II}} &= \int_0^{\delta_{\mathrm{s}}} \tau_{\mathrm{s}}(\delta)\,\mathrm{d}\delta \\ \mathcal{G}_{\mathrm{III}} &= \int_0^{\delta_{\mathrm{t}}} \tau_{\mathrm{t}}(\delta)\,\mathrm{d}\delta \end{aligned} \tag{3}$$

Note that δ_n, δ_s, and δ_τ are not independent parameters; they evolve together as a natural result of the interplay between the deformation of two joined domains and the details of the three traction-separation laws. A failure criterion is required to determine the critical values of the three components of G, $\mathcal{G}_{\mathrm{I}}^*$, $\mathcal{G}_{\mathrm{II}}^*$, and $\mathcal{G}_{\mathrm{III}}^*$ (shaded areas in Fig. A. 2), at which complete fracture of the cohesive zone occurs. The criterion used in this study is a simple one (Wang and Suo 1990):

$$\mathcal{G}_{\mathrm{I}}^*/\Gamma_{\mathrm{Ic}} + \mathcal{G}_{\mathrm{II}}^*/\Gamma_{\mathrm{IIc}} + \mathcal{G}_{\mathrm{III}}^*/\Gamma_{\mathrm{IIIc}} = 1 \tag{4}$$

where Γ_{Ic}, Γ_{IIc}, and Γ_{IIIc} are the *total* areas under the pure opening and pure shear traction–separation laws, which equal to the mode-I, mode-II, and mode-III fracture toughness in linear elastic fracture mechanics. A more detailed account of this mixed-mode cohesive zone model can be found in Yang and Thouless (2001a). The major advantage of this cohesive law is that there is no need to specify mode mixedness *a priori*. The mode mixedness and the mixed-mode toughness evolve as a numerical outcome of the local equilibrium of stresses. More importantly, this law guarantees correct mode mixedness when LEFM conditions are satisfied (Yang et al. 2010; Parmigiani and Thouless 2007; Goutianos and Sorensen 2012).

Finally, due to the strong nonlinearity in the cohesive laws, the cohesive fracture problems are often solved with incremental schemes, in which the tangential stiffness matrix of the cohesive law is needed. In this mixed-mode cohesive model, the normal and tangential tractions are not coupled, and the tangential stiffness matrix can be obtained as

$$\mathbf{D}_{\mathrm{coh}} = \begin{bmatrix} k_s(\delta_s) & 0 & 0 \\ 0 & k_t(\delta_t) & 0 \\ 0 & 0 & k_n(\delta_n) \end{bmatrix} = \begin{bmatrix} df_s/d\delta_s & 0 & 0 \\ 0 & df_t/d\delta_t & 0 \\ 0 & 0 & df_n/d\delta_n \end{bmatrix} \tag{5}$$

where

$$\begin{aligned}
k_n(\delta_n) &= \begin{cases} \hat{\sigma}/\delta_{\mathrm{n1}} & (\text{if } \delta_{\mathrm{n}} \le \delta_{\mathrm{n1}}) \\ (\hat{\sigma}_2 - \hat{\sigma})/(\delta_{\mathrm{nc}} - \delta_{\mathrm{n2}}) & (\text{if } \delta_{\mathrm{n1}} < \delta_{\mathrm{n}} \le \delta_{\mathrm{n2}}) \\ -\hat{\sigma}_2/(\delta_{\mathrm{nc}} - \delta_{\mathrm{n2}}) & (\text{if } \delta_{\mathrm{n2}} < \delta_{\mathrm{n}} \le \delta_{\mathrm{nc}}) \\ 0 & (\text{if } \delta_{\mathrm{n}} > \delta_{\mathrm{nc}}) \end{cases} \\
k_{\mathrm{s}}(\delta_{\mathrm{s}}) &= \begin{cases} \hat{\tau}_s/\delta_{\mathrm{s1}} & (\text{if } |\delta_{\mathrm{s}}| \le \delta_{\mathrm{s1}}) \\ (\hat{\tau}_{s2} - \hat{\tau}_s)/(\delta_{\mathrm{sc}} - \delta_{\mathrm{s2}}) & (\text{if } \delta_{\mathrm{s1}} < |\delta_{\mathrm{s}}| \le \delta_{\mathrm{s2}}) \\ -\hat{\tau}_{s2}/(\delta_{\mathrm{sc}} - \delta_{\mathrm{s2}}) & (\text{if } \delta_{\mathrm{s2}} < |\delta_{\mathrm{s}}| \le \delta_{\mathrm{sc}}) \\ 0 & (\text{if } |\delta_{\mathrm{t}}| > \delta_{\mathrm{tc}}) \end{cases} \\
k_{\mathrm{t}}(\delta_{\mathrm{t}}) &= \begin{cases} \hat{\tau}_t/\delta_{t1} & (\text{if } |\delta_{\mathrm{t}}| \le \delta_{\mathrm{t1}}) \\ (\hat{\tau}_{t2} - \hat{\tau}_t)/(\delta_{\mathrm{tc}} - \delta_{\mathrm{t2}}) & (\text{if } \delta_{\mathrm{t1}} < |\delta_{\mathrm{t}}| \le \delta_{\mathrm{t2}}) \\ -\hat{\tau}_{t2}/(\delta_{\mathrm{tc}} - \delta_{\mathrm{t2}}) & (\text{if } \delta_{\mathrm{t2}} < |\delta_{\mathrm{t}}| \le \delta_{\mathrm{tc}}) \\ 0 & (\text{if } |\delta_{\mathrm{s}}| > \delta_{\mathrm{sc}}) \end{cases}
\end{aligned} \tag{6}$$

As emphasized earlier, the cohesive zone models allow for correct accounting of the nonlinear interaction of different fracture processes, while the linear elastic

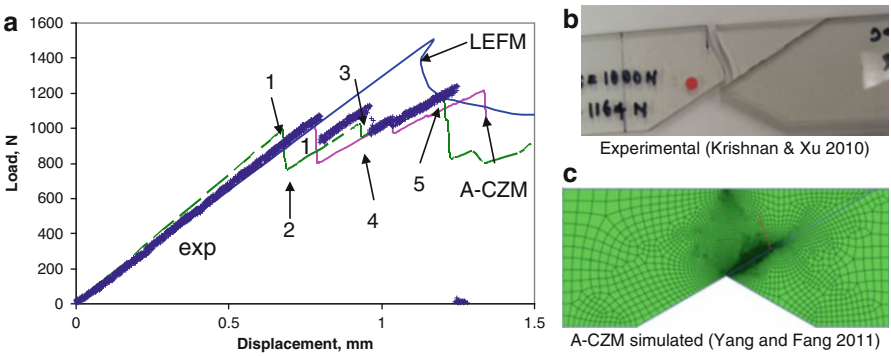


Fig. 3 (**a**) Comparison of experimentally measured, CZM simulated, and LEFM predicted load–displacement curve of an adhesively bonded, pre-notched PC specimen under three-point bending. (**b**) Experimental observation of kinked crack and (**c**) simulated kinking using nonlinear cohesive zone model

fracture mechanics (LEFM), due to its nonphysical nature at the limit of small crack length, cannot guarantee such correct coupling (Cox and Yang 2006; Yang and Cox 2005; Wisnom and Chang 2000; Xie et al. 2006; Song and Waas 1995; Yang et al. 1999, 2001; Moës and Belytschko 2002; Corigliano 1993; de Borst 2003; Elices et al. 2002; Needleman 1990; Remmers et al. 2003; Parmigiani and Thouless 2006; Yang and Thouless 2001b). Here an example that supports this view point is given. In Fig. 3a the load–displacement of a bonded polycarbonate (PC) specimen with a preexisting notch with root angle 120° under three-point bending test is shown. The bond line is at an angle of 30°to the specimen longitudinal direction and the fracture toughness along the bond line is at least three times smaller than that of the PC. The LEFM, which simply compares the ratio of the interface crack energy release rate (ERR) to its toughness ($\mathcal{G}_i/\Gamma_{ic}$) against the ERR ratio of a local mode-I kink crack ($\mathcal{G}_k/\Gamma_{kc}$) (He and Hutchinson 1989), predicts a complete interfacial crack propagation. However, repeated tests showed that kinking crack would form as shown in Fig. 3b. Only with cohesive zone modeling and with proper consideration of crack branching (Fang et al. 2011b; Yang et al. 2013) were the fracture behavior and load–displacement curves correctly predicted (Fig. 3a, c).

This example on bonded bulk PC specimens has a significant implication on composite fracture analyses because such crack kinking and coalescence processes are quite universal in laminated or textile composites. For example, crack jumping from one interface to another (Ling et al. 2011) involves both transverse crack kinking (from one interface into a transply) and crack merging (from transply into another interface). One of the important messages from this study is that crack branching is a competition between two fracture process zones and careful numerical treatment is needed to ensure proper stress and deformation coupling between the two fracture process zones. The LEFM view that a kinking process is purely determined by the two competing ERR-toughness ratios is nonphysical. Both the ERRs and cohesive strengths play an important role in determining whether an

interface crack should kink into the neighboring ply (a similar observation has also been made by Thouless and colleagues in Parmigiani and Thouless 2007; Parmigiani and Thouless 2006). The above successful simulation was done using the augmented cohesive zone (A-CZ) element recently developed by Fang et al. (2011b), which will be introduced in more detail shortly. The common practice of coupling X-FEs or A-FEs with *standard* CZ elements, as done in many previous analyses, cannot yield correct response in this case.

While the cohesive zone models are excellent in describing the fracture processes in composites, they need to be incorporated into numerical platforms such as FE programs as special elements. This implementation has a major shortcoming because it needs a crack path to be known a priori so that the CZ elements can be properly defined in a numerical model. However, in composites such information is not available. For example, the location and spacing of the transverse cracks in Fig. 1 cannot be known before analysis. Recent developments in advanced numerical methods for introducing arbitrary discontinuity in a continuum have made it possible to include most cohesive descriptions of major crack systems in structural models to account for the progressive damage. This will be reviewed in more detail in the next section.

Augmented Finite Element Method (A-FEM)

Basic Formulation of the A-FEM

In the classical finite element method (FEM), the displacement field in an element is approximated as

$$\mathbf{u}(\mathbf{x}) = \sum_i N_i(\mathbf{x})\mathbf{d}_i \tag{7a}$$

or in matrix form

$$\mathbf{u}(\mathbf{x}) = \mathbf{N}(\mathbf{x})\mathbf{d} \tag{7b}$$

where $N_i(\mathbf{x})$ is the standard FE shape function and $\mathbf{d}_i$ is the nodal displacement vector at node i. Since $N_i(\mathbf{x})$ is a continuous function within an element, Eq. 7a can only describe a continuous displacement field within the element. Therefore, the standard FE formulation cannot treat an intra-element discontinuity such as a cohesive crack or a material boundary.

In the A-FEM method with double nodes (one set of physical nodes 1–4 and another set of ghost nodes 1′–4′) developed by Ling et al. (Ling et al. 2009) and shown in Fig. 4(a), the two severed physical domains can be separately approximated by two mathematic elements (MEs) as shown in Fig. 4b, c. The active material domain in each ME is indicated by the shaded area. Stiffness and force integrations are performed only over the respective active domain in each mathematical element. The two mathematical elements are then connected by a cohesive failure description along the crack line.

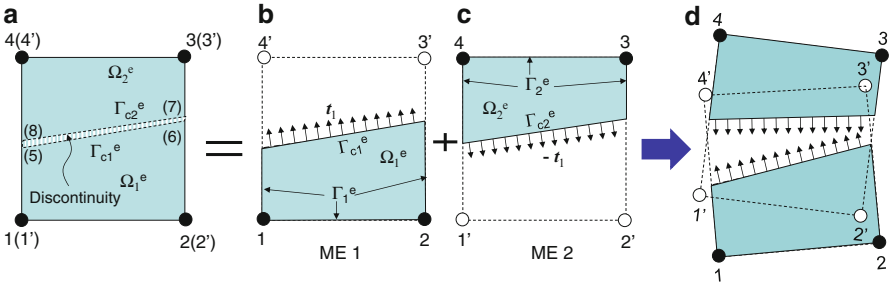


Fig. 4 (**a**) An A-FEM with double nodes traversed by an intra-element cohesive crack. This element can be treated by defining two mathematical elements (**b** and **c**), each with the same geometrical dimension as the A-FEM but with different physical material domains for stiffness integration (Ling et al. 2009). The mathematic elements are then packed into one A-FE with 8 nodes (**d**)

The discontinuous displacement field can be adequately described as

$$\mathbf{u}(\mathbf{x}) = \left(\sum_{i=1,2,3',4'} N_i(\mathbf{x}) \mathbf{d}_{1i} \right) \phi_1(\mathbf{x}) + \left(\sum_{i=1',2',3,4} N_i(\mathbf{x}) \mathbf{d}_{2i} \right) \phi_2(\mathbf{x}) \tag{8a}$$

where $\phi_1(\mathbf{x})$ and $\phi_2(\mathbf{x})$ are two functions defined below to ensure only the physical domains in the two mathematical elements are used in stiffness and force integration.

$$\phi_1(\mathbf{x}) = \frac{1 + H(\mathbf{x})}{2}; \quad \phi_2(\mathbf{x}) = \frac{1 - H(\mathbf{x})}{2}; \quad H(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in \Omega_1{}^e \\ -1 & \mathbf{x} \in \Omega_2{}^e \end{cases}$$

Here H(x) is generalized Heaviside function. In matrix form, Eq. 8a can be written as

$$\mathbf{u}(\mathbf{x}) = \mathbf{N}(\mathbf{x})\, \mathbf{d}_1\, \phi_1(\mathbf{x}) + \mathbf{N}(\mathbf{x})\, \mathbf{d}_2\, \phi_2(\mathbf{x}) = \left\{ \mathbf{N}(\mathbf{x})\phi_1(\mathbf{x}) \quad \mathbf{N}(\mathbf{x})\phi_2(\mathbf{x}) \right\} \left\{ \mathbf{d}_1 \quad \mathbf{d}_2 \right\}^T \tag{8b}$$

where the superscript "^{T}" denotes matrix transposition. Equation 8 allows for different displacement fields in physical domains $\Omega_1{}^e$ and $\Omega_2{}^e$, making it possible to account for discontinuous displacements (or displacement jumps) across the cohesive crack Γ_c. The displacement jumps can be conveniently computed as

$$\mathbf{w}(\mathbf{x}) = \mathbf{u}_1(\mathbf{x}) - \mathbf{u}_2(\mathbf{x}) = \left\{ \mathbf{N} \quad -\mathbf{N} \right\} \left\{ \mathbf{d}_1 \quad \mathbf{d}_2 \right\}^T (\mathbf{x} \in \Gamma_c) \tag{9}$$

If an element does not host a cohesive crack and is not influenced by a crack, either $\phi_1(\mathbf{x})$ or $\phi_2(\mathbf{x})$ is zero in the respective mathematical element. Then, Eq. 8 degenerates to the standard FE shape function interpolation of Eq. 7. Numerically, this can be simply reinforced by tagging all the DoFs of the ghost nodes with the associated physical nodes.

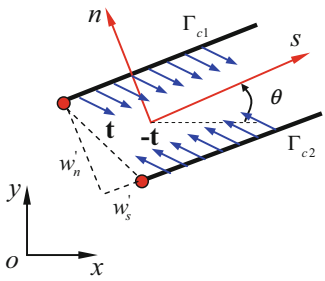


Fig. 5 Local coordinates (n-o-s) defined by the cohesive crack with $\boldsymbol{n}$ and $\boldsymbol{s}$ being normal and tangential directions with respect to the cohesive crack, θ is the angle of rotation between the global and local cohesive plane coordinate system

Since the cohesive tractions are typically nonlinear functions of cohesive opening displacements measured in local coordinates defined by the local crack direction and its perpendicular direction, it is advantageous to express the displacement jump in the local coordinates as shown in Fig. 5. The normal (along the n-direction) and shear (along the s-direction) displacement components can be obtained using a rotational matrix $\mathbf{R}$, i.e.,

$$\boldsymbol{\delta} = \{\delta_s, \delta_n\}^T = \mathbf{R}\mathbf{w} \tag{10}$$

where $\mathbf{R}$ is the rotational matrix between the global and local coordinates, i.e.,

$$\mathbf{R} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{11}$$

Similarly, the cohesive traction along the discontinuity can be expressed in local coordinates as

$$\mathbf{t}' = \{\tau_s(\delta_s), \sigma_n(\delta_n)\}^T = \mathbf{t}'(\boldsymbol{\delta}) \tag{12}$$

The weak form of the momentum equation for static analysis without body force can be derived from the virtual work principle (Ling et al. 2009)

$$\int_{\Omega} (\mathbf{L}\delta\mathbf{u})^{\mathrm{T}}\boldsymbol{\sigma}\ \mathrm{d}\Omega + \int_{\Gamma_c} (\delta\mathbf{w})^{\mathrm{T}}\mathbf{t}\ \mathrm{d}\Gamma = \int_{\Gamma_F} (\delta\mathbf{u})^{\mathrm{T}}\mathbf{F}\mathrm{d}\Gamma \tag{13}$$

where $\mathbf{L}$ is the differentiation operator as below

$$\mathbf{L} = \begin{bmatrix} \partial/\partial x & 0 \\ 0 & \partial/\partial y \\ \partial/\partial y & \partial/\partial x \end{bmatrix}, \tag{14}$$

and $\boldsymbol{\Gamma}_{\mathrm{F}}$ is the boundary upon which an externally applied traction acts (or Neumann boundary); $\boldsymbol{t} = \boldsymbol{t}(\mathbf{w})$ is the cohesive traction measured in the global coordinate system. In global coordinates, $\boldsymbol{t}$ can be obtained by

$$\mathbf{t}(\mathbf{w}) = \mathbf{R}^T \mathbf{t}'(\boldsymbol{\delta}) \tag{15}$$

Substituting the displacement field of Eqs. 8, 10, and 15 into Eq. 13,

$$\left\{ \delta\mathbf{d}_1^{\mathrm{T}} \quad \delta\mathbf{d}_2^{\mathrm{T}} \right\} \left(\int_{\Omega} \begin{Bmatrix} \mathbf{B}^T \boldsymbol{\sigma} \phi_1(\mathbf{x}) \\ \mathbf{B}^T \boldsymbol{\sigma} \phi_2(\mathbf{x}) \end{Bmatrix} \mathrm{d}\Omega + \int_{\Gamma_c} \begin{Bmatrix} \mathbf{N}^{\mathrm{T}} \mathbf{t} \\ -\mathbf{N}^{\mathrm{T}} \mathbf{t} \end{Bmatrix} \mathrm{d}\Gamma - \int_{\Gamma_F} \begin{Bmatrix} \mathbf{N}^{\mathrm{T}} \mathbf{F} \phi_1(\mathbf{x}) \\ \mathbf{N}^{\mathrm{T}} \mathbf{F} \phi_2(\mathbf{x}) \end{Bmatrix} \mathrm{d}\Gamma \right) = 0 \tag{16}$$

Since the virtual displacement array $\left\{ \delta\mathbf{d}_1^{\mathrm{T}} \quad \delta\mathbf{d}_2^{\mathrm{T}} \right\}$ is arbitrary, it follows that

$$\int_{\Omega^e} \begin{Bmatrix} \mathbf{B}^T \boldsymbol{\sigma} \phi_1(\mathbf{x}) \\ \mathbf{B}^T \boldsymbol{\sigma} \phi_2(\mathbf{x}) \end{Bmatrix} \mathrm{d}\Omega + \int_{\Gamma_c} \begin{Bmatrix} \mathbf{N}^{\mathrm{T}} \mathbf{t} \\ -\mathbf{N}^{\mathrm{T}} \mathbf{t} \end{Bmatrix} \mathrm{d}\Gamma = \int_{\Gamma_F} \begin{Bmatrix} \mathbf{N}^{\mathrm{T}} \mathbf{F} \phi_1(\mathbf{x}) \\ \mathbf{N}^{\mathrm{T}} \mathbf{F} \phi_2(\mathbf{x}) \end{Bmatrix} \mathrm{d}\Gamma \tag{17}$$

which is the equilibrium equation. The discretized form of Eq. 17 remains nonlinear in general because the cohesive traction, $\boldsymbol{t}(\mathbf{w})$ in the second integral is typically a nonlinear function of the crack opening $\mathbf{w}$. But it is straightforward to linearize the equation using an incremental scheme. The linearized equation can be written as

$$\begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix} \begin{Bmatrix} \Delta\mathbf{d}_1 \\ \Delta\mathbf{d}_2 \end{Bmatrix} = \begin{Bmatrix} \int_{\Gamma_F} \mathbf{N}^{\mathrm{T}} \Delta\mathbf{F} \phi_1(\mathbf{x}) \mathrm{d}\Gamma \\ \int_{\Gamma_F} \mathbf{N}^{\mathrm{T}} \Delta\mathbf{F} \phi_2(\mathbf{x}) \mathrm{d}\Gamma \end{Bmatrix} \tag{18}$$

where $\Delta\mathbf{d}_\alpha$ are incremental nodal displacements and $\Delta\mathbf{F}$ is the external load increment. The matrices $\mathbf{K}_{ij}$ are

$$\begin{aligned} \mathbf{K}_{11} &= \int_{\Omega^e} \mathbf{B}^T \mathbf{R}^T \mathbf{D}(\mathbf{x}) \mathbf{R} \mathbf{B} {\phi_1}^2(\mathbf{x}) \mathrm{d}\Omega + \int_{\Gamma_c} \mathbf{N}^{\mathrm{T}} \mathbf{R}^T \mathbf{D}_{\mathrm{coh}} \mathbf{R} \mathbf{N} \ \mathrm{d}\Gamma; \\ \mathbf{K}_{22} &= \int_{\Omega^e} \mathbf{B}^T \mathbf{R}^T \mathbf{D}(\mathbf{x}) \mathbf{R} \mathbf{B} {\phi_2}^2(\mathbf{x}) \mathrm{d}\Omega + \int_{\Gamma_c} \mathbf{N}^{\mathrm{T}} \mathbf{R}^T \mathbf{D}_{\mathrm{coh}} \mathbf{R} \mathbf{N} \ \mathrm{d}\Gamma; \\ \mathbf{K}_{12} &= \mathbf{K}_{21} = -\int_{\Gamma_c} \mathbf{N}^{\mathrm{T}} \mathbf{R}^T \mathbf{D}_{\mathrm{coh}} \mathbf{R} \mathbf{N} \ \mathrm{d}\Gamma \end{aligned} \tag{19}$$

and $\mathbf{D}(\mathbf{x})$ is the ply stiffness matrix in material coordinates, i.e.,

$$\mathbf{D}(\mathbf{x}) = \begin{bmatrix} 1/E_1 & -\nu_{12}/E_1 & 0 \\ -\nu_{12}/E_1 & 1/E_2 & 0 \\ 0 & 0 & 1/G_{12} \end{bmatrix} \tag{20}$$

$\mathbf{D}_{\mathrm{coh}}$ is the instantaneous tangential stiffness matrix of the cohesive law, which is given in Eq. 5. In obtaining the last equation of Eq. 19, the condition $\phi_1(\mathbf{x}) = \phi_2(\mathbf{x}) = 0$ has been used. Also note that $\mathbf{D}(\mathbf{x})$ can be different in Ω_1^e and Ω_2^e.

Once a cohesive crack is inserted or a preexisting discontinuity is detected in an element, subdomain integration (SDI) over the active material domains is necessary to obtain the stiffness matrix and force array (Eq. 18). The SDI is carried out in the

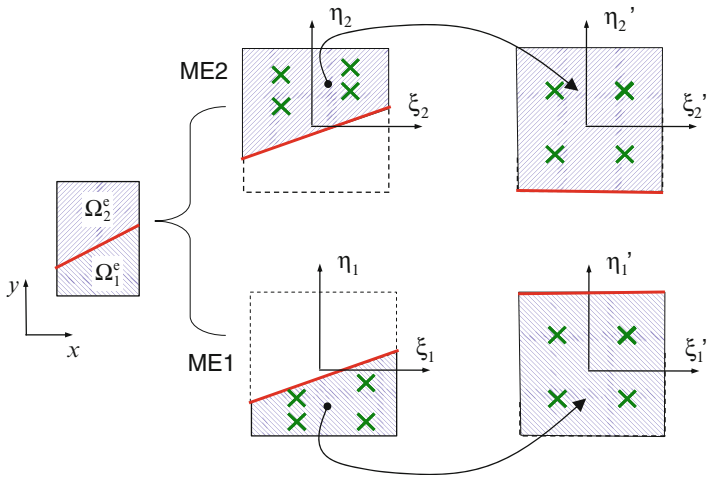


Fig. 6 The subdomain integration scheme: isoparametric interpolation of displacement within each mathematical element (*ME*); the corresponding physical domain in an ME is then mapped to another isoparametric domain for stiffness and cohesive force integration

isoparametric space of a physical material domain, as illustrated in Fig. 6. In this way, the stiffness matrix for an augmented element can be concisely written as

$$\mathbf{K}_e = \sum_{\alpha=1}^{2} \int_{-1}^{1} \int_{-1}^{1} \mathbf{B}\mathbf{D}_\alpha \mathbf{B} J_\alpha J_\alpha' \mathrm{d}\xi_\alpha' \mathrm{d}\eta_\alpha' \tag{21}$$

where J_α is the Jacobian that relates the entire domain of a mathematical element and its mapped domain in isoparametric space and J_α' is the Jacobian that relates the physical domain (shaded area) in a mathematical element to its mapped domain in isoparametric space.

Implementation of the A-FEM into ABAQUS as a User-Defined Element

The A-FEM element described above has been integrated as a user-defined plane element in the general purposed commercial finite element package, ABAQUS (v6.8). The A-FEM element is defined by 8 corner nodes as shown in the example of Fig. 7, of which the first four are real nodes (1–4). The second four nodes (5–8) are ghost nodes that will be used only when an element is traversed by a cohesive crack. The ghost nodes share the same geometric location with their corresponding real nodes. Two archetypical augmentation cases are shown in Fig. 7a, b.

The element flow chart is given in Fig. 8.

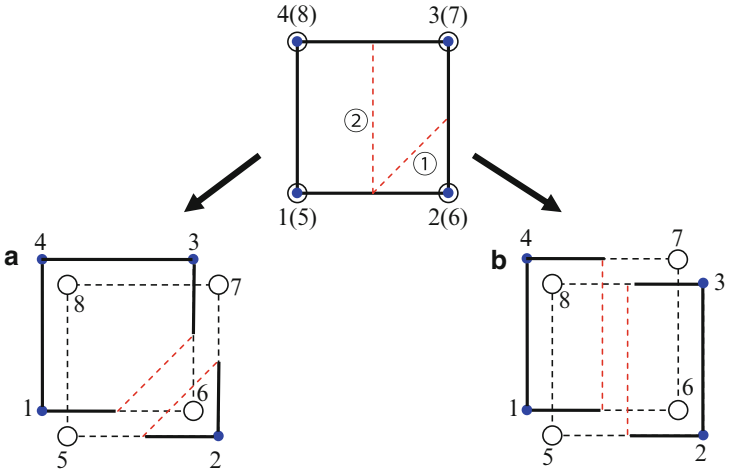


Fig. 7 Element connectivity of an A-FEM element and two possible cracking configurations: (**a**) the mathematic element definitions when the A-FEM element is cut into one pentagon and one triangle and (**b**) the mathematic element definitions when the crack cuts the A-FEM into two quadrilaterals

Augmented Cohesive Zone (A-CZ) Element for Crack Coupling

The 2D A-FEM element described in section "Augmented Finite Element Method (A-FEM)" can effectively handle the arbitrary cracking in each individual plies (or intraply cracks) as will be shown in the demonstration cases later. However, in typical composite laminates, the intraply cracks are often coupled with interlaminar delamination cracks. Various CZM-based interface elements have been developed for modeling the delamination cracks, and they are in general very effective when only delamination cracks are considered (Turon et al. 2006; Hallett and Wisnom 2006a; Yang and Cox 2005; Xie et al. 2006). However, a recent study by the authors has pointed out that when advanced elements such as A-FEM or X-FEM are used together with such traditional cohesive elements, significant numerical error can occur (Fang et al. 2010). This is due to the fact that although a traditional cohesive element can effectively account for the displacement discontinuity across the interface, it cannot resolve the displacement discontinuity within the element along the interface introduced by cracking of its abutting solid elements, as shown in Fig. 9.

The numerical error source and numerical treatment for 2D problems have been addressed by developing a 2D augmented cohesive zone (A-CZ) element in Fang et al. (2010). The detailed formulation for 2D and 3D A-CZ element is given below.

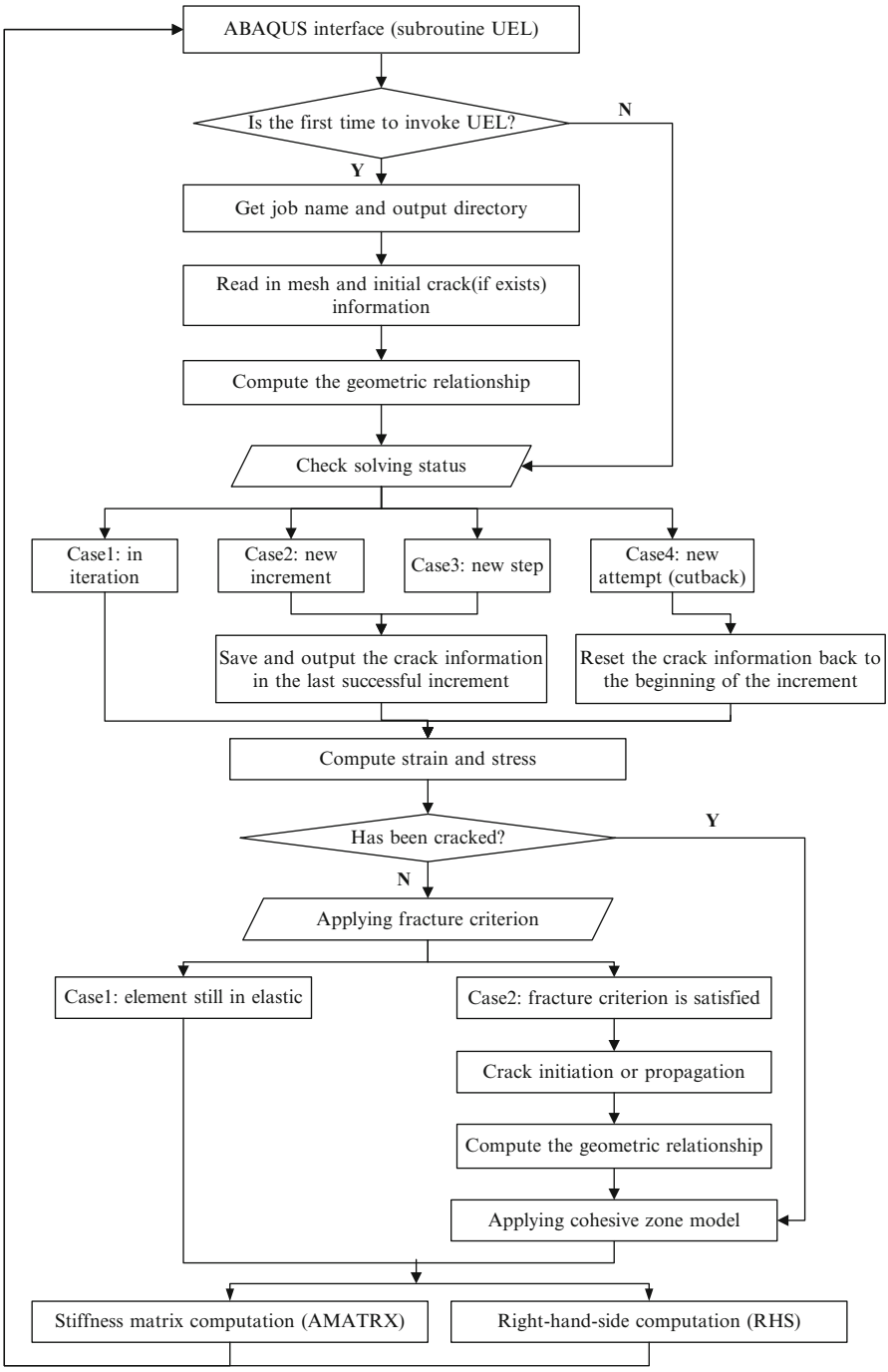


Fig. 8 The flow chart of the UEL subroutine for the 2D A-FEM when implemented into ABAQUS

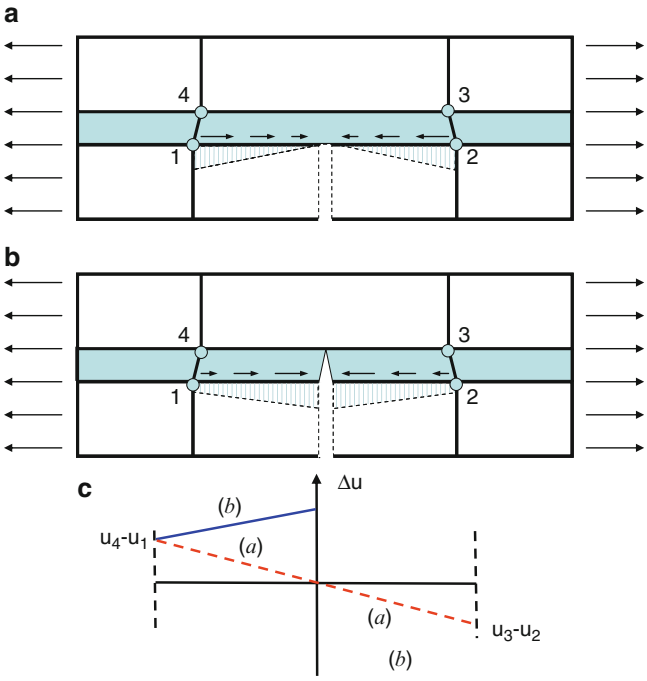


Fig. 9 (**a**) Physical interface shear stress distribution due to cracking of a bonded solid element; (**b**) inaccurate shear stress distribution by the traditional CZ element; (**c**) physically consistent shear displacement jump along the interface element (*solid segments*) as compared to the linearly interpolated shear displacement jump from the traditional CZ (*dashed line*)

2D A-CZ Formulation

The 2D A-CZ element is defined with two sets of corner nodes shared with the abutting solid A-FEs (e.g., 1-2-3-4-1′-2′-3′-4′ as shown in Fig. 10a). If neither of the solid A-FEs is cracked, all the ghost nodes are tied with the corresponding real nodes, and both the solid A-FEs and A-CZ element behave like standard FE and traditional CZ element. However, if either or both of the abutting solid elements are cracked and the crack tip(s) reaches the interface, the cohesive element will be augmented accordingly. Figure 10b shows the augmented configuration when both solid A-FEs have cracked. In this case, the two MEs associated with top solid A-FE are 4-3′-8′-7 and 4′-3-8-7′; and the two MEs for the bottom solid A-FE are 5-6′-2′-1 and 5′-6-2-1′. Accordingly, the A-CZ can then be augmented into two MEs for cohesive stress integration with 1-2′-3′-4 for ME-1 and 1′-2-3-4′ for ME-2 as shown in Fig. 10b, c. The respective cohesive stress integration domains for the MEs are shown in the figure by shaded areas.

For each ME-α(α = 1 or 2), a local coordinate system can be established by defining the tangential direction ($\mathbf{s}_\alpha$) along the line that connects the midpoints of edge 14 and 2′3′ for ME-1 or that connects the midpoints of 1′4′ and 23 for ME-2, as

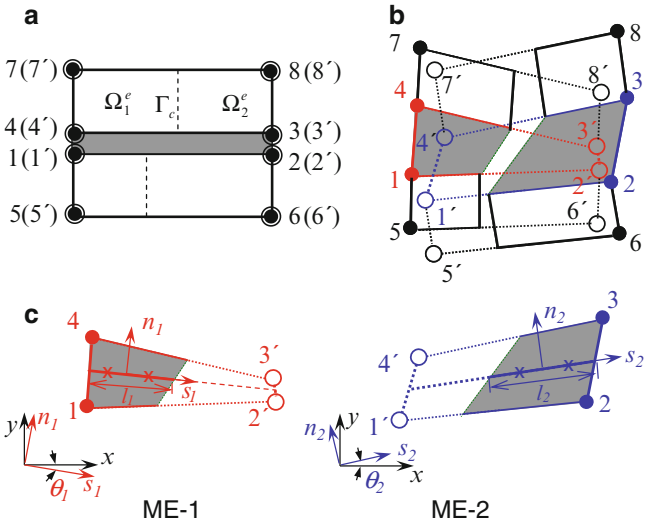


Fig. 10 (**a**) Two A-FEs bonded by an A-CZ element before deformation, all with doubly assigned nodes for possible intra-element cracking. (**b**) After deformation (assuming both solid A-FEs are cracked), the A-CZ will be augmented into two mathematical elements with their own local coordinates as shown in (**c**). For each ME, the two Gaussian points are indicated by the two symbols "×," which are located in the ME's active cohesive integration domain (*shaded area*). The interface is exaggerated for illustration purposes (Reprinted from International Journal for Numerical Methods in Engineering, X. J. Fang, Q. D. Yang, B. N. Cox, Z. Q. Zhou, *An augmented cohesive zone element for arbitrary crack coalescence and bifurcation in heterogeneous materials*, Pages No. 841–861, Copyright 2011, with permission from John Wiley and Sons)

shown in Fig. 10c. The respective normal directions ($\mathbf{n}_\alpha$) are perpendicular to $\mathbf{s}_\alpha$. Thus the rotational matrix for each ME can be obtained as $\mathbf{Q}_{i\alpha} = \text{Diag}\{\mathbf{R}_{i\alpha}, \mathbf{R}_{i\alpha}, \mathbf{R}_{i\alpha}, \mathbf{R}_{i\alpha}\}$. Here $\mathbf{R}_{i\alpha}$ is computed from Eq. 11 using the rotation angle θ_α of cohesive ME-α as shown in Fig. 10c. Note that from here on, all quantities associated with the A-CZ interface element are denoted with subscript "*i*".

The nodal displacements of ME-α in local coordinates can be obtained as

$$\mathbf{d}_{i\alpha}{}' = \mathbf{Q}_{i\alpha}\mathbf{d}_{i\alpha} \tag{22}$$

where $\mathbf{d}_{i\alpha}$ and $\mathbf{d}_{i\alpha}{}'$ are nodal displacement arrays in global and local coordinates, respectively. The local displacement jumps (crack displacements) across the ME-α can be obtained as

$$\boldsymbol{\delta}_{i\alpha} = \{\,\delta_{si}, \quad \delta_{ni}\,\}_\alpha^{\,T} = \mathbf{N}\mathbf{d}_{i\alpha}{}' = \mathbf{N}\mathbf{Q}_{i\alpha}\mathbf{d}_{i\alpha} \tag{23}$$

where

$$\mathbf{N} = \begin{bmatrix} -N_1 & 0 & -N_2 & 0 & N_1 & 0 & N_2 & 0 \\ 0 & -N_1 & 0 & -N_2 & 0 & N_1 & 0 & N_2 \end{bmatrix}_\alpha$$

is the interpolation matrix consisting of standard nodal shape functions N_i (i = 1 ~ 2). Note that the $\boldsymbol{\delta}_{i\alpha}$ is now the interface separation displacement (i.e., the differences in displacement between the upper and lower interfaces).

The interface cohesive stresses in the MEs can be obtained from the traction–separation relations of the interfacial cohesive law

$$\mathbf{t}_i' = \{\, \tau_i \quad \sigma_i \,\}^T = \{\, g_{si}(\delta_{si}) \quad g_{ni}(\delta_{ni}) \,\}^T \tag{24}$$

where τ_i and σ_i are shear and normal cohesive stresses in local coordinates, respectively, and $g_{ti}(\delta_{ti})$ and $g_{ni}(\delta_{ni})$ are traction–separation relations that govern the cohesive interface, respectively.

From the principle of virtual work, the equilibrium equation of the ME-α can be derived straightforwardly:

$$\mathbf{F}_{i\alpha} = \int_{l_\alpha} \mathbf{Q}_{i\alpha}{}^T \mathbf{N}^T \mathbf{t}' \mathrm{d}l, \tag{25}$$

where l_α is the active cohesive integration segment in ME-α as shown in Fig. 10c. The incremental form of Eq. 25 is

$$\left(\int_{l_\alpha} \mathbf{Q}_{i\alpha}{}^T \mathbf{N}^T \mathbf{D}_{\mathrm{i\ coh}} \mathbf{N}\, \mathbf{Q}_{i\alpha} \mathrm{d}l \right) \Delta \mathbf{d}_{i\alpha} = \Delta \mathbf{F}_{i\alpha} \tag{26}$$

The integral in "()" is the tangential stiffness of the ME-α. In this study, Gaussian integration is used. For each ME-α, the two Gaussian points are located in the active cohesive integration segment only and they are indicated in Fig. 10c by the symbol "×". Assembly of the two MEs into one A-CZ element is straightforward. The equivalence of the A-CZ element to the traditional treatment with two standard cohesive elements (if the crack path is known a priori) can be proven following a procedure similar to that given in Ling et al. (2009).

The proposed A-CZ element has been implemented into a general purpose commercial FE code ABAQUS (V6.9, SIMULIA. Providence, RI, USA) as a user-defined element, and it works jointly with the A-FE described in section "Augmented Finite Element Method (A-FEM)."

3D Augmented Cohesive Zone Element formulation

The above formulation for 2D A-CZ element can be extended to 3D straightforwardly. In the following, the formulation of a 16-node 3D A-CZ element as shown in Fig. 11a is given.

In 3D configuration, a crack can cut an interface from either top or bottom surface. There are three possible cut configurations in this case: (1) if the crack cut the interface cohesive element into two brick-like domains, the assignment of real and virtual nodes is shown in Fig. 11b, c; (2) if the crack cut the interface element

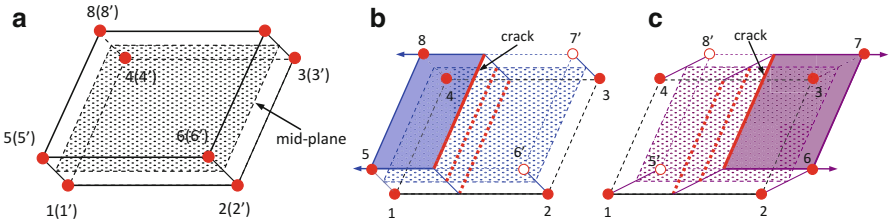


Fig. 11 (**a**) Definition of the 3D augmented cohesive zone element with 16 nodes (eight real nodes and eight virtual nodes denoted with superscript '"). If the element is cut into two due to cracking of the top abutting solid element, two mathematic elements with properly assigned real and virtual nodes as shown in (**b**) with element connectivity 1-2-3-4-5-6′-7′-8 and in (**c**) with connectivity 1-2-3-4-5′-6-7-8′ can be defined

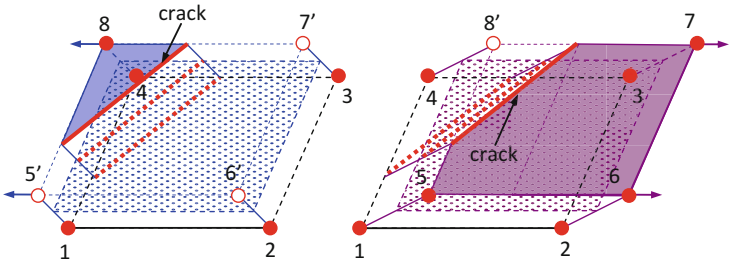


Fig. 12 Assignment of real and virtual nodes when the 3D interface element is cut into a triangular subdomain and a pentagon subdomain from one surface (*left*: 1-2-3-4-5′-6′-7′-8; *right*: 1-2-3-4-5-6-7-8′)

into one triangular and one pentagonal shape, the assignment of real and virtual nodes is shown in Fig. 12; and (3) if the interface element is cut simultaneously from both top and bottom surfaces, the element is considered as completely failed, i.e., both stiffness matrix and right-hand side are given zero values. (This is a simple numerical treatment. It can be done following A-FEM procedure but another copy of virtual nodes is needed.)

If there is no crack cutting the element, it behaves like a regular cohesive element and the crack displacements can be expressed as

$$\mathbf{w}(\mathbf{x}) = \begin{bmatrix} \mathbf{N}(\mathbf{x}) & -\mathbf{N}(\mathbf{x}) \end{bmatrix} \begin{Bmatrix} \mathbf{d}^t \\ \mathbf{d}^b \end{Bmatrix} \tag{27}$$

where $\mathbf{w}(\mathbf{x})$ is the relative displacement vector at position $\mathbf{x}$ measured in global coordinates, $\mathbf{d}^t$ and $\mathbf{d}^b$ denote the nodal displacement vectors of top and bottom surface, respectively. $\mathbf{N}(\mathbf{x}) = [\mathbf{N}_1(\mathbf{x}) \quad \mathbf{N}_2(\mathbf{x}) \quad \mathbf{N}_3(\mathbf{x}) \quad \mathbf{N}_4(\mathbf{x})]$ is the shape function matrix and each sub-matrix associated node i takes the form as

$$\mathbf{N}_i(\mathbf{x}) = \begin{bmatrix} N_i & 0 & 0 \\ 0 & N_i & 0 \\ 0 & 0 & N_i \end{bmatrix} \tag{28}$$

where N_i are standard bilinear shape functions.

When the element is cut by a crack from either the top or the bottom surface, the respective surface, which has four corners, needs to be augmented. The relative displacements of the A-CZ can be written as

$$\mathbf{w}(\mathbf{x}) = \mathbf{N}(\mathbf{x})\mathbf{d}(\mathbf{x}) = \begin{bmatrix} \mathbf{N}^{t1}(\mathbf{x}) & \mathbf{N}^{t2}(\mathbf{x}) & -\mathbf{N}^{b1}(\mathbf{x}) & -\mathbf{N}^{b2}(\mathbf{x}) \end{bmatrix} \begin{Bmatrix} \mathbf{d}^{t1} \\ \mathbf{d}^{t2} \\ \mathbf{d}^{b1} \\ \mathbf{d}^{b2} \end{Bmatrix} \tag{29}$$

The index $\alpha = 1$ and 2 denote the real and virtual node displacements, respectively. The detailed expressions of the shape function matrix do not have the same general form but differ from nodes to nodes:

$$\mathbf{N}_i^{t1}(\mathbf{x}) = \begin{bmatrix} N_i H_i^t & 0 & 0 \\ 0 & N_i H_i^t & 0 \\ 0 & 0 & N_i H_i^t \end{bmatrix} \tag{30a}$$

$$\mathbf{N}_i^{t2}(\mathbf{x}) = \begin{bmatrix} N_i\left(1 - H_i^t\right) & 0 & 0 \\ 0 & N_i\left(1 - H_i^t\right) & 0 \\ 0 & 0 & N_i\left(1 - H_i^t\right) \end{bmatrix} \tag{30b}$$

In the above expressions, $H_i^t(\mathbf{x})$ are the step functions associated with the position of node i as

$$H_i^t = \frac{1 + H^t(\mathbf{x}_i)H^t(\mathbf{x})}{2} \tag{31a}$$

where $H^t(\mathbf{x})$ is the generalized Heaviside function that takes different values in the two sides divided by crack across the domain, i.e.,

$$H^t(\mathbf{x}) = \begin{cases} +1 & \mathbf{x} \in \Omega^{t+} \\ -1 & \mathbf{x} \in \Omega^{t-} \end{cases} \tag{31b}$$

Similarly, the shape function matrix corresponding to the degree of freedoms (DoFs) of the bottom surface can be written as

$$\mathbf{N}_i^{b1}(\mathbf{x}) = \begin{bmatrix} N_i H_i^b & 0 & 0 \\ 0 & N_i H_i^b & 0 \\ 0 & 0 & N_i H_i^b \end{bmatrix} \tag{32a}$$

$$\mathbf{N}_i^{b2}(\mathbf{x}) = \begin{bmatrix} N_i\left(1 - H_i^b\right) & 0 & 0 \\ 0 & N_i\left(1 - H_i^b\right) & 0 \\ 0 & 0 & N_i\left(1 - H_i^b\right) \end{bmatrix} \tag{32b}$$

The displacement and cohesive stress are to be expressed in local coordinates. Denoting $\mathbf{Q}$ as the rotation matrix between global and local coordinates, the local crack displacements can be expressed as

$$\boldsymbol{\delta}(\mathbf{x}) = \mathbf{Q}\mathbf{w}(\mathbf{x}) \tag{33}$$

The cohesive stresses in local coordinates are

$$\mathbf{t}' = \{\tau_{\mathrm{s}}, \tau_{\mathrm{t}}, \sigma_{\mathrm{n}}\}^T = \{g_{\mathrm{s}}(\delta_s), g_{\mathrm{t}}(\delta_t), g_{\mathrm{n}}(\delta_n)\}^T \tag{34}$$

Finally, based on the principle of virtual work, the incremental form of the equilibrium equation can be written as

$$\Delta\mathbf{F} = \iint_{\Omega} \mathbf{Q}^T\mathbf{N}^T\Delta\mathbf{t}'(\boldsymbol{\delta})d\Omega = \left(\iint_{\Omega} \mathbf{Q}^T\mathbf{N}^T\mathbf{D}_{\mathrm{coh}}\mathbf{N}\mathbf{Q}d\Omega\right)\Delta\mathbf{d} \tag{35}$$

where the tangential stiffness matrix of cohesive laws, $\mathbf{D}_{\mathrm{coh}}$, can be derived from Eq. 5 by replacing $f_s(\delta_s)$, $f_t(\delta_t)$, and $f_n(\delta_n)$ with $g_s(\delta_s)$, $g_t(\delta_t)$, and $g_n(\delta_n)$, respectively. The above formulation has been implemented in ABAQUS (v.6.9) as a user subroutine and works together with A-FEM in section "Augmented Finite Element Method (A-FEM)."

Numerical Examples

In this part, several numerical examples to demonstrate the predictive capability of the A-FE and the A-CZ in analyzing the progressive failure phenomena in laminated composites are presented.

Cracking in Composite Beams Under Three-Point Bending

The three-point bending test illustrated in Fig. 13 is widely used in the engineering materials community to measure the transverse strength of unidirectional composites because of its simplicity. However, owing to the lower values of transverse modulus (E_2) for typical PMCs, beams with relatively small aspect ratio (L/h) are preferred to avoid large deflection before failure. Experimental observations show that for this type of test, the composites usually fail by the initiation of a transverse crack (mode I) from the maximum stress location (mid-bottom point in Fig. 13),

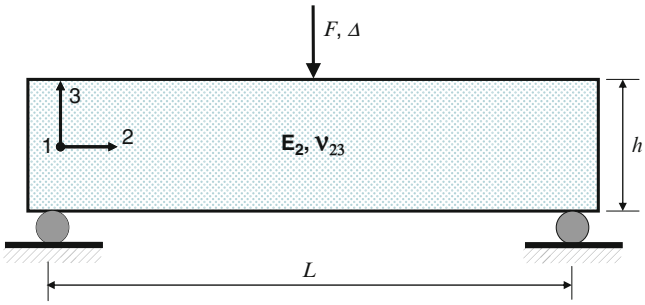


Fig. 13 Three-point bending test configuration of the composite beam

which then propagates along a straight line towards the central loading point on the top surface (O'Brien et al. 2003). The propagation process can be stable or unstable, depending upon the mode-I toughness and beam strength as well as the beam dimensions.

Since the deformation and loading are all within the isotropic plane of the transversely isotropic unidirectional composite, the problem can be treated as a plane stress problem with transverse modulus, E_2, and Poisson's ratio, ν_{23}. A simple dimensional analysis shows that the normalized load ($F/\hat{\sigma}h$) versus normalized displacement (Δ/h) should have the following functional form:

$$\frac{F}{\hat{\sigma}h} = f\left(\frac{\Gamma_{Ic}}{\hat{\sigma}h}, \frac{\hat{\sigma}}{E_2}, \nu_{23}, \frac{L}{h}\right)\frac{\Delta}{h} \tag{36}$$

where $f(\bullet)$ is a dimensionless function that relates the normalized load and displacement. It is a function of the normalized mode-I toughness ($\Gamma_{Ic}/\hat{\sigma}h$), the normalized cohesive strength ($\hat{\sigma}/E_2$), the normalized beam length or span (L/h), and Poisson's ratio (ν_{23}).

In the A-FEM model, there is no need to prescribe the crack path. But a crack initiation criterion is needed. In this numerical example, since the crack is initiated along the symmetric midsection of the specimen, the initiation criterion is simply set to be

$$\overline{\sigma}_{22}/\hat{\sigma} = 1 \tag{37}$$

where $\overline{\sigma}_{22}$ is the opening stress averaged in an element (before it is augmented). The crack growth direction is perpendicular to $\overline{\sigma}_{22}$, i.e., vertically upwards in Fig. 13. The mode-I cohesive law used is of triangular type with a very large initial stiffness.

The A-FEM predicted normalized load versus displacement curves for typical fiber reinforced composites with $\hat{\sigma}/E_2 = 0.0044$ and $\nu_{23} = 0.3$ is plotted in Fig. 14a. Normalized beam lengths, $L/h = 5$, and five normalized mode-I toughnesses, $\Gamma_{Ic}/\hat{\sigma}h = 0.0004$, 0.004, 0.01, 0.02, and 0.04, were investigated. A wide range of normalized toughness values were used to investigate the possible brittle-to-ductile transition of a composite material.

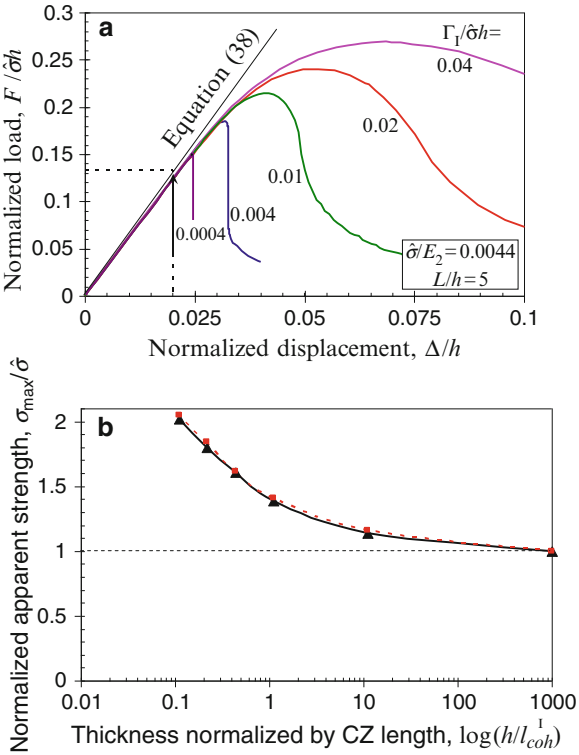


Fig. 14 (**a**) A-FEM predicted normalized load versus normalized load-point displacement in a three-point bending test for normalized beam span of $L/h = 5$. (**b**) Dependence of normalized apparent strength on normalized thickness for three-point bending test

For comparison purposes, analytical results using simple beam theory with shear correction are superimposed in the figure. The analytical beam theory result for normalized load versus normalized deflection with shear correction is

$$\frac{F}{\hat{\sigma}h} = \left[\frac{1}{4}\left(\frac{L}{h}\right)^3 + \frac{1+\nu_{23}}{2}\left(\frac{L}{h}\right)\right]^{-1}\left(\frac{\hat{\sigma}}{E_2}\right)^{-1}\frac{\Delta}{h} \tag{38}$$

According to this simple beam theory, the specimen should fail once the maximum bending stress reaches the cohesive strength, i.e., $\overline{\sigma}_{22}/\hat{\sigma} = 1$. Thus the peak value for the normalized load can be obtained as

$$\left(\frac{F_{\max}}{\hat{\sigma}h}\right) = \frac{2}{3}\left(\frac{L}{h}\right)^{-1} \tag{39}$$

The normalized analytical peak load depends only on the normalized beam span (L/h) and is not a function of toughness. In Fig. 14a, Eq. 38 is shown by the straight line for beam span $L/h = 5$. The corresponding peak load is marked by the arrow.

The A-FEM simulation, however, predicts a strong dependence of the peak normalized load ($F_{\max}/\hat{\sigma}h$) on the normalized toughness ($\Gamma_{Ic}/\hat{\sigma}h$); $F_{\max}/\hat{\sigma}h$ increases

as $\Gamma_{Ic}/\hat{\sigma}h$ increases, and the difference between analytical and numerical peak loads increases with $\Gamma_{Ic}/\hat{\sigma}h$. This is the well-known size effect phenomenon that has been appreciated in concrete materials research (Carpinteri and Colombo 1989; Bazant and Planas 1998). Since $F_{max}/\hat{\sigma}h$ increases with $\Gamma_{Ic}/\hat{\sigma}h$, the apparent strength evaluated using beam theory and the peak load, i.e., $\sigma_{max} = 3F_{max}L/2h^2$, which in normalized form is $\frac{\sigma_{max}}{\hat{\sigma}} = \frac{3}{2}\frac{F_{max}}{\hat{\sigma}h}\frac{L}{h}$, depends on specimen thickness. That is, if one keeps L/h fixed while varying h, the apparent strength increases with decreasing specimen thickness, h, as shown in Fig. 14b.

The root cause of the dependence of apparent strength on toughness is the development of the fracture process zone (or cohesive zone) before complete fracture of the specimen. More precisely, the load increase associated with the R-curve effect for a developing cohesive zone (Yang et al. 2006a) is the direct cause of the increase of apparent strength. To better elucidate the point, four snapshots of the numerically predicted bending strain (ε_{22}) for the case of $L/h = 5$ and $\Gamma_{Ic}/(E_2h) = 0.004$ are shown in Fig. 15.

Figure 15(i) shows the elastic bending strain just before cohesive damage initiation. The symmetric strain distribution with respect to the midplane (neutral plane) is obvious and is in good agreement with beam theory prediction, which demonstrates the effectiveness of the A-FEM in handling a damage-free elastic continuum. Figure 15(ii) corresponds to a point where the bending stress at the mid-bottom has just reached the cohesive strength and a cohesive crack is just initiated. From this point on, the bending strain starts to localize. However, the cohesive crack remains bridged (i.e., cohesive stress has not fallen to zero) and can still transfer bending stress. Further increase of the external load is needed to open the cohesive crack continuously. The peak load is reached only when the cohesive crack opening at the mid-bottom surface reaches the critical displacement in the cohesive law. The cohesive opening within the cracking element induces large strain localization in the element. However, as the cohesive stress becomes zero, the bending strain in those elements adjacent to the cracking element starts to release, which is evident in Fig. 15(iii). Beyond this point, the fully developed cohesive zone starts to move towards the loading point on the opposite surface and the load starts to drop (Fig. 15(iv)).

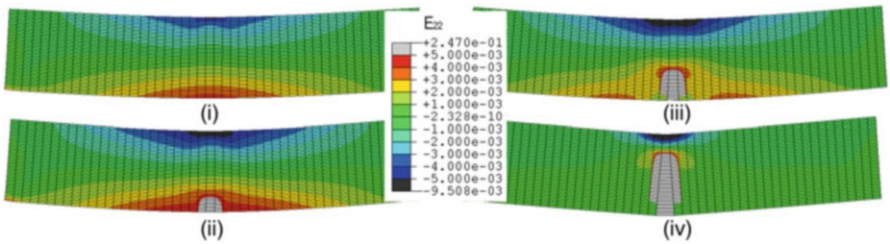


Fig. 15 A-FEM predicted mode-I fracture of a beam under three-point bending. Images from (**i**–**iv**) show sequentially progressive damage initiation from (**i**) damage-free pure elastic strain field, to (**ii**) damage initiation when the cohesive strength is reached at mid-bottom surface, to (**iii**) fully developed cohesive zone before it starts to travel towards the loading surface, and to (**iv**) final crack opening just before complete fracture

The increase of the apparent strength over the cohesive strength is directly associated with the ratio of beam thickness to the cohesive zone length. For a mode-I crack in an infinite domain, the cohesive zone length can be estimated as

$$l_{\text{coh}}^{\text{I}} = E_2 \Gamma_{\text{Ic}} / \hat{\sigma}^2 \tag{40}$$

For typical unidirectional composites in the transverse direction with $\hat{\sigma}$ ranging from 50 to 100 MPa, E_2 from 10 to 15 GPa, and Γ_{Ic} from 100 to 500 J/m^2, the cohesive zone length is 0.1 to 3 mm. For typical laminar composites, the thickness is also on the order of mm, and therefore size effects may be non-negligible, especially for composites with toughened matrices.

The A-FEM predicted normalized apparent strength ($\sigma_{\max}/\hat{\sigma}$) as a function of thickness normalized by the cohesive zone length ($h/l_{\text{coh}}^{\text{I}}$) is plotted in Fig. 14b. The deviation of $\sigma_{\max}/\hat{\sigma}$ from unity increases rapidly with decreasing $h/l_{\text{coh}}^{\text{I}}$. Even with $h/l_{\text{coh}}^{\text{I}}$ as large as 10, $\sigma_{\max}/\hat{\sigma}$ is still about 15 % larger than unity. This helps explain the experimental observation that the transverse strength values measured from three-point bending tests are usually systematically higher than those measured from simple tension tests (O'Brien et al. 2003; O'Brien and Salpekar 1993; Adams et al. 1990). This is usually explained by Weibull's theory, but recent test data of O'Brien et al. (2003) and O'Brien and Salpekar (1993) showed that this theory did poorly in explaining the thickness dependence of apparent transverse strength.

The A-FEM predicted curves for fixed L/h in Fig. 14a also show the transition from brittle-to-ductile fracture with continuously increasing normalized toughness ($\Gamma_{\text{Ic}}/\hat{\sigma}h$). Again, this is a well-documented behavior in the concrete fracture research community, but less appreciated in the composite materials research literature. Following the terminology used in concrete research, the normalized toughness, $\Gamma_{\text{Ic}}/\hat{\sigma}h$, can be renamed as the "ductility number," which characterizes the ductility of a specimen: the larger the $\Gamma_{\text{Ic}}/\hat{\sigma}h$, the more ductile the specimen appears. Note that, in addition to the fracture parameters (Γ_{Ic} and $\hat{\sigma}$), the ductility number depends strongly on the structural dimension h (specimen thickness). That is, a specimen with smaller thickness will appear more ductile than a specimen with a larger thickness.

The numerical results in Fig. 14a suggest that the brittle-to-ductile transition occurs at $\Gamma_{\text{Ic}}/\hat{\sigma}h \approx 0.0004$. For a typical fiber reinforced composite with $\Gamma_{\text{Ic}} = 200$ J/m^2 and cohesive strength of $\hat{\sigma} = 60$ MPa, the critical thickness corresponding to this transition is $h \approx 8$ mm. This is potentially significant in three-point bending tests of laminated composite specimens, especially those with relatively small thickness. If the specimen thickness is small enough that the ductility number exceeds the transition value of 0.0004, the apparent strength could be significantly larger than the cohesive strength, as shown in Fig. 14b.

Single-Cantilever Beam Bending Test

This example is a demonstration of the combined use of 2D A-FEM and 2D A-CZ elements for a coupled interface delamination and crack jump (kinking) from one

interface to another interface, which is a common damage pattern in laminated composites. Crack jumping is one of the key mechanisms that govern the through-the-thickness damage propagation of composite panels under transverse impact loading (Choi and Chang 1992; Wang and Crossman 1980; Finn et al. 1993). However, direct analyses of the problem are few in the composite research literature, largely due to lack of effective analytical tools that can adequately consider the direct coupling of all possible damage events. In particular, the transply cracking has been almost completely ignored in the majority of the previous studies and only delamination has been addressed. Analytically, the critical technical barrier in considering the delamination jump is that the initiation sites of transply cracks cannot be determined *a priori*.

Here a single-cantilever beam test used by NASA is modeled to reveal the detailed crack jumping process. The single-cantilever beam geometry chosen for analysis is shown in Fig. 16. The beam chosen as a study case is composed of 60 unidirectional plies with ply thickness of 0.127 mm (0.005 in.), with stacking sequence, geometry, boundary conditions, and loading shown in Fig. 16. Material properties are chosen to be typical of an aerospace grade carbon/epoxy composite, e.g., IM7/8552. An initial delamination crack is introduced between the interface of 30th (0° ply) and 31st (90° ply) plies, which is immediately above a group of four transverse (90°) plies. The entire layup of 60 plies was carefully designed so that (1) the flexure properties (orthotropic elasticity) of the sub-beams above and below the initial delamination crack are as symmetric as possible; (2) upon loading, the interface delamination crack will tend to propagate along the interface for a certain distance before possibly deflecting into the four 90° plies (31st to 34th ply); and (3) the kinking crack would eventually impinge on the interface between 34th ply (90°) and 35th ply (0°) and possibly initiate a new interface crack.

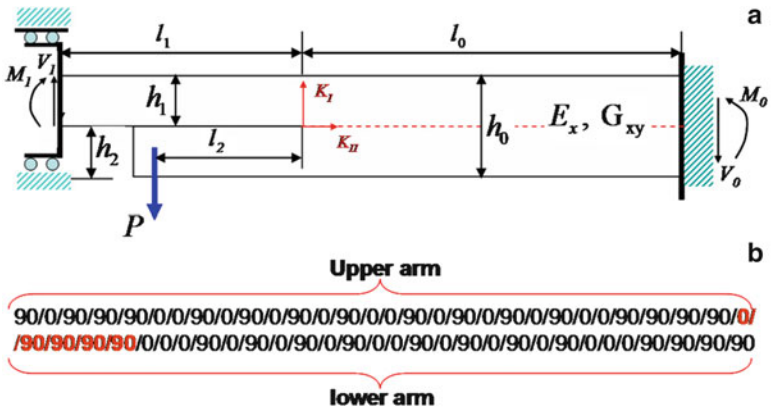


Fig. 16 (**a**) sketch of the single cantilever beam (*SCB*) for possible experimental investigation of delamination crack jump and (**b**) stacking sequence (from *top* to *bottom*) of the composite beam analyzed in this study (Ling et al. 2011) (With permission from ASCE)

The dimensions of the SCB are $h_1 = h_2 = 3.8$ mm, $l_1 = 28.0$ mm, $a_0\,(=l_2) = 12.7$ mm, and $l_0 = 35.6$ mm. These particular dimensions and boundary conditions mimic those proposed by the specimen's developers. However, the conclusions to be reached in this study will be pertinent to cases with rescaled dimensions; the important physical length parameters are the thickness of the laminae and the length of the pre-crack relative to the length scales of the nonlinear fracture process zones. The latter, for both the delamination and transverse ply cracks, are of the same order of magnitude as the ply thickness for typical polymer composites and for the case studied (see below).

The upper sub-beam of the SCB is roller supported at the left end so that it can move only in the horizontal direction, and the entire beam is rigidly supported at the right end (built-in end condition). A transverse load is applied on the lower sub-beam at a fixed distance from the initial crack tip, as shown in Fig. 16a. The laminar properties assumed for each unidirectional ply are given in Ling et al. (2011).

The detailed mesh and magnified views of important regions are shown in Fig. 17. The mesh is sufficiently fine to resolve individual plies (thickness 0.127 mm). In particular, the four 90° plies directly below the interface of the initial delamination were modeled by eight layers of A-FEM elements, allowing resolution of stress variations pertinent to a propagating transverse ply crack. The A-FEM elements accommodate transverse ply crack initiation at any location and orientation where the local stress meets the initiation criterion discussed below.

Individual ply properties were assigned to each ply with the proper orientation. The plies were assumed rigidly bonded except at the two interfaces joining the central $(/90/_4)$ layer to the immediately neighboring 0° plies. These interfaces were

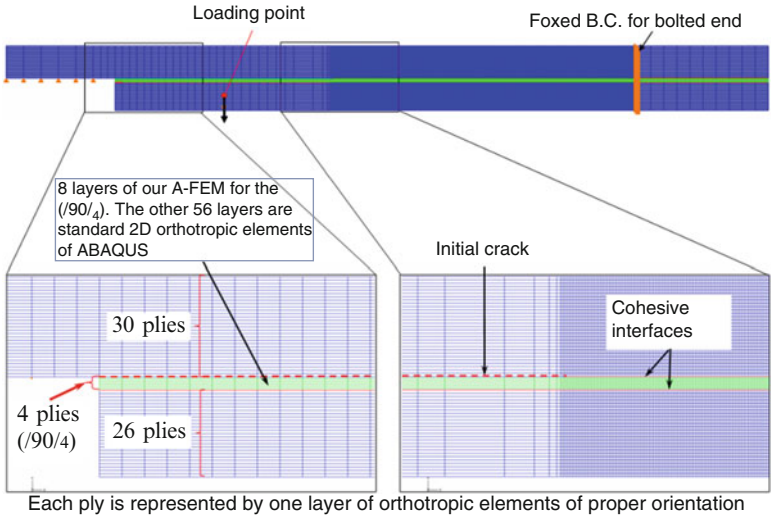


Fig. 17 Numerical mesh and modeling details for the single cantilever beam (Ling et al. 2011). (With permission from ASCE)

modeled using the 2D A-CZ elements with the following cohesive parameters. The mode-I and mode-II toughnesses, i.e., the entire areas under the mode-I and mode-II traction–separation curves, are assigned the values $\Gamma_{\mathrm{Ic}} = 200$ J/m^2 and $\Gamma_{\mathrm{IIc}} = 1300$ J/m^2, respectively, and the tensile and shear cohesive strengths the values $\hat{\sigma} = 80$ MPa and $\hat{\tau} = 90$ MPa. These values have been determined empirically (by matching theory to data) and are representative of typical aerospace composites such as IM7/8552 (Hallett and Wisnom 2006a).

