ASM Handbook®

Volume 22A
Fundamentals of Modeling for Metals Processing

Prepared under the direction of the
ASM International Handbook Committee

D.U. Furrer and **S.L. Semiatin**, Volume Editors

Charles Moosbrugger, Editor — Content Developer
Steve Lampman, Senior Content Developer
Eileen DeGuire, Content Developer
Ann Britton, Editorial Assistant
Madrid Tramble, Senior Production Coordinator
Patty Conti, Production Coordinator
Diane Whitelaw, Production Coordinator
Scott D. Henry, Senior Product Manager — Content Development
Bonnie R. Sanders, Manager of Production
Dave Celznick, Project Manager

Editorial Assistance
Elizabeth Marquard
Buz Riley

ASM INTERNATIONAL

**The Materials
Information Society**

Materials Park, Ohio 44073-0002
www.asminternational.org

This book is a collective effort involving hundreds of technical specialists. It brings together a wealth of information from worldwide sources to help scientists, engineers, and technicians solve current and long-range problems.

Great care is taken in the compilation and production of this Volume, but it should be made clear that NO WARRANTIES, EXPRESS OR IMPLIED, INCLUDING, WITHOUT LIMITATION, WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, ARE GIVEN IN CONNECTION WITH THIS PUBLICATION. Although this information is believed to be accurate by ASM, ASM cannot guarantee that favorable results will be obtained from the use of this publication alone. This publication is intended for use by persons having technical skill, at their sole discretion and risk. Since the conditions of product or material use are outside of ASM's control, ASM assumes no liability or obligation in connection with any use of this information. No claim of any kind, whether as to products or information in this publication, and whether or not based on negligence, shall be greater in amount than the purchase price of this product or publication in respect of which damages are claimed. THE REMEDY HEREBY PROVIDED SHALL BE THE EXCLUSIVE AND SOLE REMEDY OF BUYER, AND IN NO EVENT SHALL EITHER PARTY BE LIABLE FOR SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES WHETHER OR NOT CAUSED BY OR RESULTING FROM THE NEGLIGENCE OF SUCH PARTY. As with any material, evaluation of the material under end-use conditions prior to specification is essential. Therefore, specific testing under actual conditions is recommended.

Nothing contained in this book shall be construed as a grant of any right of manufacture, sale, use, or reproduction, in connection with any method, process, apparatus, product, composition, or system, whether or not covered by letters patent, copyright, or trademark, and nothing contained in this book shall be construed as a defense against any alleged infringement of letters patent, copyright, or trademark, or as a defense against liability for such infringement.

Comments, criticisms, and suggestions are invited, and should be forwarded to ASM International.

www.iran−mavad.com

مرجع مهندسی مواد و متالورژی

Foreword

This Volume 22A, *Fundamentals of Modeling for Metals Processing*, represents an expansion of the Handbook series in response to the expressed needs of members of the modeling and simulation community.

ASM International is indebted to the Co-Editors, David Furrer and S. Lee Semiatin, who had the vision for a comprehensive presentation of modeling of metals processing. They moved this vision from inception to this unified collection of content in a remarkably short time, through tireless effort. They recruited world renowned modeling experts who contributed entirely new content. We are likewise indebted to the approximately 120 volunteer authors and reviewers who fulfilled their commitments, squeezing this time intensive activity into their lives busy with family, career, and community commitments.

While this Handbook serves as an organizing vehicle for acquiring modeling knowledge, ASM International is pleased to have the means to disseminate this outstanding source of information in forms most attractive and most readily available to its members and to the technical community.

Modeling is an important aspect of "everything material." One can model at the submicroscopic scale where atomic structure is predominant; at an intermediate, or mesoscale at which grain size/grain structure effects are important; and at the macroscopic, continuum level at which bulk properties are typically determined. Through ASM's strategic content development efforts, specific needs for high-quality materials modeling information are met. Further enhancement will be forthcoming as the Co-Editors complement this work with Volume 22B, *Metals Process Simulation*.

The need for modeling metallurgical behavior during processing has long been recognized and ASM has been a forum for exchange of these ideas. Through mechanistic and phenomenological approaches, solidification and deformation processes can be optimized, the resulting mechanical properties controlled, and defects minimized. As computing power has increased and its cost decreased, more sophisticated simulation of metallurgical processes has enabled material scientists and engineers to maintain competitive advantage over those not willing or able to change.

As an organization of material scientists and engineers, ASM International is pleased to offer this content to practitioners and students of modeling as they continue their exciting journey of tailoring materials and processes to meet future functional needs. This new Handbook, in its printed and electronic forms, also moves us closer to achieving a strategic objective that will shape our society for the next fifty years: to accumulate, review, and distribute comprehensive materials information and to become the global resource for quality materials information.

Roger Fabian
President
ASM International

Stanley C. Theobald
Managing Director
ASM International

Preface

Scientists and engineers have always been curious about cause and effect relationships within nature. This is also the case relative to metals and materials. The understanding of the physics of metals has greatly increased from the earliest days of the field of metallurgy. The discovery of mechanisms that influence and control the behavior of metals has spurred continued research and further discovery. Initial understanding and description of controlling mechanisms were substantially phenomenological, based on observations and perceived interactions of material and process variables on resultant metallic material microstructure, mechanical properties and behavior. The conversion of mechanistic relationships into mathematical expressions is now the field of materials modeling.

The development of models and modeling methods is now allowing more rapid discovery of new alloy systems with greater optimization and application potential. Models are being integrated into computational tools for design and simulation of component processing and manufacture. The successful application of models by industry is also resulting in further pull for even further development of models that are more accurate and predictive. The study of mechanisms that control the evolution and behavior of metallic materials is continuing today at an even more aggressive pace.

Mechanistic models that more accurately describe the physics of metallurgical processes, such as grain growth, precipitation, phase equlibria, strength and deformation as examples are of great interest and importance to science and industry alike. Greater understanding of the physics of metals to the atomistic level, along with increased computational power, has resulted in further discovery and growth in the field of modeling and simulation.

This Handbook provides a review of the models that support the understanding of metallic materials and their processing. An accompanying volume will provide details of the integration of these models into software tools to allow simulation of manufacturing processes. The distinctly different, but complementary fields of Modeling and Simulation are providing new and increased capabilities for metallic materials for components and systems. The future of the metals industry is moving toward an integrated computational materials engineering (ICME) approach as a result of the hard work and dedication of the individuals, teams and organizations that have and continue to provide the needed models and simulation tools that are capable of providing engineers with accurate predictive guidance and direction.

D.U. Furrer, FASM
Roll-Royce Corporation

S.L. Semiatin, FASM
Air Force Research Laboratory

Contents

Introduction

Introduction to Fundamentals of Modeling for Metals Processing

D.U. Furrer, Rolls-Royce Corporation
S. L. Semiatin, Air Force Research Laboratory

THE FORMULATION AND APPLICATION of modeling and simulation methods for metallic materials and manufacturing process design and development is rapidly increasing. Classic models that predict the behavior of metals under processing conditions are continuing to be used and enhanced with greater understanding of the mechanisms that control the evolution of microstructure, texture, and defects. New theories and associated mathematical models are being developed and applied to metallic alloy systems for existing and new processing methods. To complement empirical descriptions of material behavior during processing, so-called first-principles approaches, such as those based on atomistic and molecular dynamics calculations, are now being developed to provide fundamental understanding of the mechanisms that control observed behaviors of existing and emerging alloys, such as those for unique or highly demanding applications.

The recent development of the integrated computational materials engineering (ICME) discipline allows fundamental research to be linked to industrial applications (Ref 1). See the article "Integrated Computational Materials Engineering" in this Volume. The refinement of models and modeling methods results in greater capability and accuracy of metallurgical predictions, such as phase equilibria, microstructure, and subsequent mechanical properties. The ability to rapidly apply fundamental models to practical component design and manufacturing applications has spurred unprecedented collaboration between universities, research laboratories, software companies, and industrial users of modeling and simulation tools. The linkage of component design, alloy design, and component manufacture through modeling and simulation methods will allow for continued advancement in the area of alloy research, advanced process and equipment development, and enhanced component capability.

Modeling and simulation activities are increasing within the materials field as well as other science and engineering disciplines. Development, enhancement, and implementation of computational modeling and analysis technologies to describe and predict physics-based processes are occurring globally within universities and research centers in nearly every country (Ref 2). In addition to increased collaboration between industry and academia, there are increased numbers of multinational collaborative efforts aimed at increasing the capabilities and state-of-the-art of material and process modeling. Fundamental understanding of metallic materials and processing is increasing rapidly and becoming available globally through these and other science and engineering efforts.

Volumes 22A and 22B of the *ASM Handbook* series describe the current state-of-the-art of modeling and simulation for metals processing. The present Volume (22A), *Fundamentals of Modeling for Metals Processing,* focuses on mathematical descriptions of the behavior of metallic materials during processing (and resultant properties) as well as process modeling per se. These models may be broadly characterized into three types on the basis of approach and mathematical expression utilized as being phenomenological (based primarily on direct observations/measurements), mechanistic/physics based, or a combination of the two.

Phenomenological and mechanistic models are also separated by application type, including materials models or process models. Materials models for metals can also be considered metallurgical process models, where grain growth, precipitation, recrystallization, or dislocation impedance are examples of processes within metals and alloys. Materials models describe how, for example, microstructure, crystallographic texture, and defects evolve as a function of local metallurgical process variables, including history and path dependences.

Process models, on the other hand, describe manufacturing processes that require understanding of external independent parameters and boundary conditions and provide information about macroscopic component changes and/or information for metallurgical process models. The nomenclature within this field, however, can cause some confusion, so it is important to understand and clearly define the type of models being discussed.

Manufacturing process modeling addresses the integration of material behavior models with the description of specific processes (e.g., forging, investment casting), which typically include equipment/process characteristics and interface effects (such as heat transfer and friction). The equations describing the physical phenomena in material behavior models and/or their coupling with process models are often too complex for solution by analytical means, except under somewhat restrictive (albeit often insightful) conditions. This is usually the case for real-world industrial processes. In such cases, numerical simulation methods must be used to describe material behavior, process mechanics, and processing-structure-mechanical behavior relationships. Volume 22B, *Metals Process Simulation,* deals with the state-of-the-art of the simulation of specific materials processes, including associated input-data requirements.

Various major sections of this Volume summarize the fundamentals of materials modeling, including modeling of microstructure and texture evolution, modeling of damage and defect evolution, modeling of mechanical properties, and material-specific models for industrial alloys. This article provides a brief historical perspective, a classification of metallurgical processes that are discussed within this Volume, basic model development efforts, and an overview of the potential future directions for the modeling of metals processing.

Historical Perspective

Models of various forms have been developed nearly since the beginning of time. Models can be very simple, such as an "if/then" relationship, or extremely complex mathematical expressions with numerous parameters, some of which may be easily measured and well understood and others that are not physically observable or readily inferred. All models have one thing in common: They attempt to

provide improved understanding of the nature and the variables that influence and control the results of processes, whether a naturally occurring process or man-made. The study of processes leads to theories and subsequent models that can and are used to predict future applications of the studied process. Figure 1 shows a schematic flow path for the development of models that provide improved understanding of processes, where metallurgical or manufacturing process-based.

Models are developed due to need or curiosity. The drive to increase understanding of nature and metallurgical processes has provided a substantial foundation for materials and process modeling. For the most part, the development of material behavior and process models has been spearheaded by metallurgical and mechanical engineers, respectively. Over the last two decades, however, a great emphasis has been placed on the need for coupling material behavior and process models, leading to work that is truly interdisciplinary in nature.

Material behavior models for processing have evolved from ones that are largely empirical in nature to those that incorporate physics-based mechanisms. For example, work in the first half of the 20th century led to engineering (phenomenological) models of:

- Deformation and strain hardening, such as Schmid's law and the Hollomon equation
- Kinetics of recrystallization (such as the Avrami equation), grain growth (such as the Beck equation), and precipitation/phase transformation
- Ductility for solid-state processes

Similarly, phenomenological models of macrostructure evolution during solidification and evaporation during vapor processing were developed for a number of alloy systems. For the most part, these phenomenological models were applicable only to the range of process parameters for which measurements had been made and hence were essentially methods of fitting experimental data. The development of a fundamental understanding of the mechanisms underlying these phenomena, such as dislocation glide/climb and diffusion, led to rudimentary physics-based models, such as those for the large-strain deformation of polycrystalline aggregates, phase transformations based on classical nucleation theory and diffusional growth, and grain growth, among others, in the 1940s to1960s. Beginning in the 1980s, the incorporation of the fundamental concepts of thermodynamics and thermal physics (due to, for example, Gibbs, Ising, and others) led to a great expansion of mechanism-based models, such as those designed to predict phase equilibria (e.g., Calphad), recrystallization and grain growth (Monte Carlo and cellular-automaton techniques), and precipitation and solidification problems (e.g., phase-field methods). The successful implementation of these newer techniques was made possible by the advent of inexpensive computers and ever-increasing computing power beginning in the 1990s.

Modern process models have evolved from relatively simple analytical and numerical techniques. These include the slab, upper-bound, and slip-line field methods (for predicting loads/forming pressures during metal forming, for instance) and the solution of relatively simple partial-differential equations (for various heat flow, solidification, and diffusion problems) that evolved during the 1920s to 1980s. The development of mainframe, mini-, and then microcomputers and associated software beginning in the 1970s and 1980s led to the ability to simulate much more complex processing problems, often based on finite-difference or finite-element numerical techniques. These latter approaches have also enabled the simulation of coupled phenomena, such as those involving simultaneous deformation/solidification, energy transport, and mass transport.

Classes of Material Behavior Models

Material behavior models can be grouped broadly into three classes: statistical, phenomenological, and mechanism based.

Statistical models typically require large amounts of experimental data to derive a mathematical relationship between independent/controlled process parameters and predictions of metallurgical process results. For example, linear regression analysis is often used to "fit" pairs or a series of data to determine relationships (Ref 4). This approach has pros and cons. One advantage is that data generated during the manufacture of components can often be used to generate models, but the available data often do not contain all of the required parameters. In other words, the data may mask second-order or confounding parameters, making difficult the establishment of statistical models that capture the fundamental relationships. Other disadvantages consist of not knowing a priori what metallurgical or processing parameters influence the specific results of a process, and the issue that not all metallurgical processes and relationships are linear. A specific type of statistical analysis, known as artificial neural-network (ANN) modeling, overcomes some of these drawbacks for the development of multivariable, nonlinear relationships, but this approach also requires large amounts of data (Ref 5). In addition, ANN, like other statistical approaches, suffers from not being able to predict results outside the range of data used for "training" the model. Since the physics of the metallurgical process being modeled are not known, extrapolating statistical-based models outside the parameter range in which they are trained is not advisable and can lead to errors and large deviations from reality if the physics of the process change are not captured within the model.

Phenomenological models typically rely on equations that define the relationship between process variables and resulting microstructure, properties, and so on. These types of models can be used to describe phenomena such as recrystallization, grain growth, and creep of metallic materials. For example, Avrami (sigmoidal-type) equations have been used frequently to fit observations of the kinetics of static (and dynamic) recrystallization and other phase transformations during metals processing (Ref 6–8). Such relationships quantify the nucleation and growth mechanisms that lead to an initial slow incubation period, followed by a rapid rate of increase in the
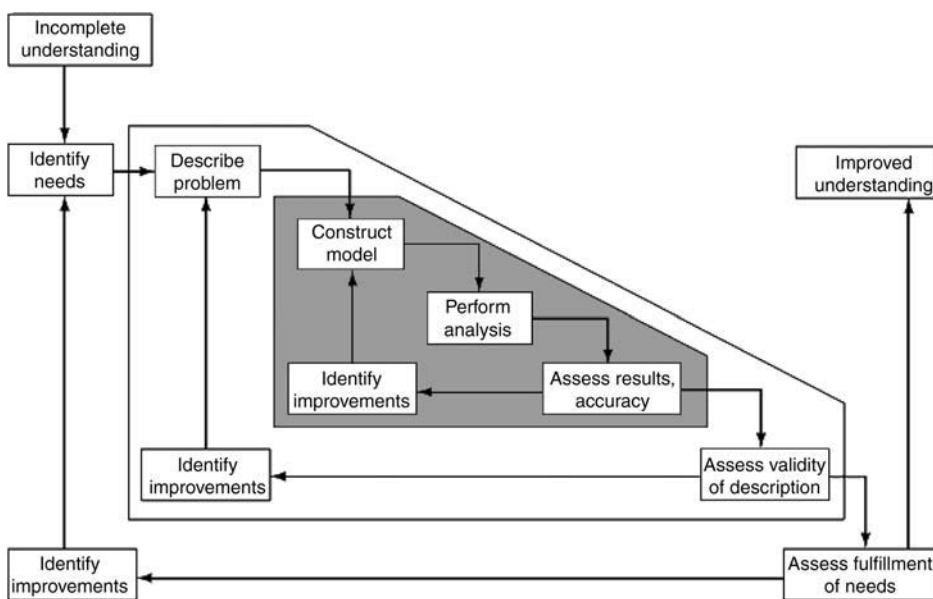
Fig. 1 Process flow for development of models to increase understanding of metallurgical and manufacturing processes. Source: Ref 3

recrystallized/transformed volume fraction, and then a final reduction in the rate of the process until the reaction or metallurgical process completes. Typical applications include the rate of decomposition of austenite in steels that underlie time-temperature-transformation and cooling-transformation diagrams (Ref 9). Similarly, Arrhenius rate-type equations are often used to describe the temperature dependence of metallurgical phenomena, such as the plastic flow of metals at high temperature (using the temperature-compensated strain rate, or Zener-Hollomon parameter), grain growth (Ref 10), and creep (using the Larson-Miller parameter) (Ref 11).

In phenomenological models, the underlying mechanism(s) that control the rate of the metallurgical process are often not known precisely nor explicitly incorporated. This type of modeling approach may also have several inherent limitations:

- Experimental data are often needed to establish or calibrate the relationship, even for mathematical expressions in which the form is known.
- If the mechanism controlling the process changes, the form of the fit assumed in the model may not be appropriate.

Mechanistic models are often called physics based due to their ability to include all of the relevant physical parameters that influence the outcome of a process to a high degree of fidelity. As such, mechanism-based models tend to be the most robust.

Examples of mechanistic models comprise phase-field approaches for microstructure evolution, such as solidification, grain growth, and solid-state precipitate growth. In each of these cases, the mechanisms that control the migration of grain or interface boundaries, such as bulk diffusion or interface reaction, and

suitable input parameters, such as grain/interface boundary energy, must be known (Ref 12, 13). An example of a mechanistic mechanical property relationship is the Hall-Petch model for yield strength as a function of grain size of single-phase alloys (Ref 14). This model relates the applied stress for plastic flow to the slip length within a grain and the stress concentration for slip transmission into neighboring grains. This simple model is effective and is readily used for metallurgical understanding of tensile property control and optimization (Ref 15). However, the grain size is not the only microstructural feature that can influence the yield strength. Additional models are required to describe contributions from solid-solution, precipitation/dispersion, and dislocation strengthening (strain hardening).

Mechanistic models do provide the ability to predict behavior outside the range for which they were developed, provided the controlling mechanism is unchanged. In some cases, the failure of a mechanism-based model to describe observations may provide the impetus for new fundamental understanding. For instance, such was the case for the discovery of microtwinning during creep of nickel-base superalloys (Ref 16). Existing mechanistic creep models did not accurately capture the behavior within a specific temperature-stress regime. After careful focused observations, the new mechanism was discovered. Figure 2 shows micrographic evidence of the new creep mechanism, microtwinning, within nickel-base superalloys.

This Volume contains articles on a range of metallurgical and materials processes. Table 1 lists the various formulations and mathematical methods to predict the relationships and interactions within metallurgical processes. Further development of modeling tools will continue to be seen in both mathematical expression and mathematical formulation for calculation of predictions.

Future Outlook

Modeling and simulation technology will continue to grow because it provides a means

to assess "what-if" scenarios and to conduct virtual experiments. Material behavior and process models provide useful information to both design and manufacturing engineers. Currently, industry is developing holistic component and product design approaches that include metallurgical and process models. In this regard, ICME provides a means of mathematically linking complex metallurgical processes to the mechanical design of components. ICME will be the prevailing method in the future for industry to design and optimize components, alloys, and manufacturing processes in a fully-coupled manner.

Another trend that is gaining momentum is the increasing interest of academic and research institutions in the development of fundamental material behavior and process models for the manufacture of metal products. This may be due to factors such as:

- Widespread availability of inexpensive computer hardware and software
- Limited mechanistic understanding of metallurgical phenomena under processing conditions for alloys of commercial importance and hence a strong technology pull from industry
- Increased use of modeling and simulation within industry develops a need for engineers and scientists with knowledge and background in this area
- Acceptance of modeling and simulation in general as a worthy research subject
- Application of modeling and simulation as a teaching method within multidisciplinary engineering fields
- Shift toward near-term investments and the concomitant reduction in research staff within many companies

Despite the reduction in modeling research within industry, the application of models in the commercial milieu is creating an increased demand for modeling tools, engineering analysis capabilities, and engineers that can apply modeling and simulation tools to industrial challenges. These needs are being addressed by academia in a very synergistic
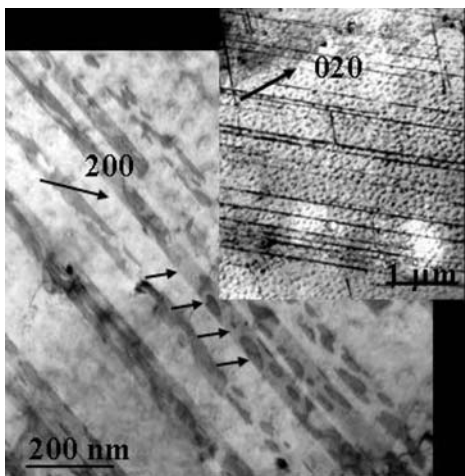
Fig. 2 Recent discovery of microtwins in nickel-base superalloys after creep deformation has led to further investigation and development of a model that describes this new mechanism. Courtesy of M. Mills, The Ohio State University

Table 1   Examples of mathematical modeling approach and applications for metals processing

Tool	Typical application(s)
Slab, upper-bound, slip-line field models	Loads, metal flow, gross defects in bulk forming
Finite-element models deformation processes	Strain/metal flow, strain rate, temperature in deformation processes; powder consolidation
Finite-element/finite-difference models solidification processes	Fluid flow, solidification in solidification processes
Nucleation and growth models	Plethora of phase transformations, recrystallization, etc.
Crystal-plasticity models	Deformation-texture evolution
Cellular-automata models	Recrystallization, solidification microstructure
Phase-field models	Precipitation, grain growth, solidification microstructure
Monte Carlo models	Grain growth, recrystallization
Vertex models	Grain growth
Defect/failure models solid-state processes	Cavity nucleation, growth, coalescence; failure maps
Defect/failure models — solidification	Hot tearing, porosity
Constitutive equation models	Material behavior, metal flow
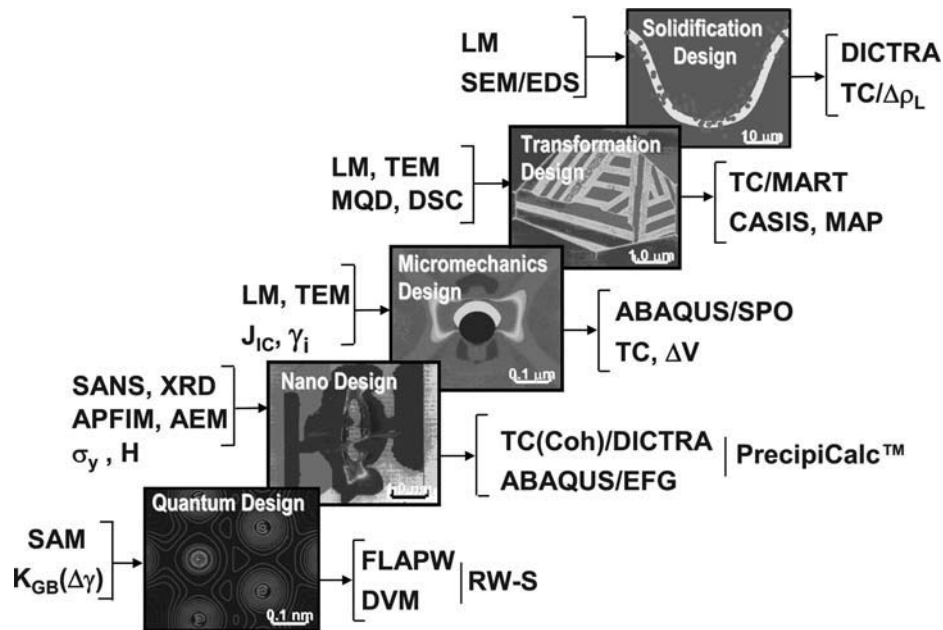
Fig. 3 Length scales of metallurgical processes result in different tools for process observation and methods for modeling and simulation. Models must be developed and used over these length scales to enable accurate, mechanistic prediction of macroscopic effects. Courtesy of C. Kuehmann, QuesTek Innovations, LLC

and collaborative manner. The development and use of modeling tools is bringing the needs of industry and the capabilities of academia and research laboratories closer together. Stronger ties between industry and academia associated with model develop and utilization will increase the rate of modeling and simulation technology development and deployment. Further work is needed to continue the path forward for ICME, but the journey has begun (Ref 1).

The great improvement in the speed of computing systems and the efficiency of mathematical methods is also instrumental in giving rise to the increased application of modeling and simulation tools for complex industrial problems. The near-standard rule of thumb within industry for the application of models is that the computational time of the model should be targeted at no more than overnight, although this is exceeded for many complex applications. Models that are being deployed today (2009) would have taken weeks or months to run only a short time ago due to slower computer speeds, making numerical approaches impractical for complex engineering problems. For this reason, models were previously simplified to allow increases in computational speed, but this resulted in a lack of predictive accuracy. Complex models with increasing detail of the physics of metallurgical processes are now being employed with less simplification and greater predictive accuracy within manageable engineering timescales.

The rate of development of models will continue to increase based on increased understanding of underlying physics of metallurgical problems. As models are established that accurately predict the behavior of metallurgical processes, further discovery of unknown mechanisms will occur when these models do not fit specific examples for new materials or processing applications. The understanding of metallurgical processes is expanding the range of metallurgical modeling over a large range of length scales (Fig. 3). The new frontier of atomistic and first-principles modeling is shedding new light on the understanding of fundamental metallurgical mechanisms. Furthermore, advances in these modeling methods will provide further capabilities for other types of models that require measurement or estimation of fundamental material parameters, such as boundary and interface energies.

REFERENCES

1. *Integrated Computational Material Engineering—A Transformational Discipline for Improved Competitiveness and National Security*, National Academies Press, 2008.
2. S.C. Glotzer et al., "WTEC Panel Report on International Assessment of Research and Development in Simulation-Based Engineering and Science," World Technology Evaluation Center, Inc.
3. K.-J. Bathe, Some Issues for Reliable Finite Element Analysis, *Reliability Methods for Engineering Analysis, Proc. First International Conference*, K.-J. Bathe and D.R.J. Owen, Ed., Pineridge Press, Swansea, U.K., 1986, p 139.
4. S. Weisberg, *Applied Linear Regression*, John Wiley and Sons, 2005.
5. A.I. Galushkin, *Neural Networks Theory*, Springer, 2007.
6. G. Shen, S.L. Semiatin, and R. Shivpuri, Modeling Microstructural Development during the Forging of Waspaloy, *Metall. Trans. A*, Vol 26A, 1995, p 1795–1802.
7. H.J. McQueen and J.J. Jonas, Recovery and Recrystallization during High Temperature Deformation, *Treatise on Materials Science and Technology*, Vol 6, Academic Press, New York, 1975, p 393–493.
8. C. Devadas, I.V. Samarasekera, and E.B. Hawbolt, The Thermal and Metallurgical State of Strip during Hot Rolling, Part III: Microstructural Evolution, *Metall. Trans. A, Vol 12A*, 1991, p 335–349.
9. J.W. Cahn and W.C. Hagel, in *Decomposition of Austenite by Diffusional Processes*, V.F. Zackay and H.I. Aaronson, Ed., Interscience, New York, 1962, pp. 131–191.
10. G. Shen, J. Rollins, and D. Furrer, Microstructure Modeling of Forged Waspaloy Discs, *Superalloys 1996*, R.D. Kissinger et al., Ed., TMS, 1996, pp. 613–620.
11. F.R. Larson and J. Miller, *Trans. ASME*, Vol 74, 1952, pp. 765–771.
12. Y. Wang et al., Predicting Phase Equilibrium, Phase Transformation, and Microstructure Evolution in Titanium Alloys, *JOM*, 2005, pp. 32–39.
13. N. Ma et al., *Acta Mater.*, Vol 52, 2004, p. 3869.
14. N.J. Petch, J. *Iron Steel Inst. London*, Vol 173, 1953, p. 25.
15. D.U. Furrer, A Review of U720LI Alloy and Process Development, *Materials Design Approaches and Experiences*, J.-C. Zhao et al., Ed., TMS, 2001, pp. 281–296.
16. G.B. Viswanathan, P. Sarosi, D. Whitis, M.F. Henry, and M.J. Mills, Investigation of Creep Deformation Mechanisms at Intermediate Temperatures in René 88 DT Superalloys, *Acta Mater.*, Vol 53, 2005, pp. 3041–3057

Integrated Computational Materials Engineering*

John E. Allison, Mei Li, and XuMing Su, Ford Motor Company

INTEGRATED computational materials engineering refers to the use of computer simulations that integrate mathematical models of complex metallurgical processes with computer models used in component and process design. Currently, computational simulation is used in both product design and manufacturing process development with tools such as finite-element analysis (FEA) and computational fluid dynamics (CFD), as described in other articles in this Volume. On the design side, for example, FEA and CFD methods are quite well developed, yielding accurate predictions of important design variables such as the stresses, strains, and temperatures to which a component is subjected. Simulation of manufacturing processes is also conducted using FEA or CFD, and these simulation tools provide a means for determining whether the component is manufacturable and for optimizing the manufacturing process to reduce the manufacturing time or improve the quality of a component. The physical phenomena associated with product performance and manufacturing are generally governed by a few general principles (e.g., elasticity, heat transfer, mass transport) that have been the subject of study and development for centuries. Thus, product performance and manufacturing simulations have evolved to be quite sophisticated and computationally efficient.

In contrast, computational tools in the simulation of metallurgical and materials processes (on micro-, meso-, and macroscopic scales) have lagged behind developments in other engineering disciplines. The behavior of materials involves complex processes, especially when the processes involve a multitude of discrete interactions that are not adequately described by the continuum models of behavior on macroscopic or mesoscopic scale. That is, computational modeling of material behavior includes many processes with time and length scales that are not adequately described by the well-developed continuum models for computer simulation of engineering product and process designs.

This is not to say that major advances have not been made in the tools of computational materials science. However, these tools are generally designed for use in scientific research to further mechanistic understanding rather than for use in engineering decision making. The computational materials tools are often directional in nature rather than quantitative, and, in most cases, they require some form of empirical calibration due to a lack of comprehensive physics-based understanding of the complex phenomena at play. Most importantly, the computational tools are not designed to be integrated with one another nor with manufacturing simulation or product performance codes. In addition, modern component design requires consideration of many different properties, which again adds a multiplier effect on the complexity of the computational problem.

Since the late 1990s with advances in computer simulation, new computational tools have been developed to include more complex materials phenomena into the product and process simulations. This approach is called integrated computational materials engineering (ICME) (Ref 1, 2), which has been defined as "the integration of materials information, represented in computational tools, with engineering product performance analysis and manufacturing-process simulation" (Ref 2). For metals, this materials information is in the form of quantitative processing-microstructure-property relationships linked with thermodynamic phase equilibria databases. The information can be from theoretical models, empirical relationships, or databases; however, this information also is captured in models that can be integrated with one another and with simulations of manufacturing and product performance. By holistically integrating simulations of materials, design, and manufacturing, ICME is revolutionizing the way the materials community provides input to the engineering and scientific communities.

This article, based on a previous publication (Ref 3) by the authors, briefly outlines an example of a suite of computer-aided engineering (CAE) tools developed and implemented at Ford Motor Company for quickly developing durable cast aluminum power train components. The CAE program, called virtual aluminum castings (VAC), captures extensive knowledge of metallurgical processes that cover all key length scales and puts this knowledge into advanced, computationally efficient materials models. These tools link analysis of manufacturing processes and design in a holistic computational environment enabling simultaneous engineering of critical aluminum castings. This has resulted in major benefits in reduced time and costs for product creation and optimization.

The VAC system stands as a key demonstration of the ICME concept as a means of mathematically linking complex metallurgical processes with the mechanical design of components. Future advances will require formation of dedicated teams of experts in materials, mechanics, physics, engineering, computer science, and information technology by industrial firms or government agencies wishing to develop tools for specific applications. A recent report has made extensive recommendations that would lead to a widespread ICME capability (Ref 2). Although ICME is still in its infancy, it has great potential for producing more robust simulations to optimize components, alloys, and manufacturing processes in a fully-coupled manner. Other articles in this Volume also provide further details on the refinement of models and modeling methods that result in greater capability and accuracy of metallurgical predictions, such as phase equilibria, microstructure, and subsequent mechanical properties.

Virtual Aluminum Castings

Simulation of materials in both product performance analysis and manufacturing simulation often involves use of nominal properties, which have been previously measured experimentally. The net result of this lack of

---

computational capability is that the design and manufacturing cycle must be iterative and involves a testing phase or phases to ensure that the material properties and the component meet the designers' intent. The influence of manufacturing history on material performance is accounted for by using minimum values for material properties that contribute to the iterative process and result in overly conservative and inefficient designs. The lack of computational materials engineering tools also results in a development process for new materials that is empirical, slow, and can be disruptive. Nowhere is this more evident than in the development of designs and manufacturing processes for cast aluminum engine blocks and cylinder heads. Increasing engine performance requirements coupled with stringent weight and packaging constraints are pushing aluminum alloys to the limits of their capabilities.

To provide high-quality blocks and heads at the lowest possible cost, engineers at Ford Motor Company developed and implemented a CAE process that tightly couples manufacturing and design analysis of cast aluminum power train components. The system, referred to as VAC, is a suite of integrated computational tools that enables the rapid development of durable, cost-effective cast aluminum power train components. The VAC system is based on advanced material models that bridge the many key dimensional scales from the atomistic level to the component level. By using the VAC system, virtual components are now designed and cast, heat treated, and tested for durability, all on a workstation long before components are fabricated.

The VAC system is a holistic manufacturing and design analysis system with robust computational models that help bridge the gaps between program needs and engineering with "up-front" analysis (i.e., providing analysis before parts are cast). It demonstrates the capability and benefits of the ICME paradigm. The development of the VAC system is the culmination of years of comprehensive research on cast aluminum. It has been accomplished by a combination of theoretical, experimental, and computational technologies and has involved development of a deep, fundamental understanding of dozens of separate phenomena. This theoretical and empirical knowledge has been captured in a computationally efficient software package.

The VAC system is the rare technological innovation that can be used to simultaneously reduce cost, improve quality, save time, and reduce weight. It is not asset-intensive and uses existing CAE infrastructures, so it costs relatively little to implement. The VAC system has been successfully implemented within the Ford power train design, manufacturing, and CAE communities, has significantly reduced product development time, and has saved Ford millions of dollars.

General Description of the VAC System. The central objective of the VAC system is to significantly reduce the time required to develop and optimize new cast aluminum components and casting/heat treatment processes. The product creation process for cast aluminum blocks and heads has traditionally been a costly method of design, build, test, redesign, build,

and retest. Manufacturing analysis is traditionally conducted after the design is complete. Subtle manufacturing changes made late in the product development process can lead to engine durability problems and, as a result, delays in launching new products. By providing a holistic analysis environment, the VAC system enables manufacturing and product engineering to work together simultaneously to solve these problems long before components are cast and engines tested.

This objective is accomplished by developing and validating a computational capability with four interdependent parts (Fig. 1):

- Accurate simulation of the thermal history of an aluminum component through casting and heat treatment
- Prediction of the microstructure that evolves during these manufacturing processes at all locations in a casting (i.e., the local microstructure)
- Prediction of critical local mechanical properties that result from these local microstructures
- Coupling predicted local properties and new damage and residual-stress models with FEA methods to predict the durability of engine components

Commercial casting simulation software, specifically MagmaSoft and ProCast, and FEA software, specifically ABAQUS, provide the basic framework for development of these computational capabilities. This methodology provides Ford with an unsurpassed capability
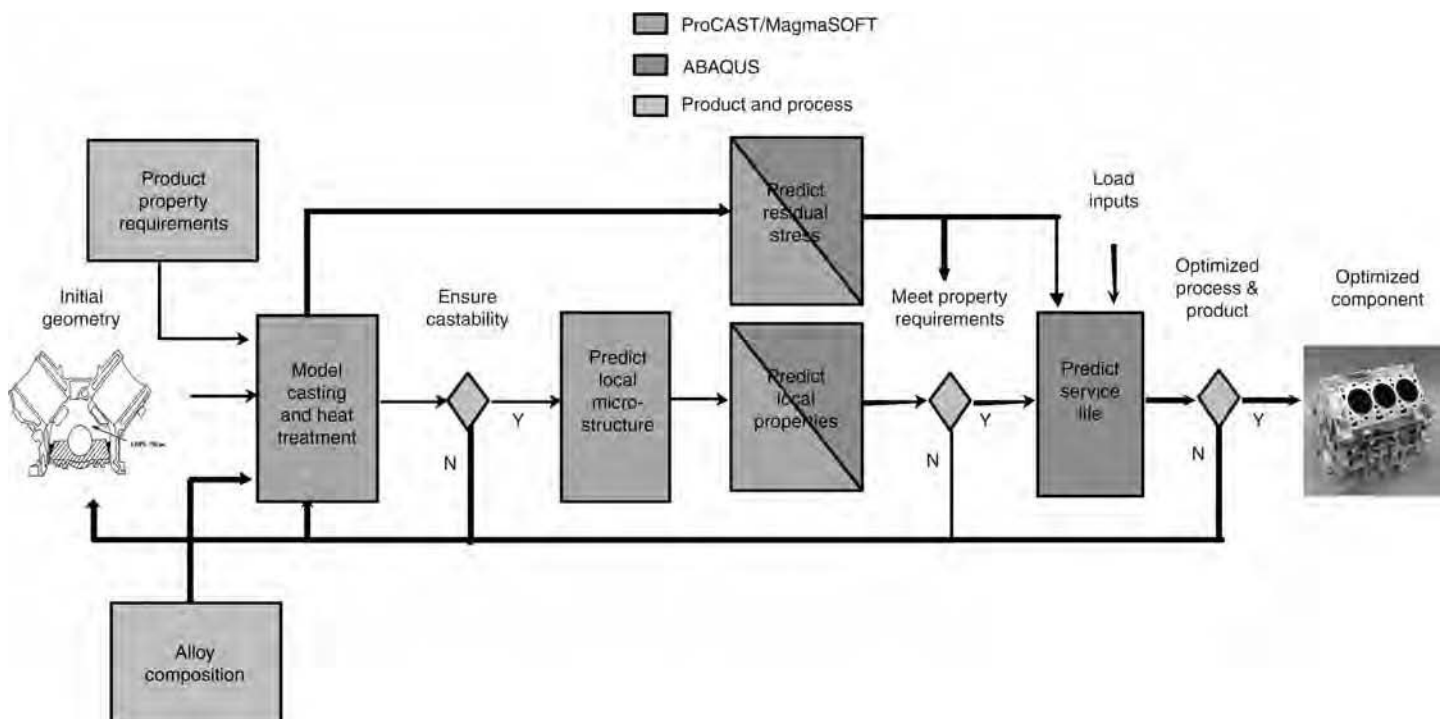
Fig. 1 Virtual aluminum castings methodology flowchart

for quickly engineering low-cost, durable aluminum castings.

The development of VAC required activities in several key areas, specifically:

- Use of advanced materials models to integrate analytical tools for simulating the casting and heat treatment processes with analysis of component durability
- Linking fundamental models for microstructural evolution with fundamental models for property prediction
- Integration of fundamental knowledge of phenomena occurring at a wide variety of length scales into complete and coherent models
- Validation of the integrated models
- Incorporation of these software tools into a simultaneous manufacturing and product engineering process, stressing computational efficiency and having the right information available at the right time

These areas also comprise the key areas of the ICME methodology described elsewhere (Ref 1, 2). Figure 2 shows the key processing, microstructure, and property knowledge nodes required for cast aluminum alloys. In developing the VAC tools, the focus was as much on linking the models that describe these individual knowledge nodes as it was on the development of the individual models themselves. The interactions between these knowledge areas and, in particular, the need to optimize multiple properties as well as cost constraints demonstrate the complexity of this problem and thus the need to conduct this optimization in an integrated modeling environment.

One of the challenges in developing VAC tools was to develop a predictive approach capturing the influence of the manufacturing process history on the mechanical properties. This is a significant challenge because these models must account for metallurgical phenomena that occur at vastly different length and time scales (Fig. 3). For instance, solute diffusion and precipitation in alloys is inherently an atomistic process but can manifest itself via changes in macroscopic properties, for example, yield strength or thermal growth. Hence, in constructing properties models in VAC, modeling tools have been used and linked from the atomistic scale through nanostructure and microstructural length scales, all the way to the macroscopic dimensions, as depicted in Fig. 3. Metallurgical features at each of these length scales influence properties in a wide variety of complex ways; thus, it was necessary to develop a fundamental and quantitative understanding of the manner in which specific properties were influenced by specific metallurgical features acting singularly or in combination. However, to ensure computational efficiency, it was important to model only those metallurgical processes and length scales that are critical to the desired outcome. Development of these models required a unique mix of research expertise including experimentalists, theoreticians, numerical modeling, metallurgy, physics, and engineering mechanics.

Model Development

To illustrate the needs and requirements for ICME development, key aspects of integrated computational engineering are described here for Ford's VAC program. Other examples are given in Ref 2. The section "Benefits of Virtual Aluminum Castings" in this article describes how the VAC system has been used in process selection, process optimization, and improved component design criteria.

Linking Manufacturing Process to Microstructure

Commercial software such as ProCAST and MagmaSoft has been used in the past for casting process simulation; however, their primary use has been for determining the castability of geometries. These codes generally do not predict the local microstructures that evolve during casting and heat treatment processes nor do they predict local mechanical properties resulting from these local microstructures. For cast aluminum, the key microstructural features are microporosity, eutectic phases, and precipitate phases.

To predict these local microstructural features, it is critical to first be able to accurately predict thermal history during casting and heat treatment processes. Commercial casting simulation codes have limited success in predicting microstructure, in part due to limitations in their heat-transfer coefficient databases. For decades, researchers have attempted to obtain heat-transfer coefficients for different metal-mold interfaces and processes experimentally, analytically, or with inverse modeling approaches. However, for a variety of reasons, it is usually difficult to apply the interfacial heat-transfer coefficients (IHTCs) from the literature directly. Chief among these reasons is the dependence of the IHTC on the casting geometry and detailed differences in mold material, quench media, and casting or quenching processes. Therefore, a specialized optimization routine was developed, called OptCast (Ref 4), that couples MagmaSoft or ProCAST with an optimization program based on an inverse modeling approach. OptCast enabled the development of accurate IHTC for a wide variety of geometries, casting processes, and heat treatment (water quenching) processes. These improved heat-transfer coefficients substantially increase the accuracy of thermal histories predicted by the casting and heat treatment simulations.

Prediction of microstructure and microsegregation during solidification in a multicomponent alloy is a crucial step in understanding and simulating mechanical properties and subsequent in-service performance of cast components. In 319-type aluminum alloys (Al-Si-Cu), the eutectic θ or Al_2Cu phase is of particular interest because it affects the subsequent evolution of the precipitation-strengthening phase, θ' (also Al_2Cu), during the aging process. During solution treatment, these eutectic phases slowly dissolve, so it is also important to include this effect. A software tool called MicroMod was developed to couple solid-state diffusion, dendrite arm coarsening, and dendrite tip undercooling directly with a commercial multicomponent phase diagram (CALPHAD) computation tool, Pandat. Results from phase-transformation kinetic models, such as the commercial tool Dictra, were used for predicting
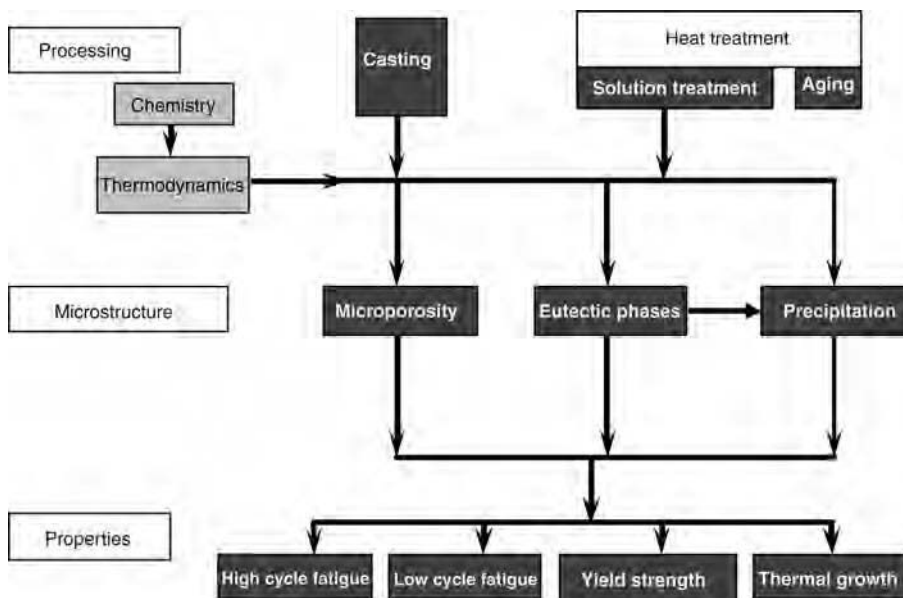
Fig. 2 Key processing-structure-property linkages and knowledge domains for cast aluminum alloys
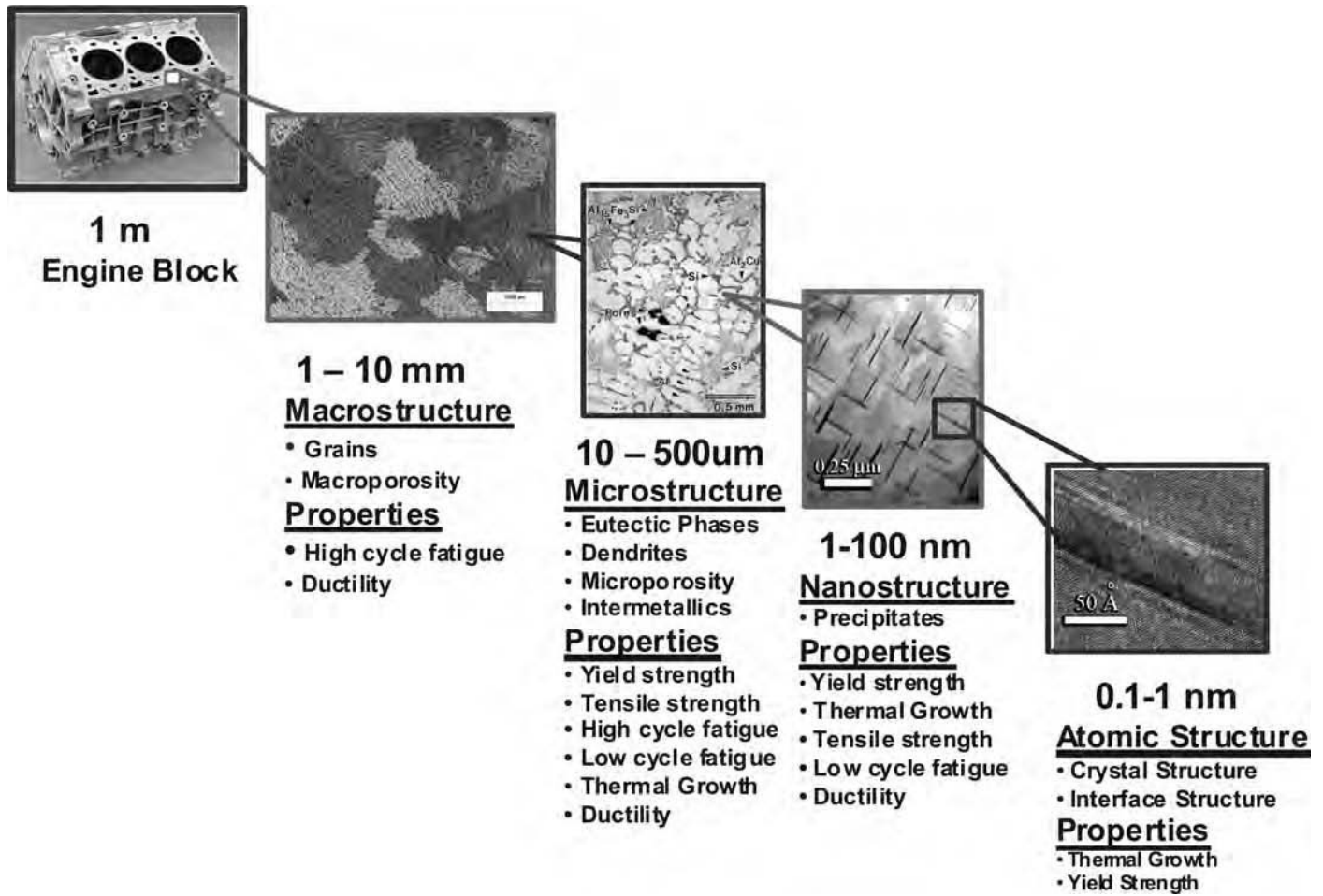
Fig. 3 Metallurgical length scales that dominate properties in cast aluminum alloys

phase dissolution. MicroMod is capable of predicting secondary dendrite arm spacing and, more importantly, the amount and type of the eutectic phases that evolve during casting and the dissolution of these phases during subsequent thermal treatments.

At the nanoscale, accurately predicting the amount and morphology of the precipitation-strengthening phase, θ′, is critical. Similar to MicroMod, this precipitate prediction requires an approach that links different modeling techniques. The resulting model, called NanoPPT, was accomplished by linking first-principles atomistic calculations based on density functional theory for the calculation of stable and metastable thermodynamic functions, thermodynamic phase equilibria calculations such as ThermoCalc or Pandat for phase stability, and microstructural evolution models for precipitate kinetics and morphology (Ref 5, 6). The microstructural evolution models were heavily dominated by empirically-derived relationships (Ref 7). Ongoing developments are aimed at the incorporation of a multiscale first-principles/phase-field approach, developed to reliably predict θ′ morphologies (Ref 8, 9), thereby eliminating one of the key empirical components of the current version of NanoPPT.

Microporosity is a common microstructural feature that can have a profound influence on properties such as fatigue. The VAC tool MicroPore, a subroutine for commercial casting simulation codes, uses the calculated casting thermal histories to quantitatively predict the relevant characteristics of microporosity. It incorporates the complex, nonlinear physics of nucleation and growth of pores. It models both the macroscopic phenomena that control microporosity, including fluid flow and pressure variations in the melt, and microscopic phenomena such as segregation of hydrogen/alloying elements to accurately predict the local pore size.

Linking Microstructures to Mechanical Properties

As described previously, the ICME philosophy was applied to develop the microstructural evolution models in VAC through linking models operating at different length scales. It was also used in development of the models for prediction of mechanical properties resulting from these microstructures.

Yield Strength. The authors' approach for modeling the age-hardening yield strength behavior of Al-Si-Cu cast alloys uses micromechanical models of precipitation strengthening that connect key microstructural parameters for realistic precipitate morphologies (e.g., {100} plates) with the age-hardening response (Ref 7). The microstructural parameters of the key strengthening θ′ plates measured by transmission electron microscopy and a combined first-principles/computational-thermodynamics model of θ′ volume fraction (Ref 10, 11) are used in the micromechanical model to predict precipitation strengthening, producing a completely parameter-free model of this contribution to the strength. This yield strength model, called LocalYS, is linked with microstructural evolution models (MicroMod and NanoPPT), producing a model of the macroscopic, location-dependent yield strength behavior throughout the cast part.

Thermal Growth. In Al 319 castings aged for peak strength (e.g., T6), a macroscopic, irreversible dimensional change (termed thermal

growth) occurs during extended high-temperature exposure (Ref 12). Hence, heat treatment schedules are often devised in an effort to stabilize the casting with respect to in-service dimensional changes. The unique combination of first-principles atomistic calculations, computational thermodynamics, and experimental measurements used to construct NanoPPT was similarly used to produce a model, called LocalTG, of thermal growth in Al 319 (Ref 6). The precipitation of Al_2Cu (θ') is the major contributor to thermal growth, and the model, based on θ' and θ evolution, provides a quantitative and accurate predictor of measured thermal growth. Like the yield strength model, LocalTG was also made possible only via a linkage with the microstructural models, NanoPPT and MicroMod (Ref 5).

As an illustrative example of the integration of microstructural evolution, length-scale information, and properties, the general expression for thermal growth $g(t,T)$ is examined. Growth is given as a function of time and temperature (Ref 6) as:

$$g(t,T) = (1-\gamma)\left[\frac{\delta V'_\theta}{3V} f'_\theta(t,T) + \frac{\delta V_\theta}{3V} f_\theta(t,T)\right]$$

(Eq 1)

In this relationship, $\delta V_{\theta'}$ and δV_θ are the volume changes associated with copper atoms going from solid solution to precipitate phases θ and θ', respectively. These volume changes are determined using first-principles atomic-scale calculations. The symbols f_θ and $f_{\theta'}$ are the fractions of copper involved in θ and θ' phases as a function of time and temperature. These quantities are determined from NanoPPT using a unique combination of first principles,

CALPHAD, and empiricism developed for the alloy in question. The factor γ accounts for the fraction of copper lost to eutectic θ phase and is predicted from the solution treatment dissolution model contained within MicroMod.

By incorporating this thermal growth model into a time- and temperature-dependent "swelling" module within ABAQUS, one is able to predict not only the thermal growth that occurs in cast aluminum engine blocks and heads as a function of heat treatment and in-service temperature and time but also the stresses that develop due to growth in these parts. Combining these tools with residual-stress models (see subsequent text) produces a tool within the VAC suite for not only process optimization of heat treatment times and temperatures but also design optimization of blocks and heads. Thus, thermal growth provides a valuable example where important phenomena occur and must be accurately modeled on scales ranging from angstroms to meters.

Fatigue Properties. Fatigue strength is one of the most important properties impacting cast aluminum engine components. The ability to predict the influence of casting history on the local fatigue strength is therefore a key capability within the VAC toolset. The authors' local fatigue strength model, LocalFS, is based on a novel short-crack growth fatigue model (Ref 13) that relates the size of micropores to the high-cycle fatigue response. The local pore size characteristic predicted by the microporosity model described previously is used in turn to predict the local fatigue strength in the critical regions. Figure 4 illustrates the procedural flow necessary to calculate pore size with MicroPore for a casting component and links it to the prediction of local fatigue properties.
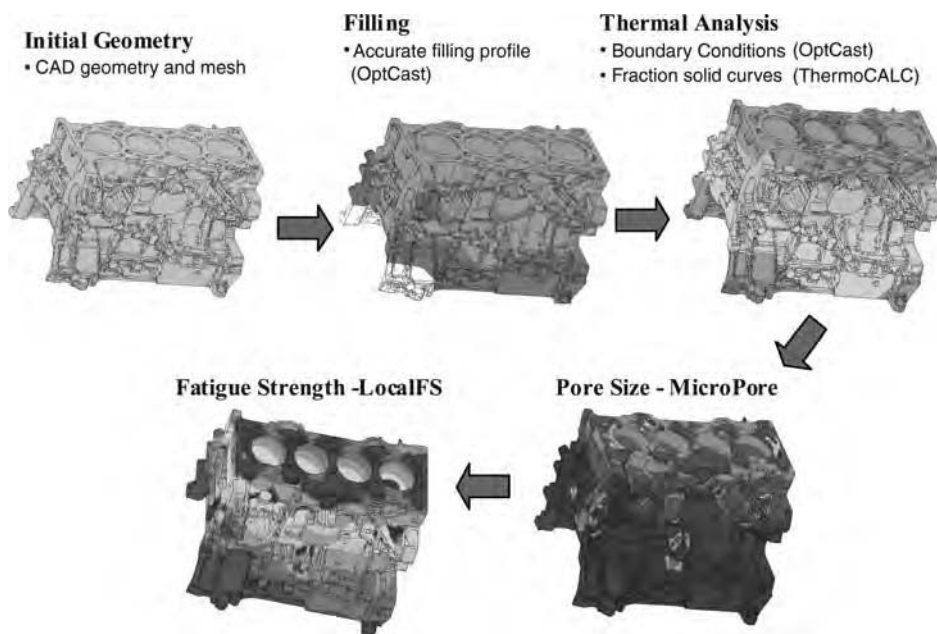
Linking Material Properties to Performance Prediction

The material properties obtained previously are integrated into engineering analysis by another link, that is, the coupling of material properties to product performance predictions, in this case, durability. For that purpose, another critical VAC component, the prediction of residual stresses, was required.

Residual-Stress Analysis. The prediction of residual stresses in complex castings represents a formidable engineering challenge. The residual stresses most concerned in VAC are those formed during heat treatment of cast aluminum cylinder head and blocks. The QuenchStress VAC module was developed to predict residual stresses due to casting and heat treatment.

Residual stresses in a heat treated cast aluminum engine component are mainly introduced during the quench step following the solution treatment, when the component is quickly cooled from approximately 500 °C to much lower temperatures by immersing the component in water or polymer quenchant or by using forced air. The temperature gradient from the surface to the interior of the component leads to nonuniform thermal expansion and nonuniform plastic deformation and residual stresses. The residual stresses generated during quenching are relaxed partially during the aging step of heat treatment. Any attempt to predict residual stresses in a heat treated aluminum component thus involves thermal analysis for the transient temperature field during quenching and stress analysis for both quench and aging.

The thermal analysis of forced-air quench processes can be handled reasonably conveniently by commercial computational fluid dynamics (CFD) codes (Ref 14). Water quenching is much more difficult to simulate, because it involves boiling and vaporization of water on metal surfaces. Due to the limitations of current CFD codes for solving this problem, the authors adopted an inverse modeling approach using the previously described OptCast software. In this instance, the heat-transfer problem is treated as a boundary value problem, using temperature-dependent IHTC on the metal surface to represent the complicated thermal activity between water and the metal surface (Ref 15). The influence of geometrical features within a component on the IHTC is accounted for by dividing the geometric surface into several groups. A database of temperature-dependent IHTCs is obtained for different water and component conditions.