For the SCB studied here, a local mode-I criterion (or equivalently, the maximum hoop stress criterion) was used for determining the triggering of transply cracks in the 90° plies. This criterion is appropriate because the kinking crack occurs in the isotropic plane of the $(/90_4)$ layer. For other crack orientations, the criterion would depend on the angle between the crack plane and the fiber direction. The failure criterion is implemented in the A-FEM by testing for the condition

$$\overline{\sigma}_{p1}/T_2 = 1 \tag{41}$$

where $\overline{\sigma}_{p1}$ is the maximum principal stress derived from the spatially averaged stress state of an element and T_2 is the transverse ply strength, which in this study is assumed to be identical to the tensile cohesive strength (i.e., $T_2 = \hat{\sigma}$). Using a gauge-averaged criterion with averaging performed over an element is justified when the nonlinear fracture process zone is larger than the element size, which is the case in this work. The initiation criterion implies that the material strength is bounded by the maximum traction that can be supported by a fracture process zone (Bao et al. 1992). Once this initiation condition is satisfied, a cohesive zone is inserted in the element with its direction perpendicular to the principal direction of $\overline{\sigma}_{p1}$. This is consistent with experimental observations that if the ply thickness is relatively small, then the kinking crack is in near mode-I condition (Choi and Chang 1992; Clark 1989).

In the A-FEM formulation, a kinking crack is allowed to initiate and propagate automatically according to the initiation criterion of Eq. 41 and the propagation rule. There is no need to prescribe the kinking crack location or orientation in advance.

The predicted load versus load-point displacement curves for the SCB with an initial crack length of $a_0 = 12.7$ mm (0.5 in.) and a transverse ply strength $T_2 = 80$ MPa is plotted in Fig. 18. In this figure, the dashed curves are the load versus load-point displacement curves with interface cracking confined to the upper or lower interface only (i.e., no transverse ply kinking cracks allowed). This was done by switching off the crack nucleation capability in the A-FEM model. These curves are in good agreement with the analytical results derived using beam theory with crack-tip rotations, validating the fidelity of the A-FEM in dealing with delamination cracks. They provide approximate bounds to the load if delamination jumping is allowed: the load should follow the upper interface-delamination-only curve before the development of the kinking process and the lower interface-delamination-only curve after the kinking process is completed and the crack is propagating along the lower interface.

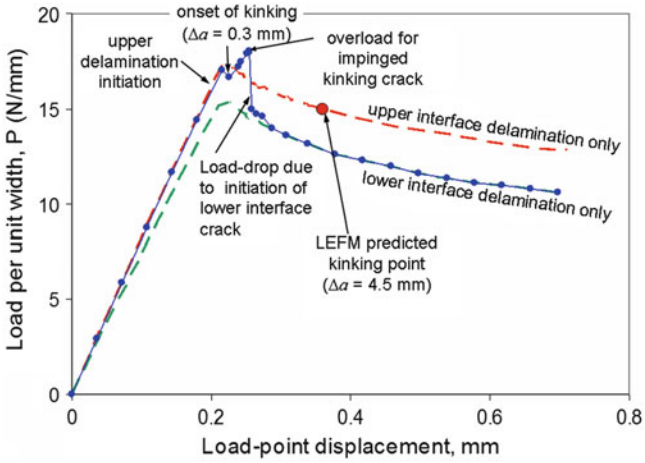


Fig. 18 The load–deflection curve predicted by A-FEM (*solid line* with dots) is almost always bounded by the interface-delamination-only curves for delaminations on the *upper* and *lower* interfaces. Immediately after the transverse ply crack initiation, however, the curve shows a distinct load-increasing phase with significant overload due to the arrest of the kinking crack before it reaches the *lower* interface (Ling et al. 2011) (With permission from ASCE)

The solid line with filled dots is the curve predicted by the A-FEM with ply jumping permitted. In this particular geometry, the initial crack propagated only a very short distance along the upper interface before the kinking process started ($\Delta a = 0.3$ mm). The kinking point is indicated in Fig. 18 by an arrow. In another simulation using initial crack length of $a_0 = 7.6$ mm, the delamination crack propagated much further along the upper interface ($\Delta a = 2.7$ mm) before kinking into the ($/90/_4$) layer.

The predicted curve including the jump process agrees well with the interface-delamination-only curves prior to and shortly after the points where transverse ply cracks initiate. Immediately after the transverse ply crack initiation, however, the curve shows a distinct load-increasing phase before abruptly dropping and conforming to the lower interface-delamination-only curve. Moreover, the peak load value immediately after the kinking point exceeds the upper interface-delamination-only curve quite significantly. A close examination of the transverse ply crack initiation and propagation process shows that the sharp load increase is associated with slowing of the transverse ply crack as it approaches the lower interface. Because the transverse ply crack arrests before reaching the lower interface in the SCB, initiating damage in the cohesive elements of the lower interface crack requires a higher load than it would in the presence of an impinging crack.

The predicted arrest of a transverse crack approaching an interface has been validated by the NASA test and it has also been reported in other material systems. For example, a crack in a flexible film rigidly bonded to a hard substrate experiences decreasing energy release rate as it approaches the interface

(see, e.g., Ye et al. (1992) for an account of this phenomenon within linear elastic fracture mechanics). That a higher load is then required for delamination initiation has also been reported in the analysis of a center cracked bi-material beam under four-point bending (Zhang and Suo 2007).

The coupled evolution of delamination and transverse ply cracks shows global–local coupling. The simulations predict that before one transverse ply crack initiates and fully traverses the transverse ply, a sequence of trial transverse ply cracks can develop as the upper interface delamination crack propagates. But these kinks are unable to propagate beyond one or two elements before the delamination crack resumes propagation along the upper interface, so that the trial transverse ply cracks become unloaded and arrest.

The initiation of the trial transverse ply cracks is due to the local stress (or strain) concentration that accompanies the moving delamination fracture zone. Within this stress concentration zone, the cohesive crack initiation criterion (Eq. 41) is locally satisfied and hence a cohesive zone is created by the A-FEM code as a possible transverse ply crack. However, the global driving force (energy release rate) is not sufficient to propagate the transverse ply crack all the way across the transverse plies. A transverse ply crack is fully developed only when the global energy release rate equals Γ_{Ic}. The implication of this global–local interaction is that the kinking process is relatively insensitive to local small flaws, i.e., the kinking process is more likely to be deterministic, because it is driven to completion by global conditions rather than local (and statistically fluctuating) conditions.

The global–local coupling is further illustrated in Fig. 19, which gives four snapshots of the local opening strains (ε_{22}) at the delamination crack tip as it propagates and eventually deviates into the ($/90/_4$) layer. The four snapshots correspond to delamination crack initiation, the onset of the transverse ply crack, the arrest of the transverse ply crack before reaching the lower interface, and delamination along lower interface after the completion of the kinking process. The instants corresponding to these snapshots are marked by arrows on the load versus load-point displacement curve in Fig. 18.

Figure 19 shows that the delamination crack has a nonlinear fracture process zone of 1.2 mm. For the kinking crack, the nonlinear fracture process zone is larger than 0.5 mm (but bounded by the two interfaces with the neighboring 0° plies), which is about four times the single ply thickness (0.127 mm) and therefore comparable to the total thickness of the ($/90/_4$) layer (Fig. 16). At the delamination initiation point (Fig. 19a), the opening strain across the interface is continuous and locally concentrated around the delamination crack tip. As the delamination propagates, a transverse ply crack is generated within the stress concentration zone of the delamination crack tip (Fig. 19b). However, this crack does not propagate downwards, and it becomes unloaded as the delamination crack moves away. When the delamination crack has propagated for approximately another 0.3 mm, a new transverse ply crack is generated and succeeds in propagating to the lower interface (Fig. 19c). Even at this point, the transverse ply crack is not completely traction-free and still transfers stress across the cohesive crack plane. But the opening strain at the delamination crack tip starts to release from this point

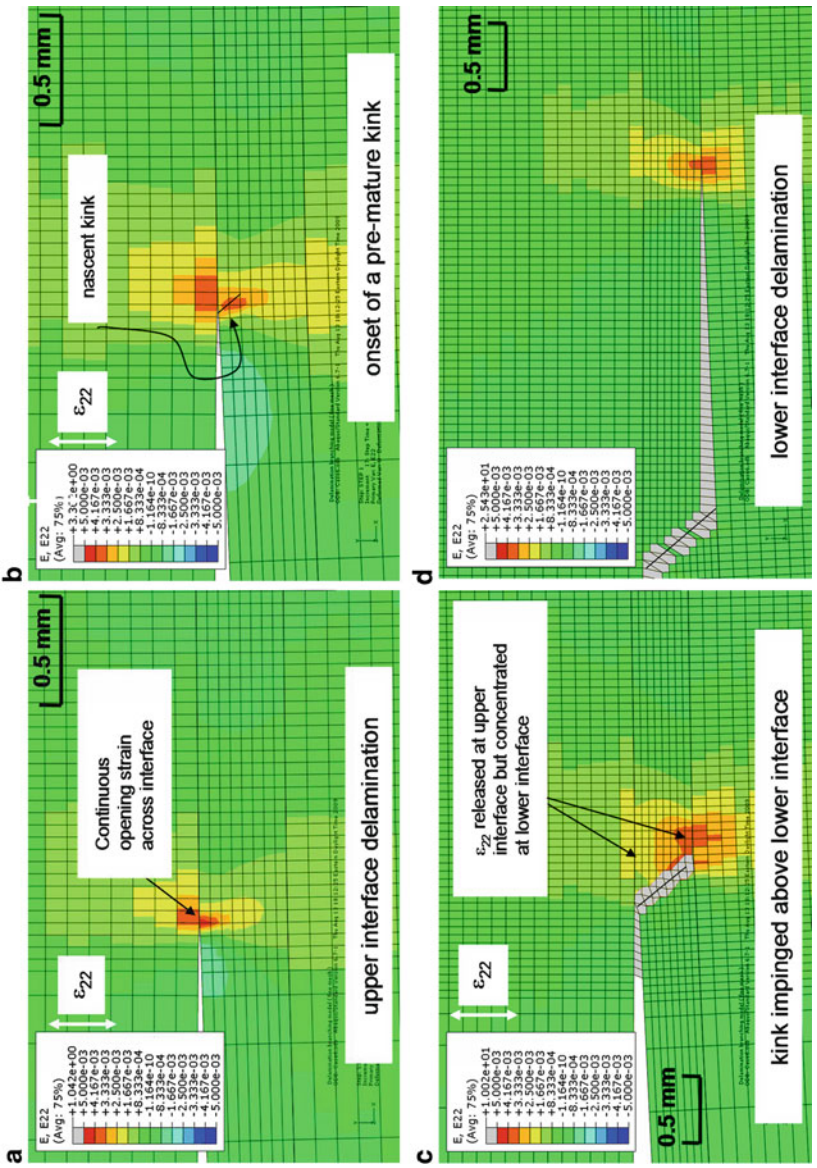


Fig. 19 Four snapshots showing different stages of the coupled delamination and kinking and the corresponding instants in the load–deflection curve (Ling et al. 2011) (With permission from ASCE)

on. As a result, the delamination crack stops propagating and further load increase drives only the transverse ply crack. The transverse ply crack, once initiated with sufficient global driving force, quickly propagates towards the lower interface but is again arrested before reaching the lower interface (Fig. 19c). A significant load spike is needed to complete its union with a new delamination crack on the lower interface. During the load spike, the opening strain at the lower interface builds up quickly. Finally, at peak load, the remaining uncracked ligament ($1 \sim 2$ elements in the $(/90/_4)$ layer) and a certain length of lower interface fail together in a single increment, indicating unstable damage propagation (Fig. 19d). This process is reflected in the load–deflection curve by the sudden drop of the load. Thereafter, the crack propagates along the lower interface in a stable fashion.

Multiple-Crack Evolution in a $[0/90]_s$ Double-Notched Tension Specimen

In this section, the 2D A-FEM and 3D A-CZ elements are applied to model the coupled multiple-cracking systems in a double-notched orthogonal composites under tension reported by Hallett and Wisnom (2006a). The experimentally observed damage evolution in these specimens has been shown in Fig. 1a. It will be demonstrated that, through a careful material properties calibration process mostly using literature data, successful predictions to the entire damage evolution process, including coupled delamination and intraply cracking (longitudinal splits and transverse ply cracking), can be achieved with unprecedented high fidelity. On the other hand, it will also be shown through a parametric study that if the fracture parameters are not well calibrated, only a subset of the macroscopic composite behavior may be predicted due to their insensitivity to the material properties, but others may be completely missed out. This demonstrates the necessity for high-fidelity material characterization for microscopic damage processes.

The specimen geometry and notch configuration are reproduced in Fig. 20a (Fang et al. 2011a). The two dots on the specimen surface are used as reference points for measuring displacements through non-contact optical means, so that experimental load–displacement data are not influenced by test machine compliance. Displacement-controlled in-plane tension is applied to the specimen.

Due to symmetry in specimen geometry, ply stacking and loading, only a quarter of the specimen is modeled with symmetric boundary conditions applied at the left and bottom edges, as shown in Fig. 20b. The possibility that damage might develop asymmetrically is therefore not considered; damage in the tests being simulated did remain close to symmetric. Further, in this study since the deformation in the planar DNT specimen under simple tension is mostly within the x-y plane, each ply is then modeled by plane stress elements with ply thickness properly assigned. The out-of-plane stresses are therefore neglected. The ply thickness is 0.127 mm (0.005 in.).

Transverse ply cracks and splitting cracks were modeled using 2D augmented elements within the plies: the flexibility of A-FEM therefore allows these two crack systems to arise at any location the initiation criterion is satisfied during

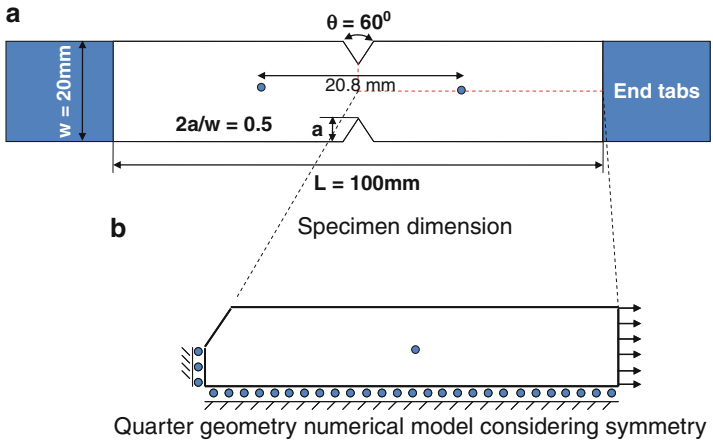


Fig. 20 (**a**) Geometry of the double-notched tension specimen tested by Hallett and Wisnom (2006a). The two filled dots in the specimen were used as markers for direct optical measurement of the displacement (so that the machine compliance could be eliminated from experimental data). (**b**) The quarter geometry numerical modeled used in our A-FEM simulation (Reprinted from Journal of the Mechanics and Physics of Solids, vol 59, X.J. Fang, Z.Q. Zhou, B.N. Cox, Q.D. Yang, *High-fidelity simulations of multiple fracture processes in a laminated composite in tension*, Pages No. 1355–1373, Copyright 2011, with permission from Elsevier)

simulations. Since the potential fracture plane of the delaminations is known, they were treated differently by pre-locating 3D A-CZ elements along all ply interfaces. This mixed meshing strategy allows the complex interaction between intraply cracks and inter-ply delamination cracks to be represented with appropriate realism.

Because the modeling strategy uses plane stress elements for plies, it implicitly depicts any intraply cracks (transverse or splitting crack) as immediately penetrating the entire ply thickness, once initiated. It therefore starts to interact immediately with interface cohesive elements. This treatment is valid as long as the mode-I toughness (which controls transverse cracking in 90° plies) is smaller than the mode-II toughness (which controls delamination), which is the case in most polymer composites (including the subject composite). A detailed study of a single transverse ply crack and associated delaminations shows that given this ordering of the toughness values, the intraply cracks do span the ply before delamination is triggered (Zhou et al. 2010).

Mesh Sensitivity

Before proceeding to presenting the prediction results, a mesh sensitivity study is carried out. Note that with cohesive modeling, mesh-independent results can only be expected in a nonlinear fracture analysis if the mesh is sufficiently fine to depict the cohesive damage process zone adequately. An adequate mesh refinement can be estimated using the analytical approximations for the cohesive zone size that are summarized in Yang and Cox (2005). For the calibrated cohesive parameters in Table 1 below, the estimated cohesive zone length for mode-I fracture (transverse

Table 1 Transversely isotropic laminar elasticity and baseline cohesive parameters (Reprinted from Journal of the Mechanics and Physics of Solids, vol 59, X.J. Fang, Z.Q. Zhou, B.N. Cox, Q.D. Yang, *High-fidelity simulations of multiple fracture processes in a laminated composite in tension*, Pages No. 1355–1373, Copyright 2011, with permission from Elsevier)

Laminar elastic properties	E_1 (GPa)	E_2 (=E_3) (GPa)	ν_{12}(=ν_{13})	ν_{23}	G_{12}(=G_{13}) (GPa)	
	43.9	15.4	0.3	0.3	4.34	
Baseline CZM parameters	Mode I		Mode II		Mode III	
	Γ_{Ic} (J/m^2)	$\hat{\sigma}_n$ (MPa)	Γ_{IIc} (J/m^2)	$\hat{\tau}_s$ (MPa)	Γ_{IIIc} (J/m^2)	$\hat{\tau}_t$ (MPa)
	250	120	900	80	900	80

intraply cracks) is $\sim l_{coh}^{I} = 0.16$ mm and that for mode-II fracture (e.g., splitting cracks) is $l_{coh}^{II} = 0.53$ mm. Accordingly, mesh dependence was studied using the three mesh resolutions of $l_e = 0.1$, 0.2, and 0.4 mm to represent the numerical domain in Fig. 20b. Comparisons were made for the predicted stress–extension curve (Fig. 21a), the splitting crack growth versus applied nominal stress curve (Fig. 21b), and the transverse crack density (Fig. 21c–e).

Specimen Compliance and Splitting Crack Growth

From Fig. 21a, b, the stress–displacement curves and splitting crack length versus applied stress curves obtained for all three meshes are very consistent. This is consistent with the previous conclusion that A-FEM results are mesh independent provided that the mesh size is smaller than the cohesive zone size (Ling et al. 2009).

Transverse Crack Spacing

In the present theory, the locations and spacing of transverse intraply cracks are determined automatically according to the computed stress state at any time in a simulation; unlike other recent works (Van de Meer et al. 2010; Van de Meer and Sluys 2009b), no constraint is imposed on crack spacing other than the fact that one element and its nodes can support only one crack. Therefore, another crack cannot propagate into any of the (neighboring) elements that share the nodes of the already cracked element, and the limit crack spacing in the mesh of Fig. 21 is that any two transverse cracks must be separated by at least one element. That is, the minimum possible transverse crack spacing is 0.2, 0.4, and 0.8 mm for meshes $l_e = 0.1$, 0.2, and 0.4 mm, respectively.

The predicted cracking systems for the three meshes are shown in Fig. 21c–e. The predicted transverse crack spacings for meshes with $l_e = 0.1$, 0.2, and 0.4 mm are, respectively, 0.9 mm ($9l_e$), 1.0 mm ($5l_e$), and 1.2 mm ($3l_e$). The changes in crack spacing are small given the large variation in mesh resolution and acceptably small in that they are less than variations that might be expected from other factors, such as stochastic local material strength.

In addition, the predicted delaminations (dark regions in Fig. 21c–e) are also all very consistent. The demonstrated mesh independence also implies that the numerical formulation presented in sections "Nonlinear Fracture Models for Composite

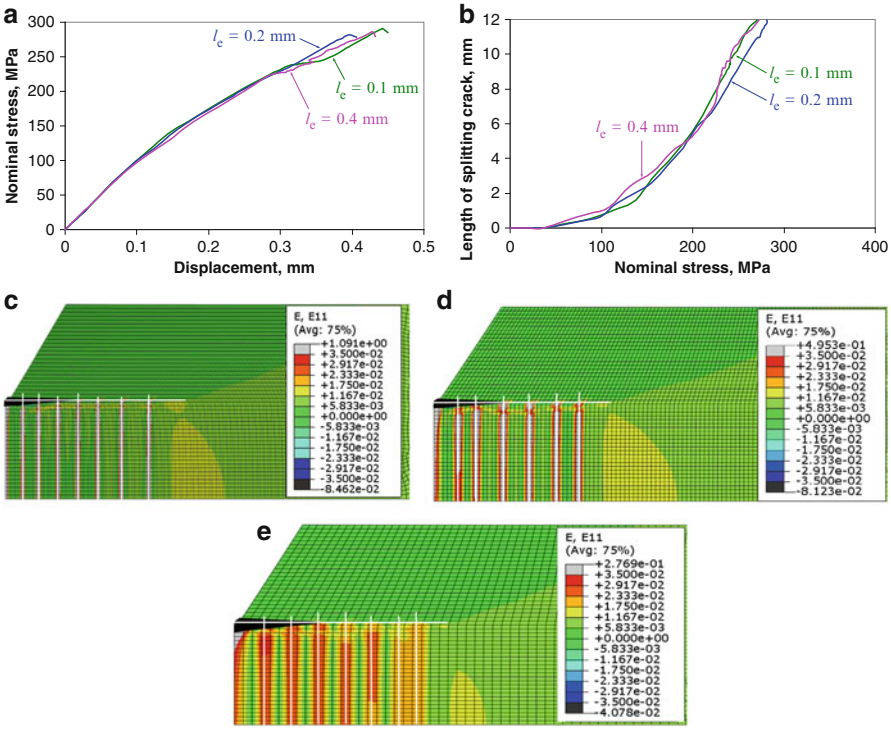


Fig. 21 Comparison of predicted numerical results with three different mesh resolutions. (**a**) Stress–displacement curves; (**b**) splitting crack length versus applied nominal stress curves; and (**c**–**e**) coupled multiple cracks include transverse cracks (*vertical lines*), splitting cracks (*horizontal lines*), and interlaminar delamination (*dark triangular zones*) at a stress level of 238 MPa (Reprinted from Journal of the Mechanics and Physics of Solids, vol 59, X.J. Fang, Z.Q. Zhou, B.N. Cox, Q.D. Yang, *High-fidelity simulations of multiple fracture processes in a laminated composite in tension*, Pages No. 1355–1373, Copyright 2011, with permission from Elsevier)

Materials," "Augmented Finite Element Method (A-FEM)," and "Augmented Cohesive Zone (A-CZ) Element for Crack Coupling" is able to capture correctly the nonlinear damage coupling among the splitting, transverse, and delamination cracks.

Nonlinear Damage Models and Their Calibration

The progressive damage in the DNT specimen was reported in detail by Hallett and Wisnom (2006a). There were five main types of damage processes evolving in a coupled manner before final failure occurred (Fig. 1a): (1) splitting cracks in the 0° ply emanating from the notch tip and propagating along the loading direction, which started when the applied nominal stress level reached about 25 % of the final strength (~260–290 MPa); (2) multiple transverse cracking in the 90° ply with a saturation cracking density of about 2 cracks/mm; (3) triangular delaminations propagating between the 0° and 90° plies in a self-similar fashion with wedge

angle of ~7–10°; (4) shear nonlinearity; and (5) fiber rupture responsible for ultimate failure of the specimen. In addition, data for the load versus the displacement between the two reference points indicated in Fig. 21a (filled dots), and the load versus the splitting crack length as a function of applied nominal stress, were available for comparison. All these observations and data will be used to test predictions. Each observed mechanism is capable of arising in the simulations, but its presence is not prescribed in advance.

This section details the constitutive models that were used to represent elasticity and the various damage modes. All parameters in each constitutive model were calibrated using data that were *independent* of the simulated tests.

Elasticity

Each ply is assumed to be homogeneous and transversely isotropic.

Calibration: The transversely isotropic ply properties are determined by comparing simulations to experimental measurements of composite stiffness. For the subject materials (E-Glass/913), this calibration was carried out in Cui et al. (1992), whose results are summarized in Table 1.

Fiber Rupture

Fiber rupture in 0° plies is introduced by a material degradation model that modifies the affected computational element

$$E = \begin{cases} E_1 & (\varepsilon_1 \leq \varepsilon_f) \\ \sim 0 & (\varepsilon_1 > \varepsilon_f) \end{cases} \tag{42}$$

where E_1 is the Young's modulus along the fiber direction and ε_f is a threshold strain value along the fiber direction at which fiber rupture occurs.

Calibration: The critical strain, ε_f, is chosen to be 3.5 %, a value determined empirically for an E-Glass/913 composite by Hallett and Wisnom (2006a). It is typical of glass fibers.

Fine-Scale Shear Deformation

A tension test of a ±45° laminate generates shear stress, τ_{12}, aligned with the fibers within individual plies (Jelf and Fleck 1994). The simple relation $\tau_{12} = \sigma_a/2$ where σ_a is the applied stress allows a constitutive law for the ply material in shear to be deduced from the global stress–strain data of the test (see also the comparison with Iosipescu shear tests in Kumosa and Odegard (2002)). Prior research on kink band formation encourages the use of this calibration: the critical shear stress level deduced from ±45° tension data can be used to predict the onset of kink bands in composites in compression.

The constitutive law calibrated by a ±45° test for the subject material is shown in Fig. 22 (Wisnom 1994). The observed softening sums the combined effect of fine-scale plasticity of the type as well as discrete delaminations and intraply cracks, whose size and spacing are associated with larger spatial scales. However, the delaminations and intraply cracks contribute to nonlinearity mainly at the higher

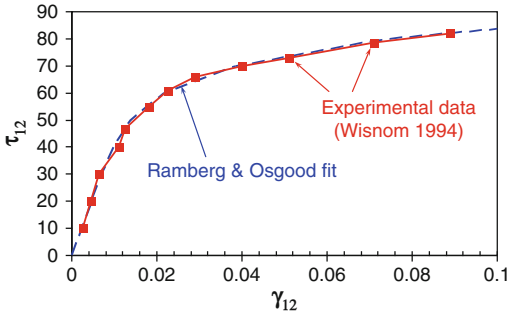


Fig. 22 Shear stress–strain data from Wisnom (1994) fitted by a Ramberg–Osgood law. The shear stress plotted refers to an axis system aligned with the fibers in a 45° ply and is therefore one-half the applied tensile stress. The strain is the engineering shear strain

strain levels in Fig. 22, whereas the simulations presented below (e.g., Fig. 25) predict that plastic strains are small (<0.02) over most of the material. Furthermore, the domain in Fig. 25 that contains larger plastic strains (but still <0.04) is small in spatial extent compared to the predicted spacing of transverse intraply cracks. Therefore, the mechanism of continuously distributed plasticity represents phenomena that are either occurring at strains beneath those at which large-scale cracks act in the calibration test or are too spatially confined to be included in the effects of transverse cracks. Therefore, including shear plasticity and transverse cracks separately in the simulations does not lead to double counting of a single phenomenon. Consistently, the characteristic of the data of Fig. 22 that dominates the fitting procedure of the calibration is the strain at which the knee occurs (≈0.02).

The data of Fig. 22 are well approximated by the Ramberg–Osgood law

$$\gamma_{12} = \mathrm{sgn}[\tau_{12}]\left(\frac{|\tau_{12}|}{G_{12}^0} + \alpha\frac{\tau_{12}^0}{G_{12}^0}\left(\frac{|\tau_{12}|}{\tau_{12}^0}\right)^n\right) \tag{43}$$

where G_{12}^0 is the undamaged in-plane shear modulus and τ_{12}^0 is the limiting elastic shear stress.

Calibration: Calibration of a constitutive law for distributed nonlinear shear deformation was achieved by fitting Eq. 43 to the data from Wisnom (1994). In the fitting function, the undamaged in-plane shear modulus G_{12}^0 was assigned the value 4.34 GPa and the limiting elastic shear stress τ_{12}^0 (which is related to the yield or flow stress) was assigned the value 30 MPa, using data from Wisnom (1994). With G_{12}^0 and τ_{12}^0 chosen thus, fitting Eq. 43 to the data of Fig. 22 yielded $\alpha = 0.01$ and $n = 6.8$. The stress–strain data are fitted within 2 % over the entire test range.

Cohesive Model for Delamination, Splitting, and Transverse Intraply Cracks

The mixed-mode cohesive model detailed in section "Nonlinear Fracture Models for Composite Materials" was employed to represent the nonlinear process zone in all types of cracks. A triangular traction–separation relation is assumed for each mode. Thus, for each mode, two key parameters, the cohesive strength and the fracture toughness, are needed to be calibrated.

In applying the cohesive law to splitting cracks and transverse ply cracks, which may arise at any location, an initiation criterion must also be stated. The initiation criterion is specified using parameters of the cohesive law, because it refers to the same material physics. Once any crack is initiated, its propagation is governed by the cohesive law.

In applying the cohesive law to delamination cracks, no initiation criterion need be specified, because the cohesive elements used to model delamination are already present on the known potential fracture plane.

The initiation criterion used for splitting and transverse cracks accounts for the effects of tension normal to the fracture plane and in-plane shear, viz.,

$$(\langle \overline{\sigma}_{\mathrm{n}} \rangle / \hat{\sigma}_n)^2 + (\overline{\tau}_s / \hat{\tau}_s)^2 = 1 \tag{44}$$

where $\overline{\sigma}_{\mathrm{n}}$ and $\overline{\tau}_s$ are the normal tensile stress and the in-plane shear stress, averaged over an element, with the coordinates (x_1, x_2) aligned with the local fiber direction and lying in the plane of the laminate. The Macaulay bracket $\langle \bullet \rangle = \max[\bullet, 0]$ appears so that normal compression ($\overline{\sigma}_{\mathrm{n}} < 0$) cannot induce crack initiation. The form of Eq. 44 reflects the fact that, in laminated polymer composites, splitting and transverse cracks tend to follow the local fiber direction (Fig. 1a) and propagate under constrained, mixed-mode conditions. The material parameters, $\hat{\sigma}_n$ and $\hat{\tau}_s$, are the mode-I and mode-II cohesive strength, respectively, in the calibrated cohesive law.

Calibration: Cohesive laws and thus the parameters appearing in the initiation criterion were calibrated by using published data to specify the material toughness (the area under the cohesive law) for each of mode I and mode II and the material strengths $\hat{\sigma}_n$ and $\hat{\tau}_s$ (Eq. 44 and Fig. 2). The mode-III law was set to be identical to the mode-II law. Given the area under the curve and the maximum traction $\hat{\sigma}_n$ or $\hat{\tau}_s$, the maximum displacement for nonzero tractions is also determined.

Identical calibration parameters were assumed for all three crack types, delamination cracks, splitting cracks, and transverse ply cracks.

The mode-I toughness of the ply material (E-Glass/913) was determined empirically using tensile fracture tests on single plies with the load axis perpendicular to the fiber direction, yielding $\Gamma_{\mathrm{IC}} = 250$ J/m^2 (Hallett and Wisnom 2006a). The mode-II toughness for a laminate made from the same materials was determined by analyzing end-notch flexure delamination tests, yielding two reported values $\Gamma_{\mathrm{IIC}} =$ 900 and 1,040 J/m^2 (Hallett and Wisnom 2006b, c). The baseline calibration values were taken to be $\Gamma_{\mathrm{IC}} = 250$ and $\Gamma_{\mathrm{IIC}} = 900$ J/m^2, with expected variance of 10 %.

In the subject problem, the mode-I strength, $\hat{\sigma}_n$, primarily influences transverse ply cracks, since both splitting and delamination cracks are shear dominated. In the layup studied, the transverse 90° plies are surface plies constrained only on the side that is bonded to the interior 0° plies. Therefore, transverse ply cracks appear as surface channeling cracks, whose behavior has been studied by both the composites and thin film communities (Dvorak and Laws 1987; Thouless 1990; Camanho et al. 2006; Davila et al. 2005). According to Camanho et al. (2006), the local

stress in the transverse ply required to initiate a surface channeling crack is approximately

$$\hat{\sigma}_n = \sqrt{\frac{4\Gamma_{Ic}}{\pi h(1/E_2 - \nu_{21}{}^2/E_1)}} \tag{45}$$

where h is the ply thickness (0.127 mm). Since the elastic constants are known, Eq. 45 predicts $\hat{\sigma}_n$ from the calibrated value of the mode-I toughness, yielding $\hat{\sigma}_n =$ 141 MPa. However, this value should be regarded as an upper-bound estimate, because Eq. 45 assumes a perfectly bonded interface between neighboring plies. Furthermore, the transverse strength of unidirectional plies of E-Glass/913 exhibits specimen-to-specimen discrepancy of ~15 % (Wisnom and Jones 1996). Summing these considerations, the value 120 MPa was used as a baseline value in the initiation criterion, Eq. 44, and in the mode-I cohesive law, with expected variance of 15 %.

This same calibration was also used for the mode-I component of the cohesive zone in splitting cracks. Since the splitting crack is an embedded crack (in the central 0° plies), its configuration is a tunneling crack, rather than the channeling crack to which Eq. 45 refers. However, in the particular layup studied here, the 0° plies are exactly twice the thickness of the 90° plies; and the effect of the different thickness on the expected critical stress exactly cancels the difference expected due to the fact that the splitting crack tunnels rather than channels (Camanho et al. 2006). Therefore, use of the same mode-I calibration is consistent.

Prior estimates of the shear cohesive strength, $\hat{\tau}_s$, for the subject E-Glass/913 material have been based on the shear deformation data of Fig. 22, with the value 75 MPa picked out from Hallett and Wisnom (2006a) and 70 MPa from Wisnom and Chang (2000). The uncertainty in the appropriate value reflects the range of the shear stress beyond the value at the knee of the data (Fig. 22). In this work, $\hat{\tau}_s$ was assigned the baseline calibration value of 80 MPa, with expected uncertainty of 10 %. The same value was used for all crack types.

The set of baseline calibration values for cohesive law parameters is summarized in Table 1. Variation of the parameters from these values will be made to assess sensitivities.

Numerical Prediction and Validation

Crack Pattern, Spacing, and Delamination Shape

The experimental images taken at three different load levels show the evolution of three major cracking systems (Fig. 23a). The damage evolution predicted at the same load levels using the baseline calibration values of Table 1 for the cohesive parameters is shown in Fig. 23b. In this figure, the splitting crack is shown as a solid white line, the transverse cracks as fine white or red lines, and the delamination by the black zone attached to the splitting crack. The A-FEM simulation successfully reproduces all the important features of the three cracking systems. The self-similar

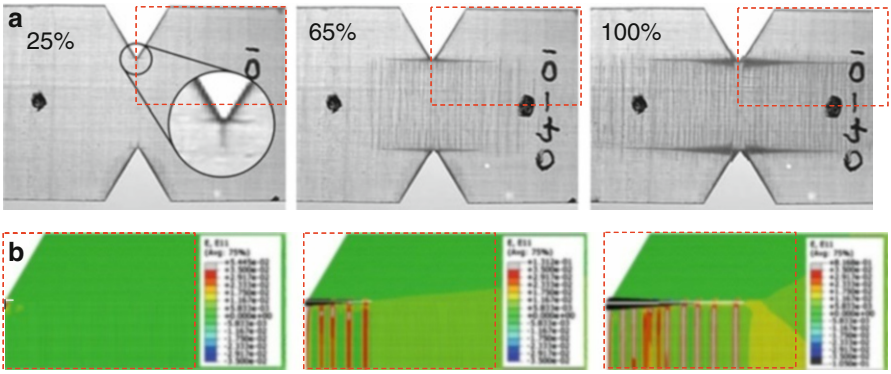


Fig. 23 (**a**) Three snapshots of the DNT specimen under different load levels (25 %, 65 %, and 100 %), showing the evolution of the coupled transverse cracks (*vertical lines*), splitting cracks (*horizontal line*), and delamination zones (*triangular dark zones*). (**b**) The A-FEM predicted crack evolution at approximately the same load levels (Reprinted from Journal of the Mechanics and Physics of Solids, vol 59, X.J. Fang, Z.Q. Zhou, B.N. Cox, Q.D. Yang, *High-fidelity simulations of multiple fracture processes in a laminated composite in tension*, Pages No. 1355–1373, Copyright 2011, with permission from Elsevier)

delamination zone angle is predicted to be 8°, in agreement with the experimental value of 7–10°. The splitting and multiple transverse cracks are also well captured.

The predicted transverse crack density is ∼1 crack/mm, which is about half of the measured density. This can be explained as a consequence of assuming symmetry in the damage in simulations. In experiments, transverse cracks are spawned from each of the two splitting cracks, so that the final crack density in the region between the two splitting cracks can be up to twice the initiation density along either crack. This possible doubling effect is absent from the simulations, because only a quarter of the specimen, containing one splitting crack, was modeled.

Load and Splitting Crack Length Versus Applied Displacement

The simulations predict the entire stress–displacement curve up to final failure, and the splitting crack length as a function of applied nominal stress, to within anticipated material variance.

The predicted stress as a function of the displacement is compared directly with experimental data in Fig. 24a. The simulation with arbitrary cracking reproduces the experimental curve very well: it follows the entire experimental curve, including the mild nonlinearity beyond the initial elastic regime, and, for this particular specimen, predicts the ultimate strength quite accurately.

The splitting crack growth as a function of applied stress is shown in Fig. 24b. In this figure, each of the experimental data points is an average of five tests reported in Hallett and Wisnom (2006b). The predicted splitting crack growth curve with all mechanisms enabled in the simulations again follows closely the experimental data during the entire loading stage.

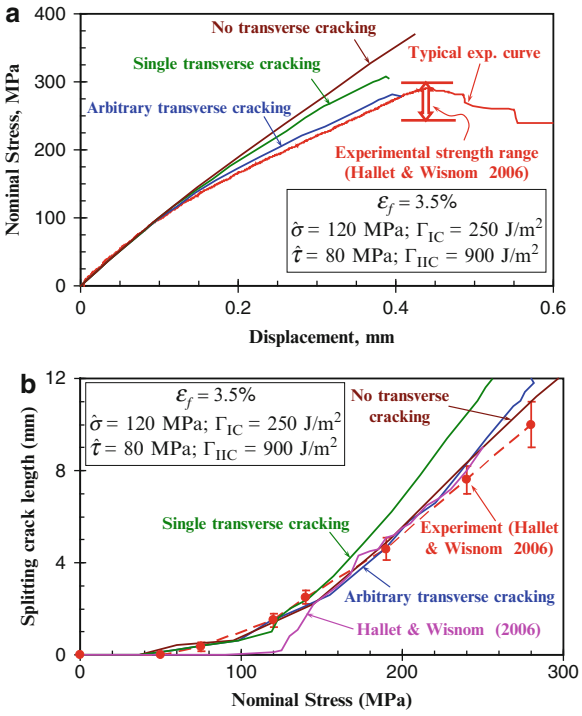


Fig. 24 Comparison of experimental results reported in Hallett and Wisnom (2006b) with predictions. (**a**) Nominal stress versus displacement curves for three cases: (1) with arbitrary transverse cracks, (2) with single transverse cracks at notch tip, and (3) no transverse crack. (**b**) Splitting crack length as functions of nominal stress for the same three cases. The numerical results obtained in Hallett and Wisnom (2006a) using preplanted CZM elements at the notch tip are also shown (Reprinted from Journal of the Mechanics and Physics of Solids, vol 59, X.J. Fang, Z.Q. Zhou, B.N. Cox, Q.D. Yang, *High-fidelity simulations of multiple fracture processes in a laminated composite in tension*, Pages No. 1355–1373, Copyright 2011, with permission from Elsevier)

Finally, based on the above-demonstrated high-fidelity simulation, investigation of damage coupling and sensitivity to cohesive failure parameters can be conducted in detail, and such a detailed parametric study can be found in Fang et al. (2011a).

Damage Coupling and Sensitivity to Cohesive Parameters

Influence of Transverse Ply Cracks

To show the necessity of including transverse cracks in strength analyses, predictions of stress versus strain are also shown in Fig. 24a for two modified cases, viz., (1) without considering transverse cracking and (2) with only one transverse crack permitted (a simplification considered in Wisnom and Chang 2000; Hallett and Wisnom 2006a). Whereas the prediction with arbitrary transverse cracking is accurate, if no or only one transverse crack is allowed, the global stiffness reduction beyond the proportional limit cannot be adequately captured.

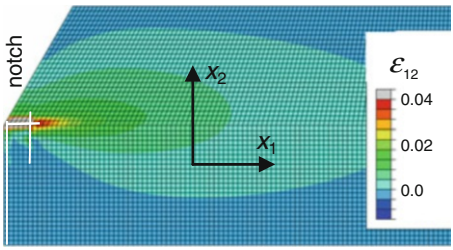


Fig. 25 Snapshot of distributed fine-scale shear deformation in the 90° ply. The strains are meaningful ahead of the tip of the splitting crack (*horizontal white line*). Two transverse cracks (*vertical white lines* – the one at notch tip is not visible) are present at this load level. Similar distributions are also predicted in the 0° ply. The large apparent shear strains (*grey zones*) immediately adjacent to the splitting crack are an artifact of a plotting deficiency of ABAQUS and should be ignored (Reprinted from Journal of the Mechanics and Physics of Solids, vol 59, X.J. Fang, Z.Q. Zhou, B.N. Cox, Q.D. Yang, *High-fidelity simulations of multiple fracture processes in a laminated composite in tension*, Pages No. 1355–1373, Copyright 2011, with permission from Elsevier)

In contrast, the propagation of the splitting crack is not very sensitive to the transverse crack details, with similar predictions for arbitrary transverse cracking, a single transverse crack, or no transverse cracks (Fig. 24b). This perhaps explains why the simulation in Hallett and Wisnom (2006b), in which a single cohesive crack was preplanted at the notch tip, also did a reasonable job of capturing the overall splitting crack growth. However, if cohesive elements are not placed accurately at the location of maximum stress, splitting crack initiation can be significantly delayed (Fig. 24b). Further simulations in the present study confirmed the sensitivity of splitting crack initiation to assumed location.

Influence of Shear Nonlinearity

Fine-scale distributed shear deformation influences all fracture modes. The shear strain distributions in the surface 90° ply are shown in Fig. 25, with splitting and transverse cracks superimposed. The largest shear strains are concentrated in a narrow region that extends ahead of the splitting crack tip. As demonstrated by the single transverse crack seen in Fig. 25, which has just initiated and is incompletely propagated, any transverse crack is initiated in material that is already damaged by distributed shear deformation. Therefore, the transverse cracks are initiated under mixed-mode conditions, rather than the approximately mode-I conditions that would exist if they initiated in undamaged material far from the splitting crack.

When distributed shear deformation is omitted from simulations, the predicted damage pattern is qualitatively wrong. Figure 26c shows the coupled damage evolution for three load levels similar to those in Fig. 23b without shear nonlinearity. The delamination zone is much larger, fewer transverse cracks have developed, and they lag significantly behind the splitting crack tip. Without shear nonlinearity, energy release is directed more towards delamination, rather than to the initiation of transverse cracks. The extensive delamination results in less energy

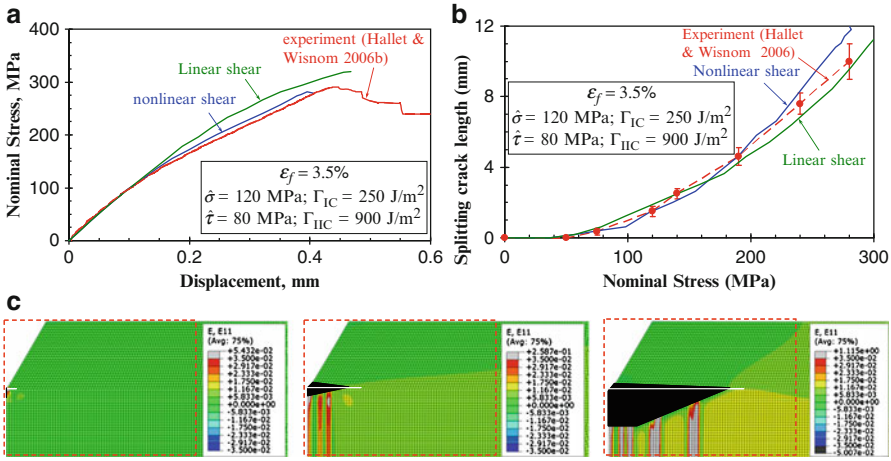


Fig. 26 Influence of shear nonlinearity on (**a**) the stress versus displacement relation and (**b**) splitting crack growth. (**c**) the A-FEM predicted crack evolution without shear nonlinearity at three load levels. The splitting crack is shown as the white *solid line*. The transverse cracks are indicated by the *white* or *red colored lines*, while the delamination is shown by the *black zone* attached to the splitting crack (Reprinted from Journal of the Mechanics and Physics of Solids, vol 59, X.J. Fang, Z.Q. Zhou, B.N. Cox, Q.D. Yang, *High-fidelity simulations of multiple fracture processes in a laminated composite in tension*, Pages No. 1355–1373, Copyright 2011, with permission from Elsevier.)

being available for driving splitting crack growth, and thus slower splitting crack growth at large stress levels. Similar numerical observations were reported in Wisnom and Chang (2000) for another material system (T300/914), although in that study transverse cracking was not considered.

Shear nonlinearity has a correspondingly strong influence on the stress–displacement curve (Fig. 26). Comparing the experimental data with the stress–displacement curves obtained with the calibrated shear nonlinearity included or excluded, one may conclude that the initial specimen stiffness is not very sensitive to shear nonlinearity (Fig. 26a). This is plausible because localized shear around the splitting crack is not expected to contribute significantly to the global stiffness of the specimen. However, the decrease in specimen stiffness beyond the proportional limit is poorly replicated if shear nonlinearity is not modeled. This may be due to transverse cracks not being reproduced faithfully, because it has been established previously that the gradual decrease of specimen stiffness is due largely to transverse cracking.

The splitting crack growth curve obtained without shear nonlinearity is compared in Fig. 26b with the curve obtained from the calibrated shear nonlinearity and the experimental curve. With shear nonlinearity, the splitting crack grows even faster at relatively large applied stress, which may be surprising because shear nonlinearity should relieve strain energy and lower the crack driving force for the splitting crack. However, the greater effect is the reduction of delamination when

shear nonlinearity is included, which keeps more energy available for splitting crack growth. This has also been noted recently in Van de Meer et al. (2010).

Influence of Cohesive Strengths

The influences of tensile and shear cohesive strength on the specimen stiffness and splitting crack growth are shown in Fig. 27. Figure 27a shows the stress–displacement curves obtained with tensile strength varying from $\hat{\sigma}_n =$ 50 to 150 MPa and shear strength from $\hat{\tau}_s = 75$ to 100 MPa.

The ultimate strength increases with $\hat{\sigma}_n$. For example, when $\hat{\sigma}_n = 50$ MPa, the predicted strength is 235 MPa. As $\hat{\sigma}_n$ increases to 150 MPa, the strength is increased to 306 MPa. A smaller tensile cohesive strength results in a larger transverse crack density and earlier development of transverse cracking near the notch tip (Fig. 27b), which shifts more load from 90° plies to the 0° plies. A larger stress concentration in the 0° plies leads to earlier fiber rupture. From the discussion presented above on calibration procedures, a reasonable estimate of the uncertainty in $\hat{\sigma}_n$ is $\pm$ 20 MPa.

The shear cohesive strength does not affect the stress–displacement relation significantly although a slight increase of stiffness is associated with shear strength increase from 75 to 85 MPa. However, the splitting crack growth rate is very sensitive to the shear strength. This is shown in Fig. 27b: the larger the shear cohesive strength, the slower the splitting crack growth. This is attributed to the relatively small hardening modulus in the shear stress–strain relation (Fig. 22): a small increase of shear strength requires a fairly large increase in the shear strain required for criticality in the elements that are to generate the splitting crack. This in turn requires a larger applied stress to propagate the splitting crack.

Influence of Mode-I and Mode-II Fracture Toughness Values

The mode-I and mode-II toughness values were varied by 50 % from the baseline calibration values, with the constraint that the toughness values in modes II and III remained equal. Changing mode-I toughness (Γ_{IC}) around the calibration value by $\pm$50 % does not significantly change the stress–strain relation or the splitting crack

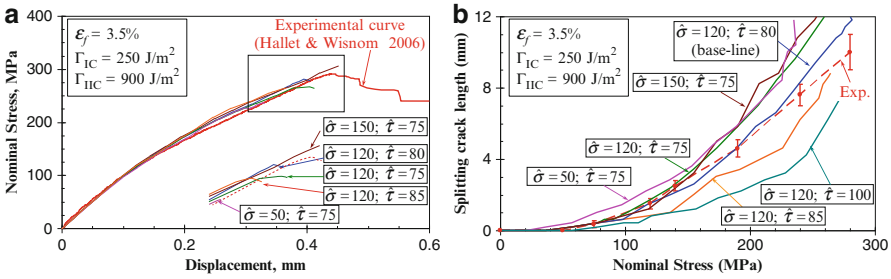


Fig. 27 Influence of cohesive strength on (**a**) nominal stress–displacement relation and (**b**) on splitting crack growth (Reprinted from Journal of the Mechanics and Physics of Solids, vol 59, X.J. Fang, Z.Q. Zhou, B.N. Cox, Q.D. Yang, *High-fidelity simulations of multiple fracture processes in a laminated composite in tension*, Pages No. 1355–1373, Copyright 2011, with permission from Elsevier)

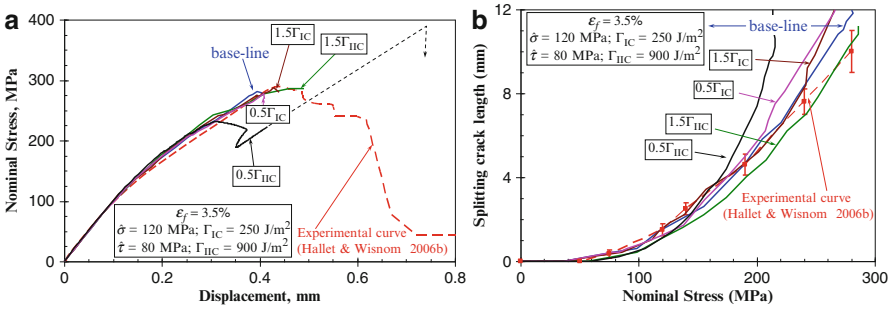


Fig. 28 Influence of fracture toughnesses, varied around baseline values (*inset*), on (**a**) nominal stress–displacement relation and (**b**) on splitting crack growth (Reprinted from Journal of the Mechanics and Physics of Solids, vol 59, X.J. Fang, Z.Q. Zhou, B.N. Cox, Q.D. Yang, *High-fidelity simulations of multiple fracture processes in a laminated composite in tension*, Pages No. 1355–1373, Copyright 2011, with permission from Elsevier)

growth rate and has a mild influence on ultimate strength (Fig. 28). In regard to transverse ply cracking, changing Γ_{IC} alone is equivalent, for the subject layup, to changing the specimen thickness while keeping other material parameters fixed (Eq. 45); the effect is similar to changing the cohesive strength. However, the influence on ultimate strength of changing Γ_{IC} by ± 50 % is much smaller than changing the normal cohesive strength by the same proportion (Eq. 45).

The mode-II fracture toughness affects both the splitting crack, which is approximately a shear (mode-II) crack, and delamination. The splitting crack propagates at a rate that is negatively correlated with the value of Γ_{IIc} (Fig. 28b), as also noted by Wisnom and colleagues in their study that employed preplanted cohesive zone elements for the splitting crack and a single transverse crack (Wisnom and Chang 2000; Hallett and Wisnom 2006a).

In the present study, lowering Γ_{IIc} by 50 % not only caused more rapid splitting crack growth (Fig. 28b) but also led to a transition in the failure sequence because of its effect on delamination. Early splitting crack growth was accompanied by rapid spreading of a delamination towards the center of the specimen (bottom edge in Fig. 23b) as well as along the split. The rapid delamination deters the development of transverse cracking, which remain few and concentrated near the notch tip with a density of ~ 0.8 mm^{-1} (slightly smaller than ~ 1.0 mm^{-1} for the baseline case). Shortly after the delamination spreads across the specimen, a steady-state delamination front is formed, which propagates towards the loading edge in a numerically unstable fashion. The above process is illustrated by the four snapshots in Fig. 29. In these figures, the fiber-direction strains (ε_{11}) in the 0° ply at four different stress levels are shown as contour plots. The delamination fronts and splitting crack tips are indicated. Throughout this coupled splitting and delamination process, the maximum fiber-direction strain within the delamination zones is ~ 2.2 %, well below the fiber-failure strain of 3.5 % (Table 1).

This coupled damage evolution is different from all other cases and results in a feature-rich, non-monotonic load–displacement curve (Fig. 28a). The initial

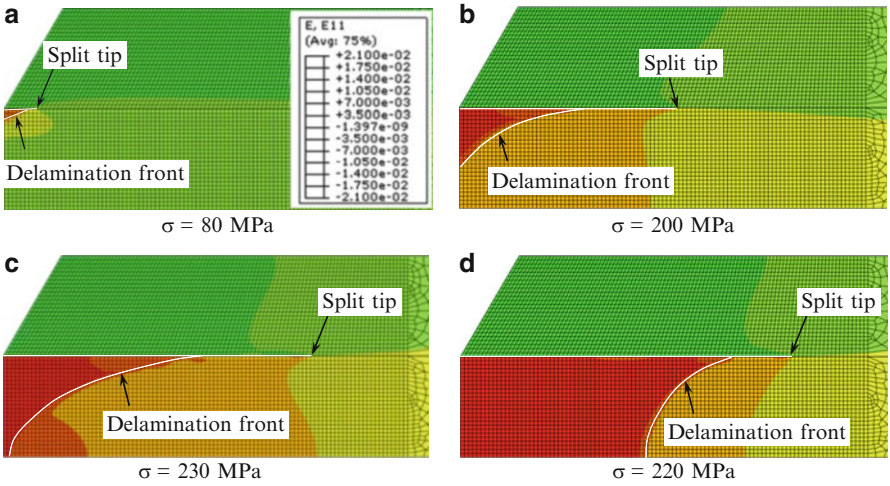


Fig. 29 Coupled evolution of splitting delamination in a specimen with Γ_{IIc} being reduced by 50 % (0.5 Γ_{IIc}). Compared to the baseline results in Fig. 26, both the splitting crack growth rate and delamination rate are much higher at similar stress levels. The delamination zone is no longer a sharp triangular wedge (**a**, **b**). Rather, it propagates quickly across the specimen width bounded by the splitting crack (**b**, **c**) and quickly establishes a steady-state delamination front that propagates towards the loading edge (**d**). The final stage is numerically unstable under displacement-controlled loading (Reprinted from Journal of the Mechanics and Physics of Solids, vol 59, X.J. Fang, Z.Q. Zhou, B.N. Cox, Q.D. Yang, *High-fidelity simulations of multiple fracture processes in a laminated composite in tension*, Pages No. 1355–1373, Copyright 2011, with permission from Elsevier)

softening of the curve is similar to other cases. However, at a stress level of about 230 MPa, the load decreases as the delamination front is established. The snapback in the load–displacement curve corresponds to the delamination front passing the displacement recording points (dots in Fig. 23). As the delamination front and splitting crack propagate further away from the notch tip (with the splitting always well ahead of the delamination front; see Fig. 29), the section of the 0° ply that passes between the notches is increasingly decoupled from the rest of the composite and carries an increasingly complete share of the load. The stress concentration from the notch tips disappears. While the simulation fails due to numerical instability, the anticipated specimen response and ultimate failure can be estimated as that of a single 0° ply with a width of 0.5 W (10 mm), which is indicated by the dashed line in Fig. 28a, terminating at the fiber-failure strain $\varepsilon_f = 3.5$ % ($\sigma_f = 385$ MPa).

Further Discussions

Mesh Requirements

Mesh independence is achieved provided the element size is less than the length of the nonlinear fracture process zones of the different crack systems.

When cohesive parameters are varied so that the process zone length becomes smaller, mesh refinement may be necessary to maintain accuracy. One example of this requirement was observed in the study case where Γ_{IIC} was reduced by 50 % from the best-calibration value: the mesh size also had to be reduced to capture the transition in failure mode and expected reduction of ultimate strength correctly.

Calibration: The calibration procedures used in this paper are not ideal, but are a pragmatic use of available data for the subject E-glass/913 composite materials, supplemented, in the absence of experiments for calibrating the mode-I parameter, $\hat{\sigma}_n$, by the micromechanical model of Eq. 45. The search for tests for calibrating cohesive laws, with at least the peak traction determined as well as the fracture toughness values, remains an active research topic. In some applications, it may also be useful to determine characteristics of the shape of the cohesive law; what details of the shape influence engineering fracture behavior is not yet known, but can be addressed in due course by high-fidelity simulations such as those presented here.

Statistical Variance in Model Parameters and Covariance with Engineering Properties

By considering either the experiments that yielded calibration data or the assumptions that underlie Eq. 45, estimates have been made of the uncertainty in the calibration values of the parameters in the cohesive laws for mode-I and mode-II fracture. These are listed in Table 2. Since they arise from deviations in the data from tests on typical specimens, the variances are likely to be material variations, rather than errors in analysis that could be reduced by more accurate procedures. If so, they imply unavoidable variance in predicted engineering properties, which will be manifested as deviance in the test data used to qualify materials. Such deviance is essential information for design.

Table 2 Estimates of uncertainty in cohesive law parameters and the covariance matrix between the set of parameters and three key metrics of composite failure (Reprinted from Journal of the Mechanics and Physics of Solids, vol 59, X.J. Fang, Z.Q. Zhou, B.N. Cox, Q.D. Yang, *High-fidelity simulations of multiple fracture processes in a laminated composite in tension*, Pages No. 1355–1373, Copyright 2011, with permission from Elsevier)

		Sensitivity matrix, $\frac{\theta_0}{\phi_0}\frac{\partial\phi}{\partial\theta}$		
		ϕ = engineering property; θ = cohesive parameter		
Cohesive parameter	Estimated uncertainty in cohesive parameter (%)	Ultimate strength	Global stiffness in nonlinear regime	Splitting crack growth rate
Γ_{IC}	10	0.1	Small	0
Γ_{IIC}	10	0.25[a]	Small	-1[a]
$\hat{\sigma}_n$	15	0.3	0.1	0
$\hat{\tau}_s(=\hat{\tau}_t)$	10	0.3	0.2	-1

[a]very nonlinear dependence

The expected variance in predicted properties can be estimated by varying the cohesive parameters used in simulations and recomputing predictions, thus evaluating the partial derivatives

$$\frac{\theta_0}{\phi_0}\frac{\partial \phi}{\partial \theta} \tag{46}$$

where θ denotes a cohesive law parameter (Γ_{IC}, Γ_{IIC} etc.), θ_0 its best-choice calibration value, ϕ a property (ultimate strength, strain etc.), and ϕ_0 its value when all cohesive law parameters take their best-choice calibration values. The partial derivatives of Eq. 46 are measures of the sensitivity of engineering properties to the material properties of fracture mechanisms. Normalization with respect to θ_0 and ϕ_0 allows the sensitivity measures to be independent of units. A matrix of their estimated values appears in Table 2 for the following engineering properties: the composite ultimate strength, the global stiffness averaged over the nonlinear regime, and the growth rate of the splitting cracks averaged as the crack length grows from zero to its maximum. Other engineering properties might be chosen, but these three cover the main features of the global performance of this material in the notched tension test. The sensitivity matrix is only qualitatively indicative of trends in performance and does not substitute for complete analysis of different cases. For variations in Γ_{IIC}, in particular, the transition in failure sequence described above leads to highly nonlinear dependence in the selected engineering properties; for example, the ultimate strength is almost independent of Γ_{IIC} for increasing Γ_{IIC} above the best-calibration value, but drops significantly for decreasing Γ_{IIC}, when it is measured by the peak load prior to unstable behavior (Fig. 28a).

The predicted variance of any quantity in experimental tests should be predicted by the product of the estimated variance in a cohesive parameter and the pertinent element of the sensitivity matrix. Combining the effects of variance in all four cohesive parameters, Table 2 predicts, for example, that the variance in composite ultimate strength should be approximately 10 %. The actual variance reported for double-notched tension tests of the subject composites is close to this: the strengths reported by Hallett and Wisnom (2006b) range from 243 to 291 MPa.

Transverse and Delamination Crack Interactions

In the present simulations, it is implicitly assumed that once a transverse crack is initiated, it immediately propagates as a steady-state crack, because the in situ strength is derived from steady-state crack analysis. For polymer matrix composites, which typically have mode-II toughness several times larger than mode-I toughness, this is justified. A recent 3D analysis of an individual transverse crack, which includes the influence of possible accompanying delamination cracks, has shown that if $\Gamma_{IIC}/\Gamma_{IC} > 1$, the initiation stress is only 5 % less than the steady-state propagation stress (c.f. Zhou et al. 2010 and works cited therein). In the same analysis of Zhou et al. (2010), it has also been shown that if $\Gamma_{IIC}/\Gamma_{IC} > 1$, the in-plane stress required for any delamination initiated by a transverse crack to propagate away from the transverse crack is significantly higher than the transverse

crack propagation stress. Consistently, for the best-calibration cohesive parameters, the present simulations show that no delaminations were triggered by transverse cracks. However, when a Γ_{IIC} was reduced by 50 % in the present simulations, large-scale delamination occurred with minimal transverse ply cracking.

Concluding Remarks

In this paper, the urgent need of advanced numerical methods that can explicitly resolve the multiple-damage processes and their nonlinear coupling at various scales is highlighted. The recent development of advanced numerical methods that can represent various composite damage modes explicitly with embedded nonlinear fracture models such as cohesive zone models has been reviewed. These methods include the eXtended Finite Element Method (X-FEM), the phantom node methods (PNM), and the Augmented Finite Element Method (A-FEM).

The elemental formulation of the A-FEM has been presented as a standard element that is fully compatible with any FE programs including commercial codes without access to source code. The capability of the A-FEM to account for arbitrary, nonlinearly coupled, multiple-cracking systems typical in laminated composites with stress concentrators has been demonstrated by replicating the damage systems observed in several composite tests including (1) the three-point bending of PMCs, (2) the crack jumping in a single-cantilever composite beam, and (3) the in-plane tension tests of an orthogonal double-notched specimens. Mesh independence is achieved provided the element size is less than the length of the nonlinear fracture process zones of the different crack systems. The required mesh size can be estimated in advance using analytical results for the process zone lengths.

A combination of independent data and a micromechanical model was used to calibrate parameters in cohesive models that are used to represent nonlinear fracture processes and the distributed nonlinear shear deformation. With this calibration, quantitative agreement is obtained to within experimental variance between measured and predicted fracture test behavior, including the entire nonlinear stress–strain curve, the ultimate strength, the growth rates of delamination and splitting cracks, and the density of transverse ply cracks.

Analyses show that the variance in fracture properties expected for the estimated variance in material parameters matches the variance in the measured ultimate strength. Thus, future elaboration of the simulations developed here can be the basis for statistical predictions of material performance.

Comparative simulations show unequivocally the necessity of including all the major damage processes explicitly to reproduce all observed damage characteristics. If incomplete nonlinear processes were included, a simulation may still be able to reproduce some aspects of the experiments, but not all in a consistent way. For example, if the shear nonlinearity is not considered, one may still obtain the stress–displacement curve and splitting crack growth curve more or less right, but the delamination is predicted incorrectly. If multiple transverse ply cracking is not

included, the predicted delamination and splitting crack growth may still be close to experimental measurements, but the gradual stiffness reduction in the stress–displacement curve would be missed.

References

D.F. Adams, T.R. King, D.M. Blackketter, Evaluation of the transverse flexure test method for composite materials. Compos. Sci. Technol. **39**, 341–353 (1990)

G. Bao et al., The role of material orthotropy in fracture specimens for composites. Int. J. Sol. Struct. **29**, 1105–1116 (1992)

G.I. Barenblatt, The formation of equilibrium cracks during brittle fracture: general ideas and hypotheses, axially symmetric cracks. Appl. Math. Mech. **23**, 622–636 (1959)

G.I. Barenblatt, The mathematical theory of equilibrium cracks in brittle fracture, in *Advances in Applied Mechanics*, ed. by H.L. Dryden, T. Von Karman (Academic, New York, 1962), pp. 55–129

Z.P. Bazant, J. Planas, *Fracture and Size Effect in Concrete and Other Quasibrittle Materials* (CRC Press, Boca Raton, 1998)

P.P. Camanho, C.G. Davila, M.F. De Moura, Numerical simulation of mixed-mode progressive delamination in composite materials. J. Compos. Mater. **37**, 1415–1438 (2003)

P.P. Camanho et al., Prediction of in situ strengths and matrix cracking in composites under transverse tension and in-plane shear. Compos. Part A Appl. Sci. Manuf. **37**, 165–176 (2006)

A. Carpinteri, G. Colombo, Numerical analysis of catastrophic softening behaviour(snap-back instability). Comput. Struct. **31**, 607–636 (1989)

A. Carpinteri, G. Ferro, Fracture assessment in concrete structures, in *Concrete Structure Integrity*, ed. by I. Milne, R.O. Ritchie, B. Karihaloo (Elsevier Science, Amsterdam, 2003)

S.W. Case, K.L. Reifsnider, *MRLife 12 Theory Manual – Composite Materials* (Materials Response Group, Virginia Polytechnical Institute and State University, Blacksburg, 1999)

J.L. Chaboche, P.M. Lesne, J.F. Maire, Continuum damage mechanics, anisotropy and damage deactivation for brittle materials like concrete and ceramic composites. Int. J. Damage Mech. **4** (1), 5–22 (1995)

J.L. Chaboche, R. Girard, P. Levasseur, On the interface debonding models. Int. J. Damage Mech. **6**, 220–256 (1997)

K.Y. Chang, S. Liu, F.K. Chang, Damage tolerance of laminated composites containing an open hole and subjected to tensile loadings. J. Compos. Mater. **25**, 274–301 (1991)

H.Y. Choi, F.K. Chang, A model for predicting damage in graphite/epoxy laminated composites resulting from low-velocity point impact. J. Compos. Mater. **26**, 2134–2169 (1992)

G. Clark, Modeling of impact damage in composite laminates. Composites **20**, 209–214 (1989)

A. Corigliano, Formulation, identification and use of interface models in the numerical analysis of composite delamination. Int. J. Sol. Struct. **30**, 2779–2811 (1993)

B.N. Cox, Q.D. Yang, In quest of virtual tests for structural composites. Science **314**, 1102–1107 (2006)

W.C. Cui, M.R. Wisnom, N. Jones, Failure mechanisms in three and four point short beam bending tests of unidirectional glass/epoxy. J. Strain. Anal. **27**(4), 235–243 (1992)

C.G. Davila, P.P. Camanho, C.A. Rose, Failure criteria for FPR laminates. J. Compos. Mater. **39**, 323–345 (2005)

R. de Borst, Numerical aspects of cohesive-zone models. Eng. Fract. Mech. **70**, 1743–1757 (2003)

R. de Borst et al., On gradient-enhanced damage and plasticity models for failure in quasi-brittle and frictional materials. Comput. Mech. **17**(1–2), 130–141 (1995)

R. de Borst, J.J.C. Remmers, A. Needleman, Mesh-independent discrete numerical representations of cohesive-zone models. Eng. Fract. Mech. **73**(2), 160–177 (2006)

J. Dowlbow, M. A. Kahaleel, J. Mitchell, *Multiscale Mathematics Initiative: A Roadmap. A Report to Department of Energy* Report PNNL-14966 (2004)
D.S. Dugdale, Yielding of steel sheets containing slits. J. Mech. Phys. Sol. **8**, 100–104 (1960)
G.J. Dvorak, N. Laws, Analysis of progressive matrix cracking in composite laminates. II. First ply failure. J. Compos. Mater. **21**, 309–329 (1987)
M. Elices et al., The cohesive zone model: advantages, limitations and challenges. Eng. Fract. Mech. **69**, 137–163 (2002)
X.J. Fang, Q.D. Yang, B.N. Cox, An augmented cohesive zone element for arbitrary crack coalescence and bifurcation in heterogeneous materials. Int. J. Numer. Meth. Eng. **88**, 841–861 (2010)
X.J. Fang et al., High-fidelity simulations of multiple fracture processes in a laminated composites in tension. J. Mech. Phys. Sol. **59**, 1355–1373 (2011a)
X.J. Fang et al., An augmented cohesive zone element for arbitrary crack coalescence and bifurcation in heterogeneous materials. Int. J. Numer. Meth. Eng. **88**, 841–861 (2011b)
A. Fawcett, J. Trostle, S. Ward, in *International Conference on Composite Materials*, Gold Coast, 1997
S.F. Finn, Y.F. He, G.S. Springer, Delaminations in composite plates under transverse impact loads – experimental results. Compos. Struct. **23**, 191–204 (1993)
J. Fish, A. Ghouali, Multiscale analysis sensitivity analysis for composite materials. Int. J. Numer. Meth. Eng. **50**, 1501–1520 (2001)
C. Gonzalez, J. LLorca, Multiscale modeling of fracture in fiber-reinforced composites. Acta Mater. **54**, 4171–4181 (2006)
S. Goutianos, B.F. Sorensen, Path dependence of truss-like mixed mode cohesive laws. Eng. Fract. Mech. **91**, 117–132 (2012)
S. Hallett, M.R. Wisnom, Numerical investigation of progressive damage and the effect of layup in notched tensile tests. J. Compos. Mater. **40**, 1229–1245 (2006a)
S.R. Hallett, M.R. Wisnom, Experimental investigation of progressive damage and the effect of layup in notched tensile tests. J. Compos. Mater. **40**, 119–141 (2006b)
A. Hansbo, P. Hansbo, A finite element method for the simulation of strong and weak discontinuities in solid mechanics. Comput. Meth. Appl. Mech. Eng. **193**, 3523–3540 (2004)
M.-Y. He, J.W. Hutchinson, Crack deflection at an interface between dissimilar materials. Int. J. Sol. Struct. **25**, 1053–1067 (1989)
A. Hillerborg, M. Modéer, P.E. Peterson, Analysis of crack propagation and crack growth in concrete by means of fracture mechanics and finite elements. Cement. Concr. Res. **6**, 773–782 (1976)
E.V. Iarve, D. Mollenhauer, R. Kim, Theoretical and experimental investigation of stress redistribution in open-hole composite laminates due to damage accumulation. Compos. Part A **36**, 163–171 (2005)
H.M. Inglis et al., Cohesive modeling of dewetting in particulate composites: micromechanics vs. multiscale finite element analysis. Mech. Mater. **39**, 580–595 (2007)
P.M. Jelf, N.A. Fleck, The failure of composite tubes due to combined compression and torsion. J. Mater. Sci. Lett. **29**, 3080 (1994)
A.S. Kaddorur, M.J. Hinton, P.D. Soden, A comparison of the predictive capabilities of current failure theories for composite laminates: additional contributions. Compos. Sci. Technol. **64**, 449–476 (2004)
M.S. Kafkalidis et al., Deformation and fracture of an adhesive layer constrained by plastically-deforming adherends. Int. J. Adhes. Sci. Technol. **14**, 1593–1646 (2000)
M. Kumosa, G. Odegard, Comparison of the +/−45 tensile and Iosipescu shear tests for woven fabric composites. J. Compos. Technol. Res. **24**, 3–15 (2002)
P. Ladeveze, Multiscale modelling and computational strategies. Int. J. Numer. Meth. Eng. **60**, 233–253 (2004)
I. Lapczyk, J. Hurtado, Progressive damage modeling in fiber-reinforced materials. Compos. Part A **38**, 2333–2341 (2007)

F. Laurin, N. Carrere et al., A multi-scale progressive failure approach for composite laminates based on thermodynamical viscoelastic and damage models. Compos. Part A **38**, 198–209 (2007)
D.S. Ling, Q.D. Yang, B.N. Cox, An augmented finite element method for modeling arbitrary discontinuities in composite materials. Int. J. Fract. **156**, 53–73 (2009)
D.S. Ling et al., Nonlinear fracture analysis of delamination crack jumps in laminated composites. J. Aerosp. Eng. **24**, 181–188 (2011)
J. LLorca, C. González, Multiscale modeling of composite materials: a roadmap towards virtual testing. Adv. Mater. **23**, 5130–5147 (2011)
P. Maimi et al., A continuum damage model for composite laminates: Part I – Constitutive model. Mech. Mater. **39**, 897–908 (2007)
A. Matzenmiller, J. Lubliner, R.L. Taylor, A constitutive model for anisotropic damage in fiber-composites. Mech. Mater. **20**, 125–152 (1995)
L.N. McCartney, Physically based damage models for laminated composites. J. Mater. Des. Appl. **217**(3), 163–199 (2003)
J. Mergheim, E. Kuhl, P. Steinmann, A finite element method for the computational modeling of cohesive cracks. Int. J. Numer. Meth. Eng. **63**, 276–289 (2005)
N. Moës, T. Belytschko, Extended finite element method for cohesive crack growth. Eng. Fract. Mech. **69**, 813–833 (2002)
N. Moes, J. Dolbow, T. Belytschko, Finite element method for crack growth without remeshing. Int. J. Numer. Meth. Eng. **46**, 131–150 (1999)
A. Needleman, An analysis of decohesion along an imperfect interface. Int. J. Fract. **42**, 21–40 (1990)
T.K. O'Brien, S.A. Salpekar, Scale effects on the transverse tensile strength of carbon/epoxy composites. Compos. Mater. Test. Des. **11**(ASTM STP 1206), 23–52 (1993)
T.K. O'Brien et al., Influence of specimen configuration and size on composite transverse tensile strength and scatter measured through flexure testing. J. Compos. Technol. Res. **25**, 50–68 (2003)
J.T. Oden, K. Vemaganti, N. Moes, Hierarchical modeling of heterogeneous solids. Comput. Method. Appl. Mech. Eng. **172**, 3–25 (1999)
J.T. Oden et al., *Simulation-Based Engineering Science – Revolutionizing Engineering Science through Simulation* (NSF, 2006)
C. Oskay, J. Fish, Eigendeformation-based reduced order homogenization for failure analysis of heterogeneous materials. Comput. Methods Appl. Mech. Eng. **196**, 1216–1243 (2007)
J. Parmigiani, M.D. Thouless, The roles of toughness and cohesive strength on crack deflection at interfaces. J. Mech. Phys. Sol. **54**, 266–287 (2006)
J. Parmigiani, M.D. Thouless, The effects of cohesive strength and toughness on mixed-mode delamination of beam-like geometries. Eng. Fract. Mech. **74**, 2675–2699 (2007)
S.T. Pinho, P. Robinson, L. Iannucci, Fracture toughness of the tensile and compressive fibre failure modes in laminated composites. Compos. Sci. Technol. **66**, 2069–2079 (2006)
S. Ramanathan, D. Ertaz, D.S. Fisher, Quasistatic crack propagation in heterogeneous media. Phys. Rev. Lett. **79**, 873–876 (1997)
J.N. Reddy, Multiscale computational model for predicting damage evolution in viscoelastic composites subjected to impact loading technical report to U.S. Army Research Office, 1-31 (2005)
J.J.C. Remmers, R. de Borst, A. Needleman, A cohesive segments method for the simulation of crack growth. Comput. Mech. **31**(1–2), 69–77 (2003)
S. Rudraraju et al., In-plane fracture of laminated fiber reinforced composites with varying fracture resistance: experimental observations and numerical crack propagation simulations. Int. J. Sol. Struct. **47**, 901–911 (2010)
S. Rudraraju et al., Experimental observations and numerical simulations of curved crack propagation in laminated fiber composites. Compos. Sci. Technol. **72**, 1064–1074 (2011)
K.W. Shahwan, A.M. Waas, Non-self-similar decohesion along a finite interface of unilaterally constrained delaminations. Proc. Roy. Soc. Lon. **A 453**, 515–550 (1997)
M.M. Shokrieh, L.B. Lessard, Progressive fatigue damage modeling of composite materials, Part I: Modeling. J. Compos. Mater. **34**(13), 1056–1080 (2000)

S.J. Song, A.M. Waas, Energy-based mechanical model for mixed mode failure of laminated composites. AIAA J. **33**, 739–745 (1995)
J.H. Song, P.M.A. Areias, T. Belytschko, A method for dynamic crack and shear band propagation with phantom nodes. Int. J. Numer. Meth. Eng. **67**, 868–893 (2006)
R. Talreja, Multiscale modeling in damage mechanics of composite materials. J. Mater. Sci. **41**, 6800–6812 (2006)
X.D. Tang et al., Progressive failure analysis of 2x2 braided composites exhibiting multiscale heterogeneity. Compos. Sci. Technol. **66**, 2580–2590 (2006)
T.-E. Tay, Characterization and analysis of delamination fracture in composites: an overview of developments from 1990 to 2001. Appl. Mech. Rev. **56**(1), 1–32 (2003)
M.D. Thouless, Crack spacing in brittle films on elastic substrates. J. Am. Ceram. Soc. **73**, 2144–2146 (1990)
M.D. Thouless, Q.D. Yang, A parametric study of the peel test. Int. J. Adhes. Adhes. **28**, 176–184 (2008)
A. Turon et al., A damage model for the simulation of delamination in advanced composites under variable-mode loading. Mech. Mater. **38**, 1072–1089 (2006)
A. Turon et al., An engineering solution for mesh size effects in the simulation of delamination using cohesive zone models. Eng. Fract. Mech. **74**, 1665–1682 (2007)
F.P. Van de Meer, L.J. Sluys, Continuum models for the analysis of progressive failure in composite laminates. J. Compos. Mater. **43**, 2131–2156 (2009a)
F.P. Van de Meer, L.J. Sluys, A phantom node formulation with mixed mode cohesive law for splitting in laminates. Int. J. Fract. **158**, 107–124 (2009b)
F.P. Van de Meer, C. Oliver, L.J. Sluys, Computational analysis of progressive failure in a notched laminate including shear nonlinearity and fiber failure. Compos. Sci. Technol. **70**, 692–700 (2010)
A.S.D. Wang, F.W. Crossman, Initiation and growth of transverse cracks and delaminations. J. Compos. Mater. **14**, 71–87 (1980)
J.S. Wang, Z. Suo, Experimental determination of interfacial toughness using Brazil-nut-sandwich. Acta Metall. **38**, 1279–1290 (1990)
M.R. Wisnom, The effect of fibre rotation in +/−45 degree tension tests on measured shear properties. Composites **26**, 25–32 (1994)
M.R. Wisnom, F.-K. Chang, Modelling of splitting and delamination in notched cross-ply laminates. Compos. Sci. Technol. **60**, 2849–2856 (2000)
M.R. Wisnom, M.I. Jones, Size effects in interlaminar tensile and shear strength of unidirectional glass fibre/epoxy. J. Reinf. Plast. Compos. **15**, 2–15 (1996)
D. Xie et al., Discrete cohesive zone model to simulate static fracture in 2D tri-axially braided carbon fiber composites. J. Compos. Mater. **40**, 2025–2046 (2006)
Q.D. Yang, B.N. Cox, Cohesive zone models for damage evolution in laminated composites. Int. J. Fract. **133**(2), 107–137 (2005)
Q.D. Yang, M.D. Thouless, Mixed mode fracture of plastically-deforming adhesive joints. Int. J. Fract. **110**, 175–187 (2001a)
Q. Yang, M.D. Thouless, Mixed mode fracture of plastically-deforming adhesive joints. Int. Fract. **110**, 175–187 (2001b)
Q.D. Yang, M.D. Thouless, S.M. Ward, Numerical simulations of adhesively-bonded beams failing with extensive plastic deformation. J. Mech. Phys. Sol. **47**, 1337–1353 (1999)
Q.D. Yang, M.D. Thouless, S.M. Ward, Elastic–plastic mode-II fracture of adhesive joints. Int. J. Sol. Struct. **38**, 3251–3262 (2001)
Q.D. Yang et al., Fracture and length scales in human cortical bone: the necessity of nonlinear fracture models. Biomaterials **27**, 2095–2113 (2006a)
Q.D. Yang et al., Re-evaluating the toughness of human cortical bone. Bone **38**, 878–887 (2006b)
Q.D. Yang et al., An improved cohesive element for shell delamination analyses. Int. J. Numer. Meth. Eng. **83**(5), 611–641 (2010)
Q.D. Yang et al., Virtual testing for advanced aerospace composites: advances and future needs. J. Eng. Mater. Technol. **133**, 11002–11008 (2011)

Q.D. Yang, X. J. Fang, Revisiting crack kinking in cohesive materials. Unpublished results, 2013
T. Ye, Z. Suo, A.G. Evans, Thin film cracking and the roles of substrate and interface. Int. J. Sol. Struct. **29**, 2639–2648 (1992)
Z. Zhang, Z. Suo, Split singularities and the competition between crack penetration and debond at a bimaterial interface. Int. J. Struct. **44**, 4559–4573 (2007)
Z.Q. Zhou et al., The evolution of a transverse intra-ply crack coupled to delamination cracks. Int. J. Fract. **165**, 77–92 (2010)

Section XI

Nuclear Damage Characterization

Irradiation Damage in Nuclear Power Plants

43

Wolfgang Hoffelner

Contents

W. Hoffelner (✉)
RWH consult GmbH, Oberrohrdorf, Switzerland

Nuclear Energy and Safety Research Department, Formerly, Laboratory for Nuclear Materials, Paul Scherrer Institute (PSI), Villigen PSI, Switzerland
e-mail: wolfgang.hoffelner@rwh.ch; wolfgang.hoffelner@bluewin.ch

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_36

Abstract

The chapter gives an introduction into the main processes occurring in metals and alloys under neutron irradiation. Displacement damage, phase reactions, swelling, irradiation creep, and transmutation are the main physical effects changing mechanical properties and microstructure of materials used in nuclear power plants. As results radiation hardening/embrittlement, enhanced stress corrosion cracking, changes in geometry, degradation of creep properties, and other damage can occur. Consequences of such material degradation are discussed for current nuclear plants where a 50-year operation experience exists. Advanced, future nuclear plants are expected to consist also almost exclusively of metals and alloys, and they are expected to undergo in principle the same types of damage. However, changes in other operational parameters (higher temperatures, fast neutrons, other coolants) might change also the degree of radiation damage (e.g., thermal creep in addition to radiation creep). Advanced modeling and testing techniques can be considered as a tool to balance the missing long-term experience with next-generation nuclear plants.

Introductory Remarks

Central components of nuclear plants are usually exposed to the coolant, radiation, and elevated temperatures. These conditions lead to degradation of components during service and limit therefore the lifetime of plants. In current light water reactors (LWRs), embrittlement of the reactor pressure vessel, irradiation-assisted stress corrosion cracking of reactor internals, and irradiation creep of claddings are typical degradation mechanisms caused by neutrons. Advanced reactors, like Generation IV plants, are expected to be exposed to more damaging fast neutrons (higher energy), higher temperatures, and coolants different from water. Although the physics behind radiation damage is expected to remain the same, some differences in damage development between LWRs and advanced reactors can be expected.

In the first part of this chapter, the basic phenomena of irradiation damage of reactor materials will be discussed. The focus will be on metals and alloys which are not only the key materials in current reactors but which will stay also key materials for future reactors. Considerations for ceramics will remain limited to graphite (as moderator for British advanced gas reactors and future high-temperature gas-cooled reactors) and to SiC/SiC composites which are considered for control rod parts or claddings in advanced plants.

In the second part of this chapter, radiation damage occurring in current nuclear plants will be highlighted with examples from reactor pressure vessels, reactor internals, and fuel claddings. Expectations for future plants will be briefly touched upon. Advanced methods of materials science will be introduced as a tool for better understanding of irradiation-related risks in future reactors.

Phenomenology of Radiation Damage

Introduction

Radiation damage is basically the result of the interaction of energetic particles with matter. The consequences of these interactions which depend primarily on energy can be manifold. The considerations of this chapter will be limited mainly to neutrons and ions on the particle side and to metals and alloys and selected ceramics on the matter side. Only the most important effects will be discussed; for more details, I would like to refer the reader to the numerous very good textbooks in this field, e.g., Schilling and Ullmaier (1994), Ullmaier and Schilling (1980), and Was (2007). The types of damage are (Schilling and Ullmaier 1994):

- Elastic collisions where bombarding particles (neutrons, ions, electrons) transfer recoil energy T to the lattice atoms. If T exceeds the threshold energy T_{th} for displacement, a vacancy-interstitial pair (Frenkel defect) is created.
- Nuclear reactions where fast particles produce considerable concentrations of foreign elements within the material. In particular, the inert gas helium which is produced by (n,α) reactions plays an important role for the behavior of metals and alloys under fast-neutron irradiation.
- Electronic excitations are of only very limited importance for metals and the irradiation damage process considered here.

The results of these damage events have an impact on radiation-exposed components like reactor pressure vessels (RPVs) or reactor internals (including fuel claddings). A summary of the damage events in the material and its consequences for components is given in Table 1. The different effects and its influence on the materials will be discussed more in detail in the following.

Types of Radiation Damage

Displacement Damage

Displacement damage starts usually with a bombarding particle that transfers a recoil energy T by elastic collision to a lattice atom. If the recoil energy exceeds a material-dependent threshold energy for displacement, E_{th}, the atom jumps from its original site to an interstitial position creating a vacancy-interstitial pair which is called "Frenkel pair." If the recoil energy is significantly higher than E_{th} (e.g., in the case of fast neutrons), the atom firstly hit by the neutron, the "primary knock-on atom" (PKA) or "primary recoil atom" (PRA) is able to transfer energy by moving further into the crystal creating further Frenkel pairs and a so-called displacement cascade (see Fig. 1). When the energetic particle is heavy and energetic enough, and the material is dense, the collisions between the atoms may occur so near to each other that they cannot be considered independent of each other. In this case the

Table 1 Different types of radiation damage and resulting technical consequences (replotted from Hoffelner (2012))

Effect	Consequence in material	Kind of degradation in component
Displacement damage	Formation of point defect clusters and dislocation loops	Hardening, embrittlement
Irradiation-induced segregation	Diffusion of detrimental elements to grain boundaries	Embrittlement, grain boundary cracking
Irradiation-induced phase transitions	Formation of phases not expected according to phase diagram, phase dissolution	Embrittlement, softening
Swelling	Volume increase due to defect clusters and voids	Local deformation, eventually residual stresses
Irradiation creep	Irreversible deformation	Deformation, reduction of creep life
Helium formation and diffusion	Void formation (inter- and intracrystalline)	Embrittlement, loss of stress rupture life, and creep ductility

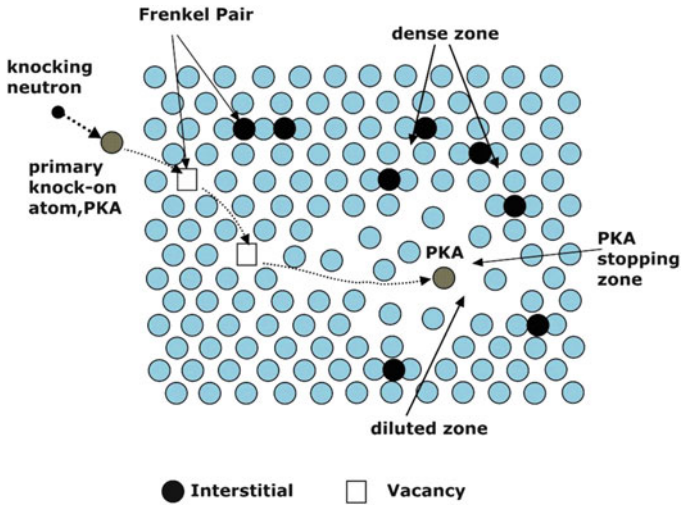


Fig. 1 Development of a collision cascade. The primary knock-on atom starts to move as a result of the energy transfer from the neutron. It creates Frenkel pairs, and it finally ends in a damaged zone with a diluted part where many vacancies exist and a dense part where many interstitials exist (Replotted from Hoffelner (2012), source Seeger (1962))

process becomes a very complicated many-body interaction between very many atoms which can only be treated with molecular dynamics modeling. A heat spike is created which is characterized by the formation of a transient diluted region in the center of the cascade and densified region around it. After the cascade, the densified

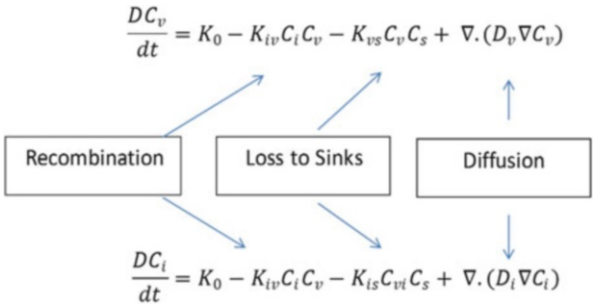


Fig. 2 Point defect rate equations

region becomes a region of interstitial defects, and the diluted region typically becomes a region of vacancy defects.

Elastic collisions produce radiation damage in three different process regimes at different timescales :

- The initial stage of the radiation damage process ($t < 10^{-8}$ s)
- Physical effects of radiation damage ($t > 10^{-8}$)
- Mechanical response of the material to radiation-induced effects

The initial stage of the radiation damage process has just been discussed. The phase of development of physical effects is mainly determined by diffusion and reactions of the created point defects. They can recombine, form agglomerations, or diffuse to sinks. The development of the concentrations of vacancies C_v and of interstitials C_i can be described by rate equations as shown schematically in Fig. 2. These rate equations can be solved for different boundary condition which allows a prediction of the development of the radiation-induced microstructure (Wiedersich 1991a, b). The supersaturation of point defects in irradiated matter leads to high diffusion coefficients already at lower temperatures. With increasing temperature the influence of irradiation on diffusion diminishes, and for temperatures higher than about 600 °C, thermal diffusion becomes the relevant diffusion process in steels. This can be seen from Fig. 3 (Zinkle et al. 1993) where the development of displacement-induced defects is shown as a function of temperature taking an austenitic steel as an example. Although there are some differences for different classes of metals and steels, the main message that point defects can agglomerate to different defects being obstacles for dislocation movement and that displacement damage vanishes with increasing temperature remains valid for several metals and alloys.

Units for Irradiation Damage

Before discussing further radiation-induced defects, a few aspects about the measure for damage relevant doses shall be mentioned. Very often particle flux which refers to the number of particles passing through an area in a certain interval of time (commonly measured in neutrons/cm^2·s) or particle fluence (neutron flux integrated over a certain time period, measured in neutrons/cm^2) are used. However,

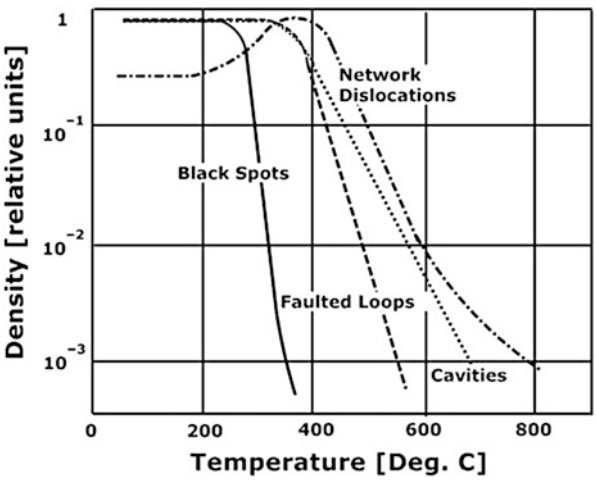


Fig. 3 Influence of irradiation temperature on the formation of different obstacles taking an austenitic steel as an example (Source Zinkle et al. (1993))

investigations of radiation hardening of annealed 316 stainless steel showed that even for the same type of material, different results were found when correlating radiation hardening (change in yield stress) with neutron fluence (Greenwood 1994). Therefore, another measure for radiation exposure or dose is frequently used which is based on the total number of displacements that the PKA will create in the solid. An important quantity is the number of displacements per unit time and per unit volume produced by a flux, $\Phi(E_i)$, of incoming particles of energy E_i. The displacement rate or number of displacements per atom (dpa) per unit time (dpa/s) describes in good approximation the energy-dependent response of an irradiated material. Typical displacement rates in reactors are 10^{-9}–10^{-7} dpa/s. Using this unit hardening for 316 stainless steel, mentioned before, could be very well correlated (Greenwood 1994).

Point Defect-Related Irradiation Damage Other Than Displacement Damage

In the previous section damage due to production, diffusion, and agglomeration of different point defects was considered. The high point defect supersaturation can also lead to other phenomena known from thermal diffusion processes like:

- Radiation-induced segregation
- Radiation-induced precipitation
  - Incoherent precipitate nucleation
  - Coherent precipitate nucleation
- Radiation-induced dissolution
- Radiation-induced phase reactions
  - Radiation disordering
  - Metastable phases
  - Amorphization

Radiation-Induced Segregation (RIS)

Thermal-induced segregation is a temperature-dependent redistribution of alloy constituents at point defect sinks such as grain boundaries. Temper embrittlement of steels is a very well-known example for segregation-related deterioration of toughness. Elements like phosphorus, sulfur, or manganese diffuse to grain boundaries. The cohesion along the grain boundaries is weakened leading to a reduction of toughness (reduction of fracture toughness or increase of ductile-to-brittle fracture appearance transition temperature). Such grain boundaries can also act as preferential corrosion sites leading to stress corrosion cracking as discussed later.

Radiation-induced segregation describes a similar effect, however, driven by radiation-induced point defects. A flow of vacancies into one direction is equivalent to material flow into the opposite direction. It can be understood in terms of the so-called inverse Kirkendall effect (Marwick 1978). This inverse Kirkendall effect refers to cases where an existing flux of point defects affects the interdiffusion of atoms of type A and atoms of type B. Irradiation segregation in a homogeneous AB alloy occurs because the irradiation has produced excess point defects which lead to a point defect flux. Figure 4 explains the mechanism for a binary alloy more in detail. The ordinates represent the concentrations of vacancies and interstitials, respectively, in arbitrary units. The x-axis gives the distance from the grain boundary. Movement of a vacancy into one direction is equivalent with the movement of an atom into the other direction. Therefore, the arrow of the vacancy flow J_v points in another direction than the arrows of the material flows J_A and J_B. In case of the movement of interstitial atoms, the directions of J_i and J_A and J_B are the same. The differences in the diffusion coefficients of A and B lead to a dilution of the

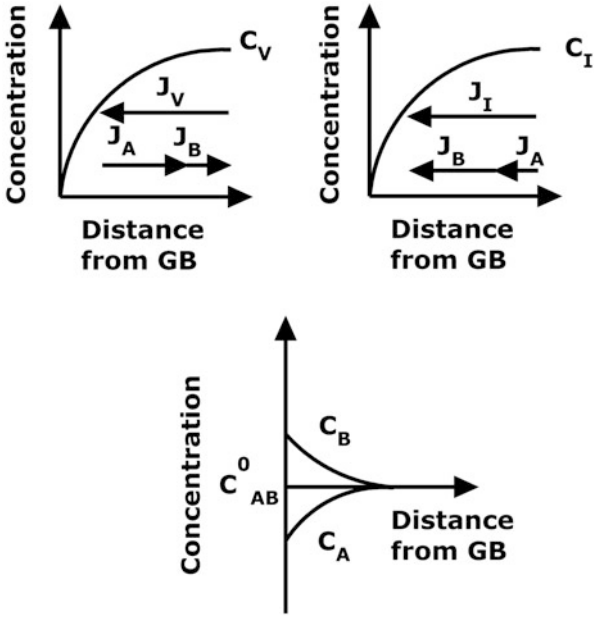


Fig. 4 Principle of radiation-induced segregation for a binary alloy (After Was (2007))

concentration of atoms A and to an increase of the concentration of atoms of type B towards the grain boundaries. As an example a diminishing chromium concentration combined with an increasing nickel concentration at the grain boundaries was found for proton-irradiated austenitic steel (304 SS) (Bruemmer et al. 1999a). Radiation-induced segregation depends on the temperature as well as on dose rate as a diffusion-driven effect. Once the temperature is too low, vacancies can move only slowly, and recombination will become the predominant mechanism. At temperatures where thermal effects become important, radiation effects become negligible. Radiation-induced segregation can therefore only happen in a temperature window between these two conditions (Was et al. 2006). Radiation-induced segregation plays an important role for irradiation-assisted stress corrosion cracking in light water reactors as will be discussed later.

Radiation-Induced Phase Transformations

Other diffusion-controlled irradiation phenomena are radiation-induced phase transformations or phase reactions which can result in precipitation of phases not expected at operation temperatures, dissolution of phases, and amorphization of phases. The driving force behind these microstructural changes is – like for RIS – the presence of large supersaturation of point defects, especially at temperatures between 250 and 550 °C or the inverse Kirkendall effect. Irradiation-induced point defect sinks like interstitial loops, helium bubbles, and voids can also give raise to precipitation.

Radiation-induced precipitation is one phenomenon belonging to this class of radiation damage. Coherent and incoherent precipitates can be formed. Coherent particles fully or partially match the lattice structure of the matrix, and incoherent particles don't. Coherent particles act as sinks for solute atoms, whereas incoherent particles allow solute atoms to be trapped and also to be released (Was 2007). Figure 5a shows an example of radiation-induced precipitates. In this figure also radiation-induced voids are visible. This phenomenon will be discussed in the next section under "void swelling." Radiation-induced nanosized precipitates

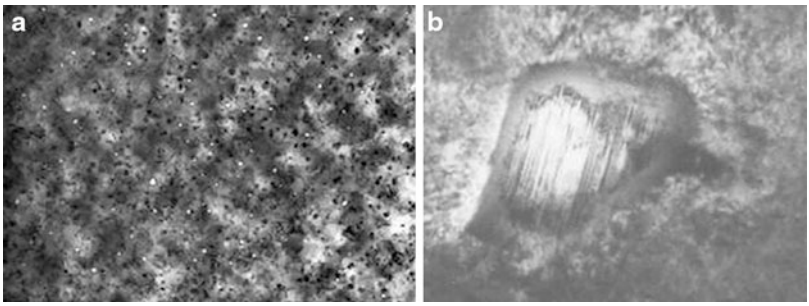


Fig. 5 (**a**) Concurrent void formation and $M_{23}C_6$ precipitation after irradiation at a PWR-relevant dpa rate of 1.8×10^{-7} dpa/s in the experimental breeder reactor II EBR-II in Argonne at 379 °C, after Isobe et al. (2008). (**b**): $Zr(Cr, Fe)_2$ precipitate neutron irradiated at a temperature of 510 K to a fluence of 8 dpa, showing formation of amorphous layer Motta et al. (1991)

(e.g., nanoclusters) are responsible for an increase in strength acting as obstacles for dislocation movement. This increase in strength reduces in turn ductility and toughness of the material which can be considered as life-limiting materials degradation as shown later for reactor pressure vessels.

Radiation-induced dissolution means that due to the presence of a high density of point defects, particles start to dissolve. This is rather similar to processes happening without irradiation during a solution treatment of alloys at high temperatures.

Amorphization: Amorphous metals do not have an ordered atomic-scale structure. They can be produced by very rapid cooling, and they are often referred to as metallic glasses. Amorphization can also occur during mechanical alloying or physical vapor deposition. Radiation-induced amorphization is a result of the radiation-induced high point defect density. Amorphization under irradiation not only happens for metals and alloys, it can also be found in intermetallics and ceramics like graphite or silicon carbide. Figure 5b (Motta et al. 1991) shows a partially amorphized second phase particle in a zirconium-based LWR cladding material (Zircaloy). Amorphization occurs often together with decomposition or dissolution of second phase particles in Zircaloys. The related change in matrix composition can improve the oxidation behavior of these materials in reactor environments.

Radiation-induced disordering can happen in an ordered lattice like an intermetallic phase or alloy when radiation-induced diffusion processes support lattice disordering already below temperatures where it would occur due to thermal reasons.

Metastable phases can form when thermal and radiation-induced phase formations are in competition and that phases appear at conditions where they should be thermally not stable or they disappear when they should remain thermally stable.

The Production of Foreign Atoms

Radiation-induced microstructural changes discussed until now happen at lower temperatures, and they disappear once the temperature exceeds about 600 °C. Radiation-induced production of foreign atoms (see, e.g., Schilling and Ullmaier (1994)) is another important type of radiation damage. Particularly interesting are reactions where gases are generated (e.g., alpha particles or protons) which can further react with the material. This is very important because gaseous atoms, especially helium (i.e., α), can seriously degrade the long-term mechanical integrity of some reactor components. This has already been recognized in the mid-1960s of the last century during the development of alloys for core components of fast breeder reactors (Barnes 1965; Harries 1966). Nuclear reactions where helium can be produced in metals (M) by fast neutrons (n^f) can be the following:

$$ {}^{A}_{Z}M + {}^{1}_{0}n^{f} \rightarrow {}^{A-3}_{Z-2}M' + {}^{4}_{2}He \quad \text{(some MeV)} $$

$$ {}^{A}_{Z}M + {}^{1}_{0}n^{f} \rightarrow {}^{A-4}_{Z-2}M'' + {}^{1}_{0}n^{f} + {}^{4}_{2}He \quad \text{(some MeV)} $$

Nickel has the highest cross sections for such reactions, and the problems for fast reactors even increase for fusion reactors which will be exposed to 14 MeV neutrons.

Also thermal neutrons nth can lead to helium formation although to a lesser extent (lower cross sections). A typical reaction (2 steps) for thermal neutrons is the following:

$$^{58}_{28}Ni + ^{1}_{0}n^{\text{th}} \rightarrow ^{59}_{28}Ni + \gamma$$

$$^{59}_{28}Ni + ^{1}_{0}n^{\text{th}} \rightarrow ^{56}_{26}Fe + ^{4}_{2}He \quad (4.67\ \text{MeV})$$

The problem with helium gas in the metal is that it can form intragranular bubbles as well as intergranular bubbles. Intergranular bubbles lead to strong reduction of creep ductility and sometimes also of creep rupture time. This is the reason why nickel-base superalloys, which are basically the high temperature materials of choice, cannot (or only limited) be used for in-core applications at high temperatures in fast reactors.

Influence of Radiation on Mechanical Properties

In the previous sections the basic principles of radiation damage were discussed. The microstructural changes have an effect on the macroscopic behavior of materials and consequently also on the performance of components.

Strength and Toughness

The presence of radiation-induced obstacles for dislocation movement (point defect clusters, dislocation loops, stacking fault tetrahedra, helium-filled pores) has an influence on the mechanical properties. Radiation hardening is generally accompanied by a reduction in uniform elongation under tensile test conditions due to highly localized plastic flow. A second consequence of radiation hardening that is particularly important for body-centered cubic (BCC) alloys is reduction in fracture toughness and a potential shift in the ductile–brittle transition temperature to values that are above the operating temperature. Figures 6, 7, and 8 show examples for irradiation hardening and embrittlement. Stress–strain curves after irradiation at different temperatures are shown in Fig. 6. The curves were shifted along the strain axis to make the results better visible. In comparison with the yield stress of the un-irradiated material, a significant increase (up to more than a factor of 2) was found. Impact tests (shown in Fig. 7) reveal a very pronounced shift in the brittle-to-ductile transition temperature, and also the upper shelf energy is significantly reduced. At temperatures above 400 °C hardening starts to disappear as a result of annealing. Operation of structural materials in the "lower shelf" fracture toughness regime is usually not feasible based on safety considerations, because this could lead to premature shutdown of the reactor before the design operating lifetime is achieved as discussed later. Embrittlement can also be seen from the temperature dependence of the fracture toughness as shown in Fig. 8.

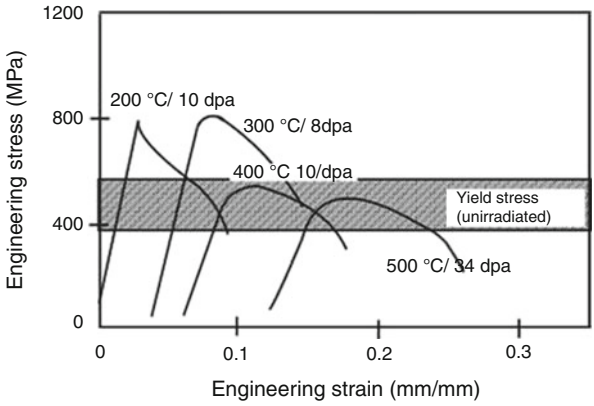


Fig. 6 Irradiation hardening in a ferritic–martensitic steel. At temperatures above 400 °C, hardening starts to disappear as a result of annealing (Replotted from Hoffelner (2012), source Robertson et al. (1997)

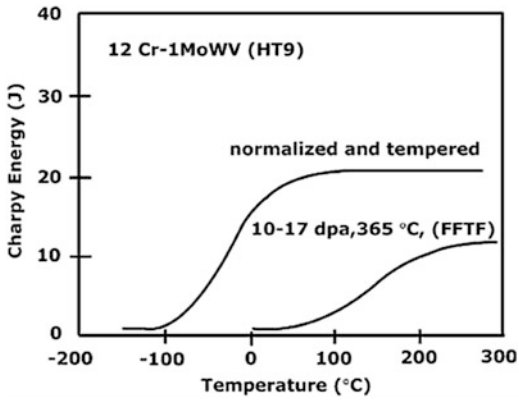


Fig. 7 Shift in the fracture appearance transition temperature as a result of irradiation embrittlement. FFTF: Fast Flux Test Facility, Hanford (Replotted from Hoffelner (2012), source Klueh and Alexander 1992)

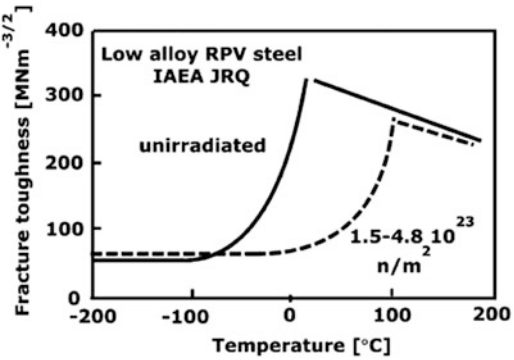


Fig. 8 Influence of neutron irradiation on the fracture toughness (Replotted from Hoffelner (2012). Source (Havel et al. 1993))

Radiation effects in ferritic–martensitic steels for temperatures where irradiation hardening/embrittlement occurs (T <= 450 °C) are well investigated. Analyses of embrittlement of steels which can be attributed to irradiation-enhanced precipitation are only scarcely available. An extended analysis of embrittlement in the

absence of radiation hardening for different steels was reported in Klueh et al. (2008). In this investigation nine different irradiated steels (ferritic–martensitic, ferritic, low activation) were analyzed that were embrittled in the absence of irradiation hardening at temperatures exceeding 450 °C. The embrittlement was attributed to irradiation-enhanced precipitation. Precipitates that were concluded to cause the observed behavior varied for the different steels and included $M_{23}C_6$, α', χ, and Laves phase. The observed effects were explained by postulating irradiation-enhanced or irradiation-induced precipitation and/or irradiation-enhanced precipitate coarsening that produced large precipitates acting as crack nuclei for fracture initiation.

Influence of Irradiation on Fatigue and Fatigue Crack Growth

Irradiation increases the yield strength, and it decreases the ductility of metallic materials. This leads to an increase of the fatigue limit (as a result of higher strength) at high number of cycles. However, under low cycle fatigue conditions, the number of cycles to failure decreases as a result of radiation-induced loss of ductility as shown in Fig. 9 and as discussed, e.g., in Hoffelner (2012) in more detail.

Fatigue crack growth rates as a function of the cyclic stress intensity range ΔK remain usually up to temperatures where the environment has only a negligible effect more or less independent from the temperature. Also microstructure has no very pronounced effect, and therefore, no significant effect of irradiation on fatigue crack growth rates is expected. This could be confirmed for an austenitic steel as shown in Fig. 10 (Lloyd et al. 1982). Similar insignificant influence of irradiation on fatigue crack growth rates also reported low-alloy reactor pressure vessel steels (James and Williams 1973).

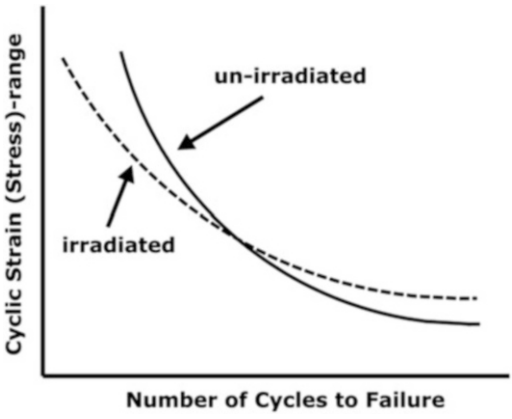


Fig. 9 The influence of irradiation on the fatigue curve (schematically). See also Hoffelner (2012)

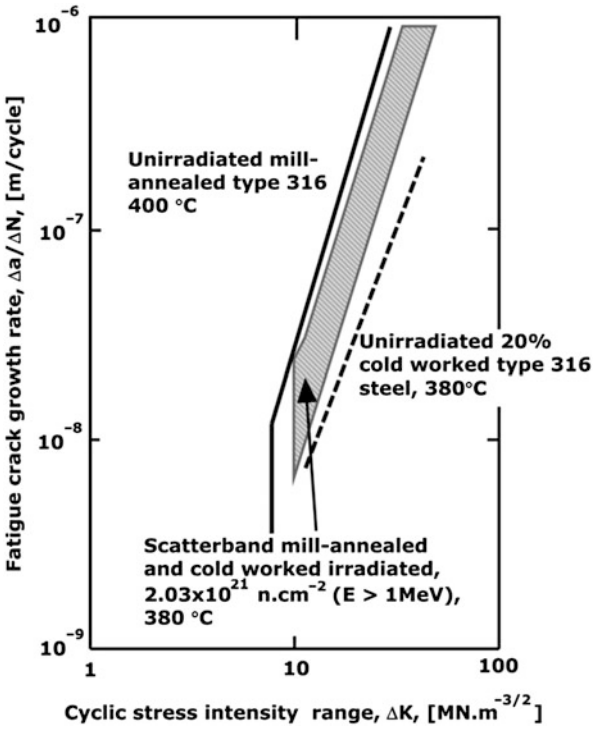


Fig. 10 Fatigue crack growth rates for different qualities of type 316 austenitic steel in un-irradiated and irradiated conditions (Replotted from Lloyd et al. (1982))

Influence of Irradiation on Thermal Creep

Application of load at elevated temperatures under radiation can lead to two types of creep: thermal creep and radiation-induced creep which will be discussed later. Degradation of stress rupture life as a result of pre-irradiation has been reported, e.g., in Bloom and Stiegler (1972). The technical relevance of such data is – as far as only displacement damage is concerned – questionable because usually in a nuclear plant irradiation and thermal creep happen synchronously. Specific attention must be paid to helium at high temperatures. The presence of helium bubbles at grain boundaries is expected to contribute synergistically to creep damage forming also voids along the same sites. Helium bubbles at grain boundaries can therefore deteriorate stress rupture ductility as well as creep rupture strength. Some in-pile creep data for an austenitic steel are shown in Fig. 11 from which the influence of irradiation creep becomes clearly visible (Puigh and Hamilton 1987). An exhaustive treatment of creep-irradiation interactions for an austenitic steel can be found in the literature (Wassiliew et al. 1986). This temperature-dependent damage pattern is also reflected in creep–fatigue interactions.

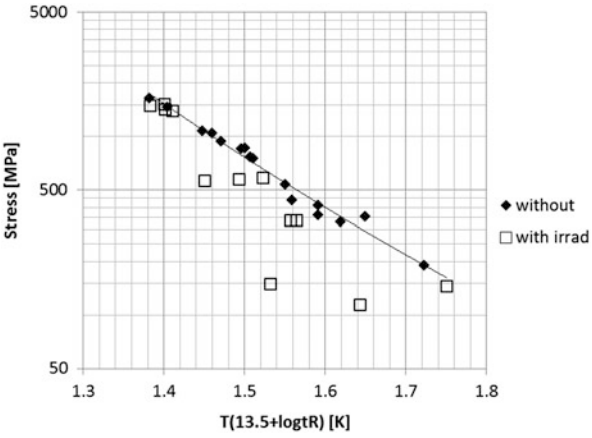


Fig. 11 Influence of irradiation on stress rupture performance of an austenitic steel. The influence of irradiation has been attributed to helium effects (Replotted from Hoffelner (2012), source Puigh and Hamilton (1987)

Radiation-Induced Dimensional Changes

Void Swelling

Voids or bubbles containing either vacuum (vacancy clusters) or gas (helium) can develop under irradiation in a crystal (see Fig. 5a as an example). According to Garner (2010a) one defining feature for discrimination between void or bubble is that bubbles tend to grow slowly by gas accumulation, while voids are either totally or partially vacuum filled, but which are free to grow rapidly via vacancy accumulation without further gas addition. Without going further into detail concerning the growth mechanisms, it is obvious that holes in a body usually increase its volume. Void swelling is the effect which leads to a three-dimensional change of the material during irradiation in a temperature interval of 0.3 $T_m <= T <= 0.5\ T_m$. Two phases must be considered for void formation: void nucleation and void growth (Russel 1971; Katz and Wiedersich 1971). The fact that voids form although its formation is energetically not very favorable is attributed to the fact that additional heterogeneities like very small helium gas bubbles are present during irradiation which promote clustering of vacancies. Void growth is quantitatively better understood than nucleation. In contrast to interstitials which tend to migrate to dislocations, the vacancies are rather attracted by voids. This net flux of vacancies to voids causes them to grow which leads macroscopically to swelling. Three stages can be discriminated: transient period, steady state swelling, and saturation. During the first period nucleated voids start to grow until a steady state is reached during which an almost linear relation between dose and volume swelling exists. With further increasing void size, the relative contribution of radiation-induced defects to macroscopic swelling decreases leading to saturation. The duration of the transient regime of swelling in austenitic and high-nickel steels is exceptionally sensitive to irradiation parameters, composition, heat treatment, and mechanical processing (Garner 2010a).

Irradiation Creep

Void swelling is a three-dimensional change of the volume which occurs without mechanical load. Superposition of radiation and mechanical load leads to deformation of the material at stresses far below the yield stress and at temperatures where thermal creep cannot be observed. An extensive review on irradiation creep phenomena is given in Garner (1994). For cold-worked and recrystallized austenitic steel, in-beam fatigue tests with hold times at 300 and 400 °C were performed, and a clear influence of the radiation was found which was attributed to irradiation creep–fatigue interaction (Scholz and Mueller 1996).

Figure 12 shows irradiation creep during helium implantation of the ferritic oxide dispersion strengthened (ODS) steel PM 2000 at 500 °C as an example. The response to thermal creep is also shown for comparison, and no indications for thermal creep can be seen.

Swelling and irradiation creep are not really separate processes. Both phenomena are caused by the presence of point defects as a result of radiation. While swelling attempts to be isotropic, irradiation creep redirects mass flow anisotropically. Irradiation creep can operate before the onset of swelling but is accelerated when swelling begins. Radiation creep is traditionally described in terms of transient contributions saturating at 1 dpa, stress-enhanced creep (proportional to void swelling), and the creep compliance B_0 in the absence of swelling. For many high exposure applications, the transient can be ignored (Garner 1994). Neglecting also possible effects related to the void swelling rate, the irradiation creep compliance B_0 remains the most important contribution. It can be written as:

$$\dot{\varepsilon} = B_0 \sigma K$$

which says that the irradiation creep rate $\dot{\varepsilon}$ is proportional to irradiation displacement damage rate K and to stress σ (at least for moderate stresses). It is interesting

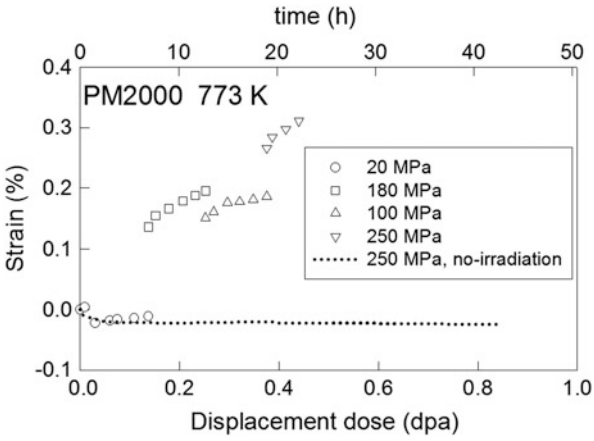


Fig. 12 Irradiation creep under helium implantation of a commercial ODS alloy Chen et al. (2008)

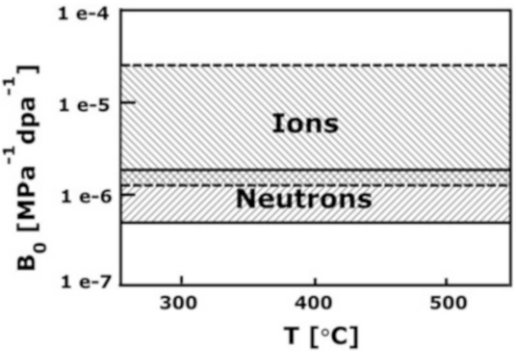


Fig. 13 Comparison of irradiation creep compliances B_0 as a function of irradiation temperature T. Ions refer to light ion irradiations (protons/helium) and represent the materials: ODS PM2000, 19Cr-ODS, ODS Ni-20Cr-1ThO2, and martensitic steel. Neutrons refer to neutron irradiations to doses below 25 dpa and above 25 dpa and represent the materials: ODS MA957, HT9, F82H, and Fe-16Cr (Sources Hoffelner (2012), Chen et al. 2010))

to notice that in this creep law, the stress exponent is 1 which is also the case for diffusion-controlled thermal creep. This is compatible with the fact that radiation creep is also a diffusion-controlled process. Irradiation creep is important for temperatures below which thermal point defects become predominant. This has been shown for austenitic and ferritic steels, and it was also found for advanced nuclear materials like ODS alloys or titanium aluminides. Some discussion is ongoing concerning the influence of type of the energetic particle on irradiation creep. Figure 13 compares irradiation creep compliances of several types of alloys. A typical value for the irradiation creep compliance of alloys under neutron irradiation is about 7.10^{-7} MPa^{-1}. dpa^{-1}. For light ions qualitatively a similar behavior was found; however, the average value was about five times higher. Possible reasons for this difference could be:

- Real influence of the type of radiation
- Radiation rate effect (because light ion irradiation is usually performed typically with 0.1 dpa/h compared with 0.003–0.004 dpa/h in a fast reactor)
- Dependence on total dose (ion irradiation tests go usually up to 1–2 dpa only)
- Influence of state of stress (multiaxiality)

Even if a quantitative explanation is still missing, it should be pointed out that the qualitative results are the same. This means that creep tests under ion irradiation allows a relative comparison between different materials which is very important for material development. Although the phenomenology of irradiation creep is quite consistent, a thorough physical understanding is still missing.

Radiation Damage in Nonmetallic Structural Materials

Graphite

Graphite is of concern for some reactor types like the British AGR or the high-temperature gas reactor. Therefore, graphite has been frequently investigated, and the mechanisms of the damage which graphite undergoes on neutron irradiation are quite well understood (IAEA 2000). However, many processes have not been correlated with the properties of the pristine graphites. In other words, the behavior of a new graphite cannot be quantitatively predicted. Certain behaviors may be anticipated, but this is an insufficient basis for a designer. This is the reason why worldwide projects on irradiation damage of graphite are underway. The basic radiation damage mechanisms for graphite are comparable with metals. A displacement cascade creates vacancies and interstitials which rearrange in the graphite lattice forming interstitial loops and vacancy loops (Ball 2008; Burchell 1999). The essential processes which happen in graphite under irradiation are the following (Fig. 14). As a result of vacancy creation and formation of vacancy clusters, the crystal undergoes an a-axis shrinkage. In contrast to this shrinkage, agglomeration of interstitials leads to an expansion along the c-axis. At irradiation temperatures $T_{irr} < 400$ °C, damage accumulates rapidly (lack of vacancy mobility), and the crystal changes start to interact with the porosity. At high temperatures ($T_{irr} > 300$ °C) shrinkage with turnaround to swelling at higher doses is observed. This turnaround into volume swelling due to incompatibility of crystal strains causes new pore generation. The radiation-induced microstructural changes lead not only to swelling and shrinking, but they also affect the physical properties of graphite. Thermal creep in graphites is negligible at temperatures up to ~2,000 °C. Irradiation creep is significant at all temperatures. Application of external load leads to irradiation creep of graphite similar to metals Fig. 15. Without external stress the graphite follows the "unstressed" line which shows shrinkage converting to

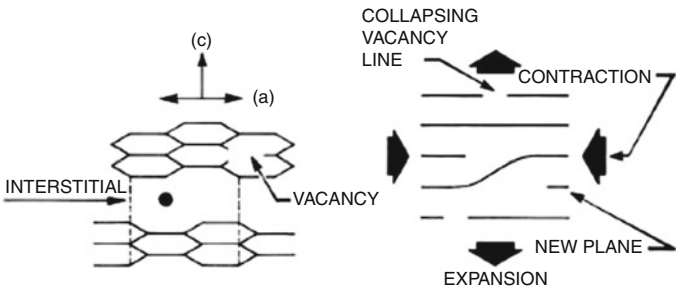


Fig. 14 Dimensional changes of graphite as a result of point defect reactions (Courtesy Burchell TD ORNL, Ball (2008), Burchell (1999))

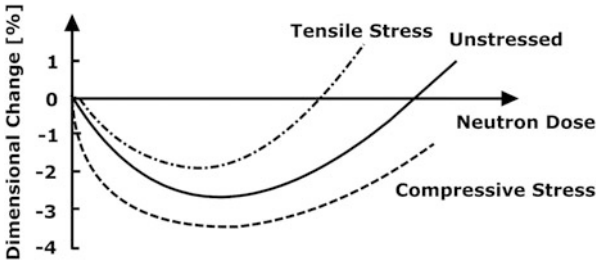


Fig. 15 Irradiation-induced dimensional change in stresses and in unstressed graphite (Courtesy Burchell TD ORNL, Ball (2008), Burchell (1999))

swelling with increasing radiation. Addition of a tensile load enhances swelling, whereas addition of a compressive load diminishes swelling.

Silicon Carbide

Fiber-reinforced materials like SiC/C or SiC/SiC are candidates for structural applications in fusion as well as advanced fission plants. They were mainly investigated with respect to fusion (Ozawa et al. 2010). Silicon carbide shows different types of radiation damage depending on temperature:

- Amorphization (up to about 200 °C)
- Point defect swelling (between 200 and 1,000 °C)
- Void swelling (above 1,000 °C)

Significant improvements with respect to resistance against irradiation could be made for SiC fibers. Also strong improvements of the matrix could be achieved with advanced compaction techniques. Indications exist that the strength of irradiated advanced fiber material could remain unchanged up to at least 10 dpa and perhaps higher. Further advances will likely require tailoring the interface swelling characteristics to compensate for differential swelling between the fiber and matrix. An exhaustive review of state of the art in ceramics for nuclear applications can be found in Katoh et al. (2007). Although this report is entitled as "Assessment of Silicon Carbide Composites for Advanced Salt-Cooled Reactors," it is a broad review of literature and results on radiation damage of SiC/SiC covering particularly fusion developments. For some advanced reactor applications like control rod or structural parts of a VHTR, the radiation damage of commercially available (German MAN today MT Aerospace AG, German DLR) ceramic composites (SiC/SiC, SiC/C) was investigated. Irradiation was performed in the SINQ neutron spallation source of PSI (2013) (up to 27 dpa, 2,300 appm He, up to 550 °C) (Pouchon et al. 2011). Under these conditions the chemical vapor-infiltrated (CVI) SiC with amorphous carbon fibers showed the best radiation resistance (almost no loss of strength). The surprisingly inferior behavior of SiC/SiC might be attributed to the fact that in the material investigated no radiation optimized SiC fibers were used (Pouchon et al. 2011).

Irradiation Damage of Components

Light Water Reactors

Pressure Vessels

Light water reactor pressure vessels are made of low-alloy steel with the inside covered with austenitic steel cladding (against corrosion) and weldments of flanges and penetrations. The aging behavior of the RPV is particularly important because of its enormous safety relevance. Low-alloy steels exhibit a brittle–ductile temperature transition. Above a characteristic temperature RPV steels are tough which means that they have a relatively high fracture toughness. Below this characteristic temperature the fracture toughness is low and fracture is dominated by cleavage. Embrittlement is characterized by an increase in the ductile-to-brittle transition temperature as well as a reduction in the fracture toughness in the ductile fracture regime. The lower fracture toughness for the embrittled material reduces the allowable (critical) crack length and therefore reduces the safety margin as shown in Fig. 16. The lower line refers to the real crack length and how it develops with time. The upper line refers to the critical crack length at which the component fails. The critical crack length is not a constant because effects like thermal embrittlement or thermal aging can reduce the fracture toughness and therefore also reduce the critical crack length. Nondestructive evaluation performed in intervals determined by the expected subcritical crack growth rates (ISI) and condition-based monitoring are therefore extremely important safety measures. Main parameters for irradiation damage of RPVs are material and its chemical composition, temperature, neutron flux, energy spectrum of neutrons, irradiation time, and neutron fluence.

Radiation damage of light water reactor pressure vessels has been summarized and thoroughly reviewed in Odette and Lucas (2001), Hashmi et al. (2005), and

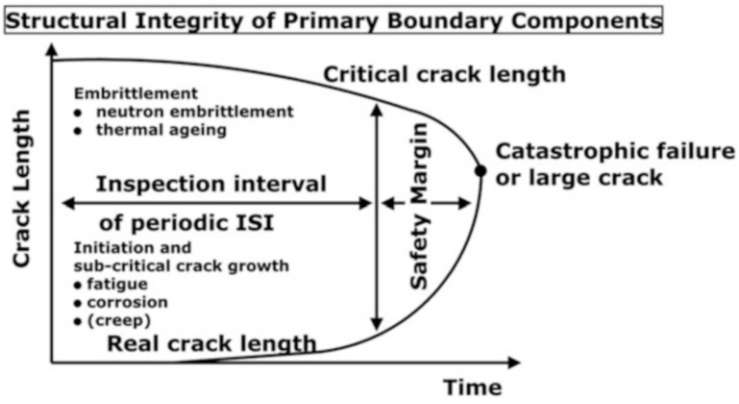


Fig. 16 Schematic of damage development in nuclear plants; ISI means "in-service inspection" (Source Bakirov (2010))

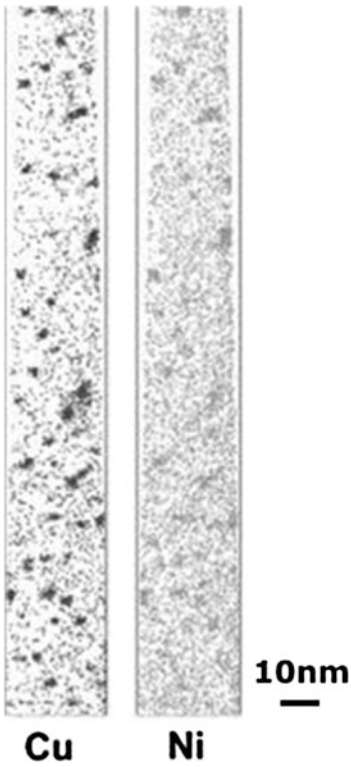


Fig. 17 Atom probe tomography (APT) image of Cu–Ni nanocluster in irradiated RPV-steel (Source (Miller and Russel 2007))

Steele (1993). It is typically caused by displacement damage and irradiation-induced nano-precipitates:

- Displacement damage: Point defect clusters and loops acting as pinning points for dislocations increasing the strength and decreasing ductility
- Radiation-induced phase transformations: Precipitation of Cu nanoclusters (see Fig. 17) or "manganese–nickel-rich precipitates" (MNPs) or "late blooming phases" (LBPs) which additionally contribute to hardening and embrittlement

Copper impurity has long been recognized as the dominant detrimental element in reactor pressure vessel (RPV) steels at copper levels in excess of about 0.1 wt%. The formation of copper-rich precipitates gives rise to considerable hardening and embrittlement at levels of neutron fluence well below the design end-of-life (EoL) of RPVs of operating nuclear power plants. The role of copper content on the embrittlement of RPV steels was thoroughly investigated with small angle neutron scattering (SANS) and tensile tests (Bergner et al. 2009). From the 1990s there has been increasing evidence of clusters enriched with manganese and nickel appearing in low-Cu steels (Cu 0.1 wt%).

The terms "manganese–nickel-rich precipitate" (MNP) or "late blooming phase" (LBP) (Odette and Wirth 1997) emphasize different aspects of this phenomenon. MNPs were first predicted by thermodynamic arguments (see, e.g., Odette and

Wirth (1997, Odette and Lucas (1998)) and then confirmed by means of several experimental techniques including atom probe tomography (APT) and positron annihilation spectroscopy. The vessel walls in the reactor beltline regions are subjected to the highest fluences and degradation due to irradiation embrittlement. Therefore, the welds within that region become possibly the weakest link since the welds are likely to contain defects that can become cracks. Additionally, the higher copper (and nickel) content in many of the older vessel welds has led to much higher radiation damage sensitivity. The base metals should not be ignored, since the copper content in older plates and forgings was not controlled to a minimum level; however, there appears to be less irradiation embrittlement in base materials as compared to welds with the same copper/nickel concentrations. A quantitative understanding of irradiation embrittlement of the reactor pressure vessels is extremely important for assessments of residual safe life of a nuclear plant. Condition-based monitoring of degree of damage belongs therefore to key tasks in several plant life extension concepts.

Surveillance specimen can help to assess the degree of radiation damage of an RPV. Those samples were built into the vessel at locations where they were exposed to neutron damage. After removal the samples could be further analyzed (Charpy tests, metallographic studies, etc.) for an assessment of the condition of the RPV. Fracture toughness can be assessed from shifts of the ductile-to-brittle transition temperature using the "master-curve""approach." It is based on the observation that at least for low-alloy steels, a good relationship between a "reference temperature" and the fracture toughness, K_{JC}, exists:

$$K_{JC} = 30 + 70.\exp\left[0.019\left(T - T_0\right)\right] \; MN\, m^{-3/2}$$

where T is the temperature of interest and T_0 is the ductile-to-brittle transition temperature (Wallin 1991; IAEA 2009). To improve the accuracy of residual life assessments, also more detailed analyses with a combination of mechanical testing of miniaturized samples, extended microstructural analyses, and advanced material modeling tools are considered (see, e.g., Hoffelner (2012)).

Reactor Internals

Radiation-induced material changes and susceptibility to intergranular failure of light water reactor core internals were summarized in Bruemmer et al. (1999b). Such failures have occurred after many years of service in boiling water reactor (BWR) core components and, to a lesser extent, in pressurized water reactor (PWR) core components. These failures occur in stainless iron- and nickel-base alloys exposed to a significant flux of neutron radiation in the reactor coolant environment.

Stress corrosion cracking (SCC) without radiation is an unexpected sudden failure of ductile metals subjected to a tensile stress in a corrosive environment. The stresses can be the result of the crevice loads due to stress concentration, or they can be caused by the type of assembly or residual stresses from fabrication (e.g., cold working). SCC cracks are predominantly intergranular in nature. Reactor

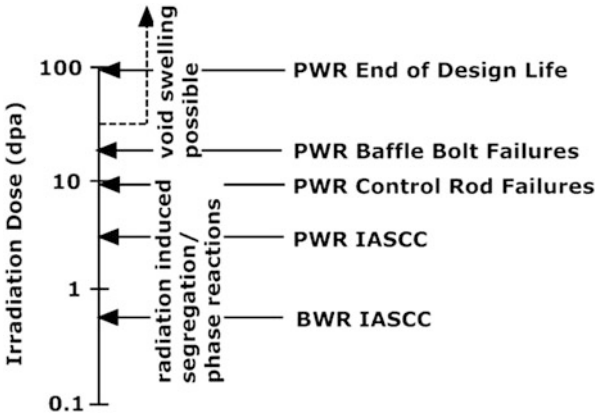


Fig. 18 Possible damage of LWR components due to irradiation-assisted stress corrosion cracking dependent on time of exposure, given as irradiation dose (Source Bruemmer et al. (1999b))

internals do not belong to the group of pressure boundaries, and therefore, the mechanical loads are lower than in RPV and pressure piping. Internals are mainly exposed to the coolant, and therefore, corrosion resistance is more important than strength. Under these conditions austenitic steels are much better suited than the low-alloyed ferritic/bainitic steels used for RPVs. During service cracking susceptibility often results as a combination of radiation, stress, and a corrosive environment. The resulting failure mechanism has therefore been termed irradiation-assisted stress corrosion cracking (IASCC). Intergranular(IG) SCC is promoted in austenitic stainless steels when a critical threshold fluence is reached (see, e.g., Was (2007), Hoffelner (2012)). This time dependence leads to occurrence of cracking after some time of operation as shown in Fig. 18 (Bruemmer et al. 1999b). Therefore, it took quite a while until this type of damage became obvious in nuclear plants. As in classical stress corrosion cracking (SCC), the aqueous environment chemistry and component stress/strain conditions also strongly influence observed cracking.

Several important metallurgical, mechanical, and environmental aspects that are believed to play a role in the cracking process (Bruemmer et al. 1999b; Andresen et al. 1990; Was and Andresen 1992; Scott 1994; Ford and Andresen 1994) are shown schematically in Fig. 19.

The damage happens most probably in the following form:

- Irradiation damage leads to hardening of the matrix making basically the grain boundaries more attractive as path for growing cracks. This is what often happens as a result of hardening also without irradiation.
- Irradiation is also responsible for changes in grain boundary compositions by radiation-induced segregation (primarily chromium depletion) which can further weaken the cohesion between them.
- The surface of the cracks (particularly at the crack tip) is exposed to the radiolysis products which lead to chemical corrosion attack.
- Additionally, the crack can act as crevice supporting crevice corrosion.

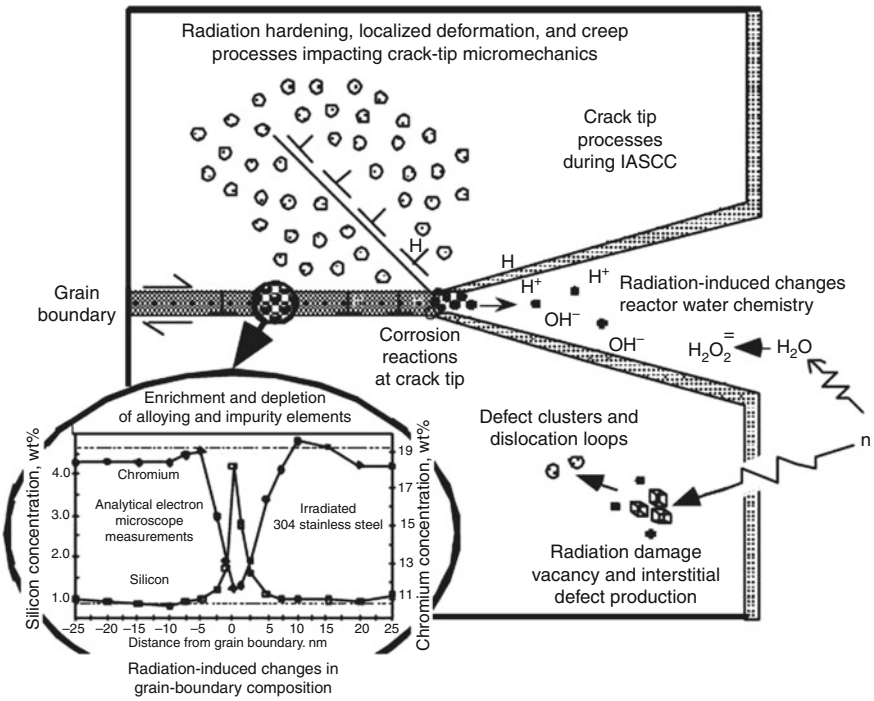


Fig. 19 Main mechanisms of irradiation-assisted stress corrosion cracking; reprinted with permission of ASM International. All rights reserved. www.asminternational.org (Source see also Was (2007))

All these facts together contribute to enhanced crack growth along the grain boundaries. As cracks in the reactor internals do not have the same damage potential as cracking of primary boundary components, measures were developed to stop or to slow down crack growth by chemistry (MacDonald et al. 1995; Hettiarachchi et al. 1995). A successful concept is the technique of noble metal chemical addition (NMCA) which has been commercially applied since 1996 (Hettiarachchi et al. 1997).

With respect to life extension and residual life assessments of reactors, discussions are ongoing if void swelling and eventually even helium effects might become relevant for very long times of operations (see Fig. 18). Void swelling was of concerns and even application limiting for earlier fast reactors. But void swelling has not yet been of real concern for LWR internals. Due to the very low dose (2–3 dpa maximum) expected in the shrouds of BWRs, void swelling per se is not considered to be a license extension issue for BWRs. However, as recently discussed (Garner et al. 2005; Garner 2010b), there is a growing body of evidence that swelling and irradiation creep might become important for life extensions of LWRs to 60 years and beyond. The most swelling-vulnerable locations (>5 %) are expected to be concentrated in small volumes of the reentrant corners of PWR

baffle-former assemblies constructed from AISI 304 stainless steel. Even at lower swelling levels, however, differential swelling of annealed 304 baffle-former plates and cold-worked 316 baffle bolts is being considered as a possible contributor to corrosion and cracking of bolts.

Zircaloy Claddings

Claddings are the structural parts which are mostly exposed to irradiation and therefore to irradiation damage. The effects of neutron irradiation on microstructure and properties of Zircaloy were summarized by Adamson (Adamson 2000). For Zircaloy with a hexagonal crystal structure black dots, dislocation loops (often related to the basal c-plane) and microstructural changes leading to swelling and irradiation creep are the most significant types of damage. After service exposure in LWRs, a high density of black spots is present (see Fig. 20). The spots are very small, and a further analysis of its nature is not easily possible.

Irradiated Zircaloy undergoes also swelling and irradiation creep both being important for design because they are responsible for structural changes during service which always bears some risk for failure. Swelling of Zircaloy is of high importance for CANDU reactors where also the pressure pipes are made of Zircaloy. Swelling phenomena of Zircaloy are today reasonably well understood which is a result of the long experience with light and heavy water reactors. However, increasing burnup and possible effects of service exposure on fuel rods after its active life during transport or at final storage trigger active research in this field still today. Swelling is a function of fluence, microstructure, and temperature but also on hydrogen content and other parameters.