The key to success of a robust residual-stress analysis is the material constitutive relationship. The material response to loading during the quench process is strongly temperature and strain-rate dependent. Experimental measurements of stress and strain relationships were carried out at isothermal conditions for various temperatures and strain rates. These test data were captured in a unified material relationship

Fig. 4 Virtual aluminum castings process flow for predicting local characteristic pore size and local fatigue strength

(Ref 14, 16). A computationally efficient, user-defined material subroutine, called QuenchStress, was developed to interface the material relationship with the commercial finite-element code, ABAQUS.

Durability Prediction. Accurate durability prediction routines required the ability to predict the material stress-strain response and fatigue response during the complex thermomechanical cycling that occurs in cylinder heads and blocks. For this purpose, the HotStress subroutine was developed. It is based on a unified viscoplastic material relationship (Ref 16, 17) and integrates the results of the viscoplasticity model with output from the previously described local yield strength model. It also accounts for the impact of material aging during both heat treatment and engine operation.

The final module of the VAC toolset is a durability model, Hotlife, which predicts how the component responds to a wide variety of loading conditions. These durability models (Ref 14, 17) have been implemented into an ABAQUS postprocessor, using the VAC-predicted local properties and residual-stress models as inputs. These durability models predict the response of the structure to complex high-cycle, low-cycle, and thermomechanical fatigue loading sequences. Normally, such durability models are based on a database approach that assumes an average material property with no manufacturing influence on the mechanical properties. The VAC tools are unique in that they account for the influence of the casting and heat treat process on the variation in local properties as well as the local residual stresses that are produced during the manufacturing processes.

Model Validation and Integration into the Engineering Process

Despite the many benefits of a tool such as VAC, convincing manufacturing and product engineers to move from a physical world to a virtual world requires them to have a strong sense of confidence in the methodology. Thus, an essential and integral step of VAC tools development is the experimental validation of these integrated models. This step involved the development of many novel experimental techniques for quantification of such factors as residual stresses (Ref 14, 15) and component durability (Ref 18) and a comprehensive experimental validation from castings manufactured under a wide variety of casting and heat treatment process conditions. Figure 5 shows a typical validation of the local yield strength model. An important aspect of validation is determining, with the target user, the range over which the product may operate and an acceptable degree of correlation. For the yield strength example shown in Fig. 5, the model predictions have an excellent correlation with the experimental measurements in the normal production region and a reasonable correlation even far beyond this region.

LIVE GRAPH
Click here to view
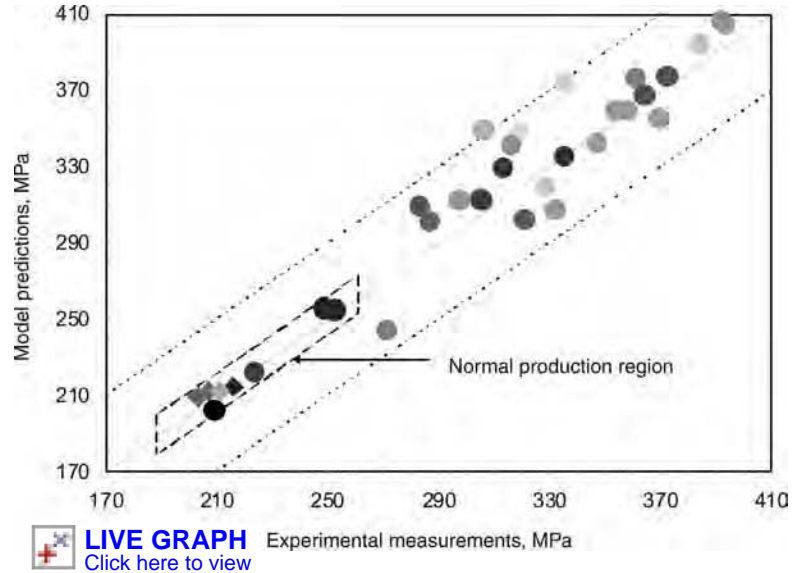
Fig. 5 Experimental validation of virtual aluminum castings local yield strength predictions for a wide range of components, heat treatments, and casting conditions

Another critical success factor was the integration of the VAC toolset into an efficient engineering methodology. This was another challenging aspect of the VAC development. Typically, manufacturing simulation is conducted after performance modeling and is used to ensure that a well-designed (from a mechanical perspective) part can be manufactured. For VAC to succeed, it was necessary to reverse these procedures; that is, the manufacturing simulation must come prior to prediction of the engine component durability. Organizational cultural changes were required as well as development of timing plans and procedures to ensure that critical manufacturing information was available earlier in the process than was previously typical. Ensuring that program timing could either be met or accelerated requires that these complex computations can be completed efficiently and that handoffs between manufacturing CAE and product (performance) CAE are organized and efficient. To accomplish this, substantial effort was expended in development of computationally efficient, proprietary algorithms and procedures as well as specialized software linking the outputs of casting simulations codes to FEA codes.

Benefits of Virtual Aluminum Castings

The VAC methodology implemented at Ford is used in a variety of engine programs with a number of important benefits. It provides a common tool for use by the global Ford power train CAE community, and this common tool captures a comprehensive knowledge in casting technology, product design, metallurgy, physics, and mechanics of cast aluminum alloys. A key benefit is a 15 to 25% reduction in the time

it takes to develop a new cylinder head or block. This was accomplished by minimizing or eliminating costly and time-consuming iterations, the need for which would previously have been discovered only during engine testing. Although engine durability testing remains a key requirement of the design verification, the use of VAC has enabled cost-savings by reducing the number of specialized component tests required to assure product durability. In addition to cost-savings resulting from an improved product and process development process, the VAC tools have been used to optimize the key economic aspects of a casting or heat treatment process (e.g., cycle time) while ensuring that the component is manufacturable and that component durability and quality are not compromised. Since its inception, it is estimated that the use of VAC tools has saved Ford millions of dollars in direct cost-savings or cost-avoidance. Examples are given as follows.

Manufacturing Process Selection and Optimization

The VAC system provides engineers a tool to explore different manufacturing processes and then select the most economical manufacturing process to produce components that meet the property requirements. One example is the application of the prediction of local fatigue properties in the selection of a manufacturing process for a cylinder head. In a demonstration of the VAC capability, two different casting processes were evaluated for fabricating a cylinder head. Figure 6 shows the comparison of local fatigue strength predictions in the cylinder head produced by the two different casting processes, noted here as process A and B. It can be seen that the fatigue strength in the valve bridge

(a critical region of cylinder head durability) was higher using process A, a lower-cost process, than that produced by process B. In addition to mapping local fatigue strength and yield strength, residual stresses could also be determined. By carefully and simultaneously analyzing all critical regions and properties, it could be determined that casting process A provided improved product performance at a substantially reduced cost.

Virtual aluminum castings also provide a valuable tool to engineers for optimizing the manufacturing process. A hypothetical example is the use of the local yield strength model to optimize the heat treatment process for a cylinder block. In this example, the property target for the key bolt boss shown in Fig. 7 was given to be 220 MPa. An initial heat treatment process for this block included a 5 h aging at 240 °C. As shown in Fig. 7(a), the yield strength at this bolt boss was predicted to be 210 MPa, which is below the property target. Using VAC, a new heat treatment process, based on aging for 3 h at 250 °C, was predicted to increase the yield strength to the desired level. The optimized aging process not only allows the component to meet the property target but also makes the process faster by reducing the cycle time by 2 h.

Design Improvement and Optimization

Another key benefit of the VAC tools is the opportunities for design improvement and optimization made possible by the improvement in durability predictions. These predictions are improved specifically by the incorporation of two key factors currently missing from other durability predictions:

- Property predictions that are dependent on the location through the cast part and sensitive to manufacturing history
- Incorporation of residual stresses

One example of the way VAC tools led to a design improvement is illustrated by the ability to predict the spatial variation in fatigue properties and the influence of the casting process on these location-dependent properties. The use of these local properties is vastly superior to the use of average or nominal properties in calculations. As shown in Fig. 6, the fatigue properties in a complex cast component can vary by 30 to 40%, in this case, from 56 MPa to over 84 MPa. The use of a nominal property on the upper end of this range in fatigue calculations would result in an overly optimistic calculated life and the potential for a component that does not pass the engine durability phase of the product development cycle. This could require a change in the geometry or the casting process, resulting in costly and time-consuming iterative rework of the casting tooling. In contrast, use of a nominal fatigue property on the low end of this range could lead to an overly conservative design. While this would be a durable design, it would be a heavier and thus more costly design than required to meet the design intent of the component. In contrast, use of the VAC-predicted local fatigue properties leads to an accurate understanding of the influence of casting on local properties, and this enables a more robust durability prediction. The use of VAC tools also provides a capability for putting the right properties where they are needed throughout the component. This capability leads to substantial improvements in product development timing and an optimum (e.g., lightweight and low-cost) design.

Residual stresses can also play an important role in determining the durability of engineering components. By influencing the mean stresses during high-cycle fatigue events, residual stresses can radically influence the fatigue resistance of a component. However, due to the difficulty in calculating these residual stresses, they are often excluded from durability design calculations. When residual stresses are calculated using VAC tools, fatigue durability estimates can be more realistic. For example, without accounting for residual stresses, the predicted life of a cylinder head is estimated at over 10^7 cycles, a typical design life for a cylinder head. However, when the residual stresses are calculated using the VAC tools, the predicted life of the head would be an order of magnitude less, thus necessitating design changes. The VAC tools thus offer a capability to quantify and account for this life-limiting factor long before components are cast and assembled into a running engine.
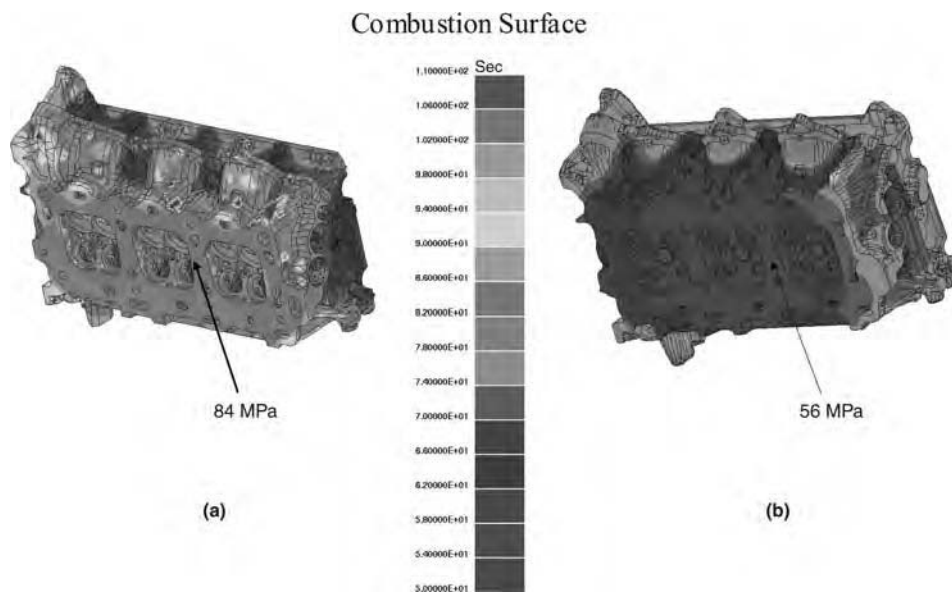
Fig. 6 Using virtual aluminum castings predictions of local fatigue strengths in the selection of a manufacturing process for a cylinder head. (a) Casting process A. (b) Casting process B
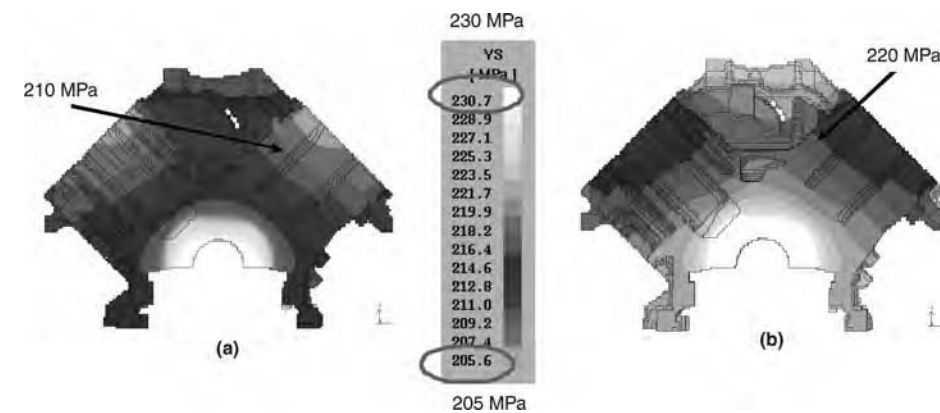
Fig. 7 Using virtual aluminum castings predictions of the local yield strength in a hypothetical cylinder block to optimize the heat treatment process. (a) Initial heat treatment process: 5 h age at 240 °C. (b) Optimized heat treatment process: 3 h age at 250 °C. Note: Property target in bolt boss area indicated by arrow is 220 MPa.

Benefits and Outlook

The benefits demonstrated by the VAC methodology are concrete proof that the promise of the ICME concept can be realized in practice. Linking manufacturing, materials, and design in a holistic CAE environment has resulted in a tool with unsurpassed capabilities. Having established the fundamental framework of VAC, it can be upgraded quickly and relatively inexpensively to other alloys, new casting processing methods, and non-power train (e.g., body and chassis) cast aluminum components. Ford is currently extending this technology to high-pressure die castings, a relatively new process for Ford aluminum engine blocks and magnesium castings. The ICME concept is also under investigation to determine its applicability to areas ranging from sheet metal to plastics and paints.

The ability to integrate manufacturing processes and component design through advanced materials models has much promise. As this article and other work (Ref 2, 19) indicate, ICME tools are feasible with significant value where they have been implemented. Although this value has been demonstrated, the field of ICME is in its infancy, and ICME tools are not yet commercially available. Future advances will require formation of dedicated teams of experts in materials, mechanics, physics, engineering, computer science, and information technology by industrial firms or government agencies wishing to develop tools for specific applications. A recent report has made extensive recommendations that would lead to a widespread ICME capability (Ref 2).

REFERENCES

1. J. Allison, D. Backman, and L. Christodoulou, Integrated Computational Materials Engineering: A New Paradigm for the Global Materials Profession, *JOM,* Nov 2006, pp. 25–27.
2. T.M. Pollock, J.E. Allison, D.G. Backman, M.C. Boyce, M. Gersh, E.A. Holm, R. LeSar, M. Long, A.C. Powell IV, J.J. Schirra, D.D. Whitis, and C. Woodward, *Integrated Computational Materials Engineering: A Transformational Discipline for Improved Competitiveness and National Security,* The National Academies Press, 2008.
3. J. Allison, M. Li, C. Wolverton, and X.M. Su, Virtual Aluminum Castings: An Industrial Application of ICME, *JOM,* Nov 2006, pp. 28–35.
4. M. Li and J. Allison, in *Simulation of Aluminum Casting Process: From Alloy Design to Mechanical Properties,* Q. Wang, M.J.M. Krane, and P. Lee, Ed., TMS, Warrendale, PA, 2006, p. 91.
5. M. Li, R. Vijayaraghavan, C. Wolverton, and J.E. Allison, in *Metallurgical Modeling for Aluminum Alloys,* M. Tiryakioglu and L. Lalli, Ed., ASM International, 2003, p. 79.
6. C. Wolverton and J. Allison, U.S. Patent 6,858,103, 2005.
7. S. Weakley, W. Donlon, C. Wolverton, J.W. Jones, and J.E. Allison, *Metall. Trans.,* Vol 35A, 2004, p. 2407.
8. V. Vaithyanathan, C. Wolverton, and L.-Q. Chen, *Acta Mater.,* Vol 52, 2004, p. 2973.
9. V. Vaithyanathan, C. Wolverton, and L.-Q. Chen, *Phys. Rev. Lett.,* Vol 88, 2002, p. 125503.
10. C. Wolverton, X.-Y. Yan, R. Vijayaraghavan, and V. Ozolins, *Acta Mater.,* Vol 50, 2002, p. 2187.
11. C. Wolverton and V. Ozolins, *Phys. Rev. Lett.,* Vol 86, 2001, p. 5518.
12. J. Boileau, C. Cloutier, L. Godlewski, P. Reeber-Symanski, C. Wolverton, and J. Allison, SAE Paper 2003-01-0657, Society of Automotive Engineers, Warrendale, PA, 2003.
13. M. Caton, J. Jones, and J. Allison, in *Fatigue Crack Growth Thresholds, Endurance Limits and Design*, STP 1372, J.C. Newman and R.S. Piascik, Ed., ASTM, West Conshohocken, PA, 2000.
14. X. Su, J. Jan, J. Lasecki, and J. Allison, in *Thermal and Residual Stress Analysis for Engine Block with Cast-in Liners*, ASM Materials Solution and Exhibition, Oct 15–17, 2003 (Pittsburgh, PA), 2003.
15. U. Weiss, X. Su, J. Lasecki, and J. Allison, in *Proceedings of Haus der Technik*, Nov 26–27, 2003 (Essen, Germany), 2003.
16. H. Sehitoglu, X. Qing, T. Smith, H. Maier, and J. Allison, *Metall. Mater. Trans. A,* Vol 20A, 1989, p. 1755.
17. X. Su, M. Zubeck, J. Lasecki, C. Engler-Pinto, C. Tang, H. Sehitolgu, and J. Allison, SAE 2002-01-0657, Society of Automotive Engineers, Warrendale, PA, 2002.
18. J. Lasecki, X. Su, and J. Allison, SAE-2006-01-0324, Society of Automotive Engineers, Warrendale, PA, 2006.
19. D. Whitis, ICME at GE: Accelerating the Insertion of New Materials and Processes, *JOM,* Nov 2006, pp. 36–41.

Model Quality Management

Charles Kuehmann and Herng-Jeng Jou, QuesTek Innovations LLC

MODEL QUALITY is a key factor in the success or failure of any computational materials modeling activity. Unfortunately, no single or simple measure of model quality is available but must be referenced in the context of the intended use of the model. For instance, a model used to predict variation in manufacturing for a fixed product must exhibit high accuracy to small changes in process and material input parameters but does not need the ability to extrapolate accurately over large variations of composition or process variables. A model used to design a new alloy for a range of potential end uses must discern relative responses across large composition and process variations but may not need high absolute accuracy. Quality relates to the purpose of a model and is defined by the intended use and user.

To our benefit, many other disciplines have forged ahead of materials science and engineering in the application of computational modeling, and these disciplines have developed significant groundwork for the quality assessment of computational models. Organizational research was an early pioneer of computational modeling to predict the responses of large, complex systems, including companies, markets, social systems, and even whole economies, to variations in populations and public policy. Computational fluid dynamics is another pioneering application, and as early as 1998, the American Institute of Aeronautics and Astronautics published a guide for model verification and validation (Ref 1). A similar guide exists for solid mechanics, published in 2006 by the American Society of Mechanical Engineers (Ref 2). Additionally, because computational models are, by nature, software tools, the Institute of Electrical and Electronics Engineers has many available resources to assist in the verification of software codes (Ref 3). Because the objective of this article is to provide guidelines for the assessment of model quality in materials science and engineering, and considering the vast scope that materials models potentially entail, the reader is highly encouraged to review the aforementioned references to further explore the topic.

Fundamentals of Model Quality

Because the key assessment of the quality of a computational model is the ability of the model to achieve its intended purpose, the first cornerstone of that assessment is a clear statement of the intended purpose. The modeler must clearly indicate the intended use and objective of the model or simulation, including the range of parameters over which the model is expected to serve its' intended function. The underlying fundamental basis of the model is also a critical feature and must be communicated to the user, including key assumptions and simplifications that are employed in implementation. Lastly, details of the model calibration protocol and the verification and validation status are also requisite for a complete assessment of model quality. These issues are summarized in this article, and further details can be found in the appropriate references.

There is an intrinsic interplay between model building and experimental data development for calibration, verification, and validation purposes. This interplay is best represented by the Sargent circle in Fig. 1. The first step involves the development of an appropriate mathematical model based on the relevant mechanisms involved in the phenomena to be modeled. The mathematical model is influenced by the observations of reality in terms of the selection of appropriate mechanisms and applicable simplifying assumptions. Programming techniques are used to translate the mathematical model into computer code, and verification is employed to ensure the code accurately reflects the mathematical model. Predictions from the computational model can then be compared with observations of the real system to provide validation. Such verification may further provide insights into the true nature of the real system and allow further refinement of the operating mechanisms and their incorporation into the mathematical model. As such feedback is accomplished, the circular nature of the overall process is preserved.
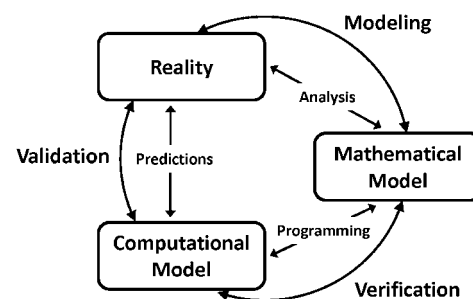
Fig. 1 Sargent circle for model verification and validation. Source: Adapted from Ref 4

Calibration of Mechanistic Material Models

Mechanistic models for materials play a critical role in materials science, design, and engineering. A mechanistic model, either process-structure model or structure-property model, captures the operating physical mechanisms in the mathematical formulation, based on a set of assumptions. Use of mechanistic material models allows extrapolation beyond the original experimental calibration/validation dataset, as long as the utilized mechanisms operate and the assumptions are valid. In addition, mechanistic material models are often hierarchical, building on top of other fundamental mechanistic submodels. An advantage of this approach is to promote model reuse.

Calibration protocols for mechanistic models must start with the calibration of any fundamental models incorporated into the modeling framework. An efficient protocol employs independent experimental measurements to decouple and/or minimize the cross interaction between adjustable model parameters, allowing the determination of the parameters with high fidelity and minimal overfitting. The protocol must also logically and sequentially address the data and parameters used by the model. The development of an experimental procedure to produce calibration data must avoid any

significant risk of invalidating the model assumptions. During the calibration, it is useful to estimate the accuracy of the model parameters, based on instrumental precision, experimental variation, and variation from fundamental models used. The requirement for model parameter accuracy will depend on the application of the model. When more accuracy is required, it is desirable to extend and refine the calibration procedure. A calibration is considered complete when satisfactory accuracy of the parameters is obtained. Several examples are provided in the following discussion to demonstrate the principles outlined previously.

Example: Martensitic and Bainitic Transformation Kinetics in the Austenitic Decomposition of Steels. A mechanistic model for prediction of martensitic and bainitic kinetics in multicomponent steels is valuable for alloy design of high-performance steels (Ref 5). Conventional empirical models are limited to composition ranges within a range of known experimental experience. Ghosh and Olson (Ref 6) established a foundation of mechanistic martensitic kinetics, later extended by Y. Mitsutani (Ref 7) for bainitic kinetics. These models are built hierarchically, using other models for thermodynamics (CALPHAD based) (Ref 8), elastic modulus (Ref 9), and distributed defect potency (Ref 10). Calibration of such a complex model begins with subcalibration of the individual models from independent experimental data, such as thermodynamic T_0 temperatures, composition-dependent shear moduli, and pre-existing defect-size distribution data from small-particle experiments. The few remaining unknown model parameters can then be calibrated by a limited set of experimental martensitic and bainitic start temperatures. This model is described in more detail later in this article.

Example: Mechanistic Precipitation Models. Precipitates represent an important class of microstructure in alloys, due to their significant impact on properties such as strength, toughness, and fatigue. PrecipiCalc (Ref 11) was created with a goal of modeling the evolution of precipitate microstructure in multicomponent alloys during complex heat treatment in reasonable computation speed. Two major submodels of PrecipiCalc are CALPHAD thermodynamics and mobility databases, and a key PrecipiCalc model parameter is matrix/precipitate interfacial energy. To decouple the interaction of these model parameters, an attempt was made to develop a robust calibration protocol (Ref 12). In the procedure, the underlying thermodynamics and mobility databases were selected based on the desirable accuracy of model parameters. Conventionally, coarsening experiments were often used to determine the interfacial energy based on the well-established correlation of the interfacial energy to the coarsening-rate constant. However, the coarsening rate depends both on interfacial energy and diffusivity. By using a novel nucleation-onset temperature experiment, the interfacial energy can be determined while

minimizing the potential impact due to uncertainties in the available diffusivity data. The nucleation-onset temperature depends exponentially on interfacial energy but only linearly on the diffusivity. With such a technique, it is possible to decouple the model parameters and allow them to be independently calibrated with experimental data. The extent to which model parameters can be independently calibrated and validated improves model quality.

Example: Analytical Precipitation-Strengthening Model. Many high-performance alloys depend on fine precipitates to provide superior strength. A useful mechanistic and analytical precipitation-strengthening model takes precipitation microstructure (size, fraction, and distribution) as an input and provides an estimate of yielding stress. Depending on the strengthening mechanisms, submodels may include the elastic moduli of the matrix and precipitates, antiphase-boundary (APB) energies, lattice mismatch, and others. The calibration of a precipitation-strengthening model should start with these submodels to ensure that desirable accuracy is available for these key input parameters. Error and uncertainty from submodel inputs have a strong effect on the overall accuracy of higher-order complex models. If the submodel accuracy is insufficient and independent data are unavailable to further improve it, the submodel parameters can become a part of the calibration of the higher-order model. For example, a lack of accurate APB energy predictions for ordered precipitate phases requires this quantity to be calibrated experimentally with strength data within a precipitation-strengthening model. A further complication here is the determination of the precipitation-strengthening contribution of the overall strength in alloys. In addition to precipitation strengthening, alloys are often strengthened by multiple mechanisms, including solid-solution, Hall-Petch (grain size), and dislocation strengthening. Accurately backing out the precipitation-strengthening contribution from the overall alloy strength requires careful experimental design and reliable strengthening models for the other relevant mechanisms.

Model Verification

Verification of computational science software determines the accuracy of the software implementation against the original mathematical models (Fig. 1). In high-performance computational science, several published articles address recommended best practices in numerical algorithm verification, numerical error estimation, and software quality assurance. Materials scientists and engineers should certainly apply these established guidelines and formal software engineering methodologies whenever possible to verify the established materials modeling software. In this section, the authors share some observations on the area of computational materials science software.

Materials models can be a simple algebraic formula or a set of complex differential/integral equations. When there is a software development need, the method of implementation typically falls into one or more of the following categories:

- Commodity software, such as spreadsheets
- Scripts or macros running within general-purpose scientific software, such as MatLab and Mathematica
- Scripts or macros running within specific-purpose scientific software, such as Thermo-Calc and ABAQUS
- General programming languages (such as Python, C++, and Java), often using special-purpose application programming interfaces (APIs) such as the NAG or IMSL numerical libraries. (Commercially available NAG and IMSL libraries contain large collections of well-tested numerical algorithms in several different programming languages. These reliable numerical libraries can significantly reduce the amount of work in software verification when developing a materials model.)

With many choices of available software programs and components to assist materials model implementation, the implementation method depends on the complexity of mathematical equations, familiarity of available software, computational efficiency, end-user requirements, and, sometimes to a lesser extent, verification procedures. Materials modeling software developers often rely on well-established third-party programs and software components to deal with the extensive numerical solving portion of the product. However, the verification protocol and software quality assurance of the final overall product is usually not in the original consideration when choosing implementation methods, and it could result in a software product that is difficult to test for accuracy and to maintain the quality. It is highly recommended that materials model developers take into account the model verification consideration during software implementation planning to identify suitable programs, software components, and programming languages.

Model Validation

Validation is the processes used to determine that a model possesses suitable accuracy for use in the context of its intended purpose. It is not an absolute state but a spectrum based on the confidence of the intended user population that can vary with time and in comparison to other options to achieve similar results. Validation essentially involves applying tests of the model to explore its limits of accuracy and applicability. It is important to understand that failure of any single test in validation is not a failure of the model itself but is essential in defining limits and boundaries of validity, in

essence, improving the usefulness of a model. Validity tests can take many forms, a number of which are discussed subsequently.

Boundary-value tests exercise the limits of a model at the extremes of potential input parameters to determine if the results conform to expected behaviors. An example of this approach could be precipitate growth at very long times, where classical particle coarsening would be expected. When a model behaves well at these extreme conditions, confidence in the model results for other conditions is enhanced.

Degenerate Problem Tests. Comparing the results of the model for well-known, simplified degenerate problems for which exact analytical solutions may be available is another key test of validity. In this case, for example, a multicomponent model can be exercised for a simpler binary system, where analytical solutions may exist. Such tests can also be applied to degenerate problems for which other model predictions have been well validated and established. The degree to which the complex model can accurately depict the behavior of the simpler analytical or model-based solution increases the expectation that the solutions to more complex problems are valid.

Sensitivity tests determine the response of the system to variation in input variables and model parameters. If the model response to such variation is consistent with expected behavior and general physical principles, higher confidence in the model results is achieved. Anomalies or unexpected sensitivities contrary to expectations must be further explored before a model is deemed invalid. Nonlinear behavior in materials phenomena can make it extremely difficult to discern true "normal" behavior in complex systems, which is precisely the reason that computational modeling has a large utility in this field.

Comparisons with Real-World Observations. Comparisons of model outputs against real-system observations are the gold standard for model validation. It is important to consider that empirical data are only another representation of the real system and are subject to their own biases, errors, and complexities. In comparing data to simulation or model results, quantification of the fidelity of the dataset is key to making a strong case for or against model validation. Significant difficulties can arise when test conditions are not well known, controlled, and documented. In a perfect case, all simulation inputs and boundary conditions would be well controlled and reported for the dataset used for validation, but such a situation is atypical. Often, missing conditions must be assumed, and it is worthwhile to exercise the model for a range of expected values of such conditions to estimate the range of possible outcomes from the model based on the variation.

Commonly, data-driven validation efforts collect as much data as possible from available sources and compare these values with model predictions, using statistical methods to remove the effects of so-called outliers. A more appropriate method, however, evaluates the empirical dataset for inconsistencies, errors, and biases prior to comparison with model predictions. If outliers exist after closely evaluating the available data, they must be closely studied for possible insight into deficiencies in the fundamental structure or assumptions of the model. When a valid dataset is achieved, then statistical analysis techniques can be used to evaluate the degree of correlation of the data-driven estimate of the system response to the computational model-derived estimate.

Engineers seeking stronger validation of computational models are highly encouraged to produce unique data specifically for the validation exercise. Such validation experiments can be tightly controlled and all model input parameters closely measured. Such data can effectively span the range of expected conditions for which the model will be used. In addition, sample characterization and data relating to underlying assumptions of the model can be developed, further enhancing the validation process.

Benchmarking is a subset of comparative validation that can use either well-known and validated experimental datasets or model results that, in themselves, have been well validated with multiple techniques and methods. The utility of benchmarking is directly related to the availability of appropriate benchmark cases for the phenomena of interest. Thus, it is imperative that, as computational materials modeling gains stronger adoption, the community allocates appropriate resources to publish and evaluate modeling results, thus establishing appropriate and important benchmarks. Benchmarking can go beyond validation and also measure relative computational speed, accuracy, and breadth for available modeling approaches and implementations, providing valuable information for users

to discern the best models and for modelers to further refine their efforts.

Example of Model Calibration, Verification, and Validation—Martensite Start Temperature Prediction for Steels

To provide readers with a complete model-development cycle, an example based on mechanistic martensite start temperature (M_s) is provided in this section.

M_s Model Calibration. The mechanistic nature of a martensite kinetics model demands a multilevel modeling hierarchy, as shown in Fig. 2. This hierarchy establishes the fundamental mechanisms involved in the heterogeneous nucleation of martensite. Using this modeling hierarchy, the M_s modeling diagram is described in Fig. 3 and establishes the mathematical representation to be used to describe the fundamental mechanisms in Fig. 2. Thus, the calibration of model parameters and submodels requires independent calibration in various structural levels:

- Multicomponent CALPHAD thermodynamics database (Ref 8)
- Composition- and temperature-dependent austenite shear modulus (Ref 9)
- Potency distribution of autocatalytic nucleation sites (Ref 10)
- Potency distribution of initial nucleation site (Ref 13)

After calibration of the aforementioned sublevel models, calibration of the critical driving force, ΔG_{crit}, which includes surface energy and frictional work by solute strengthening and dislocation, was established in Ref 5 and
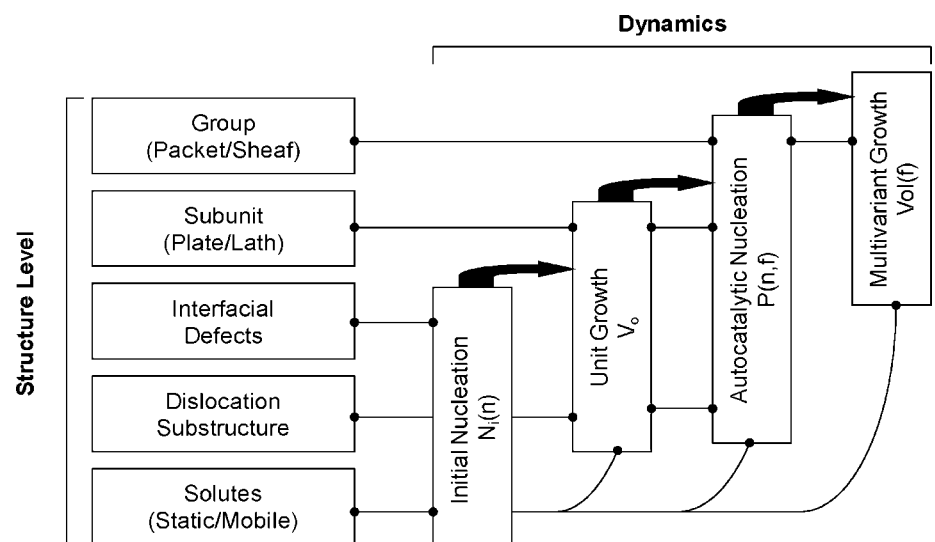
Fig. 2 System diagram for the dynamic structural hierarchy for heterogeneous nucleation of martensite
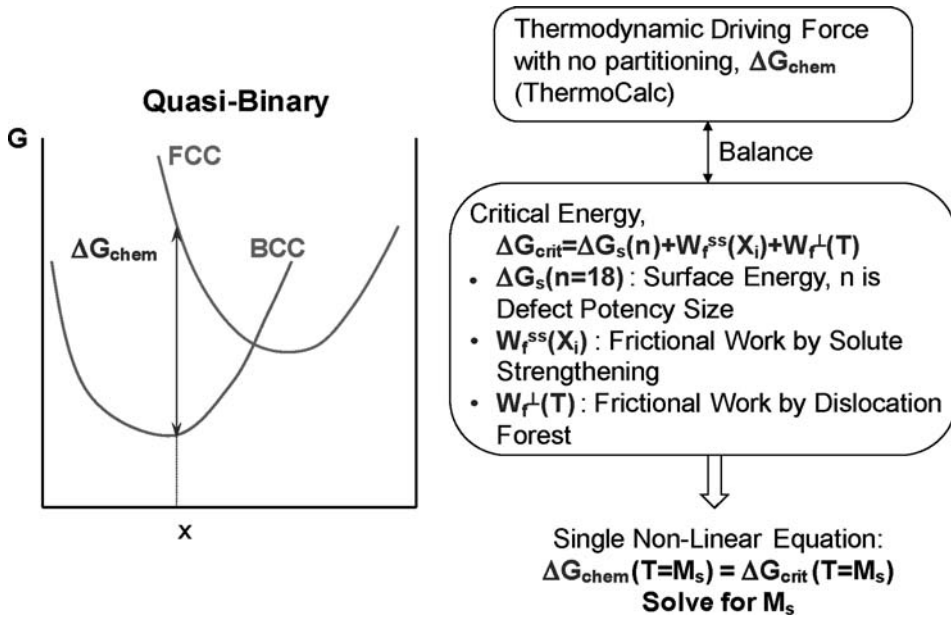
Fig. 3 Architecture of mechanistic martensite start temperature (M$_s$). fcc, face-centered cubic; bcc, body-centered cubic. Source: Computational model, Ref 6
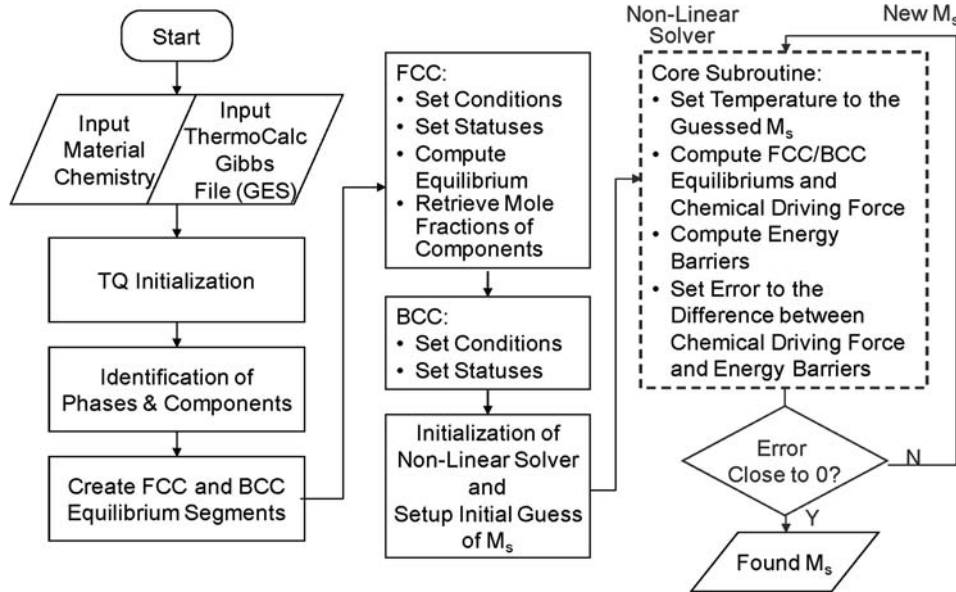
Fig. 4 Software implementation flow diagram of a computational martensite start (M$_s$) temperature model. fcc, face-centered cubic; bcc, body-centered cubic

6, using approximately 120 experimental M$_s$ measurements published in the literature.

M$_s$ Model Verification. Implementation of the M$_s$ model required two key software components:

- Thermo-Calc TQ API (Ref 14) for calculation of the nonpartitioning Gibbs free energies for the body-centered cubic and face-centered cubic phases using the CALPHAD database calibrated and established earlier (Ref 8)

- The NAG library was used to iteratively search for the root of a nonlinear energy balance equation of one unknown, the M$_s$ temperature. The numerical accuracy of the solution was set to be less than 1 °C and is more than adequate, considering the experimental error is at least 10 °C.

The overall software architecture is presented in Fig. 4. The results of the software code implemented according to this architecture were compared to an earlier implementation used in the Thermo-Calc software to ensure the software achieved consistent accuracy.

M$_s$ Model Validation. Using the verified M$_s$ software implementation, predictions of M$_s$ temperature were made for a large dataset with approximately 350 steel compositions. Figure 5 (a) shows a correlation plot of the experimental measurements to the model predictions. The points on the solid diagonal line represent a perfect correlation. The two diagonal dashed lines represent a departure of 30 °C on either side of the perfect 1 to 1 correlation. The data represent many classes of steels, including low-alloy steels, stainless steels, and cobalt-nickel steels. Figure 5(b) is a probability plot of the M$_s$ prediction error, showing a standard deviation of 35 °C, similar to the experimental error. Considering the broad range of steel compositions, the mechanistic M$_s$ computational model is found to provide acceptable prediction capability and has been used in the design of a number of new martensitic alloys (Ref 15, 16).

REFERENCES

1. "Guide for the Verification and Validation of Computational Fluid Dynamics Simulations," AIAA-G-077-1998, American Institute of Aeronautics and Astronautics, Reston, VA, 1998
2. "Guide for Verification and Validation in Computational Solid Mechanics," ASME V&V 10-2006, American Society of Mechanical Engineers, Fairfield, NJ, 2006
3. *Standard for Software Verification and Validation,* Institute of Electrical and Electronics Engineers, June 2004
4. R.G. Sargent, Verification and Validation of Simulation Models, *Proceedings of the 2003 Winter Simulation Conference,* S. Chick, P.J. Sanchez, D. Ferrin, and D.J. Morrice, Ed., The Society for Computer Simulation International (SCS), San Diego, CA, 2003, p 37–48
5. G.B. Olson, K.C. Hsieh, and H.K.D.H. Bhadeshia, On the Nucleation Driving Force for Displacive Transformations in Steels, *Microstructures LCS '94,* Iron and Steel Inst. of Japan, Tokyo, 1994
6. G. Ghosh and G.B. Olson, *Acta Metall.,* Vol 42, 1994, p 3361
7. Y. Mitsutani, "The γ/α Phase Transformations in Low-Carbon Steels," Ph.D. thesis, Northwestern University, Evanston, IL, 2006
8. G. Ghosh and G.B. Olson, *J. Phase Equilibria,* Vol 22 (No. 3), 2001, p 199
9. G. Ghosh and G.B. Olson, *Acta Metall.,* Vol 50, 2002, p 2655
10. M. Lin, G.B. Olson, and M. Cohen, *Metall. Trans. A,* Vol 23, 1992, p 2987
11. H.-J. Jou, P. Voorhees, and G.B. Olson, Computer Simulations for the Prediction of Microstructure/Property Variation in Aeroturbine Disks, *Superalloys 2004,* K.A. Green, T.M. Pollock, H. Harada,

Mean= −0.54 +/− 3.52(95%), Std Dev= 35.29
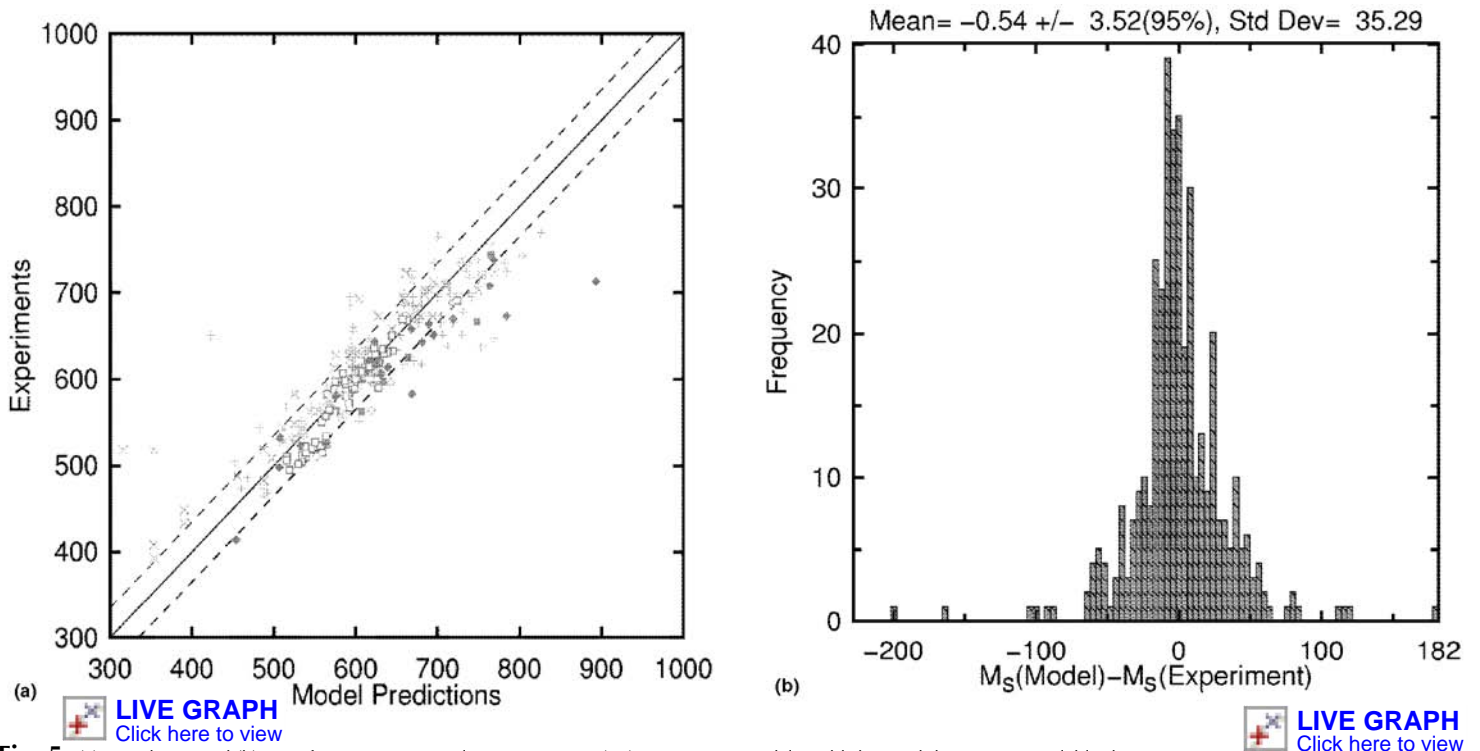
Fig. 5 (a) Correlation and (b) error for a computational martensite start (M_s) temperature model establishing validity against available datasets

T.E. Howson, R.C. Reed, J.J. Schirra, and S. Walston, Ed., The Minerals, Metals and Materials Society, 2004, p 877–886

12. G. Olson, H. Jou, J. Jung, J. Sebastian, A. Misra, I. Locci, and D. Hull, Precipitation Model Validation in 3rd Generation Aeroturbine Disc Alloys, *Superalloys 2008*, R.C. Reed, K.A. Green, P. Caron, M. Fahrmann, T. Gabb, E. Huron, and S. Woodard, Ed., The Minerals, Metals and Materials Society, 2008, p 923–932

13. G.B. Olson and M. Cohen, New Aspects of Martensitic Transformation, *Jpn. Inst. Met.,* 1977, p 93–98

14. Thermo-Calc Software, http://www.thermocalc.com/Products/TQ.html

15. C.E. Campbell and G.B. Olson, Systems Design of High Performance Stainless Steels: I. Conceptual Design, *J. Comput.-Aided Mater. Des.,* Vol 7, 2001, p 145–170

16. C.J. Kuehmann and G.B. Olson, Computational Materials Design and Engineering, *Mater. Sci. Technol.,* Vol 25 (No. 4), 2009 p 472–478

Fundamentals of Process Modeling

Modeling of Deformation Processes— Slab and Upper Bound Methods

Rajiv Shivpuri, The Ohio State University

THE GOVERNING RELATIONS OF PLASTICITY used in metal forming are highly nonlinear, with complex boundary conditions. They can only be solved for a limited number of idealized processes with known boundary conditions. Several examples of the ideal solutions are included in textbooks by Hill (Ref 1), Johnson and Mellor (Ref 2), Hosford and Caddell (Ref 3), and Thomsen et al. (Ref 4). Due to the inherent assumptions, these solutions have limited industrial applications and are primarily used to calibrate approximate models, such as slab, upper bound, or finite-element approaches. To solve more industrially relevant problems, approximate techniques have been developed that model more geometrically complex problems with industrially acceptable boundary conditions. The most popular of these methods are the slab method (SM), the upper bound method (UBM), and the finite-element method (FEM). The slab method and upper bound method of analysis assume a known deformation field in the plastically deforming body and assume this field either satisfies the force equilibrium (SM) or power-conservation equations (UBM) to determine the stress field or the external load applied. In the slab analysis, the stress equilibrium is achieved by assuming a simple homogeneous deformation field and ignoring deformation compatibility. In the UBM, deformation and compatibility equations are satisfied by the assumed kinematically admissible velocity fields, while the stress equilibrium equations are not. Therefore, the slab method often provides a lower bound and the UBM an upper bound to the actual plastic field.

The attractiveness of these methods is that they provide closed-form solutions to metal-forming problems, and therefore, these equations can be used:

- For process optimization
- For preform design using inverse techniques
- As a benchmark for validating predictions of higher-fidelity numerical models

Both these methods have been extended to more complex deformation fields by discretizing the deforming body into finite-volume elements; examples of these are the finite slab element method (FSEM) and the upper bound elemental technique (UBET).

Several researchers have compared various modeling techniques to determine their efficacy in predicting metal flow and forming forces. For example, Shivpuri and Chou (Ref 5) provided a comparison of the SM, UBM, and FEM approaches in predicting force and torque in cold flat rolling. They found that each approach had its own region of superior performance. More recently, Hsu and Tzou (Ref 6) compared slab and UBM approaches in modeling double-layer clad sheet compression forming. In this, the metal flow is analyzed using velocity fields (UBM) and the die pressures using balance of forces (SM). They found the two approaches to provide complementary results.

The Slab Method

The slab method was made popular in the early 1900s by Sachs (Ref 7), Siebel (Ref 8), von Karman (Ref 9), and Orowan (Ref 10) by extending the free-body equilibrium approach to differential elements in the deforming field. They applied it to the simple deformation fields of forging (upsetting), drawing, and rolling by using the assumptions that the principal axes of deformation are normal to the applied loads, and the effects of friction do not change the direction of principal axes or distort the deformation field. This results in homogeneous deformation, with plane sections remaining plane. These assumptions converted the partial differential equations of plasticity to ordinary differential equations of first order that can be solved by simple force-boundary conditions on the die. Typically, the slab analysis procedure consists of the following steps (Ref 3):

- Assume a homogeneous deformation field in the plastically deforming field defined by the principal axes of strain and stresses.
- Isolate a slab (differential element) by taking two cuts perpendicular to the principal direction, X_i, with thickness dX_i and the other dimensions defined by the tooling.
- Derive equilibrium equations (differential equations) by balancing the forces on this differential element in the principal directions.
- Solve the differential equations by integrating with the force-boundary conditions.
- Obtain an equation that relates the dimensions of the deforming zone to the applied forces.

An example of application of the slab method to plane-strain drawing is shown in Fig. 1. In this, the force balance on the differential elements yields the following equation:

$$\frac{\sigma_d}{2k} = \frac{1+B}{B}\left[1 - \left(\frac{t_e}{t_o}\right)^B\right] \qquad \text{(Eq 1)}$$

where σ_d is the drawing stress, k is the shear yield stress, $B = \mu \cot \alpha$, μ is the friction coefficient, α is the half-die angle, t_e is the exit strip thickness, and t_o is the entry strip thickness. This nonlinear equation can be used for process design, such as determining the optimum die angles for a given reduction. In this solution, the friction is kept constant over the entire die-workpiece contact surface, the material does not work harden, and the half-die angle is a constant. However, it is possible to extend this method to work-hardening material, discrete variations in die angle, and nonconstant friction by dividing the plastic zone into multiple zones. Then, force conservation is applied to each zone, with the unknown constant determined by the stress compatibility at the interface of these zones. This approach has been extended to other steady-state metalforming problems, such as rod drawing, tube drawing, plane-strain rolling (bloom, slab, or strip), and plane-strain strip deep drawing (sheet forming).
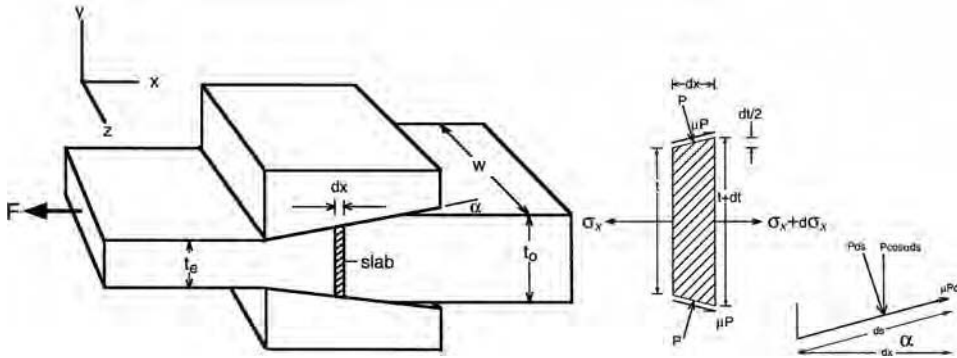
Fig. 1 Schematic of the slab element in plane-strain drawing of a plate (left) and the balance of forces on the selected slab, where F = drawing force, t_e = thickness of plate at exit, t_o = thickness of plate at entry, α = half-angle of the die, w = width of the plate, dx = thickness of differential element, σ_x = drawing stress, $d\sigma_x$ = incremental stress, dt = incremental thickness, p = die pressure, and μ = friction factor. Source: Ref 3

Fig. 2 Essentials for a slab force balance. Source: Ref 14

Selected examples of application of this method to metalforming processes are included in the following.

Example 1: Drawing of Plates, Rods, and Tubes. The slab method is ideally suited to the plane-strain or axisymmetric drawing of plates, rods, and tubes. In plane-strain drawing, the slab element is a rectangular plate (Fig. 1), while in axisymmetric drawing, it is a round disc (in rod drawing) or a ring (in tube drawing). In either case, the forces (or stresses) are balanced on the differential element, with friction taken as uniform at the tool-slab boundary. This results in a friction hill that is linear for the friction factor model and nonlinear for the coefficient of friction model, with the die pressure increasing from the open boundary toward the center and reaching a

maximum at the neutral surface. Due to the force-balancing approach, the slab method permits the inclusion of forward and back tensions in the model formulation. In a recent paper, Rojas et al. (Ref 11) used the slab method to determine tensions in the drawing of symmetric profiles using various work-hardening models. Because the work-hardening models are nonlinear (and coupled), the drawing problem was solved incrementally. The tension predictions from the slab analysis compared favorably with results from the analytical, FEM (Ref 12), and experimental results (Ref 13).

Example 2: Plane-Strain Compression of a Block. Figure 2 illustrates the plane-strain compression of a block, where $h < b$, and sliding friction with a constant coefficient, μ, prevails

at the interfaces. A force balance on the slab in the x-direction is $\sigma_x h + 2\mu P\, dx - (\sigma_x + d\sigma_x)h = 0$, which simplifies to:

$$2\mu Pdx = hd\sigma_x \qquad \text{(Eq 2)}$$

Again, σ_x and σ_y (taken as $-P$) are considered as principal stresses. Note that with low μ, the shear influence is again considered to be small with respect to the directions of the principal stresses. For plane strain, $\sigma_x - \sigma_y = 2k$ or $\sigma_x + P = 2k$, so $d\sigma_x = -dP$. Now, Eq 2 can be expressed as:

$$2\mu\ Pdx = h\ dP \quad \text{or}$$
$$\frac{dP}{P} = \frac{2\mu}{h}dx \qquad \text{(Eq 3)}$$

where at $x = 0$, $\sigma_x = 0$ and $P = 2k$. Integrating:

$$\ln\frac{P}{2k} = \frac{2\mu x}{h} \qquad \text{(Eq 4)}$$

or

$$\frac{P}{2k} = \exp\frac{2\mu x}{h} \qquad \text{(Eq 5)}$$

This is valid from the surface ($x = 0$) to the centerline ($x = b/2$), where P is a maximum:

$$\left(\frac{P}{2k}\right)_{\text{max}} = \exp\frac{\mu b}{h} \qquad \text{(Eq 6)}$$

It should be noted that there is a limitation on the values of μ for which this development is valid. The shear stress at the interface, which in the derivation is taken as μP, cannot exceed the shear strength of the work material, k. Because $\mu P \leq k$, Eq 4 and 5 are valid only for:

$$x \leq \frac{-h}{2\mu}\ln 2\mu \qquad \text{(Eq 7)}$$

and

$$\frac{b}{h} \leq -\frac{1}{\mu}\ln 2\mu \qquad \text{(Eq 8)}$$

The rise of P toward the centerline shown in Fig. 3 is referred to as the friction hill.

Of great interest is the mean pressure on the tool-workpiece interface. This pressure, P_a, times the contact area is the force that must be applied.

Average Pressure during Plane-Strain Compression. The force acting on the left side of the centerline is:

$$F_y = \int_0^{b/2} Pdx = \int_0^{b/2} 2k\exp\frac{2\mu x}{h}dx$$
$$= 2k\left[\frac{h}{2\mu}\exp\left(\frac{\mu b}{h}\right) - 1\right] \qquad \text{(Eq 9)}$$

so the average pressure is F_y divided by the area on which it acts, $P_{\text{av}} = F_y/(b/2)$, or:

$$\frac{P_{av}}{2k} = \frac{h}{\mu b}\left[\exp\left(\frac{\mu b}{h}\right) - 1\right] \qquad \text{(Eq 10)}$$

Expanding, $\exp(\mu b/h) - 1 = 1 + (\mu b/h) + (\mu b/h)^2/2! + \cdots - 1$, so for small values of $\mu b/h$:

$$\frac{P_{av}}{2k} = 1 + \frac{\mu b}{2h} + \cdots \approx 1 + \frac{1}{2}\frac{\mu b}{h} \qquad \text{(Eq 11)}$$

Numerical Example. Plane-strain compression is conducted on a slab of metal whose yield shear strength, k, is 15,000 psi. The width of the slab is 8 in., while its height is 1 in. Assuming the average coefficient of friction at each interface is 0.10:

1. Estimate the maximum pressure at the onset of plastic flow.
2. Estimate the average pressure at the onset of plastic flow.

Solution. For problem 1, from Eq 6:

$$P_{max} = 2k\exp\left(\frac{\mu b}{h}\right) = 30,000\exp\left(\frac{0.1 \times 8}{1}\right)$$

so, $P_{max} = 30,000(2.226) = 66,800$ psi.

For problem, first use the exact solution given by Eq 10:

$$P_a = \frac{30,000}{0.8}(e^{0.8} - 1) = 46,000\,\text{psi}$$

Next, use the approximation given by Eq 11:

$$P_a = 30,000(1 + 0.4) = 42,000\,\text{psi}$$

This indicates that a value of 0.8 for $(\mu b)/h$ should not be considered small in this context. For instance, if b were 4 in., h were 1.5 in., and μ were 0.05, then $(\mu b)/h$ would be 0.133. The respective values of P_a using Eq 10 and 11 would then be 32,100 and 32,000 psi, which are nearly the same.

Example 3: Flat and Shape Rolling. The slab method has been widely applied to the flat rolling process for the determination of roll loads, roll friction, roll tensions (backward and forward, Ref 15), and roll deformations. Some slab method solutions have been combined with other solution methods, such as UBM and FEM, to handle deformation gradients, while others have been used in tension algorithms for gage control. A comprehensive analysis of the roll tensions in flat rolling was provided by Bland and Ford (Ref 16) and later extended by Alexander (Ref 17). The approach used by Bland and Ford, included in the textbook by Rowe (Ref 18), is shown in Fig. 4. In this approach, the forces on the slab element are calculated as it enters the roll gap and proceeds toward the exit. The location of the neutral plane (which determines the direction of friction) is determined by balancing the forces in the horizontal direction, including the forward and backward tension. These tensions affect the position of the neutral point and, in doing so, determine the amount of

Fig. 3 Friction hill in plane-strain compression with a constant coefficient of friction. Source: Ref 14

forward and backward slip between the rolls and the sheet. The well-known solution for the roll pressure, attributed to Bland and Ford, is:

$$\ln\left(\frac{p}{S}\right) = \ln\left(\frac{h_a}{R} + \alpha^2\right) \pm 2\mu, 1/\sqrt{(h_1\alpha/R)}\tan^{-1}$$
$$\left(\frac{\alpha}{\sqrt{h_1\alpha/R}}\right) + \text{constant} \qquad \text{(Eq 12)}$$

where p is the roll-separating pressure, S is the yield strength of the sheet material, h_a is the exit thickness of the sheet, R is the roll radius, α is the angle of the bite, and μ is the coulomb coefficient of friction. The positive sign (+) applies to the exit side of the neutral plane and the negative sign (−) to the entry side of the roll bite. Front and back tensions ($\pm t$) can easily be incorporated in this solution by replacing the front and back pressures at the two ends as $p = S - t$. In either case, roll tensions reduce the roll pressures considerably.

More recently, the slab approach has been applied to cold rolling of thin foils (Ref 19) and combined with the rigid-plastic FEM algorithms to model shape rolling (Ref 20, 21) and flat rolling of wire (Ref 22). In the shape rolling models of Shin and Shivpuri, the deformation of the slab element normal to the plane of the

slab (longitudinal direction) is neglected. Hence, plane sections remain plane, and force balance can be applied (Fig. 5). However, deformation in the transverse plane due to the shape of roll passes (caliber) is included, using rigid-plastic FEM. This incorporation of slab element in the FEM formulation considerably speeds up the calculation procedures. For example, on a standard personal computer using an Intel Core 2 processor, an eight-pass shape rolling sequence using FSEM can be simulated in half an hour, while it takes several days using the fully coupled three-dimensional FEM model. More recently, semiempirical relations of microstructural evolution have been incorporated in the FSEM formulation to simultaneously model the metal flow and thermomechanical history during the shape rolling process. This simplification permits the inclusion of microstructural evolution equations in the formulation (Ref 23).

Example 4: Extrusion. The application of the slab method to extrusion has not been as widespread as in plane-strain drawing and flat rolling. This is primarily due to the complexity of deformation in extrusion, with considerable distortion. Due to this deformation complexity, the inherent assumptions of plane sections remaining plane in the slab method are not valid in extrusion. Other modeling methods,

Fig. 4 Geometric representation of plane-strain deformation in the rolling of a plate, strip, or sheet. (a) Slab element and boundary conditions. (b) Force equilibrium on the slab element. Source: Ref 18. The symbols are self-explanatory.

connecting rod forging (Ref 26), cold heading (Ref 27), axisymmetric closed-die forging (Ref 28), sintered billets (Ref 29), and spur gear preforms (Ref 30). In some of these applications, the slab method has been compared with UBM and FEM approaches and found to be sufficiently accurate for load determination.

Upper Bound Method

The upper bound method is based on the principle of virtual work or virtual power, which states that for a closed system, work/power is conserved. In an isothermal deforming medium, this can be stated as:

> Internal work (Internal power)
> = External work (External power)

Often in metalforming, these work elements can be expressed as:

> Internal work = Homogeneous work
> + Redundant work
> External work = Work by applied forces
> + Work by frictional forces

In the application of this approach to isothermal metal forming, a common procedure is (Ref 3):

1. Assume a velocity field inside the deforming body consisting of a finite number of velocity discontinuities or nonlinear velocity expressions with unknown parameters.
2. Represent internal work in terms of this velocity field.
3. Represent external work in terms of different velocity fields.
4. Formulate the work balance (or power balance) equations based on the virtual work principle.
5. Solve for forces required, for given workpiece deformation, or for unknown velocity parameters if the external forces are known.
6. In the aforementioned calculations, friction is assumed as a penalty factor on the velocity at the interface.

When the internal energy dissipation is distributed over a finite number of internal planes or surfaces, as shown in Fig. 6, the governing equation inside the deforming body can be expressed as (Ref 3):

$$dW/dt = \sum_i k.S_i \cdot V_i^*$$
$$= ((kV_\perp(12)^\top *)/V_\perp x)SV_\perp(x) = kSV_{12}^* \tag{Eq 13}$$

where the planes and velocity discontinuities are shown in Fig. 6(a), and the velocity compatibilities are satisfied through the velocity vector diagram (Fig. 6b), also known as a hodograph. This discrete description of velocity

such as the upper bound method, are more suited to extrusion processes. However, in selected cases with more uniform deformation (rods, tubes, and plates), the slab method may be applicable. Such a case of axisymmetric extrusion of bimetallic tube has been modeled by Chitkara and Aleem (Ref 24, 25) using the slab method. They assumed two different slabs for the two materials and then sought force balance at the bimetallic boundary. The results of the analysis compare favorably with the experiments carried out on bimetallic tubes comprising high-conductivity copper and commercially pure aluminum. They included both the copper-aluminum and the aluminum-copper pair.

Example 5: Forging. In non-steady-state deformation, such as forging, the slab analysis is often applied by assuming the deformation to consist of a series of equilibrium states (quasi-static states) where steady state can be assumed. This time-history approach has been used effectively, with time taken as the integration constant. Example applications include calculation of forging load for

In the continuum formulation, the upper bound method is expressed on the volume of the deforming continua. In this approach, a virtual displacement is imposed on the deforming field while keeping the forces constant. For a small volumetric element, this is equivalent to imposing a first-order variation on the strain-energy field, given by:

$$dE = \int \sigma_{ij} \, d\varepsilon_{ij} \, dV = \int F_i du_i \, dS \qquad \text{(Eq 14)}$$

where E is the strain energy in the deforming differential volume, dV; σ_{ij} is the stress field; $d\varepsilon_{ij}$ is the strain increment; F_i is the component of the external force applied; du_i is the incremental displacement, and dS is the external surface area of the body with volume V. In a nonstationary field, the deformation is time dependent, and the displacement increment can be replaced by velocity, giving the equation for power balance as:

$$\dot{E} = \int \sigma_{ij} \, \dot{\varepsilon}_{ij} \, dV + \int F_i \, v_i \, dS + \int \tau_i \, v_i \, dS \quad \text{(Eq 15)}$$

Because the stress is held constant during the deformation as the deformation increases from zero to the current value, it provides an upper bound on the power required for deformation. This is also the reason why the theory based on UBM is often called the flow theory of plasticity. The procedure for applying the upper bound to metalforming problems is:

1. Assume a family of admissible velocity fields that satisfy the continuity conditions, the conditions of incompressibility (compatibility conditions), and velocity boundary conditions.
2. Calculate the power (energy rate) for the deformation that consists of homogeneous deformation, redundant shear (inside body), and frictional shear (on the surface).
3. The velocity field that minimizes the power is the closest to the actual deformation field.
4. Calculate the external load needed for the deformation by dividing the power with the relative velocity between the die and the deforming material.

Early examples of the application of UBM to steady-state metalforming processes, especially extrusion and drawing, are included in the books by Avitzur (Ref 31–33). Later, UBM was applied successfully to non-steady-state deformation, such as in forging, by treating these processes as a series of quasi-steady-state processes and iterating over time (Fig. 7).

Recently, the UBM methods have been extended to numerical analysis using the following approaches:

- Break the body into finite elements and define the velocity fields in a piecewise continuous manner. This approach, often known as the upper bound elemental technique (UBET), has been applied successfully to

5 . 71 = A		
5 . 38 = B		
5 . 05 = C		
4 . 72 = D		
4 . 39 = E		
4 . 06 = F		
3 . 73 = G		
3 . 40 = H		
3 . 07 = I		
2 . 74 = J		
2 . 41 = K		
2 . 08 = L		
1 . 75 = M		
1 . 42 = N		

Generated by tasks V1.1

Ub — Incoming velocity
Uf — Outgoing velocity
Sb — Back tension
Sf — Front tension
Uj — Velocity of the j-th section
ε_{zj} — Longitudinal strain (rate) of the j-th section

Fig. 5 Formulation of the finite and slab element method for the shape rolling process. In the top figure, the deforming workpiece is cut into a finite number of slabs. The force balance is applied on the slab element (J_{th} cross section), while the transverse deformation is handled through the finite-element method. Source: Ref 12, 20

discontinuities in the deforming medium is widely used in solving two-dimensional metalforming problems with a simplified description of friction, flow work hardening, and heat transfer. In most cases, these phenomena are taken to be constant on a particular discontinuity plane.

Fig. 6 (a) Change in metal flow across velocity discontinuities in a deforming medium (*Y-Y'*). (b) Hodograph or vector diagram for the velocity discontinuities used in the work calculations. Source: Ref 3

forging, profile extrusion, and complex shape rolling.

- Use the UBM as a starting point in a finite-element formulation (FEM), where the Newton-Raphson algorithm is used for numerical integration. Such an approach is used in the popular commercial software DEFORM.
- Use UBM to speed up FEM calculations, especially if the three-dimensional field can be simplified to 2.5 dimensions by assuming a known velocity field in a selected direction. Such an approach is used in CADEX (Computer-Aided Design of Extrusions), a software for die design in shape extrusion (Ref 35).

The UMB method and its incremental formulation (UBET) have been widely applied to simplified industrial problems by metalforming researchers. Selected examples are included.

Example 6: Extrusion of Tubes and Profiles. Upper bound methods have been extensively applied to the extrusion of tubes and profiles, where the metal flow can easily be represented with continuous or piecewise continuous functions. Manibu Kiuchi in Japan was one of the first to develop a systematic UBM-based procedure for modeling three-dimensional extrusion of simple profiles with constant thickness. Kiuchi and Ishikawa (Ref 36) demonstrated its application to the extrusion of a T-profile using a complex functional form for the velocity field with unknown parameters that were determined by power minimization. Selected results from their paper are presented in Fig. 8. Later, Stahlberg and Hou (Ref 37) applied UBET to the extrusion of complex aluminum profiles such as T-, U-, and I-shapes, and Shivpuri and Momin (Ref 38) and Damodaran and Shivpuri (Ref 39) applied UBET to the hot extrusion of high-temperature material. In their model, they were able to include flow division as the metal enters various ports. Later, Chitkara and Aleem (Ref 24,

25) used UBM to model extrusion of bimetallic tubes of copper-aluminum and aluminum-copper material pairs. Their predictions compared well with experiments on tubes of short length. Recently, Ebrahimi et al. (Ref 40) applied UBM to the extrusion of tubes using flat dies. They calibrated their model by comparing it to load-stroke curves from extrusion of commercially pure aluminum tubes. Then, the calibrated model was used to examine material flow and compare predictions to FEM models.

Example 7: Rolling and Ring Rolling. While there are many applications of UBM to plane-strain flat rolling, including slabs, plates, strips, and sheets (Ref 41–43), applications to ring rolling are rare due to the transient nature of the process, especially the continuously evolving state of the ring profile as it is being rolled. In 1994, Hahn and Yang (Ref 44) presented a UBM formulation for ring rolling by dividing the ring profile into multiple sections, choosing different piecewise continuous velocity functions for these sections, and then assuring flow continuity at the interface boundary of these sections. They compared the predictions of roll torque and profile height for different ring configurations with the measurements on a table-top ring rolling setup.

Example 8: Forging. In 1986, Kiuchi developed UBET-based solvers (CAS-UBET: Computer-Aided System-Upper Bound Elemental Technique) to model forging of simple shapes, such as disks with simple flanges (Ref 45). The main objective of this modeling was to determine forging load as a function of die stroke for various geometric parameters (Fig. 9). He extended this approach to more complex three-dimesional shapes, including ones with protruding bosses (Fig. 10). In his approach, die underfilling was minimized by varying processing conditions and geometric parameters. He was able to handle multiple

preform design using this elemental approach. Later, Hou and Stahlberg (Ref 46) applied UBET to forging with grooved dies, Wang et al. (Ref 47) to the forging of a blade, Ranatunga et al. (Ref 48) to the design of flash gap in the forging of an aeroengine disk, Alfozan and Gunasekhera (Ref 49) to a disk with extrusion forging of many ribs, and Hwang et al. (Ref 50) to nonaxisymmetric forging of flanged parts with heads of various shapes, including hex and gear type. Bramley (Ref 51) further extended the UBET approach (Tetrahedral Element Upper Bound Analysis, or TEUBA) to the preform design and flash land determination for forging of simple parts, (Fig. 11).

Example 9: Glass Lubrication in Hot Extrusion. Glass lubrication plays a pivotal role in the hot extrusion of steel, titanium, and superalloy shapes using the Sojournet process. In this case, the glass not only provides for low friction but also determines the die profile by flowing together with the hot metal during extrusion. To model this multimaterial three-dimensional system, Shivpuri and Damodaran (Ref 52) developed a UBM approach that combines one-dimensional heat transfer with the glass and metal rheologies to model the extrusion process. The software CADEX, based on this approach, is used at many extrusion companies in the United States.

Example 10: Hybrid Approaches. As presented earlier, the slab method has been combined with FEM to create a computationally efficient model (Ref 20, 21) that not only predicts metal flow but also microstructural prediction for multipass shape rolling. As with slab methods, UBM has also been combined with FEM by Liu et al. (Ref 53) to carry out preform design of an aeroengine disk by using reverse simulation. In this approach, the design is carried out by UBM, and verification of this design is by FEM simulations. Forgings from the original design and the modified design are presented in this article, but the results are not compared to those predicted.

Summary

While FEM has provided metalforming engineers a formidable tool for predicting metal flow and microstructural prediction, its application is often limited to process parameter determination for processes with known problems. The use of FEM in process design and/or preform design (reverse engineering) is limited by the incremental nature of the solution. Approximate closed-form analytical methods, such as slab and upper bound methods, are computationally efficient and can be used for forward and inverse design of metalforming problems. The inclusion of elemental-incremental algorithms (slab element method or UBET) has widened the application of slab and upper bound approaches to transient problems, and the inclusion of piecewise continuous functions has increased their application to
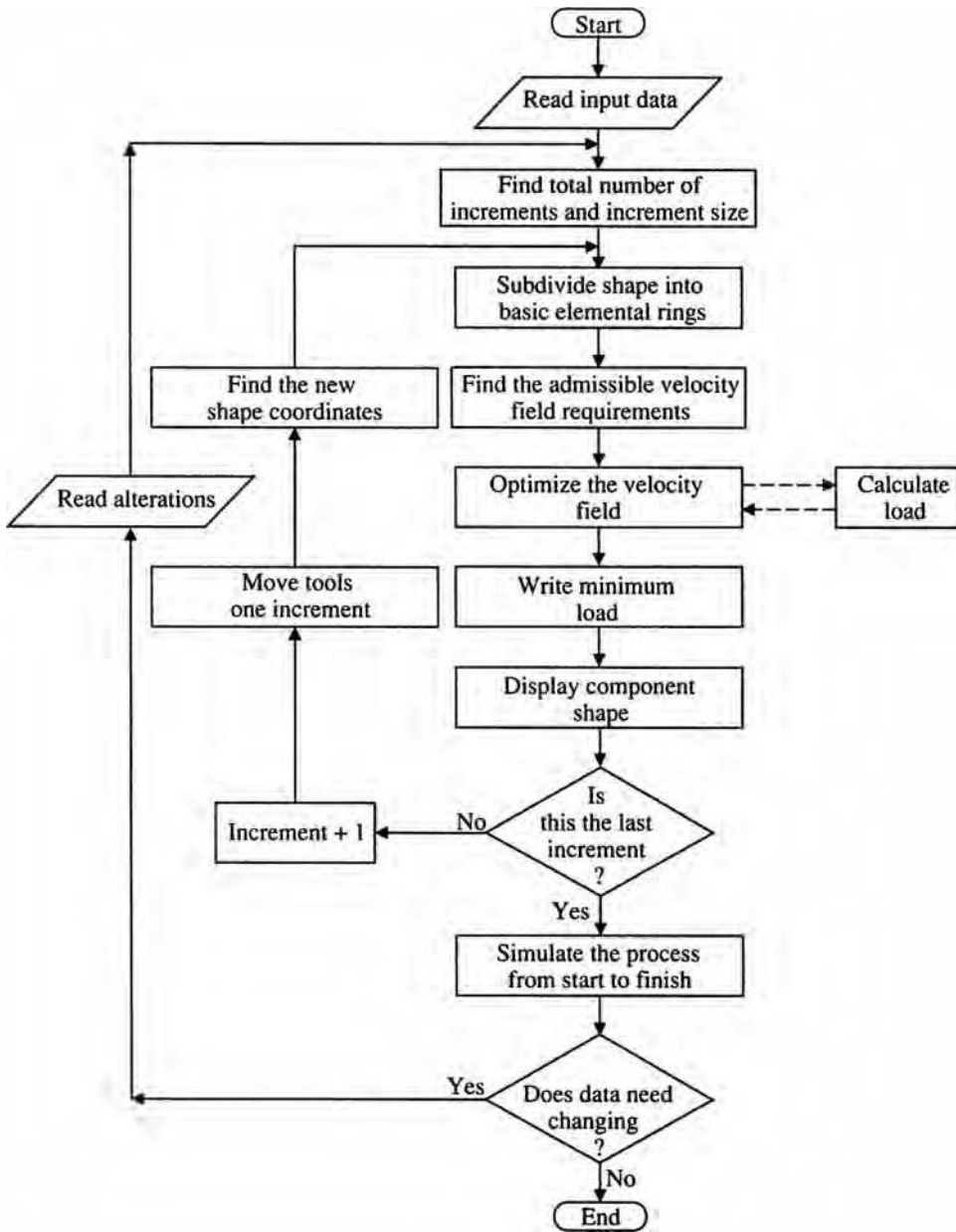
Fig. 7 Computational algorithm for the incremental upper bound method. The discrete elements are chosen to represent workpiece geometry—plane strain, axisymmetric, or three-dimensional (blocks)—and the calculation updated as the die compresses the workpiece incrementally. Source: Ref 34

more complex geometries. A typical SLEM or UBET simulation cycle takes a few minutes for even a complex problem. Recently, these approaches have been combined with FEM for geometries that have 2.5-dimensional characteristics.

REFERENCES

1. R. Hill, *Mathematical Theory of Plasticity,* Oxford University Press, 1950
2. W. Johnson and P.B. Mellor, *Engineering Plasticity,* Van Nostrand Reinhold, 1973
3. W.F. Hosford and R.M. Caddell, *Metalforming: Mechanics and Metallurgy*, 3rd ed., Cambridge University Press, 2007
4. E.G. Thomsen, C.T. Yang, and S. Kobayashi, *Mechanics of Plastic Deformation in Metal Processing,* McMillan, 1965
5. R. Shivpuri and P.C. Chou, A Comparative Study of Slab, Upperbound and Finite Element Methods for Predicting Force and Torque in Cold Rolling, *Int. J. Mach. Tools Manuf.,* Vol 29 (No. 3), 1989, p 305–322
6. H.H. Hsu and G.-Y. Tzou, Two Analytical Models of Double-Layer Clad Sheet Compression Forming Based on the Upper Bound and the Slab Methods, *J. Mater. Process. Technol.,* Vol 140, 2003, p 604–609
7. G. Sachs, On the Theory of the Tension Test, *Z. Angew. Math. Mech.,* Vol 7, 1927, p 235–236 (in German)
8. E. Siebel, Forces and Material Flow in Plastic Deformation, *Stahl Eisen,* Vol 45, 1925, p 1563–1566 (in German)
9. T. Von Karman, Contributions to the Theory of Rolling, *Z. Angew, Math. Mech.,* Vol 5, 1925, p 139–141 (in German)
10. E. Orowan, The Calculation of Roll Pressure in Hot and Cold Flat Rolling, *Proc. Inst. Mech. Eng.,* Vol 150, 1943, p 140–167
11. H.A. González Rojas, J.V. Calvet, and V.I. Bubnovich, A New Analytical Solution for Prediction of Forward Tension in the Drawing Process, *J. Mater. Process. Technol.,* Vol 198, 2008, p 93–98
12. N. Kim, S.M. Lee, W. Shin, and R. Shivpuri, Simulation of Square-to-Oval Single Pass Rolling Using a Computationally Effective Finite and Slab Element Method, *J. Eng. Ind. (Trans. ASME),* Vol 114, Aug 1992, p 329–335
13. J.G. Wistreich, Investigation of the Mechanics of Wire Drawing, *Proc. Inst. Mech. Eng.,* Vol 169, 1955, p 654–665
14. W.F. Hosford and R.M. Caddell, *Metal Forming—Mechanics and Metallurgy,* 2nd ed., Prentice Hall, Englewood Cliffs, N.J., 1993
15. E.M. Rubio, A.M. Camacho, L. Sevilla, and M.A. Sebastian, Calculation of Forward Tension in the Drawing Processes, *J. Mater. Process. Technol.,* Vol 162–163, 2005, p 551–557
16. D.R. Bland and H. Ford, The Calculation of Rolls Force and Torque in Cold Strip Rolling with Tensions, *Proc. Inst. Mech. Eng.,* Vol 159, 1948, p 144–153
17. J.M. Alexander, On the Theory of Rolling, *Proc. R. Soc. Lond. A,* Vol 326, 1972, p 525–563
18. G.W. Rowe, *Principles of Industrial Metalworking Processes,* Edward Arnold, London, 1977
19. D. Kumar and U.S. Dixit, A Slab Method of Strain Hardening and Friction Effects in Cold Foil Rolling Process, *J. Mater. Process. Technol.,* Vol 171, 2006, p 331–340
20. W. Shin, S.M. Lee, R. Shivpuri, and T. Altan, Finite-Slab Element Investigation of Square-to-Round Multipass Shape Rolling, *J. Mater. Process. Technol.,* Vol 33, 1992, p 141–154
21. R. Shivpuri and W. Shin, A Methodology for Roll Pass Optimization for Multi-Pass Shape Rolling, *Int. J. Mach. Tools Manuf.,* Vol 32 (No. 5), 1992, p 671–683
22. M. Kazeminezhad and A.K. Taheri, Applicability of the FSEM for Analyzing Wire Flat Rolling Process, *Eng. Comput.: Int. J. Comput.-Aided Eng. Software,* Vol 23 (No. 5), 2006, p 515–524
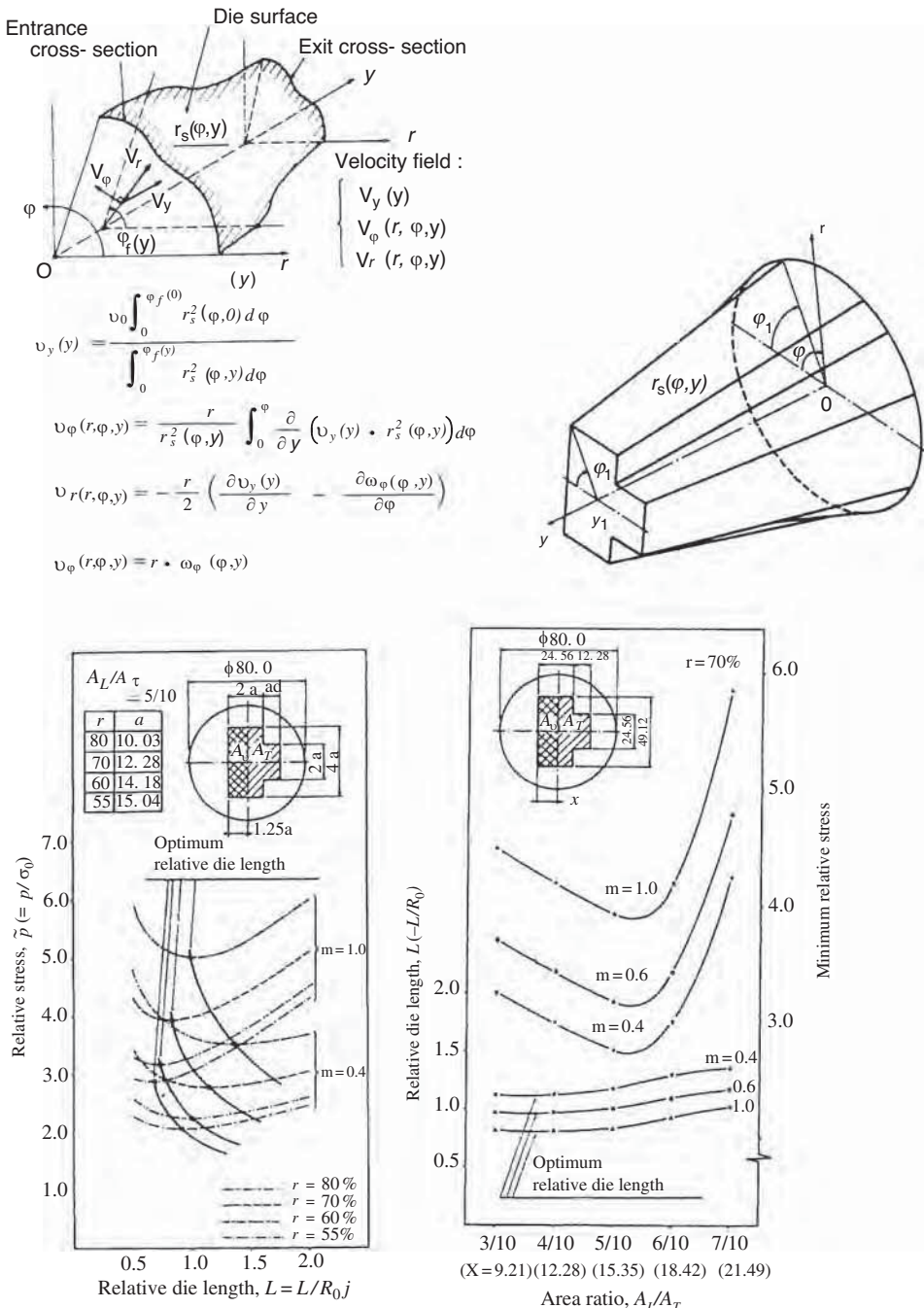
Fig. 8 Velocity fields for the extrusion of a T-shaped profile using the upper bound method, and the predicted results for the relative stress and relative die length between the various segments for the various area ratios of the profile. Source: Ref 36

LIVE GRAPH
Click here to view

23. P. Pauskar and R. Shivpuri, Microstructure and Mechanics Interaction in the Modeling of Hot Rolling of Rods, *Ann. CIRP,* Vol 48 (No. 1), 1999, p 191–194

24. N.R. Chitkara and A. Aleem, Extrusion of Axi-Symmetric Bi-Metallic Tubes from Solid Circular Billets: Application of a Generalized Upper Bound Analysis and Some Experiments, *Int. J. Mech. Sci.,* Vol 43, 2001, p 2833–2856

25. N.R. Chitkara and A. Aleem, Extrusion of Axi-Symmetric Bi-Metallic Tubes: Some Experiments Using Hollow Billets and the Application of a Generalized Slab Method of Analysis, *Int. J. Mech. Sci.,* Vol 43, 2001, p 2857–2882

26. T. Altan, S.-I. Oh, and H. Gegel, *Metal Forming: Principles and Applications,* American Society for Metals, 1983

27. M. O'Connell, B. Painter, G. Maul, and T. Altan, Flashless Closed-Die Upset Forging—Load Estimation for Optimal Cold Header Selection, *J. Mater. Process. Technol.,* Vol 59, 1996, p 81–94

28. F. Fereshteh-Saniee and M. Jaafari, Analytical, Numerical and Experimental Analyses of the Closed Die Forging, *J. Mater. Process. Technol.,* Vol 125–126, 2002, p 334–340

29. A.G. Mamalis, G.L. Petrossian, and D.E. Manolakos, Open-Die Forging of Sintered Cylindrical Billets: An Analytical Approach, *J. Mater. Process. Technol.,* Vol 96, 1999, p 112–116

30. N.R. Chitkara and M.A. Bhutta, Shape Heading of Splines and Solid Spur Gear Forms: ANA Analysis and Some Experiments, *Int. J. Mech. Sci.,* Vol 43, 2001, p 1073–1106

31. B. Avitzur, *Metal Forming: Processes and Analysis,* McGraw-Hill, New York, 1968

32. B. Avitzur, *Metal Forming: Application of Limit Analysis,* Marcel Dekker, New York, 1980

33. B. Avitzur, *Handbook of Metal Forming,* John-Wiley & Sons, New York, 1983

34. F.H. Osman and A.N. Bramley, Metal Flow Prediction in Forging and Extrusion Using UBET, *Proc. 18th Int. Mach. Tool Des. Res. Conf.,* J.M. Alexander, Ed., Sept 1977, p 51–59

35. D. Damodaran and R. Shivpuri, A Simple Numerical Model for Real Time Determination of Temperatures and Pressures during Glass Lubricated Hot Extrusion, *Trans. NAMRI/SME,* Vol XXV, 1997, p 25–30

36. M. Kiuchi and M. Ishikawa, *J. Jpn. Soc. Technol. Plastic. (JSTP),* Vol 24 (No. 270), 1983, p 722

37. U. Stahlberg and J. Hou, UBET-Simulation Meant for Basic Understanding of the Extrusion of Aluminum Profiles, *J. Eng. Ind. (Trans. ASME),* Vol 117, 1995, p 485–493

38. R. Shivpuri and S. Momin, Computer-Aided Design of Dies to Control Dimensional Quality of Extruded Shapes, *Ann. CIRP,* Vol 41, (No. 1), 1992, p 275–278

39. D. Damodaran and R. Shivpuri, "An Upperbound Model for Glass Lubricated Hot Extrusion of Titanium Shapes Using Flat Faced Shear Dies," McNU'97, 1997 Joint ASME/ASCE/SES Summer Conference, June 29–July 2, 1997 (Evanston, IL)

40. R. Ebrahimi, M. Reihanian, M. Kanaani, and M.M. Moshksar, An Upper Bound Analysis of the Tube Extrusion Process, *J. Mater. Process. Technol.,* Vol 199, 2008, p 214–220

41. B. Avitzur, An Upper Bound Approach to Cold Strip Rolling, *Trans. ASME B,* Vol 86, 1964, p 31–48

Rectangular element

$$\dot{W}=C_1\ Y+C_2$$
$$\dot{U}=-12C2\ \ C_1R+C_3/R$$

Triangular element

$$W=C_1Y\ \ R+f(R)$$
$$U=-C_1(1+C_2/R)$$

d=6
D=16
$H_C=H_L=6$
m=1.0

LIVE GRAPH
Click here to view
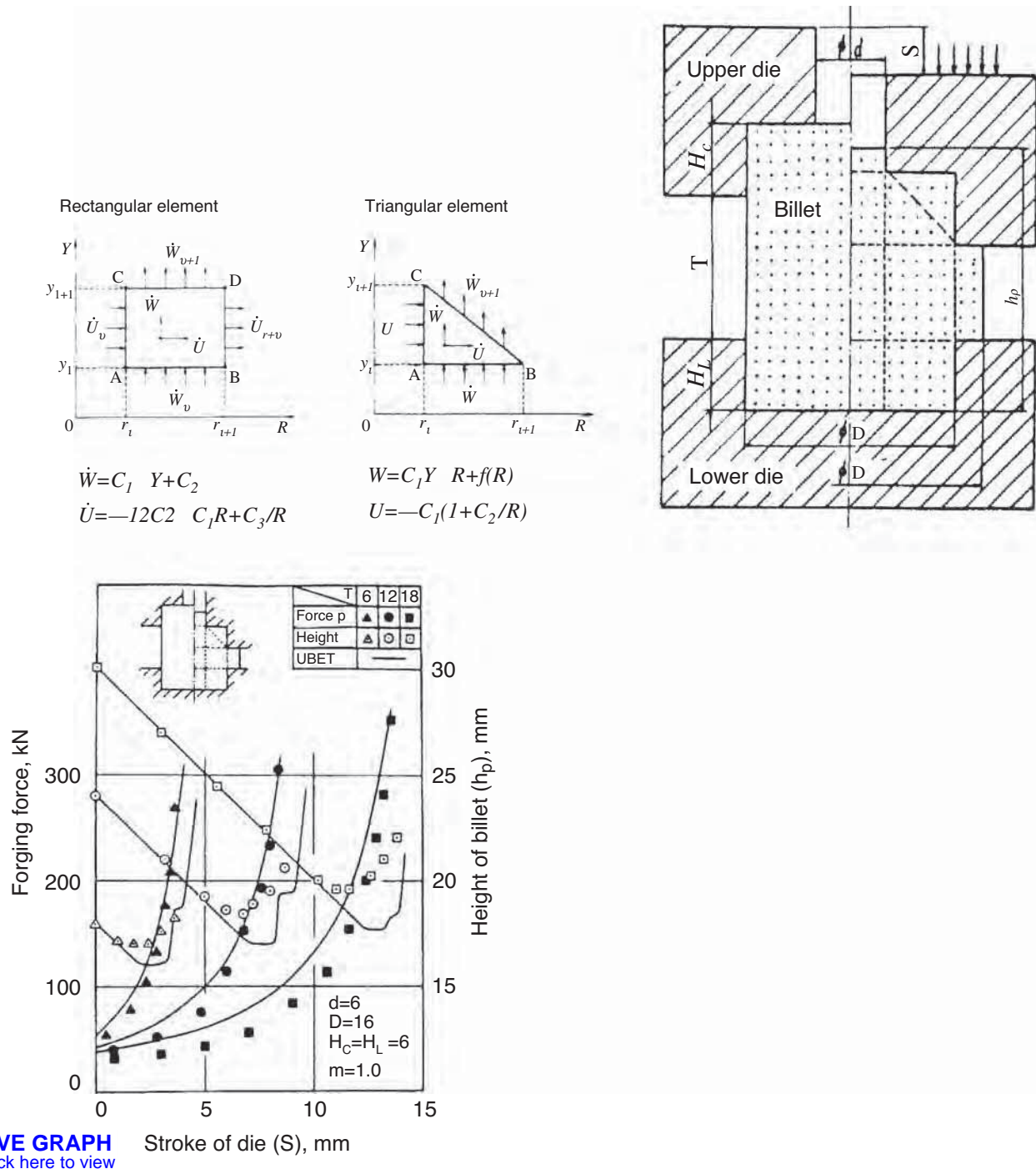
Fig. 9 Elemental description in the Computer-Aided System-Upper Bound Elemental Technique (CAS-UBET) formulation (top left), discretization of axisymmetric forging geometry (top right), and selected results for forging force with various flash ratios (bottom). Source: Ref 45

42. A.N. Dogruoglu, On Constructing Kinematically Admissible Velocity Fields in Cold Sheet Rolling, *J. Mater. Process. Technol.,* Vol 110, 2001, p 287–299

43. K. Komori, An Upper Bound Method for Analysis of Three Dimensional Deformation in the Flat Rolling of Bars, *Int. J. Mech. Sci.,* Vol 40, 1998, p 479–491

44. Y.H. Hahn and D.Y. Yang, UBET Analysis of the Closed-Pass Ring Rolling of Rings Having Arbitrary Shaped Profiles, *J. Mater. Process. Technol.,* Vol 40, 1994, p 451–463

45. M. Kiuchi, Complex Simulation System for Forging Based on UBET, *Ann. CIRP,* Vol 35 (No. 1), 1986, p 147–150

46. J. Hou and U. Stahlberg, A Three-Dimensional UBET Model Applied to the Forging of Square Blocks Between Flat Dies and Grooved Dies of Simple Geometry, *J. Mater. Proc. Technol.,* Vol 62 (No. 1–3), 1996, p 81–89

47. Z. Wang, K. Xue, and Y. Liu, Backward UBET Simulation of the Forging of a

Blade, *J. Mater. Process. Technol.*, Vol 65, 1997, p 18–21

48. V. Ranatunga, J.S. Gunasekhera, W.G. Frazier, and K.-D. Hur, Use of UBET for Design of Flash Gap in Close-Die Forging, *J. Mater. Process. Technol.*, Vol 111, 2001, p 107–112

49. A. Alfozan and J.S. Gunasekhera, An Upperbound Elemental Technique Approach to the Process Design of Axisymmetric Forging by the Forward and Backward Simulation, *J. Mater. Process. Technol.*, Vol 142, 2003, p 619–627

50. B.C. Hwang, S.J. Hong, and W.B. Bae, An UBET Analysis of the Non-Axisymmetric Extrusion/Forging Process, *J. Mater. Process. Technol.*, Vol 111, 2001, p 135–141

51. A.N. Bramley, UBET and TEUBA: Fast Methods for Forging Simulation and Preform Design, *J. Mater. Process. Technol.*, Vol 116, 2001, p 62–66

52. R. Shivpuri and D. Damodaran, Effect of Glass Lubricant Behavior on the Surface Quality of Extrudates in Glass Lubricated Hot Extrusion, *Ann. CIRP*, Vol 46 (No. 1), 1997

53. Q. Liu, W. Shichun, and S. Sheng, Preform Design in Axisymmetric Forging by New FEM-UBET Method, *J. Mater. Process. Technol.*, Vol 74, 1998, p 218–222

Fig. 10 Division of the complex workpiece geometry into simple geometrical elements in the upper bound elemental technique formulation, formulation of a velocity field for these elements, and selected results for relative pressures as a function of die filling and geometrical features. Source: Ref 45
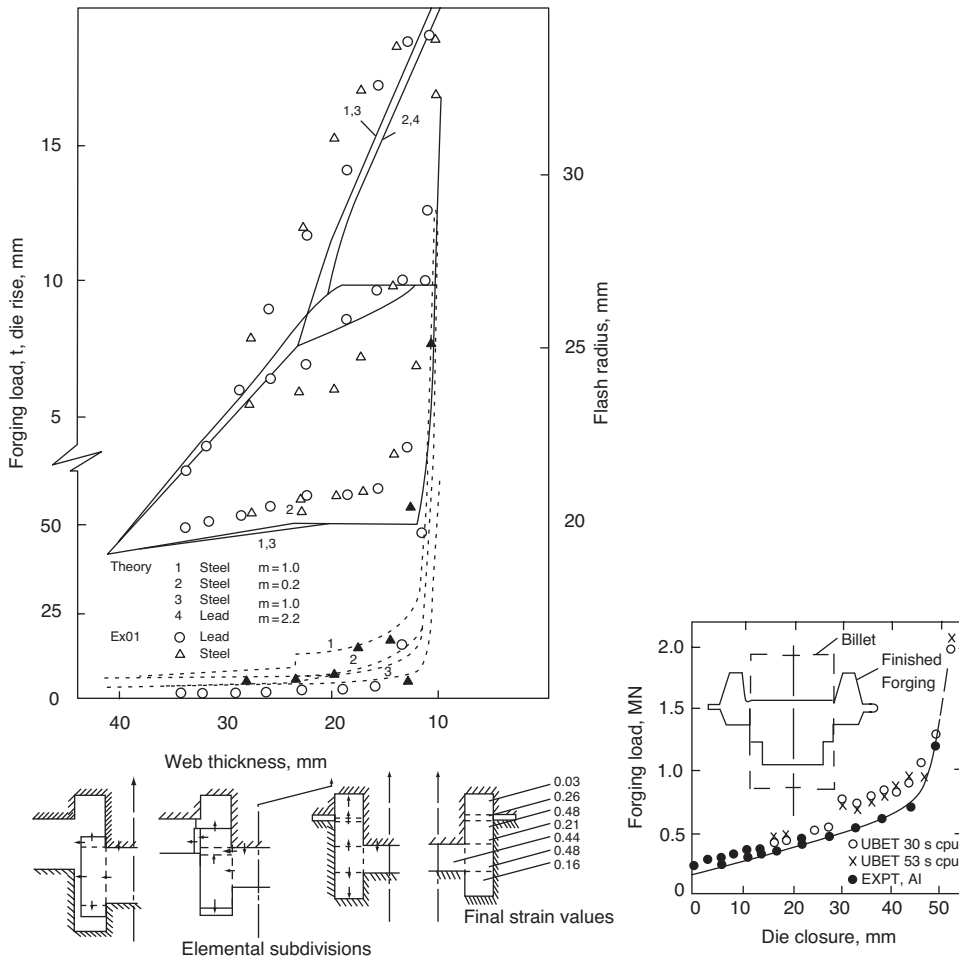
Fig. 11 Experimental and upper bound elemental technique (UBET) results for flow and strain predictions for a rig-web type of axisymmetric forging for various web radii and flash thicknesses (left), and validation of the UBET incremental model with forging load as a function of die closure using the iterative UBET solver (right). Note the relatively small use of computational resources for the UBET simulation. Source: Ref 51

LIVE GRAPH
Click here to view

Modeling with the Finite-Element Method

ENGINEERING MODELS are often used to determine the response of structure, component, process, or system to a set of conditions. The models may be based on closed-form equations, which allow analytical solutions to the response of behavior of a system. In many cases, however, analytical (closed-form) solutions are not possible for complex systems. In this case, numerical methods are needed to provide approximate solutions in modeling the behavior of a system. Numerical methods are used in many types of engineering models, and the widespread availability and computing capability of computers has increased the precision, flexibility, and accuracy of approximate numerical methods.

In general, engineering models are defined in terms of differential equations that quantify relationships in the physical response or behavior of a system. Examples include the differential equations used in continuum mechanics, heat transfer, mass transport, fluid dynamics, and thermodynamics. Differential equations also can be categorized as either partial or ordinary. Partial differential equations are required when the behavior is a function of time and space or of more than one space variable. However, if the problem can be simplified to one independent variable (time or one space variable), ordinary differential equations can be used in understanding the effect of certain parameters on the response or behavior of the process or object.

There are several techniques for solving differential equations. Numerical algorithms to solve differential equations consist of *lumped-parameter methods* or the so-called *meshed-solution methods* (such as finite-element analysis). If the problem can be simplified for the use of ordinary differential equations, then a lumped-parameter model may be used. A lumped-parameter model may help in understanding the effect of certain parameters on the process. However, these methods do not model spatial variation directly, and the parameters do not necessarily always have direct physical significance.

Typically, partial differential equations are required to describe the process in terms of time, space, field variables, and internal states. Several methods have been developed for the numerical solution of partial differential equations. This typically involves meshed-solution methods such as the finite-element method (FEM), the finite-difference method (FDM), and the boundary-element method (BEM). Each has different suitability for different types of

problems. For example, the FDM is often adopted in fluid mechanics but is seldom used in solid and structural mechanics. Finite-element methods are the most common for linear and nonlinear continuum mechanics, although the BEM has advantages in some applications of continuum mechanics.

This article briefly introduces the methods of so-called meshed solutions, with an emphasis on the FEM. As background on the general concepts of numerical methods in modeling engineering systems, an introduction is given of some basic differential equations that are used to model the responses of structures, components, processes, or systems. An emphasis is placed on continuum mechanics as it relates to metal deformation or forming. The mathematical principles of solving differentials then are outlined (only in broad terms) as a way to illustrate why FEMs play such an important role in the modeling of systems and processes. Again, an emphasis is placed on the processes of metal deformation, although some linear (elastic) structural problems are briefly reviewed to help illustrate the concept of the FEMs. The selected references also provide more details on the principles and applications of finite-element modeling.

Differential Equations

As noted, engineering models are often based on differential equations that describe the physical behavior of a system. Partial differential equations are required when the behavior is a function of more than one variable. When the system response can be simplified to one independent variable, then ordinary differential equations can be used. The basic types of partial differential equations are described in this section, with a focus on continuum mechanics as an example. Modeling of other phenomena is described in more detail in other articles in this Volume.

Partial differential equations can be divided into three categories: hyperbolic, parabolic, and elliptic. Standard hyperbolic equations include the *wave equation*:

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}$$

where c is the wave speed. Parabolic partial differential equations include the *diffusion equation*:

$$D\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}$$

where D is the diffusivity.

Elliptic equations are usually used to model steady-state phenomena. When hyperbolic or parabolic equations are assumed to be invariant with time, then they reduce to elliptic equations. For example, when the time dependence is removed from the wave equation or the diffusion equation, they reduce to the Laplace or Poisson equations for steady-state heat conduction in solids with constant properties:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0 \qquad \text{(Laplace equation)}$$

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} + \frac{g(x, y, z)}{k} = 0$$
$$\text{(Poisson equation)}$$

where the Possion equation includes a heat-source function, $g(x, y, z)$, with k as the thermal conductivity of the material.

Continuum Mechanics Equations. Continuum mechanics models are based on steady-state equations of mechanical equilibrium and constitutive equations for mechanical flow. The complete set of equations serves as the foundation for continuum mechanics models of bulk working processes. A model for each case is developed by imposing appropriate boundary conditions and initial conditions (tool and workpiece geometry, temperature, heat flow, etc.) on the solutions for the set of equations. In addition, models of thermo-mechanical processes may also require description of thermophysical behavior and the contact interface between the tool and worked material.

Equilibrium and Compatibility (Strain-Displacement) Equations. Steady-state equations include the forces acting on an element and the compatibility of strain displacements. The steady-state equations in describing the various forces acting on an element in mechanical equilibrium are:

$$\frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} + \frac{\partial \tau_{xy}}{\partial z} = -F_x$$
$$\frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \sigma_y}{\partial y} + \frac{\partial \tau_{yz}}{\partial z} = -F_y$$
$$\frac{\partial \tau_{xz}}{\partial x} + \frac{\partial \tau_{yz}}{\partial y} + \frac{\partial \sigma_z}{\partial z} = -F_z \qquad \text{(Eq 1)}$$

where σ is the normal stress component, τ is the shear stress component, and F is the body

force/unit volume component. Similarly, the strain-displacement relationships are given as:

$$\varepsilon_x = \frac{\partial u}{\partial x} \quad \gamma_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}$$
$$\varepsilon_y = \frac{\partial v}{\partial y} \quad \gamma_{yz} = \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}$$
$$\varepsilon_z = \frac{\partial w}{\partial z} \quad \gamma_{zx} = \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \qquad \text{(Eq 2)}$$

where ε is the normal strain, γ is the shear strain, and u, v, and w are the displacements in the x, y, and z directions, respectively. In addition to boundary conditions, the solution of these equations may be complicated further by time dependence of the force functions, the nonlinear stress-strain behavior of plastic deformation (i.e., constitutive equations), and the flow rules of plastic deformation (such as the Von Mises yield criterion) under combined stresses.

Constitutive Equations. For a given material, the relations between stress components in Eq 1 and strain components in Eq 2 are given by the constitutive equations representing the behavior of that material. The simplest example of a constitutive equation is the well-known Hooke's law in the elastic regime:

$$\sigma = \varepsilon E \qquad \text{(Eq 3)}$$

where E is the elastic modulus of the material, which is measured in a simple tension test or by ultrasonic means.

During plastic deformation of most metallic materials, the stress-strain curve becomes nonlinear, because hardening (or less frequently softening) of the material can occur when continuing plastic strain is built up. The strain rate can also influence the hardening or softening of a material. The general form of the constitutive equation for deformation processing is:

$$\bar{\sigma} = f(\bar{\varepsilon}, \dot{\bar{\varepsilon}}, T) \qquad \text{(Eq 4)}$$

where $\bar{\sigma}$ is the equivalent (or effective) combined stress, $\bar{\varepsilon}$ is the equivalent true strain, $\dot{\bar{\varepsilon}}$ is the equivalent true strain rate, and T is the processing temperature. Most software packages for bulk forming modeling have options to input the testing data in a tabular form or as a constitutive equation. The tabular form is easy to use but is not based on metallurgical principles, as with some constitutive equations.

The most frequently used constitutive equation is:

$$\bar{\sigma} = K\bar{\varepsilon}^n, \dot{\bar{\varepsilon}}^n + Y \qquad \text{(Eq 5)}$$

where n is the strain-hardening exponent, m is the strain-rate sensitivity, and Y and K are coefficients. Strain-rate sensitivity is important at elevated temperatures, while it has little influence at room temperature for most metallic materials. In contrast, the importance of the strain-hardening exponent becomes more significant with decreasing temperature.

Equation 5 does not reflect the influence of temperature. For each temperature, there is a

set of equations. A more fundamentally sound equation has been proposed by Sellars and Tegart (Ref 1) by assuming materials flow during deformation as a thermally activated process:

$$\dot{\bar{\varepsilon}} = A[\sinh(\alpha\bar{\sigma})]^{n'} \exp\left(-\frac{Q}{RT}\right) \qquad \text{(Eq 6)}$$

where A, α, and n' are constants determined by fitting empirical data, and Q is the apparent activation energy. At low stresses ($\alpha\bar{\sigma} < 0.8$), the equation reduces to a power law:

$$\dot{\bar{\varepsilon}} = A_1\bar{\sigma}^{n'} \exp\left(-\frac{Q}{RT}\right) \qquad \text{(Eq 7)}$$

At high stresses ($\alpha\bar{\sigma} > 1.2$),, the equation reduces to an exponential form:

$$\dot{\bar{\varepsilon}} = A_2 \exp(\beta\bar{\sigma}) \exp\left(-\frac{Q}{RT}\right) \qquad \text{(Eq 8)}$$

where $\beta = \alpha n'$.

Other constitutive relations have been proposed to describe dynamic recovery and dynamic recrystallization, such as the Laasoul-Jonas model (Ref 2) and the internal variable model (Ref 3). In all cases, constitutive equations are empirical-based relations derived from the reduction of test data.

For more details on constitutive equations, see the article "Constitutive Equations" in *Metalworking: Bulk Forming,* Volume 14A, *ASM Handbook,* 2005, page 563.

Yield Criteria (Flow Rules). Continuum mechanics also include rules or models for when a material yields or flows. In structural analysis, yield criterion may be characterized as a "failure theory," because plastic deformation is an undesired outcome in structural design. In bulk deformation, yielding of the workpiece is intended, and yield criteria are used in the modeling of flow under combined stresses.

The continuum mechanics of metallic materials includes several theories for yielding, as described in a historical sketch of continuum plasticity theory with an introduction on computational methods in solid mechanics (Ref 4). The first yield criterion for metals was proposed by Henri Tresca in the 1860s. The Tresca criterion is based on the premise that yielding is dependent on just shear stresses, whereby plastic flow begins when the shear stresses exceed the shear yield strength of the metallic material. Although the Tresca yield criterion is adequate, it neglects the intermediate principal stress, σ_2.

The Levy-von Mises yield criterion is considered to be a more complete and generally applicable yielding criteria. It is based on the theory of Richard von Mises that incorporated a proposal by M. Levy, which stated that the tensor components of plastic-strain increments are in proportion to each other just as are the tensor components for deviatoric stress. It is based on the second tensor invariant of the deviatoric stresses (that is, of the total stresses

minus those of a hydrostatic state with pressure equal to the average normal stress over all planes) (Ref 4). The von Mises yield criterion is thus expressed as:

$$2\sigma_0^2 = (\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2 \quad \text{(Eq 9)}$$

where σ_0 is the uniaxial yield (flow) stress measured in tension or compression, and σ_1, σ_2, σ_3 are the three principal stresses.

Fracture Criteria. Continuum mechanics can also be used in conjunction with multiaxial fracture criteria to gain insights and solutions on the prevention of flow-related cracks. Various fracture (or microstructural damage) criteria have been developed to evaluate workability (e.g., Ref 5–9), but not all are easily incorporated into continuum mechanics models of processes. One criterion that gives a very accurate description of workability and is easily implemented in the models is that due to Crockcroft and Latham (Ref 5):

$$\int_0^{\bar{\varepsilon}_f} \sigma_1 d\bar{\varepsilon} \geq C \qquad \text{(Eq 10)}$$

where σ_1 is the maximum principal stress, $\bar{\varepsilon}_f$ is the equivalent strain at fracture, and C is a constant representing the workability of the material. If $\sigma_1 < 0$, then there are only compressive stresses, and no fracture occurs. The Crockcroft-Latham criterion is phenomenological based rather than mechanistic, but it captures the physical concept and intuitive understanding that damage accumulation, or workability, is dependent both on the degree of plastic deformation (represented by the integral over effective strain) and tensile stress (represented by the maximum principal stress). Furthermore, both of these physical quantities are readily available outputs of continuum mechanics models and can be evaluated at every localized region or point throughout the material to determine potential sites of crack initiation. As with any constitutive relation, the value of incorporating this workability criterion into process analyses depends on accurate representation of the actual material behavior, represented by the coefficient C in this case.

A more recent fracture criterion is theorized as (Ref 8):

$$\sigma_{1f} \cdot \frac{\bar{\varepsilon}}{\bar{\sigma}} \geq C' \qquad \text{(Eq 11)}$$

where σ_{If} is the maximum principal stress at fracture, and C' is a workability constant. This criterion can be related directly to bulk workability tests. For incremental loading, Bandstra (Ref 9) proved that the criterion can be expressed as:

$$\int_0^{\bar{\varepsilon}_f} \frac{\sigma_1}{\bar{\sigma}} d\bar{\varepsilon} \geq C' \qquad \text{(Eq 12)}$$

A review of fracture criteria can be found in Ref 8, and some examples are given in the article "Models for Fracture during Deformation Processing" in this Volume.

Methods of Solution

As noted previously, there are several techniques for solving differential equations. Numerical algorithms to solve differential equations consist of lumped-parameter methods or the so-called meshed-solution methods (such as finite-element analysis). If the problem can be simplified for the use of ordinary differential equations, then a lumped-parameter model may be used. A lumped-parameter model may help in understanding the effect of certain parameters on the process. However, these methods do not model spatial variation directly, and the parameters do not necessarily always have direct physical significance.

Numerical methods for solving the governing equations of a model are required when the simplifying assumptions of methods described in the preceding are not justified. The general types of numerical techniques used in the solution of partial differential equations include:

- The finite-element method (FEM)
- The boundary-element method (BEM)
- The finite-difference method (FDM)

With the continued improvement in computer capabilities, these numerical techniques have become very effective in the engineering analysis of static problems, dynamic conditions (where the calculation of inertial and/or damping forces involves derivatives with respect to time), or quasi-static conditions where rate-dependent plasticity may require a realistic estimation of time (but where inertial forces may still be neglected).

In applications involving continuum mechanics, the FEM is the most common technique, although the BEM has advantages in some applications of continuum mechanics. The FDM has proven to be useful in fluid and thermal problems but is seldom used in solid and structural mechanics. Nonetheless, the FDM is a simple and efficient method for solving ordinary differential equations in problem regions with simple boundaries. For each node of the mesh, the unknown function values are found, replacing the relevant differential equation, (i.e., $dy = f[x,y]dx$) by a difference equation:

$$\Delta y = f(x + \Delta x/2, y + \Delta y/2)\Delta x \qquad \text{(Eq 13)}$$

where Δx Δy are steps in an interative procedure. Hyperbolic and parabolic partial differential equations are often solved using a hybrid of the FEM and FDM. The spatial variables are modeled using the FEM, and their variation with time is modeled by the FDM.

Other numerical techniques include:

- Finite-volume methods, which are important in highly nonlinear problems of fluid mechanics
- Spectral methods, which are based on transforms that map space and/or time dimensions to spaces where the problem is easier to solve
- Mesh-free methods, which are a recent development of FDMs with arbitrary grids

These techniques and the FDM are only mentioned for reference without further discussion. The FEM and BEM are more common in the modeling of bulk deformation processes, as described later in this article in more detail. Application of these computer models has become an essential tool in meeting product requirements of dimensional tolerances, surface finish, and consistency of properties.

Meshed (discrete-element) numerical methods include various numerical methods, such as FEM, BEM, and FDM. These numerical methods provide approximate solutions by converting a complex continuum model into a discrete set of smaller problems with a finite number of degrees of freedom. The meshes are created by using structured elements such as rectilinear blocks or unstructured meshes with variable-shaped elements (e.g., tetrahedra, bricks, hexahedral, prisms, and so forth) for better fidelity to the macroscopic conditions or boundaries.

The choice of a method depends on the different types of equations and boundary conditions. Within these methods, one can use a structured or an unstructured mesh. Structured meshes are created by using rectilinear, brick-like elements. It is easy to use this type of mesh; however, fine geometry details may be missed. Unstructured meshes can be of any shape—tetrahedra, bricks, hexahedral, prisms, and so forth. Many of the disadvantages of using a structured mesh are eliminated through this type of a mesh.

Once a discrete-element model has been created, mathematical techniques are used to obtain a set of equilibrium equations for each element and the entire model. By applying various boundary conditions and loads to the model, the solution of the simultaneous set of equations provides the resulting responses anywhere in the model while still providing continuity and equilibrium. The process of solution is essentially a computer-based numerical method, where interpolation functions (polynomials) are used to reduce the behavior at an infinite field of points to a finite number of points.

Meshed-solution models have advantages over the typical closed-form solutions, because they more readily give solutions to irregular shapes, variable material properties, and irregular boundary conditions. Construction of a discrete meshed model for numerical solution may be necessary if the modeled volume:

- Has a complex shape (as is common in bulk forming)
- Contains different phases and grains
- Contains discontinuous behavior such as a phase change
- Has a nonlinear physical process such as when the heat-transfer coefficient is a nonlinear function of the temperature

In many instances, meshed models are supplemented by some nonmeshed symbolic or analytical modelling. This is done in order to decide on appropriate boundary conditions for the meshed part of the problem, because it is the boundary conditions that effectively model the physical problem and control the form of the final solution. Analytic models are always useful for distinguishing between mechanisms that have to be modeled separately or modeled as a coupled set.

Boundary-Element Method

The BEM is a technique for representing a complex structure or component as a computer model in order to determine its response to a set of given conditions. Like the FEM, the model is formed by subdividing the structure into small elements to form the overall model. However, unlike the FEM, only the surface (or boundary) of the problem requires subdivision, thereby reducing the dimensionality of the problem and thus dramatically reducing the computational effort in obtaining a solution.

The BEM has a more restricted range of application than FEM has. In general it is applicable mainly to linear elliptic partial differential equations. It also requires reformulation of the governing partial differential equations into a Fredholm integral equation, which applies to a range of physical problems. For example, elliptic partial differential equations such as the Laplace or Helmholtz equations can be reformulated as Fredholm integral equations and then solved by the BEM. The advantage is that the mesh need only cover the boundaries of the domain.