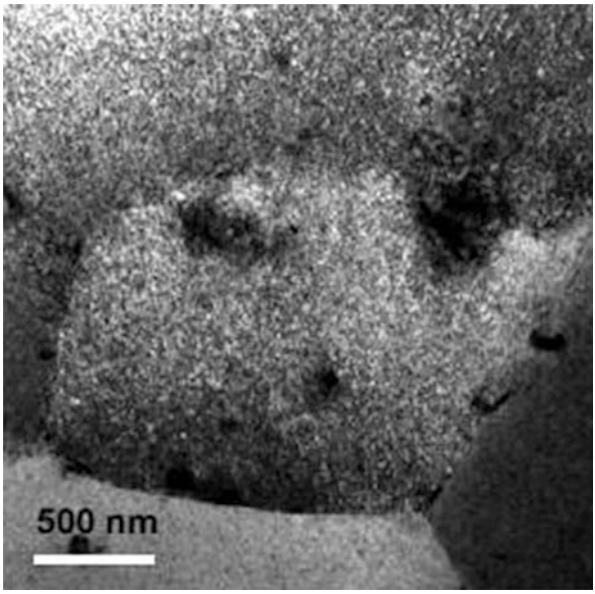


Fig. 20 Irradiation damage in service-exposed Zircaloy cladding. Black dots are clearly visible (TEM bright field micrograph replotted from Hoffelner (2012))

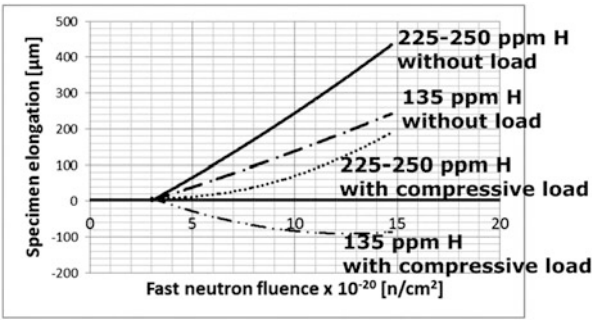


Fig. 21 Irradiation creep of Zircaloy guide tubes tested in the Halden test reactor (Source McGrath et al. (2010))

Swelling and creep of samples from several sections of pre-exposed, recrystallized Zircaloy-4 guide tubes from 2 commercial PWRs were studied in the Halden test reactor (McGrath et al. 2010). The load for the creep test was applied by the loop pressure squeezing a sealed bellow which applied a compressive axial force on guide tubes. The results are shown in Fig. 21. The radiation creep effect due to compressive load is clearly visible. It also looks like increasing hydrogen content would increase swelling and radiation creep. According to the authors of this study, a quantitative explanation needs to consider several factors including the prehistory of the claddings.

Phase changes are another effects encountered in Zircaloy under irradiation as previously mentioned already. Zircaloy contains (mainly intermetallic) precipitates, so-called second phase particles which play an important role for oxidation. Common precipitates are $Zr(Fe,Cr)_2$ and $Zr_2(Fe,Ni)$ which can amorphize and decompose during service (Valizadeh et al. 2010; Herring and Northwood 1988). Simultaneously, iron (and at a much slower rate, chromium) can be lost to the matrix. Decomposition of second phase particles changes the local chemical composition of the matrix and plays therefore an important role for the oxidation of Zircaloys.

Advanced Nuclear Plants

Basic Considerations

Although current LWRs undergo permanent improvements, also other reactor types are considered for the future. Six reactor concepts shown in Table 2 were chosen for further consideration within the International Generation IV initiative (Roadmap 2002). Cooling media different from water (except for the supercritical water reactor) as well as the use of a fast-neutron spectrum (in contrast to the thermal spectrum of current light water reactors) are characteristics of these systems. Thermal neutrons have typically energies ranging from about 1 eV to 0.1 MeV. Fast neutrons have energies higher than 0.1 MeV. The coolants allow higher operation temperatures and provide therefore also options for combined cycle plants generating electricity and process steam or heat. The fast spectrum allows

Table 2 Generation IV types of reactors and nuclear plants. Nuclear fusion is often referred to as Generation V

System	Neutron spectrum	Coolant	Outlet temperature °C	Fuel cycle	Typical size (MW_e)
VHTR (very-high-temperature reactor)	Thermal	Helium	900–1,000	Open	250–300
SFR (sodium-cooled fast reactor)	Fast	Sodium	500–550	Closed	50–150 300–1,500 600–1,500
SCWR (supercritical-water-cooled reactor)	Thermal/fast	Water	510–625	Open/closed	300–700 1,000–1,500
GFR (gas-cooled fast reactor)	Fast	Helium	850	Closed	1,200
LFR (lead-cooled fast reactor)	Fast	Lead	480–570	Closed	20–180 300–1,200 600–1,000
MSR (molten salt reactor)	Thermal/fast	Salts fluoride	700–800	Closed	1,000

closed fuel cycles which would be a real step towards waste minimization and resource saving. A fast spectrum is obtained when the reactor operates without moderators like water or graphite. Together with advanced recycling techniques, fast reactors could use spent fuel currently considered as long-life high-level radioactive waste as mixed fuel. The residuals for final storage would be only fission products with much shorter half-life than plutonium and minor actinides which are currently of concern for final repositories. Also uranium could be used far more efficiently as fuel.

Except the SCWR and GFR, several reactor concepts have been already studied earlier (Hoffelner et al. 2011). The German HTR was a high-temperature reactor with direct cycle helium turbine, and the French "Phenix" and "Superphenix" were sodium fast reactors. Experience with lead-cooled reactors exists from Russian submarines. MSR considerations were reported already in the 1960s of the last century by Oak Ridge National Laboratory (ORNL).

Besides the six GenIV concepts, also other advanced nuclear power options are under consideration: Quite some interest found the sodium-cooled traveling wave reactor of Intellectual Ventures (US) in which the active part penetrates slowly (several years) through pipes (similar to a burning cigarette) (Wald 2013) and Accelerator-Driven Systems (ADS) where the neutrons are created by directing protons from an accelerator towards a target, creating neutrons this way (Rubbia et al. 1995). These neutrons are coupled preferentially to a lead-cooled fast reactor. Finally nuclear fusion should be mentioned as advanced nuclear system (2013), however, without further description.

Fast neutrons are (according to higher energy) more damaging to radiation-exposed materials than the predominantly thermal neutrons in current LWRs. In contrast to LWRs only scarce information concerning in-service radiation damage

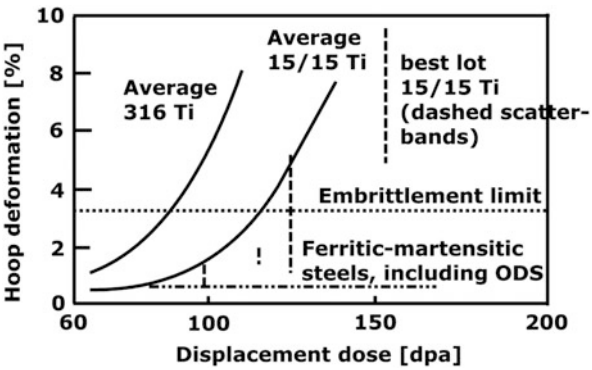


Fig. 22 Swelling of different cladding materials in French fast reactors. Ferritic–martensitic steels show the best behavior, but the titanium-modified austenitic steels could provide an alternative (Source Yvon and Carré (2009))

of components is available. Environments other than water and higher operation temperatures are additional challenges for materials and components.

Vessel embrittlement is expected to stay as an important issue also for advanced reactors. The advantage for SFR, LFR, and MSR types is that they don't need a pressure vessel, only a vessel which is preferentially made of austenitic steel or a nickel-base alloy. Vessel embrittlement is not expected to have the same impact on safety like in case of LWRs. SCWR, GFR, and VHTR vessels are pressurized vessels necessitating ferritic–martensitic steels (low-alloy steels, martensitic 9 % Cr steels) which are expected to show comparable displacement damage like current RPVs. Eventually other long-term degradation mechanisms need to be investigated.

Swelling and irradiation creep could become life-limiting factors for claddings and internals of fast reactors. Ferritic–martensitic steels show a better swelling behavior than austenitic steels (see Fig. 22). Austenitic steels based on 316 with distinct Ti additions (15/15 Ti) show acceptable swelling behavior and are therefore considered as claddings for SFRs at least for moderate burnup.

Development of helium voids and bubbles is considered as a major damage cause particularly with the high temperatures causing additionally thermal creep. Helium produced during service can move to grain boundaries forming bubbles and reducing creep ductility and creep rupture strength (see Fig. 11). Oxide dispersion strengthened (ODS) steels contain nanosized oxide dispersoids which can trap the helium and therefore prevent it from moving to the grain boundaries. Unfortunately the production of ODS materials is rather difficult, and additionally they are extremely expensive which limits its potential for future applications. Interestingly, ODS steels do not show significant improvements in irradiation creep although they show much better thermal creep resistance compared with non-ODS grades as shown in (Toloczko et al. 2004; Chen and Hoffelner 2009).

Advanced Material Characterization Tools

To gain more confidence on possible material degradation in service besides testing and evaluation of traditional samples as methods of advanced material research (materials modeling and model validation) could make an important contribution to safe operation of advanced nuclear plants. Figure 23 summarizes current testing,

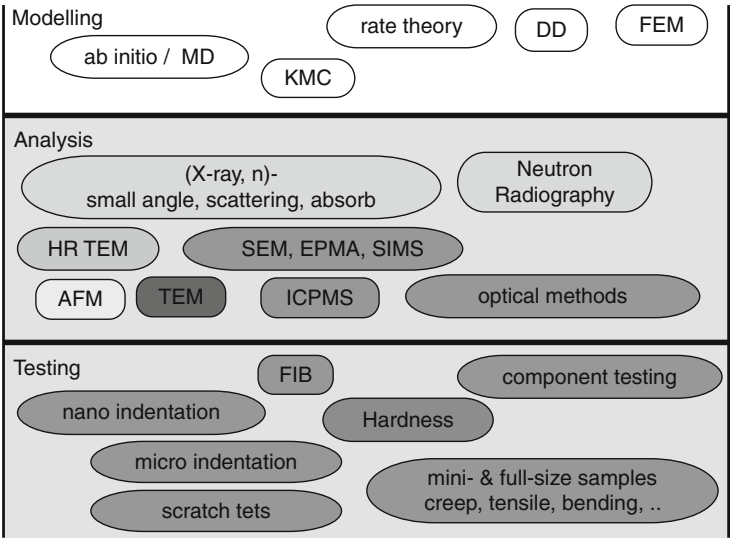


Fig. 23 Testing and analysis techniques used to characterize material behavior on different scales (MD molecular dynamics, DD dislocation dynamics, KMC kinetic Monte Carlo, n neutrons, HR high resolution, TEM transmission electron microscope, SEM scanning electron microscope, AFM atomic force microscope, ICPMS inductively coupled mass spectroscopy, SIMS secondary ion mass spectroscopy, EPMA electron probe micro analysis, FIB focused ion beam, FEM finite element analysis) (Source Hoffelner (2012))

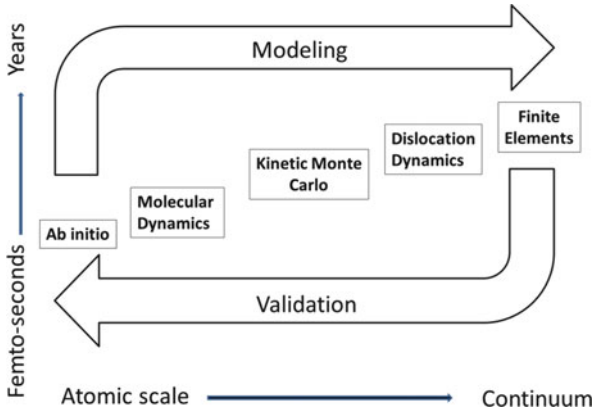


Fig. 24 Scale dependence of different modeling techniques (Source Hoffelner (2012))

analysis, and modeling methods which are used for materials and damage characterization. A detailed description of the different techniques with respect to assessment of irradiation damage in advanced plants goes far beyond the scope of this chapter, and only a few important aspects shall be highlighted. The most important modeling methods are shown in Fig. 24.

Ab initio calculations investigate the physics of matter. They are mainly based on the application of density functional theory (DFT). This allows (with a number of approximations) the determination of the energy of the ground state of a system of interacting particles. Basically this would need a many-body solution of the quantum mechanics Schrödinger equation. DFT provides a reformulation of this problem to a single-body problem. Due to computational restrictions, such calculations are currently limited to a small number of (up to 1,000) atoms. Most calculations are static, thereby neglecting dynamic effects. Methods for dynamic calculations basically exist, but they are extremely computing time-consuming and expensive. Despite these limitations which do not allow the investigations of temperature effects or the behavior of larger systems, the method allows to gain insight into the basic atomistic behavior of a solid.

Details about the arrangement of atoms and similar microstructural issues can be studied with molecular dynamics (MD) simulations. Molecular dynamics implements the potentials to describe the movement of atoms in space and time as results of interatomic and external forces. The equations of motion of classical mechanics in combination with the potentials determined with quantum mechanics are solved for a set on N interacting atoms starting from assigned initial conditions. Controlling variables like temperature or pressure can be conveniently introduced as constraints.

MD simulation can accurately describe the atomistic behavior, but the total simulation time is typically limited to less than 1 ms. On the other hand, the important damage processes in structural materials usually occur on much longer timescales. These processes include reactions between atoms, adsorption-desorption on the surface, occasional transitions from one state to another, and especially diffusion and annihilation of defects after a cascade event in an irradiation experiment. Such effects can be studied using a combination of MD and Kinetic Monte Carlo (KMC). The KMC method is a probabilistic approach that enables the prediction of longer-term damage evolution. The output data of MD is used in KMC in order to determine the probabilistic motion and reaction between defects and atoms (Dalla Torre et al. 2005; Barbu et al. 2005; Domain et al. 2004) where motion and clustering of point defects are the dominant mechanism.

Models based on reaction *rate theory* have been broadly and successfully applied to simulate radiation-induced microstructural evolution radiation damage (Stoller and Greenwood 1999; Stoller et al. 2008). The use of these models involves the simultaneous solution of a modest number of differential equations to predict phenomena such as void swelling, irradiation creep, or embrittlement. The timescale of interest for these processes is determined by atomic diffusion rates and the desired in-service lifetime of irradiated components. Rate theory is well suited to span this time range from seconds to years and a size scale from micrometers to macroscopic dimensions. However, the source term in the rate equations is dictated by atomic displacement cascades, events that occur on a timescale of a few tens of picoseconds and a few tens of nanometers in space.

Discrete dislocation dynamics (DDD) simulation is a mesoscopic tool to study plastic deformation and the interaction between dislocations as well as between

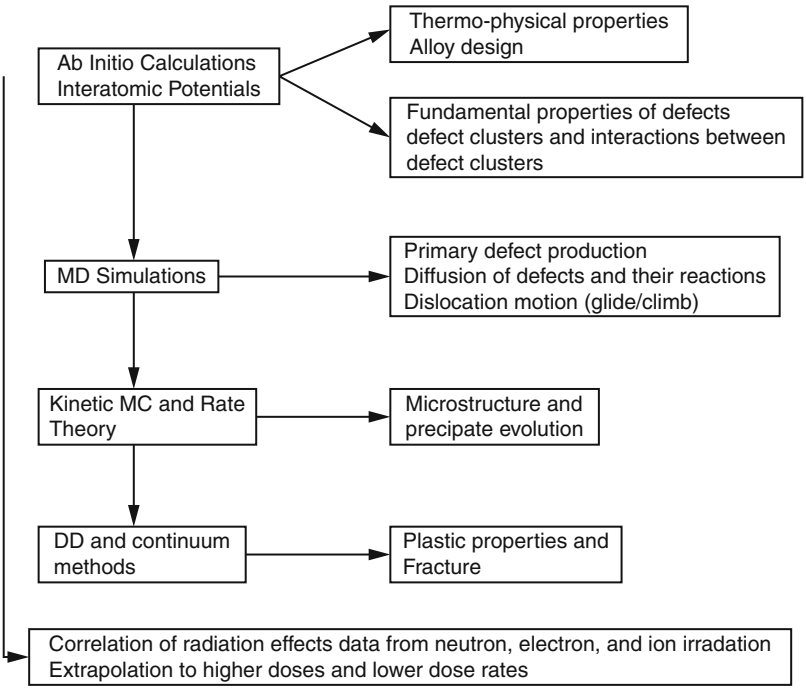


Fig. 25 Elements of a full-scale model description of radiation damage in metals (Source Stoller and Mansur (2005))

dislocations and obstacles. It directly simulates the dynamic, collective behavior of individual dislocations and their interactions. In a numerical implementation, dislocation lines are represented by connected discrete line segments that move according to driving forces including dislocation line tension, dislocation interaction forces, and external loading. Discrete dislocation dynamics operates in single crystals. There are also developments of dislocation field dynamics modeling where dislocations are characterized by stress fields rather than discrete dislocation line elements (Ghoniem et al. 2000; Ortiz 1999; Koslowskia et al. 2002). Such models operating on larger scale than discrete dislocation dynamics are able to go beyond the limitation of single crystals.

Constitutive equations and finite element calculations describe the macroscopic behavior of a material. Figure 25 (Stoller and Mansur 2005) shows schematically how the different modeling techniques can contribute to the understanding of radiation damage.

Model validations can be done with advanced analytical techniques like high-resolution TEM or beamline techniques (X-rays, neutrons) and with advanced mechanical testing. In contrast to testing of large laboratory samples, advanced mechanical testing uses micro- or nanosize testing equipment. This allows the

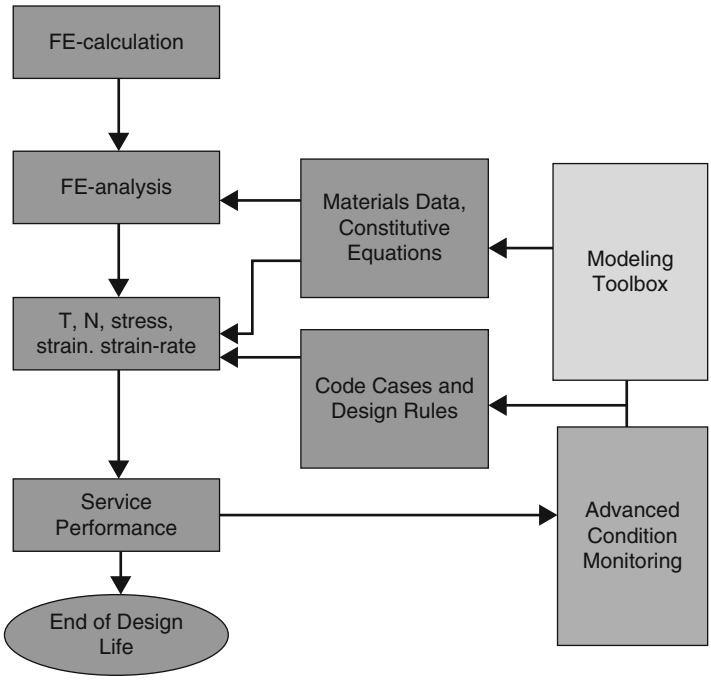


Fig. 26 Possible interaction of advanced modeling methods and advanced condition monitoring techniques with traditional design (Source Hoffelner (2012))

determination of mechanical properties of small volumes which is necessary for a good correlation between microstructure and strength. Experimental investigation of the early stage of radiation damage needs tools with extremely high time resolution and extremely short detection ranges. It can be expected that the new generation of beamline techniques (free electron lasers (Swissfel 2011; Jefferson Lab 2013)) can bring significant contributions to experimental studies of early irradiation damage.

Determination of material response under neutron irradiation requires very time-consuming, difficult, and expensive tests under neutron irradiation. Tests with other energetic particles (ions, electrons) are therefore an important tool to study radiation damage (though quantitative differences to neutrons might punctually exist).

Although all these advanced methods are remarkable improvements for the basic understanding of materials, they can currently not deliver real design data. For those data still testing of large samples and service experience in test beds and demonstrators are necessary. However, it can be expected that traditional design methods together with advanced materials science and service experience can strongly enhance damage and lifetime assessments in future nuclear plants as schematically shown in Fig. 26.

Conclusions

Irradiation damage of structural materials is a complex particle–matter interaction which shows a wide range of effects at different timescales:

- Scattering effects leading to point defect supersaturation
- Point defect reactions leading to obstacles for dislocation movement
- Nuclear reactions leading to phase transitions and phase changes
- Nuclear reactions leading to transmutation of elements

On component level these effects lead to changes in geometry (swelling), changes of mechanical properties (hardening/embrittlement), and changes of chemical behavior (irradiation-assisted stress corrosion cracking, oxidation resistance). Irradiation together with mechanical loads can lead to irradiation creep at temperatures far below which thermal creep occurs. Particularly important are transmutation reactions which can lead to alpha-radiating isotopes which are sources for gas development in the material which can severely degrade mechanical properties.

Irradiation effects and its consequence on material properties and safety are quite well understood for current light water reactors with service experience of about 50 years. New future nuclear plants will in principle undergo the same types of damage. However, the fast-neutron spectrum together with higher temperatures and new working environments need a careful assessment of damage and safety in such plants. Advanced tools and methods of materials science are expected to bring significant contributions there.

References

R.B. Adamson, Effects of neutron irradiation on microstructure and properties of Zircaloy, in *ASTM International in STP 1354, Zirconium in the Nuclear Industry: Twelfth International Symposium*, American Society for Testing and Materials, Philadelphia PA, (2000), pp. 15–31

P.L. Andresen, F.P. Ford, S.M. Murphy, J.M. Perks, in *Proceedings of the Fourth International Symposium on Environmental Degradation of Materials in Nuclear Power Systems – Water Reactors*, eds. by D. Cubicciotti, G.J. Theus GJ (National Association of Corrosion Engineers, 1990), pp. 1–83

M. Bakirov, Impact of operational loads and creep, fatigue corrosion interactions on nuclear power plant systems, structures and components (SSC), in *Understanding and mitigating ageing in nuclear power plants*, ed. by P.G. Tipping (Woodhead, Oxford, 2010), pp. 146–188

D.R. Ball, *Graphite for High Temperature Gas-Cooled Nuclear Reactors* (ASME LlC STP-NU-009, New York, 2008)

A. Barbu, C.S. Becquart, J.L. Bocquet, J. Dalla Torre, C. Domain, Comparison between three complementary approaches to simulate large fluence irradiation: application to electron irradiation of thin foils. Philos. Mag. **85**(4–7), 541–547 (2005)

R.S. Barnes, Nat. (Lond.) **206**, 1307 (1965)

F. Bergner, A. Ulbricht, H.W. Viehrig, Acceleration of irradiation hardening of low copper reactor pressure vessel steel observed by means of SANS and tensile testing. Philos. Mag. Lett. **89**(12), 795–805 (2009)

E.E. Bloom, J. Stiegler, Effect of irradiation on the microstructure and creep-rupture properties of type 316 stainless steel. ORNL (1972), http://www.osti.gov/bridge/servlets/purl/4632343-ATLvL5/4632343.pdf. Accessed 20 Sept 2013

T.D. Burchell, *Carbon Materials for Advanced Technologies* (Elsevier, Oxford, 1999). ISBN 0080426832/0-08 042683-2

J. Chen, W. Hoffelner, Irradiation creep of oxide dispersion strengthened (ODS) steels for advanced nuclear applications. J. Nucl. Mater. **392**, 360–363 (2009)

J. Chen, P. Jung, M.A. Pouchon, T. Rebac, W. Hoffelner, Irradiation creep and precipitation in a ferritic ODS steel under helium implantation. J. Nucl. Mater. **373**, 22–27 (2008)

J. Chen et al., Paul Scherrer Institut NES Scientific Highlights, Paul Scherrer Institut, Villigen PSI, (2010), pp. 46–47

J. Dalla Torre, J.L. Bocquet, N.V. Doan, E. Adam, A. Barbu, JERK an event-based KMC model to predict microstructure evolution of materials under irradiation. Philos. Mag. **85**(4–7), 549–558 (2005)

C. Domain, C.S. Becquart, L. Malerba, Simulation of radiation damage in Fe alloys: an object kinetic Monte Carlo approach. J. Nucl. Mater. **335**, 121–145 (2004)

F.P. Ford, P.L. Andresen, Corrosion in nuclear systems: environmentally assisted cracking in light water reactors, in *Corrosion Mechanisms*, ed. by P. Marcus, J. Ouder (Marcel Dekker, New York, 1994), pp. 501–546

F.A. Garner, Chapter 6: Irradiation performance of cladding and structural steels in liquid metal reactors, in *Materials Science and Technology: A Comprehensive Treatment* (10A VCH, 1994), B.R.T. Frost Ed., Wiley VCH Weinheim, pp. 419–543

F.A. Garner, S.I. Porollo, V. Yu, Y.V. Konobeev, O.P.Maksimkin, Void swelling of austenitic steels irradiated with neutrons at low temperatures and very low dpa rates, in *Proceedings of the 12th international conference on environmental degradation of materials in nuclear power system – water reactors*, eds. by T.R. Allen, P.J. King, L. Nelson (TMS The minerals metals & materials society, 2005), pp. 439–448

N. Ghoniem, S. Tong, L. Sun, Parametric dislocation dynamics: a thermodynamics based approach to investigations of mesoscopic plastic deformation. Phys. Rev. B **61**(2), 913–927 (2000)

L.R. Greenwood, Neutron interactions with recoil spectra. J. Nucl. Mater. **216**, 29–44 (1994)

D.R. Harries, J. Brit. Nucl. Energy Soc. **5**, 74 (1966)

M.F. Hashmi, S.J. Wu, X.H. Li, Neutron irradiation embrittlement modeling in RPV steels- an overview, in *18th International conference on structural mechanics in reactor technology (SMiRT 18)*, Beijing, 7–12 Aug 2005. SMiRT18-F01-8

R. Havel, M. Vacek, M. Brumovsky, Fracture properties of irradiated A533B, Cl.1, A508, Cl.3, and 15Ch2NMFAA reactor pressure vessel steel, in *Radiation embrittlement of nuclear reactor pressure vessel steels*, ed. by L. Steele. STP, vol. 1170 (ASTM, Philadelphia, 1993), pp. 163–171

R.A. Herring, D.O. Northwood, Microstructural characterization of neutron irradiated and post-irradiation annealed Zircaloy-2. J. Nucl. Mater. **159**, 386–396 (1988)

S. Hettiarachchi et al., in *Proceedings of the 7th International Symposium on Environmental Degradation of Materials in Nuclear Power Systems – Water Reactors*, Beckenridge Colorado, (1995), p. 735

S. Hettiarachchi et al., in *Proceedings of the 8th International Symposium on Environmental Degradation of Materials in Nuclear Power Systems – Water Reactors*, Beckenridge Colorado, (1997), p. 535

W. Hoffelner, R. Bratton, H. Mehta, K. Hasegawa, D.K. Morton, New generation reactors, in *Energy and Power Generation Handbook-Established and Emerging Technologies*, ed. by K.R. Rao (ASME, New York, 2011)

W. Hoffelner, *Materials for Nuclear Plants, from Save Design to Residual Life Assessments* (Springer, London, 2012). ISBN 978-1-4471-2915-8

IAEA, *Irradiation Damage in Graphite Due to Fast Neutrons in Fission and Fusion Systems* (IAEA-TECDOC-1154, Vienna, 2000)

IAEA, *Master Curve Approach to Monitor Fracture Toughness of Reactor Pressure Vessels in Nuclear Power Plants*. IAEA-TECDOC, vol. 1631 (IAEA, Vienna, 2009)
Y. Isobe, M. Sagisaka, F.A. Garner, Precipitate Evolution in Low-Nickel Austenitic Stainless Steels During Neutron Irradiation at Very Low Dose Rates. ORNL (2008), http://www.ornl.gov/sci/physical_sciences_directorate/mst/fusionreactor/pdf/dec2008/6_AUSTENITIC/6.1.Isobe_052-060.pdf. Accessed 20 Sept 2013
L.A. James, J.A. Williams, The effect of temperature and neutron irradiation upon the fatigue crack propagation behavior of ASTM A533-B steel. J. Nucl. Mater. **47**, 17–22 (1973)
Jefferson Lab (2013) http://www.lightsources.org/images/posters/jlabposter3.jpg Accessed 20 Sept 2013
Y. Katoh, D.F. Wilson, C.W. Forsberg, Assessment of Silicon Carbide Composites for Advanced Salt-Cooled Reactors, (2007). ORNL Oak Ridge Tennessee. ORNL/TM-2007/168 Revision 1
J.L. Katz, H. Wiedersich, Chem. Phys. **55**, 1414 (1971)
R.L. Klueh, D.J. Alexander, in *Effects of Radiation on Materials: 15th International Symposium*, eds. by R.E. Stoller, A.S. Kumar, D.S. Gelles. ASTM STP 1125 (American society for testing and materials, Philadelphia, 1992), p. 1256
R.L. Klueh, K. Shiba, M.A. Sokolov, Embrittlement of irradiated ferritic/martensitic steels in the absence of irradiation hardening. J. Nucl. Mater. **377**, 427–437 (2008)
M. Koslowskia, A.M. Cuitino, M.A. Ortiz, Phase-field theory of dislocation dynamics, strain hardening and hysteresis in ductile single crystals. J. Mech. Phys. Solids **50**, 2597–2635 (2002)
G.J. Lloyd, J.D. Walls, J. Gravenor, Low temperature fatigue crack propagation in neutron-Irradiated type 316 steel and weld metal. J. Nucl. Mater. **101**, 251–257 (1982)
D.D. MacDonald, T.K. Yeh, A.T. Motta, (1995) Corrosion, NACE International, Houston, TX, (1995) Paper No. 403
A.D. Marwick, Segregation in irradiated alloys: the inverse Kirkendall effect and the effect of constitution on void swelling. J. Phys. F Metal Phys. **8**, 9 (1978)
M.A. McGrath, S. Yagnik, H. Jenssen, Effects of pre-irradiation on irradiation growth & creep of re-crystallized Zircaloy-4, in *16th International Symposium on Zirconium in the Nuclear Industry*, Chengdu, Sichuan Province, 9–13 May 2010, http://www.astm.org/COMMIT/B10_Zirc_Presentations/6.5_ASTM-2010-creep-growth.pdf. Accessed 20 Sept 2013
M.K. Miller, K.F. Russel, Embrittlement of RPV steels: an atom probe tomography perspective. J. Nucl. Mater. **371**(1–3), 145–160 (2007)
M.T. Motta, F. Lefebvre, C. Lemaignan, Amorphization of precipitates in Zircaloy under neutron and charged-particle irradiation, in *Zirconium in the Nuclear Industry*. Ninth International Symposium, eds. by C.M. Eucken, A.M. Garde. ASTM STP 1132 (American Society for Testing and Materials, Philadelphia, 1991), pp. 718–739
G.R. Odette, G.E. Lucas, Rad. Eff. Def. Sol. **144**, 189 (1998)
G.R. Odette, G.E. Lucas, Embrittlement of nuclear reactor pressure vessels. JOM **53**(7), 18–22 (2001)
G.R. Odette, B.D. Wirth, J. Nucl. Mater. **251**, 157 (1997)
M. Ortiz, Plastic yielding as a phase transition. J. Appl. Mech. Trans. ASME **66**(2), 289–298 (1999)
K. Ozawa, Y. Katoh, L.L. Snead, T. Nozawa, Effect of neutron irradiation on fracture resistance of advanced SiC/SiC composites. Fusion materials semiannual progress report (2010). DOE-ER-0313/47
M.A. Pouchon, T. Rebac, J. Chen, Y. Dai, W. Hoffelner, Ceramics composites for next generation nuclear reactors, in *Proceedings of GLOBAL 2011*, Makuhari, 11–16 Dec,2011 Paper No. 358363
R.J. Puigh, M.L. Hamilton, In-reactor creep rupture behavior of the D19 and 316 alloys, in Influence of Radiation on Material Properties. 13th International Symposium Part II ASTM STP 957, eds. by F.A. Garner, C.H. Henager, N. Igata (ASTM, 1987)
GENIV Roadmap (2002), http://www.gen-4.org/PDFs/GenIVRoadmap.pdf. Accessed 20 Sept 2013
J.P. Robertson, R.L. Klueh, K. Shiba, A.F. Rowcliffe, Radiation Hardening and Deformation Behaviour of Irradiated Ferritic-Martensitc Steels (1997), ORNL Oak Ridge Tennessee http://web.ornl.gov/sci/physical_sciences_directorate/mst/fusionreactor/pdf/dec1997/paper24.pdf

C. Rubbia et al., CERN-AT-95-44-ET, in Accelerator-Driven Transmutation Technologies and Applications. Proceedings of the Conference Las Vegas, July 1994. AIP Conference Proceedings 346, (American Institute of Physics, Woodbury, 1995), p. 44
K.C. Russel, Acta Metall. **19**, 753 (1971)
W. Schilling, H. Ullmaier, Physics of radiation damage in metals, in *Materials Science and Technology* 10B (VCH, 1994), B.R.T. Frost Ed., Wiley VCH Weinheim, p. 187
R. Scholz, R. Mueller, Irradiation creep-fatigue interaction of type 3 16 L stainless steel. J. Nucl. Mater. **233–237**, 169–172 (1996)
P.M. Scott, A Review of irradiation assisted stress corrosion cracking. J. Nucl. Mater. **211**, 101 (1994)
A. Seeger, *Radiation Damage in Solids 1* (IAEA, Vienna, 1962), p. 101
SINQ PSI (2013), http://www.psi.ch/sinq/. Accessed 20 Sept 2013
L.E. Steele (ed.), Radiation embrittlement of nuclear reactor pressure vessel steels: an international review (third volume, 1993), ASTM STP 1170 American Society for Testing and Materials, Philadelphia PA
R.E. Stoller, L.R. Greenwood, From molecular dynamics to kinetic rate theory: a simple example of multiscale modeling, in *Multiscale modeling of materials*, ed. by V.V. Butalov, R.T. de la Diaz, P. Phillips, E. Kaxiras, N. Ghoniem (Materials Research Society, Warrendale, 1999), pp. 203–209
R.E. Stoller, L.K. Mansur, An Assessment of Radiation Damage Models and Methods. ORNL/TM-2005/506, 31 May
R.E. Stoller, S.I. Golubov, C. Domain, S. Becquart, Mean field rate theory and object kinetic Monte Carlo: a comparison of kinetic models. J. Nucl. Mater. **382**, 77–90 (2008)
Swissfel (2011), http://www.psi.ch/swissfel/why-swissfel. Accessed 20 Sept 2013
The ITER project (2013), http://www.iter.org/. Accessed 20 Sept 2013
M.B. Toloczko, D.S. Gelles, F.A. Garner, R.J. Kurtz, K. Abe, J. Nucl. Mater. **329–333**, 352 (2004)
H. Ullmaier, W. Schilling, *Radiation Damage in Metallic Reactor Materials*. Physics of modern materials, vol. 1 (IAEA, Vienna, 1980)
S. Valizadeh, R.J. Comstock, M. Dahlbäck, G. Zhou, J. Wright, L. Hallstadius, J. Romero, G. Ledergerber, S. Abolhassani, D. Jädernäs, E. Mader,Effects of secondary phase particle dissolution on the in-reactor performance of BWR cladding, in 16th Zr International Symposium, Chengdu (2010), http://www.astm.org/COMMIT/B10_Zirc_Presentations/5.3_Valizadeh_-_SPP_BWR.pdf. 9–13 May 2010
M.L. Wald, TR10: Traveling-Wave Reactor, http://www.technologyreview.com/biomedicine/22114. Accessed 20 Sept 2013
K. Wallin, Fracture toughness transition curve shape for ferritic structural steels, in *Joint FEFG/ICF International Conference on Fracture of Engineering Materials and Structures* (1991), Elsevier London
G. Was, *Fundamentals of Radiation Materials Science* (Springer, Berlin/Heidelberg, 2007)
G.S. Was, P.L. Andresen, Irradiation-assisted stress-corrosion cracking in austenitic alloys. J. Appl. Meteorol. **44**(4), 8–13 (1992)
G.S. Was, J. Busby, P.L. Andresen, Effect of irradiation on stress-corrosion cracking and corrosion in light water reactors, in *ASM Handbook 13C Corrosion Environments and Industries ASM international* (2006), pp. 386–414. doi:10.1361/asmhba0004147
C. Wassiliew, W. Schneider, K. Ehrlich, Creep and creep-rupture properties of type 1.4970 stainless steel during and after irradiation. Radiat. Eff. **101**, 201–219 (1986)
H. Wiedersich, Effects of the primary recoil spectrum on microstructural evolution. J. Nucl. Mater. **1799181**, 70–75 (1991)
H. Wiedersich, Evolution of Defect Cluster Distribution During Irradiation (1991), ANL/CP – 72655
P. Yvon, F. Carré, Structural materials challenges for advanced reactor systems. J. Nucl. Mater. **385**, 217–222 (2009)
S.J. Zinkle, P.J. Maziasz, R.E. Stoller, Dose dependence of the microstructural evolution in neutron-irradiated austenitic stainless steel. J. Nucl. Mater. **206**, 266–286 (1993)

Section XII

Recent Trends in Damage and Healing Mechanics

Healing, Super Healing, and Other Issues in Continuum Damage Mechanics

44

George Z. Voyiadjis, Peter I. Kattan, and Navid Mozaffari

Contents

Abstract

In this chapter, after a quick introduction on the literature of healing and super healing concept, the damage/healing mechanics principles are investigated. The concept of super healing of materials is then introduced into the framework of continuum damage mechanics (CDM). Super-healed material can be seen as a strengthened material by further healing when the whole damage is recovered by healing of a damaged material. Therefore, in this chapter the process of healing beyond what is necessary for damage recovery is called super healing. Super material is the final objective of the super healing process when the material

G.Z. Voyiadjis (✉) • P.I. Kattan • N. Mozaffari
Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA
e-mail: voyiadjis@eng.lsu.edu; pkattan@orange.jo; nmozaf2@lsu.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_45

achieves more stiffness at the end of super healing process. Then, by introducing the anisotropic super healing concept, these concepts are generalized in tensorial form to be used in anisotropic damage and healing of materials. Consequently, three fundamental issues in CDM are discussed. Nature of the damage process is investigated by dissecting the expression of the effective stress into an infinite geometric series. Several stages of damage are introduced which are termed primary, secondary, tertiary, etc., using this expression. New definition of the damage variable is then introduced for small damage cases. The new concept of undamageable materials that maintain a zero value of the damage variable throughout the deformation process is introduced. Finally and in the last section of the chapter, the forming of a singularity which leads to initiation of the process of fracture is shown in a continuous region within the framework of CDM. The internal damage processes leading to a singularity are illustrated mathematically. This section potentially provides a crucial link between the damage and fracture mechanics.

Introduction

Self-healing process in the material from damaged state has paid attention increasingly in the damage mechanics literature. Undergoing research shows two different self-healing mechanisms: one is an active or autonomous system as a coupled system in which damaging triggers self-healing (Pang and Bond 2005; Toohey et al. 2007; White et al. 2001); another is passive system as a decoupled system in which healing occurs after damage identification using external detection (John and Li 2010; Li and John 2008; Li and Muthyala 2008; Li and Nettles 2010; Li and Uppu 2010; Liu and Chen 2007; Nji and Li 2010a, b; Varley and van der Zwaag 2008; Zako and Takano 1999). Some significant new results on damage and healing of materials are presented by Pavan et al. (2010), Yuan and Ju (2012), and Zaïri et al. (2011). This process is observed experimentally even in nanoscale in wide range of materials (George and Warren 2002; Nemat-Nasser 1979, 1983; Voyiadjis and Park 1996; Wang and Sekerka 1996). Healing in constitutive models is used in two different approaches: one is used to characterize the healing process usually by a phenomenological approach (Miao et al. 1995), and the other is simple model (Adam 1999; Simpson et al. 2000). Thermodynamic-based damage and healing models are also introduced recently and are available in the literature (Barbero et al. 2005; Miao et al. 1995). However, constitutive modeling of self-healing material is still in progress since experimental aspect of healing has significant difficulties. During the past decades, progress has been made in the damage mechanics of various materials including elastoplastic models (Chaboche 1991; Ginzburg 1955; Kattan and Voyiadjis 1993; Lee et al. 1985; Naderi et al. 2012; Voyiadjis 1988; Voyiadjis and Kattan 1990, 1992; Voyiadjis et al. 2012); elasto-viscoplastic models (Chaboche 1997; Lemaitre and Chaboche 1990); continuum damage models (Kachanov 1958; Voyiadjis and Kattan 2009); materials surface degradation models including rolling, sliding contact fatigue, fretting fatigue, and

adhesive wear (Loginova et al. 2001; Singer-Loginova and Singer 2008; Wheeler et al. 1993); and coupled elastoplastic damage models (Chow and Jie 2009; Lemaitre 1985; Voyiadjis et al. 2009). The damage variable in scalar or second-order tensor forms shows the average material degradation (loss of stiffness). This variable lumps all kinds of defects such as micro-cracks, voids, and micro-cavities at the microscale level (Lubarda and Krajcinovic 1993; Voyiadjis and Kattan 2009). It is shown that in the case of isotropic damage, two independent damage variables are necessary to predict damage level (Cauvin and Testa 1999; van der Waals 1979). It has been argued that sufficient accuracy can be obtained to find certain parameters of damaged materials under the assumption of isotropic damage (Lemaitre 1984).

The representative volume element (RVE) is widely used in continuum damage mechanics in which the discontinuities (micro-voids, micro-cracks, etc.) are not considered explicitly in the RVE. The discontinuous and discrete elements of damage effects are lumped together through the use of a macroscopic internal variable. Phenomenological approach is adopted usually, and the consistent formulation is derived using sound mechanical and thermodynamic principles. Thermodynamically consistent framework is achieved using the concept of macroscopic internal variables which is used to lump the effect of all defects (Ginzburg and Landau 1965; Hansen and Schreyer 1994; Landau and Ter Haar 1965; Miao et al. 1995; Murakami 1983; Voyiadjis and Park 1997; Voyiadjis and Kattan 2006, 2012b; Voyiadjis and Park 1995; Voyiadjis et al. 2009). The concept of effective stress for uniaxial tension was first introduced by Kachanov (1958) and Rabotnov (1968). It has been shown that the isotropic damage assumption is sufficiently enough to predict the load carrying capacity and the number of cycles or the time to local failure in structural components (Kattan and Voyiadjis 2001; Voyiadjis and Kattan 2005, 2006). However, anisotropic damage propagation has been observed experimentally (Lee et al. 1985; Sidoroff 1981) even in an initially isotropic solid. The damage variable is considered in scalar form in the case of isotropic damage mechanics, and the evolution equations can be handled easily (Voyiadjis and Kattan 2009). The concept of an undamageable material was proposed recently by Voyiadjis and Kattan (2012b, 2013c, d). This kind of material is considered as a hypothetical material that cannot be damaged during the loading process. Furthermore, decomposition of the damage tensor into two damage components, one due to cracks and one due to voids, is developed by Kattan and Voyiadjis (2001). Finally, a conceptual framework for general damage processes operating in series and in parallel is introduced by Voyiadjis and Kattan (2012a).

In this chapter, after a quick introduction on the literature of healing and super healing concept, the damage/healing mechanics principles are investigated. The concept of super healing of materials is then introduced into the framework of continuum damage mechanics. Super-healed material can be seen as a strengthened material by further healing when the whole damage is recovered by healing of a damaged material. Therefore, in this chapter the process of healing beyond what is necessary for damage recovery is called super healing. Super material is the final

objective of the super healing process when the material achieves more stiffness at the end of super healing process. Consequently, by introducing the anisotropic super healing concept, these concepts are generalized in tensorial form to be used in anisotropic damage and healing of materials. Mechanics of damage/healing and super healing are illustrated using an example for the case of plane stress. Consequently, the characteristics of the super material are outlined in the following section. Furthermore, three fundamental issues in continuum damage mechanics are discussed, and fracture mechanics and damage mechanics are linked through the internal damage process leading to singularity in continuous regions.

Review of Damage/Healing Mechanics

In this section, the damage/healing mechanics principles are reviewed considering a fictitious undamaged material configuration as shown in Fig. 1. Any nonzero component of the effective Cauchy stress tensor $\overline{\sigma}_{ij}$ such as the tension component $\overline{\sigma}$ can be obtained by the following relation (Kachanov 1958; Rabotnov 1963; Sidoroff 1981; Voyiadjis and Kattan 2006, 2009):

$$\overline{\sigma} = \frac{\sigma}{1-\varphi} \tag{1}$$

σ is the corresponding component of the Cauchy stress tensor and φ is the isotropic damage variable. The damage variable φ changes between 0 and 1. It is worth to mention that the value $\varphi = 0$ shows the undamaged state while complete failure (fracture) happens when the value of φ tends to 1.

Defining an intermediate configuration as partially healed material which is indicated in Fig. 2 and considering this configuration as a combination of damage and healing between undamaged and damaged state as shown in Fig. 3, the effective stress $\overline{\sigma}$ is written as (Chow and Wang 1987; Park and Voyiadjis 1998; Voyiadjis and Park 1997; Voyiadjis et al. 2012)

$$\overline{\sigma} = \frac{\sigma}{1-\varphi(1-h)} \tag{2}$$

where h is the healing variable. The healing variable h is also changes 0 and 1. It is worth to mention that the value $h = 0$ shows the absence of healing; therefore, Eq. 1 can be obtained from Eq. 2 by substituting $h = 0$ in this case. On the other hand, the value $h = 1$ describes complete healing, i.e., recovery of all the damage. In this case, material goes back to the beginning of the loading (undamaged state), and the actual stress and the effective stress are equal in Eq. 2. Comparison between Eqs. 1 and 2 confirms the replacement of the damage variable φ in Eq. 1 by the variable $\varphi\ (1 - h)$ in Eq. 2 in self-healing materials. Both damage and healing effects are combined through this new variable in one single parameter

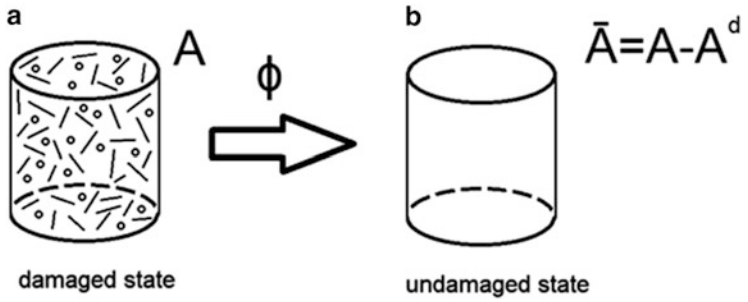


Fig. 1 Damage configurations in terms of cross-sectional area reduction (After Voyiadjis and Kattan 2013a)

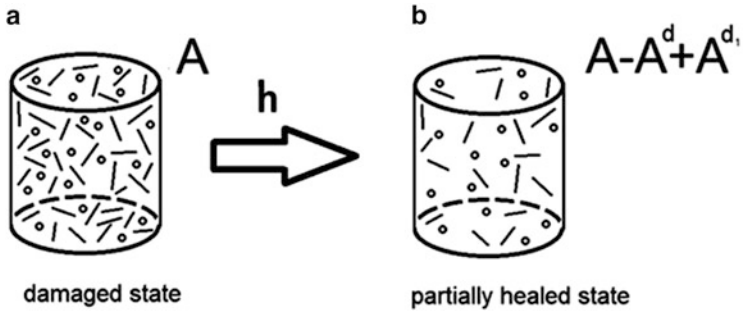


Fig. 2 Healing configurations in terms of cross-sectional area reduction (After Voyiadjis and Kattan 2013a)

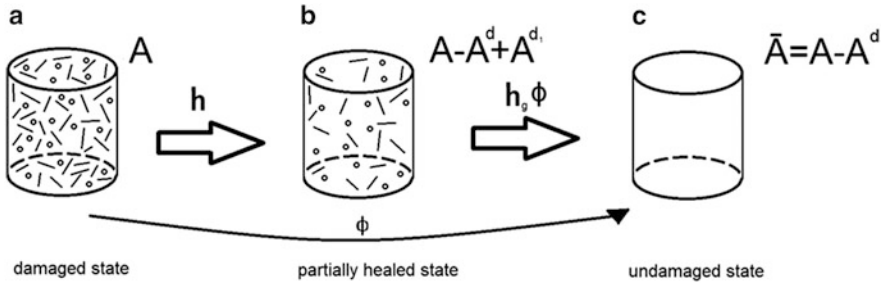


Fig. 3 Damage and healing configurations in terms of cross-sectional area reduction (After Voyiadjis and Kattan 2013a)

which is called here the combined damage/healing variable (Chow and Wang 1987) termed this variable as the effective damage variable. However, Eq. 2 is not written explicitly by Park and Voyiadjis (1998) and Voyiadjis and Park (1997), and it is not recognized as the combined damage/healing variable as a single parameter.

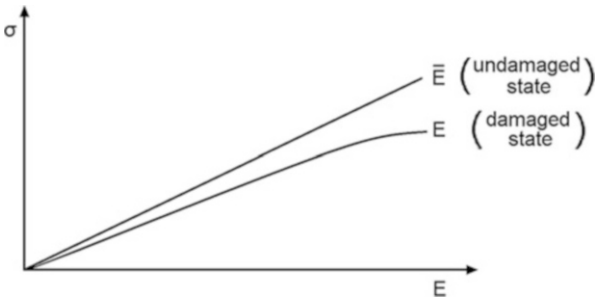


Fig. 4 Damage states in terms of elastic stiffness degradation (After Voyiadjis and Kattan 2013a)

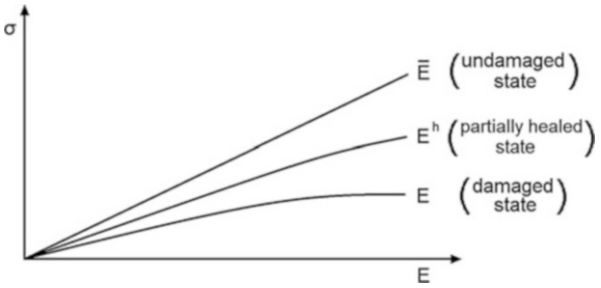


Fig. 5 Damage and healing states in terms of elastic stiffness degradation (After Voyiadjis and Kattan 2013a)

Based on Eq. 2, the complete failure occurs when the value of the new variable (combined damage/healing parameter $\varphi\ (1 - h)$) tends to 1. Two different situations can occur in the absence of damage when the new variable is zero: first one is when $\varphi = 0$ (undamaged virgin material), and the other one is when $h = 1$ (completely healed material). Schematic stress–strain curves are shown in Figs. 4 and 5 for damaged and damaged/healed materials in elastic region, respectively. It is worth to mention that reformulation of the damage and healing principles (Eqs. 1 and 2) in terms of the elastic stiffness as shown in Figs. 4 and 5 is possible, but this is beyond the scope of this chapter.

Experimental test results on bituminous materials are shown in Figs. 6 and 7. These experiments show the capabilities of self-healing materials and the relationship between the healing time and the healing percentage which were conducted by Murray et al. (1995). In their work, Qiu et al. elucidated several characteristics of bituminous materials as related to healing and damage.

It should be noted that the healing parameter h corresponds exactly to the healing percentage which are shown in Figs. 6 and 7. The experimental results fully conform to the theoretical damage/healing mechanics framework which is summarized in this section. Upon further examination of Fig. 7, it is finally clear that the healing effect is limited in the first 10 h, but healing improves significantly during the 100th and 1,000th hours. Certain biological materials like bones can be considered as other applications on healing of materials.

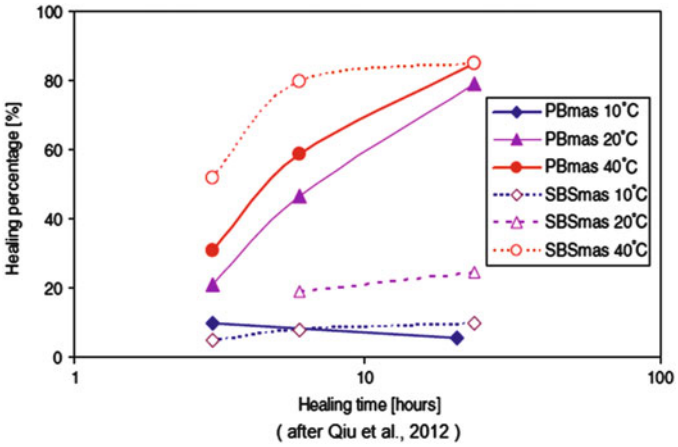


Fig. 6 A practical example of healing time versus healing percentage (After Voyiadjis and Kattan 2013a)

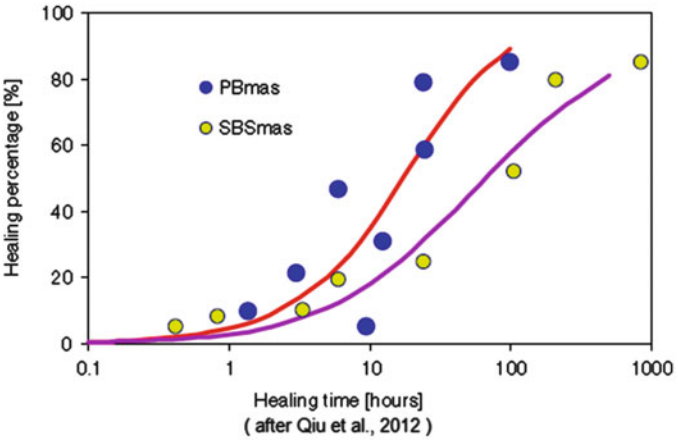


Fig. 7 The relation of healing time to healing percentage in a practical example (After Voyiadjis and Kattan 2013a)

Introduction to Super Healing

In classical damage/healing mechanics, the healing parameter h changes between 0 and 1 in which $h = 0$ implies zero healing and $h = 1$ implies complete healing. Assume that after complete damage recovery, i.e., when $h = 1$, the healing process continues beyond $h = 1$. This will enable us to use large values for h, i.e., h is allowed to take values like 2, 3, 4, ... $n + 1$. This special case is called super

healing in which some form of strengthening or enhancing the properties of the material occurs instead of further healing since the material is now undamaged (was completely healed when healing variable reaches to 1 $h = 1$). The values of n are limited to integer values only in the following derivation, and it is not allowed to take real values at this stage. The details for this issue are outlined in section "Characteristics of the Super Material." Therefore, in the super healing phase, the values of the healing parameter h start at 1 and increase to take the values 2, 3, 4, ... $n + 1$. It is obvious that by high increasing of the value of the healing parameter h, a strengthened material will be obtained. Hypothetical material which is called here the super material is postulated as the value of h approaches infinity. Full characterization of the super material is not the intention of this chapter, but only the way to its realization is pointed out. Some of the characteristics of the super material are postulated in section "Characteristics of the Super Material." It is hoped that future technology will be able to manufacture strong and self-healing materials that come as close as possible to the proposed theoretical super material. Establishing governing equations for the proposed super healing materials is the main aim of this chapter. In this work, what interests us is the mechanics of the process of super healing and not the final super material which remains theoretical and hypothetical at this time.

Equation 2 changes to the following form by assuming the value of the healing parameter h is increased through super healing to $n + 1$.

$$\overline{\sigma} = \frac{\sigma}{1 + n\varphi} \tag{3}$$

Equation 3 can be considered as the main expression of the process of super healing. Based on Eq. 3, it can be seen that as the value n tends to infinity, the value of the effective stress tends to zero. Thus, the first characteristic of the super material is concluded as its effective stress is zero, and it does not depend on the value of either the damage parameter φ or the healing parameter h. This process is called super healing of order n based on Eq. 3. Super-healed material can be obtained by continuing super healing at different stages as the value of h increases from 1 to 2 to 3 and so on. The super-healed material of order n will be finally obtained when the process of super healing of order n takes its due course.

In damage mechanics, the effective stress becomes infinity which indicates complete rupture of the material when $\varphi = 1$ (see Eq. 1). However, the effective stress does not explode but takes a finite value in super-healed materials when $\varphi = 1$ (see Eq. 3). This means that rupture does not occur in super-healed materials even when the value of the damage parameter approaches to infinity. This can be seen as the significance of the super healing process. Based on Eq. 2, it can be seen clearly that the effective stress explodes when the value of the combined damage/healing parameter $\varphi\ (1 - h)$ tends to one. This critical case cannot occur in super-healed materials since by setting $\varphi\ (1 - h) = 1$ the relation

for $h = \frac{\varphi - 1}{\varphi}$ is obtained and based on this relation by assuming the value of the damage parameter $\varphi = 1$, then the obtained value for h is $h = 0$ which is not possible. The value of the healing parameter will be negative if the value of the damage parameter is less than 1 which is not possible again. Another possibility is when both the damage and healing parameters take values larger than 1; this case cannot occur since the damage parameter is bounded to be less or equal to 1. From the beginning of this chapter up to this section, a certain mechanism for the super healing process is proposed. By admitting higher values for h larger than 1 up to $n + 1$, the process of super healing is achieved. One single healing mechanism is employed in the first approach and another approach which employs multiple healing mechanisms operating in parallel (i.e., at the same time) for the process of super healing to occur. The equivalency of the two approaches is shown in the following part. Multiple healing parameters h_1, h_2, h_3, ..., h_n characterizing multiple healing mechanisms that are operating at the same time are utilized instead of one single healing parameter h. In this case, the value of each healing parameter is limited to 1 only since there is no need to go beyond 1 in this approach for super healing. Therefore, Eq. 2 can be rewritten in the following form:

$$\overline{\sigma} = \frac{\sigma}{1 - \varphi\,(1 - h_1 - h_2 - ..... - h_n)} \tag{4}$$

In order to derive Eq. 4, one follows the same procedure used in deriving Eq. 2. For more details, check the references by Chow and Wang (1987), Park and Voyiadjis (1998), and Voyiadjis and Park (1997).

The proposed process of super healing can be achieved by utilizing Eq. 4 characterizing multiple healing parameters (n parameters) operating at the same time and by using the value 1 for each separate healing parameter. Furthermore, the elusive super material is considered again if the number of healing parameters approaches infinity (with each one limited to 1). Equivalency of the two approaches to super healing of Eqs. 2 and 4 is shown in as follows. Resultant healing parameter h is defined as the sum of the multiple healing parameters h_1, h_2, h_3,....., h_n as follows:

$$h = h_1 + h_2 + h_3 + ... + h_n \tag{5}$$

Therefore, the process of super healing can be obtained in two different ways: first by using one single large healing parameter h with values 1, 2, 3, ... n and second by using multiple small healing parameters h_1, h_2, h_3, ..., h_n operating together but with limit value of one for each parameter. It can be seen that by substituting the value of 1 for each single parameter in Eq. 4, governing equation of super healing (Eq. 3) can be obtained, and thus both approaches are the same. Finally, the following question needs to be asked: What happens by allowing each multiple healing parameter h_1, h_2, h_3, ..., h_n to have a value larger than 1? This

will open the way for more healing and super healing. For instance, allowing each healing parameter to take the value of 2, then Eq. 3 reads

$$\overline{\sigma} = \frac{\sigma}{1 + 2n\varphi} \tag{6}$$

This is called Level 2 Super Healing characterized by Eq. 6. In general, the following expression can be obtained which characterizes the process of super healing in the case where each healing parameter is allowed to take the value of n

$$\overline{\sigma} = \frac{\sigma}{1 + n^2\varphi} \tag{7}$$

This is called Level n Super Healing characterized by Eq. 7.

Anisotropic Damage/Healing Mechanics

The theory of damage/healing mechanics of section "Review of Damage/Healing Mechanics" is generalized to anisotropic damage/healing in this section (Chow and Wang 1987; Park and Voyiadjis 1998; Voyiadjis and Park 1997). For this purpose, tensors are used instead of scalars. In the following, capital letters are used to denote fourth-rank tensors, and it is assumed that tensors are represented by matrices. Let M denote the fourth-rank damage effect tensor of continuum damage mechanics. The exact relationship between the fourth-rank damage effect tensor M and the scalar damage variable φ is extensively investigated in the literature (Sidoroff 1981; Voyiadjis and Kattan 2006, 2009).

Let H denote a fourth-rank healing tensor corresponding to the healing parameter h. For the exact relationship between H and h, see Park and Voyiadjis (1998) and Voyiadjis and Park (1997). In the case of anisotropic healing and damage, Eq. 2 can be generalized as follows (Park and Voyiadjis 1998; Voyiadjis and Park 1997):

$$\overline{\sigma}_{ij} = \left[M^{-1}_{ijkl} + \left(I_{ijmn} - M^{-1}_{ijmn} \right) : H^{-1}_{mnkl} \right]^{-1} \sigma_{kl} \tag{8}$$

where I_{ijmn} is the fourth-rank identity tensor. In Eq. 8, σ_{kl} and $\overline{\sigma}_{ij}$ are second-rank stress tensors and are represented by vectors in section "Damage/Healing and Super Healing in Plane Stress." Based on Eq. 8, it is shown that the combined damage/healing parameter φ $(1 - h)$ is generalized to $(I_{ijmn} - M_{ijmn}{}^{-1})$ $(I_{mnkl} - H_{mnkl}{}^{-1})$. The latter expression for combined damage and healing can be derived directly from Eq. 8 considering the facts that the main component of M corresponds to $\frac{1}{1-\varphi}$ and the main component of H corresponds to $\frac{1}{h}$. The fourth-rank healing tensor H clearly satisfies some mathematical properties such as the components of this tensor are either positive or zero, and both the trace and norm of the tensor are positive. However, the tensor H may not necessarily be positive definite. These properties are illustrated with the example of plane stress which is given in section "Damage/Healing and Super Healing in Plane Stress."

Anisotropic Super Healing

The process of super healing for anisotropic damage and healing mechanism is shown in this section. Following the outlined reasoning in section "Introduction to Super Healing" for the proposed mechanism of super healing, the components of the fourth-rank healing tensor H is allowed to increase gradually beyond the values of the components of the fourth-rank identity tensor I, i.e., set $H_{ijkl} = (n + 1)\, I_{ijkl}$. In another way, previous relation can be written between the norms of the two fourth-rank tensors. The following expression can be obtained by substituting previous relation in Eq. 8 and simplifying

$$\overline{\sigma}_{ij} = \left[n \left(I_{ijkl} - M^{-1}_{ijkl} \right) + I_{ijkl} \right]^{-1} \sigma_{kl} \tag{9}$$

The process of anisotropic super healing is characterized by the expression obtained in Eq. 9. Based on Eq. 9, when the value of n tends to infinity, the value of the effective stress goes to zero. The same conclusion was obtained for super healing in the case of scalar damage and healing in previous sections. It should be noted that by implementing the appropriate constraints, the anisotropic super healing in Eq. 9 reduces to the scalar super healing in Eq. 3.

Damage/Healing and Super Healing in Plane Stress

Plane stress is solved to illustrate the processes of damage, healing, and super healing in this section. For this special case, the tensors of Eq. 8 are represented by vectors and matrices (Voyiadjis and Kattan 2006):

$$\{\sigma\} = \begin{Bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{12} \end{Bmatrix} \tag{10a}$$

$$\{\overline{\sigma}\} = \begin{Bmatrix} \overline{\sigma}_{11} \\ \overline{\sigma}_{22} \\ \overline{\sigma}_{12} \end{Bmatrix} \tag{10b}$$

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{10c}$$

$$M = \frac{1}{\Delta} \begin{bmatrix} \psi_{22} & 0 & \varphi_{12} \\ 0 & \psi_{11} & \varphi_{12} \\ \dfrac{\varphi_{12}}{2} & \dfrac{\varphi_{12}}{2} & \dfrac{\psi_{11} + \psi_{22}}{2} \end{bmatrix} \tag{10d}$$

where $\psi_{11} = 1 - \varphi_{11}$ and $\psi_{22} = 1 - \varphi_{22}$. The denominator Δ in Eq. 10d is given by Voyiadjis and Kattan (2006):

$$\Delta = \psi_{11}\psi_{22} - \varphi_{12}^2 \tag{10e}$$

It should be noted that Eq. 10d is obtained based on the symmetrization procedure given by the following expression (Voyiadjis and Kattan 2006):

$$\overline{\sigma}_{ij} = \frac{1}{2}\left[\sigma_{ip}\left(\delta_{pj} - \varphi_{pj}\right)^{-1} + \left(\delta_{ip} - \varphi_{ip}\right)^{-1}\sigma_{pj}\right] \tag{10f}$$

The following 3×3 matrix representation can be written for the healing tensor of Eq. 8 for the case of plane stress:

$$H^{-1} = \begin{bmatrix} h_{11} & 0 & h_{12} \\ 0 & h_{22} & h_{12} \\ h_{12} & h_{12} & \dfrac{h_{11} + h_{22}}{2} \end{bmatrix} \tag{11}$$

Based on Eq. 11, the inverse healing tensor satisfies certain mathematical properties such that the components of this tensor are either positive or zero and the trace of this tensor $\left(h_{11} + h_{22} + \frac{h_{11}+h_{22}}{2}\right)$ is positive. Also, the norm of this tensor $\left(\sqrt{h_{11}^2 + h_{22}^2 + h_{12}^2}\right)$ is positive. Since the expression ${h_{11}}^2 + {h_{22}}^2 - {h_{12}}^2$ is not necessarily positive, this tensor is not necessarily positive definite.

It can be seen that the form of the inverse healing matrix which is postulated in Eq. 11 is similar to the form of the damage effect matrix in Eq. 10d. Substituting Eqs. 10f and 11 into Eq. 8 results in

$$\{\sigma\} = [X]\{\overline{\sigma}\} \tag{12}$$

where the components of the fourth-rank tensor $[X]$ are given by (the MATLAB Symbolic Math Toolbox is used to carry out the algebraic manipulations)

$$\begin{aligned} X_{11} = \big(&-2 + 3\varphi_{11} + \varphi_{22} - \varphi_{11}\varphi_{22} - \varphi_{11}^2 + \varphi_{12}^2 - 2h_{11}\varphi_{11} + h_{11}\varphi_{11}\varphi_{22} + h_{11}\varphi_{11}^2 \\ &- h_{11}\varphi_{12}^2 - 2h_{12}\varphi_{12} + 2h_{12}\varphi_{11}\varphi_{12}\big)/(\varphi_{11} + \varphi_{22} - 2) \end{aligned} \tag{13a}$$

$$X_{12} = \varphi_{12}(-\varphi_{12} + h_{22}\varphi_{12} - 2h_{12} + 2h_{12}\varphi_{11})/(\varphi_{11} + \varphi_{22} - 2) \tag{13b}$$

$$\begin{aligned} X_{13} = \big(&2\varphi_{12} - 2\varphi_{11}\varphi_{12} - 2h_{12}\varphi_{11} + h_{12}\varphi_{11}^2 + h_{12}\varphi_{11}\varphi_{22} - h_{11}\varphi_{12} - h_{22}\varphi_{12} \\ &+ h_{11}\varphi_{11}\varphi_{12} + h_{22}\varphi_{11}\varphi_{12}\big)/(\varphi_{11} + \varphi_{22} - 2) \end{aligned} \tag{13c}$$

$$X_{21} = \varphi_{12}(-\varphi_{12} + h_{11}\varphi_{12} - 2h_{12} + 2h_{12}\varphi_{22})/(\varphi_{11} + \varphi_{22} - 2) \tag{13d}$$

$$X_{22} = \big(-2 + \varphi_{11} + 3\varphi_{22} - \varphi_{11}\varphi_{22} - \varphi_{22}^2 + \varphi_{12}^2 - 2h_{22}\varphi_{22} + h_{22}\varphi_{11}\varphi_{22} + h_{22}\varphi_{22}^2 - h_{22}\varphi_{12}^2 - 2h_{12}\varphi_{12} + 2h_{12}\varphi_{22}\varphi_{12}\big)/(\varphi_{11} + \varphi_{22} - 2) \tag{13e}$$

$$X_{23} = \big(2\varphi_{12} - 2\varphi_{22}\varphi_{12} - 2h_{12}\varphi_{22} + h_{12}\varphi_{22}^2 + h_{12}\varphi_{11}\varphi_{22} - h_{11}\varphi_{12} - h_{22}\varphi_{12} + h_{11}\varphi_{22}\varphi_{12} + h_{22}\varphi_{22}\varphi_{12}\big)/(\varphi_{11} + \varphi_{22} - 2) \tag{13f}$$

$$X_{31} = \big(\varphi_{12} - \varphi_{11}\varphi_{12} - h_{12}\varphi_{11} - h_{12}\varphi_{22} - h_{11}\varphi_{12} + h_{11}\varphi_{11}\varphi_{12} + 2h_{12}\varphi_{11}\varphi_{22}\big)/(\varphi_{11} + \varphi_{22} - 2) \tag{13g}$$

$$X_{32} = \big(\varphi_{12} - \varphi_{22}\varphi_{12} - h_{12}\varphi_{11} - h_{12}\varphi_{22} - h_{22}\varphi_{12} + h_{22}\varphi_{22}\varphi_{12} + 2h_{12}\varphi_{11}\varphi_{22}\big)/(\varphi_{11} + \varphi_{22} - 2) \tag{13h}$$

$$X_{33} = \frac{1}{2}\big(-4 + 4\varphi_{11} + 4\varphi_{22} - 4\varphi_{11}\varphi_{22} - h_{11}\varphi_{11} - h_{11}\varphi_{22} - h_{22}\varphi_{11} - h_{22}\varphi_{22} - 4h_{12}\varphi_{12} + 2h_{12}\varphi_{11}\varphi_{12} + 2h_{12}\varphi_{22}\varphi_{12} + 2h_{11}\varphi_{11}\varphi_{22} + 2h_{22}\varphi_{11}\varphi_{22}\big)/(\varphi_{11} + \varphi_{22} - 2) \tag{13i}$$

Special case, i.e., the case of principal components, is considered here for the sake of simplicity. For this case, by setting $\varphi_{12} = \varphi_{21} = 0$ and $h_{12} = h_{21} = 0$ in Eq. 13, the following expressions can be obtained:

$$X_{11} = \big(-2 + 3\varphi_{11} + \varphi_{22} - \varphi_{11}\varphi_{22} - \varphi_{11}^2 - 2h_{11}\varphi_{11} + h_{11}\varphi_{11}^2 + h_{11}\varphi_{11}\varphi_{22}\big)/(\varphi_{11} + \varphi_{22} - 2) \tag{14a}$$

$$X_{22} = \big(-2 + \varphi_{11} + 3\varphi_{22} - \varphi_{11}\varphi_{22} - \varphi_{22}^2 - 2h_{22}\varphi_{22} + h_{22}\varphi_{22}^2 + h_{22}\varphi_{11}\varphi_{22}\big)/(\varphi_{11} + \varphi_{22} - 2) \tag{14b}$$

$$X_{33} = \frac{1}{2}\big(-4 + 4\varphi_{11} + 4\varphi_{22} - 4\varphi_{11}\varphi_{22} - h_{11}\varphi_{11} - h_{11}\varphi_{22} - h_{22}\varphi_{11} - h_{22}\varphi_{22} + 2h_{11}\varphi_{11}\varphi_{22} + 2h_{22}\varphi_{11}\varphi_{22}\big)/(\varphi_{11} + \varphi_{22} - 2) \tag{14c}$$

It can be seen that based on Eq. 13 for this special case, all the other components of the fourth-rank tensor [*X*] vanish. Simplifying Eq. 14 and substituting into Eq. 12 lead to obtain the following simple expressions (after some tedious algebraic manipulations):

$$\overline{\sigma}_{11} = \frac{\sigma_{11}}{1 - \varphi_{11}(1 - h_{11})} \tag{15a}$$

$$\overline{\sigma}_{22} = \frac{\sigma_{22}}{1 - \varphi_{22}(1 - h_{22})} \tag{15b}$$

$$\overline{\sigma}_{12} = \frac{\sigma_{12}}{1 - \left(1 - \frac{h_{11} + h_{22}}{2}\right)\left(\frac{\varphi_{11} + \varphi_{22} - 2\varphi_{11}\varphi_{22}}{\varphi_{11} + \varphi_{22} - 2}\right)} \tag{15c}$$

It can be seen from Eq. 15 that the principal equations of damage and healing (Eqs. 15a and 15b) reduce to the expression of the scalar case of Eq. 2 for the case of plane stress and the proposed super healing process is valid in the case of plane stress. This can be achieved by substituting values for the healing parameters h_{11} and h_{22} exceeding 1 and approaching a large number, and then the effective stress tends to zero. In addition, when these values tend to infinity, the elusive super material is obtained.