The BEM can be an effective tool in the analysis of various metal-forming problems in rolling and extrusion. For example, Ref 10 demonstrates that the BEM can be used to efficiently and accurately analyze planar and axisymmetric forming problems involving both material and geometric nonlinearities, along with complicated interface conditions. Like FEM, the use of BEM in the modeling of metal-forming operations may require consideration of elastic-plastic behavior and elastic-viscoplastic problems involving large strains. Elastic strains (plastic or viscoplastic) are presumed to be large. When strains become large, the original mesh may become so distorted that the interpolation polynomials are incapable of modeling the geometry of the elements and their relevant state variables. This requires a process of remeshing.

Finite-Element Methods

Finite-element analysis (FEA) is a computer-based analytical technique that allows numerical solutions to be obtained for complex mathematical and engineering problems by creating a discrete or finite number of individual nodes

and elements. Discrete elements fill the appropriate geometry, and the method enables the systematic solution of equilibrium equations for each element and the entire model with as much fidelity to geometry as needed.

Finite-element analysis is a powerful analysis tool that is flexible for solving problems with irregular shapes, variable material properties, and irregular boundary conditions. With advancements in computer technology, the use of numerical methods such as FEA has grown. The method was originally developed for structural problems (stress displacement of complex geometries), but the same concepts and principles apply to many kinds of engineering problems. In deformation processes, FEA is a useful tool in die design and process analysis. Common problems solved by FEA include insufficient die filling, poor shape control, poor flow of material, cracks and voids that lead to fracture, and inadequate properties from microstructural variations (grain size).

The application of FEM in metal forming (e.g., Ref 11–13) has brought great changes to design methodologies that were formerly based on trial-and-error approaches. For example, shape changes during forging are easily predicted by FEM. The detailed temperature, strain, and strain-rate histories at each individual material point in a workplece are also available from FEM simulations. Simulation of thermomechanical processes can be based on models of continuum mechanics, transport phenomena (heat flow), and metallurgical phenomena (e.g., grain growth and recrystallization).

The range of FEA applications in the area of materials processing is extremely wide, and a brief review of the finite-element techniques applied to metal forming, nonmetal forming, and powder metallurgy are briefly discussed in Ref 14 with an encyclopedic view of the different possibilities in these various fields of application. Many texts (e.g., Ref 11–13) also describe finite-element theory and how it is used in forming analyses. Models may be based on continuum mechanics in the evaluation of flow and/or fracture problems supplemented by more sophisticated thermomechanical/thermophysical models for the simulation of microstructural evolution.

Model Development and Preprocessing

The process of modeling is done in different cycles. Figure 1 outlines some of the basic stages, with three loops suggesting three levels of activities in any modeling effort. The outer loop is managed by someone close to the process who understands the business context of the problem and can concentrate on specifying the objective and providing the raw data. The innermost loop (shaded) requires mostly computational skills, while the middle loop (unshaded) consists of activities balancing the other two. It may very well happen that all three of some combination of the activities can be done by the same person. However, very seldom is that the case. This highlights the need for forming modeling teams where all aspects of the problem can be addressed rigorously. It also emphasizes the importance of training and appropriate software tool development so that the input and output of the tools can be easily understood by all involved in the process.

Figure 2 shows the FEA process. *Finite-element preprocessing* defines the physical problem and converts it into a form that the computer can solve. The *preprocessing* stage consists of:

- Geometry definition and manipulation
- Specification of material properties
- Generation of the finite-element mesh
- Definition of generalized loads and displacements both at the boundary and distributed through the body (inertia and gravitational loads)

The material properties may be specified as constant or variable. When variable, they are often input as tables of properties or as mathematical functions, such as exponentials. Finite-element preprocessing may consume as much as 80% of the calendar time in the finite-element process.

Defining the Problem

Before a full-fledged FEA is undertaken, several questions must be answered:

- Is FEA appropriate?
- What is desired from the FEA?
- When are results needed?
- What are the product and process limitations?

If it is determined that FEA would provide adequate answers, information is gathered to start the modeling process. The required information includes the geometry of interest, initial and boundary conditions, material properties and material behavior models, and an approximate solution to ensure that the finite-element results are not physically absurd.

Next comes defining the physics of the process. This involves using the appropriate set of mathematical equations and the corresponding initial conditions and boundary conditions. Subsequently, the solution method must be determined. This involves choosing the appropriate algorithms to solve the numerical approximations of the mathematical equations.

The development of any model requires the definition of appropriate boundary conditions and the applicable governing equations of the major phenomena occurring in the process (e.g., convection, radiation, chemical reaction, diffusion, deformation). The complexity of the boundary conditions and the governing differential equations determine the possible methods of solution.

Boundary Conditions. Application of appropriate boundary conditions is a major part of the activity of process modeling. Boundary
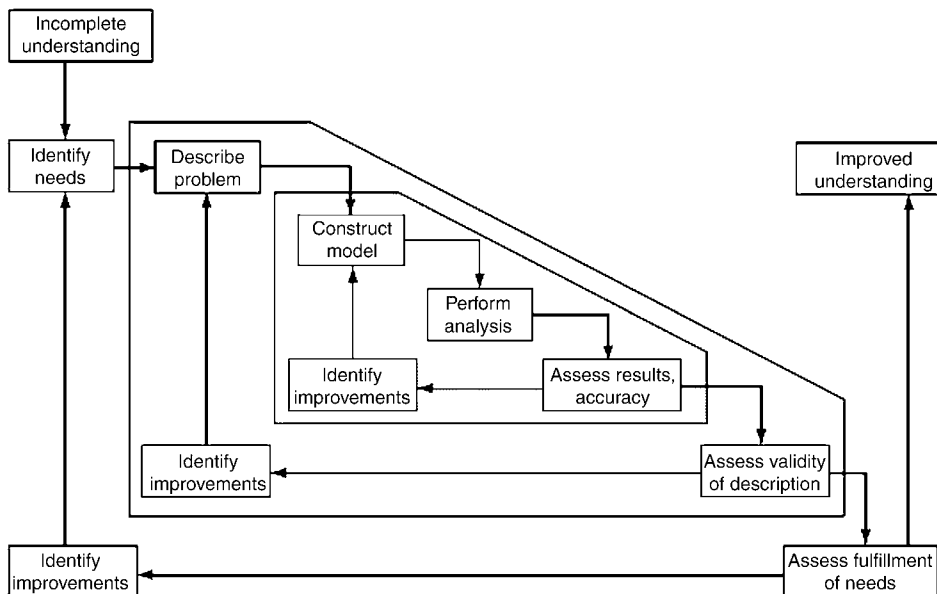
Fig. 1 Modeling cycles. Source: Ref 15
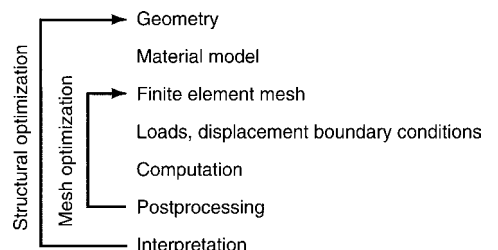
Fig. 2 Typical finite-element process

and initial conditions represent geometric factors (e.g., symmetry, tool shape, and workpiece shape), thermal factors (e.g., heat flux and temperature) and loads (e.g., tool pressure and tool friction) pertinent to the particular problem being modeled. These conditions or constraints influence the complexity of the problem and the most appropriate method or algorithm for solving the equations of the model.

For example, consider the simple example of a uniformly loaded cantilever beam (Fig. 3a). In this case (assuming uniform loading and a rigid connection), then closed-formed equations can provide analytical solutions for the deflection and bending stress over the length of the beam (Fig. 3a). However, if the boundary conditions are altered by the addition of simple supports (Fig. 3b), then the system becomes statically indeterminate (i.e., there is no longer a closed-formed solution that specifies deflection over the length of the beam). In this case, numerical techniques (such as FEA) are required to find approximate solutions of deflection and bending stresses. Likewise, the solution of dynamic problems may also require numerical techniques, depending on the complexity of the equations in the models and the appropriate boundary conditions.

In analytical solutions models, boundary conditions must be set at a very early stage. In meshed-solution models, boundary conditions are typically represented separate from the main equations and decoupled to some extent from the model itself. Therefore, sensitivity analysis can be done much easier using meshed methods.

Material Properties. All models of bulk-working processes require input of accurate material properties so that the fundamental materials behavior can be represented faithfully by the constitutive equations. Acquiring these properties can be difficult and expensive. Sensitivity analysis of the model with respect to variations in property data should be done. In many instances, it may be possible to use models with inexact material property information in order to predict trends, as opposed to determining actual values. Problems may arise if the material properties are improperly extrapolated beyond their range of validity. In deformation modeling, Young's modulus, Poisson's ratio, anisotropic behavior, and flow stress (as functions of stress, strain, strain rate, and temperature) are needed.

The requirements for particular data and the way in which it is gathered is an important step in the construction of a model. Researchers typically play down this step as an industrial implementation detail. On the other hand, industrial practitioners place a greater emphasis on data gathering, because they know the difficulties and time involved in gathering data on production-scale equipment.

Symmetry. In the development process of FEMs, the exploitation of symmetry can greatly reduce the computer resources required to solve a problem, and can sometimes move problem solution from infeasibility to feasibility. Symmetry applies in the context of both geometry and the loads (including the boundary reactions).

Symmetry arises when a geometric element is translated, rotated, mirrored, or swept. At least four major types of symmetry arise in finite-element problems: bilateral symmetry, rotational symmetry, axisymmetry, and translational symmetry.

Bilateral symmetry is synonymous with reflective or mirror symmetry. This is perhaps the simplest, most intuitive type of symmetry. It does carry with it the idea of *handedness,* that is, right hand versus left hand. If the fundamental region being reflected is itself symmetric, then right hand versus left hand need not be considered. If the small asymmetry is structurally insignificant, it is sometimes adequate to model the weaker half of the object. An example might be a connecting rod with a small oil-feed hole on one side.

Rotational symmetry exists when a fundamental part of the structure spans $360°/n$, where n is an integer. Examples of rotationally symmetric objects include things such as a stool, a marine propeller, and a fan blade. *Dihedral symmetry* is a special case of rotational symmetry. It occurs when the fundamental region to be rotated is itself bilaterally symmetric.

Axisymmetry is another special case of rotational symmetry. Every cross section through the axis of rotation of an axisymmetric object is identical. Axisymmetry can reduce a three-dimensional problem to a two-dimensional one.

Translational symmetry occurs when a fundamental unit is translated one or more times. A prismatic structure is a special case of translational symmetry. Often, structures comprise a fundamental element that is both translated and rotated. Superelement techniques are sometimes useful for such structures.

Antisymmetry refers to symmetry accompanied by a change in some property. For example, a photographic print is antisymmetrical to its negative. In the context of FEA, antisymmetry pertains to a problem whose geometry, loads, and displacement boundary conditions exhibit symmetry but the sense of whose loads are reversed.

For linear elastic structures, it is valid to superpose or add the results of multiple solutions. This permits decomposing of loads that do not have symmetry into their symmetric and antisymmetric components and superposing these component loads. This allows modeling a structure with geometrical symmetry as a reduced model, even when the loads are not symmetric (Fig. 4).

Mesh Generation

Finally, the problem must be discretized (i.e., the continuum must be broken into many smaller pieces or elements, as in Fig. 5. In the early development of FEA, when computing power was a factor limiting the generic use of FEA, models were simplified as much as possible through the use of two-dimensional (2-D) analyses, symmetry, linearity, and steady-state conditions. As computing power increased, so did the complexity of finite-element modeling and analysis techniques. The first analyses and elements available were limited to 2-D beams

Beam geometry
I = Moment of inertia
t = thickness
L = length
W_0 = Uniform load
E = Elastic modulus

$$\text{Deflection} = \frac{W_0 x^2}{24EI}(6L^2 - 4Lx + x^2)$$

$$\text{Bending stress} = \frac{W_0 x^2}{(2)}\frac{t}{(2)I}$$

(a)

(b)

Fig. 3 Effect of boundary conditions on the solution of a cantilever problem. (a) The beam deflection and bending stresses for a uniformly loaded cantilever can be solved by a closed-form equation as shown. (b) A supported cantilever beam is statically indeterminate, and numerical methods are required to approximate deflection and bending-stress conditions that are consistent with the boundary conditions established by the additional supports

Fig. 4 (a) Superposition of symmetric (b) and antisymmetric (c) loads. In symmetric loads (b), the abbreviation "SYMM" denotes that $u_x = 0$ everywhere on the cross section for continuum elements (e.g., hexagonal and tetragonal elements), or that $u_x = R_y = R_z = 0$ for line elements (e.g., beams). In antisymmetric loads (c), the abbreviation "ANTISYMM" denotes boundary conditions that are the complement of the SYMM set, e.g., $u_y = u_z = 0$ everywhere on the cross section for continuum elements.

and spars. The progression was then toward three-dimensional (3-D) spars and beams, 2-D solid and axisymmetric elements, and the full 3-D elements. Along with the progression in element capability was the advancement toward nonlinear and transient analyses. Contact elements, elastic/plastic properties, large deflection, and viscoplastic and dynamic impact analyses are now readily available.

In this stage of model development, the shape, size, and complexity of the elements must be defined and refined to provide a suitable basis for numerical approximation of the system. The factors that impact the selection of the elements are the type of the problem, geometry, accuracy desired, availability within the algorithm, nature of the physical problem, and user familiarity. Typical elements are either linear or quadratic and can be one-, two- or three-dimensional. Figure 6 schematically shows some typically used elements.

The size and number of elements are primarily determined by the various gradients (temperature, stress, etc.) in the system. For example, if the gradients are steep, a larger number of smaller-sized elements should be used. The mesh can be made coarser (fewer elements) or finer (more elements), depending on the needs of the problem. Coarser meshes use minimal computer resources in terms of storage and run time. However, because the representation is approximate, the results can be crude. Finer meshes provide a more accurate representation with improved results.

When an FEA mesh is generated, it may be obtained by acting upon a defined geometry. In this case, the mesh is sometimes referred to as top-down mesh. When the mesh itself is the only definition of the geometry, it may be referred to as bottom-up.

Top-down meshing is accomplished by one of two methods: mapped meshing or free meshing. A mapped mesh requires that the geometry be subdivided into canonical shapes, such as quadrilaterals in two dimensions and cuboids in three dimensions. A mesh is then mapped into these regions. This process generally provides for more user control than free meshing, but it is quite time-consuming.

Free meshing, by contrast, relies on one of several algorithms to fill any arbitrary geometry with elements. In two dimensions, these elements are usually triangles or quadrilaterals. In three dimensions, robust algorithms exist to fill an arbitrary space with tetrahedra. However, robust algorithms to fill an arbitrary space with hexahedra are lacking. One of the issues in finite elements is the relative performance of these two element types. Because the shape functions of these elements are analogous to those of the triangle and quadrilateral in two dimensions, performance of the low-order hexahedron may be superior over that of the low-order tetrahedron. In practice, high-order tetrahedra are often used to increase accuracy at the expense of increased computational time.

Most finite-element preprocessors offer constructs for defining geometry, but it is also desirable that geometry can be imported or exported from/to other software systems. National and international standards organizations have attempted to standardize geometry descriptions toward these ends. Geometry is sometimes described as clean or dirty, and tools to transform the latter into the former are often needed. Extraneous points or edges must be eliminated. Surfaces or edges smaller than those of the desired finite-element size must be blended or eliminated without corrupting the model. Sometimes, small features such as fillets and bosses are unimportant for preliminary analyses but become important as the design matures. For this reason, feature suppression is a desirable option.

Another issue to be considered is the number of analysis codes supported. Large organizations often deal with several commercial codes, and the needs of any organization will change with time. Preprocessors are always incomplete with regard to the number of features supported. For example, can seldom-used elements or routines be supported? Can the applied tractions be varied parabolically, sinusoidally, and so on? Can rotation be simulated? Can loads be scaled and superposed? Are contact capabilities supported? Are multiple coordinate systems conveniently supported? Are multipoint constraints supported? What about failure theories? Are composite materials modeled?

It is preferable to define loads and boundary conditions relative to the geometry rather than to the finite-element mesh. This permits remeshing of the geometry without the need to redefine loads and boundary conditions. It is also desirable to specify hard points, lines, or surfaces where parts will interact with other parts, so that the interaction (e.g., spot welding, adhesive bonding, etc.) can be properly represented. Those who must create a shell mesh from a solid geometry need a capability called midsurface extraction in order to optimally specify the defining surface for plate/shell representation.

Still another function of a preprocessor is the ability to group and categorize various items into convenient sets. Items may be organized geographically (spatially), by material, by a property such as thickness or moment of inertia, by contact pairing, by convection coefficient, or by various other attributes.

Finite-element model validation is another capability expected of the finite-element preprocessor. Elements must be investigated for geometric distortion. The traditional criteria (Fig. 7) are aspect ratio, taper, skew, warping, and edge curvature. Because even these traditional metrics are defined differently in different analysis codes, the user needs some control over their definition, their values, or the strategies imposed by the preprocessor for mesh cleanup, mesh smoothing, or element splitting. Still other validation tools are sometimes desirable: the ability to search for duplicate nodes or elements, inconsistent normals
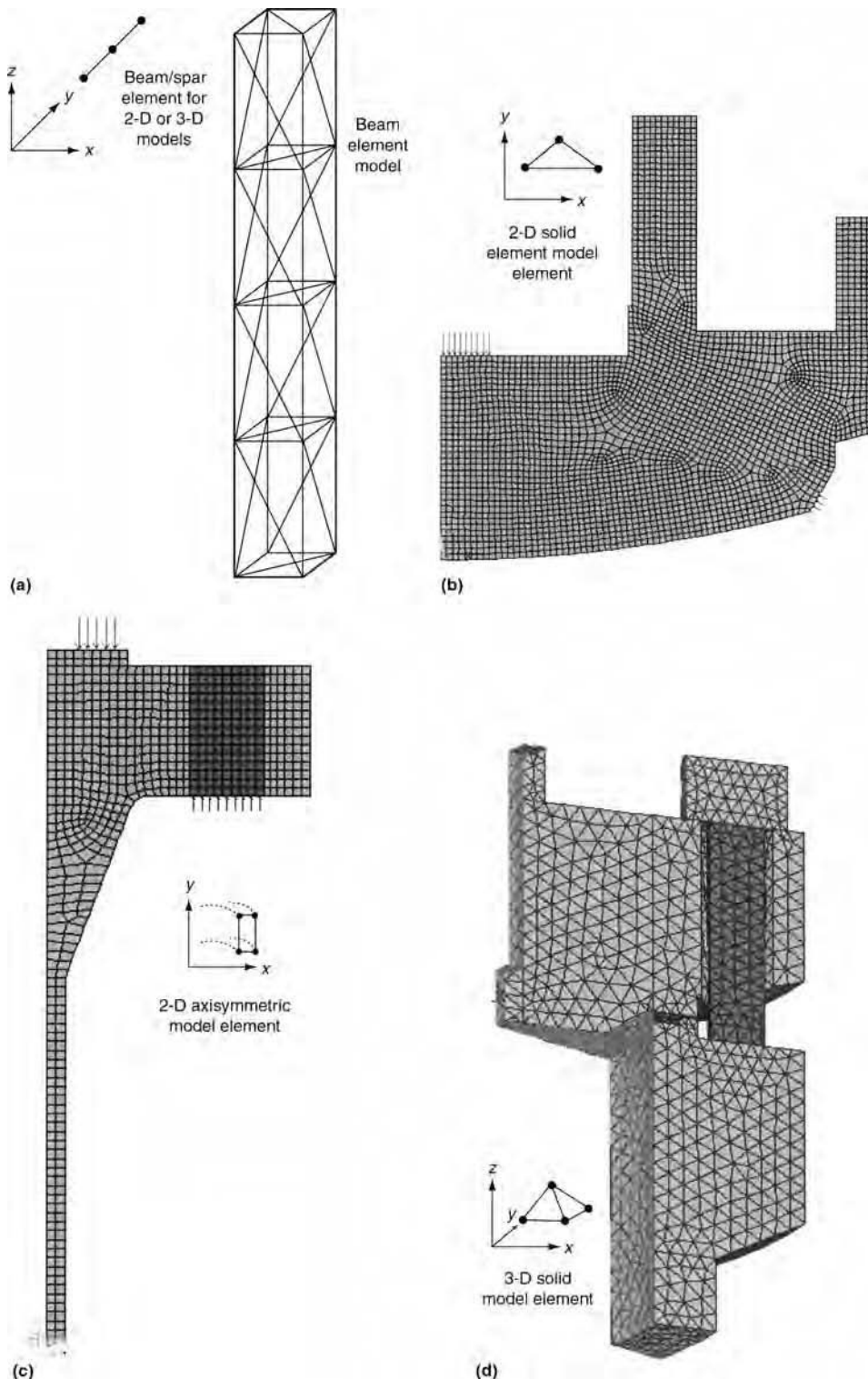
Fig. 5 Common types of finite elements used in modeling with examples. (a) Beam spar elements used to construct, for example, a beam element model. (b) Two-dimensional solid-model element with example. (c) Two-dimensional axisymmetric solid-model element with example. (d) Three-dimensional solid-model elements with simplified example

in shells, duplicate edges (implying a lack of continuity), missing elements, and so on.

Finally, returning to Fig. 2, it is apparent that the design process is iterative, that both geometry and mesh need to evolve and be updated in an appropriate manner. Macroinstruction capability in a preprocessor can greatly contribute to its usefulness. The ability to be a

part of a looping scheme where geometry is modified by the previous iteration in the loop is becoming more and more important.

The Basis of Finite Elements

To understand why some elements outperform other elements in certain problems, it is appropriate to review the *Rayleigh-Ritz procedure* (Ref 16) on which the FEM can be based. There are more general and mathematically sophisticated ways to approach finite-element theory, but they are less intuitive and less instructive to describe the underlying concept of the FEM.

The Rayleigh-Ritz procedure is an approximate method for solving structural problems based on the principle of virtual work. The principle of virtual work states that the total potential energy of an elastic system is a minimum (or stationary) when the system is in equilibrium. The Rayleigh-Ritz method reduces a continuum with infinitely many degrees of freedom (DOF) to a system with finite number of degrees of freedom. It accomplishes this by assuming the displacements of the continuum to be a function of a finite number of undetermined coefficients.

An example of solving for solution for these undetermined coefficients is illustrated as follows. This section also defines and contrasts two different types of elements: continuum and structural.

The Reduction of an Infinite DOF System to a Finite DOF System. A continuous system with infinitely many DOF may be approximated by a system with finite DOF. This can be accomplished by assuming a shape function with one or more undetermined coefficients whose values are to be determined.

For example, the simply supported beam of Fig. 8, which is loaded by a central concentrated load, P, is considered. A sinusoidal shape function is assumed, such that:

$$y = A \sin(\pi x/L) \quad 0 < x < L \qquad \text{(Eq 14)}$$

Here, A is the undetermined coefficient, and y is the displaced position of the beam centerline. When $x = 0$ and $x = L$, then $y = 0$. It should also be noted that:

$$d^2 y/dx^2 = -A(\pi/L)^2 \sin(\pi x/L)$$

This second derivative, which is proportional to beam curvature and therefore to moment, also vanishes at $x = 0$ and at $x = L$. Therefore, both the deflections and moments are zero at the ends of the beam. These are the appropriate boundary conditions.

The Rayleigh-Ritz procedure can be used to determine the value of the coefficient A. The method consists of finding an expression for the total potential energy of the system in terms of A, differentiating that expression with respect to A, and equating the results to zero. That is, one finds the deformed geometry that provides the minimum value of the total
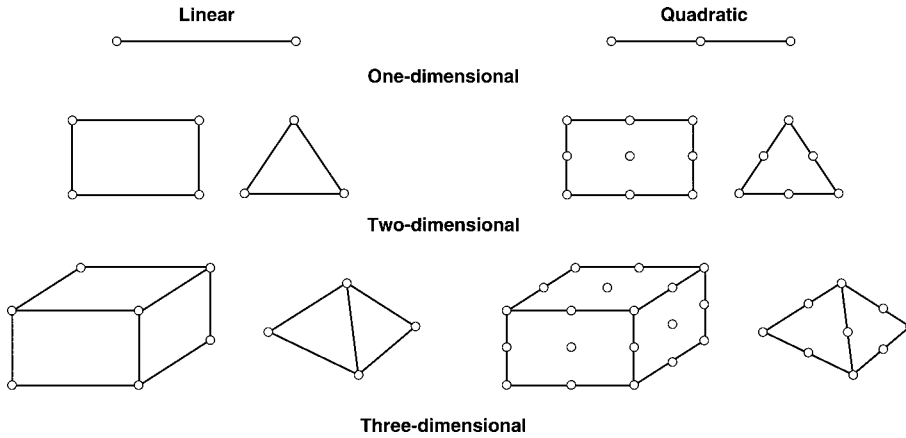
Fig. 6 Linear and quadratic elements used in typical finite-element analyses
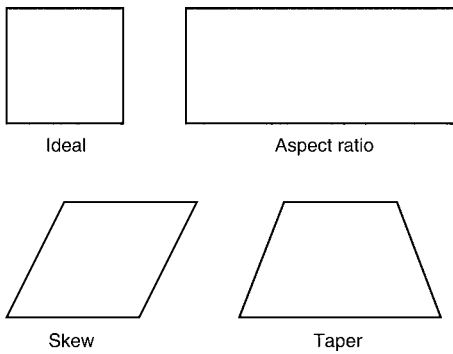
Fig. 7 Types of quadrilateral element distortion
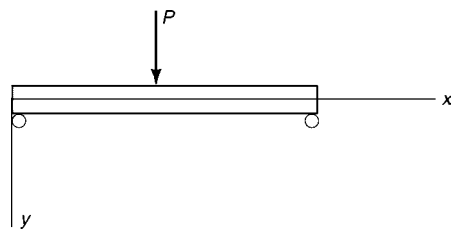
Fig. 8 Simply supported beam with central concentrated load, P

potential energy. In symbols, this is: V = elastic strain energy + gravitational potential energy, or:

$$V = U + \Omega \qquad \text{(Eq 15)}$$

where:

$$U = \left(\tfrac{1}{2}\right) \int EI\left(d^2 y / dx^2\right)^2 dx \qquad \text{(Eq 16)}$$

E is Young's modulus (material stiffness), I is section modulus (geometric stiffness), and:

$$\Omega = P(-A) \qquad \text{(Eq 17)}$$

For minimum potential energy:

$$dV/dA = 0 \qquad \text{(Eq 18)}$$

so that:

$$dV/dA = d/dA\left[\left(\pi^4 EIA^2 / 4L^3\right) - PA\right] = 0 \qquad \text{(Eq 19)}$$

and therefore:

$$A = \left(2PL^3\right)/\left(\pi^4 EI\right) \qquad \text{(Eq 20)}$$

and:

$$y = \left(2PL^3\right)/\left(\pi^4 EI\right) \sin(\pi x/L) \qquad \text{(Eq 21)}$$

A comparison of this solution with the known analytical solution shows that the deflections differ by 3% or less. However, the peak stress differs from the accepted solution by 19%.

A better approximation to this beam-bending problem should result from adding another degree of freedom, for example:

$$y = A \sin(\pi x/L) + B \sin(3\pi x/L) \qquad \text{(Eq 22)}$$

Following the same procedure as before and using $\partial V/\partial A = 0$ and $\partial V/\partial B = 0$, then:

$$A = \left(2PL^3\right)/\left(\pi^4 EI\right) \qquad \text{(Eq 23)}$$

and:

$$B = \left(2PL^3\right)/\left(27\pi^4 EI\right) \qquad \text{(Eq 24)}$$

Comparing this new solution to the analytical solution, deflections are everywhere within a $\pm 1\%$ band of the analytical solution, and peak stress differs from the accepted solution by 8%.

The FEM can be thought of as an extension to the Rayleigh-Ritz method. There are two major differences:

- In the Rayleigh-Ritz method, the structure is treated in its entirety, as one "element." In the FEM, multiple elements and nodes are used.
- In conventional finite elements, the values of the nodal displacements and rotations are the

variables; they are the undetermined coefficients. This is advantageous, because it is intuitive to understand and more convenient to specify displacements and rotational constraints at boundaries than to deal with, say, amplitudes of sine waves.

In summary, the FEM may be thought of as follows. The geometric continuum is divided into a number of elements (e.g., triangles, quadrilaterals, tetrahedra, etc.). The internal displacements of these elements are expressed in terms of the displacements at the nodes of the elements by means of interpolation functions, similar to the sinusoidal approximation (usually, the interpolation functions are polynomials) illustrated in the beam problem. An energy expression is formed and minimized in order to obtain a set of algebraic equations. The solution of these algebraic equations provides the displacements at the nodes. The values of the displacements at each node are analogs to the coefficient A in the previous example. In the FEM, the number of DOF is the product of the number of nodes multiplied by the number of unknowns per node. Knowing the displacements at each node implicitly gives the displacements and stresses throughout the continuum. As a generalization, the displacements obtained by this method are typically more accurate than the stresses.

Continuum Elements. A continuum element is one whose geometry is completely defined by its nodal coordinates. Hexahedra, pentahedra, and tetrahedra are the continuum elements commonly used in a 3-D domain. Quadrilaterals and triangles are the continuum elements commonly used in a 2-D domain.

Tetrahedra, pentahedra, and hexahedra typically have three DOF at each node. The three DOF at each node are the three translations (u, v, w) in the three spatial directions (X, Y, Z or R, Θ, Z or R, Θ, Φ). Figure 9 shows the assumed interpolation function that may be employed for each of these fundamental elements. Only the u displacement function is shown; the others have analogous form.

Triangles and quadrilaterals typically have two DOF at each node, the in-plane displacements. Triangles and quadrilaterals can also be structural elements, for example, when the bending (rotational) DOF are added. Structural elements are discussed as follows.

Figure 10 shows the possibility of enhancing continuum elements by adding interior nodes. (A node could also exist at the centroid, or multiple nodes could exist along edges.) The extra nodes supply more independent variables so that the displacement functions can contain more terms. The elements without interior nodes are sometimes called lower-order elements, and the ones with interior nodes are sometimes called higher-order elements.

Most commercial codes permit the selective omission of the interior nodes. This permits joining of lower-order elements to higher-order elements. If interior nodes are used, curved

edges and faces are possible. When the same interpolation functions are used for both the geometry and the displacements, the elements are known as isoparametric.

The optimum choice of element is code dependent and problem dependent. A generalization that is sometimes made is to use the higher-order elements for "smooth" problems, those relatively free of geometric and load discontinuities. Conversely, the lower-order element is used where there are abrupt changes in geometry, material, or loading.

Structural Elements. In addition to the continuum elements depicted in Fig. 9 and 10, there are a number of elements that behave in accordance with certain structural assumptions that predate finite-element technology. As long as these structural assumptions are valid, they provide for maximum solution efficiency (Fig. 11).

Beam elements are typically based on the so-called Euler-Bernoulli assumptions. These assumptions are that plane sections of the beam remain plane under deformation and that

bending stress is therefore directly proportional to the distance from the bending axis. The through-thickness normal stress is assumed to be zero. Shear deformations are not considered in the Euler-Bernoulli beam. The so-called Timoshenko beam adds shear deformation capability and possibly rotary inertia. The shear deformation is assumed to be constant throughout the beam depth, not parabolic. The Timoshenko assumptions tend to "soften" the response of the beam (i.e., displacements are higher and natural frequencies are lower than for the corresponding Euler-Bernoulli beam).

Plate elements are often based on the Kirchoff assumption that, after deformation, normals to the midsurface of the element remain normal to that surface. The Kirchoff assumption for plates is analogous to the Euler-Bernoulli assumptions for beams. More sophisticated assumptions are made for the Mindlin plate element, which is analogous to the Timoshenko beam. Shells, by definition, are curved surfaces. Finite-element representation of shells may be as flat facets or as curved elements. In either case, the finite elements representing shells must carry membrane as well as bending loads. The use of elements based on these various structural assumptions generally results in a very efficient model, especially for vibration and buckling problems. Accuracy of stresses may suffer at edges or discontinuities in the structure. This could be an important consideration for composite materials, subject to delamination from local though-thickness tensile stresses at the edges.

Superelements (Substructures). Several standard techniques exist that improve computational efficiency in special situations. The use of the various forms of symmetry has already been discussed. Another of these techniques is variously known as superelement or substructuring technology. The superelements are individually modeled, and the behavior of their interior nodes is determined as a function of the behavior of their external nodes. The assembly of superelements is then analyzed in

a separate run. The original motivation for this technique was to permit the solution of problems too large to be accomplished in a single pass. That is, it was an attempt to get around computer resource limitations. With the startling advances in hardware technology, that motivation has largely disappeared. However, there are several situations where superelement technology provides important advantages. These are:

- Iterative redesign of a structure will affect only one or a few of the superelements.
- Analysis of assemblies will be done by analysts in disparate locations, perhaps by individual groups on subassemblies (substructures), with results later integrated into the whole.
- Nonlinear analysis is needed in which one or more superelements retain linear behavior.
- Identical units occur repetitively within the structure.
- Mesh redesign is needed in a critical local region for more accurate results.

There are some potential pitfalls with the technology. Two examples are:

- The bandwidth (or wavefront) may be adversely affected by the substructuring process.
- The mass distribution may be adversely affected for dynamic problems, although no loss of accuracy is experienced for static
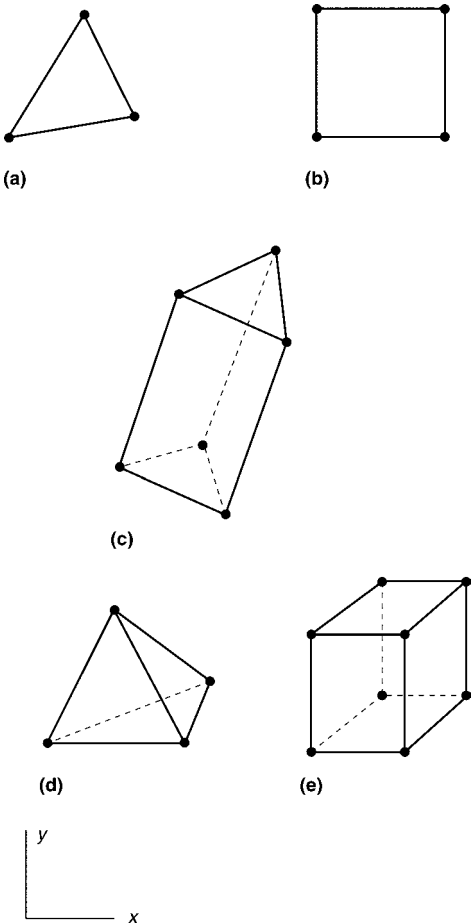
Fig. 9 Continuum elements (lower-order elements). (a) $u = a_o + a_1 x + a_2 y$; $\partial u / \partial x = a_1$; $\sigma = E\varepsilon_x = E\partial u/\partial x = Ea_1 =$ constant. (b) $u = a_o + a_1 x + a_2 y + a_3 xy$; $\partial x/\partial x = a_1 + a_3 y$; $\sigma = E\varepsilon_x = E(a_1 + a_3 y)$ = linear. (c) $u =$ directionally dependent. (d) $u = a_o + a_1 x + a_2 y$. (e) $u = a_o + a_1 x + a_2 y + a_3 xy$
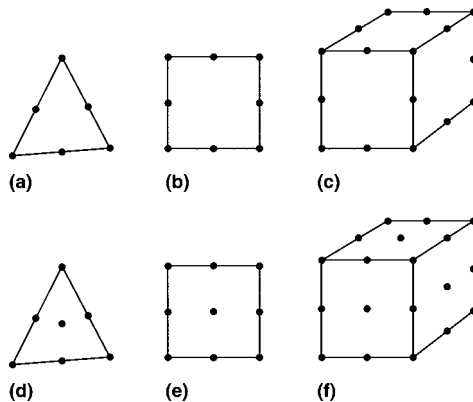
Fig. 10 Continuum elements (higher-order elements). (a) to (c) Sometimes called serendipity elements. (d) to (f) Sometimes called Lagrange elements
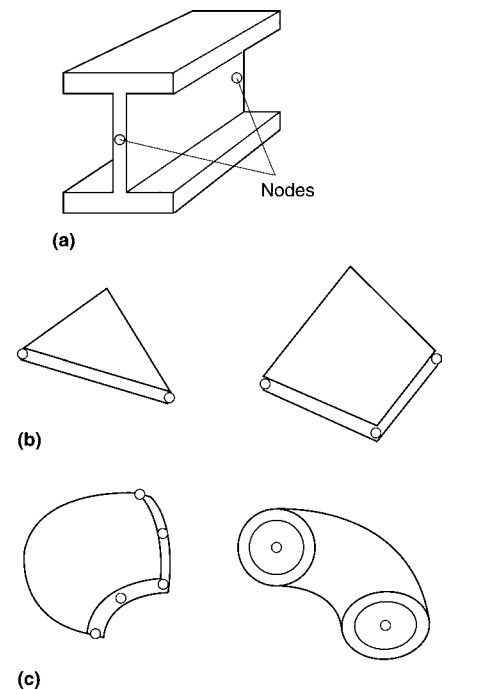
Fig. 11 Structural elements. (a) Beam element. A beam requires two or more nodes, two bending stiffnesses, torsional stiffness, orientation about the line connecting the nodes, material properties, and six degrees of freedom per node. (b) Plate elements. (c) Shell element. (d) Elbow element

problems. A technique called component mode synthesis attempts to compensate for the altered mass distribution (Ref 16).

Multipoint Constraints. In a variety of situations, it is desirable to use multipoint constraints (MPC), that is, to constrain the displacement of a node to that of another node or to a function of the displacements of other nodes. These situations include:

* Joining incompatible elements
* Simulating a rigid body
* Joining structural elements to continuum elements
* Distributing a load in a convenient fashion
* Joining dissimilar materials
* Transition of mesh density

Examples of the first three are shown in Fig. 12. Some comments on these situations follow.

Joining Incompatible Elements. The eight-noded quadrilateral, 201, is joined to the four-noded quadrilateral, 101, by constraining the midside node of 201, node 10, to displace the average of the two corner nodes. In that way,
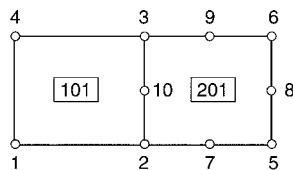
(a)
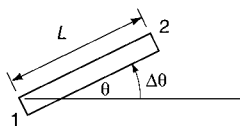
(b)
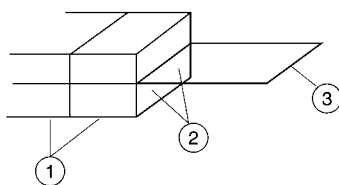
(c)

Fig. 12 Multipoint constraints. (a) Joining incompatible elements. If u and v denote displacements in the x and y directions, respectively, and subscripts refer to the nodes, then $u_{10} = (u_2 + u_3)/2$ and $v_{10} = (v_2 + v_3)/2$. This assumes that node 10 is halfway between nodes 2 and 3. (b) Simulating a rigid body. For a rigid body motion $\Delta\theta$, $u_2 = u_1 L \sin\theta\Delta\theta$ and $v_2 = v_1 + L \cos\theta\Delta\theta$. The motion of node 2 is constrained to that of point 1. (c) Joining structural elements to continuum elements. (1) Solid continuum elements with three translational degrees of freedom per node only. (2) These nodes are slaved to nodes immediately above and below them in order to prevent simulation of the mechanism (hinge). (3) Plate element with bending stiffness (three translational degrees of freedom, two rotational degrees of freedom)

no gap or interpenetration of the two elements can occur along their common edge. The MPC is an effective way to join these two different element types.

Joining Structural Elements to Continuum Elements. Shell elements typically have five DOF per node, three translations, and two bending rotations. Three-dimensional continuum elements (i.e., solids) have only the three translational DOF. Joining these two element types will produce a hinge at their interface unless care is taken. Commonly, the hinge is prevented by constraining the interface nodes to immediately adjacent nodes.

Linear Finite-Element Problems

Many finite-element problems can be solved as linear problems. In structural applications, for example, deflections are assumed to be small in the linear (elastic) stress-strain region. Under these assumptions, solution times are quite predictable. Linear algebra can apply in this case, as described subsequently for a general structural problem.

Linear Algebra Solution for a Structural Problem. For static structural problems, the FEM results in a system of equations that can be expressed in matrix form as:

$$\{\mathbf{F}\} = [\mathbf{K}]\{\mathbf{u}\} \qquad \text{(Eq 25)}$$

where $\{\mathbf{F}\}$ is a column vector of forces, $\{\mathbf{u}\}$ is a column vector of displacements, and $[\mathbf{K}]$ is a square matrix. There will be multiple $\{\mathbf{F}\}$ and $\{\mathbf{u}\}$ vectors for problems with multiple load cases. The size of the $[\mathbf{K}]$ matrix is the product of the number of nodes and the number of DOF per node. If all DOF of the problem acted completely independently of each other (e.g., a number of axial springs, each separately attached to ground but not to each other), then the stiffness matrix would be purely diagonal (that is, all terms off the diagonal in the matrix $[\mathbf{K}]$ would be zero).

Most structural problems have stiffness terms clustered, or banded, around the diagonal. This bandedness may be influenced by either the node-numbering strategy or the element-numbering strategy. Various solution strategies (solvers) are available in commercial codes, and they exploit (or ignore) bandedness in different ways. Some commercial codes offer only one solver; others may offer a half dozen or more. Be aware that some solvers are sensitive to the node-numbering strategy (e.g., sparse solver, skyline solver), while some solvers are sensitive to the element-numbering strategy (wavefront, multifront). Iterative solvers, on the other hand, are insensitive to node and element numbering and are indeed insensitive to bandedness. For this reason, the iterative solvers are excellent for large, chunky continuum problems, the kind of problems with large bandwidth. Iterative solvers are usually well suited to parallel computers (multiple central processing units), whether the computers share

memory or not. Iterative solvers, however, are not as efficient for problems with multiple load cases, because each load case requires a repetition of the iteration strategy.

Dynamic structural problems add mass and damping terms to the system of equations:

$$\{\mathbf{F}\} = [\mathbf{K}]\{\mathbf{u}\} + [\mathbf{C}]\{\mathbf{du/dt}\} + [\mathbf{M}]\left\{\mathbf{d^2u/dt^2}\right\}$$
$$\text{(Eq 26)}$$

where $\{\mathbf{du/dt}\}$ and $\{\mathbf{d^2u/dt^2}\}$ are column vectors of velocities and accelerations, respectively. The viscous damping matrix $[\mathbf{C}]$ is usually assumed to be proportional to either the $[\mathbf{K}]$ matrix or the mass matrix, $[\mathbf{M}]$, for computational efficiency. The $[\mathbf{C}]$ and $[\mathbf{M}]$ matrices are generally banded. Special techniques of lumping masses at nodes are often used to make the $[\mathbf{M}]$ matrix diagonal for computational efficiency. Dynamic problems can be solved in the time domain, using time-stepping procedures starting from known initial conditions. Because an additional solution is required for each time increment, this analysis is expensive. Dynamic problems, such as the determination of natural frequencies, can be solved in the frequency domain as opposed to the time domain. This is accomplished by assuming that the displacement vector is a harmonic function:

$$\{\mathbf{u}\} = \{\phi\}e^{i\omega t} = \{\phi\}(\cos\omega t + i\sin\omega t) \qquad \text{(Eq 27)}$$

where

$$i = \sqrt{-1}$$

and t is time, $\{\phi\}$ is the modal vector (vector of unknown amplitude at the nodes), and ω is the natural frequency associated with the mode shape. This assumption creates an eigenvalue problem (Ref 17).

Nonlinear Finite-Element Problems (Ref 18)

Many systems are nonlinear. These problems usually require iterative solution techniques and additional computer resources (e.g., disk space and solution time). One basic example of a nonlinear problem would be when stress causes plastic deformation, so that strain is not directly proportional to load. Large deflections can also result in nonlinear problems, depending on geometry. Problems of flat plates, for example, are usually assumed to be linear if the lateral deflection does not exceed the thickness of the plate. After that point, the plate stiffens due to membrane effects. Another type of large deflection problem occurs when the direction of the load follows the deformation of the structure. Still other types occur when a gap is closed or opened, or one component of an assembly strikes another.

Problems of inelastic material behavior are another class of nonlinear problems. Problems of creep, relaxation, hysteresis, phase change, and residual stress fall into this category.

Fortunately, most nonlinear material models used in finite-element solutions carry over from disciplines that predate finite-element technology (e.g., theory of plasticity). However, for problems involving a high rate of deformation (e.g., the forming of metal or the crash of a vehicle), special finite-element software based on so-called explicit integration techniques is available (Ref 19).

A problem is nonlinear if the force-displacement relationship depends on the current state, that is, on current displacement, force, and stress-strain relationship:

$$\mathbf{P} = \mathbf{K}(\mathbf{P}, \mathbf{u})\mathbf{u} \qquad \text{(Eq 28)}$$

where $\mathbf{u}$ is a displacement vector, $\mathbf{P}$ is a force vector, and $\mathbf{K}$ is the stiffness matrix. Linear problems form a subset of nonlinear problems. For example, in classical linear elastostatics, this relationship can be written in the form:

$$\mathbf{P} = \mathbf{K}\mathbf{u} \qquad \text{(Eq 29)}$$

where the stiffness matrix $\mathbf{K}$ is independent of both $\mathbf{u}$ and $\mathbf{P}$. If the matrix $\mathbf{K}$ depends on other state variables, such as temperature, radiation, and so on, but does not depend on displacement or loads, the problem is still linear. Similarly, if the mass matrix is constant, the following dynamic problem is also linear:

$$\mathbf{P} = \mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}\mathbf{u} \qquad \text{(Eq 30)}$$

There are three sources of nonlinearities: material, geometry, and boundary condition. The material nonlinearity results from the nonlinear relationship between stresses and strains due to material plasticity. Geometric nonlinearity results from the nonlinear relationship between strains and displacements or the nonlinear relationship between stresses and forces. If the stress measure is energetically conjugate to the strain measure, both sources of nonlinearity have the same form. This type of nonlinearity is mathematically well defined but often difficult to treat numerically. Boundary conditions such as contact or friction are also sources of nonlinearities. This type of nonlinearity manifests itself in several real-life situations, for instance, in metal forming, gears, interfaces of mechanical components, pneumatic tire, and crash. A load on a structure causes nonlinearity if it changes with the displacement and deformation of the structure (such as pressure loading).

The kinematics of deformation can be described by Lagrangian, Eulerian, and Arbitrary Lagrangian-Eulerian (ALE) formulations. In the Lagrangian method, the finite-element mesh is attached to the material and moves through space along with the material. In the Eulerian formulation, the finite-element mesh is fixed in space, and the material flows through the mesh. In the ALE formulation, the grid moves independently from the material, yet in a way that spans the material at any time. The Lagrangian approach can be further classified in two categories: the total and the updated

Lagrangian methods. In the total Lagrangian approach, equilibrium is expressed with the original undeformed reference frame, while, in the updated Lagrangian approach, the current configuration acts as the reference frame. In the latter, the true or Cauchy stresses and an energetically conjugate strain measure, namely, the true strain, are used in the constitutive relationships. The updated Lagrange approach is useful in:

- Analyses of shell and beam structure in which rotations are large so that the nonlinear terms in the curvature expressions may no longer be neglected
- Large strain-plasticity analyses in which the plastic deformations cannot be assumed to be infinitesimal

In general, this approach can be used to analyze structure where inelastic behavior causes large deformations. The (initial) Lagrangian coordinate frame has little physical significance in these analyses, because inelastic deformations are, by definition, permanent. Therefore, the updated Lagrangian formulation is appropriate for the simulations of sheet metal forming processes. For theses analyses, the Lagrangian frame of reference is redefined at the last completed iteration of the current increment. The variational form of the equation for the static problem in the updated Lagrangian approach is given as:

$$\int_V \frac{\partial \delta \mathbf{u}_i}{\partial \mathbf{x}_j} \sigma_{ij} dV - \int_\Gamma \delta \mathbf{u}_i f_i d\Gamma = 0 \qquad \text{(Eq 31)}$$

where V is the volume considered, Γ is the surface on which the traction components f_i are imposed, and σ_{ij} is the Cauchy stress. The linearized variational form of Eq 31 needed for the Newton-Raphson numerical solver can be written as:

$$\int_V \frac{\partial \delta \mathbf{u}_i}{\partial \mathbf{x}_j} C_{ijkl} \frac{\partial \Delta \mathbf{u}_k}{\partial \mathbf{x}_l} dV + \int_V \frac{\partial \delta \mathbf{u}_i}{\partial \mathbf{x}_j} \sigma_{ji} \frac{\partial \Delta \mathbf{u}_i}{\partial \mathbf{x}_l} dV$$
$$= -\int_V \frac{\partial \delta \mathbf{u}_i}{\partial \mathbf{x}_j} \sigma_{ij} dV + \int_\Gamma \delta \mathbf{u}_i f_i d\Gamma$$

$$\text{(Eq 32)}$$

The left side of Eq 32 corresponds to the material and geometric stiffness, while its right side is associated with the internal and external force vectors, respectively. In the FEM, the left side dominates the convergence rate, and the right side directly controls the accuracy of the solution.

Finite-Element Design

Finite-element methods are used mostly for analysis of boundary-value problems. This means that the problem is well set, with known tool and material geometries, material and interface properties, and realistic stress- and

displacement-imposed boundary conditions. This type of application of the method is referred to as FEA. However, in practice, the final shape of the product is imposed, and the manufacturing process must be designed around it. Therefore, in order to improve the conventional trial-and-error-based practices for optimizing forming processes, either by experiments or FEA, a finite-element design theory, called ideal forming theory, was proposed (Ref 20, 21). The finite-element implementation involves a time-efficient one-step code, providing not only the initial blank geometry but all the intermediate shapes and the entire load history necessary to achieve it, thus providing invaluable information about the ideal process parameters.

This application is referred to as finite-element design, or the Ideal Forming Design Theory. In this theory, materials are prescribed to deform following the minimum plastic work path. The final product shape is specified, and the initial blank shape is obtained from the global extremum plastic work criterion as a one-step backward solution. Although the theory is general enough to accommodate any other form of constraints, the underlying physical assumption of this extremum work condition is that the strain gradients are minimized on the overall part, thus departing as much as possible from plastic flow localization modes. In order to consider local thinning effects due to friction, a method based on a modified extremum work criterion has also been developed (Ref 22). The ideal forming theory has been successfully applied for sheet-forming processes to optimize flat blanks (Ref 23–25) and also for bulk forming in steady and nonsteady flows (Ref 26, 27).

Sheet Metal Forming

Nonlinear FEMs are becoming very popular in sheet metal forming process simulations (Ref 18). Nonlinearities occur in sheet metal forming processes from elastoplastic material behavior, large rotations, and contacts between the tools and the blank.

Finite-element analyses of sheet metal forming processes can be broadly classified into three categories according to the element types used:

- Membrane analysis (plane stress without bending stiffness)
- Shell analysis (plane stress with bending stiffness)
- Continuum analysis (general stress state)

For sheet metal forming simulations, the shell analysis is the most popular. Two basic approaches concerning the development of nonlinear shell finite elements include:

- Classical shell elements
- Degenerated solid elements

The classical shell elements are directly based on the governing differential equations

of an appropriate shell theory. Despite the potential economy of such elements, the development of nonlinear shell elements involves mathematical complexities. The degenerated solid element, which was initiated by Ahmad et al. (Ref 28) for the linear analysis of continuum formulation, is reduced in dimensionality by direct imposition of kinematics and constitutive constraints. The works of Ramm (Ref 29), Parish (Ref 30), Hughes and Liu (Ref 31, 32), Dvorkin and Bathe (Ref 33), and Liu et al. (Ref 34), among many others, constitute representative examples of this methodology carried out in the most general way for the nonlinear regime. The application of shell elements to sheet-forming simulations can be found in Ref 35 to 40. The kinematic formulation for a continuum-based shell is summarized in Ref 18.

Elastic-Plastic Stress Integration (Ref 18)

Most rate-independent plastic models are formulated in terms of rate-type constitutive equations, for which the integration method has a considerable influence on the efficiency, accuracy, and convergence of the solution. In the simulation of sheet-forming processes, the constitutive equation is integrated along an assumed deformation path.

Among the infinite ways to assume the deformation path, the minimum plastic work path in homogeneous deformation has been found to have several advantages. Requirements for achieving minimum plastic work paths in homogeneous deformation are well documented in Ref 41 to 44. The minimum work path, which is also the proportional logarithmic (true) strain paths, is achieved under two conditions. First, the set of three principal axes of stretching is fixed with respect to the material; second, the logarithms of the principal stretches remain in a fixed ratio.

The incremental deformation theory based on the minimum plastic work path enables convenient decoupling of deformation and rotation by the polar decomposition at each process increment. The resulting incremental constitutive law is frame-indifferent (objective), because the theory uses a materially embedded coordinate system.

The incremental deformation theory is useful for the FEM of rigid-plastic and elastoplastic constitutive formulations. In rigid plasticity, the theory was introduced for process analyses by Yang and Kim (Ref 45), Germain et al. (Ref 46), Chung and Richmond (Ref 47, 48), and Yoon et al. (Ref 49). For elastoplastic materials, the incremental deformation theory has been successfully applied for materials exhibiting isotropy (Ref 50) and planar anisotropy (Ref 40, 51).

Continuum Models. The most popular scheme for stress integration is the predictor-corrector method (often called return mapping). This method is applied in two successive steps: the prediction step, during which a trial stress state is estimated, and the corrector step, during which a flow rule is applied by return mapping procedures in order to bring (project) the stress

onto the yield surface or, in other words, to ensure the consistency condition (i.e., the stress state must be on the yield surface). During the return mapping procedure, a reasonable assumption for the deformation path must be imposed. A return mapping procedure was first introduced in the paper of Wilkins (Ref 52). The works of Ortiz and Pinsky (Ref 53) and Ortiz et al. (Ref 54) apply the closet point-projection method to perform the procedure in plane-stress conditions. Later, Ortiz and Simo (Ref 55) developed a new class of integration algorithms based on the cutting-plane approach. However, this approach has no clear physical meaning in the deformation path viewpoint. Yoon et al. (Ref 40) proposed the multistage return mapping method based on incremental deformation theory, which follows the minimum plastic work path. In this work, it was proven that, when the consistency condition and normality rule (strain increment normal to yield surface) are imposed, this new projection during the current unknown step becomes the closet point projection. The stress integration procedure is briefly summarized as follows.

The increment of the Cauchy stresses is given by applying the fourth-order elastic modulus tensor $\mathbf{C}^e$ to the incremental second-order elastic strain tensor $\Delta\boldsymbol{\varepsilon} - \Delta\boldsymbol{\varepsilon}^p$:

$$\Delta\boldsymbol{\sigma} = \mathbf{C}^e(\Delta\boldsymbol{\varepsilon} - \Delta\boldsymbol{\varepsilon}^p) \qquad \text{(Eq 33)}$$

in which $\Delta\boldsymbol{\varepsilon}$ and $\Delta\boldsymbol{\varepsilon}^p$ are the total and plastic strain increments, respectively. In order to follow the minimum plastic work path in the incremental deformation theory, the logarithmic plastic strain must remain normal to the yield surface at the representative stress state, that is:

$$\Delta\boldsymbol{\varepsilon}^p = \Delta\bar{\varepsilon}^p \frac{\partial\bar{\sigma}}{\partial\boldsymbol{\sigma}} = \gamma\mathbf{m} \qquad \text{(Eq 34)}$$

where $\mathbf{m} = \widehat{o}\bar{\sigma}/\partial\boldsymbol{\sigma}$ is a symbolic notation that represents the tensor of component

$m_{ij} = \widehat{o}\bar{\boldsymbol{\sigma}}/\partial_{ij}$. The condition stipulating that the updated stress stays on the strain-hardening curve provides the following equation:

$$\begin{aligned}
F(\lambda) &= \bar{\sigma}(\boldsymbol{\sigma}_n + \Delta\sigma) - \mathbf{h}(\bar{\varepsilon}_n^p + \lambda) \\
&= \bar{\sigma}[\boldsymbol{\sigma}_n + \mathbf{C}^e(\Delta\boldsymbol{\varepsilon} - \lambda\mathbf{m})] - \mathbf{h}(\bar{\varepsilon}_n^p + \lambda) = 0
\end{aligned}$$

$$\text{(Eq 35)}$$

where the subscript "n" denote quantities at step "n" in the simulation.

The predictor-corrector scheme based on the Newton-Raphson method is generally used to solve the nonlinear system in Eq 35 for $\lambda\Delta\bar{\varepsilon}^p$. However, while a mathematical solution to this equation does exist, it can be difficult to obtain numerically if the strain increment is not small enough. In the examples in the next section, a multistage return mapping procedure based on the control of the residual suggested by Yoon et al. (Ref 40) was employed. The proposed method is applicable to nonquadratic yield functions and general strain-hardening laws without a line search algorithm, even for a relatively large strain increment (10%). At the end of the step, when Eq 35 is solved, all kinematic variables and stresses are updated.

In order to consider the rotation of the anisotropic axes, a co-rotational coordinate system (constructed at each integration point), is defined and initially coincides with the material symmetry axes. In sheet forming, it is assumed that the orthogonality of the anisotropy axes is preserved during sheet forming under the isotropic hardening assumption. This assumption is generally considered as appropriate in sheet-forming process simulations. From the polar decomposition theorem, the deformation of a material element represented by the deformation gradient tensor, $\mathbf{F}$, is the combination of a pure rotation, $\mathbf{R}$, and a pure stretch, $\mathbf{U}$ ($\mathbf{F} = \mathbf{R}\mathbf{U}$). The rotation of the anisotropy axes is updated incrementally at every step by the rotation amount ($\mathbf{R}$) obtained from the polar decomposition (Fig. 13). For instance, if at the
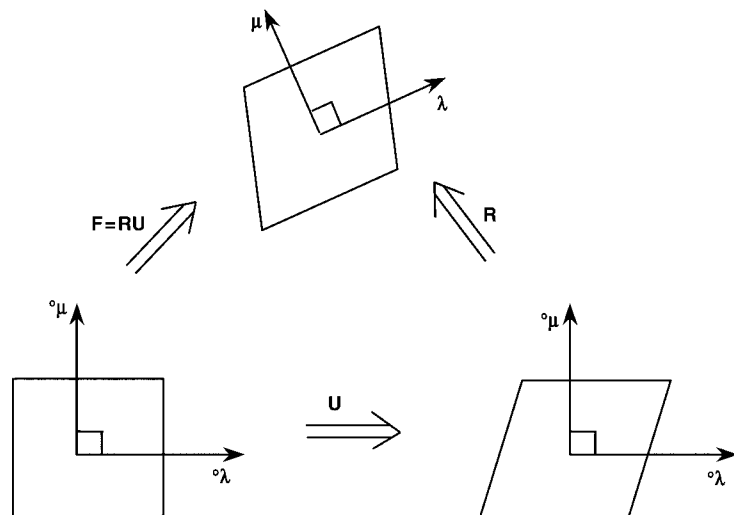
Fig. 13 Rotation of anisotropic axes during deformation. $\mathbf{F}$, deformation gradient tensor; $\mathbf{R}$, pure rotation; $\mathbf{U}$, pure stretch; $^\circ\boldsymbol{\mu}$, unit vector for transverse direction; $^\circ\boldsymbol{\lambda}$, unit vector for rolling direction

first step $^{o}\boldsymbol{\lambda}$ and $^{o}\boldsymbol{\mu}$ are the unit vectors coinciding with the rolling and transverse directions, respectively, the updated axes are given by:

$$\boldsymbol{\lambda} = \mathbf{R}^{\circ}\boldsymbol{\lambda}$$
$$\boldsymbol{\mu} = \mathbf{R}^{\circ}\boldsymbol{\mu} \qquad \text{(Eq 36)}$$

Crystal-Plasticity Models. These models account for the deformation of a material by crystallographic slip and for the reorientation of the crystal lattice. The influence of crystal symmetry on elastic constants can be included, and the strain-hardening and cross-hardening effects between the slip systems can be incorporated through the use of state variables. Furthermore, a rate-dependent approach is also typically employed to relate the shear stresses and shear strains on the different slip systems. The kinematics of the model were summarized by Dao and Asaro (Ref 56). The deformation gradient, $\mathbf{F}$, is decomposed into a plastic deformation, $\mathbf{F}^{p}$, which is the summation of the shear strain for each slip system, and a combination of elastic deformation and rigid body motion of the crystal lattice, $\mathbf{F}^{e}$, as shown in Fig. 14, that is:

$$\mathbf{F} = \mathbf{F}^{e}\mathbf{F}^{p} \qquad \text{(Eq 37)}$$

Because plasticity occurs by dislocation slip, the plastic deformation rate, $\mathbf{L}^{p}$, is determined by the summation of the shear strain contribution over all of the slip system (Ref 56):

$$\mathbf{L}^{p} = \sum_{(s)} \dot{\gamma}^{(s)} \mathbf{b}^{(s)} \mathbf{n}^{(s)} \qquad \text{(Eq 38)}$$

Here, (s) denotes a slip system, $\mathbf{n}^{(s)}$ is the normal of the slip plane, and $\mathbf{b}^{(s)}$ is the vector in the slip direction. Both vectors are orthogonal, that is, $\mathbf{n}^{(s)} \times \mathbf{b}^{(s)} = \mathbf{0}$, and are assumed to rotate with the elastic spin of the lattice.

In general, these vectors are not unit vectors like in a phenomenological model because they are allowed to stretch. Using Eq 38, the symmetric and skew symmetric parts of $\mathbf{L}^{p}$, the rate of deformation tensor $\mathbf{D}^{p}$, and plastic spin $\mathbf{W}^{p}$, respectively, can be written as:

$$\mathbf{D}^{p} = \tfrac{1}{2}(\mathbf{L}^{p} + \mathbf{L}^{pT}) = \sum_{(s)} \dot{\gamma}^{(s)} \tfrac{1}{2}\left(\mathbf{b}^{(s)}\mathbf{n}^{(s)} + \mathbf{n}^{(s)}\mathbf{b}^{(s)}\right)$$

$$\text{or } \mathbf{D}^{p} = \sum_{(s)} \dot{\gamma}^{(s)} \mathbf{P}_{(s)} \qquad \text{(Eq 39)}$$

$$\mathbf{W}^{p} = \tfrac{1}{2}(\mathbf{L}^{p} + \mathbf{L}^{pT}) = \sum_{(s)} \dot{\gamma}^{(s)} \tfrac{1}{2}\left(\mathbf{b}^{(s)}\mathbf{n}^{(s)} - \mathbf{n}^{(s)}\mathbf{b}^{(s)}\right)$$

$$\text{or } \mathbf{W}^{p} = \sum_{(s)} \dot{\gamma}^{(s)} \mathbf{W}_{(s)} \qquad \text{(Eq 40)}$$

where the tensors $\mathbf{P}_{(s)}$ and $\mathbf{W}_{(s)}$ have been introduced for notational convenience.

Usually, the Cauchy or true stress, $\boldsymbol{\sigma}$, is employed for the stress integration. However, in this section, the Kirchhoff stress, $\boldsymbol{\tau}$, is considered. By ignoring the elastic volume change, the Cauchy stress is related to the Kirchhoff stress, $\boldsymbol{\tau}$, through the relationship $\boldsymbol{\tau} = J\boldsymbol{\sigma}$, where J is the determinant of $\mathbf{F}^{e}$. Writing the rate of deformation tensor as the sum of elastic and plastic parts, an objective stress rate, $\boldsymbol{\tau}$, can be expressed as:

$$\breve{\boldsymbol{\tau}} = \mathbf{K} : \mathbf{D} - \sum_{(s)} \dot{\gamma}^{(s)}\left(\mathbf{K} : \mathbf{P}_{(s)} + \mathbf{W}_{(s)} \cdot \boldsymbol{\tau} - \boldsymbol{\tau} \cdot \mathbf{W}_{(s)}\right)$$
$$= \mathbf{K} : \mathbf{D} - \sum_{(s)} \dot{\gamma}^{(s)} \mathbf{R}_{(s)}$$

$$\text{(Eq 41)}$$

where $\mathbf{K}$ is the fourth-order elastic modulus tensor, and $\mathbf{D}$ is the rate of deformation tensor. Then, the resolved shear stress at the end of the time step becomes:

$$\boldsymbol{\tau}^{(s)}_{t+\Delta t} = \boldsymbol{\tau}^{(s)}_{t} + \Delta t \dot{\boldsymbol{\tau}}^{(s)} = \boldsymbol{\tau}^{(s)}_{t} + \Delta t \mathbf{R}_{(s)} : \mathbf{D} - \Delta t$$
$$\times \sum_{(\alpha)} \dot{\gamma}^{(a)} \mathbf{R}_{(s)} : \mathbf{P}_{(\alpha)}$$

$$\text{(Eq 42)}$$

where $\mathbf{R}_{(s)} = \mathbf{K} : \mathbf{R}_{(s)} + \mathbf{W}_{(s)} \times \boldsymbol{\tau} - \boldsymbol{\tau} \times \mathbf{W}_{(s)}$.

The resolved shear stress, $\boldsymbol{\tau}^{(s)}$, also follows the rate-dependent hardening rule:

$$\tau^{(s)} = g^{o}\left(\frac{|\dot{\gamma}^{(s)}|}{\dot{\gamma}_{o}}\right)^{m} \text{sign}\left(\dot{\gamma}^{(s)}\right) \qquad \text{(Eq 43)}$$

where g^{o} is the shear yield stress on a slip system.

Finally, the resolved shear stress defined by Eq 42 and 43 must have the same value at $t + \Delta t$:

$$E^{s}\left(\dot{\gamma}^{(s)}\right) = \tau^{(s)}_{t} + \Delta t \mathbf{R}_{s} : \mathbf{D} - \Delta t$$
$$\times \sum_{(\alpha)} \dot{\gamma}^{(a)} \mathbf{R}_{(s)} \times \mathbf{P}_{\alpha-g^{o}}\left(\frac{|\dot{\gamma}^{(s)}|}{\dot{\gamma}_{o}}\right)^{m} \text{sign}\left(\dot{\gamma}^{(s)}\right) = 0$$

$$\text{(Eq 44)}$$

Equation 44 is a nonlinear equation. It is solved with the Newton-Raphson method after linearization (Ref 57), that is:

$$E^{s}\left(\dot{\gamma}^{(s)}\right) + \sum_{(\alpha)} \frac{dE^{s}}{d\dot{\gamma}^{(\alpha)}} d\dot{\gamma}^{(\alpha)} = 0 \qquad \text{(Eq 45)}$$

Bulk Working

The FEM has become a widely used tool in materials forming industries. Some of the common metal-forming problems that are solved by FEM software are insufficient die fill, poor shape control, poor flow of materials, prediction of cracks and voids that lead to fracture, and poor final part properties. Finite-element analysis has many advantages over the typical closed-form solutions. It provides greater insights into the behavior of the product and the process, gives a good understanding of the performance of the product before the actual usage, is useful in product and process optimization, is a powerful and mature design and analysis tool, and, most importantly, gives solutions to irregular shapes, variable material properties, and irregular boundary conditions.

Historical Overview (Ref 58)

Since the 1970s, continuous improvement in finite-element software and computing technology have evolved rapidly in the analysis of 3-D metal flow as well as heat transfer and stress and strain distributions in metals forming. In the 1980s and 1990s, many advances were made in the modeling of metal-forming processes using rigid-viscoplastic FEM in modeling of complex 3-D deformation.

The rigid-plastic FEM is formulated on the basis of plasticity theory. Lee and Kobayashi first introduced the rigid-plastic formulation in
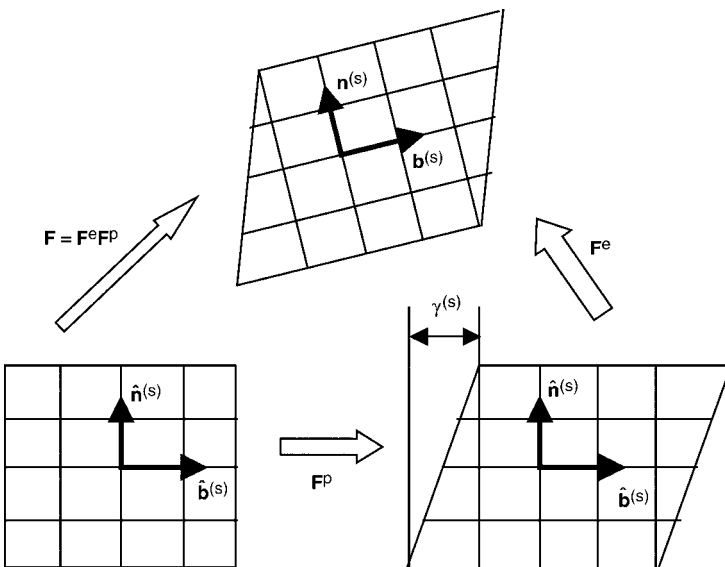
Fig. 14 Multiplicative decomposition of deformation gradient. See text for description

the 1970s (Ref 59). This formulation neglects the elastic response of deformation calculations. In the late 1970s and early 1980s, a processing science program (Ref 60) funded by the United States Air Force was performed at the Battelle Memorial Institute Columbus Laboratories to develop a process model for the forging of dual-property titanium engine disks. These disks are required to have excellent creep and high stress-rupture properties in the rim and high fatigue strength in the bore region. An FEM-based code, ALPID (Ref 61), was developed under this program.

Thermoviscoplastic FEM analyses (Ref 62) were also performed to investigate the temperature variation during hot-die disk-forging processes. The flow stress of thermo-rigid-viscoplastic material is a function of temperature, strain, and strain rate. Approximately five aerospace manufacturers pioneered the use of the code. Based on the same foundation, DEFORM was developed for 2-D applications in 1986. Due to the large deformation in the metal-forming application, the updated Lagrangian method always suffers from mesh distortion and consequently requires many remeshings to complete one simulation.

Two-dimensional metal-forming procedures became practical for industrial use when automated remeshing became available in 1990 (Ref 63). In the beginning of the 1980s, the PDP11 and the CDC/IBM mainframe computers were used. In the mid-1980s, the VAX workstation became the dominant machine for running the simulations. In the late 1980s, UNIX workstations became the primary computing facility.

Unfortunately, the majority of the metalworking processes are 3-D, where a 2-D approach cannot approximate reality satisfactorily. The initial 3-D code development began in the mid-1980s (Ref 64). One simulation with backward extrusion in a square container was reported to take 152 central processing unit hours on a VAX-11/750. In addition to the need for remeshing, a more complicated process was estimated to take several weeks. Due to the lack of computing speed in the 1980s for 3-D applications, the actual development was delayed until the 1990s (Ref 65).

Adoption of computer simulation for bulk deformation processing has been brought about by the great increases in computer power that occurred in the 1990s and the development of FEM software with automated mesh generation (AMG) capability (Ref **66**). The development of AMG has eliminated the time-consuming and error-prone process of generating a new mesh on a highly deformed body by interpolating the data from the old mesh. AMG is currently available in software specifically designed for bulk deformation processing, such as DEFORM 2D and DEFORM 3D, as well as on other commercially available FEM codes such as MARC and MacNeal-Schwendler that have forging packages. With reasonably fast computers, simulation of 2-D problems, such

as axisymmetric and plane-strain geometries, can be done in minutes or hours, but 3-D simulations may not always be cost-effective, depending on the required engineering and computer time.

Since then, many ideas to develop a practical 3-D numerical tool have been evaluated and tested. The successful ones were finally implemented. After the mid-1990s, significant computing speed improvement was seen in personal computer (PC) technology, coupled with a lower price as compared to UNIX-based machines. For this reason, the PC has become the dominant computing platform. Due to the competition for better product quality at a lower production cost, process modeling gradually became a necessity rather than a research and development tool in the production environment.

Although FEM programs were initially developed for metalworking processes, it was soon realized that metalworking is just one of the many operations before the part is finally installed. Prior to forging, the billet is made by primary forming processes, such as cogging or bar rolling from a cast ingot. After forging, the part is heat treated, rough machined, and finish machined. The microstructure of the part continuously evolves together with the shape. The residual stress within the part and the associated distortion are also changing with time. To really understand product behavior during the service, it is essential to connect all the missing links, not only the metalworking.

In the mid-1990s, a small business innovative research program was awarded by the U.S. Air Force and the U.S. Navy to develop a capability for heat treatment and machining (Ref 67). To track the residual-stress distribution, elastoplastic and elastoviscoplastic formulations were used.

Microstructural evolution, including phase transformation and grain-size evolution, was implemented. Distortion during heat treatment and material removal during machining processes can thus be predicted.

During the 1990s, most efforts were focused on the development of the FEM for computer-aided engineering applications. However, the engineer's experience still plays a major role in achieving a solution to either solving a production problem or reaching a better process design. The FEM solution-convergence speed depends highly on the engineer's experience, and the interpretation of the results requires complete understanding of the process. As the computing power continues to improve, optimization using systematic search becomes more and more attractive (Ref 68).

Methodologies

To account for the complicated thermomechanical responses to the manufacturing process, four FEM modules, as shown in Fig. 15, are loosely coupled. They are the deformation model, the heat-transfer model, the microstructural model, and, in the case of steel, a carbon-diffusion model.

Deformation Model. For metalworking applications, the formulation must take into account the large plastic deformation, incompressibility, material-tool contact, and (when necessary) temperature coupling. To avoid deformation locking under material incompressibility, the penalty method and selective integration method are usually used for the 2-D quadrilateral element and 3-D brick element, while the mixed formulation for the 3-D tetrahedral element is employed. It is generally
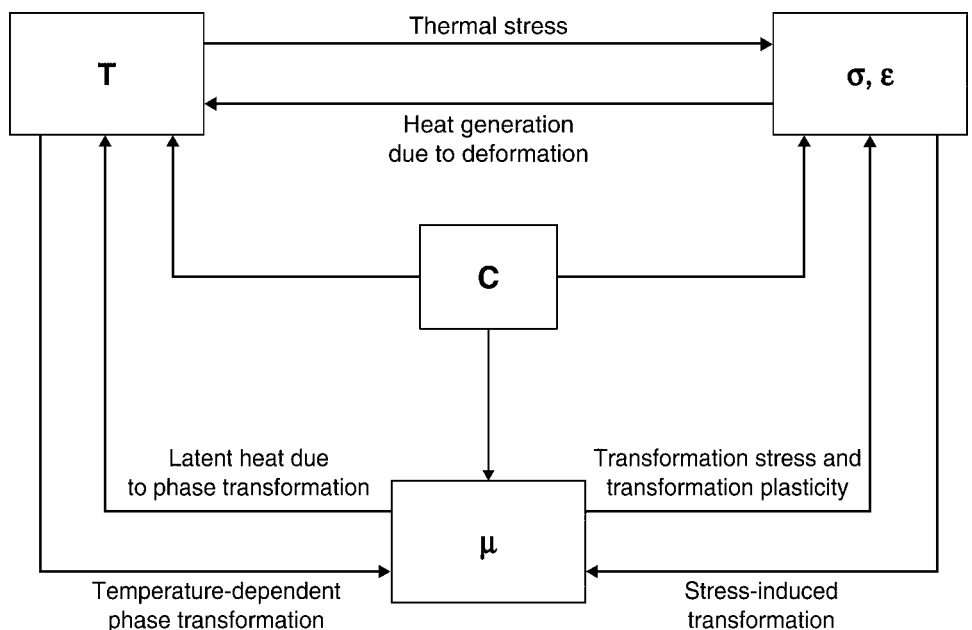
Fig. 15 View of the various coupled phenomena within metalforming

agreed that the quadrilateral element and brick element are preferred in FEM applications. Due to the difficulty in both remeshing and (frequently) the initial meshing with a brick mesh in most forming applications, a tetrahedral mesh is generally used.

Due to its simplicity and fast convergence, the rigid-plastic and rigid-viscoplastic formulations are used primarily for processes when residual stress is negligible. The elastoplastic and elastoviscoplastic formulations are important for calculating residual stress, such as in heat treatment and machining applications. However, it is very difficult to accurately characterize residual-stress evolution for forming at an elevated temperature, especially when there is significant microstructural changes, including phase transformation, precipitation, recrystallization, texture changes, and so on.

Because metal-forming processes are transient, the updated Lagrangian method has been the primary FEM method for metal-forming applications. Using this method for certain steady-state processes such as extrusion, shape rolling, and rotary tube piercing, however, may not be computationally efficient. In these special applications, the arbitrary Lagrangian Eulerian method recently has been used with great success.

Heat-Transfer Model. The heat-transfer model solves the energy balance equation. The three major modes of heat transfer are conduction, convection, and radiation. Conduction is the transfer of heat through a solid material or from one material to another by direct contact. Generally speaking, below 540 °C (1000 °F), convection has a much more pronounced effect than radiation. Above 1090 °C (2000 °F), however, radiation becomes the dominant mode of heat transfer, and convection can essentially be considered a second-order effect. Between these temperatures, both convection and radiation play an important role.

In order to predict the temperature evolution accurately during metalworking processes, several important thermal-boundary conditions must be considered:

- Radiation heat with view factor to the surrounding environment
- Convection heat to/from the surrounding environment, including the tool contact, free air, fan cool, water or oil quench
- Friction heat between two contacting bodies. It is also noted that friction heating is the primary heat source in the friction-stir welding process.
- Deformation, latent heat, and eddy current are the primary volume heat sources. Deformation heat is important for large, localized deformation and fast processes, because the adiabatic heat will increase the local temperature quickly, and material is likely to behave differently at elevated temperatures. It plays an important role in metalworking, inertial welding, translational friction welding, and the cutting process. The latent heat comes from the phase transformation or

phase change, and eddy-current heat is generated by electromagnetic fields.

Microstructural Model. Grain size is an important microstructural feature that affects mechanical properties. For example, a fine grain size is desirable to resist crack initiation, while a larger grain size is preferred for creep resistance. To obtain optimal mechanical properties, precise control of the grain size is crucial. In order to achieve a desirable microstructural distribution, as-cast materials usually undergo multiple stages of forming, such as billet conversion and closed-die forging, and multiple heat treatment steps, such as solution heat treating and aging.

During thermomechanical processing, a dislocation substructure is developed as deformation is imposed. The stored energy can provide the driving force for various restorative processes, such as dynamic recovery or recrystallization. On the completion of recrystallization, the energy can be further reduced by grain growth, in which grain-boundary area is reduced. The kinetics of recrystallization and grain-growth processes are complex. In order to predict the grain-size distribution in finished components, a basic understanding of the evolution of microstructural evolution during complex manufacturing sequences, including the primary working processes (ingot breakdown, rolling, or extrusion), final forging, and heat treatment, must be obtained. Hence, the development of microstructural evolution models has received considerable attention in recent years. Recrystallization behavior can be classified into three broad categories: static, metadynamic, and dynamic recrystallization (Ref 69).

The description of each recrystallization mode as well as static grain growth is well documented (Ref 70). Sellars' model has been used for static and metadynamic recrystallization, and the Yamada model has been used for dynamic recrystallization. Microstructural evolution in superalloys is complicated by the precipitation of γ', γ'', and δ phases (Ref 67). However, the present phenomenological approach neglects the specific effect that such phases have on the mechanisms of microstructural evolution.

Phase transformation is also another important aspect for material modeling (Ref 71). It is not only critical to achieve desirable mechanical properties but also to better understand the residual stress and the associated distortion. Phase transformation can be classified into two categories: diffusional and martensitic. Using carbon steel as an example, the austenite-ferrite and austenite-pearlite structure transformations are governed by diffusional-type transformations. The transformation is driven by a diffusion process depending on the temperature, stress history, and carbon content and is often represented by the Johnson-Mehl equation:

$$\Phi = 1 - \exp(-bt^n)$$

where Φ is the fraction transformed as a function of time, t, and b and n are material

coefficients. The diffusionless transformation from austenite to martensite usually depends on temperature, stress, and carbon content.

Defect Prediction

Simulation of metal flow in dies has advanced to the stage where it can anticipate the formation of defects, such as the fold defect shown in Fig. 16. Another example is the prediction of the suck-in defect found in extrusion (Ref 73). This defect occurs when oxide on the face of the billet is transferred by plastic deformation into the interior of the extruded part.
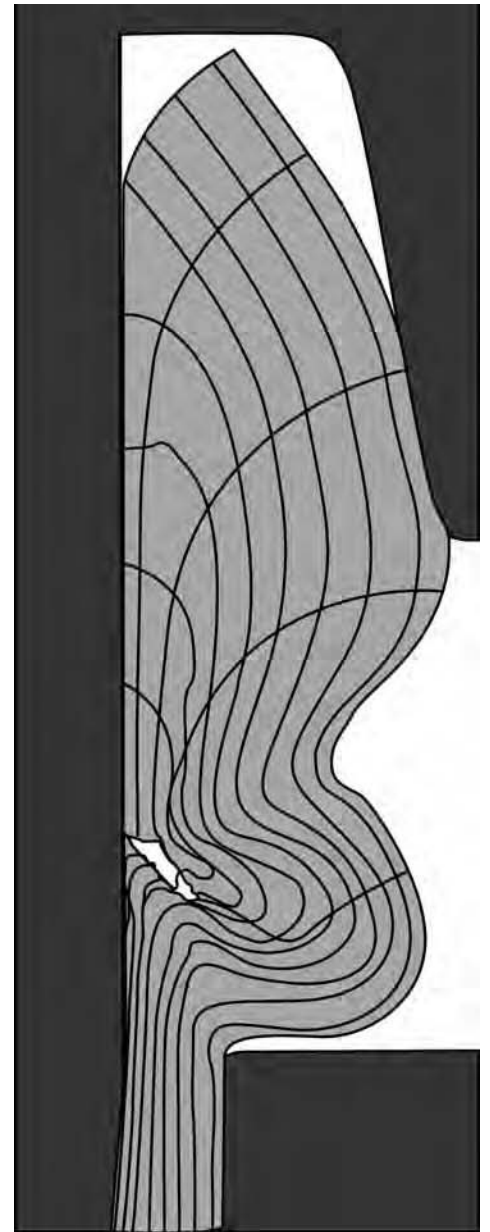
Fig. 16 Example of use of finite-element modeling software for predicting the formation of a fold defect. Source: Ref 72
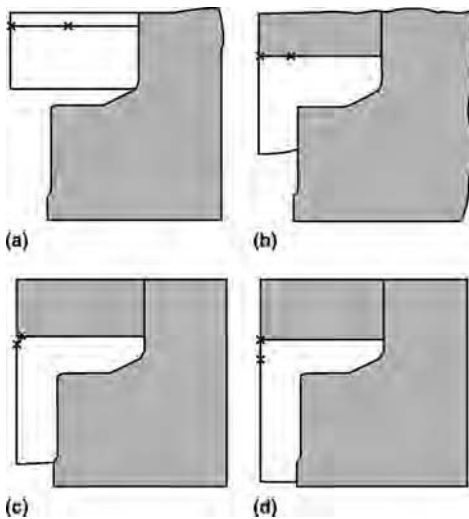
Fig. 17 Use of finite-element modeling software to predict the formation of a suck-in defect in extrusion. Source: Ref 74

Note in Fig. 17(a) the two points "x" and "x" on the top face of the billet. As extrusion proceeds, these points move closer together, and at (d) they are "sucked-in" to both lie on the centerline of the extrusion.

Predicting Onset of Ductile Fracture. An important goal of FEM software for bulk deformation processing is to predict the onset of ductile fracture. The DEFORM software incorporates a subroutine for calculating the likelihood of fracture using the Cockcroft-Latham criteria. The damage value C in Eq 10 was determined from the true stress-true strain curve measured in both compression and with a notched tension test, and then the criterion was used successfully to predict failure in multipass cold extrusion and cold compression of a cylinder with a midheight collar (Ref 75). Other studies (Ref 76) have shown that the Cockcroft-Latham criteria and the Oyane criteria (Ref 77) are equally capable of predicting the site of fracture initiation and the critical value of damage. While the Cockcroft-Latham criteria has the advantage of mathematical simplicity, it is sometimes criticized as not having a physical basis and of not being sensitive to the hydrostatic state of stress. The Oyane criteria, while being more complex, is based on a void growth model of failure and contains a term for σ_m.

Forging

Commercially available FEM software tools are used extensively by major forging suppliers because of their accuracy, speed, versatility, and ability to dramatically reduce forging process and die development cost. The FEM software tools are widely used to accurately determine press loads and die stresses, to simulate metal flow in dies, and to a lesser extent, to determine workability limits (Ref 74). Advanced applications are concerned with
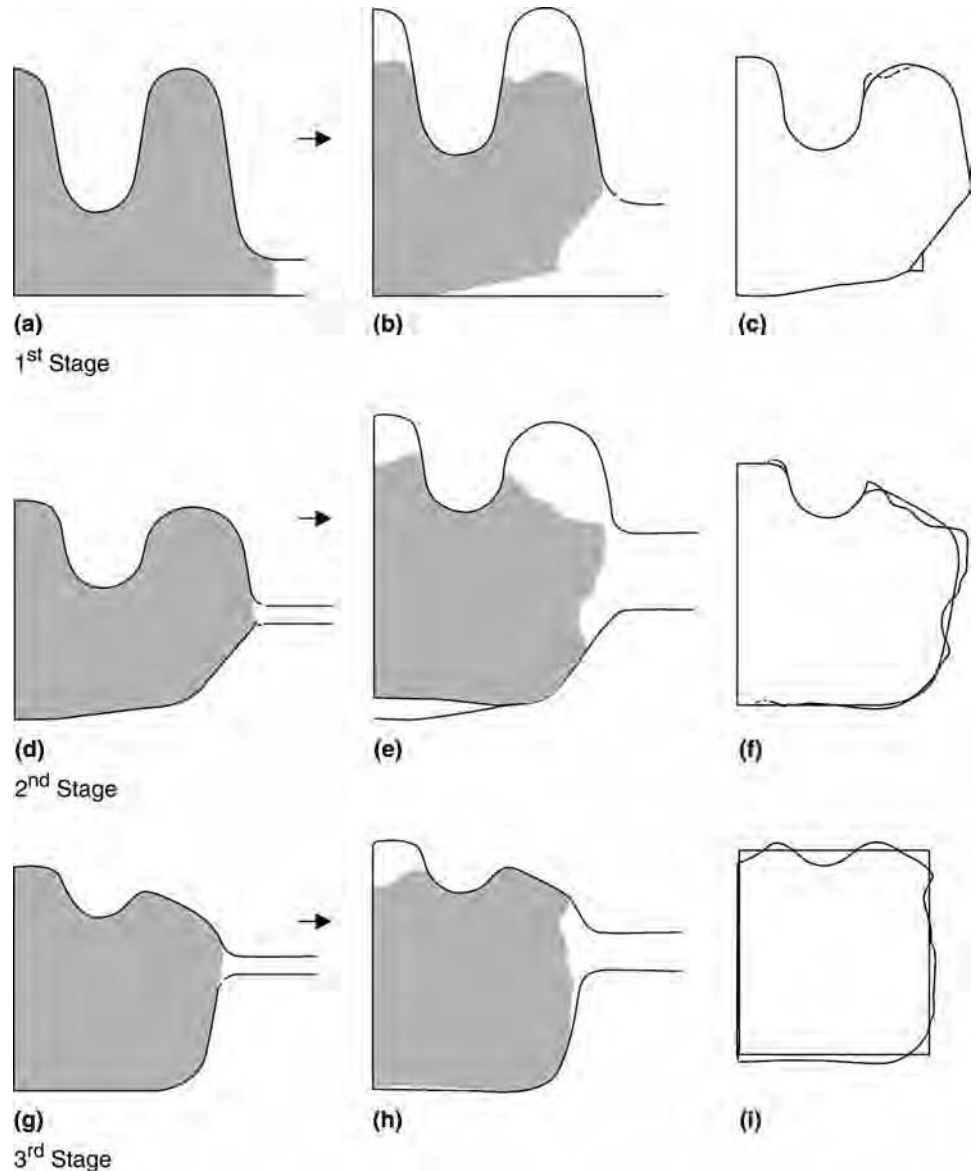
Fig. 18 Example of the use of the TEUBA software for determining forging preforms by reverse simulation. Source: Ref 72

prediction of microstructure and properties and determination of elastic recovery and residual stresses.

The dies for a hot-die or isothermally forged part can be designed today (2009) using FEA for a fraction of the cost and cycle time it took just a decade ago using trial-and-error or empirically based design rules. Finite-element analysis tools enable a designer to perform die stress analysis to ensure the forge process will not plastically deform the dies, which is especially important when considering the cost of die materials for these processes. In addition, these analytical tools enable the designer to optimize the forging shape and process to get the nearest net part shape possible without introducing forge defects such as cracks, lack of fill, or laps by ensuring the forge temperature and strain

rates stay within the optimal forge process window.

Preform Design. An important area in simulation is to design the series of intermediate shapes (preforms) to go from an initial forging billet to the final shape (Ref 72). Figure 18 shows an example of the TEUBA (Tetrahedral Element Upper Bound Analysis) simulation software, a fast preform design package (Ref 78). This provides a reverse simulation, in which the final forging is shown at (a), and the dies are forced apart, creating a series of possible intermediate shapes. As shown in Fig. 18, the forging blank (i) would be placed in the first die impression (f) that would produce the shape (d). This would be transferred to the next die (c) that would produce the final part (a).

Forging Die Design with FEM. Figure 19 shows the effective strain predictions for an isothermally forged IN 100 turbine disk. Shown are the effective strains by location once the die is completely filled. Strain and strain(−) rate predictions are useful for predicting and controlling microstructural response (Ref 79, 80) and possible defect orientations and locations. Figure 20 shows predictions of strain rate for the same part at a time step of peak strain rate, which are useful in determining whether or not the strain rates fall within the optimal process window. Successful application of FEA tools typically requires knowledge of the thermophysical properties of the die and material, flow stress as a function of temperature and strain rate, target process parameters, and friction coefficient between the part and die.

Based on successful application of these analytical tools to forging process design and development, their capability has been broadened to include microstructural evolution during the forge process (Ref 81), heat transfer during subsequent heat treatment operations, the evolution of microstructure, mechanical properties, residual stress, and distortion due to heat treatment or subsequent machining. Figures 21 and 22 show examples of γ' size and yield-strength predictions following a post-forge heat treat operation for the same IN 100 turbine disk shown in the previous figures. The analytical tool used for this work enables the user to integrate regression-based or first-principle-based equations for prediction of microstructure, mechanical properties, or creep relaxation for residual-stress prediction.

Shape Rolling

The analysis of 3-D metal flow, heat transfer, and stress and strain distributions during shape rolling is done by FEM software using the rigid-viscoplastic FEM. In modeling of complex 3-D shape rolling, the workpiece is divided into elements having simple 3-D shapes (eight-node hexahedral elements have been used most commonly for rolling problems). Most FEMs for multipass shape rolling adopt a steady-state approach (also referred to as the Eulerian approach). In the steady-state approach, stream lines and flow stress distributions are iteratively updated until numerical convergence is attained. The computation time using the steady-state approach is significantly shorter than that using the non-steady-state approach.

The rigid-plastic FEM is formulated on the basis of plasticity theory. More information about various finite-element formulations for modeling rolling is available in published literature (Ref 82–85). Figure 23 shows the structure of a finite-element program, ROLPAS-M, developed for shape rolling (Ref 86). The program, in addition to modeling metal flow, also models heat transfer in the roll bite as well as in the interstand region. Figure 24 shows the shape of the workpiece as predicted by ROLPAS-M in the roll bites of an eight-pass square to round bar rolling sequence. Such FEMs are capable of predicting the metal flow, rolling loads, and torques as well as stresses, strains, and temperatures at the nodal locations in the workpiece fairly accurately. Finite-element analysis has been used widely not only for analyzing new and existing roll pass designs but also for optimizing them (Ref 87, 88). A study (Ref 89) used FEA in combination with Taguchi Analysis to develop roll pass design guidelines for optimizing robustness of the rolling process (in other words, to make it least sensitive to process variation). More recently, FEA has been used for the simulation of microstructural evolution during hot rolling with good success (Ref 86, 90–95).

Modeling of Microstructure Evolution in Hot Rolling. Traditionally, rolling process designers have been primarily concerned with ensuring correct metal flow during the rolling process. In doing so, they normally made use of their experience as well as empirical and analytical guidelines that had been established over the years. Recently, there has been a growing emphasis on rolled product structure and properties. New grades of microalloyed steel that have been developed for cold and warm forging applications call for precise control over the product microstructure. Control over microstructure requires a good understanding of the effect of rolling mill variables on the resulting microstructure. Variables such as preheating time and temperature, rolling deformation, deformation rate, interstand cooling, and postrolling controlled cooling affect the grain size distribution, recrystallization, and phase-transformation kinetics that ultimately determine the final microstructure and mechanical properties in the rolled product. Microstructural changes occurring at different stages in the rolling process affect the final microstructure and properties of the rolled product. A typical thermomechanical cycle and microstructural evolution during the rolling process is shown in Fig. 25.

Controlled rolling to obtain desired microstructure and properties is generally known as thermomechanical control processing (TMCP). In recent years, TMCP has been effectively used in the hot rolling of plates. More recently, this technique has been applied for the hot rolling of bars, rods, and shapes using an integrated approach (Ref 86, 95).

In TMCP, history of temperature, strain, and strain rates at various locations in the workpiece obtained using FEA or other analytical methods are used in conjunction with microstructure evolution models to model microstructural changes during rolling—specifically,
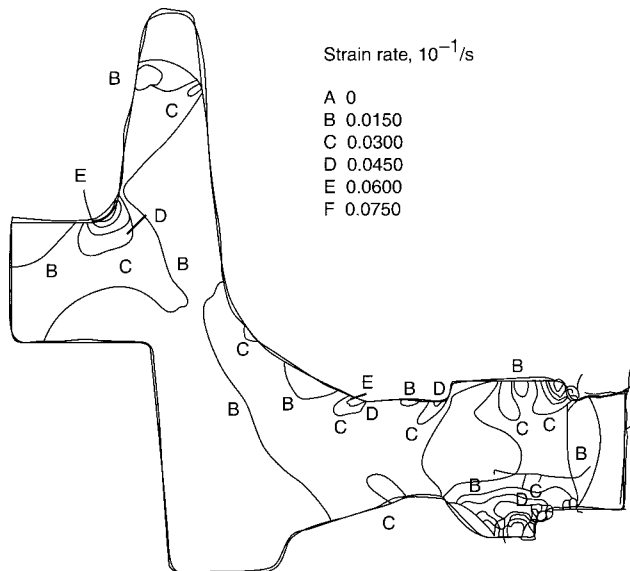
Fig. 19 Contour map of predicted effective strain for an isothermally forged turbine disk
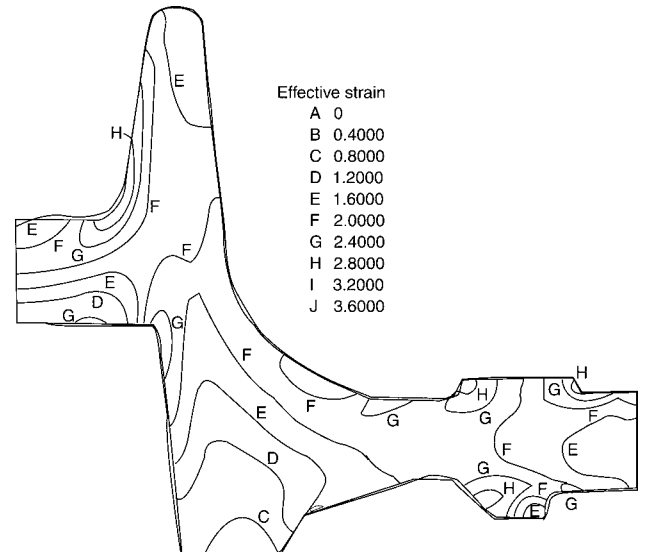
Fig. 20 Contour map of predicted strain rate for an isothermally forged turbine disk at a time point where strain rates peak at critical locations
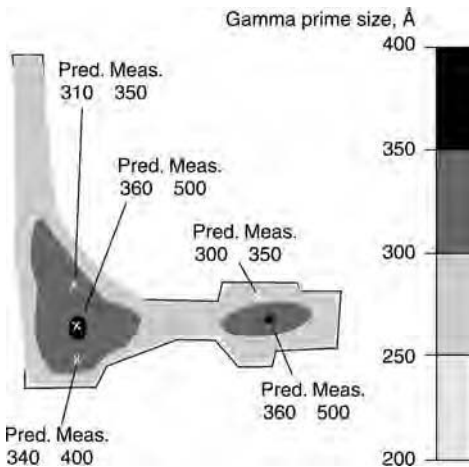
Fig. 21 Prediction of gamma prime size (angstroms) based on cooling rate from postforge solution heat treat temperature. Predicted size versus measured size
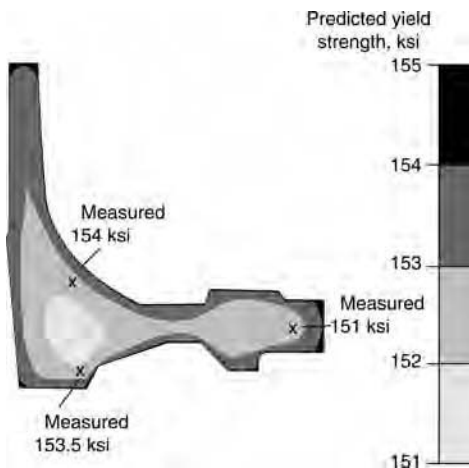
Fig. 22 Prediction of yield strength (ksi) based on predicted gamma prime size

static and dynamic recrystallization, grain growth, and phase transformation. Because of the complexity of the physical model as well as the large number of computations involved, such analyses are typically carried out using specially developed computer programs. One such program, ROLPAS-M (Ref 86), uses an approach illustrated in Fig. 26. The program consists of three main modules:

- The central feature of the integrated system is a 3-D finite-element program, ROLPAS, for simulating multipass shape rolling. The nonisothermal deformation analysis in ROL-PAS is based on rigid-viscoplastic assumption of the material behavior and uses eight-node isoparametric hexahedral elements. Deformation within the roll gap is assumed to be kinematically steady.
- Next is a microstructure evolution module, MICON, which was developed and integrated into ROLPAS to enable modeling of austenite evolution. MICON uses the thermomechanical
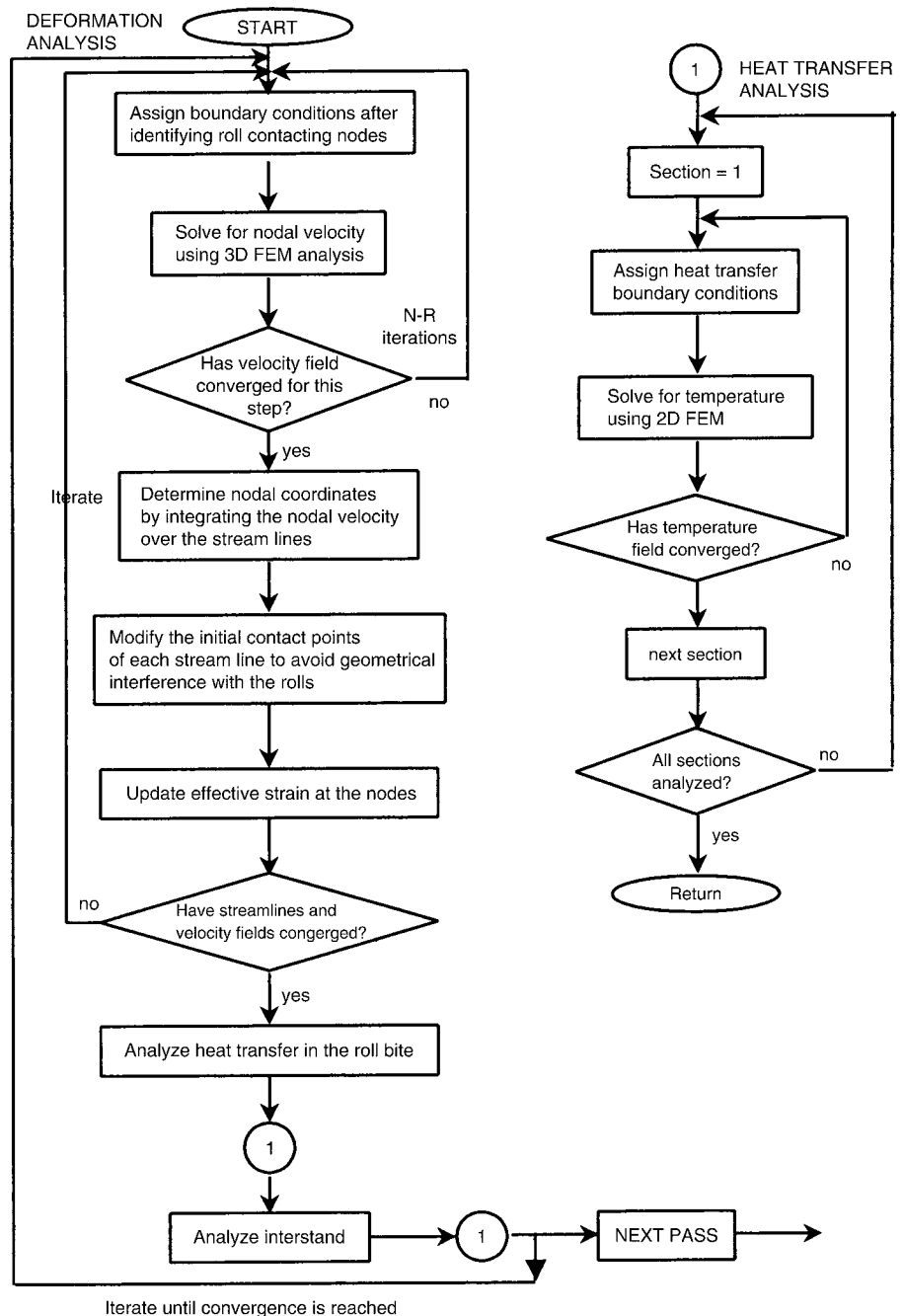
Fig. 23 Structure of program for steady-state finite-element analysis of hot rolling. Source: Ref 86

history computed by the FEM model in conjunction with microstructure evolution models to model the evolution of austenite during hot rolling. The microstructural changes occurring in bar rolling are primarily due to static recrystallization and grain growth that occur in the interstand region. In cases where accumulated strain is large enough to nucleate dynamic recrystallization, metadynamic recrystallization is modeled in the interstand region following the approach used by Sellars (Ref 93, 94). The microstructural changes in the interstand are used in

computing the retained strain and the grain sizes of the recrystallized and unrecrystallized fractions and used to compute the flow stress during the next pass.

- Last in the system is a module for modeling phase transformation, called AUSTRANS. It uses the temperature history after rolling (computed by ROLPAS) and isothermal transformation curves to model the transformation of austenite to ferrite, pearlite, bainite, and martensite. This model also uses structure-property relationships to predict the mechanical properties of the rolled product.

Pass 1 (V) Pass 2 (H)
Pass 3 (V) Pass 4 (H)
Pass 5 (V) Pass 6 (V)
Pass 7 (H) Pass 8 (V)
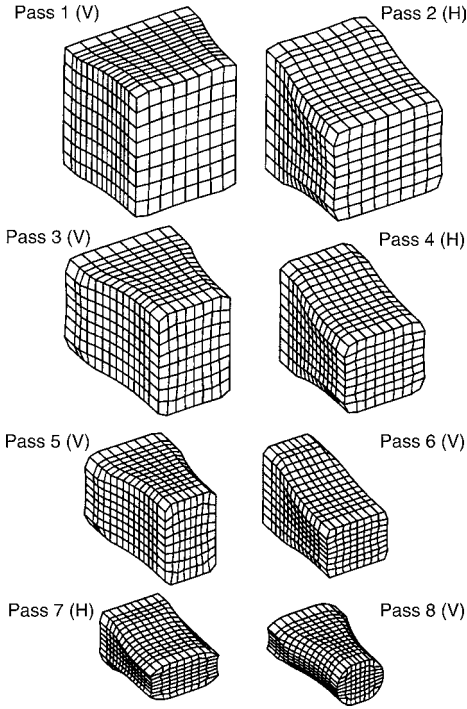
Fig. 24 Finite-element analysis of an eight-pass square to round bar rolling pass. Source: Ref 86
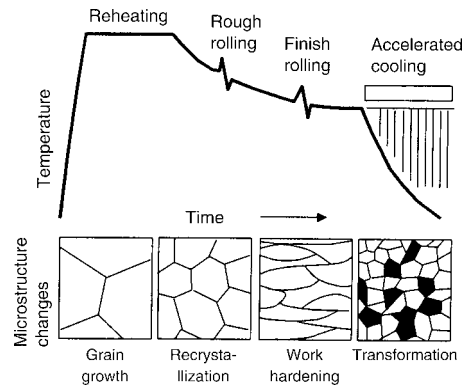
Fig. 25 Time-temperature profile and corresponding microstructural evolution in thermomechanical controlled rolling. Adapted from Ref 91

A recent study (Ref 96) has reported work on development of a system that combines a physical metallurgy model with an artificial neural network to determine microstructure and mechanical properties of hot-rolled steel in real time. Such advances are expected to yield substantial improvements in the quality and productivity of rolled products.

ACKNOWLEDGMENTS

Rolpas
Deformation and heat transfer analysis
Computes thermomechanical history

Preheating module
Ferrite grain growth
Austenite formation
Austenite grain growth

Micon
Austenite evolution
Static Rex.
Metadynamic Rex.
Grain refinement
Grain growth

Austrans
Phase transformation
Transformation to ferrite, pearlite, bainite, martensite. Estimates mechanical properties

Fig. 26 Framework of the integrated system ROLPAS-M for the modeling of microstructural evolution in hot rolling. Source: Ref 86

REFERENCES

1. C.M. Sellars and W.J. McG. Tegart, *Mem. Sci. Rev. Metall.*, Vol 63, 1966, p 731
2. A. Laasroui and J.J. Jonas, Prediction of Steel Flow Stresses at High Temperatures and Strain Rates, *Metall. Trans. A,* Vol 22, 1991, p 1545–1558
3. S.B. Brown, K.H. Kim, and L.A. Anand, An Internal Variable Constitutive Model for Hot Working of Metals, *Int. J. Plast.*, Vol 5, 1989, p 95–130
4. J.R. Rice, Foundations of Solid Mechanics, *Mechanics and Materials: Fundamentals and Linkages,* M.A. Meyers, R.W. Armstrong, and H. Kichner, Ed., John Wiley & Sons, 1999
5. M.G. Cockcroft and D.J. Latham, Ductility and Workability of Metals, *J. Inst. Met.,* Vol 96, 1968, p 33–39
6. M. Oyane, Criteria of Ductile Fracture Strain, *Bull Jpn. Soc. Mech. Eng.,* Vol 15, 1972, p 1507–1513
7. A.M. Freudenthal, *The Inelastic Behaviour of Solids,* John Wiley & Sons, 1950
8. D. Zhao, J.P. Bandstra, and H.A. Kuhn, A New Fracture Criterion for Fracture Prediction in Metalworking Processes, *Concurrent Engineering Approach to Materials Processing,* S.N. Dwivedi, A.J. Paul, and F.R Dax, Ed., TMS, 1992, p 107–119
9. J.P. Bandstra, "3D Extension of Kuhn Surface Fracture Criterion," CTC Memorandum, CTC/JPB-M0469-95, Concurrent Technologies Corporation, 1995
10. A. Chandra, Analyses of Metal Forming Problems by the Boundary Element Method, *Int. J. Solids Struct.,* Vol 31 (No. 12–13), June–July 1994, p 1695–1736
11. R.H. Wagoner and J.-L. Chenot, *Metal Forming Analysis,* Cambridge University Press, 2001
12. G.W. Rowe, C.E.N. Sturgess, P. Hartley, and I. Pillinger, *Finite-Element Plasticity and Metalforming Analysis,* Cambridge University Press, 1991
13. S. Kobayashi, S.I. Oh, and T. Altan, *Metal Forming and the Finite Element Method,* Oxford University Press, New York, 1989, p 222–243
14. N. Brannberg and J. Mackerle, Finite Element Methods and Material Processing Technology, *Eng. Comput. (WALES),*Vol 11 (No. 5), Oct 1994, p 413–455
15. K.-J. Bathe, Some Issues for Reliable Finite Element Analysis, *Reliability Methods for Engineering Analysis: Proc. First International Conference,* K.-J. Bathe and D.R.J. Owen, Ed., Pineridge Press, Swansea, U.K., 1986, p 139
16. R.D. Cook, D.S. Malkus, and M.E. Plesha, *Concepts and Applications of Finite Element Analysis,* 3rd ed., John Wiley & Sons, 1989
17. K.H. Huebner, E.A. Thornton, and T.G. Byrom, *The Finite Element Method for Engineers,* 3rd ed., John Wiley & Sons, 1995
18. J.W. Yoon and F. Barlat, Modeling and Simulation of the Forming of Aluminum Sheet Alloys, *Metalworking: Sheet Forming*, Vol 14B, *ASM Handbook,* ASM International, 2006, p 792–826
19. K.J. Bathe, *Finite Element Procedures in Engineering Analysis,* Prentice-Hall, 1982
20. K. Chung and O. Richmond, Ideal Forming, Part I: Homogeneous Deformation with Minimum Plastic Work, *Int. J. Mech. Sci.,* Vol 34, 1992, p 575–591
21. K. Chung and O. Richmond, Ideal Forming, Part II: Sheet Forming with Optimum Deformation, *Int. J. Mech. Sci.,* Vol 34, 1992, p 617–633
22. K. Chung, J.W. Yoon, and O. Richmond, Ideal Sheet Forming with Frictional Constraints, *Int. J. Plast.,* Vol 16, 2000, p 595–610
23. F. Barlat, K. Chung, and O. Richmond, Anisotropic Plastic Potentials for Polycrystals and Application to the Design of Optimum Blank Shapes in Sheet Forming,

Metall. Mater. Trans. A, Vol 25, 1994, p 1209–1216

24. K. Chung, S.Y. Lee, F. Barlat, Y.T. Keum, and J.M. Park, Finite Element Simulation of Sheet Forming Based on a Planar Anisotropic Strain-Rate Potential, *Int. J. Plast.,* Vol 12, 1996, p 93–115

25. O. Richmond and K. Chung, Ideal Stretch Forming Processes for Minimum Weight Axisymmetric Shell Structures, *Int. J. Mech. Sci.,* Vol 42, 2000, p 2455–2468

26. O. Richmond and S. Alexandrov, Non-steady Planar Ideal Plastic Flow: General and Special Analytic Solution, *J. Mech. Phys. Solids,* Vol 48, 2000, p 1735–1759

27. K. Chung, W. Lee, T.J. Kang, and J.R. Yoon, Non-Steady Plane-Strain Ideal Forming without Elastic Dead-Zone, *Fibers Polym.,* Vol 3, 2002, p 120–127

28. S. Ahmad, B.M. Irons, and O.C. Zienkiewicz, Analysis of Thick and Thin Shell Structure by Curved Finite Elements, *Int. J. Numer. Methods Eng.,* Vol 2, 1970, p 419–451

29. E. Ramm, A Plate/Shell Element for Large Deflection and Rotations, *Formulation and Computational Algorithms in Finite Element Analysis,* K.J. Bathe, J.T. Oden, and W. Wunderlish, Ed., MIT Press, 1977

30. H. Parish, Nonlinear Analysis of Shells Using Isoparametric Elements, *ASME Conference,* T.J.R. Hughes et al., Ed., New York, 1981, p 48

31. T.J.R. Hughes and W.K. Liu, Nonlinear Finite Element Analysis of Shells. Part I: Three-Dimensional Shells, *Comput. Methods Appl. Mech. Eng.,* Vol 26, 1981, p 331–362

32. T.J.R Hughes and W.K. Liu, Nonlinear Finite Element Analysis of Shells, Part II: Two-Dimensional Shells, *Comput. Method. Appl. Mech. Eng.,* Vol 27, 1981, p 167–181

33. E.N. Dvorkin and K.J. Bathe, A Continuum Mechanics Based on Four-Node Shell Element for General Non-Linear Analysis, *Eng. Comp. (Int. J. Comput.-Aided Eng. Software),* Vol 1, 1984, p 77–88

34. W.K. Liu, E.S. Law, D. Lam, and T. Belytschko, Resultant-Stress Degenerated Shell Element, *Comput. Methods Appl. Mech. Eng.,* Vol 55, 1986, p 259–300

35. R.P.R. Cardoso and J.W. Yoon, A One Point Quadrature Shell Element for Through Thickness Deformation, *Comp. Methods Appl. Mech. Eng.,* Vol 194, 2005, p 1161–1199

36. M.E. Honnor and R.D. Wood, Finite Element Analysis of Axisymmetric Deep Drawing Using a Simple Two-Nodded Mindlin Shell Element, *Numerical Methods for Nonlinear Problems,* C. Taylor et al., Ed., Pineridge Press, 1987, p 440

37. A. Honecker and K. Mattiason, Finite Element Procedures for 3-D Sheet Forming Simulation, *Proc. of the NUMIFORM '89 Conference*, A.A. Balkema, Rotterdam, 1989, p 457–463

38. E. Massoni and J.L. Chenot, 3-D Finite Element Simulation of Deep-Drawing Process, *Proc. NUMIFORM '92 Conference,* A.A. Balkema, Rotterdam, 1992, p 503–507

39. M. Kawka and A. Makinouchi, Finite Element Simulation of Sheet Metal Forming Processes by Simultaneous Use of Membrane, Shell and Solid Elements, *Proc. NUMIFORM '92 Conference,* A.A. Balkema, Rotterdam, 1992, p 491–495

40. J.W. Yoon, D.Y. Yang, and K. Chung, Elasto-Plastic Finite Element Method Based on Incremental Deformation Theory and Continuum Based Shell Elements for Planar Anisotropic Sheet Materials, *Comput. Methods Appl. Mech. Eng.,* Vol 174, 1999, p 23–56

41. R. Hill, Stability of Rigid-Plastic Solids, *J. Mech. Phys. Solids,* Vol 6, 1957, p 1–8

42. R. Hill, External Paths of Plastic Work and Deformation, *J. Mech. Phys. Solids,* Vol 34, 1986, p 511–523

43. A. Nadai, *Theory of Flow and Fracture of Solids*, Vol 2, McGraw-Hill, 1963, p 96

44. K. Chung and O. Richmond, A Deformation Theory of Plasticity Based on Minimum Work Paths, *Int. J. Plast.,* Vol 9, 1993, p 907–920

45. D.Y. Yang and Y.J. Kim, A Rigid-Plastic Finite Element Calculation for the Analysis of General Deformation of Plastic Anisotropic Sheet Metals and Its Application, *Int. J. Mech. Sci.,* Vol 28, 1986, p 825

46. Y. Germain, K. Chung, and R.H. Wagoner, A Rigid-Viscoplastic Finite Element Program for Sheet Metal Forming Analysis, *Int. J. Mech. Sci.,* Vol 31, 1989, p 1–24

47. K. Chung and O. Richmond, Ideal Forming, Part I: Homogeneous Deformation with Minimum Plastic Work, *Int. J. Mech. Sci.,* Vol 34, 1992, p 575–591

48. K. Chung and O. Richmond, Ideal Forming, Part II: Sheet Forming with Optimum Deformation, *Int. J. Mech. Sci.,* Vol 34, 1992, p 617–633

49. J.W. Yoon, I.S. Song, D.Y. Yang, K. Chung, and F. Barlat, Finite Element Method for Sheet Forming Based on an Anisotropic Strain-Rate Potential and the Convected Coordinate System, *Int. J. Mech. Sci.,* Vol 37, 1995, p 733–752

50. H.J. Braudel, M. Abouaf, and J.L. Chenot, An Implicit and Incremental Formulation for the Solution of Elasto-Plastic Problems by the Finite Element Method, *Comput. Struct.,* Vol 22, 1986, p 801–814

51. J.W. Yoon, D.Y. Yang, K. Chung, and F. Barlat, A General Elasto-Plastic Finite Element Formulation Based on Incremental Deformation Theory for Planar Anisotropy and Its Application to Sheet Metal Forming, *Int. J. Plast.,* Vol 15, 1996, p 35–67

52. M.L. Wilkins, Calculation of Elasto-Plastic Flow, *Methods of Computational Physics,* B. Alder et al., Ed., Academic Press, 1964, p 3

53. M. Ortiz and P.M. Pinsky, "Global Analysis Methods for the Solution of Elastoplastic and Viscoplastic Dynamic Problems,"

Report UCB/SESM 81/08, Dept. Civil Eng., Univ. of California, Berkley, 1981

54. M. Ortiz, P.M. Pinsky, and R.L. Taylor, Operator Split Methods for the Numerical Solution of the Elastoplastic Dynamic Problem, *Comput. Methods Appl. Mech. Eng.,* Vol 39, 1983, p 137–157

55. M. Ortiz and J.C. Simo, An Analysis of a New Class of Integration Algorithm for Elastoplastic Relations, *Int. J. Numer. Methods Eng.,* Vol 23, 1986, p 353–366