Characteristics of the Super Material

The use is made of the theories of super healing and undamageable materials to elucidate some of the characteristics of the sought super material within the continuum damage mechanics framework in this section. The concept of an undamageable material is proposed by Kobayashi (1992), Voyiadjis and Kattan (2012b), and Warren and Boettinger (1995). The value of the damage variable remains zero in this hypothetical type of material throughout the deformation process. Theoretically, such materials cannot be damaged. Constitutive equations of undamageable materials are derived through introducing a new type of material called the Voyiadjis–Kattan material of order n (Kobayashi 1992). This material type is a nonlinear elastic material which has a nonlinear strain energy form. Voyiadjis–Kattan materials of order n are based on higher-order strain energy forms that assume the general form

$$U = \frac{1}{2}\sigma\varepsilon^n \tag{16}$$

Nonlinear stress–strain relation for the given form of higher-order strain energy in Eq. 16 reads

$$\sigma = E\,\frac{1}{\varepsilon^n}\,e^{-2/\left[(n-1)\varepsilon^{(n-1)}\right]} \tag{17}$$

The reader is referred to Kobayashi (1992), Voyiadjis and Kattan (2012b), and Warren and Boettinger (1995) for full derivation and more details. This general form (Eq. 17) satisfies the initial conditions $\sigma = 0$ when $\varepsilon = 0$ since the limit of the expression for the stress tends to zero as the strain tends to zero. Equations 16 and 17 are valid for one-dimensional cases only. It can be concluded reasonably that

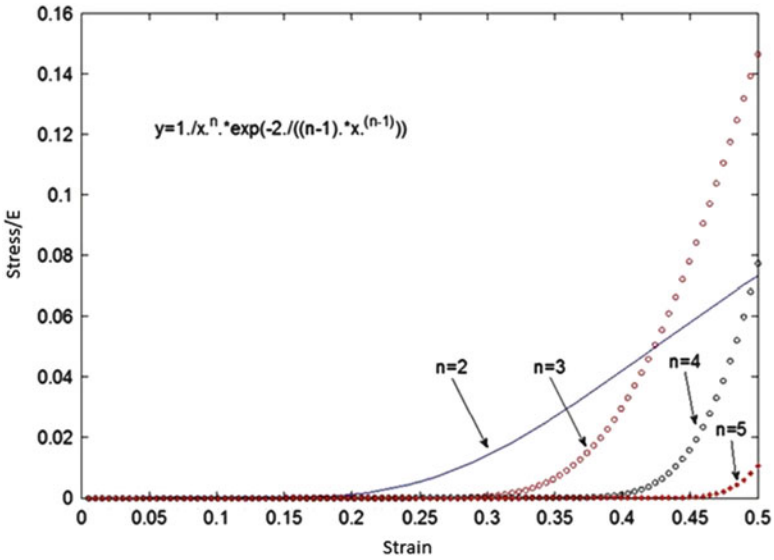


Fig. 8 Stress–strain curves based on Eq. 17 (After Voyiadjis and Kattan 2013a)

Voyiadjis–Kattan materials (and ultimately the undamageable material) can be manufactured in the future using a process based on the super healing model as outlined here. It can be seen that the Voyiadjis–Kattan material of order n is the same as a super-healed material of order n. Since the proposed higher-order strain energy forms (Eq. 16) admits integer values of the exponent n, it can be concluded that the super healing process also admits such integer values for n.

The following characteristics of the super material are outlined here based on the theory developed in sections "Introduction to Super Healing" and "Anisotropic Super Healing" in this chapter and the theory of undamageable materials as formulated by Gránásy et al. (2002), Kobayashi (1992), and Warren and Boettinger (1995).

The super material must be undamageable. Therefore, the properties of undamageable materials apply also to the super material. These properties are as outlined below. The value of the stress will remain equal to zero throughout the deformation process. The value of the damage variable will be equal to zero also throughout the deformation process. The super material has zero strain energy. This property is directly derived from above. The super material has nonzero strain values. Thus, the super material is a type of deformable material, not a rigid body. The super material is based on the proposed higher-order strain energy form of Eq. 16 taken in the limit when $n \rightarrow \infty$. The stress–strain relationship for the super material may be obtained from the elastic relation in Eq. 17 taken in the limit as $n \rightarrow \infty$. Some of the above items may be clearly deduced from the limit of Eq. 17. These characteristics are also clearly evident in Fig. 8 which was plotted based on Eq. 17.

Three Fundamental Issues in Continuum Damage Mechanics

In this section, the nature of the damage process within the framework of CDM is investigated by dissecting the expression of the effective stress into an infinite geometric series as follows:

$$ar + ar^2 + ar^3 + \ldots\ldots\ldots = \frac{a}{1-r} \tag{18}$$

The above geometric series is valid for $|r| < 1$. The effective stress given by

$$\overline{\sigma} = \frac{\sigma}{1-\varphi} \tag{19}$$

The effective stress given in Eq. 19 is the classical expression which is taken from the theory of CDM. In Eq. 19, φ is the damage variable (its value lies between zero and one), σ is the Cauchy stress, while $\overline{\sigma}$ is the corresponding effective stress. Comparing the right-hand side of the geometric series in Eq. 18 and the effective stress expression in Eq. 19 leads to the conclusion that the effective stress is equal to the sum of the infinite geometric series. It satisfies the condition since $0 < \varphi < 1$. Thus, by making analogy with the infinite geometric series in Eq. 18, the effective stress (Eq. 19) can be written as follows:

$$\overline{\sigma} = \sigma\left(1 + \varphi + \varphi^2 + \varphi^3 + \ldots\ldots\ldots\right) \tag{20}$$

Equation 20 is an infinite series exact relationship. It can be interpreted physically by considering the damage process as an infinite number of smaller damage processes or stages. Equation 20 can be rewritten in the following form since $\overline{\sigma}\overline{A} = \sigma A$, where A is the cross-sectional area and $\overline{A}$ is the effective cross-sectional area (the cross-sectional area in the fictitious effective configuration):

$$\frac{A}{\overline{A}} = 1 + \varphi + \varphi^2 + \varphi^3 + \ldots\ldots\ldots \tag{21}$$

Equation 21 shows that the damage process can be considered as the summation of several smaller damage processes or stages: the primary damage stage by taking the first two terms of the series, the secondary damage stage by taking the first three terms of the series, and the tertiary damage stage by taking the first four terms of the series. Although this process can be continued in an infinite number of smaller and smaller damage stages mathematically, but considering the first four terms of the infinite geometric series is sufficient for practical purposes.

Primary Damage Variable The first two terms of the series in Eq. 21 is considered to define the primary damage variable as follows:

$$\frac{A}{\overline{\overline{A}}} = 1 + \varphi_p \tag{22}$$

Equation 22 can be solved explicitly to obtain the following expression of the primary damage variable:

$$\varphi_p = \frac{A}{\overline{\overline{A}}} - 1 \tag{23}$$

Secondary Damage Variable The first three terms of the series in Eq. 21 are considered for defining the secondary damage variable as follows:

$$\frac{A}{\overline{\overline{A}}} = 1 + \varphi_S + \varphi_S^2 \tag{24}$$

The quadratic equation (Eq. 24) can be solved explicitly to obtain the following expression of the secondary damage variable:

$$\varphi_S = -\frac{1}{2} + \frac{1}{2}\sqrt{-3 + 4\frac{A}{\overline{\overline{A}}}} \tag{25}$$

Tertiary Damage Variable The first four terms of the series in Eq. 21 are considered to define the tertiary damage variable as follows:

$$\frac{A}{\overline{\overline{A}}} = 1 + \varphi_t + \varphi_t^2 + \varphi_t^3 \tag{26}$$

The cubic equation (Eq. 26) can be solved explicitly to obtain the following expression of the tertiary damage variable:

$$\varphi_t = -\frac{1}{3} + \frac{1}{6}\sqrt[3]{-80 + 108\frac{A}{\overline{\overline{A}}} + 12\sqrt{48 - 120\frac{A}{\overline{\overline{A}}} + 81\left(\frac{A}{\overline{\overline{A}}}\right)^2}} - \frac{4}{3\sqrt[3]{-80 + 108\frac{A}{\overline{\overline{A}}} + 12\sqrt{48 - 120\frac{A}{\overline{\overline{A}}} + 81\left(\frac{A}{\overline{\overline{A}}}\right)^2}}} \tag{27}$$

Thus, the explicit expressions for the damage variables at the primary, secondary, and tertiary damage stages have been established. In the next subsection, dissection of the damage process into the aforementioned three stages is presented mathematically.

Small Damage Processes

In this section, the problem of small damage processes is observed in details. In Voyiadjis and Mozaffari (2013), the following generalized relationship between the Cauchy stress and the effective stress is derived using the phase field method:

$$\overline{\sigma} = \frac{\sigma}{(1-\varphi)\sqrt{2\varphi+1}} \tag{28}$$

It is worth to mention that the effective stress expression given in Eq. 28 results in to a cubic formula for φ in terms of areas. The reader is referred to the work by authors for more details (Voyiadjis and Mozaffari 2013).

Equation 28 is compared with the classical expression given in Eq. 19. Considering the square root term that appears in the denominator of Eq. 28, the Taylor series expansion of the square root function is used to obtain the following approximation by taking the first two terms of the expansion:

$$\sqrt{2\varphi+1} \approx 1+\varphi \tag{29}$$

Equation 29 is valid for small values of φ, i.e., for small damage. The following expression for the effective stress of Eq. 28 can be written in the case of small damage:

$$\overline{\sigma} = \frac{\sigma}{(1-\varphi)(1+\varphi)} = \frac{\sigma}{1-\varphi^2} \tag{30}$$

The above formula (Eq. 30) corresponds to the damage variable $\varphi = \sqrt{\frac{A-\overline{A}}{A}}$ compared with $\varphi = \frac{A-\overline{A}}{A}$ for the classical case. Alternatively, the following expression for the effective stress is postulated in the case of large damage:

$$\overline{\sigma} = \frac{\sigma}{1-\sqrt{\varphi}} \tag{31}$$

which corresponds to the damage variable $\varphi = \left(\frac{A-\overline{A}}{A}\right)^2$

One can now generalize the two expressions for the effective stress (Eqs. 30 and 31) with an exponent n for the two different cases (small and large damage) when the increasing exponent goes from 2 to 3, 4, ... and ultimately to n where n tends to infinity. Therefore, the two following generalized definitions are proposed:

$$\overline{\sigma} = \frac{\sigma}{1-\varphi^n} \qquad \text{for small damage} \tag{32}$$

$$\overline{\sigma} = \frac{\sigma}{1-\varphi^{1/n}} \qquad \text{for large damage} \tag{33}$$

It can be seen that for normal (intermediate) damage then n = 1 in both cases.

The Concept of Undamageable Materials

In this section, previous issues are used to elaborate on the new undamageable material concept. These hypothetical materials were proposed recently by Voyiadjis and Kattan (2012b, 2013c, d) which can be compared with rubber materials (Arruda and Boyce 1993). Undamageable materials are compared with various nonlinear elastic materials taken from the book of Bower (2011) by Voyiadjis and Kattan (2012b, 2013c, d). Undamageable materials are assumed hypothetically to maintain a zero value for the damage variable through the loading process. Details of this formulation were presented by Voyiadjis and Kattan (2012b, 2013c, d) within the framework of CDM. Thus, it can be seen that undamageable materials are desirable since they cannot be damaged at all. It is hoped that the manufacturing technology will reach a stage in the future where the realization of this type of material can be achieved. The classical definition of the effective stress (Eq. 19) is modified to show that such materials maintain a zero value of the damage variable throughout the loading process as follows:

$$\overline{\sigma} = \frac{\sigma}{\sqrt[n]{1-\varphi}} \tag{34}$$

Performing the following derivation when n tends to infinity shows that the stress and the effective stress are equal in undamageable materials:

$$\overline{\sigma} = \frac{\sigma}{\sqrt[n]{1-\varphi}} = \frac{\sigma}{(1-\varphi)^{1/n}} = \frac{\sigma}{(1-\varphi)^{1/\infty}} = \frac{\sigma}{(1-\varphi)^{0}} = \frac{\sigma}{1} = \sigma \tag{35}$$

Thus, one obtains the undamageable material in this case. In their previous publications (Voyiadjis and Kattan 2012b, 2013c, d), they presented the concept of undamageable materials using the definition of the damage variable in terms of elastic stiffness degradation. The formulation is now supported further by presenting the concept of undamageable materials using a slightly modified form of the effective stress based on the cross-sectional area reduction as shown in Eqs. 17 and 18.

Internal Damage Processes Leading to a Singularity in a Continuous Region

Providing a possible link between the subjects of damage mechanics and fracture mechanics is the aim of this section. Phenomenological study of the various internal damage mechanisms without emphasis on the actual geometry of micro-cracks, micro-voids, or other micro-defects is the main subject in continuum damage mechanics. The study of the various forms of crack propagation and coalescence in great details but without much discussion on initiation of these defects is the main subject in fracture mechanics. Usually some form of energy threshold is used to indicate these defects initiation but without showing precisely the way that they

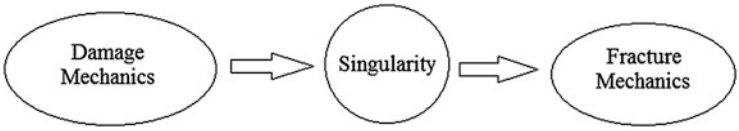


Fig. 9 The singularity could provide a crucial link between damage mechanics and fracture mechanics (After Voyiadjis and Kattan 2013b)

form. Recently, numerical simulations like the finite element method are used by some researchers to show how cracks initiate in solids (Yu et al. 2012). Currently there is no analytical closed-form solution for the precise method in which a singularity arises in a continuous region.

In this section and within the framework of continuum damage mechanics, a proposed sequence of internal damage processes within a continuous region is postulated and described mathematically. It is shown that this sequence of internal damage processes lead to a singularity in the continuous region. The resulting singularity could be interpreted in several ways. It could represent the crack tip of a forming micro-crack, the tip of a forming micro-void, or the tip of any other forming micro-defect. This emerging singularity could also provide a crucial link between the subjects of damage mechanics and fracture mechanics (see Fig. 9).

Mathematical Formulation

In this section, the principles of continuum damage mechanics are used to show how a singularity arises in a continuous region. The cross-sectional area of the damaged material is shown by A, while the corresponding damaged area is shown by A_o. Based on continuum damage mechanics, the following classic equation is common:

$$\phi_o = \frac{A_o}{A} \tag{36}$$

The force on the cross-sectional area A is equal to σA, while the force on the undamaged area $A - A_0$ is equal to $\overline{\sigma}_0(A - A_0)$. Thus, the following equation shows the equality of the forces on both damaged material (real configuration) and undamaged material (fictitious configuration):

$$\sigma A = \overline{\sigma}_0(A - A_0) \tag{37a}$$

Equating Eqs. 36 and 37a and simplifying lead to

$$\sigma = \overline{\sigma}_o(1 - \phi_o) \tag{37b}$$

where ϕ_o is the damage variable, σ is the stress in the damaged configuration, and $\overline{\sigma}_o$ is the effective stress that is associated with ϕ_o in the fictitious effective (undamaged) configuration (see Fig. 10).

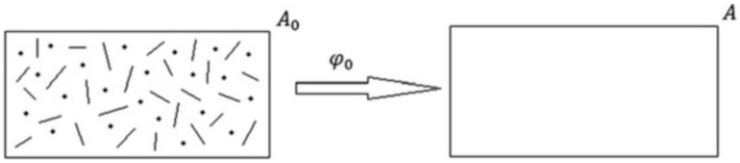


Fig. 10 The damaged state and fictitious undamaged state (After Voyiadjis and Kattan 2013b)

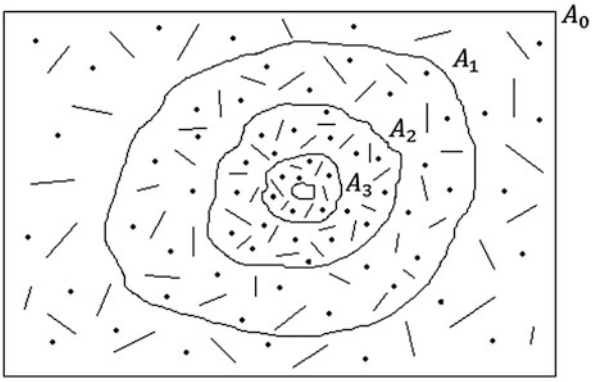


Fig. 11 The sequence of decreasing subareas leading to the singularity (After Voyiadjis and Kattan 2013b)

A subarea A_1 of the damaged area A_o where $A_1 < A_o$ (see Fig. 11) is considered and a new damage variable ϕ_1 acting on this subarea A_1 with further damage is defined as follows (in this way, A_1 is subjected to ϕ_o then followed by ϕ_1):

$$\phi_1 = \frac{A_1}{A} \tag{38}$$

The force on the cross-sectional area A is given by σA, and the forces on the undamaged areas $A - A_0$ and $A - A_1$ are given by $\overline{\sigma}_0(A - A_0)$ and $\overline{\sigma}_1(A - A_1)$, respectively. Therefore, the following equation can be written based on the force equality on both damaged and undamaged configurations:

$$\sigma A = \overline{\sigma}_0(A - A_0) + \overline{\sigma}_1(A - A_1) \tag{39a}$$

Substituting Eqs. 36 and 38 into Eq. 39a and simplifying result in

$$\sigma = \overline{\sigma}_0(1 - \phi_o) + \overline{\sigma}_1(1 - \phi_1) \tag{39b}$$

Clearly $\phi_o > \phi_1$ where ϕ_1 is a new damage variable defined on the subarea A_1 and $\overline{\sigma}_1$ is the effective stress associated with ϕ_1.

Consequently, a subarea A_2 of the damaged area A_1 where $A_2 < A_1$ (see Fig. 11) is considered and a new damage variable ϕ_2 acting on this subarea A_2 with further

damage is defined (in this way, A_2 is subjected to ϕ_o then followed by ϕ_1, then finally followed by ϕ_2). Therefore, new damage variable reads

$$\phi_2 = \frac{A_2}{A} \tag{40}$$

The force on the cross-sectional area A is obtained as σA, and the forces on the undamaged areas $A - A_0$, $A - A_1$, and $A - A_2$ are obtained by $\overline{\sigma}_0(A - A_0)$, $\overline{\sigma}_1(A - A_1)$, and $\overline{\sigma}_2(A - A_2)$, respectively. Therefore, the following equation can be written based on the force equality on both damaged and undamaged configurations:

$$\sigma A = \overline{\sigma}_0(A - A_0) + \overline{\sigma}_1(A - A_1) + \overline{\sigma}_2(A - A_2) \tag{41a}$$

Substituting Eqs. 36, 38, and 40 into Eq. 41a and simplifying lead to

$$\sigma = \overline{\sigma}_0(1 - \phi_o) + \overline{\sigma}_1(1 - \phi_1) + \overline{\sigma}_2(1 - \phi_2) \tag{41b}$$

Clearly $\phi_1 > \phi_2$ where ϕ_2 is a new damage variable defined on the subarea A_2 and $\overline{\sigma}_2$ is the effective stress associated with ϕ_2.

This process can be continued by defining n subareas $A_1 > A_2 > ..... > A_n$ along with n damage variables $\phi_1 > \phi_2 > ..... > \phi_n$.

Thus, a strictly decreasing monotonic sequence of damage variables is obtained as follows: ϕ_o, ϕ_1, ϕ_2,....., ϕ_n.

The above sequence converges to a limit based on the mathematics of sequences and series since the value of each single damage variable is less than 1. Thus, the following convergent series can be written as

$$\phi_o + \phi_1 + \phi_2 + \phi_3 + ... + \phi_n + ... = \phi \tag{42}$$

where ϕ is the limit of the sequence and sum of the series.

Based on a direct extension of Eqs. 37b, 39b, and 41b, the following equation for the stress can be written as

$$\sigma = \overline{\sigma}_0(1 - \phi_o) + \overline{\sigma}_1(1 - \phi_1) + ... + \overline{\sigma}_n(1 - \phi_n) + ... \tag{43}$$

The question that arises is what happens when the above sequence tend to infinity. Based on Eq. 48, at infinity the stress becomes infinite, while the damage variable ϕ_n becomes zero as $n \to \infty$ and the subarea A_n collapses to a point, i.e., the sought after singularity.

Example: Special Case A special case is discussed to illustrate the above concept. The following equations are valid assuming that the successive damage variables in Eq. 42 are related by a constant ratio α:

$$\phi_1 = \alpha\phi_o \tag{44}$$

$$\phi_2 = \alpha\phi_1 \tag{45}$$

where $0 < \alpha < 1$. Thus, Eq. 42 becomes

$$\phi_o\left(1 + \alpha + \alpha^2 + \alpha^3 + ... + \alpha^{n-1}\right) = \phi \tag{46}$$

The expression inside the parenthesis in Eq. 46 is a geometric series that sums to the value $\frac{1-\alpha^n}{1-\alpha}$. Thus, Eq. 46 can be written as

$$\phi_o \frac{1 - \alpha^n}{1 - \alpha} = \phi \tag{47}$$

Considering an infinite geometric series, i.e., when $n \to \infty$, then $\alpha^n \to 0$, since $0 < \alpha < 1$. Thus, Eq. 47 simplifies to the following form:

$$\frac{\phi_o}{1 - \alpha} = \phi \tag{48}$$

And since the value of ϕ is less than 1, then one obtains the constraint $\alpha < 1 - \phi_o$ on the value of the constant parameter α from Eq. 48. Therefore, the value of the damage variable ϕ at the singularity in terms of the initial value of the damage variable ϕ_o can be obtained from Eq. 48.

Conclusion

In this chapter, a new type of healing/strengthening process in materials called super healing is proposed along with a new hypothetical type of material that is called the super material. The mechanics of the super healing process has been outlined using both scalar variables and anisotropic tensors. In addition, the mechanics of scalar damage/healing is reviewed and elaborated on the mechanics of anisotropic damage/healing. For further clarification and illustration of these new concepts, the special case of plane stress was solved. Finally and in order to elucidate the characteristics of the sought super material, it is concluded that the super material has to be undamageable within the theory of elastic undamageable elastic materials framework.

The authors did not present the physical and metallurgical aspects of this theory in this work but only the theoretical mathematical formulation. This is because it is not clear yet to the authors how these types of advanced materials could be manufactured. It is hoped that the authors would be able to address the physical and metallurgical aspects in forthcoming work. The authors reiterate their viewpoint that this mathematical formulation lays a possible groundwork for any future development in this regard. The authors are still hopeful that some form of strengthened material may be realized in the near future. The various figures and equations presented here should form the basis of the future technology of undamageable materials that will be effectively indestructible.

The final question is what could be the interest (or practical use) of a material that, even if undamageable, would deliver an extremely high valued strain under a

totally negligible stress (as demonstrated on Fig. 8). The answer is that as the exponent n approaches infinity, the material becomes very soft and is not ideal for structural behavior related to practical applications. The issue is that at some value, this exponent will be presenting a reasonable stiffness for structural applications while maintaining a high fidelity for an undamaged material response. This is termed a super material in structural applications. It is also noted that in reality, the exponent n can never reach infinity in real applications. For practical applications, a finite value of n can be used although it may be very high.

It is shown that by considering a sequence of internal damage processes such that the value of the damage variable in each process is less than that in the preceding process leads to the sequence convergence to a finite value of damage along with the fact that the sequence is a strictly decreasing monotonic sequence. It is also shown considering such an infinite sequence leads to the infinite sequence of subareas which converges to a point at infinity. This emerging point can be considered as the sought singularity. This singularity may be visualized as the crack tip of an emerging micro-crack, the tip of an emerging micro-void, or the tip of any other emerging micro-defect. It is postulated that a singularity-like point will be reached at a finite but large enough value of the parameter n since the point at infinity may never be reached in practical problems. Consequently, micro-cracks, micro-voids, and other micro-defects initiate in continuous regions out of nowhere. A possible sequence of internal damage mechanisms that could produce a singularity is illustrated in the last section of this chapter which may be considered as a crucial link between the subjects of damage mechanics and fracture mechanics.

References

J.A. Adam, A simplified model of wound healing (with particular reference to the critical size defect). Math. Comput. Model. **30**, 23–32 (1999)

E.M. Arruda, M.C. Boyce, A three-dimensional constitutive model for the large stretch behavior of rubber elastic materials. J. Mech. Phys. Solids **41**, 389–412 (1993)

E.J. Barbero, F. Greco, P. Lonetti, Continuum damage-healing mechanics with application to self-healing composites. Int. J. Damage Mech. **14**, 51–81 (2005)

A.F. Bower, *Applied Mechanics of Solids* (CRC press, Boca Raton, 2011)

A. Cauvin, R.B. Testa, Damage mechanics : basic variables in continuum theories. Int. J. Solids Struct. **36**, 747–761 (1999)

J.L. Chaboche, On some modifications of kinematic hardening to improve the description of ratchetting effects. Int. J. Plast. **7**, 661–678 (1991)

J.L. Chaboche, Thermodynamic formulation of constitutive equations and application to the viscoplasticity and viscoelasticity of metals and polymers. Int. J. Solids Struct. **34**, 2239–2254 (1997)

C.L. Chow, M. Jie, Anisotropic damage-coupled sheet metal forming limit analysis. Int. J. Damage Mech. **18**, 371–392 (2009)

C. Chow, J. Wang, An anisotropic theory of elasticity for continuum damage mechanics. Int. J. Fract. **33**, 3–16 (1987)

W.L. George, J.A. Warren, A parallel 3D dendritic growth simulator using the phase-field method. J. Comput. Phys. **177**, 264–283 (2002)

V. Ginzburg, On the theory of superconductivity. Il Nuovo Cimento (1955–1965) **2**, 1234–1250 (1955)
V. Ginzburg, L.D. Landau, On the theory of superconductivity. Zh. Eksp. Teor. Fiz. **20**(1950), 1064–1082 (1965). Translation in Collected papers of L.D.Landau. Pergamon, Oxford
L. Gránásy, T. Börzsönyi, T. Pusztai, Nucleation and bulk crystallization in binary phase field theory. Phys. Rev. Lett. **88**, 206105 (2002)
N.R. Hansen, H.L. Schreyer, A thermodynamically consistent framework for theories of elastoplasticity coupled with damage. Int. J. Solids Struct. **31**, 359–389 (1994)
M. John, G. Li, Self-healing of sandwich structures with grid stiffened shape memory polymer syntactic foam core. Smart Mater. Struct. **19**, 1–12 (2010)
L.M. Kachanov, On the creep fracture time. Izv Akad. Nauk USSR Otd. Tekh. **8**, 26–31 (1958)
P.I. Kattan, G.Z. Voyiadjis, A plasticity-damage theory for large deformation of solids—II. Applications to finite simple shear. Int. J. Eng. Sci. **31**, 183–199 (1993)
P.I. Kattan, G.Z. Voyiadjis, Decomposition of damage tensor in continuum damage mechanics. J. Eng. Mech. **127**, 940–944 (2001)
R. Kobayashi, Simulations of three dimensional dendrites, in *Pattern Formation in Complex Dissipative Systems*, ed. by S. Kai (World Scientific, Singapore, 1992), pp. 121–128
L.D. Landau, D. Ter Haar, *Collected Papers of LD Landau* (Pergamon Press, Oxford, 1965)
H. Lee, K. Peng, J. Wang, An anisotropic damage criterion for deformation instability and its application to forming limit analysis of metal plates. Eng. Fract. Mech. **21**, 1031–1054 (1985)
J. Lemaitre, How to use damage mechanics. Nucl. Eng. Des. **80**, 233–245 (1984)
J. Lemaitre, Coupled elasto-plasticity and damage constitutive equations. Comput. Methods Appl. Mech. Eng. **51**, 31–49 (1985)
J. Lemaitre, J.L. Chaboche, *Mechanics of Solid Materials* (Cambridge University Press, Cambridge, 1990)
G. Li, M. John, A self-healing smart syntactic foam under multiple impacts. Compos. Sci. Technol. **68**, 3337–3343 (2008)
G. Li, V.D. Muthyala, Impact characterization of sandwich structures with an integrated orthogrid stiffened syntactic foam core. Compos. Sci. Technol. **68**, 2078 (2008)
G. Li, D. Nettles, Thermomechanical characterization of a shape memory polymer based self-repairing syntactic foam. Polymer **51**, 755–762 (2010)
G. Li, N. Uppu, Shape memory polymer based self-healing syntactic foam: 3-D confined thermomechanical characterization. Compos. Sci. Technol. **70**, 1419–1427 (2010)
Y.L. Liu, Y.W. Chen, Thermally reversible cross-linked polyamides with high toughness and self-repairing ability from maleimide- and furan-functionalized aromatic polyamides. Macromol. Chem. Phys. **208**, 224–232 (2007)
I. Loginova, G. Amberg, J. Ågren, Phase-field simulations of non-isothermal binary alloy solidification. Acta Mater. **49**, 573–581 (2001)
V.A. Lubarda, D. Krajcinovic, Damage tensors and the crack density distribution. Int. J. Solids Struct. **30**, 2859–2877 (1993)
S. Miao, M.L. Wang, H.L. Schreyer, Constitutive models for healing of materials with application to compaction of crushed rock salt. J. Eng. Mech. **121**, 1122–1129 (1995)
S. Murakami, Notion of continuum damage mechanics and its application to anisotropic creep damage theory. ASME Trans. J. Eng. Mater. Technol. **105**, 99–105 (1983)
B. Murray, A. Wheeler, M. Glicksman, Simulations of experimentally observed dendritic growth behavior using a phase-field model. J. Cryst. Growth **154**, 386–400 (1995)
M. Naderi, A. Kahirdeh, M. Khonsari, Dissipated thermal energy and damage evolution of Glass/Epoxy using infrared thermography and acoustic emission. Compos. Part B **43**, 1613–1620 (2012)
S. Nemat-Nasser, Decomposition of strain measures and their rates in finite deformation elastoplasticity. Int. J. Solids Struct. **15**, 155–166 (1979)
S. Nemat-Nasser, *On Finite Plastic Flow of Crystalline Solids and Geomaterials* (DTIC Document, Fort Belvoir, 1983)

J. Nji, G. Li, A self-healing 3D woven fabric reinforced shape memory polymer composite for impact mitigation. Smart Mater. Struct. **19**, 1–9 (2010a)
J. Nji, G. Li, A biomimic shape memory polymer based self-healing particulate composite. Polymer **51**, 6021–6029 (2010b)
J.W.C. Pang, I.P. Bond, A hollow fibre reinforced polymer composite encompassing self-healing and enhanced damage visibility. Compos. Sci. Technol. **65**, 1791–1799 (2005)
T. Park, G. Voyiadjis, Kinematic description of damage. J. Appl. Mech. **65**, 93–98 (1998)
R. Pavan, B. Oliveira, S. Maghous, G. Creus, A model for anisotropic viscoelastic damage in composites. Compos. Struct. **92**, 1223–1228 (2010)
Y.N. Rabotnov, Paper 68: on the equation of state of creep, in *Proceedings of the Institution of Mechanical Engineers, Conference Proceedings*, SAGE Publications, 1963, pp. 2–117, 112–122
Y.N. Rabotnov, Creep rupture, in *Proceedings of the XII International Congress on Applied Mechanics*, 1968, pp. 342–349
F. Sidoroff, Description of anisotropic damage application to elasticity, in *Proceedings of the IUTAM Colloquium on Physical Nonlinearities in Structural Analysis*, Berlin, 1981, pp. 237–244
A.H.R.W. Simpson, T.N. Gardner, M. Evans, J. Kenwright, Stiffness, strength and healing assessment in different bone fractures – a simple mathematical model. Injury **31**, 777–781 (2000)
I. Singer-Loginova, H. Singer, The phase field technique for modeling multiphase materials. Rep. Prog. Phys. **71**, 106501 (2008)
K.S. Toohey, N.R. Sottos, J.A. Lewis, J.S. Moore, S.R. White, Self-healing materials with microvascular networks. Nat. Mater. **6**, 581–585 (2007)
J.D. van der Waals, The thermodynamic theory of capillarity under the hypothesis of a continuous variation of density. J. Stat. Phys. **20**, 200–244 (1979)
R.J. Varley, S. van der Zwaag, Towards an understanding of thermally activated self-healing of an ionomer system during ballistic penetration. Acta Mater. **56**, 5737–5750 (2008)
G.Z. Voyiadjis, Degradation of elastic modulus in elastoplastic coupling with finite strains. Int. J. Plast. **4**, 335–353 (1988)
G.Z. Voyiadjis, P.I. Kattan, A coupled theory of damage mechanics and finite strain elasto-plasticity—II. Damage and finite strain plasticity. Int. J. Eng. Sci. **28**, 505–524 (1990)
G.Z. Voyiadjis, P.I. Kattan, A plasticity-damage theory for large deformation of solids—I. Theoretical formulation. Int. J. Eng. Sci. **30**, 1089–1108 (1992)
G.Z. Voyiadjis, P.I. Kattan, *Damage Mechanics* (CRC, Boca Raton, 2005)
G.Z. Voyiadjis, P.I. Kattan, *Advances in Damage Mechanics: Metals and Metal Matrix Composites With an Introduction to Fabric Tensors* (2nd edition), 742 p., (Elsevier, Oxford, ISBN: 0-08-044688-4, 2006)
G.Z. Voyiadjis, P.I. Kattan, A comparative study of damage variables in continuum damage mechanics. Int. J. Damage Mech. **18**, 315–340 (2009)
G.Z. Voyiadjis, P.I. Kattan, Mechanics of damage processes in series and in parallel: a conceptual framework. Acta Mech. **223**, 1863–1878 (2012a)
G.Z. Voyiadjis, P.I. Kattan, A new class of damage variables in continuum damage mechanics. J. Eng. Mater. Technol. **134**, 021016 (2012b)
G.Z. Voyiadjis, P.I. Kattan, Healing and Super Healing in Continuum Damage Mechanics. Int. J. Damage Mech. 23(2), 245–260 (2014)
G.Z. Voyiadjis, P.I. Kattan, How a Singularity Forms in Continuum Damage Mechanics. Mech. Res. Commun. **55**, 86-88 (2014)
G.Z. Voyiadjis, P.I. Kattan, Introduction to the mechanics and design of undamageable materials. Int. J. Damage Mech. **22**, 323–335 (2013c)
G.Z. Voyiadjis, P.I. Kattan, On the theory of elastic undamageable materials. J. Eng. Mater. Technol. **135**, 021002 (2013d)

G.Z. Voyiadjis, N. Mozaffari, Nonlocal damage model using the phase field method: theory and applications. Int. J. Solids Struct. **50**, 3136–3151 (2013)
G.Z. Voyiadjis, T. Park, Local and interfacial damage analysis of metal matrix composites. Int. J. Eng. Sci. **33**, 1595–1621 (1995)
G.Z. Voyiadjis, T. Park, Anisotropic damage for the characterization of the onset of macro-crack initiation in metals. Int. J. Damage Mech. **5**, 68–92 (1996)
G. Voyiadjis, T. Park, Anisotropic damage effect tensors for the symmetrization of the effective stress tensor. J. Appl. Mech. **64**, 106–110 (1997)
G.Z. Voyiadjis, Z.N. Taqieddin, P.I. Kattan, Theoretical formulation of a coupled elastic—plastic anisotropic damage model for concrete using the strain energy equivalence concept. Int. J. Damage Mech. **18**, 603–638 (2009)
G.Z. Voyiadjis, A. Shojaei, G. Li, P. Kattan, Continuum damage-healing mechanics with introduction to new healing variables. Int. J. Damage Mech. **21**(3), 391–414 (2012)
G.Z. Voyiadjis, A. Shojaei, G. Li, A generalized coupled viscoplastic–viscodamage–viscohealing theory for glassy polymers. Int. J. Plast. **28**, 21–45 (2012)
S.-L. Wang, R.F. Sekerka, Algorithms for phase field computation of the dendritic operating state at large supercoolings. J. Comput. Phys. **127**, 110–117 (1996)
J.A. Warren, W.J. Boettinger, Prediction of dendritic growth and microsegregation patterns in a binary alloy using the phase-field method. Acta Metall. Mater. **43**, 689–703 (1995)
A. Wheeler, B. Murray, R. Schaefer, Computation of dendrites using a phase field model. Phys. D Nonlinear Phenom. **66**, 243–262 (1993)
S.R. White, N.R. Sottos, P.H. Geubelle, J.S. Moore, M.R. Kessler, S.R. Sriram, E.N. Brown, S. Viswanathan, Autonomic healing of polymer composites. Nature **409**, 794–797 (2001)
H.-L. Yu, C. Lu, K. Tieu, G.-Y. Deng, A numerical model for simulation of crack initiation around inclusion under tensile load. J. Comput. Theor. Nanosci. **9**, 1745–1749 (2012)
K. Yuan, J. Ju, New strain energy–based coupled elastoplastic damage-healing formulations accounting for effect of matric suction during earth-moving processes. J. Eng. Mech. **139**, 188–199 (2012)
F. Zaïri, M. Naït-Abdelaziz, J.-M. Gloaguen, J.-M. Lefebvre, A physically-based constitutive model for anisotropic damage in rubber-toughened glassy polymers during finite deformation. Int. J. Plast. **27**, 25–51 (2011)
M. Zako, N. Takano, Intelligent material systems using epoxy particles to repair microcracks and delamination damage in GFRP. J. Intell. Mater. Syst. Struct. **10**, 836–841 (1999)

Thermodynamics of Continuum Damage Healing Mechanics

45

George Z. Voyiadjis and Amir Shojaei

Contents

Abstract

In this chapter, the governing thermodynamic laws on the damage and healing processes are revisited. The solid mechanics thermodynamic framework provides a physically consistent description for the deformation mechanisms in solids, and it has been widely examined for the plasticity and damage processes in the literature (S. Yazdani, H.L. Schreyer, Combined plasticity and damage mechanics model for plain concrete. J. Eng. Mech. **116**(7), 1435–1450 (1990); J.L. Chaboche, On some modifications of kinematic hardening to improve the description of ratchetting effects. Int. J. Plast. **7**(7), 661–678 (1991); J.L. Chaboche, Cyclic viscoplastic constitutive equations, part I: a thermodynamically consistent formulation. J. Appl. Mech. **60**(4), 813–821 (1993);

G.Z. Voyiadjis (✉)
Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA
e-mail: voyiadjis@eng.lsu.edu

A. Shojaei
Department of Mechanical and Industrial Engineering, Louisiana State University, Baton Rouge, LA, USA
e-mail: ashoja1@tigers.lsu.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_44

N.R. Hansen, H.L. Schreyer, A thermodynamically consistent framework for theories of elastoplasticity coupled with damage. Int. J. Solids Struct. **31**(3), 359–389 (1994); G. Voyiadjis, I. Basuroychowdhury, A plasticity model for multiaxial cyclic loading and ratchetting. Acta Mech. **126**(1), 19–35 (1998); J.L. Chaboche, A review of some plasticity and viscoplasticity constitutive theories. Int. J. Plast. **24**(10), 1642–1693 (2008)). Introduction of the healing process into the thermodynamic framework was formerly proposed by Voyiadjis et al. (A thermodynamic consistent damage and healing model for self healing materials. Int. J. Plast. **27**(7), 1025–1044 (2011)) where a physically consistent description for the healing process is provided.

Basically, the mathematical foundation of the thermodynamic-based solid mechanics modeling was developed formerly for capturing plasticity and damage in metallic structures and it is not directly applicable to polymeric materials. Polymers usually show strain softening after their initial yield and they show strain hardening at higher strain levels. To overcome the mathematical deficiency associated with the classical thermodynamic framework, Voyiadjis, Shojaei, and Li (A generalized coupled viscoplastic- viscodamage- viscohealing theory for glassy polymers. Int. J. Plast. **28**(1), 21–45 (2012a)) established a generalized formulation within the thermodynamic framework in which the mathematical competency for simulating the most nonlinear viscoplastic, viscodamage, and viscohealing effects in polymers was enhanced. They have successfully shown that the proposed framework is able to accurately capture the viscoplastic and viscodamage responses of polymers and the model has enough flexibility to capture the healing response in polymeric-based self-healing materials.

Introduction

Continuum damage mechanics (CDM) provides a general approach to incorporate different types of microscale damages including microcracks, micro-cavities, debonding between fiber and matrix, and breakage of bonds in polymeric materials, corresponding to the respective loading conditions and the inherent material defects. While the classical fracture mechanics provides a practical approach for dealing with macroscale cracks and voids (Ashby et al. 1979; Rice et al. 1980; Tvergaard and Hutchinson 1992), it still suffers from its limited applicability and computational difficulties for complex problems. By empowering the computational facilities in the past decade, the application of CDM-based failure studies has been increased significantly and the developed CDM models are now built into the finite element analysis (FEA) through user-defined coding. This coupling between CDM and mesh removal techniques provides a very efficient way for analyzing complex structures under complex loading. Basically, the continuum damage in materials is represented by a continuous degradation of the strength of material systems until the failure point, and this task is accomplished by developing the damage variables within the CDM framework. This concept of healing is then

introduced where the microscale damages can be removed in the self-healing systems. This task is accomplished by introducing new healing variable by Voyiadjis et al. (2011, 2012a, b, c).

The formulation for the coupled continuum damage and plasticity processes have been extensively investigated in the literature (Yazdani and Schreyer 1990; Hansen and Schreyer 1994; Chaboche 1997, 2008), although quite a few works up to present have addressed the continuum damage–healing process, and the continuum damage–healing mechanics (CDHM) is almost a newly developed topic (Miao et al. 1995; Barbero et al. 2005; Voyiadjis et al. 2011, 2012a, b, c). The CDHM framework is formulated in ▶ Chap. 46, "Continuum Damage-Healing Mechanics," and the present chapter concerns the thermodynamic consistent formulation of such a CDHM process.

This chapter is designed as follows. In section "Scalar Damage and Healing Process," the thermodynamics of the damage and healing is studied and the thermodynamic consistent constitutive relations are developed in the case of scalar damage and healing variables. In section "Anisotropic Damage and Healing Mechanics Constitutive Relations," the anisotropic damage–healing process is investigated where the tensorial form of the governing equations are formulated. This book chapter aims the material designers to describe the damage–healing process in self-healing materials within a continuum mechanics framework. The healing and the damage processes can be concurrently or individually active. In the first case, in which the damage and healing are active simultaneously, the healing system is referred to *coupled* and the other case is termed *decoupled* damage–healing system hereinafter.

Thermodynamic Consistent Damage and Healing Evolution Equations

In this section, the thermodynamic framework for the CDHM concept is reviewed. The kinematic and isotropic hardening effects are incorporated for all the three processes, i.e., plasticity, damage, and healing. The physical description for the kinematic and isotopic hardening effects for the damage and the healing is taken into account and the small strain assumption is incorporated. The damage and healing yield surfaces are then presented, and finally, the constitutive equations for a coupled elastoplastic–damage–healing process are derived.

Scalar Damage and Healing Process

The energy considerations concept is adopted by Voyiadjis et al. (2012b) to derive the evolution equations for the damage and healing variables for an *uncoupled damage–healing* process, where damage and healing processes are independent from each other. It is assumed that the material obeys the damage criterion proposed by Lee et al. (1985), which is a scalar function, i.e., g^d, and a generalized

thermodynamic force, y^d, is assumed to be the driving force for the damaging process. The healing criterion is constituted based on the same terminology in which a scalar function, i.e., g^h, is defined based on the healing thermodynamic conjugate force, i.e., y^h. These two criteria are defined as follows (Voyiadjis et al. 2012b):

$$g^d\left(y^d, s^d\right) \equiv \frac{1}{2}y^{d2} - L^d\left(s^d\right) = 0 \tag{1}$$

$$g^h\left(y^h, s^h\right) \equiv \frac{1}{2}y^{h2} - L^h\left(s^h\right) = 0 \tag{2}$$

where L^d is a damage strengthening parameter which is a function of the "overall" damage, i.e., s^d, and L^h represent the strengthening effects for the healing and s^h is the "overall" healing. Voyiadjis and co-workers introduced two sets of scalar damage–healing variables (Voyiadjis et al. 2012a, b, c) that are mentioned in section "Computational Aspect and Simulation Results" of ▶ Chap. 44, "Healing, Super Healing, and Other Issues in Continuum Damage Mechanics." The first set is a scalar damage variable, i.e., ϕ, with the corresponding scalar healing variable, i.e., h, which respectively take into account the removal and recovery of the damaged microsurfaces. The second set of the scalar damage variable, i.e., l, and the healing variable, i.e., h', is proposed to facilitate the calibration process based on the changes of the elastic moduli instead of measurement of the damaged microsurfaces which is impractical in some cases. Based on these two sets of scalar damage–healing variables, two sets of evolution equations are derived herein as follows.

Evolution Laws Based on the Damage Variable ϕ and the Healing Variable h

The elastic stiffness results in the following expression for the elastic strain energy in the damaged–healed configuration (Voyiadjis et al. 2012b):

$$U = \frac{1}{2}\sigma\epsilon = \frac{1}{2}E\epsilon^2 = \frac{1}{2}\overline{E}((1-\phi)+\phi(1-h))^2\epsilon^2 \tag{3}$$

Differentiating Eq. 3 results in (Voyiadjis et al. 2012b)

$$\begin{aligned} dU = {} & \overline{E}((1-\phi)+\phi(1-h))^2\epsilon d\epsilon - h\overline{E}((1-\phi)+\phi(1-h))\epsilon^2 d\phi \\ & - \phi\overline{E}((1-\phi)+\phi(1-h))\epsilon^2 dh \end{aligned} \tag{4}$$

The generalized conjugated thermodynamic forces corresponding to the damage and healing processes are then defined as follows (Voyiadjis et al. 2012b):

$$y^d = \frac{\partial U}{\partial \phi} = -h\overline{E}((1-\phi)+\phi(1-h))\epsilon^2 \tag{5}$$

$$y^h = \frac{\partial U}{\partial h} = -\phi\overline{E}((1-\phi)+\phi(1-h))\epsilon^2 \tag{6}$$

In order to derive the normality rule regarding the damage and healing processes, one may introduce the power of dissipation as follows (Voyiadjis et al. 2012b):

$$\Pi = -y^d d\phi + y^h dh - L^d ds^d + L^h ds^h > 0 \tag{7}$$

The energy dissipation is also subjected to the constraints from the damage and healing criteria (Eqs. 1 and 2). Then function Π is modified by incorporating the Lagrange multipliers which results in the following functional, to be minimized (Voyiadjis et al. 2012b):

$$\Psi = \Pi - d\lambda^d.\, g^d - d\lambda^h.\, g^h \tag{8}$$

Substituting Eqs. 1 and 2 and Eq. 7 into Eq. 8 yields (Voyiadjis et al. 2012b)

$$\begin{aligned}\Psi = -y^d d\phi + y^h dh - L^d ds^d - L^h ds^h - d\lambda^d.\left(\frac{1}{2}y^{d^2} - L^d\left(s^d\right)\right) \\ + d\lambda^h.\left(\frac{1}{2}y^{h^2} - L^h\left(s^h\right)\right)\end{aligned} \tag{9}$$

Applying the stationary conditions of $\partial\Psi/\partial y^d = 0$ and $\partial\Psi/\partial y^h = 0$ and $\partial\Psi/\partial L^d = 0$ and $\partial\Psi/\partial L^h = 0$, Eq. 9 results in the following relations (Voyiadjis et al. 2012b):

$$d\phi = -d\lambda^d.\, y^d \tag{10}$$

$$dh = d\lambda^h.\, y^h \tag{11}$$

$$ds^d = d\lambda^d \tag{12}$$

$$ds^h = -d\lambda^h \tag{13}$$

Equations 10 and 11 are evolution equations for damage and healing variables and Eqs. 12 and 13 are showing the relation between the damage and healing variables and their corresponding multipliers. To obtain damage and healing variables, the consistency conditions, i.e., $dg^d = 0$ and $dg^h = 0$, are applied as follows (Voyiadjis et al. 2012b):

$$\frac{\partial g^d}{\partial y^d}dy^d + \frac{\partial g^d}{\partial L^d}dL^d = 0 \tag{14}$$

$$\frac{\partial g^h}{\partial y^h} dy^h + \frac{\partial g^h}{\partial L^h} dL^h = 0 \tag{15}$$

Making use of Eqs. 1, 2, 12, and 13 and using $dL^d = ds^d(\partial L^d/\partial s^d)$ and $dL^h = ds^h(\partial L^h/\partial s^h)$ into Eqs. 14 and 15 result in (Voyiadjis et al. 2012b)

$$ds^d = d\lambda^d = \frac{y^d dy^d}{\partial L^d/\partial s^d} \tag{16}$$

$$ds^h = -d\lambda^h = \frac{y^h dy^h}{\partial L^h/\partial s^h} \tag{17}$$

To investigate the strain-damage and strain-healing relations, one may differentiate Eqs. 5 and 6 as follows (Voyiadjis et al. 2012b):

$$dy^d = \overline{E}(2h\phi - 1)\epsilon^2 dh - 2h\overline{E}(1 - \phi h)\epsilon d\epsilon + \overline{E}h^2\epsilon^2 d\phi \tag{18}$$

$$dy^h = \overline{E}(2h\phi - 1)\epsilon^2 d\phi - 2\phi\overline{E}(1 - \phi h)\epsilon d\epsilon + \overline{E}\phi^2\epsilon^2 dh \tag{19}$$

Substituting Eqs. 16 and 17 into Eqs. 10 and 11 yields the evolution equations in terms of conjugate forces and strengthening functions as follows (Voyiadjis et al. 2012b):

$$d\phi = -\frac{\left(y^d\right)^2 dy^d}{\partial L^d/\partial s^d} \tag{20}$$

$$dh = -\frac{\left(y^h\right)^2 dy^h}{\partial L^h/\partial s^h} \tag{21}$$

Making use of Eqs. 5, 6, 18, and 19 into Eqs. 20 and 21 results in the strain-damage and strain-healing relations as follows (Voyiadjis et al. 2012b):

$$\left(\partial L^d/\partial s^d\right)d\phi = h^2\overline{E}^3\epsilon^5(1 - h\phi)^2(-(2h\phi - 1)\epsilon dh + 2h(1 - \phi h)d\epsilon - h^2\epsilon d\phi) \tag{22}$$

$$\left(\partial L^h/\partial s^h\right)dh = \phi^2\overline{E}^3\epsilon^5(1 - h\phi)^2(-(2h\phi - 1)\epsilon d\phi + 2\phi(1 - \phi h)d\epsilon - \phi^2\epsilon dh) \tag{23}$$

In the case of damage without healing, by substituting $h = 1$ and $dh = 0$ into Eq. 22, it is reduced to (Voyiadjis et al. 2012b)

$$\left(\partial L^d/\partial s^d\right)d\phi = \overline{E}^3\epsilon^5(1 - \phi)^2(2(1 - \phi)d\epsilon - \epsilon d\phi) \tag{24}$$

Equation 24 is one of the CDM fundamental relationships which links the level of the damage and the strain (Voyiadjis and Kattan 2006, 2009). By simple change of variables $x = h(1 - h\phi)\epsilon^2$, Eq. 22 reduces to (Voyiadjis et al. 2012b)

$$\left(\partial L^d/\partial s^d\right)d\phi = \overline{\overline{E}}^3 x^2 dx \tag{25}$$

One may consider a linear function with the form $L^d = \overline{\overline{c}} s^d + \overline{\overline{d}}$, where $\overline{\overline{c}}$ and $\overline{\overline{d}}$ are constant, and then Eq. 25 can be solved as follows (Voyiadjis et al. 2012b):

$$\frac{\phi - \phi_0}{h^3(1-h\phi)^3} = \overline{\overline{E}}^3 \frac{\epsilon^6}{3\overline{\overline{c}}} \tag{26}$$

where ϕ_0 is the initial damage value. In a similar way, one may solve Eq. 23 by the change of variables $y = \phi(1-h\phi)\epsilon^2$ and the linear function assumption of the form $L^h = c'\, s^h + d'$, where c' and d' are constants. The resulting relation is as follows (Voyiadjis et al. 2012b):

$$\frac{h - h_0}{\phi^3(1-h\phi)^3} = \frac{\overline{\overline{E}}^3 \epsilon^6}{3c'} \tag{27}$$

where h_0 is the initial value of healing variable. Equations 26 and 27 are strain-damage and strain-healing relations for the damage variable ϕ and the healing variable h, respectively (Voyiadjis et al. 2012b).

Evolution Laws Based on the Damage Variable *l* and the Healing Variable *h*′

Substituting the elastic stiffness from Eq. 3.10 into Eq. 3.7 results in the following expression for the elastic strain energy in the damaged–healed configuration (Voyiadjis et al. 2012b):

$$U = \frac{1}{2}\frac{(1+h')\overline{\overline{E}}}{(1+l)}\epsilon^2 \tag{28}$$

Differentiating Eq. 23 results in (Voyiadjis et al. 2012b)

$$dU = \frac{\overline{\overline{E}}\epsilon^2}{2(1+l)}dh' - \frac{\overline{\overline{E}}\epsilon^2(1+h')}{2(1+l)^2}dl + \frac{(1+h')}{(1+l)}\overline{\overline{E}}\epsilon d\epsilon \tag{29}$$

The generalized conjugated thermodynamic forces corresponding to damage $y^d = \partial U/\partial l$ and healing $y^h = \partial U/\partial h'$ processes are then defined as follows (Voyiadjis et al. 2012b):

$$y^d = -\frac{\overline{\overline{E}}\epsilon^2(1+h')}{2(1+l)^2} \tag{30}$$

$$y^h = \frac{\overline{\overline{E}}\epsilon^2}{2(1+l)} \tag{31}$$

To investigate the strain-damage and strain-healing relations, one may differentiate from Eqs. 30 and 31 as follows (Voyiadjis et al. 2012b):

$$dy^d = -\frac{\overline{E}\epsilon(1+h')}{(1+l)^2}d\epsilon - \frac{\overline{E}\epsilon^2}{2(1+l)^2}dh' + \frac{\overline{E}\epsilon^2(1+h')}{(1+l)^3}dl \tag{32}$$

$$dy^h = \frac{\overline{E}\epsilon}{(1+l)}d\epsilon - \frac{\overline{E}\epsilon^2}{2(1+l)^2}dl \tag{33}$$

Equations 20 and 21 are modified for the damage variable l and healing variable h' as follows (Voyiadjis et al. 2012b):

$$dl = -\frac{\left(y^d\right)^2 dy^d}{\partial L^d/\partial s^d} \tag{34}$$

$$dh' = -\frac{\left(y^h\right)^2 dy^h}{\partial L^h/\partial s^h} \tag{35}$$

Substituting Eqs. 30, 31, 32, and 33 into Eqs. 34 and 35 results in (Voyiadjis et al. 2012b)

$$\left(\partial L^d/\partial s^d\right)dl = \left(\frac{\overline{E}^3\epsilon^5(1+h')^2}{4(1+l)^6}\right)\left((1+h')\,d\epsilon + \epsilon\,dh' - \frac{(1+h')}{(1+l)}\epsilon dl\right) \tag{36}$$

$$\left(\partial L^h/\partial s^h\right)dh' = \left(\frac{\overline{E}^3\epsilon^5}{4(1+l)^3}\right)\left(\frac{\epsilon}{2(1+l)}dl - d\epsilon\right) \tag{37}$$

Equations 36 and 37 represent the strain-damage and healing–damage relationships for the damage variable l and the healing variable h'.

Anisotropic Damage and Healing Mechanics Constitutive Relations

Thermodynamic restrictions including energy consideration are used to derive the material behavior and the evolution equations for irreversible processes (such as damage, plasticity, and healing). In the following notation, all tensorial parameters are printed in bold. Let u denote the *specific internal energy*, which is a function of entropy, s; elastic strain, ϵ^e_{ij}; damage variable tensor, ζ^d_{ij}: plastic deformation variable tensor, ζ^p_{ij}; and healing variable tensor, ζ^h_{ij}, which might be *observable* or *internal* variables. The specific internal energy u is defined as follows (Voyiadjis et al. 2011, 2012a, c):

$$u = u\left(s, \epsilon^e_{ij}, \zeta^d_{ij}, \zeta^p_{ij}, \zeta^h_{ij}\right) \tag{38}$$

where the dimension of second-order tensors ζ^d_{ij}, ζ^p_{ij} and ζ^h_{ij} is the number of internal variables which is used to describe each phenomenon, i.e., damage, plasticity, and

healing processes, respectively. The superscript "e" in the strain variable, ϵ_{ij}^{e}, simply denotes that the strain is elastic. Time derivative of Eq. 38 yields

$$\dot{u} = \frac{\partial u}{\partial s}\dot{s} + \frac{\partial u}{\partial \epsilon_{ij}^{e}}\dot{\epsilon}_{ij}^{e} + \frac{\partial u}{\partial \zeta_{ij}^{d}}\dot{\zeta}_{ij}^{d} + \frac{\partial u}{\partial \zeta_{ij}^{p}}\dot{\zeta}_{ij}^{p} + \frac{\partial u}{\partial \zeta_{ij}^{h}}\dot{\zeta}_{ij}^{h} \tag{39}$$

The *second law of thermodynamics* states that the change in entropy is always positive and it can be expressed in the *Clausius–Duhem* inequality as follows (Voyiadjis et al. 2011, 2012a, c):

$$\sigma_{ij}\dot{\epsilon}_{ij}^{e} - \rho\left(\dot{u} + s\dot{T}\right) - \frac{q_i}{T}\nabla_i T \geq 0 \tag{40}$$

where ρ is the density which is assumed to be constant, σ_{ij} is the Cauchy stress, T is the absolute temperature, and q_i and $\nabla_i T$ are, respectively, heat flux and temperature gradient. The following restrictions are used in this formulation (Lubliner 1972): (1) purely mechanical theory is used (no heat source and no heat flux in the body) and (2) the infinitesimal deformation state is considered. An additive elastic and plastic state is the result of the second statement. Substituting Eq. 39 into Eq. 40 and eliminating the heat flux term gives the following expression (Voyiadjis et al. 2011, 2012a, c):

$$\sigma_{ij}\dot{\epsilon}_{ij}^{e} - \rho\left(\left(\frac{\partial \mathrm{u}}{\partial \mathrm{s}}\dot{\mathrm{s}} + \frac{\partial u}{\partial \epsilon_{ij}^{e}}\dot{\epsilon}_{ij}^{e} + \frac{\partial u}{\partial \zeta_{ij}^{d}}\dot{\zeta}_{ij}^{d} + \frac{\partial u}{\partial \zeta_{ij}^{p}}\dot{\zeta}_{ij}^{p} + \frac{\partial u}{\partial \zeta_{ij}^{h}}\dot{\zeta}_{ij}^{h}\right) + \mathrm{T}\dot{\mathrm{s}}\right) \geq 0 \tag{41}$$

Rearranging Eq. 41 yields (Voyiadjis et al. 2011, 2012a)

$$\left(\sigma_{ij} - \rho\frac{\partial u}{\partial \epsilon_{ij}^{e}}\right)\dot{\epsilon}_{ij}^{e} + \rho\left(T - \frac{\partial u}{\partial s}\right)\dot{s} - \rho\left(\frac{\partial u}{\partial \zeta_{ij}^{d}}\dot{\zeta}_{ij}^{d} + \frac{\partial u}{\partial \zeta_{ij}^{p}}\dot{\zeta}_{ij}^{p} + \frac{\partial u}{\partial \zeta_{ij}^{h}}\dot{\zeta}_{ij}^{h}\right) \geq 0 \tag{42}$$

The conjugate thermodynamic forces related to the flux of the entropy s and the elastic strain, ϵ_{ij}^{e}, are obtained respectively as follows (Voyiadjis et al. 2011, 2012a):

$$T = \frac{\partial u}{\partial s}; \sigma_{ij} = \rho\frac{\partial u}{\partial \epsilon_{ij}^{e}}; \tag{43}$$

The power of dissipation, Γ, is expressed as follows (Voyiadjis et al. 2011, 2012a):

$$\Gamma = -\rho\left(\frac{\partial u}{\partial \zeta_{ij}^{d}}\dot{\zeta}_{ij}^{d} + \frac{\partial u}{\partial \zeta_{ij}^{p}}\dot{\zeta}_{ij}^{p} + \frac{\partial u}{\partial \zeta_{ij}^{h}}\dot{\zeta}_{ij}^{h}\right) \tag{44}$$

The dissipative power Γ is used to define the following conjugate thermodynamic forces (Voyiadjis et al. 2011, 2012a):

$$y_{ij}^{d} = \frac{\partial u}{\partial \zeta_{ij}^{d}}; y_{ij}^{p} = \frac{\partial u}{\partial \zeta_{ij}^{p}}; y_{ij}^{h} = \frac{\partial u}{\partial \zeta_{ij}^{h}} \tag{45}$$

where y_{ij}^{d}, y_{ij}^{p} and y_{ij}^{h} are damage, plasticity, and healing conjugate forces, respectively. Finally, the second law of thermodynamics reduces to (Voyiadjis et al. 2011, 2012a)

$$\Gamma \geq 0 \tag{46}$$

Helmholtz free energy function Ψ is obtained through *Legendre transformation* of the internal energy as follows:

$$\Psi = u - Ts \tag{47}$$

Using the Helmholtz free energy yields the same result for the dissipative power Γ as shown in Eq. 46. The only difference between the internal energy and the Helmholtz free energy definitions is that the internal energy is a function of entropy and mechanical variables but the Helmholtz free energy is a function of temperature and mechanical variables. For the case of isentropic processes, internal energy formulation is used, and for the case of isothermal processes, the Helmholtz potential is used.

The *first law of thermodynamics* of an infinitesimal quasi-static process states that the change of energy for a system is equal to the sum of mechanical and heat input or output. The first law based on the internal variable formulation reduces to the following expression (Voyiadjis et al. 2011, 2012a, c):

$$\dot{u} = \frac{1}{\rho}\sigma_{ij}\dot{\epsilon}_{ij}^{e} \tag{48}$$

Substituting Eq. 48 into Eq. 39 while incorporating Eq. 43 results in the following expression for entropy generation due to the damage, plasticity, and healing processes (Voyiadjis et al. 2011, 2012a, c):

$$T\dot{s} = -\left(\frac{\partial u}{\partial \zeta_{ij}^{d}}\,\dot{\zeta}_{ij}^{d} + \frac{\partial u}{\partial \zeta_{ij}^{p}}\,\dot{\zeta}_{ij}^{p} + \frac{\partial u}{\partial \zeta_{ij}^{h}}\,\dot{\zeta}_{ij}^{h}\right) \tag{49}$$

The coupled elastoplastic–damage–healing is formulated by assuming an isothermal process and the internal variable tensors ζ_{ij}^{d}, ζ_{ij}^{p} and ζ_{ij}^{h} are introduced to the system in the most generalized form. Plasticity related internal variable tensors, ζ_{ij}^{p}, correspond to three internal variables as follows: (a) second-order plastic strain tensor ϵ_{ij}^{p}, (b) second-order kinematic hardening tensor α_{ij} which accounts for the flow of residual stresses during inelastic deformation and represents the shift in the center of the yield surface f^{p} in the stress space, and (c) tensor p_{ij} to express the isotropic hardening, which shows change in the size of the yield surface f^{p} in

different directions during a plastic deformation (distortion in plasticity yield surface) (Voyiadjis et al. 2011, 2012a). In the case of the damage variable tensor, ζ_{ij}^{d}, in the most general form, the corresponding internal variables can be represented in three main classes (Voyiadjis et al. 2011, 2012a, c): (a) generalized damage tensor d_{ij} which measures overall degradation of materials, (b) damage tensor d_{ij}^{K} which describes the kinematic hardening due to the damage process and indicates the shift in the center of the damage surface f^{d}, and (c) damage tensor d_{ij}^{I} to represent isotropic hardening which shows the change in the size of the damage surface f^{d} (Hansen and Schreyer 1994; Voyiadjis et al. 2011, 2012a, c). The healing variable tensor, ζ_{ij}^{h}, is proposed in the most generalized form as follows: (a) generalized healing tensor h_{ij} which accounts for the overall healing in the material, (b) healing variable tensor h_{ij}^{K} which describes the kinematic hardening during the healing process (it is the change in the center of the healing surface f^{h}), and (c) healing tensor h_{ij}^{I} to describe the isotropic hardening/softening during a healing process, which is the change in the size of the healing surface f^{h} in different directions.

In the case of polycrystalline metallic materials, the physical description of the kinematic and isotropic hardening is correlated to the formation of the dislocation and dislocation dynamics (Asaro and Lubarda 2006; Voyiadjis et al. 2010). Changes in entropy of the molecular chains in the polymer materials can describe the hardening effects (Beloshenko et al. 2005; Voyiadjis et al. 2011, 2012a, c). Tip crack plasticity and crack arresting effect induce the hardening effect in the damage processes (Voyiadjis and Kattan 2006; Voyiadjis et al. 2011). In the case of the healing process, the hardening/softening may be induced to the system from different sources such as chemical decomposition changes after the healing, or inelastic strains during the healing process, or even diffusion of the healing agent into the microcracks. The chemical reaction of the healing microencapsulated agent may result in the higher mechanical properties after curing, or inelastic strains during programming of a shape memory-based self-healing system may result in hardening effects (Voyiadjis et al. 2011). The hardening due to plasticity, damage, and healing can be represented by hardening evolution laws based on the experimental evidences. The kinematic and isotropic concept provides the tool to model these phenomena.