56. M. Dao and R.J. Asaro, Localized Deformation Modes and Non-Schmid Effects in Crystalline Solids, Part I. Critical Conditions of Localization, *Mech. Mater.,* Vol 23, 1996, p 71–102

57. J.W. Yoon, F. Barlat, J.J. Gracio, and E. Rauch, Anisotropic Strain Hardening Behavior in Simple Shear for Cube Textured Aluminum Alloy Sheets, *Int. J. Plast.,* Vol 21, 2005, p 2426–2447

58. S.-I. Oh, J. Walters, and W.-T. Wu, Finite Element Method Applications in Bulk Forming, *Metalworking: Bulk Forming,* Vol 14A, *ASM Handbook,* ASM International, 2005, p 617–639

59. C.H. Lee and S. Kobayashi, New Solutions to Rigid-Plastic Deformation Problem Using a Matrix Method, *J. Eng. Ind. (Trans. ASME),* Vol 95, 1973, p 865

60. G.D. Lahoti and T. Altan, "Research to Develop Process Models for Producing a Dual Property Titanium Alloy Compressor Disk," Interim Annual Report, AFML-TR-79-4156, Battelle Columbus Laboratories, Dec 1979

61. S.I. Oh, Finite Element Analysis of Metal Forming Problems with Arbitrarily Shaped Dies, *Int. J. Mech. Sci.,* Vol 24, p 479

62. N. Rebelo and S. Kobayashi, A Coupled Analysis of Viscoplastic Deformation and Heat Transfer, Parts I and II, *Int. J. Mech. Sci.,* Vol 22, 1980, p 699, 707

63. W.T. Wu, S.I. Oh, T. Altan, and R. Miller, "Automated Mesh Generation for Forming Simulation—I," ASME International Computers in Engineering Conference, Aug 5–9, 1990

64. J.J. Park and S.I. Oh, "Three Dimensional Finite Element Analysis of Metal Forming Processes," NAMRC XV, May 27–29, 1987 (Bethlehem, PA)

65. G. Li, W.T. Wu, and J.P. Tang, "DEFORM-3D—A General Purpose 3-D Finite Element Code for the Analysis of Metal Forming Processes," presented at the Metal Forming Process Simulation in Industry International Conference and Workshop, Sept 28–30, 1994 (Baden-Baden, Germany)

66. W.T. Wu, S.I. Oh, T. Altan, and R.A. Miller, Automated Mesh Generation for Forming Simulation, *Proc. ASME Int. Comput. Eng.,* Vol 1, 1991, p 507

67. W.T. Wu, G. Li, J.P. Tang, S. Srivatsa, R. Shankar, R.Wallis, P. Ramasundaram, and J. Gayda, "A Process Modeling System for Heat Treatment of High Temperature

Structural Materials," AFRL-ML-WP-TR-2001-4105, June 2001

68. J. Oh, J. Yang, W.T. Wu, and H. Delgado, "Finite Element Method Applied to 2-D and 3-D Forging Design Optimization," NumiForm 2004, June 13–17, 2004 (Columbus, OH)

69. C.M. Sellars and J.A. Whiteman, Recrystallization and Grain Growth in Hot Rolling, *Met. Sci.,* Vol 13, 1979, p 187–194

70. G. Shen, S.L. Semiatin, and R. Shivpuri, Modeling Microstructure Development during the Forging of Waspaloy, *Metall. Mater. Trans. A,* Vol 26, 1995, p 1795–1803

71. K. Arimoto, G. Li, A. Arvind, and W.T. Wu, "The Modeling of Heat Treating Process," ASM 18th Heat Treating Conference, Oct 12–15, 1998 (Chicago, IL), ASM International

72. A.N. Bramley and D.J. Mynors, The Use of Forging Simulation Tools, *Mater. Des.,* Vol 21, 2000, p 279–286

73. S.I. Oh, W.Y. Wu, J.P. Tang, and A. Vedhanayagam, Capabilities and Application of FEM Code DEFORM, *J. Mater. Process. Technol.,* Vol 24, 1990, p 25

74. T. Altan and V. Vazquez, Numerical Process Simulation for Tool and Process Design in Bulk Metal Forming, *Annals CIRP,* Vol 45/2, 1996, p 599–615

75. H. Kim, M. Yamanaka, and T. Altan, Predictions and Elimination of Ductile Fracture in Cold Forgings Using FEM Simulations, *Proc. of NAMRC,* Society of Manufacturing Engineers, 1995, p 63

76. B.P.P.A. Gouveia, J.M.C. Rodrigues, and P.A.F. Martins, Fracture Predicting in Bulk Metal Forming, *Int. J. Mech. Sci.,* Vol 38, 1996, p 361–372

77. M. Oyane, Criteria of Ductile Fracture Strain, *Bull. JSME,* Vol 15, 1972, p 1507

78. C.C. Chang, J.C.A. Tildesley, F. Bonnavand, D.J. Mynors, and A.N. Bramley, Forging Simulation—TEUBA, *Metallurgia,* 1999, p 27–28

79. E. Huron, S. Srivatsa, and E. Raymond, Control of Grain Size Via Forging Strain Rate Limits for R′88DT, *Superalloys 2000,* T.M. Pollock et al., Ed., TMS-AIME, 2000, p 49–58

80. M. Soucail, M. Marty, and H. Octor, The Effect of High Temperature Deformation on Grain Growth in a P/M Nickel Base Superalloy, *Superalloys 1996,* R.D. Kissinger et al., Ed., TMS-AIME, 1996, p 663–666

81. D. Huang, W.T. Wu, D. Lambert, and S.L. Semiatin, Computer Simulation of Microstructure Evolution During Hot Forging of Waspaloy and Nickel Alloy 718, *Microstructural Modeling and Prediction During Thermomechanical Processing,* R. Srinivasan and S.L. Semiatin, Ed., Nov 4–8, 2001 (Indianapolis, IN), TMS, 2002, p 137–146

82. K. Mori and K. Osakada, Finite Element Simulation of Three-Dimensional Deformation in Shape Rolling, *Int. J. Numer. Meth. Eng.,* Vol 30, 1990, p 1431–1440

83. K. Mori and K. Osakada, Simulation of Three-Dimensional Deformation in Rolling by the Finite Element Method, *Int. J. Mech. Sci.,* Vol 26, 1984, p 515–525

84. J.J Park and S.I. Oh, Application of Three Dimensional Finite Element Analysis to Shape Rolling Processes, *Trans. ASME,* Vol 112, Feb 1990, p 36–46

85. N. Kim and S. Kobayashi, Three Dimensional Analysis and Computer Simulation of Shape Rolling by the Finite and Slab Element Method, *Int. J. Mach. Tools Manuf.,* Vol 31 (No. 4), 1991, p 553–563

86. P. Pauskar, "An Integrated System for Analysis of Metal Flow and Microstructural Evolution in Hot Rolling," Ph.D. Thesis, The Ohio State University, 1998

87. R. Shivpuri and W. Shin, A Methodology for Roll Pass Optimization for Multi-Pass Shape Rolling, *Int. J. Mach. Tools Manuf.,* Vol 32 (No. 5), 1992, p 671–683

88. K. Sawamiphakdi and G.D. Lahoti, Application of the Slab-Finite-Element Method for Improvement of Rolled Bar Surface Quality, *Annals CIRP,* Vol 43, 1991, p 219–222

89. K. Yoshimura and R. Shivpuri, "Robust Design of Roll Pass for Reduced Geometric Variance in Hot Rolling of Steel Rod," Report ERC/NSM-B-95-26, ERC for NSM, The Ohio State University, NSF Engineering Research Center for Net Shape Manufacturing, 1995

90. Y. Saito and C. Shiga, Computer Simulation of Microstructural Evolution in Thermomechanical Processing of Steel Plates, *ISIJ Int.,* 1992, p 414–422

91. A. Yoshie et al., Modeling of Microstructural Evolution and Mechanical Properties of Steel Plates Produced by Thermo-Mechanical Control Process, *ISIJ Int.,* Vol 32 (No. 3), 1992, p 395–404

92. I.V. Samarasekhara, D.Q. Jin, and J.K. Brimacombe, The Application of Microstructural

Engineering to the Hot Rolling of Steel, *38th Mechanical Working and Steel Processing Conf. Proc.,* Vol 34, ISS, 1996, p 313–327

93. C.M. Sellars and J.A. Whiteman, Recrystallization and Grain Growth in Hot Rolling, *Met. Sci.,* March–April 1979, p 187–194

94. C.M. Sellars, Modeling—An Interdisciplinary Activity, *Proc. International Symposium on Mathematical Modeling of Hot Rolling of Steel,* Aug 26–29, 1990 (Hamilton, Ontario, Canada), Canadian Institute of Mining and Metallurgy, 1990, p 1–18

95. J. Yanagimoto and T. Ito, Prediction of Microstructure Evolution in Hot Rolling, *NUMIFORM '98,* June 22–25, 1998 (Enschede, Netherlands), Netherlands Institute for Metals Research, University of Twente, OSM, Korinklijke and Boal B.V., 1998, p 359–364

96. H. Loffler et al., Control of Mechanical Properties by Monitoring Microstructure, *AISE Steel Technol.,* June 2001, p 44–47

SELECTED REFERENCES

- R.D. Cook, *Concepts and Applications of Finite Element Analysis*, Wiley, New York, 1974

- R.D. Cook, D.S. Malkus, and M.E. Plesha, *Concepts and Applications of Finite Element Analysis,* 3rd ed., John Wiley &Sons, 1989

- K.H. Huebner, E.A. Thornton, and T.G. Byrom, *The Finite Element Method for Engineers,* 3rd ed., John Wiley &Sons, 1995

- T.J.R. Hughes, *The Finite Element Method. Linear Static and Dynamic Finite Element Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1987

- S. Kobayashi, S.I. Oh, and T. Altan, *Metal Forming and the Finite Element Method*, Oxford University Press, New York, 1989, p 222–243

- G.W. Rowe, C.E.N. Sturgess, P. Hartley, and I. Pillingre, *Finite-Element Plasticity and Metalforming Analysis*, Cambridge University Press, 1991

- R.H. Wagoner and J.-L. Chenot, *Metal Forming Analysis*, Cambridge University Press, 2001

- R.E. White, *An Introduction to the Finite Element Method with Application to Nonlinear Problems*, Wiley, New York, 1985

- O.C. Zienkiewicz and R.L. Taylor, *The Finite Element Method*, McGraw-Hill, New York, 1989

Computational Fluid Dynamics Modeling*

COMPUTATIONAL FLUID DYNAMICS (CFD) is a discipline of computer-aided engineering (CAE) together with computer-aided design (CAD) and computer-aided manufacturing (CAM). In this overview, CFD is described in terms of:

- Scope of CFD and its utility for engineering
- Basic technical foundation for CFD
- How CFD is incorporated into engineering product and process design

Background and History

Computational fluid dynamics has as its objective the numerical solution of fluid-flow equations. The calculus problem of solving a coupled system of nonlinear partial differential equations (PDEs) for the variables of interest (e.g., velocity, pressure, and temperature) is transformed into an algebra problem of solving a large system of simultaneous linear equations for discrete unknowns that represent the state of a thermal-fluids system; the latter is amenable to numerical solution on a digital computer.

This is a somewhat abstract description of CFD, but it is necessary to speak in general terms when introducing a subject that encompasses such a wide variety of solution techniques. This overview discusses finite-difference, finite-volume, finite-element, spectral, and some computational particle methods. The emphasis is on the first three, because these are the methods that are primarily used in contemporary CFD codes for engineering design.

In this article, the terminology "computational fluid dynamics" is reserved for computationally intensive three-dimensional simulations of thermal-fluids systems where nonlinear momentum transport plays an important role. It does not encompass all branches of numerical analysis as applied to fluid-dynamics problems. In particular, consideration of zero- or quasi-dimensional analysis of fluids systems (Ref 1, 2) and linear heat conduction or potential flow problems (Ref 3, 4) has been excluded.

The practice of CFD began with the advent of computers; indeed, the first computer was developed, in part, to solve fluid-flow equations. It was recognized by the developers of the atomic bomb at Los Alamos National Laboratory that many fluid-dynamics problems were impossible to solve by analytic means. What was needed was a machine that could perform the massive number of calculations necessary to solve the flow equations by simple finite-difference methods. The ENIAC computer began operating shortly after World War II, and its first calculations were to test various configurations for a hydrogen bomb (Ref 5). Ensuring the safety and reliability of modern nuclear weapons remains a major impetus for the development of more powerful computers and more efficient numerical techniques for solving the fluid-flow equations (Ref 6).

Initially, most numerical solutions were limited to flows that could be approximated as spatially one- or two-dimensional; the time and expense of performing three-dimensional calculations remained prohibitive. Over the last 30 years, however, CFD calculations of three-dimensional flows have become more common. This has heightened enormously the interest in CFD among engineers, because most real flows are three dimensional. In fact, most fluid-flow problems encountered in industry are so complex that the *only* method of analysis to which they are amenable is CFD. Thus, although use of CFD began only 60 years ago, it is difficult to find problems in fluid dynamics to which computer solution has not been brought to bear.

The capability to perform three-dimensional CFD has resulted primarily from the availability of faster computers with larger memories (Fig. 1) (Ref 7). The development of parallel and massively parallel computers promises to further improve the speed and extend the applicability of CFD. A recent simulation of the oceans (Ref 8, 9) serves to illustrate the problem size and computational requirements that have been realized in modern applications. This problem was run on a 512-node massively parallel computer, requiring ten gigabytes of memory (giga = billion, one byte = eight bits). It ran at a computational speed of four gigaflops ("flops" = floating point operations per second) and required 80 days of computer time. A plot of ocean surface temperature obtained in this simulation is shown in Fig. 2. Computers with maximum performance at one teraflop (tera = trillion) are now available, and petaflop computers (peta = 1000 trillion) are being planned (Ref 10).

At the same time, improved numerical methods have yielded higher computational efficiency, that is, fewer operations and/or less memory for a given accuracy. Among the most important of these advances has been the development of faster methods for solving implicit difference approximations (see the following section). A third enabler for three-dimensional CFD has been the formation of improved finite-volume and finite-element methods that better accommodate the complex geometrical boundaries that characterize engineering flows. Examples of engineering applications are given later in this article.

The advent of three-dimensional calculations has increased the engineering relevance of CFD, but many obstacles remain to be overcome before CFD realizes its full potential as an engineering design tool. Foremost among these is spatial resolution. Most flows of practical interest have features whose relevant spatial and temporal scales span many orders of magnitude. For example, in an automotive four-stroke-cycle spark-ignited internal-combustion engine operating at 2000 rev/min, hydrodynamic scales range from 0.01 mm (0.0004 in.) (the turbulence microscale) to 100 mm (4 in.) (the bore diameter); flame thickness (stoichiometric, undiluted reactants) is in the range of 0.01 to 0.10 mm (0.0004 to 0.004 in.); and spray droplets issuing from a typical port-fuel injector have diameters as small as 0.10 mm (0.004 in.) (Ref 11). Computers do not exist, and will not exist in the foreseeable future, that can store all the numbers required to fully resolve these phenomena. Thus, the effects of small-scale, unresolvable features on the large-scale, average flow features of interest are "modeled" through modifications to the governing PDEs. Examples of models include turbulence models, combustion models, and multiphase flow models. All models necessarily introduce imprecision, and an ongoing goal of research is to improve the accuracy of these models.

Other issues for three-dimensional engineering CFD include geometry acquisition and grid generation, numerical accuracy, and diagnostics

---
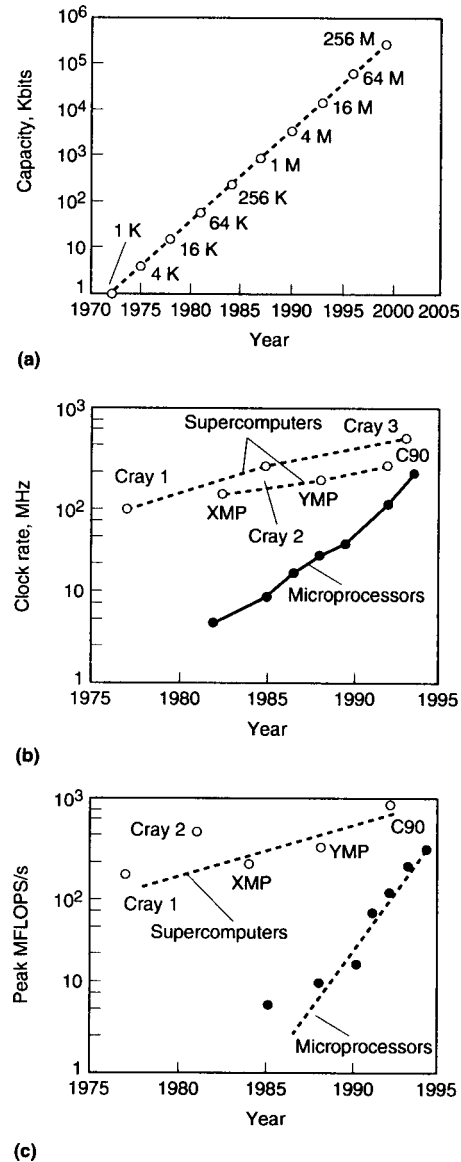
Fig. 1 Growth in computer hardware performance, 1970 to 1995. (a) Memory chip capacity doubles every 1.5 years. (b) Clock rate. (c) Peak single-process megaflops. Source: Ref 7

to extract the physical information of interest from the computations. Modeling and other issues are discussed further in subsequent sections.

Governing Equations

The governing equations of fluid dynamics and an introduction to the CFD techniques for their solution are given in this section. Also, basic terminology used by practitioners of CFD is introduced. Readers who are not interested in the technical foundation of CFD can proceed to the section "Computational Fluid Dynamics for Engineering Design" in this article.

Fig. 2 Ocean surface temperatures from a recent computational fluid dynamics simulation of the North Atlantic Ocean. Source: Ref 8

The equations of fluid dynamics can be derived from kinetic theory or continuum points of view (Ref 12–14), each of which complements the other. Kinetic theory regards the fluid as made up of molecules subject to collisions and intermolecular forces. Kinetic theory derivations are valid only for dilute gases but give detailed information about how such transport phenomena as stresses and heat fluxes arise from molecular fluctuations, which in turn are related to the average molecular properties for which the fluid equations solve. Continuum derivations regard the fluid as a continuous medium and show the applicability of the fluid equations to a much broader class of media than dilute gases but do not give detailed information about transport phenomena.

The Equations of Continuous, Compressible Media. Three basic physical principles, applicable to any continuous medium, are used in continuum derivations:

- Conservation of mass
- Newton's second law that force equals mass times acceleration
- The first law of thermodynamics, which states that total energy, in all its forms, must be conserved

These three principles lead to the following three equations of motion: the mass, or continuity, equation:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u_i}{\partial x_i} = 0 \qquad \text{(Eq 1)}$$

the momentum equation:

$$\frac{\partial \rho u_i}{\partial t} + \frac{\partial (\rho u_i u_j)}{\partial x_j} = \frac{\partial \Sigma_{ij}}{\partial x_j} + \rho F_i \qquad \text{(Eq 2)}$$

the total energy equation:

Table 1 Computational fluid dynamics nomenclature

Symbol	Definition
C_p	Specific heat at constant pressure
C_v	Specific heat at constant volume
e	Internal energy
E	Total energy
F	Body force per unit mass
h	Enthalpy
K	Turbulent kinetic energy
P	Pressure
Q	Heat flux
R	Universal gas constant
S	Rate of deformation
t	Time
T	Temperature
u	Velocity
W	Molecular weight
x	Spatial location
δ	Kronecker delta function
ε	Dissipation rate of turbulent kinetic energy
κ	Heat conductivity
λ	Second coefficient of viscosity
λ'	Bulk viscosity
μ	First coefficient of viscosity
ρ	Mass density
Σ	Stress

$$\frac{\partial \rho E}{\partial t} + \frac{\partial (\rho E u_j)}{\partial x_j} = \frac{\partial (\Sigma_{ij} u_j)}{\partial x_i} - \frac{\partial Q_j}{\partial x_j} + \rho F_i u_i \qquad \text{(Eq 3)}$$

These equations are written in Cartesian tensor notation (Ref 15), according to which the subscripts i and j take the values 1, 2, or 3 corresponding to the three Cartesian coordinate directions. A subscript that appears just once in a term takes on one of the three values 1, 2, or 3; repeated subscripts in a term denote a summation of that term over all three coordinate directions. The other notation in Eq 1 to 3 is defined in Table 1.

The total energy E is the sum of the local flow kinetic energy and its internal energy e:

$$E = e + \frac{1}{2} u_i^2 \qquad \text{(Eq 4)}$$

Alternative energy equations for e and for enthalpy $h = e + p / \rho$, where p is the pressure, can easily be derived using Eq 1 to 3. Computational fluid dynamics codes often solve internal energy or enthalpy equations, in place of Eq 3, when calculating compressible flows.

The previous equations are expressed in Eulerian form; that is, the time derivative is taken at a fixed point in space. This contrasts with the Lagrangian form, in which the time derivative is taken following a fluid element (Ref 14). Although the Eulerian form of the equations is most often used in CFD, there are CFD methods that approximate the Lagrangian equations.

When completed with constitutive relations appropriate for fluids, these are the basic equations of compressible fluid dynamics. In practice, one often encounters applications in which extensions of these equations are necessary. Among the most common are extensions to multicomponent and chemically reactive

flows (Ref 16), to magnetohydrodynamic flows (Ref 17), and to flows with radiative heat transfer (Ref 18). It is beyond the scope of this overview of CFD to give these extended equations, and the reader is referred to the aforementioned references for this information.

Constitutive Relations of Fluid Flow. To complete these equations, the stress Σ_{ij} and heat flux Q_i need to be expressed in terms of known fluid variables and their derivatives. These expressions are known as constitutive relations. A fluid is a medium for which the nonhydrostatic part of its stress depends only on its rate of deformation S_{ij}. The quantity S_{ij} is given by:

$$S_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \qquad \text{(Eq 5)}$$

Thus, a fluid has no memory of its previous configurations. This fact, together with the assumption of an isotropic medium in terms of its microstructure, allow the expression (Ref 14) of the full stress as:

$$\Sigma_{ij} = 2\mu S_{ij} + \lambda S_{kk}\delta_{ij} - p\delta_{ij} \qquad \text{(Eq 6)}$$

where δ_{ij} is the Kronecker delta function (Ref 15). Fluids with this form of stress tensor are called Newtonian fluids. The thermodynamic pressure p and the first and second coefficients of viscosity μ and λ, depend only on the local thermodynamic state of the fluid. Often, the second coefficient of viscosity in Eq 6 is replaced by the bulk viscosity λ', defined by:

$$\lambda' = \lambda + \frac{2}{3}\mu \qquad \text{(Eq 7)}$$

The heat flux Q_i depends on gradients in temperature, again assuming an isotropic fluid Q_i can be written:

$$Q_i = -\kappa\frac{\partial T}{\partial x_i} \qquad \text{(Eq 8)}$$

This is Fourier's heat conduction law, and κ is called the heat-conduction coefficient or simply the heat conductivity. Its value depends on the local thermodynamic state of the gas. When one substitutes Eq 6 for the stress tensor and Eq 8 for the heat-flux vector into Eq 1 to 3, the resulting equations are called the compressible Navier-Stokes equations.

The fluid equations are completed by the specification of the quantities p, e, μ, λ, and κ in terms of the local fluid temperature T and density ρ. The equations specifying p and e are referred to as thermal and caloric equations of state, respectively. For a so-called ideal gas, these are given by:

$$p = \rho\frac{R}{W}T \qquad \text{(Eq 9)}$$

and

$$e = \int^T C_v(T')\,dT' \qquad \text{(Eq 10)}$$

where the specific heat at constant volume C_v is a function of temperature. Alternatively, the enthalpy h is given by:

$$h = \int^T C_p(T')\,dT' \qquad \text{(Eq 11)}$$

where, from the definition of h and the thermal equation of state Eq 9, the specific heat at constant pressure C_p is given by:

$$C_p = C_v + \frac{R}{W} \qquad \text{(Eq 12)}$$

Values of C_v and C_p versus temperature are given in Ref 19 and 20.

The quantities μ, λ, and κ are called transport coefficients. How they are related to the local thermodynamic state of the fluid and its molecular properties are given in Ref 21. Given an expression for the viscosity μ, the heat conductivity κ can frequently be approximated by:

$$\kappa = \frac{\mu C_p}{Pr} \qquad \text{(Eq 13)}$$

where Pr is the Prandtl number, whose value is nearly constant and often of order unity.

Simplifications of the Fluid-Flow Equations. For certain flow situations, considerable computer time can be saved by solving simplified forms of the compressible flow equations. In this section, the steady-state, inviscid, and incompressible approximations are introduced, and the circumstances under which they may be used are described.

The steady-state approximation is obtained simply by dropping the time-derivative terms in Eq 1 to 3. While solving the steady-state equations can often save computer time, sometimes CFD-solution techniques for the steady-state equations have what are called convergence difficulties, and steady-fluid-flow solutions are more reliably obtained by calculating the long-time limits of solutions to the unsteady equations (Ref 22).

The inviscid, or Euler, equations are obtained by neglecting the viscosity and heat-conduction terms in the preceding equations. A necessary condition for the applicability of the Euler equations is that the Reynolds number Re be much greater than one, where Re is defined by:

$$Re = \frac{\rho u L}{\mu} \qquad \text{(Eq 14)}$$

In Eq 14 ρ, u, and μ are characteristic values of the density, velocity, and viscosity, respectively, of the fluid, and L is a characteristic distance over which the velocity changes appreciably, also called a gradient length. In a flow to which the Euler equations apply, L is typically the dimension of the apparatus that bounds the flow. The Reynolds number is approximately the ratio of the magnitude of the advective terms to that of the viscous terms in the fluid momentum equation, Eq 2. Thus, when Re is large, the viscous terms may sometimes be neglected.

When fluid Prandtl numbers are of order unity, smallness of the viscous terms also implies smallness of the heat-conduction terms relative to the advection terms in the energy equation.

There are many high Reynolds number flows, however, where neglect of the viscous and heat-conduction terms is not justified. Sometimes, fluid flows have broad regions over which the inviscid equations apply, coupled with thin regions (e.g., boundary layers and shocks) in which the viscous and heat-conduction terms are important. In addition, as Re is increased, many flows become turbulent, and the velocity then varies over a range of length scales L. At the smallest of these length scales, Re is of order unity, and viscosity is important because it is responsible for the dissipation of turbulent kinetic energy into heat. (See the section on turbulence that follows.)

An incompressible flow is one in which the divergence of the velocity field is identically equal to zero:

$$\frac{\partial u_i}{\partial x_i} = S_{ii} = 0 \qquad \text{(Eq 15)}$$

A necessary, but not sufficient (see Ref 14), condition that a flow be incompressible is that the Mach number M be much less than one, where M is defined by:

$$M = \frac{u}{c} \qquad \text{(Eq 16)}$$

In Eq 16, u and c are characteristic values of the velocity and sound speed of the fluid.

In combination with the continuity equation, Eq 1, Eq 15 implies that:

$$\frac{D\rho}{Dt} = \frac{\partial\rho}{\partial t} + u_i\frac{\partial\rho}{\partial x_i} = 0 \qquad \text{(Eq 17)}$$

D/Dt is the time derivative following a fluid element, and Eq 17 states that the density of each element of fluid remains a constant along its trajectory. Usually, a more restrictive assumption is made that the density of the whole fluid is equal to a constant ρ_0. In this case, the momentum equation, Eq 2, becomes:

$$\frac{\partial u_i}{\partial t} + \frac{\partial(u_i u_j)}{\partial x_j} = \frac{1}{\rho_0}\frac{\partial\Sigma_{ij}}{\partial x_j} + F_i \qquad \text{(Eq 18)}$$

The great simplification of incompressible flow equations is that the energy equation is decoupled from the momentum equation and need not be solved. This is the so-called primitive-variable form of the incompressible flow equations. Another formulation that is used in CFD calculations of two-dimensional, incompressible flows is the stream function and vorticity formulation, which can be found in Ref 3.

Turbulence and Other Models. As has been stated earlier, there are many flow situations in which flows have changes in their properties, such as their velocities, with superimposed size scales or time scales that differ by many orders of magnitude. Examples are the seemingly

chaotic motions in a turbulent flow or in a multiphase flow such as a liquid spraying into a gas. Classical theories of turbulence predict that the ratios of the largest to the smallest fluctuation length scales of turbulent flows are approximately equal to $Re^{0.75}$, where, in this case, Re is based on the velocity and size scales of the largest turbulent eddies (Ref 23). Even a low value of $Re = 10,000$ gives fluctuation length scales varying over 3 orders of magnitude. For such cases, it is impossible to resolve the detailed flow fluctuations with CFD methods, and fortunately one usually is not concerned with these detailed fluctuations. The average flow behavior is of interest, however, and it is important to account for the effects of the fluctuations on average flow variables.

There are many ways to define averaged, or filtered, flow variables. In general, space- and time-averages can be defined using a filter function $K(x_i, t)$ whose integral over all space and time is unity. In terms of K, the average of a fluid variable q, denoted by $\bar{q}$, is defined by:

$$\bar{q}(x_i, t) = \iiint\int q(y_i, t')K(y_i - x_i, t' - t)dy_i \, dt' \tag{Eq 19}$$

For example, for pure time averaging, one can take $K(x_i, t) = \delta(x_i) \, \Psi_T(t)/T$, where:

$$\Psi_T(t) = \begin{cases} 1 & \text{if } |t| < T/2 \\ 0 & \text{otherwise} \end{cases} \tag{Eq 20}$$

and $\delta(x_i)$ is the Dirac delta function. Then, the average q is defined by:

$$\bar{q}(x_i, t) = \frac{1}{T}\int_{i-T/2}^{i+T/2} q(x_i, t') \, dt' \tag{Eq 21}$$

and the filter size is said to be T. In addition to space- or time-averaging, one can also use ensemble averaging. This is defined by averaging over an imagined large set of realizations of a fluid experiment. Sometimes, ensemble-averaging is combined with space- or time-averaging. In any case, the fluctuation of quantity q from its mean value is denoted by q':

$$q'(x_i, t) = q(x_i, t) - \bar{q}(x_i, t) \tag{Eq 22}$$

There are two approaches to calculating average flow fields. In the first, called Reynolds averaging because it was first proposed by O. Reynolds (Ref 24), one is interested in predicting *the* average flow field and uses ensemble averaging or a filter size that is large compared to the scales of fluctuations. Thus, the average of the fluctuating part of q is zero:

$$\overline{(q')} = 0 \quad \text{(Reynolds averaging)} \tag{Eq 23}$$

In contrast, subgrid-scale turbulence models use filters with as small a size as possible, typically comparable to the grid size in the CFD calculation. Thus, one attempts to calculate flow fluctuations with scales larger than the

filter, or grid, size and to model only subgrid-scale fluctuations. In a subgrid-scale model, the average of the fluctuation part of q is, in general, nonzero:

$$\overline{(q')} \neq 0 \quad \text{(subgrid model filtering)} \tag{Eq 24}$$

Once the method of averaging is chosen, then equations for the averaged flow variables can be obtained by averaging the equations of the preceding sections, or simplified forms of these. In deriving the averaged equations, one finds that the rates of change of average flow variables depend on averages of the products of two fluctuating quantities, also called second-order correlations. The values of these are unknown, and if one tries to close the system of equations by deriving transport equations for the second-order correlations, it is found that these depend on third-order correlations, or averages of three fluctuating quantities. Continuing in this way, one finds that it is impossible to obtain a closed system of equations using either Reynolds-averaging or subgrid-scale-averaging (Ref 25). By using physical and dimensional reasoning and empirical information, the unknown correlation must be expressed (the word *modeled* is also used here) in terms of average flow variables that are known.

A very important example of a second-order correlation and its modeling arises when averaging the incompressible flow momentum equation (Eq 18). The Reynolds-averaged form of this equation is:

$$\rho_0\left(\frac{\partial \bar{u}_i}{\partial t} + \frac{\partial \bar{u}_i \bar{u}_j}{\partial x_j}\right) = \frac{\partial \bar{\Sigma}_{ij}}{\partial x_j} - \frac{\partial}{\partial x_j}\left(\rho_0 \overline{u_i' u_j'}\right) + \rho_0 \bar{F}_i \tag{Eq 25}$$

The second-order correlation $-\rho_0 \overline{u_i' u_j'}$ on the right side of this equation is called the Reynolds stress. The most popular turbulence models in engineering design calculations are the so-called two-equation models, in which the Reynolds stress is, with some theoretical justification (Ref 25, 26), taken to have the form:

$$-\rho_0 \overline{u_i' u_j'} = 2\mu_T \bar{S}_{ij} - \frac{2}{3}\rho_0 K \delta_{ij} \tag{Eq 26}$$

In this expression, μ_T is the turbulent viscosity, and K is the turbulent kinetic energy:

$$K = \frac{1}{2}\overline{\left(u_i'\right)^2} \tag{Eq 27}$$

By substituting Eq 26 into Eq 25, one finds that the momentum equation for turbulent flow closely resembles the momentum equation for laminar, or nonturbulent, flow. This is also the case for other averaged fluid equations, and this resemblance allows the same numerical techniques for CFD to be applied to both laminar and turbulent flows.

In two-equation turbulence models, transport equations are solved for the turbulent kinetic energy K and one other scalar that gives a local

length or time scale of the turbulence. A popular choice for this second turbulence quantity is the turbulence kinetic energy dissipation rate ε. In terms of K and ε, the turbulent viscosity is given by:

$$\mu_T = c_\mu \rho_0 \frac{K^2}{\varepsilon} \tag{Eq 28}$$

where c_μ is a dimensionless constant. The K-ε turbulence model is described in more detail in Ref 26, which also describes the use of wall functions to calculate wall heat and momentum losses in conjunction with the K-ε model. Descriptions of many two-equation turbulence models, and their relative advantages, can be found in Ref 27.

Numerical Solution of the Fluid-Flow Equations

This section introduces some common techniques for discretizing the fluid-flow equations and methods for solving the discrete equations.

Discretization of the Fluid Equations

In the process of discretization, a continuously varying fluid flowfield, which has an infinite number of degrees of freedom, is represented by a finite set of data. This section introduces the discretization techniques used by finite-difference, finite-volume, finite-element, spectral, and some particle methods, and associated concepts of numerical stability and accuracy. The discrete equations of the finite-difference, finite-volume, and finite-element techniques all look similar and are referred to generically as "difference approximations." This introduction to CFD can only scratch the surface of each method. For more in-depth information, the reader should consult Ref 3 and 28 to 32.

Finite-Difference Methods (FDMs). In FDMs, the entire fluid region of interest is divided into nonoverlapping cells, and approximate values of the fluid variables are stored in each cell. This subdivision is called a grid or a mesh. Derivatives are approximated by taking differences between the variable values in neighboring cells, using the idea of a Taylor-series expansion. Consider the simple one-dimensional example of finite-difference solution of the linear advection equation:

$$\frac{\partial q}{\partial t} + u\frac{\partial q}{\partial x} = 0 \tag{Eq 29}$$

in the spatial interval $a \leq x \leq b$. This interval is subdivided into cells of equal size $\Delta x = (b - a)/N$, where N is the total number of cells, and q_i^n denotes the approximate value of q at the center of cell i, which lies at the location, or grid point, $x_i = a + (i - 1/2)\Delta x$, at time $t = n\Delta t$, where Δt is the computational timestep. (In this section, the subscript i represents a cell number,

rather than a coordinate direction.) All the values of $q_i^n, 1 \leq i \leq N$, for a particular time $t = n\Delta t$ are stored in computer memory, and values can be computed at time $t = (n + 1)\Delta t$ by using a finite-difference approximation to Eq 29.

To approximate the spatial derivative in Eq 29, q_i^n is considered to be the value at x_i of a differentiable function $q(x, t)$ that can be expanded in a Taylor series about any grid point. Thus, the value of q at a neighboring grid point can be expressed in terms of the value of q and its derivatives at grid point i by:

$$q_{i+k}^n = q_i^n + \left(\frac{\partial q}{\partial x}\right)_i^n k\Delta x + \left(\frac{\partial^2 q}{\partial x^2}\right)_i^n \frac{(k\Delta x)^2}{2} + O(\Delta x^3) \tag{Eq 30}$$

where $O(\Delta x^m)$ represents the fact that the remaining terms in this expansion have as their lowest-order term one in which Δx is raised to the power m. Now the value of the spatial derivative in Eq 29 can be approximated by any finite combination that satisfies:

$$\sum_k a_k q_{i+k}^n = \left(\frac{\partial q}{\partial x}\right)_i^n + O(\Delta x^m) \tag{Eq 31}$$

when one substitutes from Eq 30 for the q_{i+k}, where a_k are coefficients that depend on Δx. In the approximation Eq 31, m is said to be the order of accuracy of the approximation, and all terms containing Δx to some power are said to be truncation errors. As long as $m > 0$, the approximation is said to be consistent. Examples of consistent approximations are the centered-difference approximation:

$$\frac{q_{i+1}^n - q_{i-1}^n}{2\Delta x} = \left(\frac{\partial q}{\partial x}\right)_i^n + O(\Delta x^2) \tag{Eq 32}$$

which is second-order accurate, and the one-sided approximations:

$$\frac{q_i^n - q_{i-1}^n}{\Delta x} = \left(\frac{\partial q}{\partial x}\right)_i^n + O(\Delta x) \tag{Eq 33a}$$

and

$$\frac{q_{i+1}^n - q_i^n}{\Delta x} = \left(\frac{\partial q}{\partial x_i}\right)_i^n + O(\Delta x) \tag{Eq 33b}$$

which are first-order accurate. If the advection speed u in Eq 29 is positive, then the approximation Eq 33(a) is called an upwind approximation and Eq 33(b) a downwind approximation.

Order of accuracy is one measure of the accuracy of an FDM. To test the accuracy of a finite-difference solution, one can refine the grid by reducing the cell size Δx. When Δx is reduced by a factor of 2, numerical errors will be reduced approximately by a factor of 4 when using a second-order method, but only by a factor of 2 with a first-order method. It may thus seem to be desirable to use only methods with

a very high order of accuracy. In practice, however, it is difficult to define high-order methods near boundaries, and often numerical solutions using high-order methods have oscillations in regions of steep gradients. Because of these difficulties, most modern FDMs have second- to fourth-order accuracy and sometimes drop to first-order accuracy in regions of steep gradients.

Returning to the example of the linear advection equation, the time derivative can be approximated in much the same way as the spatial derivative. Because one usually only stores the value of q_i at a single time-level in order to save computer storage, the time derivative is most often approximated by the one-sided finite-difference formula:

$$\frac{q_i^{n+1} - q_i^n}{\Delta t} = \left(\frac{\partial q}{\partial t}\right)_i^n + O(\Delta t) \tag{Eq 34}$$

When Eq 34 and one of the finite-difference formulas Eq 32, 33(a), or 33(b) is used to approximate the time- and space-derivatives in Eq 29, one obtains a consistent approximation to the linear advection equation that is first-order accurate in time.

When these finite-difference equations are used to advance the numerical solution for q in time, one finds that, in contrast to solutions to the differential equation, solutions to the finite-difference equations using Eq 32 or the downwind approximation Eq 33(b) are subject to catastrophic numerical instabilities, and solutions using Eq 33(a) are only stable if a certain condition is met. This condition, the so-called Courant condition, is that the Courant number $C = (u\Delta t)/\Delta x$ be less than one. The origin of these numerical instabilities was first discovered by J. von Neumann (Ref 33), who devised a method for analyzing the stability of linear finite-difference equations based on examining the behavior of each Fourier component of the solution.

The finite-difference approximations presented so far are explicit in the sense that the solution for q_i^{n+1} can be explicitly found by solving only the finite-difference equation at grid point i. All explicit methods, if they are stable, are subject to Courant conditions to ensure their numerical stability. Intuitively, this condition arises because when using an explicit method, information can only propagate at a speed proportional to $\Delta x/\Delta t$. In order for the numerical solution to approximate the physical solution, the numerical propagation speed must be at least as great as the physical speed. For the simple advection Eq 29, the only physical propagation speed is u. For the fluid equations, there are several physical, or characteristic, speeds. The largest of these is $u + c$, where c is the fluid speed of sound, and the Courant condition in explicit CFD calculations is based on the speed $u + c$. To overcome the Courant condition, one uses implicit FDMs, in which solution for the value of q_i^{n+1} is implicitly coupled to the solution for q^{n+1} at other grid points.

An example of an implicit finite-difference approximation to the linear advection Eq 29 is:

$$\frac{q_i^{n+1} - q_i^n}{\Delta t} = -u\frac{q_{i+1}^{n+1} - q_{i-1}^{n+1}}{2\Delta x} \tag{Eq 35}$$

which can be shown to be unconditionally stable. The disadvantage of implicit methods is that they usually require costly iterative solution. Some iterative solution techniques for implicit equations are introduced as follows.

Finite-Volume Methods (FVMs). As in FDMs, FVMs subdivide the computational region into a mesh of cells, but finite-volume cells can be arbitrary quadrilaterals in two dimensions, hexahedra in three dimensions, or indeed any shape enclosed by a set of corner points. In contrast, FDMs are defined on grids that are obtained using orthogonal curvilinear coordinate systems. Thus, FVMs have much more geometric flexibility than FDMs.

Finite-volume methods approximate forms of the fluid equations that are integrated over these cells, which are also called control volumes. An example is the finite-volume approximation of the integrated form of the mass equation, Eq 1. After integrating Eq 1 over control volume V and applying the Reynolds transport and divergence theorems (Ref 14) one obtains:

$$\frac{d}{dt}\iiint_V \rho \, d\nu + \iint_S \rho u_i n_i \, da = 0 \tag{Eq 36}$$

The quantity $\rho u_i n_i$ is the mass flux (mass per unit area and time) through surface S with unit normal n_i, and Eq 36 is a statement that the time-rate-of-change of the total mass in volume V is equal to the sum of the fluxes, times the areas, through the surface S of the volume. Thus, mass is conserved in the sense that there are no internal mass sources. Commonly, the time derivative term in Eq 36 is approximated by:

$$\frac{d}{dt}\iiint_V \rho \, d\nu = V_\nu \frac{\rho_\nu^{n+1} - \rho_\nu^n}{\Delta t} \tag{Eq 37}$$

where ν is the index of finite-volume cell, and V_ν is its volume, and the surface integral is approximated by:

$$\iint_S \rho u_i n_i \, da = \sum_\alpha \rho_\alpha (u_i)_\alpha (n_i)_\alpha A_\alpha \tag{Eq 38}$$

where the sum is over all faces α of control volume ν; ρ_α and $(u_i)_\alpha$ are approximations to ρ and u_i, respectively, on face α; $(n_i)_\alpha$ is an average unit normal vector to face α pointing out of volume V; and A_α is the area of face α.

Using finite-volume methods, one can easily construct discrete approximations that have the conservative property; that is, the discrete approximations can mimic the physical laws from which the fluid equations were derived by conserving properties such as computed mass, momentum, and energy. To be more

precise, consider the approximation to the aforementioned mass equation. A conservative approximation has the property that if ν and μ are two cells that share face α, then when one sums the finite-volume approximations to the change of mass in cells ν and μ, the contributions due to fluxes through common face α cancel each other. This will be true if ρ_α and $(u_i)_\alpha$ are defined the same way in the finite-volume approximations at nodes ν and μ, because the unit outward normal to face α relative to cell ν is minus the outward normal relative to cell μ. Conservative difference approximations have many desirable accuracy properties. For example, it can be shown that difference approximations that conserve mass, momentum, and energy will calculate the correct jump conditions across shocks without having to resolve shock structure (Ref 3).

A problem with FVMs is that it is difficult to formulate higher-order FVMs. When a FVM is specialized to a finite-difference grid, the difference approximations look very much like finite-difference approximations, and one can perform Taylor-series expansions and determine the order of accuracy of the method. When more general meshes are used, however, it is unclear whether the same accuracy can be expected.

Finite-element methods (FEMs) (Ref 28) use a consistent spatial interpolation when evaluating all the spatial derivative terms in the fluid dynamics equations. These methods have long been popular in stress-analysis problems and have recently been gaining popularity in CFD problems because of advances in the methodology. As in FVMs, the computational domain is subdivided into nonoverlapping cells that in three dimensions are either arbitrary hexahedra or tetrahedra (Fig. 3, 4). Finite-element terminology is different, however, in that the cells are called elements, and the vertices of the cells are called nodes. A function $q(x_i,t)$ is represented by an expansion of the form:

$$q(x_i, t) = \sum_\nu q_\nu(t) b_\nu(x_i) \qquad \text{(Eq 39)}$$

where the sum is over all the nodes ν in the computational domain. The $b_\nu(x_i)$ are called basis functions and have finite support, meaning that they vanish outside of some neighborhood of the node ν location $(x_i)_\nu$. They also have the properties that:

$$b_\nu((x_i)_\mu) = \delta_{\nu\mu} \qquad \text{(Eq 40)}$$

where $\delta_{\nu\mu}$ is the Kronecker delta function, and:

$$\sum_\nu b_\nu(x_i) = 1 \qquad \text{(Eq 41)}$$

for all x_i. Linear (for tetrahedra) or trilinear (for hexahedra) basis functions give rise to second-order numerical methods in the following sense: When the finite-element grid is refined in such a way that the dimensions of elements are reduced by a factor of 2, then the difference
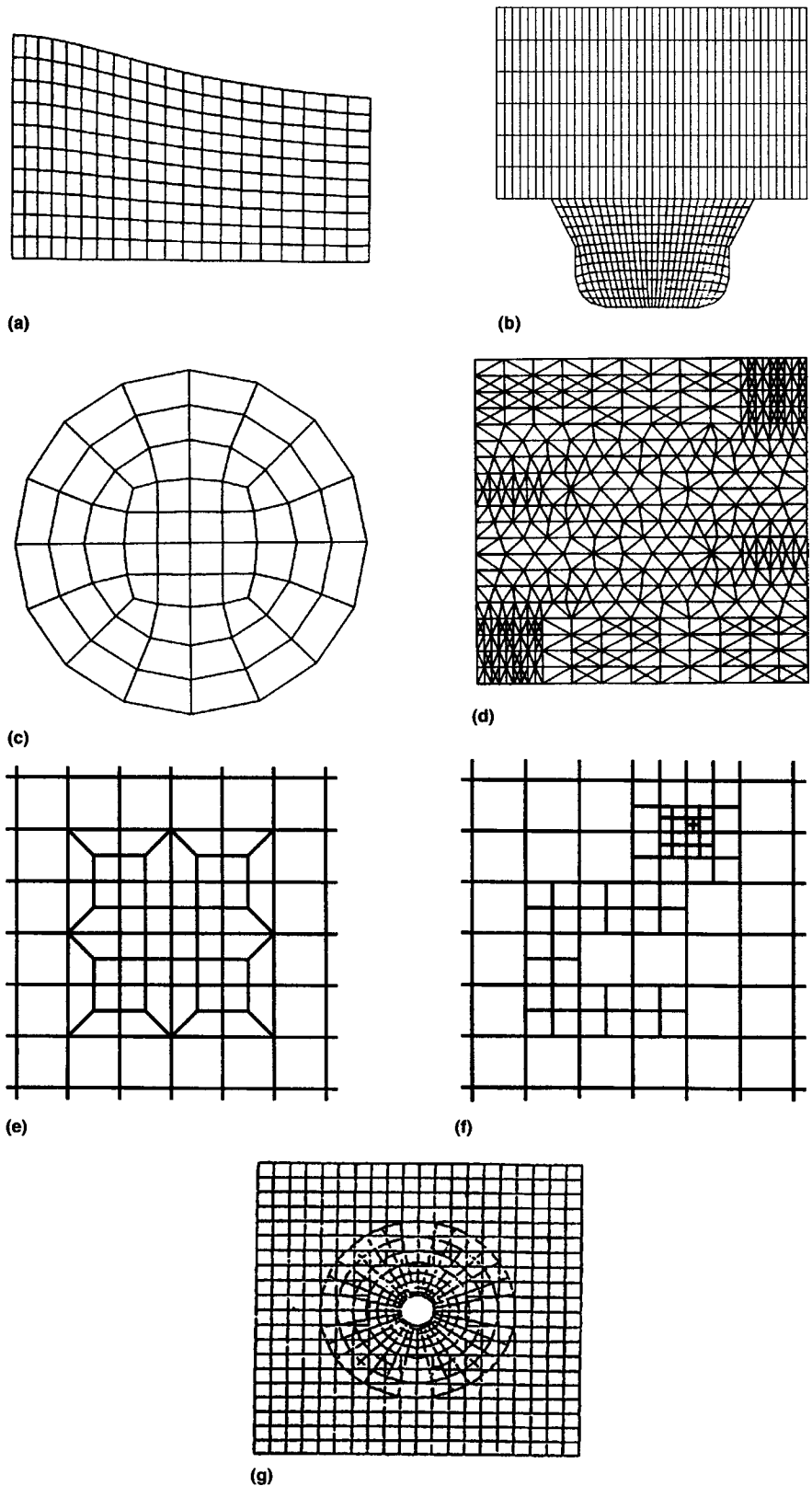
Fig. 3 Examples of grids used in computational fluid dynamics calculations. Two-dimensional examples are shown for clarity. (a) Structured grid. (b) Block-structured grid. (c) Unstructured hexahedral (quadrilateral) grid. (d) Unstructured tetrahedral (triangular) grid. (e) Local mesh refinement via a transition region on an unstructured hexahedral grid. (f) Local mesh refinement via cell splitting on an unstructured hexahedral grid. (g) Chimera grid

between the computed and exact solutions, as measured by a global integral of this difference, is reduced by a factor of 4. Higher-order FEMs can be constructed by adding midside nodes to the elements and using nonlinear basis functions that have properties Eq 40 and 41 (Ref 28). Because of Eq 40, the coefficient $q_v(t)$ is the value of q at location $(x_i)_v$ at time t.

Although there are many possibilities for determining $q_v(t)$, the most common method is that of the Galerkin finite-element method (GFEM). In GFEM, one substitutes expansions of the form Eq 39 for each function in the fluid equations. To obtain the discrete equations associated with node v, one multiplies the resulting expanded equations by basis function $b_v(x_i)$ and integrates over the entire computational domain. This gives rise to a coupled system of ordinary differential equations for the functions $q_v(t)$. Standard numerical methods for ordinary differential equations can then be used to solve for the $q_v(t)$. These ordinary differential equations involve coefficients that are integrals of products of the basis functions and their derivatives. Evaluating these coefficients can be a costly step in obtaining a GFEM solution.

Spectral Methods. Like FEMs, spectral methods (Ref 28) represent a function $q(x_i, t)$ by a finite sum:

$$q(x_i,t) = \sum_n c_n(t)\, b_n(x_i) \qquad \text{(Eq 42)}$$

Fig. 4 Principal cell or element types for computational fluid dynamics. (a) Tetrahedron: there are four vertices or nodes, four faces, and six edges for each element. (b) Hexahedron: there are eight vertices or nodes, six faces, and twelve edges for each element. Hexahedral elements generally must remain convex (angles formed by edge and face intersections must remain smaller than 180°). (c) A sampling of possible edge and/or face degeneracies for hexahedral elements

but unlike FEMs, the basis functions $b_n(x_i)$ are typically orthogonal functions with respect to some weight functions $W(x_i)$; that is:

$$\int\int\int b_n(x_i)\,b_m(x_i)W(x_i)dx_i = \delta_{nm} \qquad \text{(Eq 43)}$$

There is no grid in a spectral method. The $c_n(t)$ are no longer the values of q at nodes but simply the coefficients of the function q in an orthogonal function expansion. Ordinary differential equations for the $c_n(t)$ are obtained by a method that is similar to that of GFEM: one substitutes the expansion Eq 42 into the fluid equations and then multiplies the resulting expanded equation by $b_n(x_i)\,W(x_i)$ and integrates over the computational domain.

Spectral methods are most often used in situations where suitable basis functions can be found that satisfy the boundary conditions of a problem. When this is the case, spectral methods are very efficient for solving fluid-dynamics problems. For example, direct simulations of turbulence with periodic boundary conditions invariably use Fourier series expansions (Ref 34) because of their high accuracy. Because of the difficulty of finding suitable basis functions that satisfy boundary conditions in complex geometries, spectral methods are usually used only for simple geometries.

Computational Particle Methods. Computational particles have long been used for many purposes in CFD calculations (Ref 35). At the simplest level, they are used to follow the motion of Lagrangian fluid elements for flow-visualization purposes. At the other extreme, in some particle methods the fluid is completely represented by particles, each of which is endowed with a certain amount of mass, momentum, and energy. This is the case for particle-in-cell (PIC) methods (Ref 36, 37) and for the newer smoothed-particle-hydrodynamics (SPH) methods (Ref 38). The great advantage of the latter two methods is their Lagrangian nature. Because the Lagrangian equations are solved, numerical truncation errors are avoided that arise from finite-difference approximations to the advection terms. These are often the largest errors in approximations of the Eulerian equations. When carefully formulated, PIC and SPH method solutions can also be Galilean invariant and conserve angular momentum (Ref 37).

A disadvantage of particle methods lies in the difficulty of calculating interactions among fluid particles—which give rise, for example, to the pressure gradient terms in the momentum equation. This difficulty manifests itself, particularly in low Mach number calculations, in particle bunching and consequent fluctuations in advective transport. Possibly because of this difficulty, very few commercially available CFD codes use particle methods. An exception is a class of commonly used fluid/particle methods for calculating dispersed, two-phase flows (Ref 39, 40), such as occur when a liquid sprays into a gas. In these methods, computational

particles represent the dispersed phase entities and only interact with each other weakly, if they do so at all.

Solution of Implicit Equations

When solving difference approximations to the steady fluid equations or when solving implicit approximations to the unsteady equations, one must solve a large number of coupled algebraic equations for the unknown values of the fluid variables. When the equations are linear, an equation corresponding to the ith cell or node can be written in the form:

$$\sum_j a_{ij}q_j = s_i \qquad \text{(Eq 44)}$$

where a_{ij} are constant coefficients, and s_i is a known source term. The q_j are the unknowns to be solved for, in an unsteady problem $q_j=q_j^{n+1}$. For example, for the implicit approximation Eq 35 to the one-dimensional linear advection equation, one can take $a_{ii} = 1.0$, $a_{ii+1} = u\Delta t/(2\Delta x)$, $a_{ii-1} = -u\Delta t/(2\Delta x)$, and $s_i = q_i^n$. Equation 44 is usually written:

$$\mathbf{Aq} = \mathbf{s} \qquad \text{(Eq 45)}$$

where $\mathbf{A} = (a_{ij})$ is an $N \times N$ matrix of coefficients, N being the number of unknowns; $\mathbf{s} = (s_i)$ is a known source vector; and $\mathbf{q} = (q_j)$ is the vector of unknowns. Because the difference approximation in cell i only depends on the values of q in cell i and a small number of neighbors of cell i, only a small number of elements of the ith row of matrix $\mathbf{A}$ will be nonzero, and for this reason $\mathbf{A}$ is referred to as a sparse matrix. The basic problem of implicit fluid dynamics is to solve Eq 45 for the vector of unknowns, given a sparse matrix $\mathbf{A}$ and source vector $\mathbf{s}$.

Only for problems with small N can the matrix problem Eq 45 be solved directly by Gaussian elimination. This is because although the matrix $\mathbf{A}$ is sparse, and therefore does not require much computer storage for its nonzero elements, when Gaussian elimination is used to solve Eq 45, one finds that it is generally necessary to store in computer memory approximately N^2 nonzero coefficients, which is impossible in problems with a large number of cells.

Thus, interative methods are usually used to solve the matrix problem, Eq 45. Iterative solution methods calculate a sequence of approximations $\mathbf{q}^k$ that converge to the solution $\mathbf{q}$. The exact solution is not obtained, but one stops calculating $\mathbf{q}^k$ when either the difference between successive iterates $\mathbf{q}^{k+1}-\mathbf{q}^k$, or the residual $\mathbf{Aq}^k-\mathbf{s}$, is acceptably small. In the past, popular iterative methods have been point-successive relaxation, line-successive relaxation, and methods based on approximate decomposition of matrix $\mathbf{A}$ into a product of lower and upper triangular matrices that can each be easily inverted (Ref 30). These methods have largely been supplanted by two methods that

have greatly reduced the computer time to solve implicit equations and thereby have made implicit methods more attractive. These methods are conjugate-gradient methods (Ref 41) and multigrid methods (Ref 42).

When nonlinear finite-difference equations are solved, the aforementioned iterative methods can be used in conjunction with Newton's method (Ref 43). A nonlinear difference approximation can be written:

$$\mathbf{F(q)} = \mathbf{O} \qquad \text{(Eq 46)}$$

where $\mathbf{F}$ is a vector-valued function of the vector of unknowns $\mathbf{q}$. If $\mathbf{q}^k$ is the approximation to the solution $\mathbf{q}$ after k Newton-iteration steps, then $\mathbf{q}^{k+1} = \mathbf{q}^k + \delta\mathbf{q}$ is obtained by solving the matrix equation:

$$\frac{\partial \mathbf{F}}{\partial \mathbf{q}} \delta\mathbf{q} = -\mathbf{F}(\mathbf{q}^k) \qquad \text{(Eq 47)}$$

The matrix $\partial\mathbf{F}/\partial\mathbf{q}$ is called the Jacobian matrix. Equation 47 is of the form of Eq 45 and can be solved by one of the iterative methods for linear equations. Thus, solution for $\mathbf{q}$ involves using an iteration within an iteration. As in the solution of nonlinear equations for single variables, convergence is sometimes accelerated by under-relaxation; that is, one takes $\mathbf{q}^{k+1} = \mathbf{q}^k + \lambda\delta\mathbf{q}$ where $\delta\mathbf{q}$ is the solution to Eq 47 and λ is an underrelaxation factor whose value lies between zero and one.

Newton's method is sometimes used to solve systems of coupled difference equations arising in CFD (Ref 44), but it is often more economical for this purpose to use the simple-implicit method for pressure-linked equations (SIMPLE) method (Ref 45). In the SIMPLE method, a system of coupled implicit equations is solved by associating with each equation an independent solution variable and solving implicitly for the value of the associated solution variable that satisfies the equation, while keeping the other solution variables fixed. As is implied by the acronym SIMPLE, pressure is chosen as an independent variable, and special treatment is used to solve for pressure (Ref 45). The equations are solved sequentially, and repeatedly, until convergence of all the equations is obtained. The SIMPLE method is more efficient if the difference equations are loosely coupled, or if some independent linear combinations of the equations can be found that have little coupling.

Grid Generation for Complex Geometries

Before applying most of the CFD methods outlined previously, a computational grid must be generated that fills the flow domain and conforms to its boundaries. For complex domains with curved or moving boundaries, or with embedded subregions that require higher resolution than the remainder of the flowfield, grid generation can be a formidable task requiring more time than the flow solution itself. Two general approaches are available to deal with complex geometries: use of unstructured grids and use of special differencing methods on structured grids.

Unstructured Meshes. Figure 3 shows examples (in two dimensions) of several possible grid arrangements for CFD. In a structured three-dimensional grid (Fig. 3a), one can associate with each computational cell an ordered triple of indices (i, j, k), where each index varies over a fixed range, independently of the values of the other indices, and where neighboring cells have associated indices that differ by ± 1. Thus, if N_i, N_j, and N_k are the number of cells in the i-, j-, and k-index directions, respectively, then the number of cells in the entire mesh is $N_i\, N_j\, N_k$. Additionally, it is seen that each interior vertex in a structured grid is a vertex of exactly eight neighboring cells.

In an unstructured grid (Fig. 3c and d), on the other hand, a vertex is shared by an arbitrary number of cells. Unstructured grids are further classified according to the allowed cell or element shapes (Fig. 4). In the case of FVMs in particular, an unstructured CFD code may require a mesh of strictly hexahedral cells (Fig. 4b), hexahedral cells with degeneracies (Fig. 4c), strictly tetrahedral cells (Fig. 4a), or may allow for multiple cell types. In any case, the cells cannot be associated with an ordered triple of indices as in a structured mesh.

Intermediate between structured and unstructured meshes are block-structured meshes (Fig. 3b), in which "blocks" of structured grid are pieced together to fill the computational domain.

There are three advantages of unstructured meshes over structured and block-structured meshes. First, unstructured meshes do not require that the computational domain or subdomains be topologically cubic. This flexibility allows one to construct unstructured grids in which the cells are less distorted, and therefore give rise to less numerical inaccuracy, compared to a structured grid. Second, local adaptive mesh refinement (AMR) is naturally accommodated in unstructured meshes by subdividing cells in flow regions where more numerical resolution is required (Fig. 3e, f). Such subdivisions cannot be performed in structured meshes without destroying the logical (i, j, k) indexing. Third, in some cases, particularly when the cells are tetrahedra, unstructured grid generation can be automated with little or no user intervention (Ref 46). Thus, generating unstructured grids can be much faster than generating block-structured grids.

On the other hand, unstructured-mesh CFD codes generally demand higher computational resources. Additional memory is needed to store cell-to-cell and vertex-to-cell pointers on unstructured meshes, while this information is implicit for structured meshes. And, the implied connectivity of structured meshes reduces the number of numerical operations and memory accesses needed to implement a given solution algorithm compared to the indirect addressing required with unstructured meshes.

The relative advantages of hexahedral versus tetrahedral element shapes remain subjects of debate in the CFD community. Tetrahedra have an advantage in grid generation, because any arbitrary three-dimensional domain can be filled with tetrahedra using well-established methodologies (Ref 46). By contrast, it mathematically is not possible to tessellate an arbitrary three-dimensional domain with nondegenerate six-faced convex volume elements. Thus, each of the various automatic hexahedral grid-generation approaches that have been proposed (e.g., Ref 47, 48) either yields occasional degeneracies or shifts the location of boundary nodes, thus compromising the geometry.

Specialized Differencing Techniques. In a second general approach to computing flows in complex geometric configurations, the onus of work is shifted from complexity in grid generation to complexity in the differencing scheme (Ref 49–51). Structured and block-structured grids are used, but one of three numerical strategies is used to extend the applicability of these grids. The first strategy is to use so-called chimera grids (Ref 49) that can overlap in a fairly arbitrary manner (Fig. 3g). Solutions on the multiple grids are coupled by interpolating the solution from each grid to provide the boundary conditions for the grid that overlaps it. This is a very powerful strategy that handles naturally problems in which two flow regions meet at a boundary with a complicated shape or where one object moves relative to another. The second numerical strategy is to use so-called embedded boundaries (Ref 50). Again, structured meshes are used, but the complicated boundary of the computational domain is allowed to cut through computational cells. Special numerical methods are then used in the partial cells that are intersected by the boundary. In the third strategy, local AMR is allowed by using a nested hierarchy of grids (Ref 51). The different grids in the hierarchy are structured and have different cell sizes, but the cells in the more finely resolved grids must subdivide those of the coarser grids.

Although the second general approach affords simplicity in grid generation, it generally is less mature than the various unstructured-mesh approaches. Much development remains before these specialized differencing techniques have the robustness, generality, and efficiency to deal with the variety of problems presented in engineering applications. For the near future, then, the use of various unstructured-mesh approaches is expected to dominate in engineering applications of CFD.

Computational Fluid Dynamics for Engineering Design

Computational fluid dynamics is one of the tools available to the engineer to understand

and predict the performance of thermal-fluids systems. It is used to provide insight into thermal-fluids processes, to interpret experimental measurements, to identify controlling parameters, and to optimize product and process designs. It is the use of CFD as a design tool that is the principal focus here. In the course of a design program, an engineer may perform multiple CFD computations to explore the influence of geometry (hardware shape), operating conditions (initial and boundary conditions), and fluid properties. For CFD to be fully integrated into the design process, it must satisfy ever-tightening demands for functionality, accuracy, robustness, speed, and cost.

Most engineering CFD programs can be characterized as having high geometric complexity (domain boundaries as complex three-dimensional surfaces) and moderate physical complexity. The majority of flows considered are steady, incompressible, single-phase, and nonreacting. A common physical complexity encountered in engineering situations is turbulence, because engineering flows typically are characterized by high Reynolds number. Turbulence is modeled using a two-equation model (standard K-ε or variants, Ref 27) in most cases. Applications to transient flows with additional physical complexity and/or more sophisticated models (e.g., compressibility, multiphase, reacting, higher-order turbulence models) are increasing.

The CFD Process

An idealized component design process is shown schematically in Fig. 5. There, the left side of the flowchart depicts a hardware-based design process, while the right side represents an analysis- or math-based process. Although CFD is the single analysis tool under consideration here, the right side applies equally well to other mathematical/computational tools (e.g., finite-element structural analysis) that together fall under the heading of CAE.

Both the hardware- and analysis-based processes require the generation or acquisition of geometric data and the specification of design requirements. Here, it is assumed that a three-dimensional CAD geometry model is the preferred method for geometric representation. A hardware approach then proceeds with fabrication of prototypes, followed by testing of prototypes, and evaluation of test results. Design iterations are accomplished either by direct changes to the hardware or by modification of the CAD data set and refabrication, until the design requirements are satisfied. At that point, the original CAD data must be updated (in the case of direct hardware iterations), and the design proceeds to the next component or system level, where a similar process is repeated.

Analysis-based design (here, CFD) is not fundamentally different. Mesh generation replaces hardware fabrication, computer simulation substitutes for experimental measurement,

and postprocessing diagnostics are needed to extract relevant physical information from the vast quantity of numerical data. To the extent that relatively simple design criteria are available and the component lends itself to a parametric representation, the design-iteration loop can be automated using numerical optimization techniques (Ref 52). Automated computer optimization with three-dimensional CFD remains a subject of research; in most engineering applications, determination of the next design iteration remains largely a subjective, experience-based exercise.

Analysis-based design can be faster and less costly compared to hardware build-and-test. If this is not yet the case in a particular application, it most likely will be true at some point in the future. Thus, analysis affords the opportunity to explore more design possibilities within specified time and budget constraints. Advances in rapid prototyping systems (Ref 53) and other fabrication technology mitigate this advantage to some extent.

A second advantage of analysis is that more extensive information can be extracted compared to experimental measurements. Computational fluid dynamics yields values of the computed dependent variables (e.g., velocity, pressure, temperature) at literally thousands or even millions of discrete points in space and (in time-dependent problems) in time. From this high density of information can be extracted qualitative and quantitative pictures of flow streamlines and three-dimensional isopleths of any computed dependent variable. For time-dependent problems, animation or

"movies" reveal the time evolution of physical processes. Application-specific "figures of merit" including total drag force, wall heat flux, or overall pressure drop or rise can be computed. Experimental measurements, on the other hand, traditionally have been limited to global quantities or to values of flow variables at a small number of points in space and/or time. Thus, in principal, much more complete information is available from CFD to guide the next design iteration. An important caveat is that this additional information is useful only to the extent that it accurately and reliably represents the actual hardware under the desired operation conditions. In most applications of CFD today (2009), there are sufficient sources of uncertainty that abandonment of experimentation is unwarranted. Progress in two- and three-dimensional experimental diagnostics (e.g., particle-image velocimetry for velocity fields, Ref 54; laser-induced fluorescence for species concentrations, Ref 55) is enabling higher spatial and/or temporal measurement densities in many applications.

In Fig. 6, the CFD process is modeled as a four-step procedure: (1) geometry acquisition, (2) grid generation and problem specification, (3) flow solution, and (4) postprocessing and synthesis. Depending on the level of integration in the software selected, four (or more) distinct codes may be needed to accomplish these tasks. Some vendors offer fully integrated systems. For the purpose of exposition, each is treated separately.

Geometry Acquisition (CAD). The principal role of CAD software in the CFD process is to provide geometric definition of the bounding surfaces of the three-dimensional computational domain. The computational domain of interest in CFD generally is everything external to the solid material; this conveniently may be thought of as the negative of a finite-element structural solid model. Codes are designed primarily with the design and fabrication of three-dimensional solids in mind and have
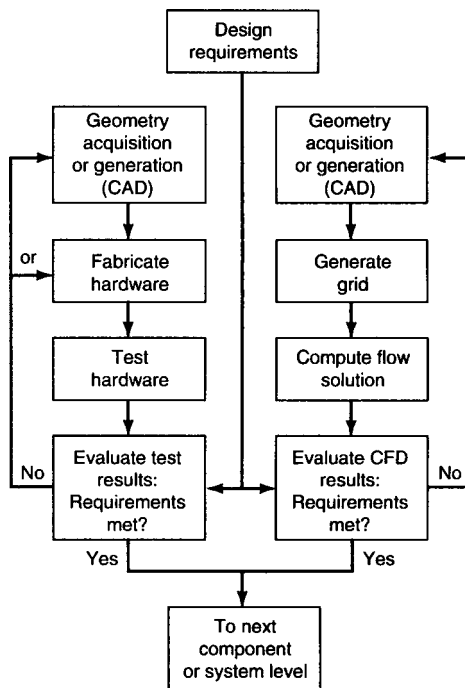
Fig. 5 Engineering component design processes. Left side depicts a hardware-based approach; right side is an analysis-(computational fluid dynamics, or CFD-) based approach. CAD, computer-aided design
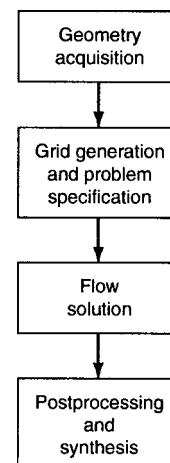
Fig. 6 The computational fluid dynamics process.

considerable functionality that is not of direct relevance for CFD (Ref 56).

The various CAD packages use different internal representations for curves (one-dimensional objects), surfaces (two-dimensional objects), and solids (three-dimensional objects). The surfaces needed for CFD, for example, may be represented using one of several tensor-product polynomial or spline representations in a two-dimensional parametric space (Ref 57, 58). Any of these representations generally suffice for CFD; most FDM, FVM, and FEM solution methodologies in current engineering CFD codes require at most linear interpolation between the discrete points (nodes or vertices) representing the surface. However, spectral-element methods (Ref 59) and some other high-order orthogonal-basis function expansions require a level of surface definition that generally is not available from current commercial CAD systems; this limits the application of such methods to simple geometric configurations at present.

The need to move geometry models among different CAD systems having different internal representations led to the establishment of standards for external geometric data exchange. An early standard supported by most CAD software is the initial graphics exchange specification (IGES, Ref 60). Most CAD-to-CFD interfaces operate by extracting the outer surfaces and writing an IGES file of "trimmed" B-spline surfaces. Newer standards such as standard for the exchange of products model data (STEP) are merging with IGES and supplanting it; existing standards are evolving rapidly, and new standards are developed as needed. Other external data formats commonly used in the CAD/CAE arena include stereolithography, where surfaces are processed into a set of triangular facets, cloud-of points (a set of random points in three-dimensional space), and DES (a set of piecewise linear curves describing a surface).

The set of raw surfaces extracted from the CAD model usually requires additional processing before it is suitable for CFD grid generation. The extracted surfaces may not define a closed three-dimensional domain (gaps), there may be more than one surface at a physical location (overlaps), and there simply may be too much geometric detail to be practical for CFD. Modern CAD and grid-generation systems provide fault tolerance and a variety of tools to "clean up" the extracted surfaces prior to grid generation. This cleanup step is labor-intensive and often is the single most time-consuming element of the CFD process.

Grid Generation and Problem Specification. The second step in the CFD process is to generate a computational mesh. This may be accomplished using the same software as for geometry acquisition, or a separate code. The grid must satisfy three general requirements:

- It must be compatible with the selected flow solver.

- It must be sufficiently fine to satisfy accuracy requirements.
- It must be sufficiently coarse to satisfy computational resource limitations.

For an unstructured mesh, the minimum information that must be provided from the grid-generation step is the location of each node or vertex and a description of connectivity among the vertices. In addition, a complete problem prescription for CFD requires the specification of initial and boundary conditions for all flow variables (e.g., velocity, pressure, temperature), fluid properties, and any model and numerical parameters. Other code- and application-specific information also may be needed. Because both geometry and grid information are available at the grid-generation stage, this is the most natural time to tag volumes for initial conditions and material properties and surfaces for boundary conditions (e.g., specify which surfaces represent walls, inflow boundaries, etc.). Specific initial values for each dependent variable at each interior cell or vertex, boundary values for each boundary element face or vertex, and fluid properties may be set either in the grid-generation software itself or in a separate "preprocessor" provided for the specific CFD code. For present purposes, the preprocessor is considered to be part of the flow solver. Model constants and numerical parameters are specified to the flow solver directly.

Fully automatic tetrahedral-mesh generation is available in a number of commercial and public-domain codes (Ref 46). Early generations of automated hexahedral, hexahedral-with-degeneracies, and hybrid hexahedral/tetrahedral strategies (requiring varying levels of manual intervention) also are available (Ref 47, 48). However, a high level of manual intervention still is required to generate high-quality meshes for CFD. This is particularly true in the case of tetrahedral meshes in the vicinity of solid walls. A "high-quality" mesh is defined here as one that yields high numerical accuracy for low computational effort (memory and central processing unit time). This is quantified by performing multiple computations of a single flow configuration using different meshes and computing the error in each with respect to a benchmark numerical or experimental solution. Discussions of modern mesh-generation techniques for CFD can be found in Ref 32 and 61.

Regardless of the specific methodology used to generate the mesh, it is important that any grid-generation software for CFD maintain separate data structures for geometry definition and for the computational mesh. This ensures that design changes (modifications to CAD surfaces) can be made without redoing the domain decomposition, that boundary conditions can be reset without regenerating the grid, and that mesh density and distribution can be changed independently of the geometry.

Flow Solution. Most contemporary CFD solvers available to the industrial design engineer use either finite-volume or finite-element

discretization, with SIMPLE-like iterative pressure-based implicit solution algorithms. Unstructured meshes of primarily hexahedral elements (with limited degeneracies) have been prevalent in most finite-volume formulations to date, although the grid-generation advantages of tetrahedra are leading to an increase in the usage of that element type.

Default or recommended values of numerical parameters are provided by each flow solver. New and/or unusual applications often require experimentation in selecting values of numerical parameters to obtain a stable, converged solution. For the solution methodologies commonly used today (2009), parameters include choice of advection scheme (e.g., the degree of upwinding), convergence criteria for linear equation solvers and pressure iterations, time-step control (for transient problems), mesh adaptation (where available), and other method-specific controls. For this reason, the CFD practitioner must have a working knowledge of the information covered in the "Governing Equations" section of this article. With these caveats, flow solution is the step requiring the least manual intervention. The engineer can monitor the solution as it progresses using the available diagnostics, which are discussed next.

Postprocessing and Synthesis. Viewing and making sense of the vast quantities of three-dimensional data that are generated in CFD is a challenging task. Many software packages have been developed for this purpose, both for structured and unstructured meshes. All provide considerable flexibility in setting model orientation, in passing cutting planes and/or lines through the computed solution, and in displaying the computed vector and scalar fields. Postprocessors have varying levels of "calculator" capability for computing quantities not supplied directly from the CFD solution, such as vorticity or total pressure. Many allow transient animation to accommodate time-dependent data. Most modern packages provide both a graphics-user interface and a save file/read file capability, the latter to allow the user to replicate a particular view of interest for multiple data sets.

Such direct inspection of the computed fields provides detailed insight into flow structure in the same sense as a high-resolution flow visualization experiment. In this respect and others, it had been argued that CFD is more akin to experiment than to theory. Features such as an undesirable flow separation, for example, may provide the engineer with sufficient information to guide a modification to the device geometry for the next design iteration. The connection between device performance or design requirements and the full three-dimensional flow field often is not obvious, however, considerable effort may be required to extract meaningful figures-of-merit from the numerical solution.

Judicious development of diagnostics is necessary to advance CFD from a sophisticated flow-visualization tool to a scientifically based design tool. Quantitative information of direct relevance to the designer is needed to drive

design changes toward satisfaction of the design requirements. Such diagnostics are application-specific and have received relatively little attention by CFD researchers and code developers. Examples of diagnostics to extract physical insight and to assess numerical accuracy can be found in Ref 62.

CFD of Manufacturing Processes. Application areas that have been particularly active in their use of CFD include aircraft and ship design, geophysical fluid flows, and flows in industrial devices that involve energy conversion and utilization. A comprehensive list of the applications of CFD would be difficult to compile, and no attempt to do so is made here. Instead, the focus here is to just briefly review CFD in casting (Ref 63–77).

In manufacturing processes such as casting, injection molding, welding, and crystal growth, heat conduction in the solid is coupled to convective heat transfer in the fluid. The solid-liquid interface moves with time, and its location needs to be tracked as a propagating three-dimensional surface in the CFD solution. Also, fluid properties may be highly temperature dependent and non-Newtonian including phase changes.

Metal casting is a key example of such an application. Casting is a process in which parts are produced by pouring molten metal into a cavity having the shape of the desired product. Figure 7(a) is a schematic of a typical sand-casting configuration. Once the two halves of the mold have been made, they are carefully aligned, one over the other, with the aid of pins and bushings in the sides of the molding boxes, to create the complete mold. Aside from the casting cavity itself, other features are also incorporated into the finished mold, such as the pouring basin, downsprue, runners, and ingates that conduct the molten metal into the casting cavity. Risers, or reservoirs of molten metal that remain molten longer than the casting, are needed with most metals and alloys that undergo liquid shrinkage as the casting solidifies. These are placed at critical locations in the mold, generally at heavier sections and areas remote from the ingates. Once the casting has been poured and allowed to cool, and after it has been withdrawn from the sand mold, these appendages are removed before the casting undergoes various finishing operations.

Fluid flow plays two important roles in the casting process. First, and most obviously, the flow of molten metal is necessary to fill the mold. Second, and less obvious, are the effect of convective fluid flow during solidification of the casting. It is the task of the foundry engineer to design grating and riser systems (Fig. 7a) that ensure proper filling and solidification, and CFD is playing an increasingly important role in this field. Proper designs result in lower scrap and less casting repair at the foundry. An example of a computational mesh and computed solidification times is given in Fig. 7(b) and (c). One CFD package that has been developed specifically for the modeling of flow and
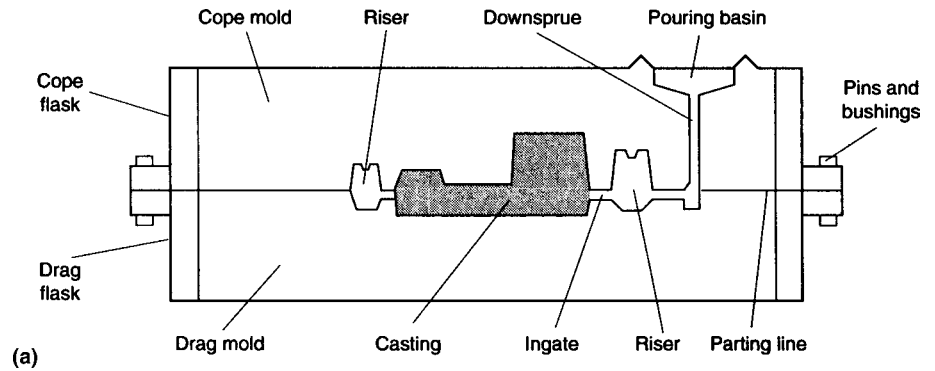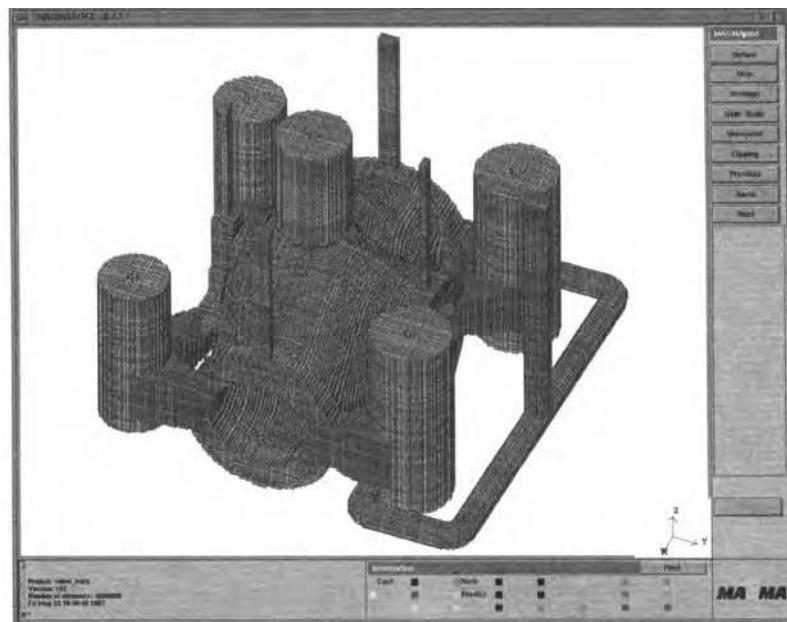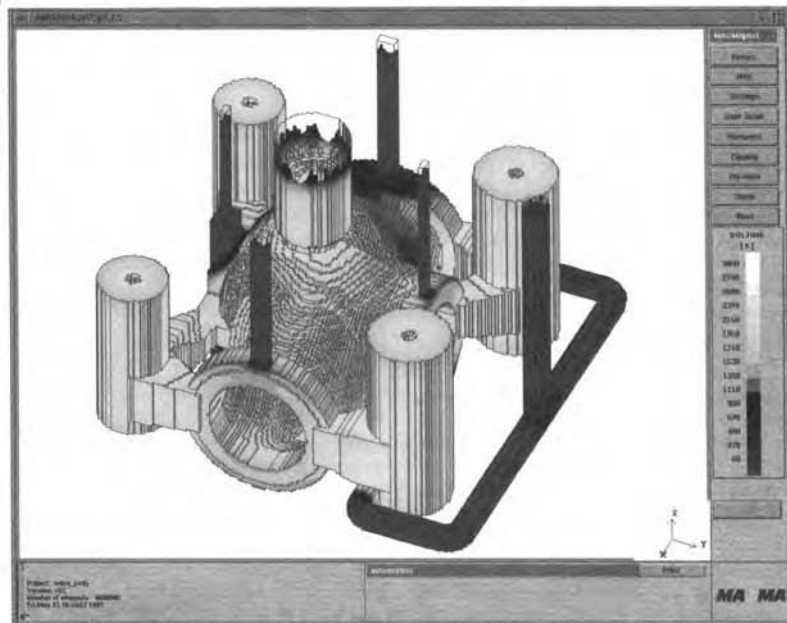
Fig. 7 Metal casting simulation. (a) Typical sand-casting configuration. (b) Automatically generated mesh (five million elements) for casting and cooling channels (Ref 78). (c) Computed local solidification times, which range from 1 to 3000 s (Ref 78)

thermal phenomena in casting applications is Magmasoft (Ref 79). References from the literature give ample evidence of the vast amount of CFD activity that is taking place in this area (Ref 78, 80).

Issues and Directions for Engineering CFD

Geometric fidelity between hardware and the computational mesh is crucial to obtaining accurate results. It is characteristic of the highly nonlinear flow equations that small geometric perturbations can result in large changes to the flow-field. One example is shown in Fig. 8 (Ref 81). There, significantly different flow structure and mixing result when the fraction-of-a-millimeter gap between piston and cylinder liner (the "top-ring-land crevice") is included in the mesh compared to when it is ignored. With a top-ring-land crevice, the flow entering the cylinder attaches to the cylinder wall and flows parallel to the wall for an extended time; in the absence of a top-ring-land crevice, the entering flow quickly adopts the port angle on entering the cylinder. This highlights the importance of maintaining a consistent three-dimensional representation of the hardware at all stages of design, analysis, and fabrication. The CFD practitioner should be wary of compromising the geometry in favor of grid-generation expediency, particularly in applications where he or she has little previous experience.

Numerical Inaccuracy. Meshes of hundreds of thousands of computational cells are common in transient engineering applications of CFD today (2009), and several millions of cells are being used in steady-state computations. Even so, numerical inaccuracy remains an issue for three-dimensional CFD. A mesh of 1 million cells corresponds to just 100 nodes in each coordinate direction in a three-dimensional calculation. With the low-order numerics that characterize engineering CFD, this is sufficient to resolve a dynamic range of approximately 1 order of magnitude (a factor of 10) in flow scales.

Rapid progress is being made both in discretization schemes for tetrahedral meshes and in automated grid generation for (primarily) hexahedral meshes; it is unclear at this time which will become dominant in engineering CFD.

The physical models used to represent turbulence, combustion, sprays, and other unresolvable phenomena are a third source of uncertainty in CFD. Turbulence modeling, in particular, is an issue that affects nearly all engineering applications. Research toward improved models continues. Much new physical insight into turbulence is itself being derived from large-scale numerical simulations (Ref 82).

In many high-Reynolds-number engineering applications where the instantaneous flow is highly transient and three-dimensional, turbulence models can be used to reduce the problem to one of steady flow, provided that the *mean* quantities of interest are time independent. This reduces the computational requirements considerably and provides results of acceptable accuracy in many cases. However, as engineering design requirements tighten, there is an increasing number of problems that demand a full three-dimensional transient treatment. Models still are needed to account for scales smaller than those that can be resolved numerically, but subgrid-scale turbulence models are used instead of Reynolds-averaged models. The resulting three-dimensional time-dependent simulations in this case are referred to as large-eddy simulations (LES) (Ref 83). The use of LES in engineering design is expected to proliferate rapidly. Examples of current applications of interest include acoustics and aerodynamic noise (Ref 84) and in-cylinder flows in engines (Ref 85).

Each of these three sources of uncertainty can, in principle, be isolated and quantified in simple configurations where a second source of data (e.g., experimental measurements) is available. It is more difficult in engineering applications of CFD to isolate and to quantify these errors to obtain meaningful estimates of error bounds. Early in the history of three-dimensional CFD, discrepancies between CFD and experiments generally were attributed to the turbulence model. The importance of the other sources of uncertainty and numerical
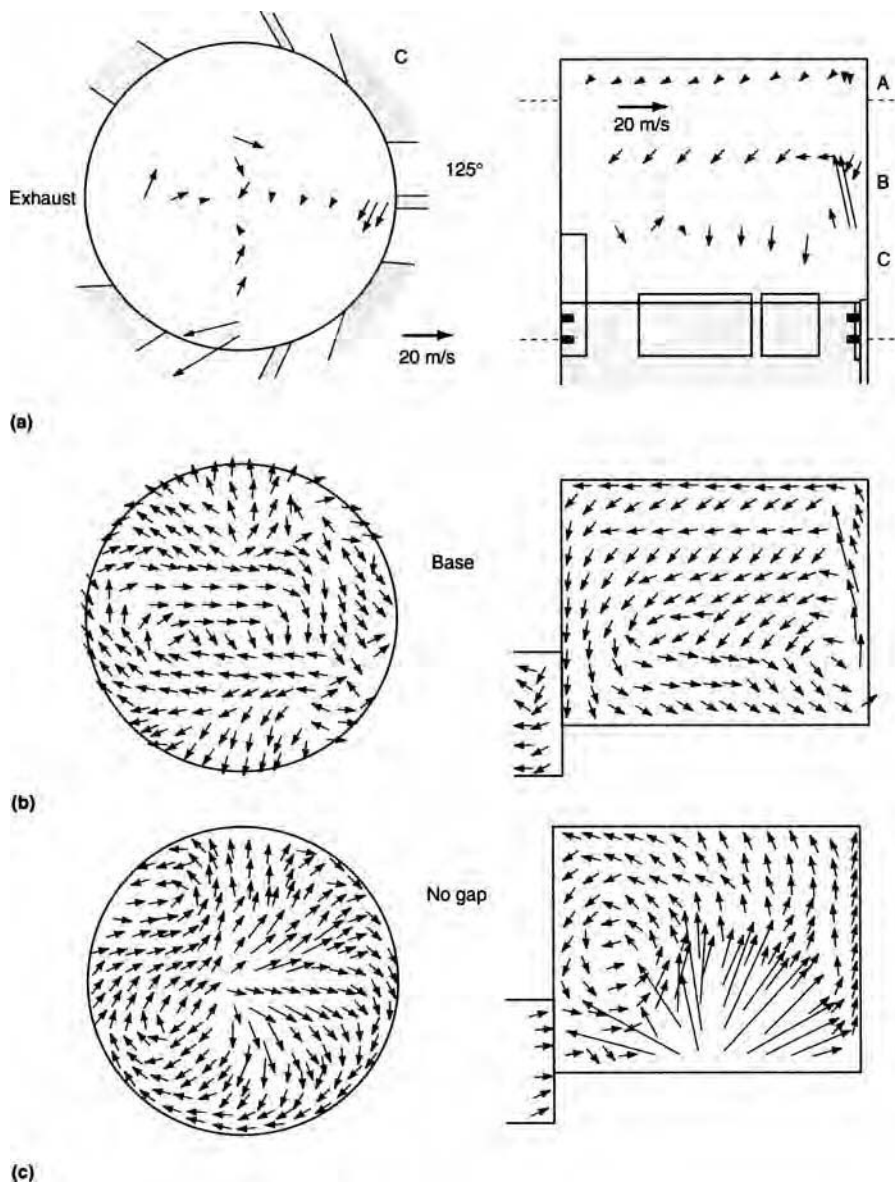
Fig. 8 Computed and measured ensemble-mean velocity fields on two-dimensional cutting planes at 125° after piston top-dead-center for a ported two-stroke-cycle engine. Computational results with and without a top-ring-land crevice are shown. (a) Measured. (b) Computational fluid dynamics (CFD) with top-ring-land crevice. (c) CFD without top-ring-land crevice

inaccuracy, in particular, has been more widely acknowledged recently (Ref 62, 86, 87). In the authors' experience, most discrepancies between computations and measurements for single-phase nonreacting flows in complex configurations are traced to geometric infidelity or to inadequate mesh resolution (in cases where they have been traced at all).

User Expertise. Computational fluid dynamics codes generally require more experience on the part of the user than other, more mature, CAE tools (e.g., linear FEM structural analysis). "General-purpose" CFD software provides a large number of numerical parameters and problem-specification options. In steady-flow problems, results should be independent of the choice of initial conditions, but different initial conditions may lead to different steady solutions when time-marching to the steady state. The choice of computational domain and specification of boundary conditions always are important, both for steady and time-dependent flows. Minimal user experience may suffice to obtain a reliable solution for steady incompressible flow in a benign geometric configuration, but considerable expertise is needed in problem specification and in results interpretation for complex flows.

CFD and Experimental Measurements. The engineering and scientific community typically accepts measurements from experiments as being more reliable than similar information generated by a CFD calculation. This is the reason for the strong emphasis placed by the profession on "validating" CFD results. While it is true that there are many sources of uncertainty in CFD, the same is true of experiments, particularly for complex systems (e.g., the in-cylinder flow in the last example). When comparing CFD results with measurements for such complex engineering problems, it is more appropriate to approach the exercise as a "reconciliation" rather than a "validation," because the latter implies that the experiment provides the "correct" value.

Interdisciplinary Analysis. In this overview, CFD has been considered as an isolated analysis tool. This is satisfactory only to the extent that one can reasonably prescribe boundary conditions that are independent of the flow solution itself.

For example, a coolant-flow analysis of temperature boundary conditions may be prescribed from a separate finite-element structural analysis, but the temperature field in the solid depends on the coolant flow itself. One can alternate through a sequence of CFD and thermal structural analyses, taking the most recent boundary conditions available at each step, to obtain a solution that effectively is coupled. A single direct computation of the coupled solution would be more satisfactory, however. In this case, a coupled fluid/heat conduction analysis is feasible because many CFD codes provide conjugate heat transfer capability.

More difficult are cases where fluids and solids interact in a manner that changes the shape of the flow domain. Flow/structure interactions including deformations are important, for example, in some aircraft design problems or in applications where there is significant thermal distortion. Interdisciplinary analysis tools are becoming available for these problems and will see more widespread use in the future.

Future of Engineering CFD. Most contemporary commercial CFD codes start from a discretization of the continuum equations of fluid mechanics and require a computational mesh of discrete cells or elements. An alternative is to approach CFD from a kinetic theory point of view. In Ref 88, for example, an (essentially) grid-free Lagrangian-particle method has been developed and implemented. It is too early, at the time of this writing, to speculate on the future of this approach for engineering design. Computations have been reported for configurations including external flow over simplified and realistic vehicles.

Research areas for CFD have included automated mesh generation, numerical algorithms for parallel computer architectures, linear equation solvers, more accurate and stable discretization schemes, automatic numerical error assessment and correction, improved solution algorithms for coupled nonlinear systems, new and enhanced physical models, more sophisticated diagnostics, interdisciplinary coupled structures/fluids analysis, optimization algorithms, and coupling of three-dimensional CFD into systems-level models.

In the ideal math-based design process, CFD is one part of a multidisciplinary CAE approach, and the full system (versus isolated component) is considered. Grid generation is fully automated to ensure a high-quality (initial) mesh. The flow solver selects all numerical parameters and provides automated solution-adaptive mesh refinement to a specified level of error or allowable computational resource (time or cost). Solution diagnostics provide information of direct relevance to the design requirements. Automated design optimization through modifications to the geometry and/or operating conditions proceeds until design requirements are met.

REFERENCES

1. A. Date, *Introduction to Computational Fluid Dynamics*, Cambridge University Press, 2005
2. J.F. Wendt, Ed., *Computational Fluid Dynamics: An Introduction*, 3rd ed., Von Karman Institute Book, Springer, 2008
3. P.J. Roache, *Computational Fluid Dynamics*, Hermosa Publishers, 1982
4. N. Grün, "Simulating External Vehicle Aerodynamics with Carflow," Paper 960679, SAE International, 1996
5. W. Aspray, *John von Neumann and the Origins of Modern Computing*, MIT Press, 1990
6. "Accelerated Strategic Computing Initiative (ASCI): Draft Program Plan," Department of Energy Defense Program, May 1996
7. L.H. Turcotte, ICASE/LaRC Industry Roundtable, Oct 3–4 , 1994; original source for some data is F. Baskett and J.L. Hennessey, Micro-processors: From Desktops to Supercomputers, *Science*, Vol 261, Aug 13, 1993, p 864–871
8. J. K. Dukowicz and R.D. Smith, *J. Geophys. Res.*, Vol 99, 1994, p 7991–8014
9. R.D. Smith, J.K. Dukowicz, and R.C. Malone, *Physica D*, Vol 60, 1992, p 38–61
10. G. Taubes, Redefining the Supercomputer, *Science*, Vol 273, Sept 1996, p 1655–1657
11. J.B. Heywood, *Internal Combustion Engine Fundamentals*, McGraw-Hill, 1988
12. F.H. Harlow and A.A. Amsden, "Fluid Dynamics," Report LA-4700, Los Alamos Scientific Laboratory, June 1971
13. W.G. Vincenti and C. H. Kruger, *Introduction to Physical Gas Dynamics*, Robert E. Krieger Publishing, 1975
14. P.A. Thompson, *Compressible-Fluid Dynamics*, McGraw-Hill, 1972
15. H. Jeffreys, *Cartesian Tensors*, Cambridge University Press, 1997
16. F.A. Williams, *Combustion Theory*, 2nd ed., Benjamin/Cummings, 1985
17. T.G. Cowling, *Magnetohydrodynamics*, Inter-science Tracts on Physics and Astronomy, No. 4, 1957
18. S. Chandrasekhar, *Radiative Transfer*, Dover, 1960
19. D.R. Stull and H. Prophet, *JANAF Thermochemical Tables*, 2nd ed., NSRDS-NBS37, National Bureau of Standards, 1971
20. B. McBride and S. Gordon, "Computer Program for Calculating and Fitting Thermodynamic Functions," NASA-RP-1271, National Aeronautics and Space Administration, 1992
21. R.B. Bird, W. E. Stewart, and E.N. Lightfoot, *Transport Phenomena*, Wiley, 1960
22. L. Crocco, A. Suggestion for the Numerical Solution of the Steady Navier-Strokes Equations, *AIAA J.*, Vol 3 (No. 10), 1965, p 1824–1832
23. L.D. Landau and E.M. Lifshitz, *Fluid Mechanics*, Pergamon Press, 1959
24. O. Reynolds, On the Dynamical Theory of Incompressible Viscous Fluids and the Determination of the Criterion, *Philos. Trans. R. Soc. London*, Series A, Vol 186, 1895, p 123
25. H. Tennekes and J.L. Lumley, *A First Course in Turbulence*, MIT Press, 1972
26. B.E. Launder and D.B. Spalding, *Mathematical Models of Turbulence*, Academic Press, 1972
27. D.C. Wilcox, *Turbulence Modeling for CFD*, DCW Industries, 1993
28. R. Peyret and T.D. Taylor, *Computational Methods for Fluid Flow*, Springer-Verlag, 1983
29. G.D. Smith, *Numerical Solution of Partial Differential Equations*, 2nd ed., Oxford University Press, 1978

30. R.D. Richtmyer and K. W. Morton, *Difference Methods for Initial-Value Problems*, 2nd ed., Interscience Publishers, 1967

31. C.A.J. Fletcher, *Computational Techniques for Fluid Dynamics*, Vol 1, *Fundamental and General Techniques*, 2nd ed., Springer-Verlag, 1991

32. C.A.J. Fletcher, *Computational Techniques for Fluid Dynamics*, Vol II, Specific Techniques for Different Flow Categories, 2nd ed., Springer-Verlag, 1991

33. G.G. O'Brien, M.A. Hyman, and S. Kaplan, a Study of the Numerical Solution of Partial Differential Equations, *J. Math. Phys.*, Vol 29, 1950, p 223–251

34. M.J. Lee and W.C. Reynolds, "Numerical Experiments on the Structure of Homogeneous Turbulence," Report TF-24, Dept. of Mechanical Engineering, Stanford University, 1985

35. J.U. Brackbill and J.J. Monaghan, Ed., *Proceedings of the Workshop on Particle Methods in Fluid Dynamics and Plasma Physics*, in *Comput. Phys. Commun.*, Vol 48 (No. 1), 1988

36. F.H. Harlow, The Particle-in-Cell Computing Method for Fluid Dynamics, *Fundamental Methods in Hydrodynamics*, B. Alder, S. Fern-bach, and M. Rotenberg, Ed., Academic Press, 1964

37. J.U. Brackbill and H.M. Ruppel, FLIP: A Method for Adaptively Zoned, Particle-in-Cell Calculations of Fluid Flows in Two Dimensions, *J. Comput. Physics*, Vol 65, 1986, p 314

38. J.J. Monaghan, Particle Methods for Hydrodynamics, *Comput. Phys. Rep.*, Vol 3, 1985, p 71–124

39. J.K. Dukowicz, A Particle-Fluid Numerical Model for Liquid Sprays, *J. Comput. Phys.*, Vol 35 (No. 2), 1980, p 299–253

40. P.J. O'Rourke, "Collective Drop Effects in Vaporizing Liquid Sprays," Ph.D. thesis, Princeton University, 1981

41. Y. Sahd and M. Schultz, Conjugate Gradient-like Algorithms for Solving Non-Symmetric Linear Systems, *Math. Comput.*, Vol 44, 1985, p 417–424

42. W.L. Briggs, *A Multigrid Tutorial*, Society for Industrial and Applied Mathematics (Philadelphia), 1987

43. W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vettering, *Numerical Recipes: the Art of Scientific Computing*, Cambridge University Press, 1987

44. D.A. Knoll and RR. McHugh, Newton-Krylov Methods Applied to a System of Convection-Diffusion-Reaction Equations, *Compt. Phys. Commun*, Vol 88, 1995, p 141–160

45. S.V. Patankar, *Numerical Heat Transfer and Fluid Flow*, McGrow-Hill, 1980

46. M.C. Cline, J.K. Dukowicz, and F.L. Addessio, "CAVEAT-GT: A General Topology Version of the CAVEATCode," Report LA-11812-MS, Los Alamos National Laboratory, June 1990

47. HEXAR, Cray Research Inc., 1994

48. G.D. Sjaardeam et al., *CUBIT Mesh Generation Environment*, Vol 1 & 2, SAND94-1100/-1101 Sandia National Laboratories, 1994

49. W.D. Henshaw, A Fourth-Order Accurate Method for the Incompressible Navier-Stokes Equations on Overlapping Grids, *J. Comput. Phys.*, Vol 133, 1994, p 13–25

50. R.B. Pember et al., "An Embedded Boundary Method for the Modeling of Unsteady Combustion in an Industrial Gas-Fired Furnace," Report UCRL-JC-122177, Lawrence Livermore National Laboratory, Oct 1995

51. J.P. Jessee et al., "An Adaptive Mesh Refinement Algorithm for the Discrete Ordinates Method," Report LBNL-38800, Lawrence Berkely National Laboratory, March 1996

52. M. Landon and R. Johnson, Idaho National Engineering Laboratory, 1995

53. S. Ashley, Rapid Concept Modelers, *Mech. Eng.*, Vol 118 (No. 1), Jan 1996, p 64–66

54. D.L. Reuss, R.J. Adrian, C.C. Landreth, D.T. French, and T.D. Fansler, "Instantaneous Planar Measurements of Velocity and Large-Scale Vorticity and Strain Rate Using Particle Image Velocimetry," Paper 890616, SAE, 1989

55. M.C. Drake, T.D. Fansler, and D.T. French, "Crevice Flow and Combustion Visualization in a Direct-Injection Spark-Ignition Engine Using Laser Imaging Techniques," Paper 952454, SAE International, 1995

56. D. Deitz, Next-Generation CAD Systems, *Mech, Eng.*, Vol 118 (No. 8), Aug 1996, p 68–72

57. G. Farin, *Curves and Surfaces for Computer-Aided Geometric Design*, Academic Press, 1990

58. M. Hosaka, *Modeling of Curves and Surfaces in CAD/CAM*, Springer-Verlag, 1992

59. D.C. Chan, A Least Squares Spectral Element Method for Incompressible Flow Simulations, *Proceedings of the Fifteenth International Conference on Numerical Methods in Fluid Dynamics*, Springer-Verlag, 1996

60. "3D Piping IGES Application Protocol Version 1.2," IGES 5.2 Standard, IGES ANS US PRO/IPO-100-1993, U.S. Product Data Association, 1993

61. S. Sengupta, J. Hauser, P.R. Eiseman, and J.F. Thompson, Ed., *Numerical Grid Generation in Computational Fluid Dynamics*, Pineridge Press, 1988

62. D.C. Haworth, S.H. El Tahry, and M.S. Huebler, A Global Approach to Error Estimation and Physical Diagnostics in Multidimensional Computational Fluid Dynamics, *Int. J. Numer. Methods Fluids*, Vol 17, 1993, p 75–97

63. S. Peng, X. Li, S. Qiu, Y. Gan, and Z. Xiao, Application of Computational Fluid Dynamics to Analysis of Metallurgical Process for Continuous Casting Tundishes, *J. Univ. Sci. Technol. Beijing* (English ed.), Vol 11 (No. 3), June 2004, p 207–211

64. N. Hudson, S.H. Bhavnani, and R.A. Overfelt, Computational Fluid-Dynamics Modeling of the Hydrodynamics of Fluidization in the Sand Surrounding a Lost-Foam Casting Pattern, *Metall. Mater. Trans. B*, Vol 33 (No. 4), Aug 2002, p 565–575

65. A. Mukhopadhyay and J. Lam, Role of CFD in Analysis and Optimization of the Continuous Casting Processes, *First MIT Conference on Computational Fluid and Solid Mechanics*, June 12–15, 2001 (Cambridge, MA), 2001, p 912–916

66. J. Ha, P. Cleary, V. Alguine, and T. Nguyen, Simulation of Die Filling in Gravity Die Casting Using SPH and MAGMAsoft, *Second International Conference on Computational Fluid Dynamics in the Minerals and Process Industries*, Dec 6–8, 1999 (Melbourne, Australia), 1999, p 423–428

67. P.W. Cleary and J. Ha, Three Dimensional Modelling of High Pressure Die Casting, *Second International Conference on Computational Fluid Dynamics in the Minerals and Process Industries*, Dec 6–8, 1999 (Melbourne, Australia), 1999, p 437–442

68. I. Hamill and T. Lucas, Computational Fluid Dynamics Modelling of Tundishes and Continuous Casting Moulds, *Fluid Flow Phenomena in Metals Processing*, Feb 28–March 4, 1999 (San Diego, CA), 1999, p 279–286

69. R.P. Tavares and R.I.L. Guthrie, Computational Fluid Dynamics Applied to Twin-Roll Casting, *Computational Fluid Dynamics and Heat/Mass Transfer Modeling in the Metallurgical Industry*, Aug 24–29, 1996 (Montreal, Quebec, Canada), 1996, p 203–217

70. J. Goldak, A. Artemev, N. Dai, D. Downey, F. Montoya, C. Shu, and J. Zhou, Casting Analysis for Design: Continuum Mechanics, Metal Physics and Computer Science, *Computational Fluid Dynamics and Heat/Mass Transfer Modeling in the Metallurgical Industry*, Aug 24–29, 1996 (Montreal, Quebec, Canada), 1996, p 68–82

71. D. Winterscheidt, C.M. Wang, and K.-O. Yu, Finite Element Fluid Flow and Heat Transfer Modeling of the Casting Process, *Computational Fluid Dynamics and Heat/Mass Transfer Modeling in the Metallurgical Industry*, Aug 24–29, 1996 (Montreal, Quebec, Canada), 1996, p 126–128

72. X. Huang and B.G. Thomas, Modeling of Transient Flow Phenomena in Continuous Casting of Steel, *Computational Fluid Dynamics and Heat/Mass Transfer Modeling in the Metallurgical Industry*, Aug 24–29, 1996 (Montreal, Quebec, Canada), 1996, p 129–145

73. C. Cheng and K.-O. Yu, Innovative Approach for Modeling the Heat Transfer during Casting Solidification, *Computational Fluid Dynamics and Heat/Mass Transfer Modeling in the Metallurgical Industry*, Aug 24–29, 1996 (Montreal, Quebec, Canada), 1996, p 56–67

74. E. Walicki and A. Walicka, Pressure Drops in the Flow of a Liquid Metal in Slots of Casting Dies, *Computational Fluid Dynamics and Heat/Mass Transfer Modeling in the Metallurgical Industry*, Aug 24–29, 1996 (Montreal, Quebec, Canada), 1996, p 177–186

75. M. Trovant and S. Argyropoulos, The Implementation of a Mathematical Model to Characterize Mold Metal Interface Effects in Metals Casting, *Computational Fluid Dynamics and Heat/Mass Transfer Modeling in the Metallurgical Industry*, Aug 24–29, 1996 (Montreal, Quebec, Canada), 1996, p 108–122

76. R.I.L. Guthrie and R.P. Tavares, Computational Fluid Dynamics Applied to Twin-Roll Casting, *Can. Metall. Q.*, Vol 37 (No. 3–4), July–Oct 1998, p 241–250

77. C.-W. Chen, H.-Y. Chou, C.-R. Li, C.-T. Shei, and W.-S. Hwang, Comparison of Different Computational Fluid Dynamics Techniques on Their Applications to the Modelling of Mold Filling in Casting, *Trans. Jpn. Foundrymen's Soc.*, Vol 12, Oct 1993, p 94–106

78. R.H. Box and L.H. Kallien, Simulation-Aided Die and Process Design, *Die Cast. Eng.*, Sept/Oct 1994

79. MAGMASOFT, developed by MAGMA Giessereitechnologie GmbH, Aachen, Germany; marketed and supported in the U.S. by MAGMA Foundry Technologies, Inc., Arlington Heights, IL

80. L. Karlsson, Computer Simulation Aids V-Process Steel Casting, *Mod. Cast.*, Feb 1996

81. D. C. Haworth, M.S. Huebler, S.H. El Tahry, and W. R. Matthes, "Multidimensional Calculations for a Two-Stroke-Cycle Engine: A Detailed Scavenging Model Validation," Paper 932712, SAF, 1993

82. P. Moin and J. Kim, Tackling Turbulence with Supercomputers, *Scientific American*, Vol 276 (No. 1), Jan 1997, p 62–68

83. B. Galperin and S.A. Orszag, Ed., *Large-Eddy Simulation of Complex Engineering and Geophysical Flows*, Cambridge University Press, 1993

84. A.R. George, Automobile Aerodynamic Noise, *SAE Trans.*, Vol 99–6, 1990, p 434–457

85. D.C. Howorth and K. Jansen, LES on Unstructured Deforming Meshes: Towards Reciprocating IC Engines, *Proceedings of the 1996 Summer Program*, Stanford University/NASA Ames Center for Turbulence Research, 1996, p 329–346

86. R.W. Johnson and E.D. Hughes, Ed., *Quantification of Uncertainty in Computational Fluid Dynamics*, FED-Vol 213, Fluids Engineering Division, American Society of Mechanical Engineers, 1995

87. I. Celik, C.J. Chen, P.J. Roache, and G. Scheuerer, Ed., *Quantification of Uncertainty in Computational Fluid Dynamics*, FED-Vol 158, Fluids Engineering Division, American Society of Mechanical Engineers, 1993

88. K. Molvig, *Digital Physics: A New Technology for Fluid Simulation*, Exa Corporation, Cambridge, MA, Aug 1993

Transport Phenomena during Solidification

Jonathan A. Dantzig, University of Illinois at Urbana-Champaign

Fundamentals

Some of the most important tools that materials scientists use regularly are equilibrium phase diagrams. These diagrams, derived from thermodynamics, define the phases present and their relative amounts in equilibrium as a function of overall alloy composition, C, and temperature, T. However, there is no information in the diagram about how a system evolves in space, $\mathbf{x}$, and time, t. Transport phenomena is the generic name used to describe the dynamical and spatial aspects of the system behavior. There are more specific names. The study of energy transport, embodied in the temperature field $T(\mathbf{x},t)$, is called heat transfer. The transport of solute, given by $C(\mathbf{x},t)$, is called mass transfer. In fluid systems, it is also necessary to consider the velocity field $\mathbf{v}(\mathbf{x},t)$, the study of which is called momentum transport or, more commonly, fluid dynamics. Clearly, these subjects are much deeper and broader than the scope of a Handbook article. Therefore, the focus here is on an introduction to the topic, all given in the context of solidification.

One begins with the balance equations for mass, energy, and solute and the necessary boundary conditions for solving problems of interest in casting and solidification. The phenomena cover a vast range of length and time scales—from atomic dimensions up to macroscopic casting size, and from nanoseconds for interface attachment kinetics to hours for casting solidification. It is essential to understand how to determine which phenomena are most important at the particular length and time scale of interest for the problem at hand. Methods for doing so are described in the section "Scaling" in this article. Finally, several examples are given of the application of transport phenomena in solidification, focusing in particular on microstructure formation.

Balance Equations

The derivation of the general forms of the balance equations is given in Ref 1 and 2. Instead, a shorter form is presented here that begins with the general balance equations and then specializes them using common models for materials behavior appropriate for metal systems. Since the subject of interest is solidification, it is necessary to examine more closely how these equations are applied when there is a moving solidification front. This should become more clear as each balance equation is developed.

Mass Balance. The mass balance for a fluid of density ρ moving at velocity $\mathbf{v}$ is given by:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \qquad \text{(Eq 1)}$$

This is written in the so-called conservative form, which is quite useful if one decides to integrate over a control volume. Doing so, and choosing a fixed control volume V with surface S, yields:

$$\frac{\partial}{\partial t}\int_V \rho \, dV + \int_S \rho \mathbf{v} \cdot \mathbf{n} \, dS = 0 \qquad \text{(Eq 2)}$$

where the divergence theorem has been applied to convert the second term to a surface integral. The surface normal vector is $\mathbf{n}$. The interpretation of Eq 2 is clear: The time rate of change of total mass in the volume, represented by the first term, is balanced by the total mass leaving the volume through the surface.

Now consider the control volume shown Fig. 1, which encloses a portion of the solid-liquid interface. The interface, solid, and liquid are moving at velocities $\mathbf{v}^*$, $\mathbf{v}_s$, and $\mathbf{v}_L$, respectively, and the interface normal vector is $\mathbf{n}^*$. (The subscripts "s" and "L" are used throughout to refer to the solid and liquid, respectively, and the superscript "*" indicates quantities on the liquid-solid interface.) Now evaluate Eq 2 as the thickness of the control volume goes to zero. The first term vanishes, and the second term then consists of four parts:

$$\rho_s(\mathbf{v}_s \cdot \mathbf{n}^* - \mathbf{v}^* \cdot \mathbf{n}^*)$$
$$+ \rho_L(\mathbf{v}^* \cdot \mathbf{n}^* - \mathbf{v}_L \cdot \mathbf{n}^*) = 0 \qquad \text{(Eq 3)}$$

The physical meaning of Eq 3 is made more clear by assuming that the solid is not moving ($\mathbf{v}_s = 0$) and by rearranging terms to obtain:

$$\mathbf{v}_L \cdot \mathbf{n}^* = -\frac{\rho_s - \rho_L}{\rho_L}\mathbf{v}^* \cdot \mathbf{n}^* = -\beta \mathbf{v}^* \cdot \mathbf{n}^* \qquad \text{(Eq 4)}$$

where β is called the solidification shrinkage. Equation 4 states that there is a net flow of liquid into the interface in order to balance the density difference between the liquid and solid phases. This flow is the underlying cause of porosity and hot tearing in castings. It is less important for the other balances. Therefore, the shrinkage flow is omitted from the development of the momentum, energy, and solute balances that follow.

Momentum Balance. The balance of linear momentum is the continuum form of Newton's second law of motion, that is, that the sum of the forces on a body is balanced by the time rate of change of linear momentum. In conservative form:

$$\frac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot (\rho \mathbf{v}\mathbf{v}) = \nabla \cdot \boldsymbol{\sigma} + \rho \mathbf{b} \qquad \text{(Eq 5)}$$

where $\boldsymbol{\sigma}$ is the stress tensor, and $\mathbf{b}$ is a body force such as gravity. The dyadic product $\mathbf{v}\mathbf{v}$ is a tensor with components $v_i v_j$. Integrating over a control volume as for the mass balance shows more clearly the relation to Newton's second law:

$$\frac{\partial}{\partial t}\int_v \rho \mathbf{v} \, dV + \int_S \rho \mathbf{v}\mathbf{v} \cdot \mathbf{n} \, dS$$
$$= \int_S \mathbf{f} \cdot \mathbf{n} \, dS + \int_v \rho \mathbf{b} \, dV \qquad \text{(Eq 6)}$$

where $\mathbf{f} = \boldsymbol{\sigma} \cdot \mathbf{n}$ is the external traction force applied on the surface of the control volume. It is often convenient to separate the stress into
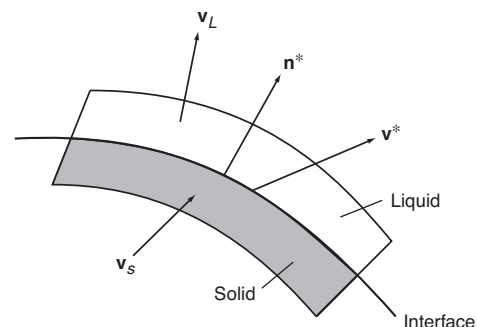
Fig. 1 Schematic of a control volume including the liquid-solid interface

hydrostatic pressure, p, and the "extra" stress tensor, $\boldsymbol{\tau}$, leading to the form $\boldsymbol{\sigma} = -p\mathbf{I} + \boldsymbol{\tau}$. Following a procedure similar to the one followed to develop the interface form of the mass balance leads to balance equations for the normal and tangential components of the force at the interface:

$$(p_L - p_s) - (\boldsymbol{\tau}_s - \boldsymbol{\tau}_L) : (\mathbf{n}^*\mathbf{n}^*) = 2\gamma\bar{\kappa} \qquad \text{(Eq 7)}$$

$$(\boldsymbol{\tau}_s - \boldsymbol{\tau}_L) : (\mathbf{n}^*\mathbf{t}^*) = \nabla_{surf}\gamma\cdot\mathbf{t}^* = 0 \qquad \text{(Eq 8)}$$

where γ is the surface energy of the liquid-solid interface, and $\bar{\kappa}$ is the mean curvature. The notation ":" implies a double-dot product, and ∇_{surf} indicates the gradient within the surface. This term captures the so-called Marangoni effect, where gradients in temperature or composition along the surface create tangential forces that can drive a flow in the liquid. These forces are very important for welding, where there is a free surface, but less so for liquids and solids. Equation 7 provides the well-known relation that the surface tension is balanced by a pressure difference across the curved interface.