Substituting the introduced internal variables into the Helmholtz free energy, one obtains the following relation (Voyiadjis et al. 2011):

$$\Psi = \Psi\left(\epsilon_{ij}, \epsilon_{ij}^{p}, \alpha_{ij}, p_{ij}, d_{ij}, d_{ij}^{K}, d_{ij}^{I}, h_{ij}, h_{ij}^{K}, h_{ij}^{I}\right) \tag{50}$$

Using Eq. 45, the thermodynamic conjugate forces associated with the discussed flux variables are obtained as follows (Voyiadjis et al. 2011):

$$\sigma_{ij}^{p} = -\rho\frac{\partial\Psi}{\partial\epsilon_{ij}^{p}}; \quad y_{ij}^{d} = -\rho\frac{\partial\Psi}{\partial d_{ij}}; \quad y_{ij}^{h} = -\rho\frac{\partial\Psi}{\partial h_{ij}}$$

$$y_{ij}^{pK} = -\rho \frac{\partial \Psi}{\partial \alpha_{ij}}; \quad y_{ij}^{dK} = -\rho \frac{\partial \Psi}{\partial d_{ij}^{K}}; \quad y_{ij}^{hK} = -\rho \frac{\partial \Psi}{\partial h_{ij}^{K}} \tag{51}$$

$$y_{ij}^{pI} = -\rho \frac{\partial \Psi}{\partial p_{ij}}; \quad y_{ij}^{dI} = -\rho \frac{\partial \Psi}{\partial d_{ij}^{I}}; \quad y_{ij}^{hI} = -\rho \frac{\partial \Psi}{\partial h_{ij}^{I}}$$

where σ_{ij}^{p} is the conjugate force for the plastic strain, and y_{ij}^{d} and y_{ij}^{h} are the conjugate forces associated with the damage and healing variables, respectively. The variables y_{ij}^{pK}, y_{ij}^{dK}, and y_{ij}^{hK} represent the conjugate forces associated with the kinematic hardening effect of plasticity, damage, and healing processes, respectively. The variables y_{ij}^{pI} and y_{ij}^{dI} indicate the conjugate forces associated with the isotropic hardening effect of the plasticity and damage processes, respectively, and y_{ij}^{hI} represents the isotropic hardening/softening conjugated force of the healing process. The Helmholtz free energy is decomposed as follows (Hansen and Schreyer 1994; Voyiadjis et al. 2011):

$$\Psi = W\left(\epsilon_{ij}, \epsilon_{ij}^{p}, d_{ij}, h_{ij}\right) + H\left(\alpha_{ij}, p_{ij}, d_{ij}^{K}, d_{ij}^{I}, h_{ij}^{K}, h_{ij}^{I}\right) + G_{dh}\left(d_{ij}, h_{ij}\right) \tag{52}$$

where the *elastic* part of the Helmholtz free energy W is defined as follows (Voyiadjis et al. 2011):

$$W\left(\epsilon_{ij}, \epsilon_{ij}^{p}, d_{ij}, h_{ij}\right) = \frac{1}{2}\left(\epsilon_{ij} - \epsilon_{ij}^{p}\right) E_{ijkl}\left(\epsilon_{kl} - \epsilon_{kl}^{p}\right) \tag{53}$$

where the fourth-order elastic stiffness tensor E_{ijkl} is a function of the damage variable d_{ij} and the healing variable h_{ij}. These parameters are defined based on their corresponding theories herein. The unknown *hardening function* $H(\alpha_{ij}, p_{ij}, d_{ij}^{K}, d_{ij}^{I}, h_{ij}^{K}, h_{ij}^{I})$ shows the effect of different hardening/softening processes. The final term $G_{dh}(d_{ij}, h_{ij})$ is related to surface energy which takes into account the microcrack and microsurfaces propagation and recovery (Voyiadjis et al. 2011). This term implies that some part of energy in a damaged material is converted to increase the surface energy and the remaining part converts to heat (Hansen and Schreyer 1994; Barbero et al. 2005; Voyiadjis et al. 2011). In the case of healing process, $G_{dh}(d_{ij}, h_{ij})$ may represent the surface energy reduction due to the diffusion process of the healing agent (Voyiadjis et al. 2011; Li and Shojaei 2012). This concept can be utilized for both the uncoupled (Li and Nettles 2010; Voyiadjis et al. 2011; Li and Shojaei 2012) and coupled self-healing systems (White et al. 2001; Kirkby et al. 2009). In other words, in a healing process, the surface energy reduction due to the diffusion of the healing agent besides considering external loading for crack closures can be shown using this function.

Using Eqs. 43, 51, and 52, the Cauchy stress σ_{ij} yields to its classical form and the conjugate force σ_{ij}^{p} is obtained as follows (Voyiadjis et al. 2011, 2012a):

$$\sigma_{ij} = E_{ijkl}\left(\epsilon_{ij} - \epsilon_{ij}^{p}\right); \quad \sigma_{ij} = \sigma_{ij}^{p} \tag{54}$$

In order to derive the evolution equations for the internal variables associated with plasticity, damage, and healing, the power of dissipation, Eq. 44, along with

the Helmholtz free energy definition, i.e., Ψ, is used to obtain the following relation (Voyiadjis et al. 2011, 2012a):

$$\begin{aligned} \Gamma = -\rho \Bigg(& \frac{\partial \Psi}{\partial \epsilon_{ij}^{p}} : \dot{\epsilon}_{ij}^{p} + \frac{\partial \Psi}{\partial \alpha_{ij}} : \dot{\alpha}_{ij} + \frac{\partial \Psi}{\partial p_{ij}} : \dot{p}_{ij} + \frac{\partial \Psi}{\partial d_{ij}} : \dot{d}_{ij} + \frac{\partial \Psi}{\partial d_{ij}^{K}} : \dot{d}_{ij}^{K} \\ & + \frac{\partial \Psi}{\partial d_{ij}^{I}} : \dot{d}_{ij}^{I} + \frac{\partial \Psi}{\partial h_{ij}} : \dot{h}_{ij} + \frac{\partial \Psi}{\partial h_{ij}^{K}} : \dot{h}_{ij}^{K} + \frac{\partial \Psi}{\partial h_{ij}^{I}} : \dot{h}_{ij}^{I} \Bigg) \end{aligned} \tag{55}$$

Substituting the thermodynamic conjugate forces from Eq. 51 into Eq. 55 results in the following relation for the dissipative power (Voyiadjis et al. 2011, 2012a):

$$\Gamma = \sigma_{ij}\dot{\epsilon}_{ij}^{p} + y_{ij}^{pK}\dot{\alpha}_{ij} + y_{ij}^{pI}\dot{p}_{ij} + y_{ij}^{d}\dot{d}_{ij} + y_{ij}^{dK}\dot{d}_{ij}^{K} + y_{ij}^{dI}\dot{d}_{ij}^{I} + y_{ij}^{h}\dot{h}_{ij} + y_{ij}^{hK}\dot{h}_{ij}^{K} + y_{ij}^{hI}\dot{h}_{ij}^{I} \tag{56}$$

Equation 56 is the most general form for the dissipative power function including the associated kinematic and isotropic hardening terms for each process. The generalized form of the yield, damage, and healing thresholds is introduced within the classical plasticity formulation framework, where the kinematic and isotropic hardening terms are represented by two separate functions (Voyiadjis and Foroozesh 1990; Chaboche 1997; Voyiadjis et al. 2011, 2012a). The coupling between all processes is captured by incorporating an additional term into all criteria which captures the generalized coupling between the different processes. The generalized form of the yield surface is introduced in the following expression (Voyiadjis et al. 2011, 2012a):

$$f_{ij}^{p} \equiv f^{p1}\left(\sigma_{ij} - y_{ij}^{pK}\right) - f^{p2}\left(y_{ij}^{pI}\right) - f^{p3}\left(y_{ij}^{d}, y_{ij}^{dK}, y_{ij}^{dI}\right) - f^{p4}\left(y_{ij}^{h}, y_{ij}^{hK}, y_{ij}^{hI}\right) - \sigma_{ij}^{y} \leq 0 \tag{57}$$

where σ_{ij}^{y} is the initial yield stress for a heterogeneous material, f^{p1} represents the plastic kinematic hardening effect, f^{p2} captures the plastic isotropic hardening effect, and f^{p3} and f^{p4} represent the effect of damage and healing processes on the plastic yield criterion, respectively. In a similar manner, the generalized damage surface is defined as follows (Voyiadjis et al. 2011, 2012a):

$$f_{ij}^{d} = f^{d1}\left(y_{ij}^{d} - y_{ij}^{dK}\right) - f^{d2}\left(y_{ij}^{dI}\right) - f^{d3}\left(\sigma_{ij}, y_{ij}^{pK}, y_{ij}^{pI}\right) - f^{d4}\left(y_{ij}^{h}, y_{ij}^{hK}, y_{ij}^{hI}\right) - \omega_{ij}^{d0} \leq 0 \tag{58}$$

where ω_{ij}^{d0} represents the initial size of the damage surface in different directions and the first two terms f^{d1} and f^{d2} represent the kinematic and isotropic hardening due to the damage process, respectively, and the last two terms f^{d3} and f^{d4} capture

the effect of the plastic deformation and the healing process on the damage criterion, respectively. The generalized healing surface is defined as follows: the first two terms f^{h1} and f^{h2} show the respective kinematic and isotropic hardening/softening due to the healing process and the last two terms f^{h3} and f^{h4} are capturing the effects of the plastic deformation and the damage on the healing criterion as follows (Voyiadjis et al. 2011, 2012a):

$$f_{ij}^{h} \equiv f^{h1}\left(y_{ij}^{d} - y_{ij}^{dK}\right) - f^{h2}\left(y_{ij}^{dI}\right) - f^{h3}\left(\sigma_{ij}, y_{ij}^{pK}, y_{ij}^{pI}\right) - f^{h4}\left(y_{ij}^{h}, y_{ij}^{hK}, y_{ij}^{hI}\right) - \omega_{ij}^{h0} \leq 0 \tag{59}$$

where ω_{ij}^{h0} is the initial size of the healing surfaces. All of these functions are required to be homogenous of order one.

Plasticity and damage yield criteria have been investigated extensively in the literature (Chaboche 1993; Voyiadjis and Abu Al-Rub 2003; Voyiadjis and Kattan 2006; Khan et al. 2010a, b) and only the healing surface concept may require more elaboration here. The healing process is activated when the consistency conditions are satisfied, that is, $f^{h} = 0$ and $\dot{f}^{h} = 0$. The healing criterion is a function of the relevant healing mechanism. For example, in the case of microencapsulated healing agent, as soon as the wall of the microcapsules is broken, the healing process is activated and the healing criterion can be formulated with respect to the required stress to fracture the wall of the microcapsules.

The principal of extremized entropy production during a thermodynamic process is applied to the power dissipation function Eq. 56 along with the constraints of Eqs. 57, 58, and 59. Assuming isotropic yield surfaces for the plasticity, damage, and healing phenomena, the resulting Lagrangian functional which should be extremized is as follows (Voyiadjis et al. 2011, 2012a):

$$\gamma^{*} = \Gamma - \dot{\lambda}^{p} f^{p} - \dot{\lambda}^{d} f^{d} - \dot{\lambda}^{h} f^{h} \tag{60}$$

where $\dot{\lambda}^{p}$, $\dot{\lambda}^{d}$, and $\dot{\lambda}^{h}$ are plasticity, damage, and healing Lagrangian multipliers, respectively, to enforce the yielding constrains for each of the processes. Applying the three stationary conditions $\partial\gamma^{*}/\partial\sigma_{ij} = 0$, $\partial\gamma^{*}/\partial\sigma_{ij}^{pK} = 0$, and $\partial\gamma^{*}/\partial\sigma_{ij}^{pI} = 0$ on the resulting functional yields the following coupled evolution equations for the set of internal variables associated with the plastic deformation (Voyiadjis et al. 2011):

$$\dot{\epsilon}_{ij}^{p} = \dot{\lambda}^{p} \frac{\partial f^{p1}\left(\sigma_{ij} - y_{ij}^{pK}\right)}{\partial\sigma_{ij}} + \dot{\lambda}^{d} \frac{\partial f^{d3}\left(\sigma_{ij}, y_{ij}^{pK}, y_{ij}^{pI}\right)}{\partial\sigma_{ij}} + \dot{\lambda}^{h} \frac{\partial f^{h3}\left(\sigma_{ij}, y_{ij}^{pK}, y_{ij}^{pI}\right)}{\partial\sigma_{ij}} \tag{61}$$

$$\dot{\alpha}_{ij} = \dot{\lambda}^{p} \frac{\partial f^{p1}\left(\sigma_{ij} - y_{ij}^{pK}\right)}{\partial y_{ij}^{pK}} + \dot{\lambda}^{d} \frac{\partial f^{d3}\left(\sigma_{ij}, y_{ij}^{pK}, y_{ij}^{pI}\right)}{\partial y_{ij}^{pK}} + \dot{\lambda}^{h} \frac{\partial f^{h3}\left(\sigma_{ij}, y_{ij}^{pK}, y_{ij}^{pI}\right)}{\partial y_{ij}^{pK}} \tag{62}$$

$$\dot{p}_{ij} = \dot{\lambda}^p \frac{\partial f^{p2}\left(y_{ij}^{pI}\right)}{\partial y_{ij}^{pI}} + \dot{\lambda}^d \frac{\partial f^{d3}\left(\sigma_{ij}, y_{ij}^{pK}, y_{ij}^{pI}\right)}{\partial y_{ij}^{pI}} + \dot{\lambda}^h \frac{\partial f^{h3}\left(\sigma_{ij}, y_{ij}^{pK}, y_{ij}^{pI}\right)}{\partial y_{ij}^{pI}} \tag{63}$$

Using the three conditions $\partial\gamma^*/\partial y_{ij}^d = 0$, $\partial\gamma^*/\partial y_{ij}^{dK} = 0$, and $\partial\gamma^*/\partial y_{ij}^{dI} = 0$ yields the following evolution equations for the set of the internal variables associated with the damage process (Voyiadjis et al. 2011):

$$\dot{d}_{ij} = \dot{\lambda}^d \frac{\partial f^{d1}\left(y_{ij}^d - y_{ij}^{dK}\right)}{\partial y_{ij}^d} + \dot{\lambda}^p \frac{\partial f^{p3}\left(y_{ij}^d, y_{ij}^{dK}, y_{ij}^{dI}\right)}{\partial y_{ij}^d} + \dot{\lambda}^h \frac{\partial f^{h4}\left(y_{ij}^d, y_{ij}^{dK}, y_{ij}^{dI}\right)}{\partial y_{ij}^d} \tag{64}$$

$$\dot{d}_{ij}^K = \dot{\lambda}^d \frac{\partial f^{d1}\left(y_{ij}^d - y_{ij}^{dK}\right)}{\partial y_{ij}^{dK}} + \dot{\lambda}^p \frac{\partial f^{p3}\left(y_{ij}^d, y_{ij}^{dK}, y_{ij}^{dI}\right)}{\partial y_{ij}^{dK}} + \dot{\lambda}^h \frac{\partial f^{h4}\left(y_{ij}^d, y_{ij}^{dK}, y_{ij}^{dI}\right)}{\partial y_{ij}^{dK}} \tag{65}$$

$$\dot{d}^I = \dot{\lambda}^d \frac{\partial f^{d2}\left(y_{ij}^{dI}\right)}{\partial y_{ij}^{dI}} + \dot{\lambda}^p \frac{\partial f^{p3}\left(y_{ij}^d, y_{ij}^{dK}, y_{ij}^{dI}\right)}{\partial y_{ij}^{dI}} + \dot{\lambda}^h \frac{\partial f^{h4}\left(y_{ij}^d, y_{ij}^{dK}, y_{ij}^{dI}\right)}{\partial y_{ij}^{dI}} \tag{66}$$

Finally, applying the three conditions $\partial\gamma^*/\partial y_{ij}^h = 0$, $\partial\gamma^*/\partial y_{ij}^{hK} = 0$, and $\partial\gamma^*/\partial y_{ij}^{hI} = 0$ results in the following evolution equations for the set of internal variables associated with the healing process (Voyiadjis et al. 2011):

$$\dot{h}_{ij} = \dot{\lambda}^h \frac{\partial f^{h1}\left(y_{ij}^h - y_{ij}^{hK}\right)}{\partial y_{ij}^h} + \dot{\lambda}^p \frac{\partial f^{p4}\left(y_{ij}^d, y_{ij}^{dK}, y_{ij}^{dI}\right)}{\partial y_{ij}^h} + \dot{\lambda}^d \frac{\partial f^{d4}\left(y_{ij}^d, y_{ij}^{dK}, y_{ij}^{dI}\right)}{\partial y_{ij}^h} \tag{67}$$

$$\dot{h}_{ij}^K = \dot{\lambda}^h \frac{\partial f^{h1}\left(y_{ij}^h - y_{ij}^{hK}\right)}{\partial y_{ij}^{hK}} + \dot{\lambda}^p \frac{\partial f^{p4}\left(y_{ij}^d, y_{ij}^{dK}, y_{ij}^{dI}\right)}{\partial y_{ij}^{hK}} + \dot{\lambda}^d \frac{\partial f^{d4}\left(y_{ij}^d, y_{ij}^{dK}, y_{ij}^{dI}\right)}{\partial y_{ij}^{hK}} \tag{68}$$

$$\dot{h}_{ij}^I = \dot{\lambda}^h \frac{\partial f^{h2}\left(y_{ij}^{hI}\right)}{\partial y_{ij}^{hI}} + \dot{\lambda}^p \frac{\partial f^{p4}\left(y_{ij}^d, y_{ij}^{dK}, y_{ij}^{dI}\right)}{\partial y_{ij}^{hI}} + \dot{\lambda}^d \frac{\partial f^{d4}\left(y_{ij}^d, y_{ij}^{dK}, y_{ij}^{dI}\right)}{\partial y_{ij}^{hI}} \tag{69}$$

Coupling in the evolution laws between the different processes is obvious in Eqs. 61–69. The unknown Lagrangian multipliers could be obtained through the consistency conditions $\dot{f}^p = 0, \dot{f}^d = 0$, and $\dot{f}^h = 0$ (Voyiadjis et al. 2011, 2012a).

Computational Aspect and Simulation Results

Generally speaking, the computational aspects of the coupled problem can be divided into two distinctive modules. One of them ensures the basic solid mechanics governing equations are satisfied while boundary conditions and geometry of

structure are enforced. Basically, these relations are equilibrium, strain–stress, and compatibility relations which result in an initial boundary value problem (IBVP). The computed state of inelastic strain, damage, and healing variables is iteratively corrected (Newton–Raphson technique) to satisfy this IBVP along with the initial and boundary conditions (Voyiadjis et al. 2012a). The second computational module deals with the flow rules and nonlinear governing equations of inelastic strain, damage, and healing variables. An incremental solution, which provides history dependency in a coupled inelastic–damage–healing problem, updates the internal variables. These updated values are utilized during the iteration process of the IBVP solution and this task continues until IBVP solution is converged. Different solution algorithms may be incorporated to solve the nonlinear governing equations of inelasticity, damage, and healing, including iterative return mapping (Simo and Hughes 1997) or non-iterative methods (Sivakumar and Voyiadjis 1997). A comprehensive description of solution algorithms for coupled as well as uncoupled inelastic–damage–healing problems can be found in Voyiadjis et al. (2011). The return mapping technique proposed by Mendelson (Mendelson and Manson 1957) is generalized here for all the three processes.

To compute the damage inside the materials, the proposed viscoplasticity theory is used to simulate the half-cycle strain-controlled compression test of the SMP where damage and plastic deformations are coupled. Consequently, the Cauchy stress σ_{ij} and stress rate $\dot{\sigma}_{ij}$ in the real damaged state can be introduced into the damage computational module from the viscoplasticity solution. Once the damage criterion is satisfied, following each inelastic strain increment, a damage increment is computed while the updated values of stress and stress rate are incorporated. The computed damage values are then used to update the elastic modulus for the next load increment. This computational procedure allows fully coupling between inelastic and damage computation (Voyiadjis et al. 2012a).

In a coupled self-healing system, the healing computational module receives the updated inelastic deformation and damages from the inelastic and damage computational modules. In the case of a coupled healing system, after each increment of the inelastic deformation and damage, the healing computation module computes the amount of the healing. While in an uncoupled self-healing system, after a certain amount of damage, all final inelastic deformations and damage variables are introduced into the healing computational module and the healing is computed based on these values. Finally, the updated elastic modulus is introduced to the viscoplasticity module for the next load step of computations. This computational procedure allows each mechanism to follow its respective governing equations while full coupling is considered (Voyiadjis et al. 2012a).

Figure 1a represents changes in the elastic modulus of the SMP sample under compression tests with a strain rate of 0.002 (s^{-1}) in which the proposed damage theory is used to capture this behavior. Figure 1b shows the compression test results of SMP at room temperature with a strain rate of 0.002 (s^{-1}) and the simulation is obtained using the proposed plasticity theory. Table 1 gives the required material constants for SMP compression. In order to show the capability of the proposed theory

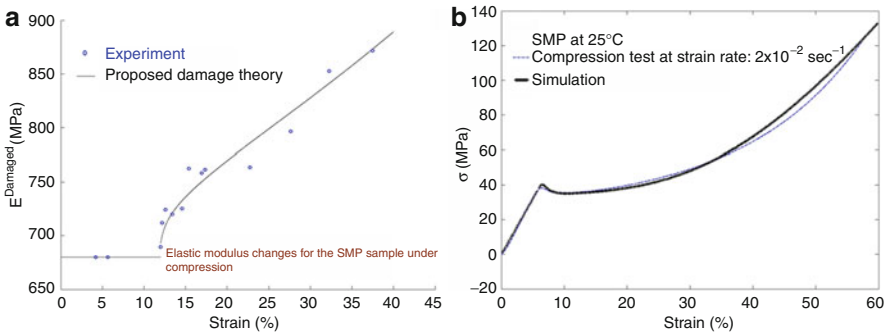


Fig. 1 Experimental and simulation results for SMP sample under compression at 0.002 (s^{-1}) loading rate: (**a**) elastic modulus changes; (**b**) inelastic response (Voyiadjis et al. 2012a, c)

Table 1 Material constant for SMP ($\bar{E} = 642$ (MPa) and $\sigma_y \cong 38$ (MPa)) compression at 25 °C, with substituting $\beta = \Lambda_{(2)} = 0$ and $k_1 = b_{(1)} = c_{(2)} = b_{(2)} = 1$ into the inelastic kinematic hardening evolution law (Voyiadjis et al. 2012a, c)

$$\dot{X}_x = \left|\dot{\epsilon}_x^p\right|\left(\tfrac{2}{3} + (X_x - X^s)\left(a_{(1)}\left|\epsilon_x^p\right|^{c_{(1)}-1}\left(z_{(1)} + \left|\epsilon_x^p\right|^{c_{(1)}}\right)^{\Lambda_{(1)}} + a_{(2)}\left(z_{(2)} + \left|\epsilon_x^p\right|\right)^{-1}\right)\right)$$

Strain spectrum	X^s	$a_{(1)}$	$z_{(1)}$	$c_{(1)}$	$\Lambda_{(1)}$	$a_{(2)}$	$z_{(2)}$
$\epsilon_f^I \le \epsilon_x < 6.3$	300	0.5	0	1	1	0	–
$6.3 \le \epsilon_x < 10$	$110X_0^{II}$	−15	0.5	0.5	5	−25	1.5
$10 \le \epsilon_x \le 60$	$1.1X_0^{III}$	−5	1	0.5	−1	8	0.5

Table 2 Material constants for SMP damage simulation at 25 °C with substituting $\beta' = \Lambda'_{(2)} = a_{(2)} = 0$ and $k'_1 = b'_{(1)} = c'_{(2)} = b'_{(2)} = 1$ into the damage kinematic hardening evolution law (Voyiadjis et al. 2012a, c)

$$\dot{y}_x^{dK} = \left|\dot{d}_x\right|\left(\tfrac{2}{3} + \left(y_x^{dK} - y_x^{dK,s}\right)\left(a'_{(1)}\left|d_x\right|^{c'_{(1)}-1}\left(z'_{(1)} + \left|d_x\right|^{c'_{(1)}}\right)^{\Lambda'_{(1)}}\right)\right)$$

Damage initiation check	$y_x^{dK,s}$	$a'_{(1)}$	$z'_{(1)}$	$c'_{(1)}$	$\Lambda'_{(1)}$
$\epsilon_f^I \le \epsilon_x$ and $\dot{\sigma}_x > 0$	0.3	20	1	1	−1

to simulate a vast range of irregular inelastic deformations of glassy polymers, additional experimental data, regarding the inelastic mechanical response of polyethylene terephthalate (PET), is presented here. It is worthwhile to indicate that precise inelastic simulations of material systems are crucial in the damage and healing computations where the stress state in the real damaged and/or healed configurations is utilized to compute the state of the damage and/or healing. Table 2 shows the material constants to simulate PET tension at room temperature, and the simulation results are shown in Fig. 2 and experiments are reported by G'Sell et al. (2002). The inflection in the experimental data points shows the respective softening and subsequent strain hardening region for PET which is captured by the proposed model. In these tables, ϵ_f^I indicates the yield strain and X_0^{II} and X_0^{III} are, respectively, the final values of X^s at the previous strain spectrum (Voyiadjis et al. 2012a, c) (Table 3).

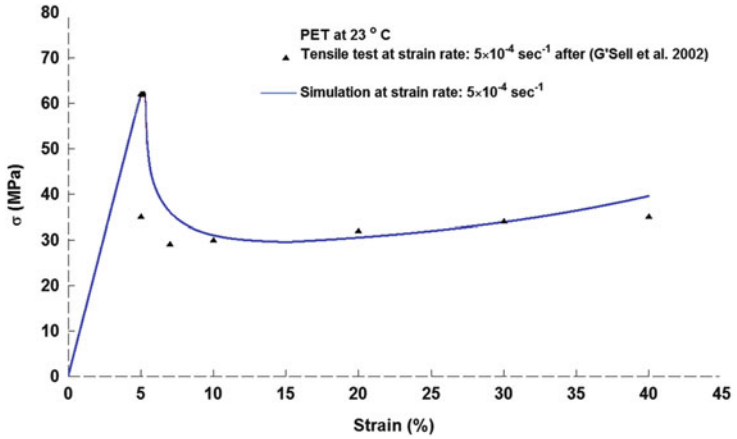


Fig. 2 Experimental and simulation results for PET at 23 °C (Voyiadjis et al. 2012b) (Experiments are after G'Sell et al. (2002))

Table 3 Material constant for PET ($\bar{E} = 1240$ (MPa) and $\sigma_y \cong 62$ (MPa)) at 23 °C, with substituting $\beta = \Lambda_{(2)} = 0$ and $k_1 = b_{(1)} = c_{(2)} = b_{(2)} = 1$ and into the inelastic kinematic hardening law (Voyiadjis et al. 2012a, c)

$$\dot{X}_x = \left|\dot{\epsilon}^p_x\right|\left(\tfrac{2}{3} + (X_x - X^s)\left(a_{(1)}\left|\epsilon^p_x\right|^{c_{(1)}-1}\left(z_{(1)} + \left|\epsilon^p_x\right|^{c_{(1)}}\right)^{\Lambda_{(1)}} + a_{(2)}\left(z_{(2)} + \left|\epsilon^p_x\right|\right)^{-1}\right)\right)$$

Strain spectrum	X^s	$a_{(1)}$	$z_{(1)}$	$c_{(1)}$	$\Lambda_{(1)}$	$a_{(2)}$	$z_{(2)}$
$\epsilon_f^I \le \epsilon_x < 5.2$	300	0.5	0	1	1	0	–
$5.2 \le \epsilon_x < 15$	$110X_0^{II}$	−8	0.5	0.5	5	−25	1.5
$15 \le \epsilon_x \le 40$	$1.04X_0^{III}$	−12	1	0.5	−1	4	1

Conclusion

The mechanics of damage and healing processes is investigated for self-healing systems, and the bases of a physically based modeling of these phenomena are constituted. Thermodynamics of the plasticity, damage, and healing is investigated and coupled constitutive equations are developed in the most generalized forms. Kinematic and isotropic hardening effects for all processes are incorporated into the formulation together with a new set of potential functions in order to achieve mathematical competency in precisely simulating each phenomenon. The irregular inelastic deformation of SMP and PET is captured using the proposed viscoplasticity theory. The developed viscoplasticity theory is based on a strain spectrum strategy where each strain spectrum is captured using a set of material constants. These strain spectrums are functions of strain rate and temperature and they should be evaluated experimentally. The performance of the proposed

damage–healing variables and coupled constitutive equations is examined for an SMP-based self-healing system. The proposed anisotropic damage–healing variables provide the designers with the ability to measure the state of the damage and healing in a highly anisotropic self-healing system, and the constitutive equations perform quite well in simulating the plastic and damage responses of glassy polymers. As shown in this work, the proposed potential functions are capable of capturing the most nonlinear inelastic deformation and the damage responses of the polymeric-based material systems. However availability of experiments for self-healing systems is limited in the literature, it is expected that the healing process shows a similar trend but inverse to that of the damaging process in which the proposed theory provides mathematical competency to capture it. The healing test setup is under preparation and will be presented in the near future together with the corresponding healing experiments. This will be reported together with the corresponding simulations, utilizing the proposed theory. It is worth noting that the proposed theory considers a continuum scale without incorporating the microstructural changes during each process. It is well known that such a macroscopic model, although may provide good correlation with a specific loading condition, may not be capable of capturing a general case of loading in which loading histories deviate from the experiments used to fit the material parameters. This opens a new field to incorporate such microstructural changes into the governing equations of these coupled processes. The microstructural physics of the problem can be incorporated in CDM framework by introducing fabric tensors.

References

R.J. Asaro, V. Lubarda, *Mechanics of Solids and Materials* (Cambridge University Press, New York, 2006)

M.F. Ashby, C. Gandhi, D.M.R. Taplin, Overview No. 3 fracture-mechanism maps and their construction for f.c.c. metals and alloys. Acta Metall. **27**(5), 699–729 (1979)

E.J. Barbero, F. Greco, P. Lonetti, Continuum damage-healing mechanics with application to self-healing composites. Int. J. Damage Mech. **14**(1), 51–81 (2005)

V.A. Beloshenko, V.N. Varyukhin, Y.V. Voznyak, The shape memory effect in polymers. Russ. Chem. Rev. **74**(3), 265 (2005)

J.L. Chaboche, On some modifications of kinematic hardening to improve the description of ratchetting effects. Int. J. Plast. **7**(7), 661–678 (1991)

J.L. Chaboche, Cyclic viscoplastic constitutive equations, part I: a thermodynamically consistent formulation. J. Appl. Mech. **60**(4), 813–821 (1993)

J.L. Chaboche, Thermodynamic formulation of constitutive equations and application to the viscoplasticity and viscoelasticity of metals and polymers. Int. J. Solids Struct. **34**(18), 2239–2254 (1997)

J.L. Chaboche, A review of some plasticity and viscoplasticity constitutive theories. Int. J. Plast. **24**(10), 1642–1693 (2008)

C. G'Sell, J.M. Hiver, A. Dahoun, Experimental characterization of deformation damage in solid polymers under tension, and its interrelation with necking. Int. J. Solids Struct. **39**, 3857–3872 (2002)

N.R. Hansen, H.L. Schreyer, A thermodynamically consistent framework for theories of elastoplasticity coupled with damage. Int. J. Solids Struct. **31**(3), 359–389 (1994)
A.S. Khan, M. Baig, Anisotropic responses, constitutive modeling and the effects of strain-rate and temperature on the formability of an aluminum alloy. Int. J. Plast. **27**(4), 522–538 (2011)
A.S. Khan, A. Pandey, T. Stoughton, Evolution of subsequent yield surfaces and elastic constants with finite plastic deformation. Part II: a very high work hardening aluminum alloy (annealed 1100 Al). Int. J. Plast. **26**(10), 1421–1431 (2010a)
A.S. Khan, A. Pandey, T. Stoughton, Evolution of subsequent yield surfaces and elastic constants with finite plastic deformation. Part III: yield surface in tension-tension stress space (Al 6061-T 6511 and annealed 1100 Al). Int. J. Plast. **26**(10), 1432–1441 (2010b)
E.L. Kirkby, V.J. Michaud, J.A.E. Månson, N.R. Sottos, S.R. White, Performance of self-healing epoxy with microencapsulated healing agent and shape memory alloy wires. Polymer **50**(23), 5533–5538 (2009)
H. Lee, K. Peng, J. Wang, An anisotropic damage criterion for deformation instability and its application to forming limit analysis of metal plates. Eng. Fract. Mech. **21**(5), 1031–1054 (1985)
G. Li, D. Nettles, Thermomechanical characterization of a shape memory polymer based self-repairing syntactic foam. Polymer **51**(3), 755–762 (2010)
G. Li, A. Shojaei, A viscoplastic theory of shape memory polymer fibres with application to self-healing materials. Proc. R. Soc. A **468**(2144), 2319–2346 (2012). doi:10.1098/rspa.2011.0628
J. Lubliner, On the thermodynamic foundations of non-linear solid mechanics. Int. J. Non-Linear Mech. **7**(3), 237–254 (1972)
S. Miao, M.L. Wang, H.L. Schreyer, Constitutive models for healing of materials with application to compaction of crushed rock salt. J. Eng. Mech. **121**(10), 1122–1129 (1995)
R.W. Rice, S.W. Freiman, J.J. Mecholsky, The dependence of strength-controlling fracture energy on the flaw-size to grain-size ratio. J. Am. Ceram. Soc. **63**(3–4), 129–136 (1980)
A. Shojaei, G. Li, G.Z. Voyiadjis, Cyclic viscoplastic-viscodamage analysis of shape memory polymers fibers with application to self-healing smart materials. J. Appl. Mech. **80**(1), 011014–011015 (2013a)
A. Shojaei, G.Z. Voyiadjis, P.J. Tan, Viscoplastic constitutive theory for brittle to ductile damage in polycrystalline materials under dynamic loading. Int. J. Plast. (2013b). doi:10.1016/j.ijplas.2013.02.009
J.C. Simo, T.J.R. Hughes, Computational inelasticity. New York, Springer (1997)
M.S. Sivakumar, G.Z. Voyiadjis, A simple implicit scheme for stress response computation in plasticity models. Journal of Computational Mechanics. **20**(6), 520–529 (1997)
V. Tvergaard, J.W. Hutchinson, The relation between crack growth resistance and fracture process parameters in elastic–plastic solids. J. Mech. Phys. Solids **40**(6), 1377–1397 (1992)
G.Z. Voyiadjis, R.K. Abu Al-Rub, Thermodynamic based model for the evolution equation of the backstress in cyclic plasticity. Int. J. Plast. **19**(12), 2121–2147 (2003)
G. Voyiadjis, I. Basuroychowdhury, A plasticity model for multiaxial cyclic loading and ratchetting. Acta Mech. **126**(1), 19–35 (1998)
G.Z. Voyiadjis, M. Foroozesh, Anisotropic distortional yield model. J. Appl. Mech. **57**(3), 537–547 (1990)
Z. Voyiadjis, P.I. Kattan, *Advances in Damage Mechanics* (Elsevier, London, 2006)
G.Z. Voyiadjis, P.I. Kattan, A comparative study of damage variables in continuum damage mechanics. Int. J. Damage Mech. **18**(4), 315–340 (2009)
G.Z. Voyiadjis, G. Pekmezi, B. Deliktas, Nonlocal gradient-dependent modeling of plasticity with anisotropic hardening. Int. J. Plast. **26**(9), 1335–1356 (2010)
G.Z. Voyiadjis, A. Shojaei, G. Li, A thermodynamic consistent damage and healing model for self healing materials. Int. J. Plast. **27**(7), 1025–1044 (2011)

G.Z. Voyiadjis, A. Shojaei, G. Li, A generalized coupled viscoplastic- viscodamage- viscohealing theory for glassy polymers. Int. J. Plast. **28**(1), 21–45 (2012a)

G.Z. Voyiadjis, A. Shojaei, G. Li, P. Kattan, Continuum damage-healing mechanics with introduction to new healing variables. Int. J. Damage Mech. **21**(3), 391–414 (2012b)

G.Z. Voyiadjis, A. Shojaei, G. Li, P.I. Kattan, A theory of anisotropic healing and damage mechanics of materials. Proc. R. Soc. A Math. Phys. Eng. Sci. **468**(2137), 163–183 (2012c). doi:10.1098/rspa.2011.0326

S.R. White, N.R. Sottos, P.H. Geubelle, J.S. Moore, M.R. Kessler, S.R. Sriram, E.N. Brown, S. Viswanathan, Autonomic healing of polymer composites. Nature **409**(6822), 794–797 (2001)

S. Yazdani, H.L. Schreyer, Combined plasticity and damage mechanics model for plain concrete. J. Eng. Mech. **116**(7), 1435–1450 (1990)

Continuum Damage-Healing Mechanics

46

George Z. Voyiadjis and Amir Shojaei

Contents

Abstract

Microscale damage mechanisms, such as microcracks or microvoids, are well-known damage process zones for the formation of the macroscale cracks. The microscale defects, which are in the order of submicrons, will coalesce and branch within the course of the deformation and gradually form the macroscale damages. Healing of the microscale damages prohibits the formation of the macroscale defect zones and increases the life of the structures. Developing new healing strategies have become a hot research topic in the field of self-healing materials during recent years and many healing strategies have been proposed. In this chapter, the mathematics of healing is investigated within the

G.Z. Voyiadjis (✉)
Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA
e-mail: voyiadjis@eng.lsu.edu

A. Shojaei
Department of Mechanical and Industrial Engineering, Louisiana State University, Baton Rouge, LA, USA
e-mail: ashoja1@lsu.edu; a.shojaei.mech.eng@gmail.com

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_46

continuum damage-healing mechanics (CDHM) framework. This aids smart material designers for the characterization of the coupled damage-healing process. Special emphasis is given on definition of new healing variables within the framework of CDHM. These novel damage-healing variables were formerly proposed by the authors and their performances have been examined in coupled damage-healing simulations (Voyiadjis et al., Int J Plast 27:1025–1044, 2011; Voyiadjis et al., Proc Roy Soc A Math Phys Eng Sci 468:163–183, 2012a; Voyiadjis et al., Int J Plast 28:21–45, 2012c). The proposed CDHM framework together with the developed thermodynamic consistent description of the microscale healing and damaging processes provide a well-structured method for accurately predicting the degradation and healing mechanisms in smart self-healing material systems.

Introduction

The healing of the damages inside the material systems and its application in smart structures have been topics of intensive research in the last decade and many of the developed healing strategies have been deployed in real applications, e.g., biomedical applications (Adam 1999; Simpson et al. 2000) or load-carrying self-healing structures (Shojaei et al. 2013; Brown et al. 2002; Plaisted and Nemat-Nasser 2007; Kirkby et al. 2008; Li and John 2008; Kirkby et al. 2009; John and Li 2010; Li and Uppu 2010; Nji and Li 2010b; Li and Shojaei 2012). In general, each one of these newly developed healing schemes has been designed to cure specific type of damage category inside a specific material system. The damages categories may be classified based on their characteristic lengths into two classes: (i) structural length scale – the length scale is in order of the structural length scales, e.g., a few (cm), and (ii) microscale – the characteristic length is in order of the microstructural length scale, e.g., a few (μm). Most of the self-healing systems are designed on the premises that healing the microscale damages will seize the damages zone to propagate and produce a structural scale damage.

In order to commercialize the self-healing schemes, the development of rigorous modeling techniques is crucial for predicting their responses to the complex thermomechanical loading conditions and also for evaluating their healing efficiency. The theoretical developments for the healing schemes aid the material designers not only to design a complex smart structure but also to optimize their designed material systems. While material plasticity and damage have been investigated in the literature over the past century, the healing concept is a new topic in solid mechanics and quite a few works have addressed this challenging topic. In the case of the damage mechanics theoretical developments, one may mention the pioneering works by Lemaitre, Murakami, Chaboche, and Voyiadjis (Murakami 1988; Lemaitre and Chaboche 1990; Voyiadjis and Kattan 2006b). While these researchers had implicitly discussed the possibility of the damage removal, however, the theoretical development for the healing process has not been accomplished until most recently. Voyiadjis, Shojaei, Li, and Kattan published papers in

Int. J. Damage Mechanics and J. of Proceeding of the Royal Society A on the coupled damage-healing process (Voyiadjis et al. 2012a, c). The proposed formulation in these two works together with another publication by these authors on the thermodynamics-based modeling of the damage-healing process (Voyiadjis et al. 2011) closes the gap between the continuum damage mechanics (CDM) and the healing process in which the continuum damage-healing mechanics (CDHM) framework is thoroughly formulated and introduced.

In the newly developed CDHM, the healing process is controlled through scalar or tensorial healing parameters, and it is well suited for the finite element analysis (FEA) implementations where user-defined coding can be utilized to introduce the CDHM formulation. The thermodynamic framework provides a physically consistent approach to formulate the constitutive behavior of the material systems (Chaboche 1989, 1991, 1993, 1997, 2008; Lemaitre and Chaboche 1990; Hansen and Schreyer 1994; Shojaei et al. 2010; Voyiadjis et al. 2011; Voyiadjis et al. 2012c). In the case of self-healing materials, the governing deformation mechanisms for the damage and healing processes are formulated within the thermodynamic consistent coupled elasto-plastic-damage-healing framework by Voyiadjis et al. (2011, 2012a). It is worth noting that FEA implementation of the developed models aims to numerically investigate the complex damage and healing processes in self-healing materials and will reduce the burden of experimental investigations.

In this chapter, first, a brief overview on existing self-healing schemes is presented in section "A Brief Overview on Existing Self-Healing Schemes." In section "Continuum-Damage-Healing Mechanics," the CDHM is elaborated, and in section "Results and Discussion," the experimental and simulation results are shown.

A Brief Overview on Existing Self-Healing Schemes

The self-healing smart materials have been introduced into the research arena a few years ago and they have been already deployed into many practical applications. These materials are designed to heal micro- and macroscale damages (Miao et al. 1995; Adam 1999; Simpson et al. 2000; White et al. 2001; Pang and Bond 2005; Trask and Bond 2006; Plaisted and Nemat-Nasser 2007; Toohey et al. 2007; Williams et al. 2007; Kirkby et al. 2008; Varley and van der Zwaag 2008; Beiermann et al. 2009; Kirkby et al. 2009; Li and Uppu 2010; Li and Shojaei 2012). The currently developed healing schemes can be in general categorized in three classes. The first class is known as *embedded liquid healing agent* which was introduced by White's pioneering work. In this class of healing scheme, the healing agent is stored in microcapsules (White et al. 2001) or it is delivered to the damaged sites through microvascular network system (Toohey et al. 2007). In this system, the healing is initiated once the wall of the microcapsule or microvascular system is cracked due to the damage mechanism. Usually, these systems are designed in such a way that the released healing agent contacts with the embedded catalyst inside the

matrix material where the solidification of the healing agent in the cracked area closes the damage. The applicability of these self-healing systems is limited due to two main drawbacks: (a) clogging of the vascular network, or shortage of microencapsulated healing agent, after the first round of healing limits the repeatability of the healing process and (b) the existence of an uncured resin inside a composite material will undermine the final product material properties. Particularly, the performance of these systems in healing macroscale damages is still an obstacle that must be overcome. This kind of healing methodology is referred to as a coupled damage-healing system which indicates that both of the damage and healing processes are active concurrently in the system (Shojaei et al. 2013; Voyiadjis et al. 2012a). The second type of healing scheme utilizes the so-called solid healing agent. These systems utilize some dispersed solid phase, e.g., thermoplastic particles (TPs), as the healing agent in which external triggering, e.g., heating, is required to activate the solid healing agent, e.g., melting in the case of TPs, and diffuse to the crack surfaces (Zako and Takano 1999; Li and Uppu 2010; Nji and Li 2010b; Li and Shojaei 2012). The third class is called thermally reversible covalent bonds where the broken chemical bonds are rejoinable (Liu and Chen 2007; Plaisted and Nemat-Nasser 2007; Varley and van der Zwaag 2008). In the last two methods, external triggering is utilized to activate the healing mechanism. These systems are referred to as decoupled damage-healing systems where the damage and healing processes are active separately as discussed by Voyiadjis et al. (2011).

There are many research groups currently working on theoretical and experimental developments of self-healing materials. Among the newly developed healing schemes, one may mention the recently established bio-inspired two-step close-then-heal (CTH) mechanism which has been developed by Li and coworkers at Louisiana State University, USA. This novel scheme provides molecular level healing for structural- and micro-length scale damages in polymeric composite materials and shows repeatability and efficiency properties for the healing process in which the healing can be implemented in a timely manner (Shojaei et al. 2013; Li and John 2008; Li and Nettles 2010; Li and Uppu 2010; Nji and Li 2010a, b; Voyiadjis et al. 2011; Li and Shojaei 2012). In these systems, the confined shape recovery of shape memory polymer (SMP) provides the essential force for crack closure and the embedded TPs are molten to heal the cracks. Molecular healing level is obtained through diffusion of TPs into the crack surfaces. The SMP sample is programmed under specific programming steps before the structural level damage is induced into the sample. These three-step thermomechanical programming for the SMP are as follows: (i) compression at a temperature above the glass transition temperature, (ii) cooling while maintaining the compressive strain constant, and (iii) removal of the applied stress at a temperature well below the glass transition temperature. This process is shown in Fig. 1 (Voyiadjis et al. 2012b). While the bulk SMP, as the matrix in self-healing systems, requires the mentioned specific programming steps, the SMP fibers are programmed with simple room temperature tension. Most recently, Li and Shojaei proposed the use of SMP fibers instead of making the whole matrix from SMP. It is shown by them that the cold-

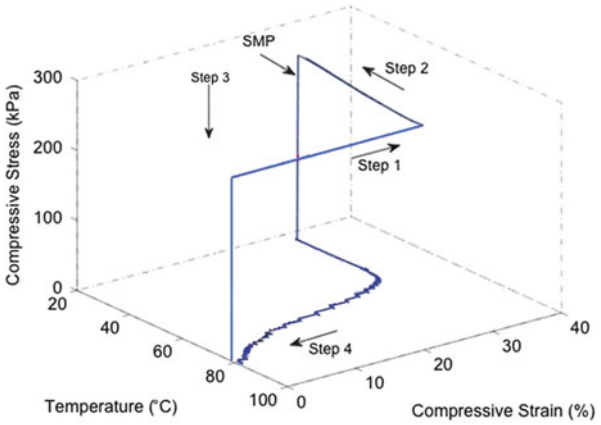


Fig. 1 Thermomechanical cycle for an SMP-based self-healing system (Picture courtesy of Voyiadjis et al. (2012a))

drawn SMP fibers are work-hardened and they show excellent strength and shape recovery property when they are embedded inside thermosetting polymeric matrix (Li and Shojaei 2012; Shojaei et al. 2013). One may also mention works by Hayes and coworkers at the University of Sheffield for developing self-healing fiber-reinforced composite materials (Hayes et al. 2007a, b) and works by Bond and Trask in University of Bristol on bio-inspired healing schemes (Bond et al. 2008).

Due to the high-performance nature of the CTH healing scheme, its concept is more elaborated in detail here. Figure 2 represents an overview on the CTH mechanism using SMP matrix as the closing agent (Voyiadjis et al. 2012a). Figure 2a shows the schematic of a damaged structure at room temperature in which a three-point single-edge notch bend test might be utilized to introduce a structural level damage into the system. The green spheres are schematics of the dispersed thermoplastic particles (TPs) in the SMP matrix. Figure 2b shows the heated damaged configuration from room temperature to an elevated temperature above the SMP glass transition temperature T_g, where the Shape Memory property of the SMP is activated. The confined boundary conditions cause the crack to be sealed. Figure 2c illustrates the state of the sample when the temperature reaches above the melting point of TPs in which the molten TP molecules diffuse into the crack surfaces and fills the microscale gap between closed crack surfaces. Figure 2d shows the state of the SMP after the cooling down process where the crack is fully healed in this configuration (Voyiadjis et al. 2012a). A schematic representation of the confined shape recovery of SMP is depicted in Fig. 2e where the configuration of rigid rod and sample is shown inside the rigid cylindrical fixture (Voyiadjis et al. 2012a). The confined recovery starts by heating the setup to a temperature above T_g while the rigid fixture and rod retain the overall shape of the sample. Consequently, the activated SM property will cause the sample to expand, and the external confinement fills the open internal cracked spaces and the cracks within the sample are closed. As shown in Fig. 3, it is experimentally confirmed that the TP in this step diffuses into the SMP matrix and provides the desired molecular level healing (Nji and Li 2010b; Li et al. 2012). Experimental observations confirm the efficiency of

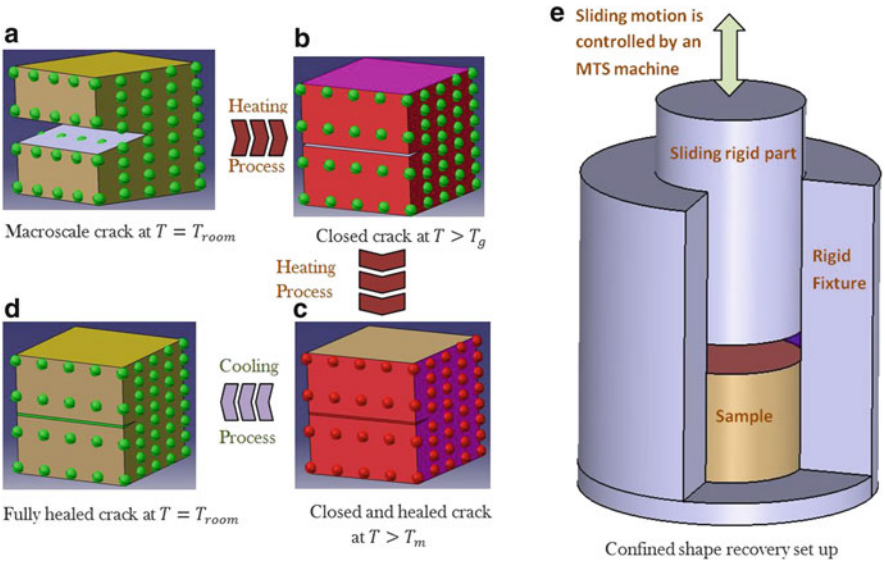


Fig. 2 Schematic representation of the healing process in an SMP-based self-healing system with embedded TPs, (**a**) Damaged state at room temperature. (**b**) Partially heated sample above T_g, where SMP remembers its original shape, and under confined boundary condition, the macroscale crack is sealed. (**c**) Upon further heating to a temperature above the melting temperature of the thermoplastic particles (T_m), the crack is filled in by the molten TP and the TP molecules diffuse into the SMP matrix. (**d**) The final healed configuration at room temperature, where the molecular level of healing is reached through physical entanglement. (**e**) A schematic representation of confined shape recovery setup (the cylinder is sliced here to show the inside configuration) (Picture is after Voyiadjis et al. (2012a))

this healing mechanism in which molecular entanglements are gained after healing and the mechanical properties of the self-healing system are fully recovered (Nji and Li 2010b; Voyiadjis et al. 2011). Figure 3 shows scanning electron microscopy and transmission electron microscopy results for a macroscale crack (Voyiadjis et al. 2012a), produced by a three-point bending test (ASTM-D5045 2007). Figure 3a shows SEM image of a macroscale crack in the self-healing system, where the openings of the crack, crack path, and crack tip are highlighted on the SEM image (Voyiadjis et al. 2012a). Figure 3b shows the healed configuration where the crack path is almost disappeared and the TEM image shows the interface between the TP and the SMP matrix after diffusion (Voyiadjis et al. 2012a).

Recently, Li and Shojaei proposed a new healing scheme for the polymeric material systems in which the programmed SMP fibers are utilized to close the cracks, while the same healing strategy, i.e., diffusion of the molten TPs, provides the molecular level of healing (Shojaei et al. 2013; Li and Shojaei 2012). This new healing scheme compensates for the high cost associated with the required high volume of SMP as matrix, and also the external confinement restriction is removed (Shojaei et al. 2013). The schematic of this healing strategy is depicted in Fig. 4 (Li and Shojaei 2012).

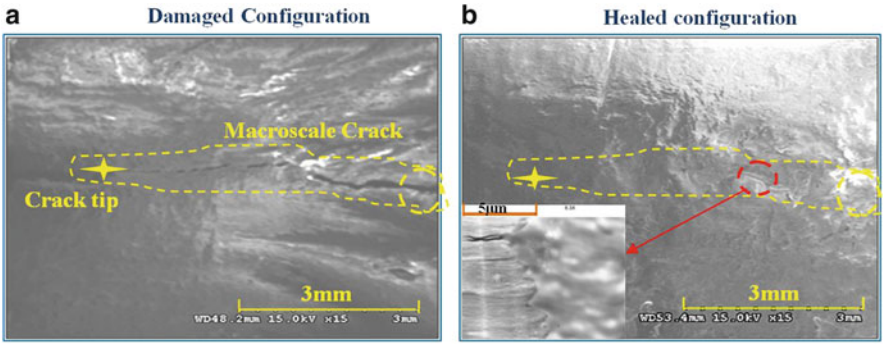


Fig. 3 (**a**) Structural-level damage. (**b**) Healed configuration (Picture courtesy of Voyiadjis et al. (2012a))

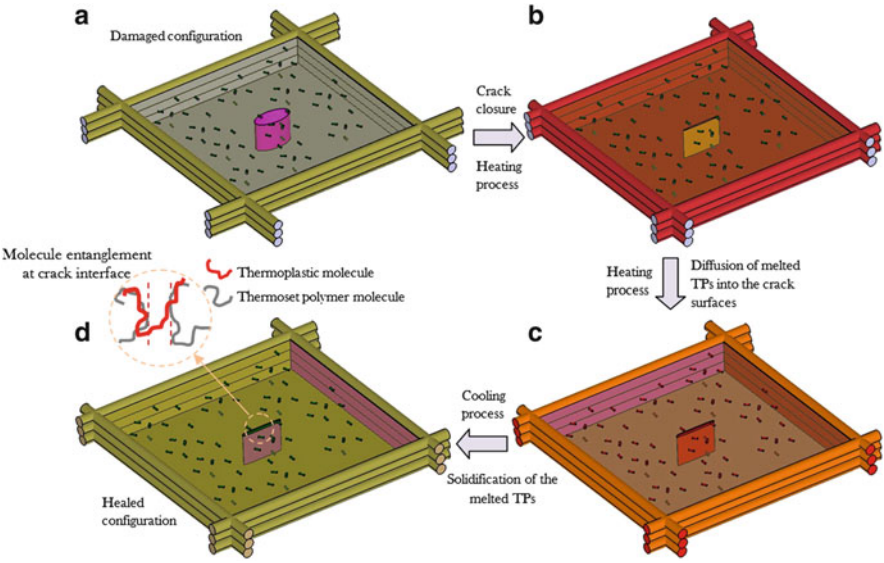


Fig. 4 Schematic of the bio-inspired healing process of the proposed composite (3-D view) (**a**) a unit cell (bay) of SMP grid (ribs and z-pins) stiffened conventional thermoset polymer dispersed with thermoplastic particles. A macroscopic crack is introduced in the unit cell, which can be identified by visual or nondestructive inspections ($T < T_g$); (**b**) crack closure process through recovery of the SMP fiber ribs and z-pins, when local heating is applied ($T > T_g$); (**c**) further temperature rising melts the thermoplastic particles which flow into the crack by capillary force and diffuse into the fractured surface by concentration gradient ($T > T_m$); (**d**) cooling down to below the glass transition temperature, solid wedge can be formed and molecular entanglement can be established ($T < T_g$). Magnified view shows the molecule entanglement at the crack interface (Picture courtesy of Li and Shojaei (2012))

Continuum Damage-Healing Mechanics

In this section, the concept of effective and real configurations in the case of damage-healing is elaborated and new scalar and tensorial healing variables are formulated within the CDM framework. In section "Scalar Damage-Healing Variables," the scalar damage variables are introduced, and in section "Generalized Damage-Healing Variables," the anisotropic damage-healing problem is evaluated in which tensorial representation of the healing variables is proposed.

Scalar Damage-Healing Variables

The major advantage of continuum damage mechanics is that it utilizes a continuous damage description in the damage process zone. Kachanov (1958) and Rabotnov (1963) were the first ones to introduce a damage variable to relate the density of defects to the overall material degradation (Voyiadjis et al. 2011, 2012a, b; Shojaei et al. 2013; Voyiadjis and Kattan 2006b, 2009, 2010). Many damage variables have been introduced later in the literature to overcome measurement difficulty of the damage densities, such as calibrating the elastic constant changes due to the damage mechanism (Kachanov 1958; Voyiadjis and Kattan 2009, 2010) or using the surface energy descriptions for generating new microsurfaces during the damaging process and correlating it to the material deterioration (Hansen and Schreyer 1994). Also, many direct and indirect damage measurement processes are introduced in the literature (Lemaitre and Dufailly 1987).

Making use of the well-developed damage formulation in the literature, the lack of the theoretical background for the healing process is addressed by Voyiadjis, Shojaei, Li, and Kattan in Voyiadjis et al. (2011, 2012a, c). In their work, a scalar healing variable for the case of isotropic healing process is proposed and a direct method to capture the healing effect is established (Voyiadjis et al. 2012c). The developed scalar healing variable measures the changes in the density of defects during the healing process and then they proposed a scalar healing variable based on the changes of the elastic moduli (Voyiadjis et al. 2012c). For this purpose, the concept of the effective configuration, in the framework of CDM, is modified to capture the effect of healing. The proposed framework provides a consistent approach for FEA implementation of self-healing and it has been adopted by many other researchers in the field of solid mechanics to investigate the healing effect.

As discussed by Voyiadjis et al. (2012c), one may consider the CDM framework to show the initial undeformed and undamaged configuration of the body by C_0 and the damaged and deformed configuration by C. The effective configuration is a fictitious state where all damages including microcracks and voids have been removed from the deformed body and it is shown by $\overline{C}$. In Fig. 5, these configurations are depicted schematically. For the case of isotropic damage and using the

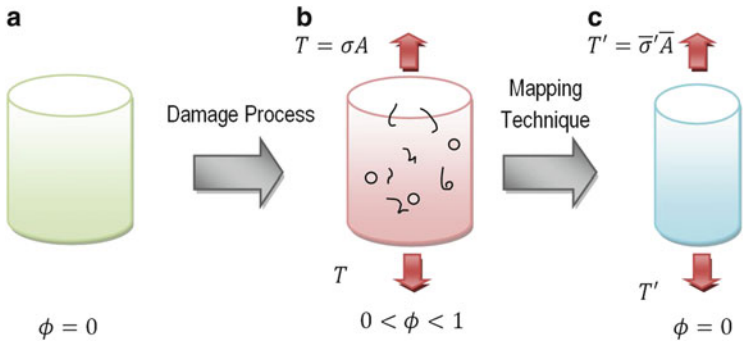


Fig. 5 (**a**) Initial undeformed and undamaged configuration C_0. (**b**) Damaged and deformed configuration, C. (**c**) Effective fictitious undamaged and deformed configuration, $\overline{C}$ (Picture courtesy of Voyiadjis et al. (2012c))

basic concept of the effective state, the damage variable is defined as a scalar in the following manner (Kachanov 1958; Voyiadjis et al. 2012c):

$$\phi = \frac{A - \overline{A}}{A} \tag{1}$$

To define the scalar healing variable, three new configurations are defined subsequently. The fictitious state of fully damaged and deformed configuration C^d, which is obtained by subtracting the fictitious undamaged and deformed cross section $\overline{A}$ in the configuration $\overline{C}$ from the total cross section A in the configuration C, is as follows (Voyiadjis et al. 2012c):

$$A^d = A - \overline{A} = A - A(1 - \phi) = A\phi \tag{2}$$

where A^d is the total damaged cross section (removed area due to damage).

The healing process is considered to be active only on the pure damaged configuration, i.e., A^d, and then the fictitious healed and deformed configuration C^h is defined by removing some portion of damages from A^d as shown in Fig. 6. The scalar healing variable is then defined as follows (Voyiadjis et al. 2012c):

$$h = \frac{A^d - A^h}{A^d}, 0 < h < 1 \tag{3}$$

where A^h is the healed portion of the cross section A^d (Voyiadjis et al. 2012c). The case of $h = 1$ corresponds to zero percent healing of the damaged area ($A^h = 0$), and $h = 0$ relates to 100 % healing of the damaged area ($A^h = A^d$).

Finally, the effective fictitious fully healed and deformed configuration $\overline{C}^h$ is obtained by removing all the remaining damages from C^h, as shown in Fig. 7.

Based on the CDHM approach, the damaged area does not sustain any further loading. However, after healing, the healed cross section A^h can carry load T'' as

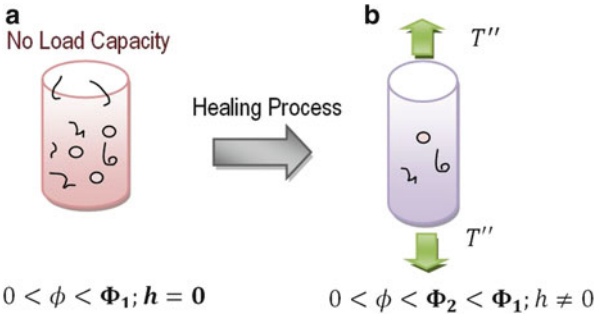


Fig. 6 (**a**) Fictitious total removed area due to damage, C^d. (**b**) Fictitious healed and damaged configuration, C^h (Picture courtesy of Voyiadjis et al. (2012c))

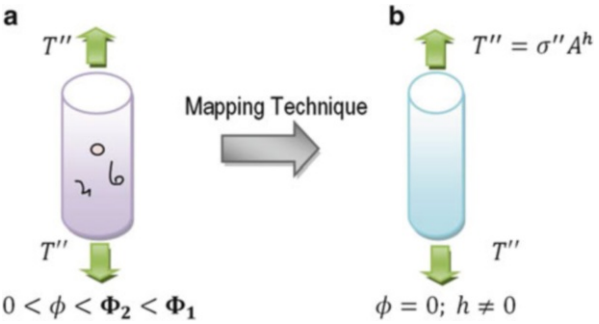


Fig. 7 (**a**) Fictitious healed and deformed configuration, C^h. (**b**) Effective fictitious fully healed and deformed configuration, $\overline{C}^h$ (Picture courtesy of Voyiadjis et al. (2012c))

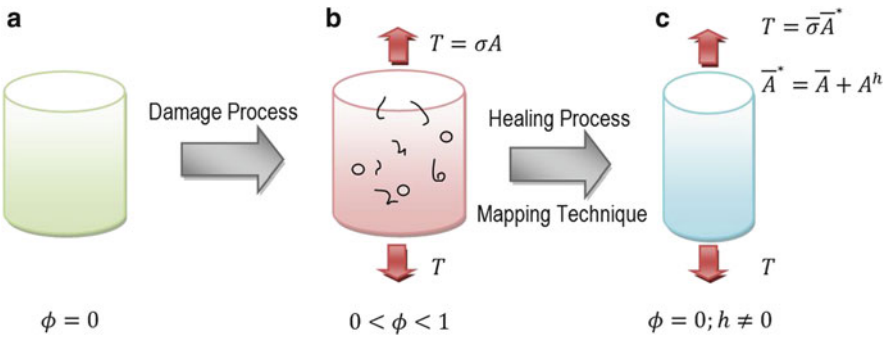


Fig. 8 (**a**) Virgin material. (**b**) Damaged and deformed configuration, C. (**c**) Hypothetical healed configuration, $\overline{C}^{healed}$ (Picture courtesy of Voyiadjis et al. (2012c))

shown in Figs. 6 and 7. The cross section of the hypothetical healed configuration $\overline{C}^{healed}$ is assumed to be the summation of the cross section of the effective fictitious healed and deformed configuration $\overline{C}^h$ and the cross section of the effective fictitious undamaged and deformed configuration $\overline{C}$ as shown if Fig. 8.

The necessary transformation equation for stresses between $\overline{C}^{healed}$ and C is derived by applying the equilibrium on the healed and damaged states as follows (Voyiadjis et al. 2012c):

$$\overline{\sigma} = \frac{\sigma}{((1-\phi)+\phi(1-h))} \tag{4}$$

In order to compute the transformation relations between the damaged-healed elastic modulus, E(ϕ, h), and the effective undamaged elastic modulus, $\overline{E}$, certain assumptions regarding the strains in the two configurations are required (Voyiadjis et al. 2012c). Accordingly, one of the following two hypotheses is followed within CDHM (Voyiadjis et al. 2012a, c):

1. *Hypothesis of elastic strain equivalence*: in this case, the strains in damaged and fictitious configurations are assumed to be the same, i.e., $\epsilon = \overline{\epsilon}$.
2. *Hypothesis of elastic strain energy equivalence*: in this case, the elastic strain energy in both configurations is assumed to be equivalent, i.e., $U = \overline{U}$.

While both of these hypotheses have been used by researchers in the field of CDM, it is believed that the hypothesis of elastic energy equivalence is more general since it evaluates the response of the system based on the energy balance. In the following, for completeness, both hypotheses are used to derive the transformation relations.

Hypothesis of Elastic Strain Equivalence for Damage Variable ϕ and Healing Variable h

The transformation laws between the *effective* and the *damaged-healed* configurations based on the hypothesis of elastic strain equivalence are established through

$$\epsilon = \overline{\epsilon} \tag{5}$$

Using the elastic constitutive equation for both the effective configuration, $\overline{C}^{healed}$, and the damaged-healed configuration, C, along with Eq. 4, the transformation equation for elasticity modulus between these configurations is computed as follows:

$$E^h(\phi,h) = \overline{E}((1-\phi)+\phi(1-h)) \tag{6}$$

where $E^h(\phi, h)$ indicates the damaged and healed elastic modulus and $\overline{E}$ is the modulus of elasticity for the virgin material. When there is no healing process $h = 1$, Eq. 6 results in the classical relation: $E^h(\phi, h = 1) = \overline{E}(1-\phi)$ (Kachanov 1958; Lemaitre and Chaboche 1990; Voyiadjis and Kattan 2006b).

Hypothesis of Elastic Energy Equivalence for Damage Variable ϕ and Healing Variable h

As an alternative approach to deriving the transformation rule of Eq. 6, one may use the hypothesis of elastic strain energy equivalence between the effective configuration, $\overline{C}^{\text{healed}}$, and the real damaged-healed configuration, C, in order to derive the elastic modulus relation between these two states as follows:

$$U = \frac{1}{2E^{h}(\phi,h)}\sigma^{2} = \frac{1}{2\overline{E}}\overline{\sigma}^{2} \tag{7}$$

Substituting $\overline{\sigma}$ from Eq. 4 into Eq. 7 results in the relation between the transformation relation for the elastic modulus between the two configurations:

$$E^{h}(\phi,h) = \overline{E}((1-\phi)+\phi(1-h))^{2} \tag{8}$$

The state of 100 % healing yields $E^{h}(\phi, h=0) = \overline{E}$, and the corresponding elastic modulus is fully recovered after the healing process. In the case when there is no healing process, Eq. 8 results in $E^{h}(\phi, h=1) = \overline{E}(1-\phi)^{2}$ which shows consistency with the classical result obtained from the continuum damage mechanics (Kachanov 1958; Lemaitre and Chaboche 1990; Voyiadjis and Kattan 2006b, 2009). Finally, there are two options to choose from for the transformation laws: those are Eq. 6 resulting from the hypothesis of elastic strain equivalence and Eq. 8 which is a direct result from utilizing the hypothesis of elastic energy equivalence.

Elastic Modulus Based Scalar Damage-Healing Variable

The damage measurement based on the area reduction is hard to obtain and it is not practical in many real applications. The measurement of the cross section area reduction due to damage or cross section area increase due to healing involves precise measurement of microcracks and microvoids which is hard to implement and requires advanced mathematical and mapping techniques. In order to facilitate the damage and healing measurement, an indirect measurement method for the damage and healing is adopted in this work (Voyiadjis et al. 2012c; Lemaitre and Dufailly 1987). This new healing variable for the uniaxial case of healing is introduced based on the elastic modulus changes. The scalar damage measure is introduced as (Voyiadjis et al. 2012c)

$$l = \frac{\overline{E} - E^{d}(l)}{E^{d}(l)} \tag{9}$$

The scalar healing variable h' is defined to measure the change in the elastic modulus during the healing process, as follows (Voyiadjis et al. 2012c):

$$h' = \frac{E^{h}(l,h') - E^{d}(l)}{E^{d}(l)} \tag{10}$$

where $\overline{E}$ is the elastic modulus for a virgin material, $E^d(l)$ is the damaged elastic modulus in the damaged and deformed configuration, C, and $E^h(l, h')$ is the elastic modulus in the healed and deformed configuration, C^h. The relation between the three moduli is as follows (Voyiadjis et al. 2012c):

$$E^d(l) \leq E^h(l, h') \leq \overline{E} \tag{11}$$

To obtain the transformation equation between the effective and healed elastic modulus, one may substitute $E^d(l)$ from Eq. 9 into Eq. 10 as follows (Voyiadjis et al. 2012c):

$$E^h(l, h') = \frac{(1 + h')\overline{E}}{(1 + l)} \tag{12}$$

Equation 12 indicates $h' = 0$ corresponds to **zero percent healing** and results in $E^h(l, h' = 0) = E^d(l) = \overline{E}/(1 + l)$ which is consistent with the result of pure damage without healing (Voyiadjis et al. 2012c). By setting $E^h(l, h' = l) = \overline{E}$, the upper bound for h' is obtained as $h' = l$. In the following, the transformation laws based on this new healing variable are derived for both cases of elastic strain equivalence and elastic energy equivalence.

Hypothesis of Elastic Strain Equivalence for Damage Variable l and Healing Variable h′

The transformation laws based on the hypothesis of elastic strain equivalence are utilized to obtain the correlation between the two configurations of $\overline{C}^{healed}$ and C. Using Eq. 5 results in the following transformation law (Voyiadjis et al. 2012c):

$$\overline{\sigma} = \sigma \frac{(1 + l)}{(1 + h')} \tag{13}$$

The correlation between damage variables l and ϕ and the healing variables h and h' is given by (Voyiadjis et al. 2012c)

$$\phi = \frac{1}{h}\left(1 - \frac{1 + h'}{1 + l}\right) \tag{14}$$

Hypothesis of Elastic Energy Equivalence

The stress transformation between the damaged state, C, and the hypothetical healed states, $\overline{C}^{healed}$, is obtained based on Eq. 7 as follows (Voyiadjis et al. 2012c):

$$\overline{\sigma} = \sigma \sqrt{\frac{1 + l}{h' + 1}} \tag{15}$$

where $\overline{\sigma}$ is the healed stress in the healed configuration and stress σ is in the damaged configuration.

The relation between the damage variables l and ϕ and the healing variables h and h′ in the case of elastic energy equivalence is obtained through substituting $\overline{E}$ from Eq. 12 into Eq. 8 (Voyiadjis et al. 2012c):

$$\phi = \frac{1}{h}\left(1 - \sqrt{\frac{(1+h')}{(1+l)}}\right) \tag{16}$$

The relationship between the classical damage variables ϕ and l is shown in Eq. 16 with new healing variables h and h′, when the hypothesis of elastic strain energy is utilized.

If the healing is eliminated from the system, it means that $h' = 0$ and $h = 1$ and Eq. 16 reduces to

$$\phi = \frac{\sqrt{(1+l)} - 1}{\sqrt{(1+l)}} \tag{17}$$

which is consistent with the result published before on this topic (Voyiadjis and Kattan 2009). Substituting Eq. 17 into Eq. 16 results in

$$\sqrt{(1+l)} - 1 = \frac{\sqrt{1+h'} - 1}{1-h} \tag{18}$$

If one uses ϕ instead of l, then the above relation becomes

$$\frac{\phi}{1-\phi} = \frac{\sqrt{1+h'} - 1}{1-h} \tag{19}$$

Note that Eqs. 18 and 19 are special cases of Eq. 16.

Generalized Damage-Healing Variables

The material degradation based on the damage density and calibrating the area reduction due to the microscale damages formation was formerly formulated by Kachanov (1958). This concept was later generalized to a multiaxial anisotropic case by Murakami (1988). Figure 9 represents the CDM approach in removing damaged material from the real configuration (Voyiadjis et al. 2012b, c). Figure 5a depicts the real state of the damaged material which is assumed to be decomposed into a fictitious effective configuration in Fig. 9b, which carries the load, and a fictitious fully damaged configuration, Fig. 9c, which cannot sustain load. The damage variable tensor ϕ_{ij} represents the transformation between the real

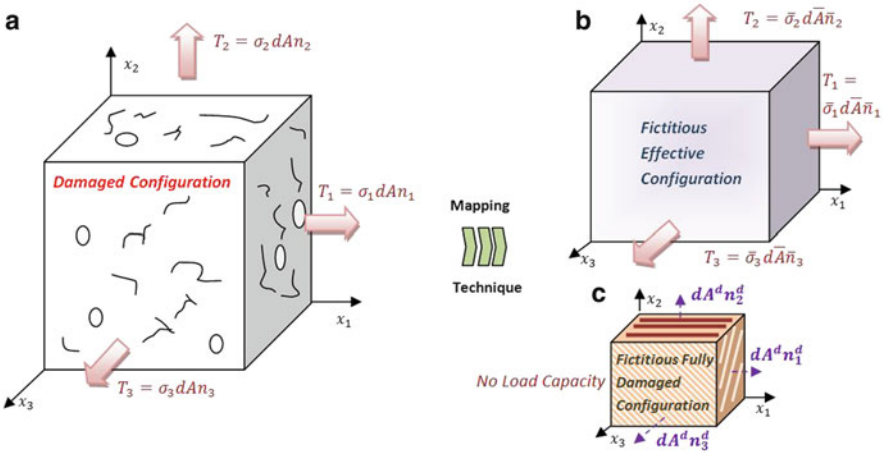


Fig. 9 Schematic representation of (**a**) real damaged configuration, (**b**) fictitious effective configuration, (**c**) fictitious damaged state (Voyiadjis et al. 2012b)

damaged area vector dAn_i (Fig. 9a) and effective fictitious area vector $d\overline{A}\overline{n}_i$ (Fig. 9b) as proposed by Murakami (1988):

$$\phi_{ij} n_i = \frac{\left(dAn_j - d\overline{A}\overline{n}_j\right)}{dA}; 0 \leq \left(\phi_{ij}\phi_{ij}\right)^{1/2} \leq 1 \tag{20}$$

The pure damaged area vector $dA^d n_i^d$ (Fig. 9c) is then obtained as a function of the damage variable tensor, ϕ_{ij}, and the area vector, dAn_j, in the real damaged configuration. Equation 20 is rearranged to obtain

$$dA^d n_i^d = dAn_i - d\overline{A}\overline{n}_i = \phi_{ij} n_j dA \tag{21}$$

Voyiadjis, Shojaei, Li, and Kattan utilized the underlying mechanism of the healing process to find physically consistent healing variables to calibrate the healing mechanism (Voyiadjis et al. 2012b, c). During the healing process, some of the microscale damages are healed and then the effective area, which carries the load, is increased. To represent this phenomenon, it is assumed that the fictitious fully damaged configuration (Fig. 9c) undergoes the healing process (Voyiadjis et al. 2012b). This anisotropic healing problem is shown in Fig. 10, where Fig. 10a shows the pure damaged sate without load-carrying capacity. Figure 10b shows the healed configuration and Fig. 10c shows the fictitious effective healed configuration. In Fig. 10d, the remaining damages after accomplishing the healing process are shown which indicates that the healing process can be partially effective in removing the damages.

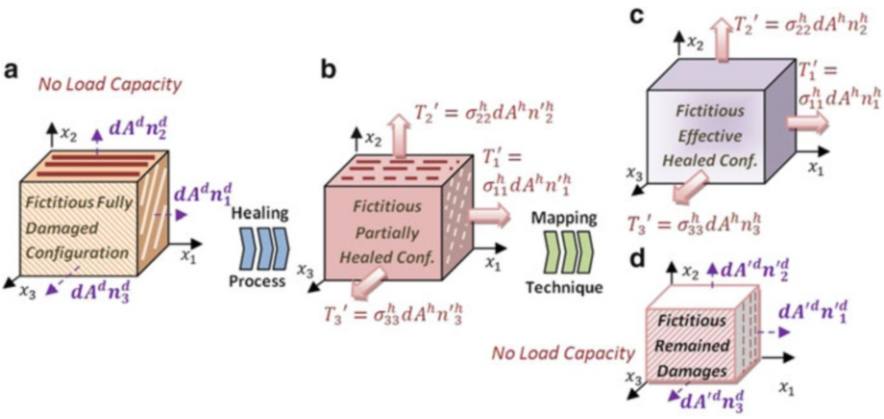


Fig. 10 (**a**) Fictitious damaged state, (**b**) fictitious healed state, (**c**) fictitious effective healed configuration, and (**d**) fictitious remained damaged state (Voyiadjis et al. 2012b)

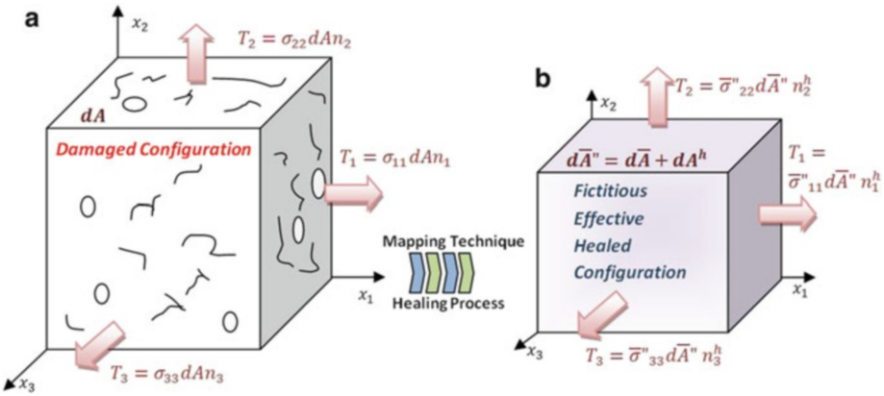


Fig. 11 (**a**) Damaged configuration, (**b**) fictitious effective configuration after healing process (Voyiadjis et al. 2012b)

According to the presented physical description of the healing process, Voyiadjis, Shojaei, and Li (Shojaei et al. 2013; Voyiadjis et al. 2012b) proposed a second-rank anisotropic healing variable tensor h_{ij} as follows:

$$h_{ij}n_i^d = \frac{\phi_{jk}dAn_k - dA^h n_j^h}{dA^d}; 0 \le \left(h_{ij}h_{ij}\right)^{1/2} \le 1 \tag{22}$$

where healing variable h_{ij} captures the transformation between the real damaged area vector dAn_i (Fig. 11a) and the fictitious healed area vector $d\overline{A}''n_i^h$ (Fig. 11b; Voyiadjis et al. 2012b). Figure 11 shows the overall mapping procedure between the real damaged and fictitious healed effective configurations (Voyiadjis et al. 2012b).

As indicated in the case of scalar damage-healing measurement techniques, the idea of area reduction due to the damage and area increasing due to the healing process is referred to as direct measurement and it can be accomplished by advanced measurement techniques (Voyiadjis et al. 2012b, c). Due to the complex mapping techniques required in measuring the damage and healing based on this defect density approach, indirect measurement methods are preferable (Voyiadjis et al. 2012b, c). In these methods, the damage and healing processes are calibrated based on elastic modulus changes (Lemaitre and Dufailly 1987). The fourth-order anisotropic damage variable tensor κ_{ijkl} based on the elastic modulus change is given as follows (Voyiadjis et al. 2012b; Lemaitre and Dufailly 1987):

$$\begin{aligned} \kappa^{(1)}_{ijkl} &= \left(\overline{E}_{ijmn} - E^{d}_{ijmn}\right)\overline{E}^{-1}_{mnkl} \\ \kappa^{(2)}_{ijkl} &= \overline{E}^{-1}_{ijmn}\left(\overline{E}_{mnkl} - E^{d}_{mnkl}\right) \end{aligned} \tag{23}$$

where the superscripts "(1)" and "(2)" indicate the two different mathematical tensorial expressions for the damage tensor. These expressions normalize the damage tensor with respect to the inverse of the undamaged elasticity tensor, $\overline{E}_{ijkl}$, and E^{d}_{ijkl} is the damaged elastic modulus (Voyiadjis et al. 2012b).

In order to measure the healing indirectly, a fourth-rank healing variable tensor h'_{ijkl} is introduced by Voyiadjis, Shojaei, and Li to measure the elastic modulus changes after accomplishing the healing process as follows (Voyiadjis et al. 2012b):

$$\begin{aligned} h'^{(1)}_{ijkl} &= \left(E^{h}_{ijmn} - E^{d}_{ijmn}\right)E^{d^{-1}}_{mnkl} \\ h'^{(2)}_{ijkl} &= E^{d^{-1}}_{ijmn}\left(E^{h}_{mnkl} - E^{d}_{mnkl}\right) \end{aligned} \tag{24}$$

where E^{h}_{ijkl} is the elastic modulus of the healed material in which $h'_{ijkl} = 0_{ijkl}$ represents no healing case and 0_{ijkl} is a fourth-order tensor with zero components. Substituting E^{d}_{ijkl} from Eq. 24 into Eq. 23 results in the following expression for the healed elastic modulus E^{h}_{ijkl} (Voyiadjis et al. 2012b):

$$\begin{aligned} E^{h}_{ijmn} &= \overline{E}_{ijmn} + \overline{E}_{klmn}\left(h'^{(1)}_{ijkl} - \kappa^{(1)}_{ijkl} - \kappa^{(1)}_{pqkl}h'^{(1)}_{ijpq}\right) \\ E^{h}_{ijmn} &= \overline{E}_{ijmn} + \overline{E}_{ijpq}\left(h'^{(2)}_{pqmn} - \kappa^{(2)}_{pqmn} - \kappa^{(2)}_{pqkl}h'^{(2)}_{klmn}\right) \end{aligned} \tag{25}$$

As shown in Fig. 11, the Cauchy stress tensor, σ_{ij}, represents the stress state in the real damaged configuration (Fig. 11a), and the effective stress tensor $\overline{\sigma}_{ij}$ shows the state of stress in the effective healed configuration (Fig. 11b). In general, two approaches are available to derive the relation between these two stress tensors. One may equate the strains in both configurations: $\epsilon_{ij} = \overline{\epsilon}_{ij}$, and the second method is the so-called equivalent elastic energy (EEE) method which is more reliable as it is energy based. In general, the EEE method relies on energy equivalence between the real and effective configurations which incorporates more physical phenomena in comparison to a simple equivalence of strains. The EEE approach has been extensively utilized in the literature to derive the relations between effective and

real configurations (Lemaitre 1985; Yazdani and Schreyer 1990; Hansen and Schreyer 1994; Voyiadjis et al. 2012b, c; Shojaei et al. 2013). Making use of the second approach, which is stated between real damaged and fictitious healed configurations, results in

$$U = \frac{1}{2} E^{h^{-1}}_{ijkl} \sigma_{ij} \sigma_{kl} = \frac{1}{2} \overline{E}^{-1}_{ijkl} \overline{\sigma}_{ij} \overline{\sigma}_{kl} \tag{26}$$

A fourth-rank damage-healing transformation tensor, Q_{ijkl}, is defined to represent the transformation between the Cauchy stress tensors, σ_{ij}, in the real state and Cauchy stress tensor, $\overline{\sigma}_{ij}$, in the effective configuration. It is expressed by the following relationship (Voyiadjis et al. 2012b):

$$\overline{\sigma}_{ij} = Q_{ijkl} \sigma_{kl} \tag{27}$$

Substituting $\overline{\sigma}_{ij}$ from Eq. 27 and E^{h}_{ijkl} from Eq. 25 into Eq. 26 provides the relationships between h'_{ijkl}, E^{d}_{ijkl}, and Q_{ijkl} as follows (Voyiadjis et al. 2012b):

$$\begin{aligned} Q_{ijuw} Q_{klpq} &= \left(h'^{(1)}_{mnij} E^{d}_{klmn} + E^{d}_{klij} \right) \overline{E}^{-1}_{uwpq} \\ Q_{ijuw} Q_{klpq} &= \left(E^{d}_{mnij} h'^{(2)}_{klmn} + E^{d}_{klij} \right) \overline{E}^{-1}_{uwpq} \end{aligned} \tag{28}$$

Applying the equilibrium between the real and effective configurations results in the relation between stresses in these two states as shown below (Voyiadjis et al. 2012b):

$$\sigma_{ij} = \left(M^{-1}_{ijkl} + \left(I_{ijmn} - M^{-1}_{ijmn} \right) H^{-1}_{mnkl} \right) \overline{\sigma}_{kl} \tag{29}$$

where I_{ijkl} is the fourth-order identity tensor. The fourth-rank damage effect tensor, M_{ijkl}, and healing effect tensor, H_{ijkl}, are defined in the following. One may find $H_{ijkl} = 0_{ijkl}$ (where 0_{ijkl} is the fourth-rank zero tensor) that indicates the unhealed configuration in which $\overline{\sigma}_{ij} = M_{ijkl} \sigma_{kl}$. The case of $H_{ijkl} = I_{ijkl}$ represents the fully healed state where $\overline{\sigma}_{ij} = \sigma_{ij}$ (Voyiadjis et al. 2012b). Consequently, the following tensorial representations for M_{ijkl} and H_{ijkl} are proposed (Voyiadjis et al. 2012b):

$$\begin{aligned} M_{ijkl} &= \left[\left(I_{ij} - \phi_{ij} \right) \left(I_{kl} - \phi_{kl} \right) \right]^{-1/2} \\ H_{ijkl} &= \left[h_{ij} h_{kl} \right]^{-1/2} \end{aligned} \tag{30}$$

Comparing Eqs. 29 and 27, one may find (Voyiadjis et al. 2012b)

$$Q_{ijkl} = \left(M^{-1}_{ijkl} + \left(I_{ijmn} - M^{-1}_{ijmn} \right) H^{-1}_{mnkl} \right)^{-1} \tag{31}$$

The relationships between the second-rank damage variable tensor, ϕ_{ij}, and the healing variable tensor, h_{ij}, with the fourth-rank damage tensor, κ_{ijkl}, and the

healing variable tensor, h'_{ijkl}, are obtained by substituting Q_{ijkl} from Eq. 31 into Eq. 28 (Voyiadjis et al. 2012b):

$$\begin{aligned}\left(M^{-1}_{ijuw}+\left(I_{ijmn}-M^{-1}_{ijmn}\right)H^{-1}_{mnuw}\right)^{-1}\left(M^{-1}_{klpq}+\left(I_{klmn}-M^{-1}_{klmn}\right)H^{-1}_{mnpq}\right)^{-1}\\ =\left(h'^{(1)}_{mnij}E^{d}_{klmn}+E^{d}_{klij}\right)\overline{E}^{-1}_{uwpq}\\ \left(M^{-1}_{ijuw}+\left(I_{ijmn}-M^{-1}_{ijmn}\right)H^{-1}_{mnuw}\right)^{-1}\left(M^{-1}_{klpq}+\left(I_{klmn}-M^{-1}_{klmn}\right)H^{-1}_{mnpq}\right)^{-1}\\ =\left(E^{d}_{mnij}h'^{(2)}_{klmn}+E^{d}_{klij}\right)\overline{E}^{-1}_{uwpq}\end{aligned} \tag{32}$$

Upon substituting M_{ijkl} and H_{ijkl} from Eq. 30 into Eq. 32, the relationships between all introduced damage-healing variables are established (Voyiadjis et al. 2012b). Consequently, knowing one set of these damage-healing variables, i.e., (ϕ_{ij}, h_{ij}) or (κ_{ij}, h'_{ij}), the other set can be obtained using Eq. 32 (Voyiadjis et al. 2012b).

Results and Discussion

The performance of the new introduced scalar healing variables is evaluated for both the coupled and the uncoupled damage-healing processes in (Voyiadjis et al. 2012c). In the case of elastic strain energy equivalence, the relation between the damage variables and the healing variables is introduced in Eq. 16. The effect of healing is studied in Fig. 12. The state of fully damaged configuration is obtained by setting the healing variables $h = 1$ and $h' = 0$ in such a way that the healing effect is omitted. The effect of healing parameter h is then evaluated for three different healing values including $h = 0.8$, 0.5 and 0.3, and the healing variable h' is kept zero. It is seen that the valid ranges for the damage variable ϕ when the damage variable l varies in the range $0 < l < 1$ is $0 \leq \phi \leq 0.293$. Consequently, the limiting maximum value for ϕ based on the definition of the stiffness reduction of l is 0.293 (Voyiadjis and Kattan 2009; Voyiadjis et al. 2012c). However, in the case of the healed configuration, this limiting value is increased and the material can sustain more damage based on the definition of ϕ with respect to the unhealed material.

In the case of the coupled damage-healing process, the physical relation between the damage and healing processes are utilized to introduce an empirical relation between the healing and damage variables (Voyiadjis et al. 2012a, c). For instance, in the case of self-healing materials containing microencapsulated healing agent, one may find a relation between the introduced damage into the system and the diffused healing agent into the microcracks (Kirkby et al. 2008, 2009). Here, an empirical function is introduced which relates the damage variable l to the healing variable h' as follows (Voyiadjis et al. 2012c):

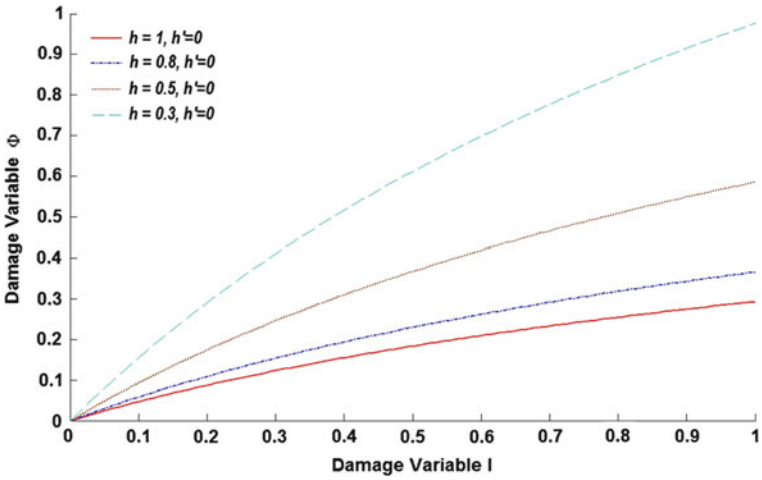


Fig. 12 Effect of healing variable h on damage parameters ϕ and l (After Voyiadjis et al. 2012c)

$$Z(l) = \alpha e^{-\beta l}, \text{ and } h' = \begin{cases} 1 & \text{if } Z(l) > 1 \\ Z(l) & \text{if } Z(l) \leq 1 \end{cases} \tag{33}$$

where α and β are two material-dependent constants which may represent the physical characterization and dispersion of microcapsules and diffusion and effectiveness of the healing agent (Voyiadjis et al. 2012c). Equation 33 performance is depicted in Fig. 13. The healing is assumed to recover all damages at the initial stages of the damage process, and with increasing the amount of the introduced damage into the system, the effectiveness of the healing process is reduced (Voyiadjis et al. 2012c).

This empirical relation is now introduced into the system and the coupled damage-healing process is evaluated. In the case of elastic strain energy equivalence and making use of Eq. 16, the effect of the healing process h' on the damage variable ϕ is depicted in Fig. 14, while the healing variable h is set to be 1 which means this healing variable is deactivated (Voyiadjis et al. 2012c). The damaged configuration and the two coupled damage-healing configurations are shown in Fig 14. Both healing processes show full recovery of the damage variable ϕ at the initial damage process and both of these healing processes show less damage value for ϕ at the point $l = 1$.

The effect of the healing variable h' in the case of elastic strain energy equivalence in a coupled damage-healing process on the stress ratio $\sigma/\overline{\sigma}$ (stress ratio between damaged and effective configuration) is depicted in Fig. 15. Equation 15 is used and both healed configurations show an initial non-damaged response. After a certain limit of damage, they show higher stress ratios, which is an indication of a strengthened material where damaged stress is closer to the effective non-damaged configuration (Voyiadjis et al. 2012c).

One may introduce the governing Eq. 33 into an uncoupled damage-healing process such as self-healing materials which undergo two-step CTH (Li and Uppu

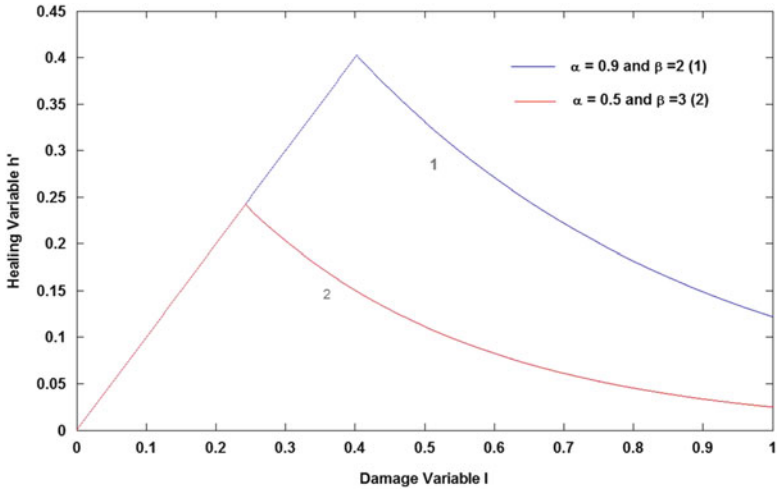


Fig. 13 Empirical relation between damage variable l and healing variable h′ in a coupled damage-healing process (After Voyiadjis et al. 2012c)

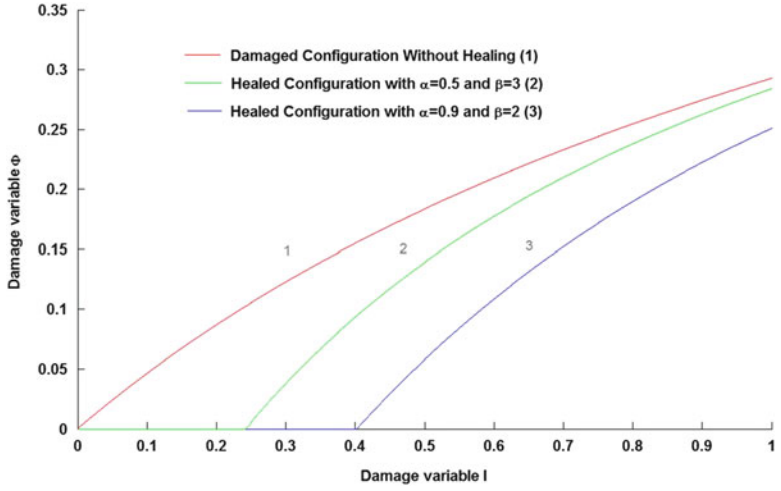


Fig. 14 Effect of healing variable h′ in a coupled damage-healing process on damage variable ϕ (After Voyiadjis et al. 2012c)

2010). In Fig. 16, this situation is investigated where the material is unloaded and healed at a certain level of damage ($l = 0.5$). Healing parameters are chosen as $\alpha = 0.9$ and $\beta = 2$, and the material is loaded after healing. In Fig. 16, the damage variable l following the healing process is reduced and the material shows higher resistance to the damage after healing.

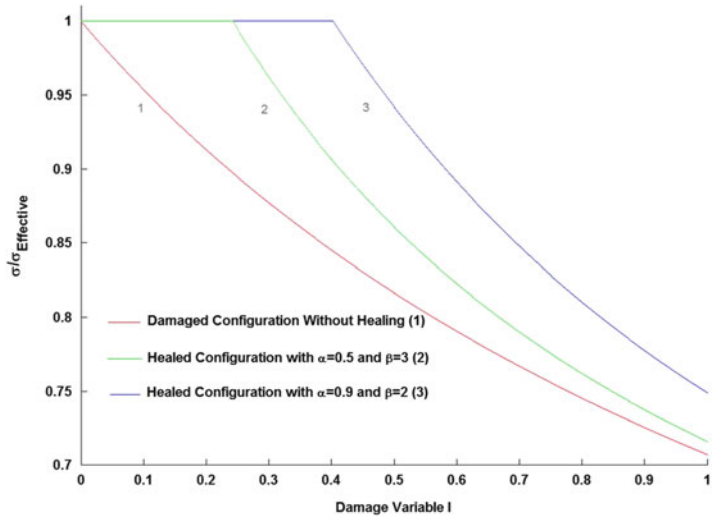


Fig. 15 Effect of healing variable h' in a coupled damage-healing process on stress ratio between damaged and effective configurations (After Voyiadjis et al. 2012c)

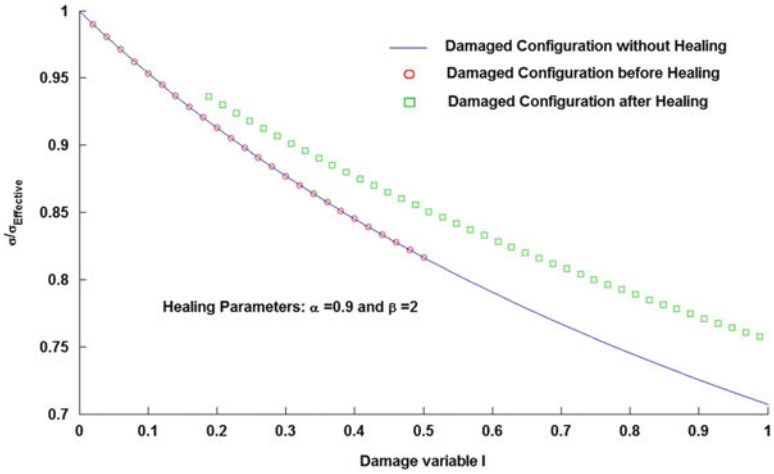


Fig. 16 Effect of healing variable h′ in an uncoupled damage-healing process on stress ratio between damaged and effective configurations (After Voyiadjis et al. 2012c)

Conclusion

The mechanics of damage and healing processes is studied in this chapter, and the bases of a physically based modeling of self-healing are presented (Voyiadjis et al. 2012a, 2012c). The healing variables for calibrating the healing process are introduced in which two new anisotropic damage-healing variables which are

defined based on the physics of the damage and healing processes are reviewed. The performances of the newly developed damage-healing variables are examined and they have been performing quite well in capturing SMP-based self-healing system responses (Voyiadjis et al. 2011, 2012b). The proposed anisotropic damage-healing variables provide the designers with the ability to measure the state of the damage and healing in a highly anisotropic self-healing system. As shown by the authors, the proposed potential functions are capable of capturing the irregular coupled and/or uncoupled damage-healing responses of the polymeric-based material systems. However availability of experiments for self-healing systems is limited in the literature, the proposed theory provides mathematical competency to capture the most complex responses. The healing test setup at Louisiana State University (USA, Baton Rouge) is utilized to investigate the performance of the developed framework and it correlates well with the observed results (Voyiadjis et al. 2011). The microstructural physics of the healing can be incorporated in CDHM framework by introducing *fabric tensors*. Fabric tensors capture the crack and void distributions in the damaged material and they have been linked to the CDM concept (Voyiadjis and Kattan 2006a, 2007; Voyiadjis et al. 2007). This will open a new field for the material researchers in associating fabric tensors and micromechanics to the concept of the healing mechanics.

The healing variables should be prescribed constitutively based upon the material system characteristics and the healing cycle. This task is accomplished by taking into account several deformation mechanisms, involved in healing process, including diffusion of healing particles into matrix and wetting of the fracture surfaces by the healing agents.

References

J.A. Adam, A simplified model of wound healing (with particular reference to the critical size defect). Math. Comput. Model. **30**(5–6), 23–32 (1999)

B.A. Beiermann, M.W. Keller, N.R. Sottos, Self-healing flexible laminates for resealing of puncture damage. Smart Mater. Struct. **18**(8), 085001 (2009)

I.P. Bond, R.S. Trask, H.R. Williams, Self-healing fiber-reinforced polymer composites. MRS Bull. **33**, 770–774 (2008)

E. Brown, N. Sottos, S. White, Fracture testing of a self-healing polymer composite. Exp. Mech. **42**(4), 372–379 (2002)

J.L. Chaboche, Constitutive equations for cyclic plasticity and cyclic viscoplasticity. Int. J. Plast. **5**(3), 247–302 (1989)

J.L. Chaboche, On some modifications of kinematic hardening to improve the description of ratchetting effects. Int. J. Plast. **7**(7), 661–678 (1991)

J.L. Chaboche, Cyclic viscoplastic constitutive equations, part I: a thermodynamically consistent formulation. J. Appl. Mech. **60**(4), 813–821 (1993)

J.L. Chaboche, Thermodynamic formulation of constitutive equations and application to the viscoplasticity and viscoelasticity of metals and polymers. Int. J. Solids Struct. **34**(18), 2239–2254 (1997)

J.L. Chaboche, A review of some plasticity and viscoplasticity constitutive theories. Int. J. Plast. **24**(10), 1642–1693 (2008)

N.R. Hansen, H.L. Schreyer, A thermodynamically consistent framework for theories of elastoplasticity coupled with damage. Int. J. Solids Struct. **31**(3), 359–389 (1994)
S.A. Hayes, F.R. Jones, K. Marshiya, W. Zhang, A self-healing thermosetting composite material. Compos. A: Appl. Sci. Manuf. **38**(4), 1116–1120 (2007a)
S.A. Hayes, W. Zhang, M. Branthwaite, F.R. Jones, Self-healing of damage in fibre-reinforced polymer-matrix composites. J. R. Soc. Interface **4**(13), 381–387 (2007b)
M. John, G. Li. Self-healing of sandwich structures with grid stiffened shape memory polymer syntactic foam core. Smart Mater. Struct. **19**(7), paper number 075013 (2010)
L.M. Kachanov, On the creep fracture time. Izv Akad. Nauk USSR Otd. Tekh **8**, 26–31 (1958)
E.L. Kirkby, J.D. Rule, V.J. Michaud, N.R. Sottos, S.R. White, J.E. Månson, Embedded shape-memory alloy wires for improved performance of self-healing polymers. Adv. Funct. Mater. **18**(15), 2253–2260 (2008)
E.L. Kirkby, V.J. Michaud, J.A.E. Månson, N.R. Sottos, S.R. White, Performance of self-healing epoxy with microencapsulated healing agent and shape memory alloy wires. Polymer **50**(23), 5533–5538 (2009)
J. Lemaitre, Coupled elasto-plasticity and damage constitutive equations. Comput. Methods Appl. Mech. Eng. **51**(1–3), 31–49 (1985)
J. Lemaitre, J.L. Chaboche, *Mechanics of Solid Materials* (Cambridge University Press, Cambridge, 1990)
J. Lemaitre, J. Dufailly, Damage measurements. Eng. Fract. Mech. **28**(5–6), 643–661 (1987)
G. Li, M. John, A self-healing smart syntactic foam under multiple impacts. Compos. Sci. Technol. **68**(15–16), 3337–3343 (2008)
G. Li, D. Nettles, Thermomechanical characterization of a shape memory polymer based self-repairing syntactic foam. Polymer **51**(3), 755–762 (2010)
G. Li, A. Shojaei, A viscoplastic theory of shape memory polymer fibres with application to self-healing materials. Proc. R. Soc. A Math. Phys. Eng. Sci. **468**(2144), 2319–2346 (2012)
G. Li, N. Uppu, Shape memory polymer based self-healing syntactic foam: 3-D confined thermomechanical characterization. Compos. Sci. Technol. **70**(9), 1419–1427 (2010)
G. Li, H. Meng, J. Hu, Healable thermoset polymer composite embedded with stimuli-responsive fibres. J. R. Soc. Interface **9**(77), 3279–3287 (2012)
Y.L. Liu, Y.W. Chen, Thermally reversible cross-linked polyamides with high toughness and self-repairing ability from maleimide- and furan-functionalized aromatic polyamides. Macromol. Chem. Phys. **208**(2), 224–232 (2007)
S. Miao, M.L. Wang, H.L. Schreyer, Constitutive models for healing of materials with application to compaction of crushed rock salt. J. Eng. Mech. **121**(10), 1122–1129 (1995)
S. Murakami, Mechanical modeling of material damage. J. Appl. Mech. **55**(2), 280–286 (1988)
J. Nji, G. Li. A self-healing 3D woven fabric reinforced shape memory polymer composite for impact mitigation. Smart Mater. Struct. **19**(3), paper number 035007 (2010a)
J. Nji, G. Li, A biomimic shape memory polymer based self-healing particulate composite. Polymer **51**(25), 6021–6029 (2010b)
J.W.C. Pang, I.P. Bond, A hollow fibre reinforced polymer composite encompassing self-healing and enhanced damage visibility. Compos. Sci. Technol. **65**(11–12), 1791–1799 (2005)
T.A. Plaisted, S. Nemat-Nasser, Quantitative evaluation of fracture, healing and re-healing of a reversibly cross-linked polymer. Acta Mater. **55**(17), 5684–5696 (2007)
Y.N. Rabotnov, On the equations of state for creep, in *The Progress in Applied Mechanics – The Prager Anniversary Volume* (Macmillan, New York, 1963), pp. 307–315
A. Shojaei, M. Eslami, H. Mahbadi, Cyclic loading of beams based on the Chaboche model. Int. J. Mech. Mater. Des. **6**(3), 217–228 (2010)
A. Shojaei, G. Li, G.Z. Voyiadjis, Cyclic viscoplastic-viscodamage analysis of shape memory polymers fibers with application to self-healing smart materials. J. Appl. Mech. **80**, 011014-1–011014-15 (2013)

A.H.R.W. Simpson, T.N. Gardner, M. Evans, J. Kenwright, Stiffness, strength and healing assessment in different bone fractures – a simple mathematical model. Injury **31**(10), 777–781 (2000)

K.S. Toohey, N.R. Sottos, J.A. Lewis, J.S. Moore, S.R. White, Self-healing materials with microvascular networks. Nat. Mater. **6**(8), 581–585 (2007)

R.S. Trask, I.P. Bond, Biomimetic self-healing of advanced composite structures using hollow glass fibres. Smart Mater. Struct. **15**(3), 704 (2006)

R.J. Varley, S. van der Zwaag, Towards an understanding of thermally activated self-healing of an ionomer system during ballistic penetration. Acta Mater. **56**(19), 5737–5750 (2008)

G.Z. Voyiadjis, P.I. Kattan, Damage mechanics with fabric tensors. Mech. Adv. Mater. Struct. **13** (4), 285–301 (2006a)

G.Z. Voyiadjis, P. Kattan, Advances in Damage Mechanics: Metals and Metal Matrix Composites With an Introduction to Fabric Tensors, 2nd edn. (Elsevier, Oxford, 2006b), p. 742, ISBN: 0-08-044688-4

G.Z. Voyiadjis, P.I. Kattan, Evolution of fabric tensors in damage mechanics of solids with micro-cracks: part I – theory and fundamental concepts. Mech. Res. Commun. **34**(2), 145–154 (2007)

G.Z. Voyiadjis, P.I. Kattan, A comparative study of damage variables in continuum damage mechanics. Int. J. Damage Mech. **18**(4), 315–340 (2009)

G.Z. Voyiadjis, P.I. Kattan, Z.N. Taqieddin, Continuum approach to damage mechanics of composite materials with fabric tensors. Int. J Damage Mech. **18**(3), 301–329 (2007)

G.Z. Voyiadjis, P.I. Kattan, Mechanics of small damage in fiber-reinforced composite materials. Compos. Struct. **92**(9), 2187–2193 (2010)

G.Z. Voyiadjis, A. Shojaei, G. Li. A thermodynamic consistent damage and healing model for self healing materials. Int. J. Plast. **27**(7), 1025–1044 (2011)

G.Z. Voyiadjis, A. Shojaei, G. Li, P.I. Kattan, A theory of anisotropic healing and damage mechanics of materials. Proc. Roy. Soc. A Math. Phys. Eng. Sci. **468**(2137), 163–183 (2012a). doi:10.1098/rspa.2011.0326

G.Z. Voyiadjis, A. Shojaei, G. Li, A generalized coupled viscoplastic- viscodamage- viscohealing theory for glassy polymers. Int. J. Plast. **28**(1), 21–45 (2012b)

G.Z. Voyiadjis, A. Shojaei, G. Li, P. Kattan, Continuum damage-healing mechanics with introduction to new healing variables. Int. J. Damage Mech. **21**(3), 391–414 (2012c)

S.R. White, N.R. Sottos, P.H. Geubelle, J.S. Moore, M.R. Kessler, S.R. Sriram, E.N. Brown, S. Viswanathan, Autonomic healing of polymer composites. Nature **409**(6822), 794–797 (2001)

H.R. Williams, R.S. Trask, I.P. Bond, Self-healing composite sandwich structures. Smart Mater. Struct. **16**(4), 1198 (2007)

S. Yazdani, H.L. Schreyer, Combined plasticity and damage mechanics model for plain concrete. J. Eng. Mech. **116**, 1435–1450 (1990)

M. Zako, N. Takano, Intelligent material systems using epoxy particles to repair microcracks and delamination damage in GFRP. J. Intell. Mater. Syst. Struct. **10**(10), 836–841 (1999)

Modeling of Nonlocal Damage Using the Phase Field Method

47

George Z. Voyiadjis and Navid Mozaffari

Contents

G.Z. Voyiadjis (✉) • N. Mozaffari
Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA
e-mail: voyiadjis@eng.lsu.edu; nmozaf2@lsu.edu

G. Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9_47

Abstract

In this chapter, the application of the phase field method (PFM) into continuum damage mechanics is discussed. It is shown that the effect of the damage gradient can be deduced using the PFM which yields a nonlocal damage model. This is derived for isotropic damage using a scalar variable. The derivation is in the elastic region and the damage rate equation shows the evolution of damage for brittle materials. However, this theory may be coupled with a plasticity model. The framework of the phase field method is discussed in a simple scalar form. After a brief review of isotropic damage, the order parameter is related to the damage variable and a free energy functional in damaged materials is derived which is capable in capturing the evolution of nonlocal damage through the Allen–Cahn equation. It is shown that there is no need to follow the conventional normality rule – which is common in previously proposed models – using this variational approach. Specific length scale due to damage is proposed and the general state of stress with scalar damage variable is discussed. Details of three different finite difference schemes are discussed and the application and regularization capabilities of the model are demonstrated by a 1D numerical example.

Introduction

The phase field method (PFM) as a powerful theoretical and computational tool is applied in many research areas. Transformation that consists of two or more different phases and the continuous change between the different phases can be represented by this method. It has been used to simulate temporal evolution specifically for important phase transformation problems like microstructure evolution, diffusion, and solidification in solid materials. This method has several applications in various research areas including microstructural evolution (Guo et al. 2005; Hu et al. 2007), solidification (Boettinger et al. 2002; Cha et al. 2001; Gránásy et al. 2004; Karma 2001; Ohno and Matsuura 2010), inhomogeneous elasticity (Boussinot et al. 2010; Hu and Chen 2001; Sankarasubramanian 2011; Wang et al. 2002; Zhu et al. 2001), stress-induced phase transformation (Levitas and Ozsoy 2009a, b; Levitas and Preston 2002), crack propagation and fracture models (Aranson et al. 2000; Karma et al. 2001; Miehe et al. 2010a, b; Spatschek et al. 2006, 2007), theory of dislocations and dislocation dynamics (Koslowski et al. 2002; Rodney et al. 2003; Wang et al. 2001), and grain growth simulation (Fan and Chen 1997; Uehara et al. 2007). It can also be used in conjunction with elasticity or a combination of elasticity and diffusion (Onuki 1989), and enhancement of this method with nonlinear mechanical behavior is discussed in a number of references (Gaubert et al. 2010; Guo et al. 2008; Yamanaka et al. 2008; Zhou et al. 2008). Multiphase phenomena can also be simulated with this approach (Moelans 2011; Ofori-Opoku and Provatas 2010; Steinbach and Apel 2006; Steinbach et al. 1996). On the other hand, damage mechanics is introduced to develop constitutive and computational models in order to predict the material

behavior which is based on the evolution of microcracks and microvoids. As it is common in solid mechanics to use finite element method, several attempts have been made to develop computational models to predict damage evolution by using the finite element method (Abu Al-Rub and Voyiadjis 2003; Dorgan and Voyiadjis 2007; Voyiadjis 1988; Voyiadjis and Deliktas 2000; Voyiadjis et al. 2009; Voyiadjis and Dorgan 2007; Voyiadjis and Kattan 1990, 2006). In parallel, it is shown that the phase field method as a powerful technique can be used in tracking the microstructure and morphological evolution in phase transformations with diffuse interfaces. Setting a general thermodynamic consistent framework that combines standard phase field approach with conventional damage mechanics theory is the motivation of this chapter. The advantages of using this theory are the relative simplicity of implementation on computer programs using the finite difference method, governing an estimate of the damaged part through the nonlocal term and also predicting temporal evolution of damage. The local behavior of each phase inside the RVE (representative volume element) is obtained by the classical damage theories as an assumption in the following sections. The RVE is large enough to consider both phases as well as averaging them in the RVE. Therefore, averaging of the microcrack areas is set and specific arrangement of phases is not assumed inside the RVE. The formulation is based on the general type of damage variable which is related to order parameter based on the physics of the problem to simulate damage growth.

The outline of this chapter is as follows: First, a quick introduction of the PFM application history in material science is given. Then, the phase field theory in a simple scalar form is discussed based on the work of Boettinger et al. (2002). Aforementioned framework provides a succinct and sound physical base of the phase field theory. The concept of isotropic damage is then reviewed in order to link the order parameter to the damage variable in continuum damage mechanics (CDM). Following that, free energy functional of one scalar order parameter is considered for two different phases in the damaged material. This functional enables one to capture the evolution of nonlocal damage through the phase field theory. Next, the time-dependent Ginzburg–Landau equation (TDGL) which is also known as the Allen–Cahn equation is used to describe the damage evolution process. It is shown that by using the Allen–Cahn equation, there is no reason to use conventional normality rule to obtain the damage evolution law which is common in previously proposed damage models. It can be seen that using free energy functional without using Allen–Cahn equation will alleviate the use of the phase field theory for damage and reduce the formulation to regular models. Specific length scales due to transformation (damage in this case) exist which show the difference between sharp interface models and diffuse interface models. These length scales capture the effect of the damaged localization zone and address the interface region in which the process of changing undamaged solid to fully damaged material (microcracks) occurs. Subsequently, the terms free energy in the phase field formulation are compared with the corresponding terms in the variational formulation, and the properties of gradient damage models through the variational formulation are demonstrated. Next, generalized state of stress in elastic behavior is considered assuming isotropic damage and the effect of the scalar phase

field-based damage variable are shown on components of stress and strain tensors. When the damage is isotropic in behavior, it evolves simultaneously at the same rate in all the three mutually orthogonal directions and it is represented by a scalar variable, but it can affect any component of the stress or strain tensors. Finally, application of the new phase field-based damage model to a one-dimensional problem is demonstrated and its numerical implementation is discussed. Three different finite difference schemes are used, and a mathematical procedure is derived to show regularization capabilities of the proposed model by means of numerical examples and the validity and usefulness of the phase field modeling approach. For the sake of simplicity, small deformation theory is assumed. Therefore, higher-order terms in the displacement field are neglected. In this chapter, any variable with an overbar indicates the effective state (undamaged material) and without the bar indicates the real damaged state. For more details of this work, readers are referred to the work by the authors (Voyiadjis and Mozaffari 2013).

General Framework of Phase Field Models

Order Parameter

Phase transition (transformation) is known as physical change in materials from one state to another state. For instance, the internal temperature rise of water by heating causes a change from solid state (ice) to liquid state. In a standard phase field model, present phases in the system can be defined using order parameter or "phase field." There is no obligation to find a macroscopic physical interpretation for order parameters. In a two-phase system, order parameter is set to zero in one phase and it is set to one in the other phase and variation of the order parameter in different phases is shown using a smooth function. Like other thermodynamic consistent methods, free energy can be defined as a function of the order parameter and other thermodynamic variables such as temperature and concentration. The system contains two separate phases, one phase is classified as ordered phase and the other is disordered phase (Elder and Provatas 2010). Ordered phase is the phase with lower number of geometric symmetries and is specified by an order parameter equal to any arbitrary value except zero. Disordered phase is a phase with higher number of geometric symmetries. Therefore, in solid–liquid transformation, solid phase is an ordered phase and liquid phase can be defined as a disordered phase. Two types of commonly used order parameters, "field variables," include conserved and nonconserved order parameters (Chen 2002; Moelans et al. 2008). The type of a function which is used to show the transition from one phase to another phase shows the type of transition. If the order parameter vanishes continuously from ordered to disordered phases at the boundary, the transition is called second-order transition, and if there is a discontinuous change in order parameter between phases, the transition is called first-order transition (Elder and Provatas 2010).

Framework of Phase Field Method

The phase field modeling approach which is based on the superconductivity theory of Ginzburg–Landau (Cahn and Hilliard 1958; Ginzburg and Landau 1965) is used to model various types of microstructure evolution as a powerful tool during the last two decades. Generally, two types of the phase field models are used (Chen 2002). The first type is introduced to avoid tracking of the interfaces during microstructural evolution, and the thermodynamic and kinetic coefficients in this type of model are selected to correspond to parameters of the model to the conventional sharp interface model. In the second type, order parameters are defined physically to incorporate field variables for transformation. The latter type of modeling has been used widely for many solid-state transformations. As a general assumption in the second type, temporal microstructure evolution during the process can be achieved by the phase field equations including Allen–Cahn (Allen and Cahn 1979; Cahn and Allen 1977; Cahn and Hilliard 1958) and Cahn–Hilliard (Cahn and Hilliard 1958; Gurtin 1996) equations. In addition, all thermodynamic and kinetic coefficients can be related to the microstructure parameters. The conventional Landau type of expansion which is given later is used to define the free energy function as a polynomial of order parameters (Elder and Provatas 2010). The first approach is used in this chapter. Evolution equations of the phase field variables are derived using the general thermodynamic and kinetic principles based on the Ginzburg–Landau theory (Ginzburg and Landau 1965; Nauman and Balsara 1989; Sethna 2006). These equations along with the free energy form the basis of the phase field method. The free energy functional F for an isothermal process can be defined as functional of phase field variables and their gradients as follows (Boettinger et al. 2002):

$$F = \int_V \left[\psi(c, \eta, T) + \frac{\epsilon_c^2}{2} |\nabla c|^2 + \frac{\epsilon_\eta^2}{2} |\nabla \eta|^2 \right] dV \tag{1}$$

where $\psi(c, \eta, T)$ is the free energy density, c is the concentration, T is the temperature, η is the order parameter, and ϵ_c and ϵ_η are gradient coefficients. The free energy functional which is given by Eq. 1 must decrease during the microstructure evolution. Determination of gradient coefficients in Eq. 1 can give an accurate description of interface properties such as the interface energy and anisotropy of interface energy. The variational derivatives of the free energy functional F (Eq. 1) with concentration c as conservative field and nonconservative field η must satisfy the following equations in equilibrium conditions assuming the gradient energy coefficients are constant:

$$\frac{\delta F}{\delta \eta} = \frac{\partial \psi}{\partial \eta} - \epsilon_\eta^2 \left(\nabla^2 \eta \right) = 0 \tag{2}$$

$$\frac{\delta F}{\delta c} = \frac{\partial \psi}{\partial c} - \epsilon_c^2 \left(\nabla^2 c\right) = \text{constant} \tag{3}$$

Constant concentration during the process (in damage process this is not the case) can ensure the last equation to be a constant. Decrease in total free energy and increase in entropy versus time are guaranteed by the Ginzburg–Landau equations during the process which are given by

$$\frac{\partial \eta}{\partial t} = -M_\eta \left[\frac{\partial \psi}{\partial \eta} - \epsilon_\eta^2 \left(\nabla^2 \eta\right)\right] \tag{4}$$

$$\frac{\partial c}{\partial t} = \nabla . \left[M_c c(1-c) \nabla \left(\frac{\partial \psi}{\partial c} - \epsilon_c^2 \left(\nabla^2 c\right)\right)\right] \tag{5}$$

M_η and M_c are positive mobility constants which are related to the kinetic coefficients. These coefficients can be obtained through experiments and characterized based on the mechanism of transformation. Comparing results of the new model to the previous proposed models is another way to find these specific coefficients. Equation 4 is the time-dependent form of Ginzburg–Landau equation and is termed the Allen–Cahn equation . Equation 5 is called the Cahn–Hilliard equation. Evolution of order parameter with respect to time is proportional to the change of the free energy functional with respect to the order parameter through the Allen–Cahn equation. Since the concentration is not conserved during the damage growth, Cahn–Hilliard equation cannot be used in the following derivation. If the phase field evolution remains in equilibrium condition for static or quasi-static loading (Hunter and Koslowski 2008), then Eq. 4 can be expressed as follows:

$$\frac{\partial \psi}{\partial \eta} = 0 \tag{6}$$

The original form of the theory will be followed for phase field evolution in the case of dynamic impact loading, by neglecting the gradient coefficient as follows:

$$\frac{\partial \eta}{\partial t} = -M_\eta \frac{\partial \psi}{\partial \eta} \tag{7}$$

The term ψ shows the free energy function as a function of the order parameter. Definition of the problem shows the type of thermodynamic function which can be used in the general equation. In a more general case, when there are n different kinds of phases in a system, an order parameter is introduced in each phase under the following constraint condition:

$$\sum\nolimits_{i=1}^{n} \eta_i = 1 \quad \text{with} \quad \eta_i \geq 0, \forall i \tag{8}$$

Landau type of free energy is preferred in such a system because all the terms are a function of the order parameter. Description of the phenomena restricts the

number of terms in Taylor expansion of order parameter to define this free energy. In a simple case with one order parameter, Landau free energy can be given by the following form and each coefficient can be a function of temperature:

$$\mathcal{L} = \mathcal{L}_0 + a\eta + b\eta^2 + c\eta^3 + d\eta^4 + e\eta^5 + f\eta^6 + \dots \tag{9}$$

Landau-type free energy formulation can be used in any other applicable forms if each term consists of the order parameter. The nature of the problem defines the type of description of free energy which can be used to describe the phenomenon. For instance, entropy can be used in isolated systems with variation of temperature. Gibbs free energy may be used in the systems with constant pressure and temperature, while for the systems with constant temperature and volume like most of the solid mechanics problems, the Helmholtz free energy may be used. In the definition of the free energy functional (Eq. 1), it can be seen that except the first term, other terms depend only on the gradient of the order parameter and concentration. These terms are equal to zero except in the interface region where $0 < \eta < 1$ and ϵ_c and ϵ_η are the gradient energy coefficients with the following definition (Cahn and Hilliard 1958):

$$\epsilon_\eta^2 = \frac{\partial^2 \psi}{\partial(|\nabla\eta|)^2} - 2\frac{\partial}{\partial\eta}\left(\frac{\partial\psi}{\partial(\nabla^2\eta)}\right) \tag{10}$$

$$\epsilon_c^2 = \frac{\partial^2 \psi}{\partial(|\nabla c|)^2} - 2\frac{\partial}{\partial c}\left(\frac{\partial\psi}{\partial(\nabla^2 c)}\right) \tag{11}$$

As a general rule in phase field modeling, free energy ψ which covers the whole domain of the order parameter $0 \leq \eta \leq 1$ should lead to the appropriate term when only one phase exists. Another way of constructing the free energy function is to use a double well potential function with the minima in the two different phase configurations and use another interpolating function. Assume that ψ_1 and ψ_2 show the energy of each phase as a function of temperature and its concentration. Therefore, the free energy ψ can be described as

$$\psi(c,\eta,T) = h(\eta)\psi_1(c_1,T) + (1-h(\eta))\psi_2(c_2,T) + Wg(\eta) \tag{12}$$

with the following constraint:

$$c_1 + c_2 = 1 \tag{13}$$

In Eq. 12, $g(\eta)$ is a well double function and $h(\eta)$ is the interpolating monotonic function between two phases. Various choices are possible for functions $g(\eta)$ and $h(\eta)$. Some expressions for these functions are given in the work by (Chen 2002)

$$g(\eta) = \eta^2(1-\eta)^2 \tag{14}$$

$$h(\eta) = \eta^3\left(6\eta^2 - 15\eta + 10\right) \tag{15}$$

$$h(\eta) = \eta^2(3-2\eta) \tag{16}$$

The effect of dissipation is shown by function $g(\eta)$; therefore, it should be a monotonically increasing function of the phase energy when $0 < \eta < 1$ and should satisfy $(0) = 0$, $g(1) = 0$, $\left.\frac{\partial g}{\partial \eta}\right|_{\eta=0} = \left.\frac{\partial g}{\partial \eta}\right|_{\eta=1} = 0$. Mathematical restrictions of function $h(\eta)$ are $h(0) = 0$, $h(1) = 1$, $\left.\frac{\partial h}{\partial \eta}\right|_{\eta=0} = \left.\frac{\partial h}{\partial \eta}\right|_{\eta=1} = 0$, $\left.\frac{\partial^2 h}{\partial \eta^2}\right|_{\eta=0} = \left.\frac{\partial^2 h}{\partial \eta^2}\right|_{\eta=1} = 0$. The proposed relation $h(\eta) = \frac{\int_0^\eta g(y)dy}{\int_0^1 g(y)dy}$ (Furukawa and Nakajima 2001) can be used to derive an appropriate function $h(\eta)$ from function $g(\eta)$. For other possibilities of functions $g(\eta)$ and $h(\eta)$, the readers are referred to the work by Wang et al. (1993). In the previous formulation (12), the functions $\psi_1(c_1, T)$ and $\psi_2(c_2, T)$ are the Helmholtz free energy densities of the two different phases. The coefficient W should be positive to be consistent with the thermodynamic laws and is used to describe the interfacial energy. Also, concentration within the interface (C_I) will vary between each phase concentration and can be obtained using the interpolation function $h(\eta)$:

$$c_I = h(\eta)c_1(\eta) + (1 - h(\eta))c_2(\eta) \tag{17}$$

Phase Field Method (PFM) Versus Continuum Damage Mechanics (CDM)

Order Parameter

For incorporating PFM to CDM, special transition in a representative volume element (RVE) is considered in which microvoids or microcracks exist in a pure solid (undamaged configuration). The evolution of cracks and voids and chemical reaction does not affect this process. Reduction of area which is continuous during the loading is considered as a nonconserved order parameter. Damage growth leads to the order parameter change during the damage process. It will be represented by η in the following formulation. Based on previous definitions, microcracks and microvoids (fully damaged material) are shown as a disordered phase with $\eta = 0$ and undamaged material (pure solid) can be considered as an ordered phase with $\eta = 1$.

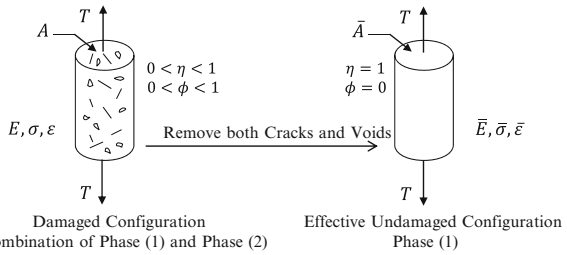


Fig. 1 Damage characterization (After Voyiadjis and Mozaffari 2013)

Therefore, combination of both phases can be seen in the damaged configuration in the conventional continuum damage mechanics. Through CDM, obviously $\phi = 0$ represents undamaged configuration and $\phi = 1$ represents fully damaged material. In addition $0 < \phi < 1$ shows interface region between fully damaged and undamaged configurations. Therefore, common continuum damage variable and the order parameter can be related as follows:

$$\begin{array}{lll} \eta = 0, \quad \phi = 1 & \rightarrow & \text{cracks, voids, fully damaged} \\ 0 < \eta < 1, \quad 0 < \phi < 1 & \rightarrow & \text{damaged configuration} \\ \eta = 1, \quad \phi = 0 & \rightarrow & \text{undamaged configuration} \end{array}$$

Fundamental definitions which are given in the previous sections are illustrated in Fig. 1. Undamaged material is represented by phase (1) and fully damaged material (integration of all cracks and voids) is represented by phase (2).

Therefore, the relation between the order parameter and the continuum damage variable is as follows:

$$\eta = 1 - \phi \tag{18}$$

Definitions which have been used in previous works to simulate crack propagation and simulation of fracture agree with the given relation (Abu Al-Rub and Voyiadjis 2003; Amor et al. 2009; Aranson et al. 2000; Borden et al. 2012; Kuhn and Müller 2010; Miehe et al. 2010b; Salac and Lu 2006; Spatschek et al. 2006; Voyiadjis et al. 2004). It is worth to mention here that in all previous models (Miehe et al. 2010b; Amor et al. 2009; Borden et al. 2012) fracture mechanics is considered and not damage mechanics. Tracking single macrocrack (fracture) in the material is the main goal of these models and they do not focus to address damage which is the integration of microcrack/microvoid area (damage quantification) in the material. The significance of the present derivation is formulating the continuous change between the phases such as the damage growth. In conventional PFM, order parameter should be a function of time and it is related to the damage variable using Eq. 18. In general, both the nonconserved order parameter and the damage variable are functions of position to track different damage levels in the material. Therefore, both variables, the phenomenological nonconserved phase field and damage variable, can be used to show which phase (undamaged or microcracked) is present at the specific material position.

Thermodynamic Formulation of Damage Mechanics Using the Phase Field Method

Nonlocal damage in a material can be described through the use of the damage gradient in the thermodynamic formulation. Helmholtz free energy is used into the phase field formalism to obtain evolution of nonlocal damage in an elastic material. Following the method which is discussed in the previous sections (Boettinger et al. 2002), the free energy function can be constructed in two steps as follows:

1. Each phase free energy is defined as follows:

$$\psi^{ud}(\overline{\varepsilon},\eta) = \frac{1}{2}\overline{E}\,\overline{\varepsilon}^2 \tag{19}$$

$$\psi^{fd}(\overline{\varepsilon},\eta) = 0 \tag{20}$$

   where ψ^{ud} and ψ^{fd} are the free energies of the undamaged and fully damaged configurations, respectively. Modulus of elasticity and the corresponding strain in the undamaged configuration are shown by $\overline{E}$ and $\overline{\varepsilon}$, respectively.
2. Double well function $g(\eta)$ and the interpolation function $h(\eta)$ along with the free energies of fully damaged $\psi^{fd}(\overline{\varepsilon}\eta)$ and undamaged configurations $\psi^{ud}(\overline{\varepsilon}\eta)$ are used in Eq. 12 to construct the free energy of the damaged configuration including both phases:

$$\psi(\overline{\varepsilon},\eta) = h(\eta)\psi^{ud}(\overline{\varepsilon},\eta) + (1-h(\eta))\psi^{fd}(\overline{\varepsilon},\eta) + Wg(\eta) \tag{21}$$

   Substituting Eqs. 19 and 20 into Eq. 21 results in

$$\psi(\overline{\varepsilon},\eta) = h(\eta)\frac{1}{2}\overline{E}\,\overline{\varepsilon}^2 + Wg(\eta) \tag{22}$$

   Neglecting the effect of damage gradient, the free energy of the damaged configuration can be obtained by using Eqs. 14 and 16 in Eq. 22 in which both phases are present:

$$\psi(\overline{\varepsilon},\eta) = \eta^2(3-2\eta)\frac{1}{2}E\,\overline{\varepsilon}^2 + W\eta^2(1-\eta)^2 \tag{23}$$

   Accordingly, the functional of the free energy containing its gradient is given as follows:

$$F = \int_V \left[\eta^2(3-2\eta)\frac{1}{2}\overline{E}\,\overline{\varepsilon}^2 + W\eta^2(1-\eta)^2 + \frac{\epsilon_\eta^2}{2}|\nabla\eta|^2\right]dV \tag{24}$$

Using Eq. 23 into Eq. 4 leads to the order parameter evolution equation. Order parameter is related to the damage variable in Eq. 18. Therefore, the damage evolution equation is obtained as follows:

$$\frac{\partial \phi}{\partial t} = M_\phi \left[3\overline{E}\,\overline{\varepsilon}^2(\phi)(1-\phi) + 2W\phi(2\phi-1)(1-\phi) - \epsilon_\phi^2\left(\nabla^2\phi\right)\right] \tag{25}$$

where W is a positive dissipation constant and ϵ_ϕ^2 is a positive damage gradient constant as a length scale to bring appropriate physical meaning to Eq. 25. These coefficients are written in the following form:

$$W = \overline{E}w \tag{26}$$

$$\epsilon_\phi^2 = \overline{E}l^2 \tag{27}$$

Substituting Eqs. 26 and 27 into Eq. 25 leads to the appropriate form for the evolution of damage:

$$\frac{\partial \phi}{\partial t} = M_\phi \overline{E}\left[3\overline{\varepsilon}^2(\phi)(1-\phi) + 2w\phi(2\phi-1)(1-\phi) - l^2\left(\nabla^2\phi\right)\right] \tag{28}$$

Equation 28 introduces a new nonlocal, gradient-based damage model for scalar damage in elastic materials using PFM. Damage evolution can be obtained using this unique equation. In Eq. 28 l represents a length scale due to damage and depends on the microstructure of the material such as the grain size. Substituting Eqs. 18, 26, and 27 in Eq. 24 results in the definition of the free energy function of the damaged material with the effect of damage gradient:

$$\psi(\overline{\varepsilon}, \phi, \nabla\phi) = \frac{1}{2}\overline{E}\,\overline{\varepsilon}^2(1-\phi)^2(2\phi+1) + \overline{E}w\phi^2(1-\phi)^2 + \frac{1}{2}\overline{E}l^2|\nabla\phi|^2 \tag{29}$$

Equations 28 and 29 are the governing equations of the damage evolution through the PFM.

Comparison Between Proposed Model and the Variational Formulation

In order to validate the formulation and obtain the mathematical restrictions on the coefficient calculations or measurements in the previous equations, the central equations (28 and 29) of the phase field modeling of damage are compared with the assumptions of the variational formulation (Pham et al. 2011). Based on their work, the following properties are mentioned as the general restrictions of any gradient-type damage model.

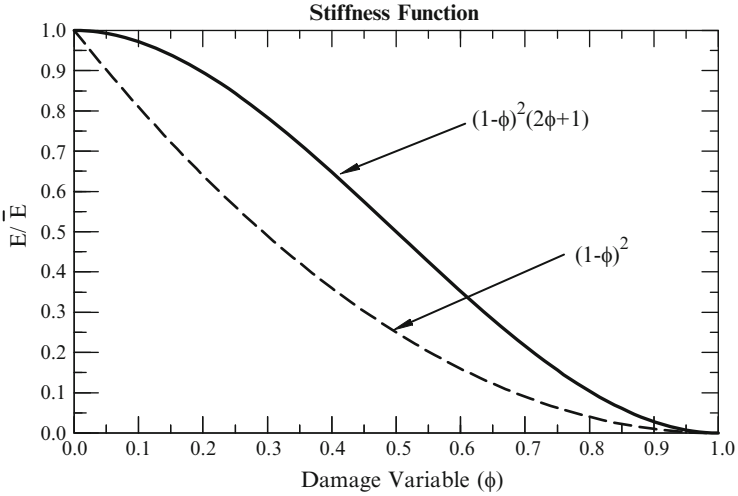


Fig. 2 Variation of the elastic stiffness with damage (After Voyiadjis and Mozaffari 2013)

Positive Elasticity

Stiffness function $E(\phi)$ which indicates the reduction of the stiffness should be positive with $E(\phi = 1) = 0$. Based on Eq. 29 the stiffness function is defined as follows:

$$E(\phi) = \overline{E}(1 - \phi)^2(2\phi + 1) \tag{30}$$

Equation 30 satisfies $E(\phi = 1) = 0$ and $E(\phi) > 0$ for $0 \leq \phi \leq 1$. This new stiffness function is compared with the conventional stiffness function in Fig. 2.

In Fig. 2, the region between the two curves shows the incorporation of the gradient term. The excessive term in the stiffness function, namely, $(2\phi + 1)$, acts as an internal hardening variable which reflects the mechanism of microcrack interaction and arresting inside the material. An inflection point occurs in the curvature of the stiffness function. This inflection point is at $\phi = \frac{1}{2}$ (i.e., $\frac{\partial^2 E}{\partial \phi^2} = 0$) which shows the practical limit of the damage variable (Voyiadjis and Kattan 2012). The damage variable may not have a practical meaning above 0.5. Lemaitre and others pointed out that this is in the neighborhood of 0.3 (Lemaitre and Desmorat 2005). This is obtained without a mathematical justification, except that the continuum is not valid beyond that value. Degradation of the material has a much faster trend after that point. In addition microcracks arrest each other at the beginning of the loading. This introduces hardening; however, after considerable loss of the starting value of stiffness, the material undergoes fast reduction in its elastic stiffness.

Decreasing Stiffness

Stiffness function needs to be a monotonically decreasing function with respect to damage parameter. Therefore we have

$$\frac{\partial E(\phi)}{\partial \phi} < 0 \tag{31}$$

Derivative of Eq. 30 with respect to damage parameter (ϕ) satisfies the aforementioned property:

$$6\overline{E}\phi(\phi - 1) < 0 \tag{32}$$

Dissipation

Positivity of dissipation function $w(\phi)$ with $w(\phi = 0) = 0$ is necessary for gradient damage models. Based on Eq. 29, dissipation function can be obtained as follows:

$$w(\phi) = \overline{E}w\phi^2(1 - \phi)^2 \tag{33}$$

This satisfies $w(\phi = 0) = 0$ and $w(\phi) > 0$ for $0 \leq \phi < 1$. Also, dissipation function (33) should be a monotonically increasing function with respect to the damage variable:

$$\frac{\partial w(\phi)}{\partial \phi} > 0 \tag{34}$$

This leads in the following criterion:

$$2\overline{E}w\phi\left(2\phi^2 - 3\phi + 1\right) > 0$$

This criterion is held unconditionally if $0 < \phi < 1/2$.

Irreversibility

Damage evolution should be positive since it is an irreversible process. Based on the derived equations, all constants in Eq. 28 such as M, w, and l are positive. The nonlocal term containing the length scale parameter, in Eq. 28, is relatively smaller than the other two terms. The summation of the other terms is always positive for $\frac{1}{2} < \phi < 1$. They remain positive based on the specific choice of the dissipation constant (w) under the condition $w < \frac{3}{2}\overline{\varepsilon}^2$. These conditions for the coefficients are followed in the numerical example that follow these conditions.

New Implicit Damage Variable

Equation 29 can be used to define a new damage variable. In general, the fictitious undamaged configuration and real damaged configuration are used to obtain the damage level in a simple way. Mapping the stage of damage from one configuration to the other at each step is possible through the use of specific functions. The stress, strain, and elasticity tensors are used to relate the different damage levels from one configuration to another which is depicted in Fig. 3 as follows:

$$\overline{\sigma}_{ij} = M(\phi)\sigma_{ij} \tag{36}$$

$$\overline{\varepsilon}_{ij} = \varepsilon_{ij}(q(\phi))^{-1} \tag{37}$$

$$\overline{E}_{ijkl} = E_{ijkl}(p(\phi))^{-1} \tag{38}$$

Effective stress coefficient $M(\phi)$ is the function that may take the form of a second- or higher-order tensor for the anisotropic damage case (even order tensor). All functions ($M(\phi)$, $q(\phi)$, and $p(\phi)$) are scalar functions and nonzero over $0 < \phi < 1$.

Strain Energy Equivalence

A more general type of the hypothesis of strain energy equivalence (Sidoroff 1981) is used instead of the hypothesis of strain equivalence. This hypothesis is used to find general mapping functions $M(\phi)$, $q(\phi)$, and $p(\phi)$ as summarized below:

$$\frac{1}{2}E_{ijkl}\varepsilon_{ij}\varepsilon_{kl} = \frac{1}{2}\overline{E}_{ijkl}\overline{\varepsilon}_{ij}\overline{\varepsilon}_{kl} \tag{39}$$

These scalar functions are capable to be used for mapping of isotropic damage in the general material. Using Eqs. 37 and 38 into Eq. 39 leads to

$$\frac{1}{2}(p(\phi))\overline{E}_{ijkl}\left(\overline{\varepsilon}_{ij}q(\phi)\right)\left(\overline{\varepsilon}_{kl}q(\phi)\right) = \frac{1}{2}\overline{E}_{ijkl}\overline{\varepsilon}_{ij}\overline{\varepsilon}_{kl} \tag{40}$$

From Eq. 40, the first specification of the mapping functions is derived as follows:

$$p(\phi)(q(\phi))^2 = 1 \tag{41}$$

The state of stress in the undamaged configuration is expressed as follows:

$$\overline{\sigma}_{ij} = \overline{E}_{ijkl}\overline{\varepsilon}_{kl} \tag{42}$$

Accordingly, the stress in the damaged configuration may be written as follows:

$$\sigma_{ij} = E_{ijkl}\varepsilon_{kl} \tag{43}$$

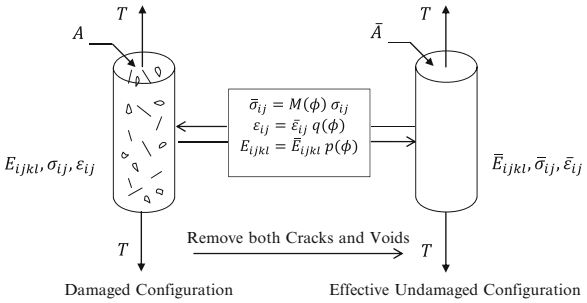


Fig. 3 Tensors for different configurations of isotropic damage (After Voyiadjis and Mozaffari 2013)

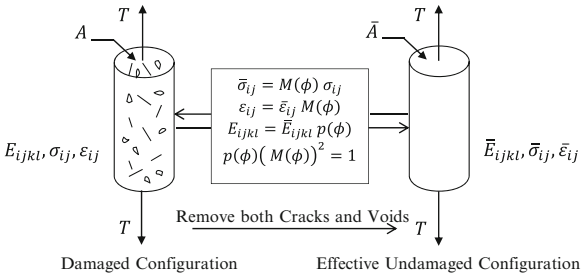


Fig. 4 Mapping functions for isotropic damage (After Voyiadjis and Mozaffari 2013)

Substitution of Eqs. 42 and 43 into Eq. 39 results in another form of the strain energy equivalence hypothesis:

$$\frac{1}{2}E^{-1}_{ijkl}\sigma_{ij}\sigma_{kl} = \frac{1}{2}\bar{E}^{-1}_{ijkl}\bar{\sigma}_{ij}\bar{\sigma}_{kl} \tag{44}$$

Substituting Eqs. 36 and 38 into Eq. 44 leads to

$$\frac{1}{2}E^{-1}_{ijkl}\sigma_{ij}\sigma_{kl} = \frac{1}{2}\left(p(\phi)E^{-1}_{ijkl}\right)\left(M(\phi)\sigma_{ij}\right)\left(M(\phi)\sigma_{kl}\right) \tag{45}$$

Thus, another mapping function specification is derived as follows:

$$p(\phi)\left(M(\phi)\right)^2 = 1 \tag{46}$$

Equations 41 and 46 prove the equality of functions $q(\phi)$ and $M(\phi)$ ($q(\phi) = M(\phi)$). Therefore, based on the strain energy equivalence hypothesis, the general mapping between the two configurations can be summarized as depicted in Fig. 4.

Function $p(\phi)$ is obtained using Eq. 30 as follows:

$$p(\phi) = (1-\phi)^2(2\phi+1) \tag{47}$$

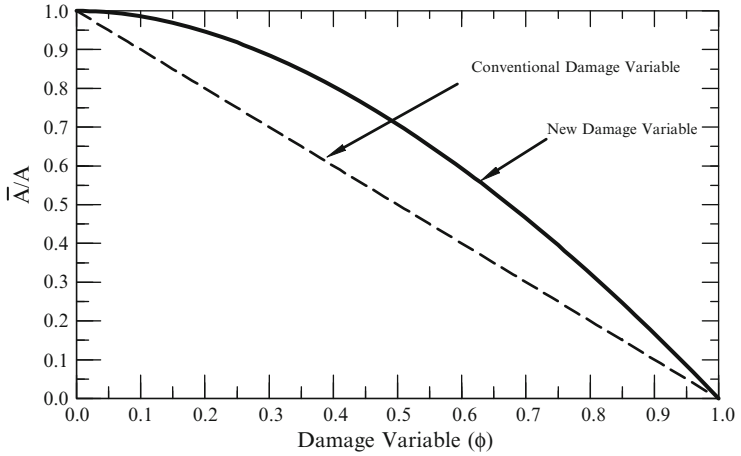


Fig. 5 Change in the cross section due to the damage variable (After Voyiadjis and Mozaffari 2013)

Using Eq. 47 into Eq. 46 leads to the definition of the effective stress function $M(\phi)$ through PFM as follows:

$$M(\phi) = \frac{1}{(1-\phi)\sqrt{(2\phi+1)}} \tag{48}$$

The additional term $\left(\sqrt{(2\phi+1)}\right)$ shows the effect of the damage gradient in the new definition of the effective stress coefficient. Equating the forces in two different configurations (Fig. 4) by using Eq. 48 results in a new implicit definition of the damage variable:

$$\frac{T}{\overline{A}} = \frac{1}{(1-\phi)\sqrt{(2\phi+1)}}\frac{T}{A} \tag{49}$$

Therefore, the implicitly defined damage variable can be deduced as follows:

$$\left(\frac{\overline{A}}{A}\right)^2 = 2\phi^3 - 3\phi^2 + 1 \tag{50}$$

The undamaged area can be computed using Eq. 50 as follows:

$$\overline{A} = A\sqrt{2\phi^3 - 3\phi^2 + 1} \tag{51}$$

The conventional damage variable, $\frac{\overline{A}}{A} = 1 - \phi$ (Kachanov 1958), and the proposed definition are compared in Fig. 5. Equation 50 is a cubic function in damage

which can be solved explicitly in terms of $\frac{\overline{A}}{A_0}$ The solution procedure and the explicit solution are given in the work by the authors Voyiadjis and Mozaffari (2013). This definition of the damage indicates the initial slow propagation of damage due to the interaction of cracks. These cracks initially assist in arresting and slowing down the damage evolution. Beyond the value of $\phi = 0.5$, there are many cracks and the arresting mechanism does not stop the damage propagation. The proposed stiffness function enables one to obtain mapping functions for strain and stress between the two configurations as follows:

$$\overline{\sigma}_{ij} = \frac{1}{(1-\phi)\sqrt{(2\phi+1)}}\sigma_{ij} \tag{52}$$

$$\overline{\varepsilon}_{ij} = \varepsilon_{ij}(1-\phi)\sqrt{(2\phi+1)} \tag{53}$$

Strain Decomposition

From the work of Nemat-Nasser (1979, 1983), one may use the strain additive decomposition for small deformation theory of damage mechanics. Following the work of Voyiadjis and Kattan (1990, 1992) and Abu Al-Rub and Voyiadjis (2003), the total reversible elastic strain ε_{ij}^{E} can be decomposed as follows:

$$\varepsilon_{ij}^{E} = \varepsilon_{ij}^{e} + \varepsilon_{ij}^{ed} \tag{54}$$

In Eq. 54 ε_{ij}^{e} is the conventional elastic strain and ε_{ij}^{ed} is the elastic damage strain. The physical interpretation of this decomposition is given in Abu Al-Rub and Voyiadjis (2003) and the same approach is used in the work of Sadowski et al. (2005) and Samborski and Sadowski (2005). Also, additive decomposition of strain into two components is used widely in phase field models. Interested readers are referred to the works by Levitas and Ozsoy (2009a, b) and Uehara et al. (2007). Decomposition due to various types of order parameter (nonconserved or conserved) is reported in Chen (2002) and Moelans et al. (2008). Strains due to the transformation are termed eigenstrains which can be found in Yamanaka et al. (2008). For instance, the strains are divided to homogeneous and nonhomogeneous parts in the inhomogeneous elasticity models (Salac and Lu 2006; Wang et al. 2002; Yu et al. 2005), and it has been widely used in numerous textbooks (Khachaturyan 1983). This approach in phase field modeling confirms the additive decomposition which is given in Eq. 54. Following the additive strain decomposition, Eq. 54, and using Hooke's law, one can obtain the elastic damage strain and the total elastic strain for isotropic damage in the damaged configuration. Therefore, the total elastic strain in the damaged

configuration (ε_{ij}) is equal to the total elastic strain in the aforementioned decomposition (ε_{ij}^{E}). Hooke's law is written in the following form:

$$\sigma_{ij} = E_{ijkl}\varepsilon_{kl}^{E} \tag{55}$$

Substituting Eq. 54 into Eq. 55 results in

$$\sigma_{ij} = E_{ijkl}\left(\varepsilon_{kl}^{e} + \varepsilon_{kl}^{ed}\right) \tag{56}$$

Using the hypothesis of strain energy equivalence leads to

$$\frac{1}{2}\sigma_{ij}\varepsilon_{ij}^{E} = \frac{1}{2}\overline{\sigma}_{ij}\overline{\varepsilon}_{ij}^{e} \tag{57}$$

One can obtain the relation between the undamaged elastic strain and the total elastic strain using Eq. 36 into Eq. 56 as follows:

$$\overline{\varepsilon}_{ij}^{e} = \frac{1}{M(\phi)}\varepsilon_{ij}^{E} \tag{58}$$

Stress–strain relations in the damaged and undamaged states are given by the following relations:

$$\sigma_{ij} = E_{ijkl}\varepsilon_{kl}^{e} \tag{59}$$

$$\overline{\sigma}_{ij} = \overline{E}_{ijkl}\overline{\varepsilon}_{kl}^{e} \tag{60}$$

Performing few manipulations using Eqs. 36, 54, 58, 59, and 60 results in obtaining the pure elastic strain and the elastic damage strain as follows:

$$\overline{\varepsilon}_{ij}^{e} = M(\phi)\varepsilon_{ij}^{e} \tag{61}$$

$$\varepsilon_{ij}^{e} = \frac{1}{\left(M(\phi)\right)^{2}}\varepsilon_{ij}^{E} \tag{62}$$

$$\varepsilon_{ij}^{ed} = \frac{1-\left(M(\phi)\right)^{2}}{\left(M(\phi)\right)^{2}}\varepsilon_{ij}^{E} \tag{63}$$

Equations 62 and 63 are valid regardless of the definition of the effective stress function ($M(\phi)$) for any scalar damage model under the strain additive decomposition assumption and by using the strain energy hypothesis. The portion of the elastic strain and elastic damage strain for the proposed model can be obtained using Eq. 48 as follows:

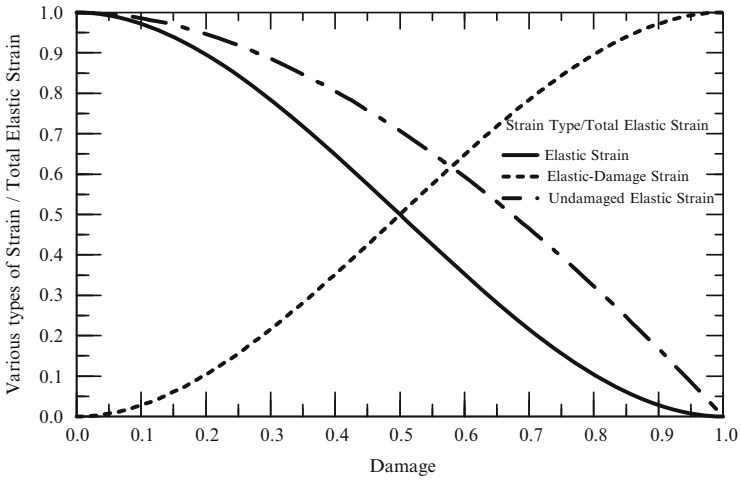


Fig. 6 Variation of the different types of elastic strains with respect to the damage variable (After Voyiadjis and Mozaffari 2013)

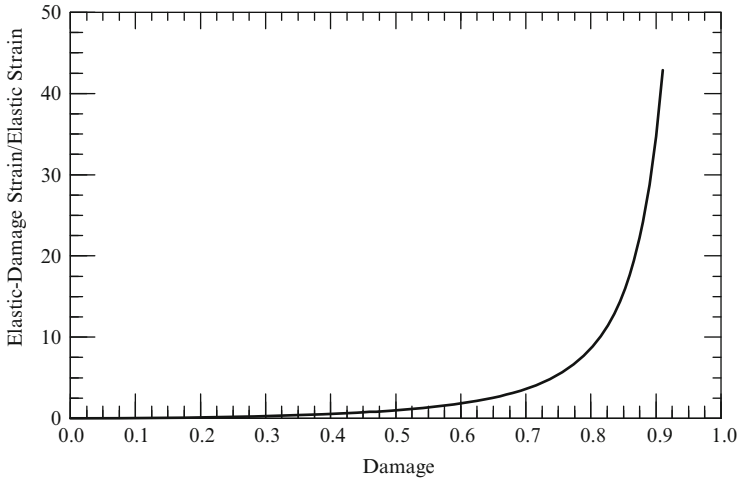


Fig. 7 Variation of elastic damage strain to elastic strain ratio with respect to damage variable (After Voyiadjis and Mozaffari 2013)

$$\varepsilon_{ij}^{e} = (1-\phi)^2(2\phi+1)\varepsilon_{ij}^{E} \tag{64}$$

$$\varepsilon_{ij}^{ed} = \Big(1-(1-\phi)^2(2\phi+1)\Big)\varepsilon_{ij}^{E} \tag{65}$$

The increase of the elastic damage strain, ε_{ij}^{ed}, and elastic strain, ε_{ij}^{e}, is confirmed by Eqs. 64 and 65. These increments are depicted in Figs. 6 and 7.

It is worth mentioning here that in the undamaged configuration, the elastic damage strain is removed; therefore, $\overline{\varepsilon}_{ij}^{ed} = 0$ and $\overline{\varepsilon}_{ij} = \overline{\varepsilon}_{ij}^{e}$.

Thermodynamic Conjugate Force Due to Damage

Consistent thermodynamic framework to describe a phenomenon consists of the internal state variables and the laws of thermodynamics. Each internal state in the thermodynamic consistent framework has its own conjugate forces. Conjugate force due to damage needs to be defined since the damage variable is considered as an internal state variable. This definition can be used to define the damage criterion which is conventionally used in damage models. Several ways have been proposed to define the conjugate damage force which describes the internal effect of the microcracks and microvoids. In this chapter, the effective stress coefficient (48) is used to obtain the damage conjugate force through the hypothesis of strain energy equivalence (Voyiadjis and Kattan 1999). The details of this formulation are given in the work by the authors Voyiadjis and Mozaffari (2013):

$$Y(\phi) = -\left(\frac{M\sigma^2}{\overline{E}}\right)\frac{\partial M}{\partial \phi} \tag{66}$$

Using Eq. 48 in Eq. 66 leads to

$$Y(\phi) = -\overline{E}\overline{\varepsilon}^2 \frac{3\phi}{(1-\phi)(2\phi+1)} \tag{67}$$

Damage Criterion

Damage initiation for uniaxial scalar damage models without the effect of kinematic hardening is checked by the proposed damage criterion which is given as follows:

$$F_d(Y,\phi) = \frac{1}{2}Y^2 - \left(l_d + q\phi_{eq}\right) \tag{68}$$

where ϕ_{eq} is the accumulated damage and in the case of scalar damage is as follows:

$$\phi_{eq} = \sqrt{\int_0^t \phi^2 dt} \tag{69}$$

In Eq. 68, Y is the thermodynamic conjugate force due to damage, q is the damage hardening modulus, and l_d is the initial damage threshold. Damage

evolution is possible if both of the following conditions hold simultaneously based on the Kuhn–Tucker conditions:

$$F_d = 0 \quad \text{and} \quad \frac{\partial F_d}{\partial Y}\dot{Y} > 0 \tag{70}$$

Boundary Conditions

In phase field modeling, directional derivative of order parameter is assumed to be equal to zero as the boundary condition. Since the order parameter is related to the damage parameter by Eq. 18, this boundary condition can be written as follows:

$$\frac{\partial \phi}{\partial n} = 0 \tag{71}$$

in which n indicates the normal vector to the boundary.

Numerical Aspects, Algorithm, and 1D Implementation

In this section, numerical aspects of the proposed damage evolution law are considered and given examples show the validity of the model. The proposed nonlocal damage evolution law (Eq. 28) is able to attribute both elastic and inelastic damages using appropriate constants for each loading region. Both rate-dependent and rate-independent materials can be modeled through this formulation. The model used here incorporates the solution of the Allen–Cahn-type equation (Bates et al. 2009; Choi et al. 2009; Del Pino et al. 2010; Feng and Prohl 2003; Kassam and Trefethen 2005; Shen and Yang 2010). For some problems, the semi-implicit Fourier spectral method is used (Chen and Shen 1998; Feng et al. 2006). In the following sections, numerical details are derived using the finite difference method in order to solve Eq. 28. The numerical algorithm is constructed and used to solve some simple uniaxial examples.

Numerical Aspects

Various types of finite difference schemes are detailed here to solve the numerical examples. Damage criterion is checked through the algorithm in order to show the damage evolution. The proposed model is presented as a special case of the well-known equation (Allen–Cahn equation) with the reaction (nonlinear) term. In the following expressions superscripts show the time step and subscripts show the position in the domain. Superscript n shows the previous step and superscript $n + 1$ shows the current step. Discretization over time should be explicit, but spatial discretization can be implicit or explicit. Readers are referred to the work by the authors Voyiadjis and Mozaffari (2013) for more details.

Explicit in Space, Explicit in Time

Numerical solution of Eq. 28 can be obtained using the finite difference method. Time step Δt is restricted by the CFL (Courant et al. 1928) condition (convergence of any explicit method is guaranteed by this condition) in order to solve the equation with an explicit scheme since the equation has a nonlinear term. This condition for the stable convergence in the 1D case reads as

$$M_\phi \overline{E} l^2 \left(\frac{\Delta t}{(\Delta x)^2} \right) \leq \frac{1}{2} \tag{72}$$

where $\Delta t > 0$ is the time step and Δx is the space step in the x direction. The nonlinear term coefficient in general phase field models ($M\overline{E}l^2$ in present formulation) is set to a very small value over the domain. Once an explicit scheme is used, then a relatively large value for the time step (i.e., Δt) can be used. If the CFL condition (72) is held, then the nonlinear term (reaction term) restricts the time step to set it to a small value. Divergence of the solution for variable ϕ occurs because of the nonlinear term especially when its value is outside the interval $0 \leq \phi \leq 1$. To resolve this issue, the forward time terms are introduced through the semi-implicit scheme (Warren et al. 2003). The discretization for the 1D case is obtained as follows:

$$\frac{\phi_i^{n+1} - \phi_i^n}{\Delta t} = -M_\phi \overline{E} l^2 \left(\frac{\phi_{i+1}^n - 2\phi_i^n + \phi_{i-1}^n}{(\Delta x)^2} \right) + \begin{cases} \phi_i^{n+1}(1 - \phi_i^n) r(\phi_i^n) & \text{for } r(\phi_i^n) \leq 0 \\ \phi_i^n (1 - \phi_i^{n+1}) r(\phi_i^n) & \text{for } r(\phi_i^n) > 0 \end{cases} \tag{73}$$

where

$$r(\phi_i^n) = M_\phi \overline{E} \left[3(\overline{\varepsilon}_i^n)^2 + 2w(2\phi_i^n - 1) \right]. \tag{74}$$

Equation 73 along with the definition of function $r(\phi_i^n)$ (Eq. 74) guarantees that ϕ remains in the desired interval $0 \leq \phi \leq 1$ even if the large time step is used. Equation 73 can be computed directly over individual nodes without solving the set of linear equations as follows:

$$\begin{cases} \left[1 - \Delta t(1 - \phi_i^n) r(\phi_i^n)\right] \phi_i^{n+1} = \phi_i^n - M_\phi \overline{E} l^2 \Delta t \left(\dfrac{\phi_{i+1}^n - 2\phi_i^n + \phi_{i-1}^n}{(\Delta x)^2} \right) & \text{for } r(\phi_i^n) \leq 0 \\ \left[1 + \Delta t \phi_i^n r(\phi_i^n)\right] \phi_i^{n+1} = \phi_i^n + \Delta t \phi_i^n r(\phi_i^n) - M_\phi \overline{E} l^2 \Delta t \left(\dfrac{\phi_{i+1}^n - 2\phi_i^n + \phi_{i-1}^n}{(\Delta x)^2} \right) & \text{for } r(\phi_i^n) > 0 \end{cases} \tag{75}$$

Implicit in Space, Explicit in Time

Fully implicit method in space discretization is governed by the same treatment on the reaction term in the Allen–Cahn equation (28) as follows:

$$\frac{\phi_i^{n+1}-\phi_i^n}{\Delta t}=-M_\phi\overline{E}l^2\left(\frac{\phi_{i+1}^{n+1}-2\phi_i^{n+1}+\phi_{i-1}^{n+1}}{(\Delta x)^2}\right)+\begin{cases}\phi_i^{n+1}(1-\phi_i^n)r(\phi_i^n) & \text{for } r(\phi_i^n)\le 0\\ \phi_i^n(1-\phi_i^{n+1})r(\phi_i^n) & \text{for } r(\phi_i^n)>0\end{cases} \tag{76}$$

Defining $A=\frac{-M_\phi\overline{E}l^2\Delta t}{(\Delta x)^2}$ and substituting in Eq. 76 lead to

$$\phi_i^{n+1}-\phi_i^n=A\left(\phi_{i+1}^{n+1}-2\phi_i^{n+1}+\phi_{i-1}^{n+1}\right)+\begin{cases}\Delta t\phi_i^{n+1}(1-\phi_i^n)r(\phi_i^n) & \text{for } r(\phi_i^n)\le 0\\ \Delta t\phi_i^n(1-\phi_i^{n+1})r(\phi_i^n) & \text{for } r(\phi_i^n)>0\end{cases} \tag{77}$$

The two coefficients $B_i=\Delta t(1-\phi_i^n)r(\phi_i^n)$ and $C_i=\Delta t\phi_i^n r(\phi_i^n)$ are defined to simplify Eq. 77:

$$\begin{cases}\phi_i^{n+1}-A\left(\phi_{i+1}^{n+1}-2\phi_i^{n+1}+\phi_{i-1}^{n+1}\right)-B_i\phi_i^{n+1}=\phi_i^n & \text{for } r(\phi_i^n)\le 0\\ \phi_i^{n+1}-A\left(\phi_{i+1}^{n+1}-2\phi_i^{n+1}+\phi_{i-1}^{n+1}\right)+C_i\phi_i^{n+1}=\phi_i^n+C_i & \text{for } r(\phi_i^n)>0\end{cases} \tag{78}$$

Equation 78 is rewritten in the following form by separating known terms from the previous step:

$$\begin{cases}-A\phi_{i+1}^{n+1}+(1+2A-B_i)\phi_i^{n+1}-A\phi_{i-1}^{n+1}=\phi_i^n & \text{for } r(\phi_i^n)\le 0\\ -A\phi_{i+1}^{n+1}+(1+2A+C_i)\phi_i^{n+1}-A\phi_{i-1}^{n+1}=\phi_i^n+C_i & \text{for } r(\phi_i^n)>0\end{cases} \tag{79}$$

Equation 79 can be written as follows in order to find the constant finite difference coefficients. Additional term can be added to the specific row of the FD matrix which is related to the node based on the sign of the function $r(\phi)$ at the desired node:

$$\begin{cases}-A\phi_{i+1}^{n+1}+(1+2A)\phi_i^{n+1}-A\phi_{i-1}^{n+1}-B_i\phi_i^{n+1}=\phi_i^n & \text{for } r(\phi_i^n)\le 0\\ -A\phi_{i+1}^{n+1}+(1+2A)\phi_i^{n+1}-A\phi_{i-1}^{n+1}+C_i\phi_i^{n+1}=\phi_i^n+C_i & \text{for } r(\phi_i^n)>0\end{cases} \tag{80}$$

Applying the boundary conditions ($\phi_0^{n+1}=\phi_1^{n+1}$ and $\phi_{n-1}^{n+1}=\phi_n^{n+1}$) results in the given matrix representation which is derived for Eq. 80 over all the nodes:

$$\begin{bmatrix} 1 & -1 & \cdots & 0 & \\ -A & 1+2A & & & \\ \vdots & & \ddots & \vdots & \\ 0 & & \cdots & 1+2A & -A \\ & & & -1 & 1 \end{bmatrix} \begin{bmatrix} \phi_0^{n+1} \\ \vdots \\ \phi_n^{n+1} \end{bmatrix} + [\text{Matrix } B \text{ or Matrix } C] \begin{bmatrix} \phi_0^{n+1} \\ \vdots \\ \phi_n^{n+1} \end{bmatrix} = \begin{bmatrix} \phi_0^{n+1} \\ \vdots \\ \phi_n^{n+1} \end{bmatrix} + [\text{Vector } C] \tag{81}$$

where the diagonal matrices B and C are defined as

$$[B] = \begin{bmatrix} 1 & -1 & \cdots & . & \\ 0 & B_1 & & & \\ \vdots & & \ddots & \vdots & \\ . & & \cdots & B_{n-1} & 0 \\ & & & -1 & 1 \end{bmatrix} \tag{82}$$

$$[C] = \begin{bmatrix} 1 & -1 & \cdots & . & \\ 0 & C_1 & & & \\ \vdots & & \ddots & \vdots & \\ . & & \cdots & C_{n-1} & 0 \\ & & & -1 & 1 \end{bmatrix} \tag{83}$$

and vector C is written as

$$[C] = \begin{bmatrix} 0 \\ C_1 \\ . \\ C_{n-1} \\ 0 \end{bmatrix} \tag{84}$$

Implicit in Space, Explicit in Time (with Crank–Nicolson Scheme in Space)

The Crank–Nicolson scheme is a well-known unconditionally stable method and is adopted just on the terms related to the derivatives in space. Reaction term is treated in the same way as the semi-implicit scheme. Crank–Nicolson discretization of the governing equation over the space results in

$$\frac{\phi_i^{n+1} - \phi_i^n}{\Delta t} = -M_\phi \overline{El}^2 \left(\frac{1}{2} \left(\frac{\phi_{i+1}^{n+1} - 2\phi_i^{n+1} + \phi_{i-1}^{n+1}}{(\Delta x)^2} \right) + \frac{1}{2} \left(\frac{\phi_{i+1}^{n} - 2\phi_i^{n} + \phi_{i-1}^{n}}{(\Delta x)^2} \right) \right) + \begin{cases} \phi_i^{n+1}(1-\phi_i^n) r(\phi_i^n) & \text{for } r(\phi_i^n) \le 0 \\ \phi_i^{n}(1-\phi_i^{n+1}) r(\phi_i^n) & \text{for } r(\phi_i^n) > 0 \end{cases} \tag{85}$$

Defining a coefficient $E = \frac{-M_\phi \overline{E}l^2 \Delta t}{2(\Delta x)^2}$ and using it into Eq. 85 lead to the following equation:

$$\begin{aligned}\phi_i^{n+1} - \phi_i^n = E\left(\phi_{i+1}^{n+1} - 2\phi_i^{n+1} + \phi_{i-1}^{n+1}\right) + E\left(\phi_{i+1}^n - 2\phi_i^n + \phi_{i-1}^n\right) \\ + \begin{cases} \Delta t\phi_i^{n+1}\left(1-\phi_i^n\right)r\left(\phi_i^n\right) & \text{for } r\left(\phi_i^n\right) \le 0 \\ \Delta t\phi_i^n\left(1-\phi_i^{n+1}\right)r\left(\phi_i^n\right) & \text{for } r\left(\phi_i^n\right) > 0 \end{cases}\end{aligned} \tag{86}$$

Using the previously defined two coefficients $B_i = \Delta t(1 - \phi_i^n)r(\phi_i^n)$ and $C_i = \Delta t\phi_i^n r(\phi_i^n)$, Eq. 86 can be written as

$$\begin{cases} \phi_i^{n+1} - E\left(\phi_{i+1}^{n+1} - 2\phi_i^{n+1} + \phi_{i-1}^{n+1}\right) - B_i\phi_i^{n+1} = E\left(\phi_{i+1}^n - 2\phi_i^n + \phi_{i-1}^n\right) + \phi_i^n & \text{for } r\left(\phi_i^n\right) \le 0 \\ \phi_i^{n+1} - E\left(\phi_{i+1}^{n+1} - 2\phi_i^{n+1} + \phi_{i-1}^{n+1}\right) + C_i\phi_i^{n+1} = E\left(\phi_{i+1}^n - 2\phi_i^n + \phi_{i-1}^n\right) + \phi_i^n + C_i & \text{for } r\left(\phi_i^n\right) > 0 \end{cases} \tag{87}$$

Equation 87 can be rearranged easily as follows:

$$\begin{cases} -E\phi_{i+1}^{n+1} + (1 + 2E - B_i)\phi_i^{n+1} - E\phi_{i-1}^{n+1} = E\phi_{i+1}^n + \left(1-2E\right)\phi_i^n + E\phi_{i-1}^n & \text{for } r\left(\phi_i^n\right) \le 0 \\ -E\phi_{i+1}^{n+1} + (1 + 2E + C_i)\phi_i^{n+1} - E\phi_{i-1}^{n+1} = E\phi_{i+1}^n + \left(1-2E\right)\phi_i^n + E\phi_{i-1}^n + C_i & \text{for } r\left(\phi_i^n\right) > 0 \end{cases} \tag{88}$$

Equation 88 can be written as follows in order to find the constant finite difference coefficients. Additional term can be added to the specific row of the FD matrix which is related to the node based on the sign of the function $r(\phi)$ at the desired node like the implicit scheme:

$$\begin{cases} -E\phi_{i+1}^{n+1} + (1 + 2E)\phi_i^{n+1} - E\phi_{i-1}^{n+1} - B_i\phi_i^{n+1} = E\phi_{i+1}^n + \left(1-2E\right)\phi_i^n + E\phi_{i-1}^n & \text{for } r\left(\phi_i^n\right) \le 0 \\ -E\phi_{i+1}^{n+1} + (1 + 2E)\phi_i^{n+1} - E\phi_{i-1}^{n+1} + C_i\phi_i^{n+1} = E\phi_{i+1}^n + \left(1-2E\right)\phi_i^n + E\phi_{i-1}^n + C_i & \text{for } r\left(\phi_i^n\right) > 0 \end{cases} \tag{89}$$

Applying the boundary conditions ($\phi_0^{n+1} = \phi_1^{n+1}$ and $\phi_{n-1}^{n+1} = \phi_n^{n+1}$) results in the given matrix representation which is derived for Eq. 89 over all the nodes:

$$\begin{aligned}&\begin{bmatrix} 1 & -1 & \cdots & 0 & \\ -E & 1+2E & & & \\ \vdots & & \ddots & \vdots & \\ 0 & & \cdots & 1+2E & -E \\ & & & -1 & 1 \end{bmatrix}\begin{bmatrix} \phi_0^{n+1} \\ \vdots \\ \phi_n^{n+1} \end{bmatrix} + [\text{Matrix } B \text{ or Matrix } C]\begin{bmatrix} \phi_0^{n+1} \\ \vdots \\ \phi_n^{n+1} \end{bmatrix} \\ &= \begin{bmatrix} 1 & -1 & \cdots & 0 & \\ -E & 1+2E & & & \\ \vdots & & \ddots & \vdots & \\ 0 & & \cdots & 1+2E & -E \\ & & & -1 & 1 \end{bmatrix}\begin{bmatrix} \phi_0^n \\ \vdots \\ \phi_n^n \end{bmatrix} + [\text{Vector } C]\end{aligned} \tag{90}$$

The diagonal matrices B and C and vector C are given in Eqs. 82, 83, and 84, respectively.

Rate-Independent Material

For this case, the time increment is set equal to one ($\Delta t = 1$). Equation 73 can then be considered as a damage increment for a specific node:

$$\Delta\phi_i^n = -M_\phi \overline{E} l^2 \left(\frac{\phi_{i+1}^n - 2\phi_i^n + \phi_{i-1}^n}{(\Delta x)^2} \right) + \begin{cases} \phi_i^{n+1}\left(1 - \phi_i^n\right) r\left(\phi_i^n\right) & \text{for } r\left(\phi_i^n\right) \le 0 \\ \phi_i^n\left(1 - \phi_i^{n+1}\right) r\left(\phi_i^n\right) & \text{for } r\left(\phi_i^n\right) > 0 \end{cases} \tag{91}$$

Therefore, the updated damage level reads

$$\phi_i^{n+1} = \phi_i^n + \Delta\phi_i^n. \tag{92}$$

Rate-Dependent Material

The time increment is not equal to unity in this case. Furthermore, Eq. 73 can be used to obtain the damage level at the current step using Eq. 75. Damage level will be updated simultaneously if the implicit method or Crank–Nicolson method is used at each increment over all nodes and it will be updated separately if the explicit method is used. Therefore, the implicit method or Crank–Nicolson schemes can give better results since they update the damage level over all nodes of the domain at the same time. This is important in specific problems when there is a great change in the stiffness. In this case, Eqs. 81 and 90 can be used to update the damage level over all nodes.

Numerical Algorithm

In order to solve Eq. 28 in 1D domain along with satisfying the boundary conditions, the following algorithm is constructed. This algorithm is used to solve a stress-driven problem using the finite difference schemes which are detailed in the previous sections. In the following algorithm, superscript $n + 1$ indicates the current load step and superscript n indicates the previous load step. Therefore, for time-dependent problem the stress at the current step reads as

$$\sigma^{n+1} = \sigma^n + \Delta t \dot{\sigma}$$

and for the case of time-independent problem, it can be written as follows:

$$\sigma^{n+1} = \sigma^n + \Delta\sigma$$

where $\dot{\sigma}$ is the stress loading rate and $\Delta\sigma$ is the stress increment. This algorithm is given exactly the same here as in the work by the authors Voyiadjis and Mozaffari (2013):

1. Initialize values of M, l, w, $\overline{E}, \sigma_0$, $\Delta\sigma$, ϕ_{cr}, and $\phi_0 = 0.001$ for all nodes.
2. Set $E^n = \overline{E}^n = \overline{E}, \sigma^n = \sigma^{n+1} = \sigma_0, \overline{\sigma}^n = \overline{\sigma}^{n+1} = \sigma_0,$ and $\phi^n = \phi_0$ at all nodes.

3. Compute $\varepsilon^n = \frac{\sigma^n}{E^n}$, $\varepsilon^{n+1} = \frac{\sigma^{n+1}}{E^{n+1}}$, $\overline{\varepsilon}^n = \frac{\overline{\sigma}^n}{\overline{E}^n}$, and $\overline{\varepsilon}^{n+1} = \frac{\overline{\sigma}^{n+1}}{\overline{E}^{n+1}}$.
4. Iterate the following steps until $\phi_{\max} < \phi_{cr}$.
5. Update the load level $\sigma^{n+1} = \sigma^n + \Delta\sigma$.
6. Compute Y^n and F_d^n using Eqs. 67 and 68.
7. Check the damage criterion (Eq. 70) at each node.
   If $F_d < 0$, then damage does not evolve and $\phi^{n+1} = \phi^n$.
   If $F_d > 0$, compute damage level ϕ^{n+1} with the desired scheme (Eqs. 75 and 81 or Eq. 90).
8. Compute ε^{n+1}, E^{n+1}, and $\overline{\varepsilon}^{n+1}$.

Numerical Examples

The dimension of each coefficient M, w, and l needs to be defined to make each term in Eq. 28 dimensionless, since the damage variable is a dimensionless parameter. For the (SI) unified system, the coefficient for mobility of microcracks M has the inverse dimension of the modulus of elasticity ($\frac{m^2}{N}$), the dissipation coefficient w is a dimensionless coefficient, and the specific length scale due to damage (capable in capturing the effects of nonlocal damage) has the length dimension (m). Various values for these coefficients are examined by means of numerical experiments. These are described later and detailed in the examples. Two sets of experiments are designed for a specific material in order to determine the appropriate coefficients. The first set is the usual experiments to evaluate the damage value at a center of a bar. This is in order to obtain the coefficient w that cancels the second-order gradient term in Eq. 28 and its corresponding coefficient (l). Another set of nonlocal experiments is designed to evaluate the damage value at several points along the length of the bar in each increment of loading for the same time. Utilizing this set of data with the determined coefficient (w) from the previous experiments provides the evaluation of the coefficient (l). The M (mobility of microcracks) can be determined by using all these two sets of data. However, it can also be considered exactly equal to the inverse of the value of the modulus of elasticity at the beginning (undamaged material) of the test in order to simplify this procedure.

In order to show the regularization capabilities of the proposed model, a uniaxial bar under tension is considered in this work. Reduction in stiffness is considered in the middle of the bar (L) based on the number of nodes which are used in this specific length (L_D). This is in order to show the nonlocal distribution of damage. The geometry of the bar is shown in Fig. 8. The material properties with the hardening parameters are given by Abu Al-Rub and Voyiadjis (2003) for 30CrNiMo8 high-strength steel: $\overline{E} = 199$ Gpa, $v = 0.3$, $\sigma_{yp} = 870$ Mpa, $q = 8.2$ MPa, and $l_d = 3.8$ MPa. Other specific parameters are $L = 1$ m, $L_D = 0.1$ L, $\sigma_0 = 10$ MPa (initial value of stress), and $\dot{\sigma} = 10\frac{\text{MPa}}{s}$. A small value of damage $\phi_0 = 0.001$ is used aver all nodes to initiate the numerical procedure. The critical value of damage to end the numerical procedure is assumed as $\phi_{cr} = 0.35$ unless it is different for a specific case. The loading is continued up to the yield point.

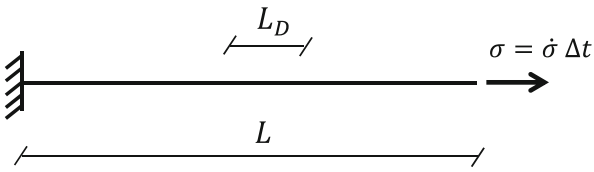


Fig. 8 Geometry of the specimen for the numerical examples (After Voyiadjis and Mozaffari 2013)

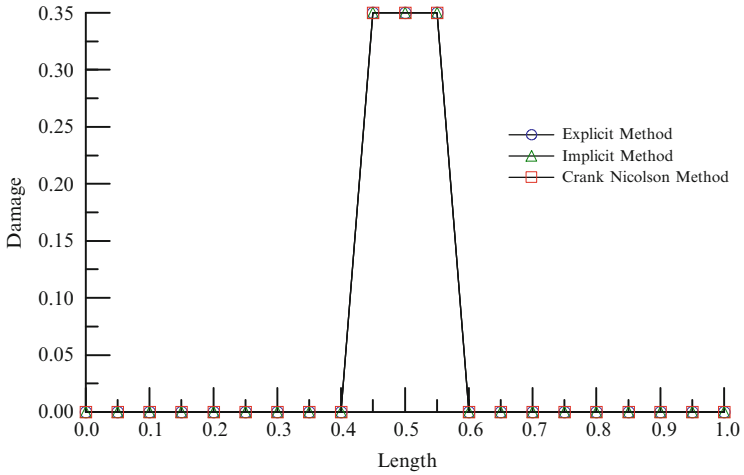


Fig. 9 Damage distribution over the bar with 21 nodes. Using three schemes: explicit, implicit, and Crank–Nicolson (After Voyiadjis and Mozaffari 2013)

The total time of loading is 86 s, and in order to hold the CFL condition and also to keep the points over the damage surface, it is assumed that $\Delta t = 0.001$ s.

Example 1: **Comparison of Different Schemes** Three different schemes along with various numbers of nodes on the length of the sample are examined in this example. The modulus of elasticity on the specific central length (L_D) has 10 % reduction comparing with the modulus of elasticity of the bar, and other coefficients are assumed as $M = 1\ (\text{MPa})^{-1} = 10^{-6}\ \text{m}^2/\text{N}$, $w = 10^{-5}$, *and* $l = 1\ \mu\text{m} = 10^{-6}$ m.

Case 1: 21 nodes over L, 3 nodes over L_D The numerical results are depicted in Fig. 9 for three different computational schemes. It can be seen that there is complete coincidence in the solution using the three different schemes.

Case 2: 41 nodes over L, 5 nodes over L_D The results are depicted in Fig. 10 for three different FD schemes. It can be seen that there is complete coincidence in the solution using the three different schemes.

Case 3: 81 nodes over L, 9 nodes over L_D The results are shown in Fig. 11 for three different FD schemes. It can be seen that there is complete coincidence in the solution using the three different schemes.

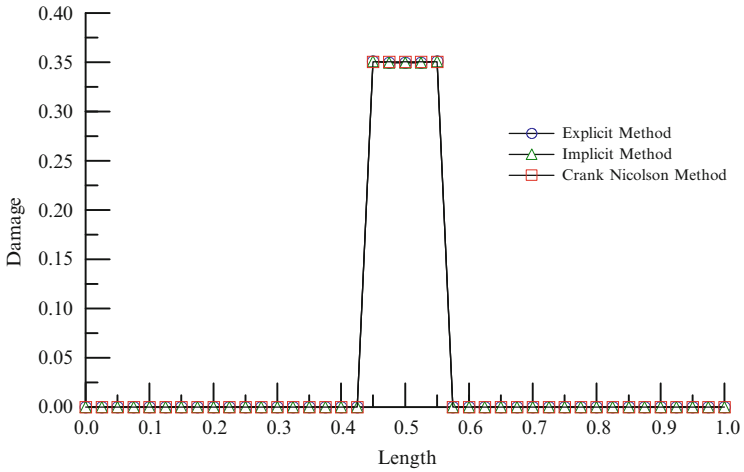


Fig. 10 Damage distribution over the bar with 41 nodes. Using three schemes: explicit, implicit, and Crank–Nicolson (After Voyiadjis and Mozaffari 2013)

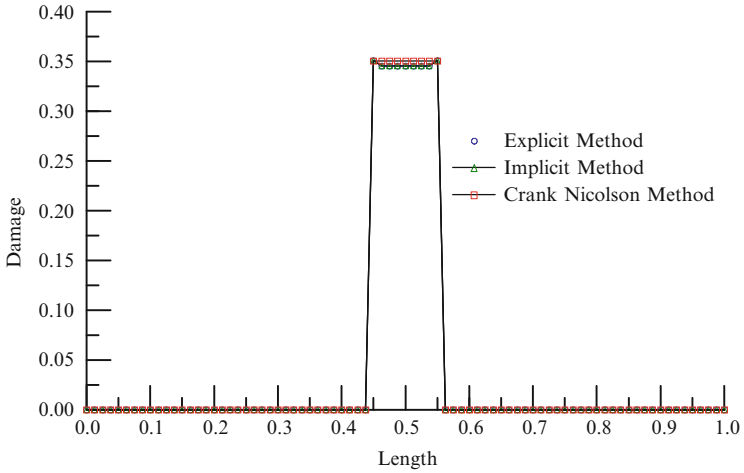


Fig. 11 Damage distribution over the bar with 81 nodes. Using three schemes: explicit, implicit, and Crank–Nicolson (After Voyiadjis and Mozaffari 2013)

Stress change trend versus damage value is shown by performing a simulation using the explicit method, but the final damage value is set to one (theoretical limit value of damage variable). Numerical results are depicted in Figs. 12 and 13.

This example shows that all three FD schemes give the same result regardless of the number of nodes. In the last case, a slight difference between the Crank–Nicolson scheme and the other two schemes can be seen. It is possible to use large time steps for both implicit and Crank–Nicolson schemes as there is no

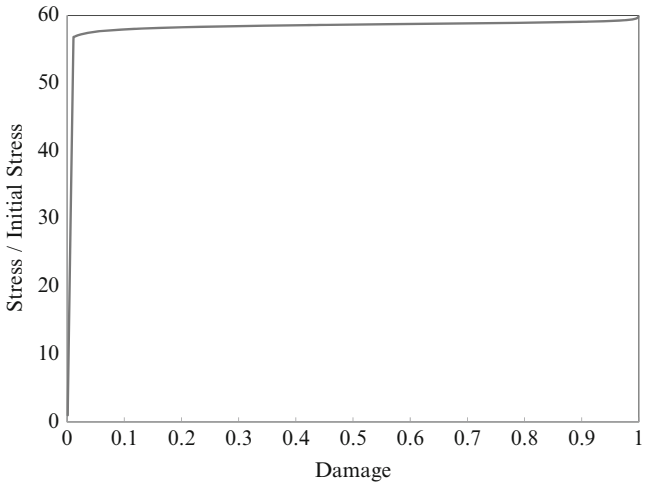


Fig. 12 Variation of stress at the central point with respect to damage (After Voyiadjis and Mozaffari 2013)

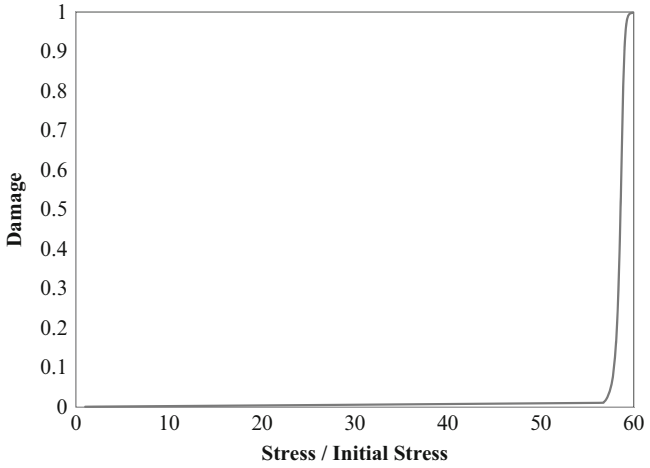


Fig. 13 Variation of damage at the central point with respect to stress (After Voyiadjis and Mozaffari 2013)

need to check the CFL condition, since they are unconditionally stable schemes. Also, these two schemes as it is mentioned before in section "Numerical Aspects" of this chapter update the damage level over all nodes simultaneously. This can be considered in specific problems in which there is a great change in the stiffness of the domain. Stress variation with respect to damage follows the trend like other types of damage models.

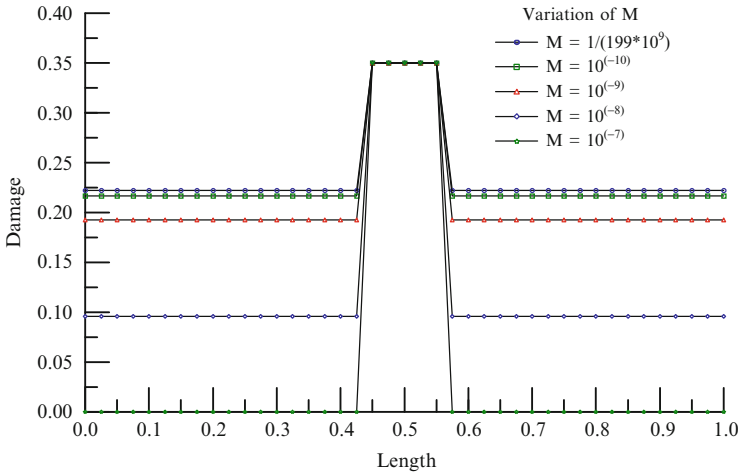


Fig. 14 Influence of various values for mobility of microcrack coefficient M on damage level for all nodes (After Voyiadjis and Mozaffari 2013)

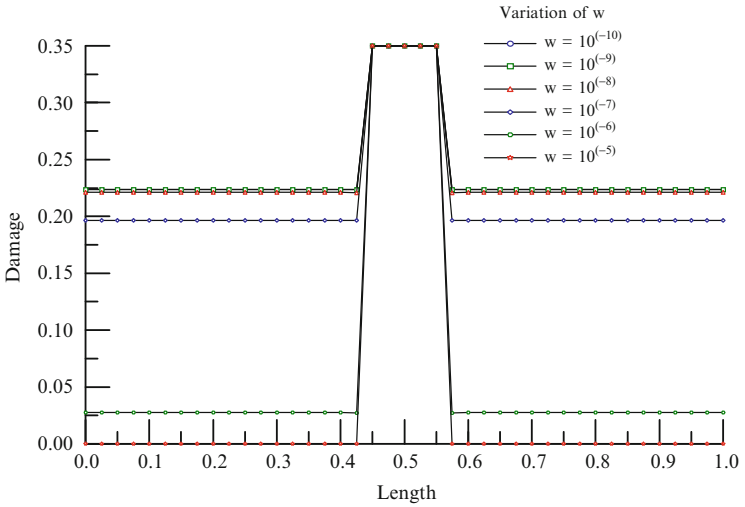


Fig. 15 Influence of various values for dissipation coefficient w on damage level for all nodes (After Voyiadjis and Mozaffari 2013)

Example 2: **Influence of Microcrack Mobility Constant (M)** The dissipation coefficient and length scale are assumed constant and equal to $w = 10^{-5}$, $l = 1$ μm $= 10^{-6}$ m in this example. Different values are examined for the coefficient M and superimposed results are shown in Fig. 14 using the explicit scheme with 41 nodes along the bar.

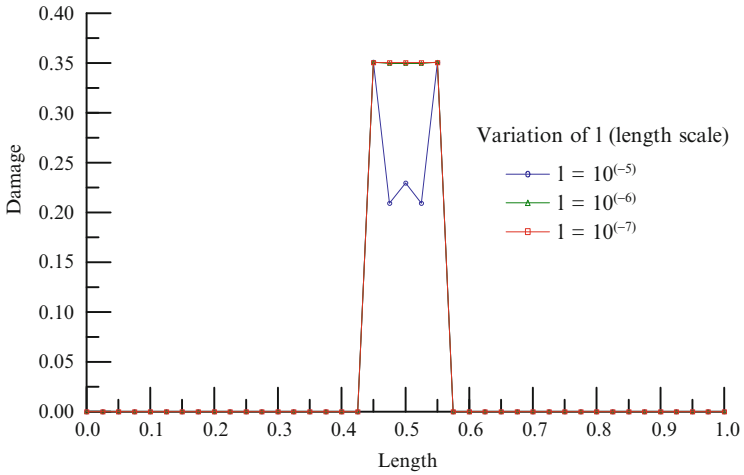


Fig. 16 Influence of various values for length scale l on damage level for all nodes (After Voyiadjis and Mozaffari 2013)

Example 3: **Influence of Dissipation Coefficient Constant (w)** The mobility of microcrack coefficient and length scale are assumed constant and equal to $M = 10^{-6}$, $l = 1\ \mu m = 10^{-6}$ m in this example. Different values are examined for the coefficient w and superimposed results are shown in Fig. 15 using the explicit scheme with 41 nodes along the bar. It can be seen that increasing the coefficient M results in decreasing the damage level on neighbor nodes.

Example 4: **Influence of Characteristic Length Coefficient (Length Scale) Constant (l)** The mobility of microcrack coefficient and dissipation coefficient are assumed constant and equal to $M = 10^{-6}$, $w = 10^{-5}$ in this example. Different values are examined for the coefficient l and superimposed results are shown in Fig. 16 using explicit scheme with 41 nodes along the bar.

It was observed that the greater value for l affects the damage value at the points with great change in stiffness (see the drop in Fig. 16 at length = 0.55) which is obvious based on the nature of length scale due to damage. Localization effects can be obtained using this parameter gradient-type models.

Conclusions

In this chapter, the application of the PFM into damage mechanics is shown. Considering damage as phase transformation leads to developing a nonlocal gradient-type damage model for ductile materials using energy minimization procedure of the PFM approach. A unique Allen–Cahn-type partial differential equation (28) as a damage evolution law is proposed. Numerical examples are used to show regularization capabilities of the proposed model. Although only elastic case

is considered here, a specific set of constants M, w, and l can be used to extend this model to the inelastic region by adding a continuum plasticity model. The effects of the material constants M (mobility of microcracks), w (dissipation constant), and l (length scale due to damage) are demonstrated in this work. Damage in both rate-dependent and rate-independent materials can be simulated using this model. It is capable of modeling viscodamage by incorporating viscoelastic or viscoplastic behavior in materials. It is shown that unconditionally stable methods like Crank–Nicholson and implicit schemes are more powerful to solve Eq. 28 as they can update the damage level over all nodes simultaneously. They are capable in involving other node damage levels in order to show the effects of gradient of damage as it is detailed here.

The effectiveness and uniqueness of this approach in damage mechanics are summarized below.

A new physically based damage evolution law is proposed through the phase field method. This is accomplished through the incorporation of the damage variable and order parameter which is general and can be used for the simulation of isotropic damage in any kind of material including rate dependent or rate independent without using regular normality rule (Eq. 28). A new implicit definition of damage variable (Eq. 51) is presented which is compared to the conventional definition (Fig. 5). The new damage effect coefficient (Eq. 48) is capable through the mathematical restrictions of the phase field theory to transform stresses and strains into the fictitious undamaged configuration. Various FD schemes are used to show the capability of the model in various coupled damage plasticity models for materials regardless of rate dependency.

References

R.K. Abu Al-Rub, G.Z. Voyiadjis, On the coupling of anisotropic damage and plasticity models for ductile materials. Int. J. Solids Struct. **40**, 2611–2643 (2003)

S.M. Allen, J.W. Cahn, A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. Acta Metall. **27**, 1085–1095 (1979)

H. Amor, J.-J. Marigo, C. Maurini, Regularized formulation of the variational brittle fracture with unilateral contact: numerical experiments. J. Mech. Phys. Solids **57**, 1209–1229 (2009)

I. Aranson, V. Kalatsky, V. Vinokur, Continuum field description of crack propagation. Phys. Rev. Lett. **85**, 118–121 (2000)

P.W. Bates, S. Brown, J. Han, Numerical analysis for a nonlocal Allen-Cahn equation. Num. Anal. Model. **6**, 33–49 (2009)

W. Boettinger, J. Warren, C. Beckermann, A. Karma, Phase-field simulation of solidification 1. Annu. Rev. Mater. Res. **32**, 163–194 (2002)

M.J. Borden, C.V. Verhoosel, M.A. Scott, T.J. Hughes, C.M. Landis, A phase-field description of dynamic brittle fracture. Comput. Methods Appl. Mech. Eng. **217**, 77–95 (2012)

G. Boussinot, Y. Le Bouar, A. Finel, Phase-field simulations with inhomogeneous elasticity: comparison with an atomic-scale method and application to superalloys. Acta Mater. **58**, 4170–4181 (2010)

J. Cahn, S. Allen, A microscopic theory for domain wall motion and its experimental verification in Fe-Al alloy domain growth kinetics. Le J. de Phys. Colloques **38**, 7–7 (1977)

J.W. Cahn, J.E. Hilliard, Free energy of a nonuniform system. I. Interfacial free energy. J. Chem. Phys. **28**, 258 (1958)

P.R. Cha, D.H. Yeon, J.K. Yoon, A phase field model for isothermal solidification of multicomponent alloys. Acta Mater. **49**, 3295–3307 (2001)
L.Q. Chen, Phase-field models for microstructure evolution. Annu. Rev. Mater. Res. **32**, 113–140 (2002)
L. Chen, J. Shen, Applications of semi-implicit Fourier-spectral method to phase field equations. Comput. Phys. Commun. **108**, 147–158 (1998)
J.W. Choi, H.G. Lee, D. Jeong, J. Kim, An unconditionally gradient stable numerical method for solving the Allen–Cahn equation. Phys.Statist. Mech. Appl. **388**, 1791–1803 (2009)
R. Courant, K. Friedrichs, H. Lewy, Über die partiellen Differenzengleichungen der mathematischen Physik. Math. Ann. **100**, 32–74 (1928)
M. Del Pino, M. Kowalczyk, F. Pacard, J. Wei, Multiple-end solutions to the Allen–Cahn equation in R2. J. Funct. Anal. **258**, 458–503 (2010)
R.J. Dorgan, G.Z. Voyiadjis, Nonlocal coupled damage-plasticity model incorporating functional forms of hardening state variables. AIAA J. **45**, 337–346 (2007)
K. Elder, N. Provatas, *Phase-Field Methods in Materials Science and Engineering*, 1st edn. (Wiley-VCH, Weinheim, 2010)
D. Fan, L.Q. Chen, Computer simulation of grain growth using a continuum field model. Acta Mater. **45**, 611–622 (1997)
X. Feng, A. Prohl, Numerical analysis of the Allen-Cahn equation and approximation for mean curvature flows. Numer. Math. **94**, 33–65 (2003)
W. Feng, P. Yu, S. Hu, Z. Liu, Q. Du, L. Chen, Spectral implementation of an adaptive moving mesh method for phase-field equations. J. Comput. Phys. **220**, 498–510 (2006)
Y. Furukawa, K. Nakajima, *Advances in Crystal Growth Research*. (Elsevier Science, Amsterdam, The Netherlands, 2001)
A. Gaubert, Y. Le Bouar, A. Finel, *Coupling Phase Field and Visco-Plasticity to Study Rafting in Ni-Base Superalloys*. (Philosophical Magazine 90, 2010), pp. 375–404
V. Ginzburg, L.D. Landau, On the theory of superconductivity, Zh. Eksp. Teor. Fiz. **20**, 1064–1082 (1950). Translation in Collected papers of L.D.Landau. (Pergamon, Oxford, 1965)
L. Gránásy, T. Pusztai, J.A. Warren, Modelling polycrystalline solidification using phase field theory. J. Phys. Condens. Matter **16**, R1205 (2004)
X. Guo, S.Q. Shi, X. Ma, Elastoplastic phase field model for microstructure evolution. Appl. Phys. Lett. **87**, 221910–221910–221913 (2005)
X. Guo, S. Shi, Q. Zhang, X. Ma, An elastoplastic phase-field model for the evolution of hydride precipitation in zirconium. Part I: smooth specimen. J. Nuclear Mater. **378**, 110–119 (2008)
M.E. Gurtin, Generalized Ginzburg-Landau and Cahn-Hilliard equations based on a microforce balance. Phys. Nonlinear Phenomena **92**, 178–192 (1996)
S. Hu, L. Chen, A phase-field model for evolving microstructures with strong elastic inhomogeneity. Acta Mater. **49**, 1879–1890 (2001)
S. Hu, M. Baskes, M. Stan, Phase-field modeling of microvoid evolution under elastic-plastic deformation. Appl. Phys. Lett. **90**, 081921–081921–081923 (2007)
A. Hunter, M. Koslowski, Direct calculations of material parameters for gradient plasticity. J. Mech. Phys. Solids **56**, 3181–3190 (2008)
L.M. Kachanov, On the creep fracture time. Izv Akad. Nauk USSR Otd. Tekh **26–31** (1958)
A. Karma, Phase-field formulation for quantitative modeling of alloy solidification. Phys. Rev. Lett. **87**, 115701 (2001)
A. Karma, D.A. Kessler, H. Levine, Phase-field model of mode III dynamic fracture. Phys. Rev. Lett. **87**, 45501 (2001)
A.K. Kassam, L.N. Trefethen, Fourth-order time-stepping for stiff PDEs. SIAM J. Sci. Comput. **26**, 1214–1233 (2005)
A.G. Khachaturyan, *Theory of Structural Transformations in Solids* (Wiley, New York, 1983)
M. Koslowski, A.M. Cuitino, M. Ortiz, A phase-field theory of dislocation dynamics, strain hardening and hysteresis in ductile single crystals. J. Mech. Phys. Solids **50**, 2597–2635 (2002)

C. Kuhn, R. Müller, A continuum phase field model for fracture. Eng. Fract. Mech. **77**, 3625–3634 (2010)
J. Lemaitre, R. Desmorat, *Engineering Damage Mechanics: Ductile, Creep, Fatigue and Brittle Failures* (Springer, New York, 2005)
V.I. Levitas, I.B. Ozsoy, Micromechanical modeling of stress-induced phase transformations. part 1. thermodynamics and kinetics of coupled interface propagation and reorientation. Int. J. Plast. **25**, 239–280 (2009a)
V.I. Levitas, I.B. Ozsoy, Micromechanical modeling of stress-induced phase transformations. Part 2. computational algorithms and examples. Int. J. Plast. **25**, 546–583 (2009b)
V.I. Levitas, D.L. Preston, Three-dimensional Landau theory for multivariant stress-induced martensitic phase transformations. I. Austenite↔ martensite. Phys. Rev. B **66**, 134206 (2002)
C. Miehe, M. Hofacker, F. Welschinger, A phase field model for rate-independent crack propagation: Robust algorithmic implementation based on operator splits. Comput. Methods Appl. Mech. Eng. **199**, 2765–2778 (2010a)
C. Miehe, F. Welschinger, M. Hofacker, Thermodynamically consistent phase-field models of fracture: variational principles and multi-field FE implementations. Int. J. Numer. Methods Eng. **83**, 1273–1311 (2010b)
N. Moelans, A quantitative and thermodynamically consistent phase-field interpolation function for multi-phase systems. Acta Mater. **59**, 1077–1086 (2011)
N. Moelans, B. Blanpain, P. Wollants, An introduction to phase-field modeling of microstructure evolution. Calphad **32**, 268–294 (2008)
E. Nauman, N.P. Balsara, Phase equilibria and the Landau–Ginzburg functional. Fluid Phase Equilib. **45**, 229–250 (1989)
S. Nemat-Nasser, Decomposition of strain measures and their rates in finite deformation elastoplasticity. Int. J. Solids Struct. **15**, 155–166 (1979)
S. Nemat-Nasser, On Finite Plastic Flow of Crystalline Solids and Geomaterials. J. Appl. Mech. **50** (4b), 1114 (1983)
N. Ofori-Opoku, N. Provatas, A quantitative multi-phase field model of polycrystalline alloy solidification. Acta Mater. **58**, 2155–2164 (2010)
M. Ohno, K. Matsuura, Quantitative phase-field modeling for two-phase solidification process involving diffusion in the solid. Acta Mater. **58**, 5749–5758 (2010)
A. Onuki, Ginzburg-Landau approach to elastic effects in the phase separation of solids. J. Phys. Soc. Jap. **58**, 3065–3068 (1989)
K. Pham, H. Amor, J.J. Marigo, C. Maurini, Gradient damage models and their use to approximate brittle fracture. Int. J. Damage Mech. **20**, 618–652 (2011)
D. Rodney, Y. Le Bouar, A. Finel, Phase field methods and dislocations. Acta Mater. **51**, 17–30 (2003)
T. Sadowski, S. Samborski, Z. Librant, Damage growth in porous ceramics. Key Eng. Mater. **290**, 86–93 (2005)
D. Salac, W. Lu, Controlled nanocrack patterns for nanowires. J. Comput. Theor. Nanosci. **3**, 263–268 (2006)
S. Samborski, T. Sadowski, On the method of damage assessment in porous ceramics, in *Conference Proceedings of 11th Conference on Fracture*, Turin, 2005
R. Sankarasubramanian, Microstructural evolution in elastically-stressed solids: a phase-field simulation. Def. Sci. J. **61**, 383–393 (2011)
J.P. Sethna, *Statistical Mechanics: Entropy, Order Parameters, and Complexity* (Oxford University Press, New York, 2006)
J. Shen, X. Yang, Numerical approximations of allen-cahn and cahn-hilliard equations. Discrete Contin. Dyn. Syst **28**, 1669–1691 (2010)
F. Sidoroff, *Description of Anisotropic Damage Application to Elasticity* (Springer, Berlin, 1981), pp. 237–244
R. Spatschek, D. Pilipenko, C. Müller-Gugenberger, E.A. Brener, Phase field modeling of fracture and composite materials. Phys. Rev. Lett. **96**, 015502 (2006)

R. Spatschek, C. Müller-Gugenberger, E. Brener, B. Nestler, Phase field modeling of fracture and stress-induced phase transitions. Phys. Rev. E. **75**, 066111 (2007)
I. Steinbach, M. Apel, Multi phase field model for solid state transformation with elastic strain. Phys. Nonlinear Phenomena **217**, 153–160 (2006)
I. Steinbach, F. Pezzolla, B. Nestler, M. Seeßelberg, R. Prieler, G. Schmitz, J. Rezende, A phase field concept for multiphase systems. Phys. Nonlinear Phenomena **94**, 135–147 (1996)
T. Uehara, T. Tsujino, N. Ohno, Elasto-plastic simulation of stress evolution during grain growth using a phase field model. J. Cryst. Growth **300**, 530–537 (2007)
G.Z. Voyiadjis, Degradation of elastic modulus in elastoplastic coupling with finite strains. Int. J. Plast. **4**, 335–353 (1988)
G.Z. Voyiadjis, B. Deliktas, A coupled anisotropic damage model for the inelastic response of composite materials. Comput. Methods Appl. Mech. Eng. **183**, 159–199 (2000)
G.Z. Voyiadjis, R.J. Dorgan, Framework using functional forms of hardening internal state variables in modeling elasto-plastic-damage behavior. Int. J. Plast. **23**, 1826–1859 (2007)
G.Z. Voyiadjis, P.I. Kattan, A coupled theory of damage mechanics and finite strain elasto-plasticity–II. Damage and finite strain plasticity. Int. J. Eng. Sci. **28**, 505–524 (1990)
G.Z. Voyiadjis, P.I. Kattan, A plasticity-damage theory for large deformation of solids–I. Theoretical formulation. Int. J. Eng. Sci. **30**, 1089–1108 (1992)
G.Z. Voyiadjis, P.I. Kattan, *Advances in Damage Mechanics: Metals and Metal Matrix Composites* (Elsevier, Oxford, ISBN 0-08-043601-3, 1999), p. 542
G.Z. Voyiadjis, P.I. Kattan, *Advances in Damage Mechanics: Metals and Metal Matrix Composites with an Introduction to Fabric Tensors*. (2nd edn.) (Elsevier, Oxford, London, ISBN: 0-08-044688-4, 2006), p. 742
G.Z. Voyiadjis, P.I. Kattan, A new class of damage variables in continuum damage mechanics. J. Eng. Mater. Technol. **134** (2012)
G.Z. Voyiadjis, N. Mozaffari, Nonlocal damage model using the phase field method: theory and applications. Int. J. Solids Struct. **50**, 3136–3151 (2013)
G.Z. Voyiadjis, R.K. Abu Al-Rub, A.N. Palazotto, Thermodynamic framework for coupling of non-local viscoplasticity and non-local anisotropic viscodamage for dynamic localization problems using gradient theory. Int. J. Plast. **20**, 981–1038 (2004)
G.Z. Voyiadjis, B. Deliktas, A.N. Palazotto, Thermodynamically consistent coupled viscoplastic damage model for perforation and penetration in metal matrix composite materials. Compos. Part B **40**, 427–433 (2009)
S.-L. Wang, R. Sekerka, A. Wheeler, B. Murray, S. Coriell, R. Braun, G. McFadden, Thermodynamically-consistent phase-field models for solidification. Phys. Nonlinear Phenomena **69**, 189–200 (1993)
Y. Wang, Y. Jin, A. Cuitino, A. Khachaturyan, Nanoscale phase field microelasticity theory of dislocations: model and 3D simulations. Acta Mater. **49**, 1847–1857 (2001)
Y.U. Wang, Y.M. Jin, A.G. Khachaturyan, Phase field microelasticity theory and modeling of elastically and structurally inhomogeneous solid. J. Appl. Phys. **92**, 1351–1360 (2002)
J.A. Warren, R. Kobayashi, A.E. Lobkovsky, W. Craig Carter, Extending phase field models of solidification to polycrystalline materials. Acta Mater. **51**, 6035–6058 (2003)
A. Yamanaka, T. Takaki, Y. Tomita, Elastoplastic phase-field simulation of self-and plastic accommodations in Cubic–> tetragonal martensitic transformation. Mater. Sci. Eng. A **491**, 378–384 (2008)
P. Yu, S. Hu, L. Chen, Q. Du, An iterative-perturbation scheme for treating inhomogeneous elasticity in phase-field models. J. Comput. Phys. **208**, 34–50 (2005)
N. Zhou, C. Shen, M. Mills, Y. Wang, Contributions from elastic inhomogeneity and from plasticity to [gamma]'rafting in single-crystal Ni-Al. Acta Mater. **56**, 6156–6173 (2008)
J. Zhu, L.Q. Chen, J. Shen, Morphological evolution during phase separation and coarsening with strong inhomogeneous elasticity. Model. Simul. Mater. Sci. Eng. **9**, 499 (2001)

Index

G.Z. Voyiadjis (ed.), *Handbook of Damage Mechanics*,
DOI 10.1007/978-1-4614-5589-9