Energy Balance. The energy balance is one of the most important balances for solidification problems. The most convenient thermodynamic variables are temperature, T, and pressure, p. The energy balance is written in terms of the enthalpy density, H, as:

$$\frac{\partial \rho H}{\partial t} + \nabla\cdot(\rho H\mathbf{v})$$
$$= \nabla\cdot(k\nabla T) - p\nabla\cdot\mathbf{v} + \boldsymbol{\tau} : \mathbf{D} + \dot{Q} \qquad \text{(Eq 9)}$$

where $\mathbf{D}$ is the rate of deformation tensor, $2\mathbf{D} = \nabla\mathbf{v} + (\nabla\mathbf{v})^T$, and τ:$\mathbf{D}$ is the work done by the shear stress, also called the viscous dissipation. For most solidification problems, the pressure effects, viscous dissipation, and internal heat generation, $\dot{Q}$, usually can be neglected, leaving the form:

$$\frac{\partial \rho H}{\partial t} + \nabla\cdot(\rho H\mathbf{v}) = \nabla\cdot(K\nabla T) \qquad \text{(Eq 10)}$$

Integrating over the control volume shows that the time rate of change of enthalpy in the control volume is balanced by that advected and conducted through the surface:

$$\frac{\partial}{\partial t}\int_V \rho h dV + \int_S \rho h\mathbf{v}\cdot\mathbf{n}dS = \int_S K\nabla T\cdot\mathbf{n}dS$$
$$= -\int_S \mathbf{q}\cdot\mathbf{n}dS \qquad \text{(Eq 11)}$$

where Fourier's law for the conductive heat flux, $\mathbf{q} = -K\nabla T$, has been applied. The interfacial balance, neglecting the shrinkage flow, is:

$$\rho_L H_L - \rho_s H_s = K_s\nabla T_s\cdot\mathbf{n}^* - K_L\nabla T_L\cdot\mathbf{n}^* \quad \text{(Eq 12)}$$

The left side of Eq 12 is the difference in enthalpy between the liquid and solid phases, the latent heat of fusion, ΔH; the right side is the net heat conduction away from the interface.

Inserting ΔH gives the Stefan condition:

$$\rho_s\Delta H = K_s\nabla T_s\cdot\mathbf{n}^* - K_L\nabla T_L\cdot\mathbf{n}^* \qquad \text{(Eq 13)}$$

Often, the ratio $K_s/K_L \approx 1$.

Solute Balance. The solute balance is developed in a similar way to the energy balance. For the sake of brevity, the treatment here is restricted to binary alloys (A and B) with no chemical reactions. In analogy with Fourier's law for conduction, Fick's law for diffusion is adopted, which gives the diffusive flux of species B as proportional to the composition gradient, $\mathbf{J} = D\nabla(\rho C)$. The composition is given as a mass fraction, that is, grams B divided by total grams, the units of the diffusivity D are m^2/s, and thus, the flux has units grams B per unit area per unit time. Note that there are numerous subtleties associated with whether the mass appears inside or outside the gradient operator when the density is not constant, but these issues are beyond the scope of this article.

The solute balance is given by:

$$\frac{\partial \rho C}{\partial t} + \nabla\cdot(\rho C\mathbf{v}) = \nabla\cdot(D\nabla\rho C) \qquad \text{(Eq 14)}$$

and the interfacial balance for solute is also analogous to the interfacial balance for energy:

$$(\rho_L C_L - \rho_s C_s)\mathbf{v}^*\cdot\mathbf{n}^* = D_s\nabla(\rho_s C_s)\cdot\mathbf{n}^*$$
$$- D_L\nabla(\rho_L C_L)\cdot\mathbf{n}^* \quad \text{(Eq 15)}$$

As opposed to heat conduction, where $K_s \approx K_L$, usually $D_s \ll D_L$, and the first term on the right side of Eq 15 is often neglected. The left side is usually modeled by assuming local thermodynamic equilibrium at the interface, and therefore the compositions C_S and C_L are given by the equilibrium phase diagram. In that case, the segregation coefficient, k, is introduced, such that $\rho_s C_s = k\rho_L C_L$. Introducing both of these simplifications in Eq 15 gives:

$$\rho_L C_L(1 - k)\mathbf{v}^*\cdot\mathbf{n}^* = -D_L\nabla(\rho_L C_L)\cdot\mathbf{n}^* \qquad \text{(Eq 16)}$$

Scaling

The governing equations, while necessary, are only the beginning. The geometry, boundary and initial conditions, and material properties also must be specified. When all of these data are gathered, one can, in principle, evolve the appropriate equations in time and find a solution to the problem at hand. This is usually done numerically, perhaps using one of the many available commercial codes. However, that is usually much more work than is either necessary or even possible, because the range of length and time scales involved with solidification processes is vast—from atomic-scale kinetic processes that take place over picoseconds to macroscopic heat flow at the size of the casting, extending over as much as an hour for large parts. Practical simulations can cover at most 3 to 4 orders of magnitude in both space and time, and this means that some phenomena must be omitted from the problem. But which ones should be kept and which ones

neglected (or are better to include via an analytical model)? Clearly, the important scales should be included, but how can one decide which those are? That is the subject of this section.

The decision about which length and time scales are important is done by scaling the governing equations. This provides a systematic means for ensuring that the important terms are kept while the negligible terms are discarded. First, the procedure is described, then a simple example is given. Scaling begins by choosing characteristic scales for variables such as length, temperature, and so on that usually come from the problem geometry and boundary conditions. Some characteristic values are unknown and are defined by the governing equations. The characteristic values are used to define dimensionless variables that are order one. For example, suppose that the temperature, T, is known (e.g., from initial and boundary conditions) to lie in the interval $T_1 \le T \le T_0$. A dimensionless temperature is defined, $\theta = (T - T_1)/(T_0 - T_1) \in [0,1]$. Unknown characteristics are just given a symbol and carried along. When the dimensionless variables are defined, they are substituted into the governing equations. Each term in the equation then has a (dimensional) coefficient multiplying a scaled (dimensionless) term. The equation is then divided by the coefficient of a term that is expected to be important. This requires some engineering judgment, but it is expected that something about the problem at hand is known! The unknown characteristics are then evaluated by setting the value of the coefficient where they appear to 1. Now that all of the terms in the equation are order 1 or less, terms whose coefficients are small compared to 1 can be safely neglected and the remaining problem solved.

A concrete example makes the scaling process seem much less mysterious. Consider a thin, circular rod of length L and radius $R \ll L$, initially at temperature T_0, quenched into a large bath at temperature T_1. Some simplifying assumptions are made in order to focus on the scaling process. Assume that the bath temperature stays at T_1, unaffected by the quench, and that the heat transfer from the surface of the rod to the bath can be characterized by a constant heat transfer coefficient, h. All of the material properties will be assumed to be constant, and the temperature distribution is axisymmetric. The problem is solved in polar coordinates. Finally, the rod is solid ($\mathbf{v} = 0$), and there is no internal heat generation ($\dot{Q} = 0$). The energy equation for the rod ($0 \le r \le R; -L/2 \le z \le L/2$) is then:

$$\frac{\partial T}{\partial t} = \alpha\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial T}{\partial r}\right) + \frac{\partial^2 T}{\partial z^2}\right] \qquad \text{(Eq 17)}$$

and the initial and boundary conditions are:

$$T = T_0 \qquad\qquad t = 0 \qquad\qquad \text{(Eq 18)}$$

$$-K\frac{\partial T}{\partial r} = h(T - T_0) \qquad r = R \qquad \text{(Eq 19)}$$

$$\pm K \frac{\partial T}{\partial z} = h(T - T_0) \qquad z = \mp L/2 \qquad \text{(Eq 20)}$$

where $\alpha = K/\rho c_p$ is the thermal diffusivity. Now the problem specification is used to define the following dimensionless variables:

$$\tilde{r} = \frac{r}{R}; \quad \zeta = \frac{z + L/2}{L}; \quad \theta = \frac{T - T_1}{T_0 - T_1} \qquad \text{(Eq 21)}$$

Time must also be scaled, but the characteristic value t_c is not yet known. Therefore, $\tilde{t} = t/t_c$ is defined, and guidance is awaited. The scaled variables are now inserted into Eq 17, and the entire equation is divided by $\alpha(T_0 - T_1)/R^2$, the coefficient of the term associated with conduction in the radial direction. The result is:

$$\frac{R^2}{\alpha t_c} \frac{\partial \theta}{\partial \tilde{t}} = \left[\frac{1}{\tilde{r}} \frac{\partial}{\partial \tilde{r}} \left(\tilde{r} \frac{\partial \theta}{\partial \tilde{r}} \right) + \frac{R^2}{L^2} \frac{\partial^2 \theta}{\partial \zeta^2} \right] \qquad \text{(Eq 22)}$$

Setting the dimensionless group on the left side to 1 gives the characteristic time $t_c = R^2/\alpha$, and the ratio $\alpha t/R^2$ is called the Fourier number. It can be seen that the geometry of the problem, where $R \ll L$, indicates that axial conduction, associated with the term $\partial^2 \theta/\partial \zeta^2$, is negligible in comparison to radial conduction. This makes the problem effectively one-dimensional, a great simplification. The initial condition is simply $\theta = 1$ at $\tilde{t} = 0$. The scaled form of the boundary conditions becomes:

$$-\frac{\partial \theta}{\partial \tilde{r}} = \frac{hR}{K} \theta \qquad \tilde{r} = 1 \qquad \text{(Eq 23)}$$

$$+\frac{\partial \theta}{\partial \zeta} = \frac{hR}{K} \frac{L}{R} \theta \qquad \zeta = 0,1 \qquad \text{(Eq 24)}$$

The dimensionless parameter hR/K is called the *Biot number*, Bi. Notice that information about the problem geometry is retained in the aspect ratio L/R, which now appears in the boundary condition.

To review concepts of the scaling: The characteristic time for conduction is R^2/α, conduction in the axial conduction can be neglected with respect to radial conduction for R/L, and the important parameter that determines the rate of heat loss to the environment is the Biot number, hR/K. The remaining one-dimensional problem can be solved, for example, using Fourier series
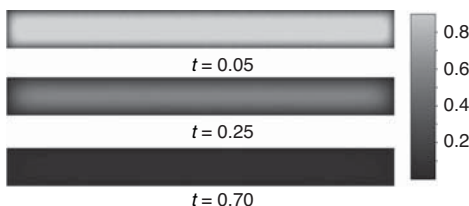
methods (Ref 3). The simplification to a one-dimensional problem came at a price, however. Since the derivatives with respect to ζ were eliminated from the differential equation, the end conditions in Eq 24 cannot be satisfied. It is expected that there will be a small region near the boundary, called a boundary layer, where the solution is two-dimensional. In this region, the axial and radial conduction terms must be comparable, and scaling should make it obvious that this layer extends a few R from each end. Figure 2 shows temperature solutions for a problem where Bi = 10 and $L/R = 20$, at several values of $\tilde{t}$. Notice that the temperature is nearly uniform at T_1 at $\tilde{t} = 0.7$—this shows the significance of the characteristic time.

Transport and Microstructure

Using the fundamentals presented in the preceding sections, solidification problems now can be discussed. Perhaps the most important thing to understand is that for most materials, $D_s \ll D_L \ll \alpha_L \approx \alpha_s$, and these large differences produce a separation of characteristic time and length scales. Because the thermal diffusivity is much larger than the chemical diffusivities, the solidification rate is controlled by heat flow, and chemical diffusion tries to follow. This eventually leads to the formation of microstructure, such as that shown in Fig. 3 (Ref 4). Several aspects of the solidification process are now considered as well as how transport phenomena are expressed in the development of microstructure.

Planar Front Growth

As discussed in other articles in this Handbook, the two processing variables that are used to correlate microstructures are the temperature gradient in the liquid at the interface, G_L, and

the solidification front speed, v_n^*. For this reason, a common method for either assessing microstructural features experimentally or actually controlling them in practice is directional solidification. In a simplified version of the experiment, a thin layer of alloy is encapsulated between microscope slides that are then placed atop controlled hot and cold blocks. The blocks are maintained at predetermined temperatures and at a fixed distance, in order to establish a desired temperature gradient, G_L. The slide assembly is then pulled at constant speed v_n^* through the gradient. The experiment is often done with a transparent organic alloy, and the solidification morphology is observed using a microscope. (Recently, similar experiments have been performed in metallic systems using synchrotron radiation to observe the microstructure (Ref 5). Next, a model of the system is built with the goal of finding out how the important parameters interact to produce the observed microstructural features shown in Fig. 3.

It is observed that if the pulling speed is slow enough, the interface remains flat. In this case, after a brief transient at the beginning of the experiment, the process reaches a quasi-steady state where the microstructure is constant in time. It is convenient to analyze the process in a frame that is fixed on the interface ($z = 0$) and to assume that in this frame the transport is independent of time. The material moves through the interface at velocity $-v_n^*$. It is assumed that the material properties are all constant (thus there is no shrinkage flow) and, for convenience, that the thermal diffusivities of the solid and liquid are equal. Finally, variations in all fields perpendicular to the pulling direction are ignored, making the problem one-dimensional. Note that all these assumptions are not essential to the analysis, and either everything could be included or the assumptions could be justified through a scaling analysis similar to that given for the quenched rod. For the sake of brevity, they are simply taken

Fig. 2 Temperature distributions through the midplane of a quenched rod for several values of the dimensionless time
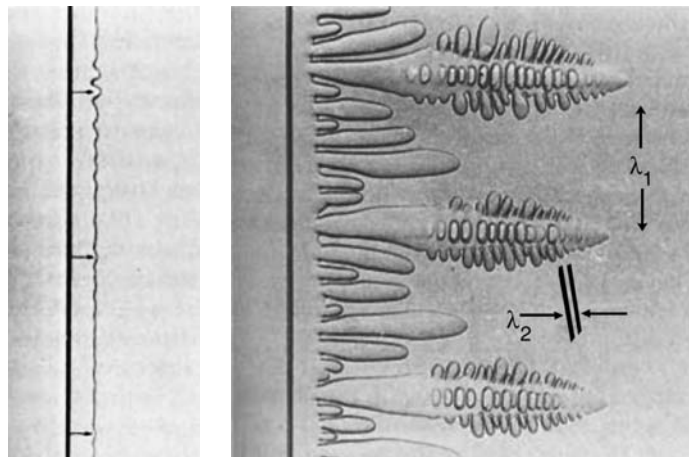
Fig. 3 Micrographs from directional solidification of a succinonitrile-acetone alloy. Left: initial planar interface with small perturbations; right: developing dendrites, with primary and secondary arms indicated. Source: Ref 4

as given. The transport equations for heat and solute now reduce to:

$$-v_n^* \frac{\partial T}{\partial z} = \alpha \frac{\partial^2 T}{\partial z^2} \quad -\infty < z < \infty \quad \text{(Eq 25)}$$

$$-v_n^* \frac{\partial C_L}{\partial z} = D_L \frac{\partial^2 C_L}{\partial z^2} \quad 0 \le z \quad \text{(Eq 26)}$$

$$-v_n^* \frac{\partial C_s}{\partial z} = D_s \frac{\partial^2 C_s}{\partial z^2} \quad z \le 0 \quad \text{(Eq 27)}$$

Looking at the micrograph in Fig. 3, it seems that the system chooses a length scale λ_1, the primary dendrite arm spacing. These equations are partially scaled by defining $\zeta = z/\lambda_1$, with the result:

$$-\left(\frac{v_n^* \lambda_1}{\alpha}\right) \frac{\partial T}{\partial \zeta} = \frac{\partial^2 T}{\partial \zeta^2} \quad \text{(Eq 28)}$$

$$-\left(\frac{v_n^* \lambda_1}{D_L}\right) \frac{\partial C_L}{\partial \zeta} = \frac{\partial^2 C_L}{\partial \zeta^2} \quad \text{(Eq 29)}$$

$$-\left(\frac{v_n^* \lambda_1}{D_S}\right) \frac{\partial C_S}{\partial \zeta} = \frac{\partial^2 C_S}{\partial \zeta^2} \quad \text{(Eq 30)}$$

Each equation has as a prefactor a variation of a dimensionless group called the Péclet number, corresponding to the ratio of advective to diffusive transport of heat or solute, depending on the equation. Actual values can be taken from the experiments for the material properties, v_n^* and λ_1, to evaluate the Péclet numbers:

$$\frac{v_n^* \lambda_1}{\alpha} \approx 10^{-2}; \quad \frac{v_n^* \lambda_1}{D_L} \approx 1; \quad \frac{v_n^* \lambda_1}{D_s} \approx 10^4 \quad \text{(Eq 31)}$$

Substituting these results into Eq 28 to 30 allows terms to be dropped that have coefficients that are much less than 1. It is also convenient to then redimensionalize the equations so that boundary conditions can be applied. The result is:

$$\frac{\partial^2 T}{\partial z^2} \approx 0 \quad \text{(Eq 32)}$$

$$-v_n^* \frac{\partial C_L}{\partial z} = D_L \frac{\partial^2 C_L}{\partial x^2} \quad \text{(Eq 33)}$$

$$-\frac{\partial C_S}{\partial x} \approx 0 \quad \text{(Eq 34)}$$

The solution to Eq 32 is a simple linear profile. Applying the boundary conditions $T = T_{cold}$ at $z = z_{cold}$ and $T = T_{hot}$ at $z = z_{hot}$ gives:

$$T = T_{cold} + \frac{T_{hot} - T_{cold}}{z_{hot} - z_{cold}} (z - z_{cold}) \quad \text{(Eq 35)}$$

The solution clearly shows how the apparatus selects the temperature gradient. The result that the temperature field is linear and undisturbed by the moving interface is sometimes called the frozen temperature approximation, but it is really a result of scaling.

The solution for the composition field in the liquid can be found after applying the boundary condition given in Eq 16 at $z = 0$ and setting $C_L = C_0$ for $z \to \infty$:

$$C_L = C_0 + C_0 \left(\frac{1-k}{k}\right) e^{-v_n^* z/D_L} \quad \text{(Eq 36)}$$

Thus, there is a solute boundary layer ahead of the interface of thickness roughly $3D_L/v_n^*$. The solution to Eq 33 is just C_s is constant. Examination of Eq 36 shows that $C_L = C_0/k$ at $z = 0$, and this permits evaluation of the constant, so that:

$$C_s = C_0 \quad \text{(Eq 37)}$$

This example, although rather simple, indicates how the transport phenomena naturally separate into different length scales due to the disparity in the rate of transport of energy and solute. This is very important for understanding the origin of microstructure during solidification. It is beyond the scope of this article to pursue the analysis much further, but it is important to understand that there is one other important aspect. On considering the stability of the planar front growth morphology, it is found that for small values of the ratio G_L/v_n^*, the interface becomes unstable to perturbations and breaks down first into cells and then into dendrites. In the stability analysis, a length scale associated with the surface energy of the liquid-solid interface enters the problem through a material property called the capillary length, $d_0 = \Gamma k/m_L C_0 (1 - k)$, where Γ is the Gibbs-Thomson coefficient, and m_L is the slope of the liquidus line. The characteristic length selected by the dendritic arms can be correlated to the geometric mean of the two length scales, that is:

$$\lambda \sim \sqrt{d_0 (D_L/v_n)}$$

Microsegregation

After the initially flat interface becomes unstable, it evolves into the dendritic pattern shown in the micrograph in Fig. 3. The structure is characterized by the primary (λ_1) and secondary (λ_2) dendrite arm spacings. In the previous analysis of transport, λ_1 is chosen as the length scale. At a finer scale, such as in the interdendritic region between secondary arms, λ_2 would be the appropriate length scale. Figure 4 illustrates the geometry.

The fundamental concepts of scaling can be applied to this new problem. As illustrated, the array of secondary arms is approximately periodic, and a control volume can be chosen as shown, with dimensions $\lambda_2 \times L$ and the expectation that $\lambda_2 \ll L$. This time, there is no velocity, because the control volume is fixed in the dendrite array. Writing the governing equations for energy and solute within this volume yields:

$$\frac{\partial T}{\partial t} = \alpha \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial T}{\partial z}\right) \quad \text{(Eq 38)}$$
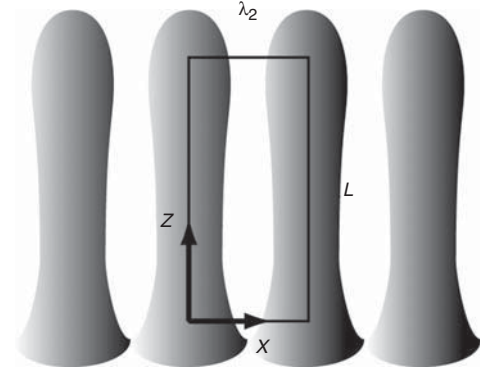
Fig. 4 Schematic drawing of the local geometry of secondary dendrite arms

$$\frac{\partial C_L}{\partial t} = D_L \left(\frac{\partial^2 C_L}{\partial x^2} + \frac{\partial C_L}{\partial z}\right) \quad \text{(Eq 39)}$$

$$\frac{\partial C_s}{\partial t} = D_s \left(\frac{\partial^2 C_s}{\partial x^2} + \frac{\partial C_s}{\partial z}\right) \quad \text{(Eq 40)}$$

Next, some scaled variables are defined:

$$\xi = \frac{x}{\lambda_2/2}; \quad \zeta = \frac{z}{L}; \quad \tau = \frac{t}{t_f};$$
$$\theta = \frac{T - T_{sol}}{T_{liq} - T_{sol}} \quad \text{(Eq 41)}$$

where t_f is the local freezing time, that is, the time it takes to go from the liquidus temperature, T_{liq}, to the solidus, T_{sol}. Substituting these scaled variables into the governing equations and invoking the assumption that $\lambda_2 \ll L$ gives:

$$\frac{\partial \theta}{\partial \tau} = \frac{4\alpha t_f}{\lambda_2^2} \frac{\partial^2 \theta}{\partial \xi^2} \quad \text{(Eq 42)}$$

$$\frac{\partial C_L}{\partial \tau} = \frac{4D_L t_f}{\lambda_2^2} \frac{\partial^2 C_L}{\partial \xi^2} \quad \text{(Eq 43)}$$

$$\frac{\partial C_s}{\partial \tau} = \frac{4D_s t_f}{\lambda_2^2} \frac{\partial^2 C_s}{\partial \xi^2} \quad \text{(Eq 44)}$$

Metallurgists often use measured correlations between λ_2 and t_f, which show that for aluminum alloys (Ref 6):

$$\lambda_2 \approx \left(10^{-5} \text{m/s}^{1/3}\right) t_f^{1/3} \quad \text{(Eq 45)}$$

For these same alloys, $\alpha \approx 10^{-6}$ m²/s, $D_L \approx 10^{-10}$ m²/s, and $D_s \approx 10^{-14}$ m²/s. Combining these data with Eq 45, one finds that for solidification times between approximately 10 s and 1 day, Eq 42 to 44 reduce to:

$$\frac{\partial^2 \theta}{\partial \xi^2} \approx 0 \quad \text{(Eq 46)}$$

$$\frac{\partial^2 C_L}{\partial \xi^2} \approx 0 \quad \text{(Eq 47)}$$

$$\frac{\partial C_s}{\partial \tau} \approx 0 \quad \text{(Eq 48)}$$

The dendritic array is periodic, so the derivatives $\partial \theta/\partial \xi$ and $\partial C_L/\partial \xi$ must be zero at both

$\xi = 0$ and $\xi = 1$. From this, it is concluded that both the temperature and the composition in the liquid are constant, and that the composition in the solid does not change with time. However, the composition at the liquid-solid interface must adjust to maintain local equilibrium. Applying the solute balance at the interface from Eq 16 gives a relation between the composition of the solid, C_s, and the solid fraction, f_s, known as the Scheil equation:

$$C_s = kC_0(1 - f_s)^{k-1} \qquad \text{(Eq 49)}$$

(This is probably not completely obvious from what has been presented here—see Ref 1 or 7 for a more thorough derivation.) Equation 49 can be recast in terms of temperature using the phase diagram, with the result:

$$f_s = 1 - \left(\frac{T - T_f}{T_{liq} - T_f}\right)^{1/(k-1)} \qquad \text{(Eq 50)}$$

This equation is very important, because it provides a very good estimate for the microsegregation in cast alloys.

Homogenization

The microsegregation described in the previous section is inevitable in normal casting practices. This is often undesirable, and many castings are heat treated to reduce the segregation. A detailed analysis of the problem could be done, but it really is not necessary to do so if one wishes only to estimate the proper homogenization time. It is known that the segregation is more or less periodic on the length scale of the secondary dendrite arm spacing. Based on scaling, then, the estimated time needed to homogenize the structure is simply λ_2^2/D_s. A conservative approach would be to homogenize the structure for two to three times this value.

This implies that the microstructure, in the form of λ_2, is known. From where does that come? It is correlated with the local solidification time, as in Eq 45, but how does one obtain t_f? This comes from solving the transport equation for heat at the macroscopic scale, for example, using a commercial casting solidification analysis package. Note that there is also a connection between the microsegregation and heat flow in the code, because the user usually must specify the enthalpy as a function of temperature, which is obtained from Eq 50 from the definition:

$$H(T) = \int_{298}^{T} c_p dT + \Delta H(1 - f_s) \qquad \text{(Eq 51)}$$

The integral begins at 298 K (77 °F), where enthalpy is taken to be zero by convention.

Summary

The governing equations for transport based on balance of mass, momentum, energy, and solute were written. Equations for these balances were developed for a moving solidification front. It was found that the differences in density, enthalpy, and composition between the liquid and solid are very important for the understanding of solidification processes. Considerable effort was spent to understand how material properties and geometry can be analyzed in the context of the governing equations, a process called scaling.

Several example problems were considered to show how the hierarchy of time and length scales associated with transport leads to important features of cast microstructures. An analysis of planar front growth revealed the important length scale associated with diffusion D_L/v_n^*. On the scale of secondary dendrite arms, the limited transport of solute in the solid, combined with rapid transport in the liquid, leads to the Scheil equation for microsegregation. Finally, the principles of transport and scaling are applied to homogenization processes to ameliorate the undesirable segregation pattern.

REFERENCES

1. J.A. Dantzig and C.L. Tucker III, *Modeling in Materials Processing*, Cambridge University Press, New York, 2001
2. R.B. Bird, W.E. Stewart, and E.N. Lightfoot, *Transport Phenomena*, 2nd ed., Wiley, New York, 2002
3. H.S. Carslaw and J.C. Jaeger, *Conduction of Heat in Solids*, 2nd ed., Oxford University Press, London, 1959
4. R. Trivedi and K. Somboonsuk, Constrained Dendritic Growth and Spacing, *Mater. Sci. Eng.*, Vol 65, 1984, p 65–74
5. R. Mathiesen and L. Arnberg, X-Ray Monitoring of Solidification Phenomena in Al-Cu Alloys, *Mater. Sci. Forum*, Vol 508, 2006, p 69–75
6. T.F. Bower, H.D. Brody, and M.C. Flemings, Measurements of Solute Redistribution in Dendritic Solidification, *Trans. AIME*, Vol 236, 1966, p 624
7. W.F. Kurz and D.J. Fisher, *Fundamentals of Solidification*, Trans-Tech, Aedermannsdorf, 1984

Modeling of Vapor-Phase Processes

Alain Dollet, Processes, Materials and Solar Energy Laboratory (PROMES) CNRS, France

VAPOR-PHASE PROCESSES (VPP) involve the building of a solid material (usually a coating) from elementary growth units that are transported by gaseous molecules or atoms onto a bulk substrate. Today (2009), VPP are widely used in industry for preparing semiconducting, insulating, or metallic solid films for microelectronic or photovoltaic devices, for protection of bulk materials against corrosion or wear, or in optical applications (e.g., spectrally selective coatings); powders and nanostructures can also be prepared. Vapor deposition processes are traditionally divided into physical vapor deposition and chemical vapor deposition processes, as further discussed in this article. Early vapor deposition equipment has been set up and optimized after long trial-and-error procedures or by using empirical approaches, but significant progress has been made in the design and optimization of vapor deposition processes during the last 25 years by using numerical models.

The goal of this article is twofold: First, it introduces the underlying physics and the basics of VPP modeling, with a special emphasis on transport phenomena; second, examples of vapor-phase models, from pioneer to state-of-the-art models, are illustrated through representative works. The article mainly focuses on phenomena and modeling approaches that are specific to VPP; however, most of the complementary information needed on general or specialized topics of interest (nucleation, microstructure evolution, Monte Carlo methods, etc.) is found in the cited references or in other articles in this Handbook.

Vapor-Phase Processes for the Synthesis of Materials

Vapor-phase processes are versatile processes that can be used for preparing a wide number of materials, such as metallic (tungsten, copper, aluminum, gold, platinum, etc.), semiconducting (silicon, germanium, diamond carbon, GaAs, CdTe, etc.), dielectric (SiO_2, Si_3N_4, etc.), and hard or refractory (SiC, ZrO_2, Al_2O_3, ZrO_2, etc.) films. Complex materials such as ternary or quaternary alloys (AlGaAs,

InAlGaN, etc.) and even metastable materials (nonstoichiometric, amorphous) can be prepared from the vapor phase. Various substrates have been used for growing materials, for example, single-crystal or polycrystalline wafers, fibers, porous materials, particles, and so on. Sometimes, solid particles can even be grown in the gas phase without solid substrates (homogeneous nucleation). Growth is usually performed under vacuum conditions and in open systems, although a small number of deposition processes operate at or slightly above atmospheric pressure and sometimes in closed systems. The gaseous precursor molecules are provided either directly from high-pressure cylinders or from liquid tanks by evaporation or from solid sources by sublimation, physical ablation, and so on. Depending on the phenomena governing the production and transport of growth units to the substrate, distinction is made between chemical vapor deposition (CVD) and physical vapor deposition (PVD). Physical vapor deposition refers to processes producing coatings from pure condensation onto a substrate, whereas CVD refers to processes producing coatings from chemical reactions (Ref 1). Several processes derived from PVD involve chemical reactions as well; they are sometimes referred to as hybrid vapor-phase processes.

Obtaining uniform coatings with controlled composition and thickness, in a reproducible way, at the highest deposition rate and at the lowest possible cost is one of the most important objectives when using industrial VPP for commercial applications. Several CVD or PVD methods may be selected for addressing this issue, and the results obtained often differ markedly depending on the exact nature of the desired material. Modeling and simulation are powerful tools that can be used for better design and optimization of the selected CVD or PVD equipment.

Chemical Vapor Deposition and Related Processes

In a typical thermal CVD process, gaseous reactants are introduced in the reaction chamber, where thermal energy initiates gas phase

and surface reactions. If thermodynamic conditions are favorable (see section "Physical Vapor Deposition and Related Processes" in this article), chemical reactions lead to the formation of a stable solid film on a heated substrate. Reaction products and nonconverted reactants are pumped outward from the reactor. By accurately controlling and monitoring the deposition conditions, it is possible to grow solid films with the desired composition, structure, and thickness. In addition to accurate control of the deposition parameters, the design of the reactor is also of major importance in CVD, because its geometry strongly influences vapor transport and film growth.

Although its principle is quite simple, CVD involves many complex phenomena that are often strongly coupled. Depicted in Fig. 1 is a schematic representation of the elementary steps typically involved in CVD, which has been quite well accepted for approximately 40 years. The gas phase usually undergoes significant physical and chemical changes because of homogeneous and heterogeneous reactions.

These changes are also strongly dependent on momentum and heat-transfer phenomena, which are themselves strongly coupled to mass transport. Numerous surface processes take place at the gas-solid interface, where gaseous species are transported and adsorbed. Adsorbed species may diffuse to the highest energy sites at the surface or join existing nuclei before being incorporated into the growing film. By-products or nonreacting adsorbed molecules desorb from the film surface before being transported away from the reactor with other by-products formed in the gas phase. Interdiffusion of atoms between film and substrate is usually negligible at low or moderate temperature but may become very important at high temperature. As illustrated in Fig. 1, CVD involves so many complex coupled phenomena that it is almost impossible to have a comprehensive view of the influence of each phenomenon on deposition without using a predictive simulation model in addition to experimental measurements.

Variants of CVD Processes. Many variants of conventional thermal CVD processes have been proposed and are often used in industry
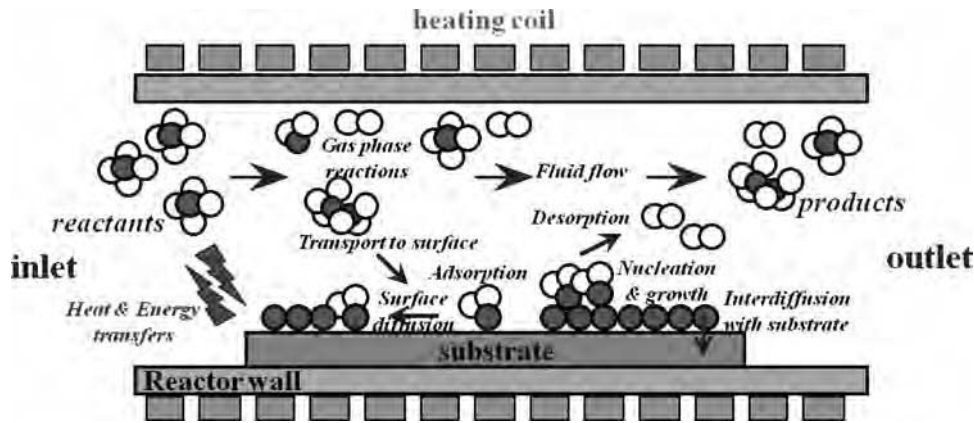
Fig. 1 Elementary steps involved in chemical vapor deposition

for producing various types of films. Sometimes, the only difference arises from the operating pressure range; some authors refer to atmospheric-pressure CVD (APCVD), low-pressure CVD (LPCVD), or ultrahigh-vacuum CVD (UHVCVD), although these processes remain basically conventional CVD. In some cases, the processes only differ from the conventional one by the nature of reactants or substrates. For example, MOCVD refers to a CVD process using organometallic precursors. Chemical vapor infiltration (CVI) consists of performing deposition inside porous preforms or foams. Fluidized-bed CVD (FB-CVD) is a special case of deposition on particles forming a bed that expands when swept by a gas mixture. These modified CVD processes are often referred to as assisted CVD processes. In most cases, thermal energy is replaced by another activation source in order to initiate chemical reactions, or another activation process is added to thermal activation. Among the most famous assisted CVD processes are plasma-enhanced chemical vapor deposition (PECVD) or plasma-assisted CVD, where activation is done through an electrical discharge in the gas), photo-assisted CVD (chemical reactions are initiated by photons), flame-assisted CVD (chemical precursors, often in a liquid state, are injected into a flame, where chemical reactions or combustion occurs), aerosol-assisted CVD (precursors are atomized into submicrometer liquid droplets), and atomic layer deposition (deposition is performed by sequentially saturating surface conditions to form only one monolayer per reaction sequence).

Physical Vapor Deposition and Related Processes

In most PVD processes, film growth units are single atoms or ions emitted from a solid source or target that is simply sublimated by heating or sputtered by energetic ion beam (liquid sources can also be used and evaporated by heating). For deposition of binary, ternary, and so on alloys, several solid sources may be used at the same time in the deposition chamber, or alternatively, a single crucible of specific composition can be prepared; in both cases, control of film composition often turns out to be difficult because the evaporation or sputtering rates are not equal for the various elements. The temperature of solid sources as well as the energy and flux of ions must be accurately controlled because energetic ions may release important amounts of energy, displace bulk atoms, or cause ablation of substrate surface atoms, which may affect film composition and properties. Conversely, substrate heating is usually not mandatory for deposition. Collisions in the gas phase are usually negligible in PVD systems operating under ultrahigh vacuum but not completely in systems operating in the pressure range of 0.1 to 1 Pa, in the transition regime (see the section "Continuous versus Rarefied Gas Flows" in this article). At very low pressure, the exact geometry of the chamber is not usually a crucial point, except for the position of the substrate respective to the solid crucibles, because the flux of depositing species is usually highly directional.

Most of the elementary steps depicted in Fig. 1 for CVD are involved in PVD, except that there is usually no thermal activation of the vapor phase, that is, no gas-phase reactions and no (or few) chemically driven desorption or adsorption steps at the film surface. Hence, growth is mainly driven by a pure condensation process rather than by reactions.

Main Types of PVD Processes. Evaporation is one of the most widely used PVD processes. Source materials are simply vaporized by resistance, radiation, or electron-beam heating in an ultrahigh vacuum (UHV) chamber; then, the evaporated particles impinge the growing film surface with a rather low energy. A special case of evaporation is molecular beam epitaxy (MBE), in which a collision-less beam obtained by evaporation or sublimation of a source in a UHV chamber is used to grow a high-purity epitaxial film on a single-crystal substrate.

In sputtering processes, a solid crucible is bombarded by energetic ions of a few hundreds eV to a few keV ($1 \ eV \approx 1.6 \times 10^{-19}$ J) that transfer part of their energy by a cascade of collisions inside the crucible material to atoms of the very first upper layers, which can then be ejected. In planar diode sputtering, ions (usually Ar^+) are formed in a direct current (dc) plasma discharge between the target electrode and a biased substrate. The atoms ejected from the target surface (with typical energy of a few eV to several tens of eV) move toward the substrate but lose a large part of their energy in collisions with gaseous molecules, because the pressure is typically of a few Pa. Ionization efficiency is not very good in dc sputtering systems, and the resulting deposition rate is usually rather low. Moreover, sputtering of insulating materials cannot be performed in dc systems, and radio-frequency plasmas must be generated in this case. To overcome the limitations of dc systems, magnetron sputtering processes have been developed that consist of superimposing a magnetic field to the electric field so that electrons are constrained to follow the magnetic field lines (spirals). Much higher ion densities are obtained in magnetron sputtering, then higher deposition rates, as compared to dc sputtering.

Ion-assisted deposition processes are also used, such as ion plating, which combines electron-beam evaporation of the coating material and a plasma discharge at a pressure of approximately 0.1 to 1 Pa. Ions generated from the sputtered atoms also participate in the growth process but release energy at the film surface, which is usually beneficial to adhesion and other film properties.

Laser ablation is a quite complicated but efficient process that can be advantageously used for depositing alloys. A pulsed laser beam is used in this case to extract atoms from a solid target; a plasma plume forms, and the depositing atoms move toward the substrate, where film growth occurs.

Several hybrid processes combining CVD and PVD features have also been developed, such as ion-beam-induced chemical vapor deposition, where the main role of the ion beam is to dissociate a gaseous precursor molecule (usually an organometallic compound). A well-known and widely used hybrid CVD-PVD process is reactive magnetron sputtering. In this case, some of the film components, such as metals, are sputtered from a solid target, whereas a reactive gas mixture (e.g., O_2, N_2, CH_4) is used to provide the other chemical elements. A large variety of ceramic materials, such as carbides, oxides, or nitrides, have been prepared using reactive magnetron sputtering. In the hybrid gas-source MBE process, reactive gases are used in addition to solid sources, providing, for example, a more precise control of the dopant incorporation compared to conventional MBE.

PVD versus CVD

It is not relevant to claim that CVD is superior to PVD or vice versa; both families of the process have advantages and drawbacks, and it is difficult to find a rule of thumb to perform a choice. The main drawbacks and limitations of PVD compared to CVD are:

- PVD processes require high-vacuum equipment and often expensive power supplies.
- Controlling film stoichiometry is often difficult (except with laser ablation).
- Obtaining uniform coatings on complex-shaped substrate is usually problematic with line-of-sight PVD processes.

Conversely, the limitations and drawbacks of CVD are:

- Hazardous (toxic or flammable) chemical precursors are sometimes used.
- Complex chemistry and coupled transport phenomena may lead to difficulty in controlling or understanding the key steps of the process.
- Deposition on thermally sensitive substrates is sometimes impossible (except with PECVD, photo-CVD, and sometimes MOCVD).

It is out of the scope of this article, dedicated to modeling and simulation, to further describe technical aspects of CVD, PVD, and related processes. The reader may refer to several excellent books or reviews on these topics, for example, in Ref 2 to 6. However, prior to addressing the modeling of VPP, it is of fundamental importance to know and understand the underlying physics. The two subsequent sections introduce most of the basic notions of importance for modeling transport phenomena as well as reactions and surface processes.

Transport Regimes and Transport Equations

Transport phenomena in gases are of major importance for VPP modeling. This section reviews some basic notions of molecular collisions and gas flows, and presents the transport equations in gases.

Basics of Molecular Collisions

In a sample of pure gas, two moving molecules, approximated by hard spheres of diameter d, approaching each other collide when the distance between their centers decreases to d. A total collision cross section (σ_T) is defined for the colliding pair of molecules as:

$$\sigma_T = \pi d^2 \qquad \text{(Eq 1)}$$

Let n be the number of gas molecules per unit volume, so that the mean molecular spacing is

$\cong n^{-1/3}$. For a dilute gas, this spacing is $>>d$; hence, most collisions are binary collisions. For collisions between molecules of species i and molecules of species j in a mixture, Eq 1 becomes:

$$\sigma_{T_{ij}} = \pi d_{ij}^2, \text{ with } d_{ij} = \frac{(d_i + d_j)}{2} \qquad \text{(Eq 2)}$$

The relative velocity ($\mathbf{c_r}$) between two molecules of mass m_1 and m_2 traveling, respectively, at velocity $\mathbf{c_1}$ and $\mathbf{c_2}$ before colliding is:

$$\mathbf{c_r} = \mathbf{c_1} - \mathbf{c_2} \qquad \text{(Eq 3)}$$

After collision, these molecules have new velocities, $\mathbf{c_1}^*$ and $\mathbf{c_2}^*$, and relative velocity, $\mathbf{c_r}^*$:

$$\mathbf{c_r^*} = \mathbf{c_1^*} - \mathbf{c_2^*} \qquad \text{(Eq 4)}$$

In case of elastic collisions, momentum and energy are conserved, and it can be shown that $c_r^* = c_r$. It is almost straightforward to establish that the mean collision rate (ν), that is, the collision frequency in the gas phase, is:

$$\nu = n\overline{\sigma_T c_r} \qquad \text{(Eq 5)}$$

For the special case of hard spheres, Eq 5 becomes:

$$\nu = n\pi d^2 \overline{c_r} \qquad \text{(Eq 6)}$$

and for molecular collisions between species i and j in a mixture:

$$\nu_{i,j} = n_j \overline{\sigma_{T_{i,j}} c_{r_{i,j}}} \qquad \text{(Eq 7)}$$

The mean collision rate, ν_i, for molecules i is first obtained by summing over all collision partners j, then the mean collision rate, ν, for the mixture is obtained by averaging over all species i:

$$\nu_i = \sum_j n_j \overline{\sigma_{T_{i,j}} c_{r_{i,j}}} \qquad \text{(Eq 8a)}$$

$$\nu = \sum_i \left(\frac{n_i}{n}\right) \nu_i \qquad \text{(Eq 8b)}$$

The frequency of molecular bombardment per unit area (z) on one side of any stationary surface in an equilibrium gas is (Ref 7):

$$z = \frac{1}{4} n\overline{c'} \qquad \text{(Eq 9)}$$

The mean free path, λ, is the average distance traveled by a molecule between successive collisions. It is equal to the mean thermal speed, $\overline{c'}$ (Eq 21), divided by the collision frequency:

$$\lambda = \frac{\overline{c'}}{\nu} \qquad \text{(Eq 10)}$$

For a simple gas at equilibrium, it can be shown that (Ref 7):

$$\overline{c'} = \frac{\overline{c_r}}{\sqrt{2}} \qquad \text{(Eq 11)}$$

Hence, in the simplest case where gas molecules are approximated by hard spheres of diameter d and for an equilibrium gas:

$$\lambda = \frac{1}{\sqrt{2}\pi d^2 n} \qquad \text{(Eq 12)}$$

Expressions of λ in other cases can be found in molecular theories of gases (Ref 7, 8). Equation 12 can easily be generalized to gas mixtures. The molecular mean free path for species i is:

$$\lambda_i = \frac{1}{\sum_j \frac{n_j \overline{\sigma_{T_{i,j}} c_{i,j}}}{\overline{c_i'}}} \qquad \text{(Eq 13)}$$

and the mean free path for the mixture becomes:

$$\lambda = \sum_i \left(\frac{n_i}{n}\right) \lambda_i \qquad \text{(Eq 14)}$$

The hard sphere model is a useful but oversimplified molecular model for defining the interaction force (or potential) between particles. There is only a repulsive term, which is equal to zero for distances $r > d$ and equal to ∞ for $r = d$. Of course, more sophisticated and realistic models have been developed, often including both a repulsive and an attractive term, for example, Lennard-Jones potentials:

$$\Phi_{L-J} = \frac{K_1}{r^{\alpha_1}} + \frac{K_2}{r^{\alpha_2}} \qquad \text{(Eq 15)}$$

The 6-12 Lennard-Jones potentials ($\alpha_1 = 6$, $\alpha_2 = 12$) have been widely used.

In the general case, energy transfers must be considered in molecular collisions that no longer can be considered as elastic. This is especially true for high-energy particles or reacting molecules that usually exchange large amounts of energy in collisional processes.

Continuous versus Rarefied Gas Flows

The different flow regimes in gases are conventionally classified into three categories (Fig. 2) according to the value of the Knudsen number (Kn), a dimensionless number defined as the ratio of the mean free path of molecules in the gas phase (λ) to a characteristic system dimension (D):

$$Kn = \lambda/D \qquad \text{(Eq 16)}$$

For an equilibrium gas composed of hard spheres of diameter d, λ is given by Eq 12.

At sufficiently high operating pressure and for rather large characteristic dimensions (reactor length, wafer diameter or spacing, etc.), Kn is small (<0.01) and the gas behaves as a continuous medium. Conversely, at very low pressure and for very short characteristic dimensions, Kn is large (>10), which means
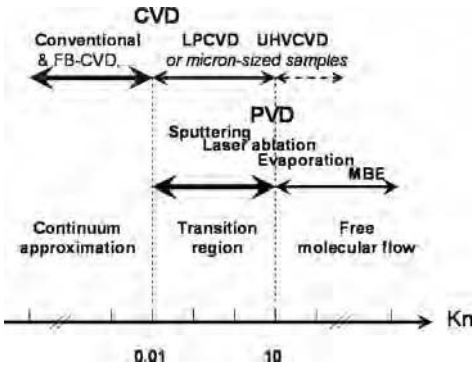
Fig. 2 Flow regimes in physical vapor deposition (PVD) and chemical vapor deposition (CVD) versus Knudsen number (Kn). FB-CVD, fluidized-bed CVD; LPCVD, low-pressure CVD; UHVCVD, ultrahigh-vacuum CVD; MBE, molecular beam epitaxy

that the gas phase is collision-less or in free molecular regime (wall collisions are much more frequent than gas-phase collisions). Between these two extreme regions, that is, for $0.01 < Kn < 10$, a transition region exists where the flow is neither continuous nor collision-less.

Many CVD processes operate at pressures greater than 10^1 to 10^2 Pa; hence, they can be modeled by using a continuum formulation. However, simulation of low- and very-low-pressure processes (LPCVD and UHVCVD) as well as simulation of growth on micrometer-sized features (very small D) must be based on noncontinuum equations. In fact, the last section of this article shows that a full modeling of CVD processes requires combining several models simulating different length scales (i.e., continuum and noncontinuum models). Unlike CVD processes, most PVD processes operate in the transition regime or in the free molecular regime; thus, PVD models are almost never based on continuum equations, at least for describing mass transport.

Transport Equations in Gases

Molecular Velocities and Velocity Distribution Functions. Consider a sample of gas in physical space containing n molecules per unit volume. A typical molecule has a velocity, $\mathbf{c}$, with components (u,v,w) in a Cartesian coordinate system (x,y,z). The fraction of molecules within a velocity space element $d\mathbf{c} = du\,dv\,dw$ is:

$$\frac{dn}{n} = \mathrm{f}(\mathbf{c})d\mathbf{c}, \text{ with } \int_{-\infty}^{+\infty} \mathrm{f}dc = 1 \qquad \text{(Eq 17)}$$

where $\mathrm{f}(\mathbf{c})$ is the single particle distribution function in velocity space. In case of a pure equilibrium gas at $T > 0$ with no macroscopic motion (average mass velocity $\mathbf{v} = \bar{\mathbf{c}} = \mathbf{0}$), a thermal (or peculiar) molecular motion exists at the microscopic scale: Molecules move in

random directions with a thermal speed c', and c = c' in this particular case.

Macroscopic properties can be derived by calculating moments of the distribution function. The average value of a given molecular quantity, $\mathbf{Q}$, is:

$$\bar{\mathbf{Q}} = \int_{-\infty}^{+\infty} \mathbf{Q}\mathrm{f}(\mathbf{c})d\mathbf{c} \qquad \text{(Eq 18)}$$

For example, for $\mathbf{Q} = \mathbf{c}$:

$$\bar{\mathbf{c}} = \int_{-\infty}^{+\infty} \mathbf{c}\mathrm{f}(\mathbf{c})d\mathbf{c} \qquad \text{(Eq 19)}$$

From the Boltzmann equation presented in the following section, an expression of the equilibrium distribution function (f_0) for a nonmoving equilibrium gas can be established (Ref 7):

$$\mathrm{f}_0(\mathrm{c}') = \left(\frac{m}{2\pi k_\mathrm{B}T}\right)^{3/2} \exp\left(-\frac{m\mathrm{c}'^2}{2k_\mathrm{B}T}\right) \qquad \text{(Eq 20)}$$

with k_B the Boltzmann constant and m the molecular weight. Starting from Eq 20, the average thermal speed ($\bar{\mathrm{c}}'$) can be derived (Ref 7):

$$\bar{\mathrm{c}}' = \sqrt{\frac{8k_\mathrm{B}T}{\pi m}} \qquad \text{(Eq 21)}$$

The thermal velocity $\mathbf{c}'$ is more generally defined as the velocity, $\mathbf{c}$, of a molecule relative to the macroscopic gas velocity, $\mathbf{v}$ (i.e., $\mathbf{c}' = \mathbf{c} - \mathbf{v}$). Because $\bar{\mathbf{c}} = \mathbf{v}$, the mean value of $\mathbf{c}'$ in a pure gas is:

$$\bar{\mathbf{c}}' = \bar{\mathbf{c}} - \mathbf{v} = \mathbf{0} \qquad \text{(Eq 22)}$$

Note that Eq 22 holds for $\bar{\mathbf{c}}'$, which is a vector (bold character), but not for its modulus, $\bar{\mathrm{c}}'$, which is a scalar quantity defined in (Eq 21). In the case of a mixture containing s species, the mass average velocity, $\mathbf{v}$, is obtained from a weighted mean accounting for the different masses of the s species. The mean thermal velocity of a particular species i is then given by:

$$\overline{\mathbf{c'}_i} = \bar{\mathbf{c}}_i - \mathbf{v} \qquad \text{(Eq 23)}$$

where $\overline{\mathbf{c'}_i}$ is called the diffusion velocity of species i in the mixture.

The Boltzmann Equation. This well-known equation derived by Boltzmann establishes the relationships between the distribution functions and their related variables. For a simple dilute gas, the equation is written:

$$\frac{\partial}{\partial t}(n\mathrm{f}) + \mathbf{c}\cdot\frac{\partial}{\partial \mathbf{r}}(n\mathrm{f}) + \mathbf{F}\cdot\frac{\partial}{\partial \mathbf{c}}(n\mathrm{f})$$
$$= \int_{-\infty}^{+\infty}\int_0^{4\pi} n^2(\mathrm{f}^*\mathrm{f}_1^* - \mathrm{f}\,\mathrm{f}_1)c_r\sigma d\Omega dc_1 \qquad \text{(Eq 24)}$$

In this equation, binary collisions of molecules of class $\mathbf{c}$ with molecules of class $\mathbf{c_1}$ are considered.

c_r denotes the relative velocity of colliding particles ($\mathbf{c_r} = \mathbf{c} - \mathbf{c_1}$), and f_1 is the value of f at $\mathbf{c_1}$. Post-collision quantities are identified by asterisks symbols (*). $\sigma d\Omega$ is the differential collision cross section. The first term of the left part of Eq 24 is the total rate of change of the number of molecules per unit volume; the second term is the convection flow of molecules of class $\mathbf{c}$ across the surface of volume $d\mathbf{r}$ due to the velocity $\mathbf{c}$; the third term is the convection flow of molecules across the surface $d\mathbf{c}$, due to the external force per unit mass, $\mathbf{F}$. The right part of the equation is the collision term, that is, the total rate of increase of molecules of class $\mathbf{c}$ as a result of collisions.

In the case of a gas mixture containing s species, the Boltzmann equation for a given species p is:

$$\frac{\partial}{\partial t}(n_p\mathrm{f}_p) + \mathbf{c}_p\cdot\frac{\partial}{\partial \mathbf{r}}(n_p\mathrm{f}_p) + \mathbf{F}\cdot\frac{\partial}{\partial \mathbf{c}}(n_p\mathrm{f}_p)$$
$$= \sum_{q=1}^{s}\int_{-\infty}^{+\infty}\int_0^{4\pi} n_p n_q(\mathrm{f}_p^*\mathrm{f}_{1_q}^* - \mathrm{f}_p\mathrm{f}_{1_q})c_{r_{pq}}\sigma_{pq}d\Omega dc_{1_q}$$
$$\text{(Eq 25)}$$

In this equation, distribution functions are defined for each species. More general forms of the Boltzmann equation can be derived in order to account for the presence of internal degrees of freedom, which have not been considered in the previous equations.

It is rather straightforward to see that the main difficulty associated with the resolution of the Boltzmann equation is the collision term, even in the assumption of molecular chaos. A particular case of interest is, of course, the one for which collisions can be neglected; that is, when $Kn \gg 10$, because the right-side term simply drops out of the equation. Approximate solutions of the collision term have been proposed, such as in the BGK equation (Ref 7) where the collision term of Eq 24 is replaced by:

$$n\nu(\mathrm{f}_0 - \mathrm{f}) \qquad \text{(Eq 26)}$$

where ν is a collision frequency.

Even if the problem of the collision term can be circumvented, analytical solutions of the Boltzmann equations remain difficult to obtain for problems involving complex geometries, and direct numerical solutions are also very complicated to obtain for multidimensional flows (see the section "Direct Boltzmann Method in CFD" in this article). For a restricted set of problems in which the distribution function is perturbed by only a small amount from equilibrium, the Chapman-Enskog theory provides a solution of the Boltzmann equation. This theory is particularly used for the calculation of some transport parameters in ideal gases, such as viscosity, heat conductivity, or diffusion coefficients, as seen in the section "Some Important Macroscopic Properties of Vapors from the Molecular Model" in this article.

Macroscopic Conservation Equations in Pure Gases. Moles are usually used instead of molecules at the macroscopic scale. Molar

quantities are related to their molecular counterparts by the Avogadro's number $\mathcal{N}(\cong 6.022 \times 10^{23})$, including the universal ideal gas constant, R, which is simply related to the Boltzmann constant by $R = k_B \mathcal{N}$. Therefore, the product (nm), which is defined as the gas density (ρ) at the molecular level, can also be expressed as the product of the molar concentration (C) and the molar weight (M) of the gas. In most situations encountered in VPP, gases can be considered as ideal; thus, the following equation of state can be used for ρ:

$$\rho = \frac{PM}{RT} \qquad (Eq\ 27)$$

A moment of the Boltzmann equation can be obtained as for a moment of the distribution function (Eq 18), by multiplying the equation by the quantity $\mathbf{Q(c)}$ and integrating over all velocity space. By doing so, the transfer equation (also referred to as the equation of change) is obtained for $\mathbf{Q}$:

$$\frac{\partial}{\partial t}(n\bar{\mathbf{Q}}) + \nabla \cdot (n\mathbf{c}\bar{\mathbf{Q}}) - n\mathbf{F} \cdot \frac{\partial \bar{\mathbf{Q}}}{\partial \mathbf{c}} = \Delta[\bar{Q}] \qquad (Eq\ 28)$$

where the right-side term, called the collision integral, has replaced the collision term in Eq 24. It can be shown (Ref 7) that for $\mathbf{Q} = m$, $m\mathbf{c}$ and $1/2\ mc^2$, the collision integral is zero; mass, momentum, and energy of a molecule are collisional invariants. Replacing $\mathbf{Q}$ by these terms in Eq 28 leads to the conservation equations of gas dynamics. The first one is the conservation of mass, or continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \qquad (Eq\ 29)$$

in which $\bar{\mathbf{c}}$ has simply been replaced by the average mass velocity, $\mathbf{v}$. Introducing the substantial derivative, that is, the time derivative for a path following the fluid motion:

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \qquad (Eq\ 30)$$

Equation 30 may be rewritten as:

$$\frac{D\rho}{Dt} = -\rho(\nabla \cdot \mathbf{v}) = 0 \qquad (Eq\ 31)$$

The second equation of gas dynamics, which is a vector equation, is the conservation of momentum:

$$\frac{\partial}{\partial t}(\rho \mathbf{v}) + \nabla \cdot (\rho \mathbf{vv}) = \rho \frac{D\mathbf{v}}{Dt} = -\nabla P - \nabla \cdot \mathbf{\tau} + \rho \mathbf{g} \qquad (Eq\ 32)$$

where P is the scalar pressure, and $\mathbf{\tau}$ is the viscous stress tensor. In Eq 32, gravity $(\mathbf{g})$ has been considered as the sole external force $(\mathbf{F})$. In the particular case of a Newtonian fluid (i.e., for most low-density gases), the viscous stress tensor in Eq 32 becomes:

$$\mathbf{\tau} = -\mu(\nabla \mathbf{v} + (\nabla \mathbf{v})^T - \frac{2}{3}\mathbf{\delta}\nabla \cdot \mathbf{v}) \qquad (Eq\ 33)$$

where μ is the fluid viscosity, and $\mathbf{\delta}$ is the unit tensor (i.e., a tensor whose components are

unity for the diagonal elements and zero for the others). The superscript "T" is for the transpose of a tensor (or a dyadic product). For constant ρ and μ, Eq 32 becomes the famous Navier-Stokes equation:

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla P + \mu \nabla^2 \mathbf{v} + \rho \mathbf{g} \qquad (Eq\ 34)$$

The third equation of gas dynamics is the conservation of energy:

$$\rho \frac{D(\hat{U} + \hat{K})}{Dt} = -\nabla \cdot \mathbf{q} - \nabla \cdot (\mathbf{\pi} \cdot \mathbf{v}) + \rho(\mathbf{v} \cdot \mathbf{g}) \qquad (Eq\ 35)$$

where $\hat{U}$ and $\hat{K}$ are, respectively, the internal energy and the kinetic energy of the fluid per unit mass. $\mathbf{\pi}$ is the pressure tensor, defined as:

$$\mathbf{\pi} = \mathbf{\tau} + \mathbf{\delta}P \qquad (Eq\ 36)$$

For a nonmoving gas or for a coordinate system moving with the gas, heat is transferred by conduction according to Fourier's law:

$$\mathbf{q} = -k\nabla T \qquad (Eq\ 37)$$

where $\mathbf{q}$ is the heat flux, and k is the thermal conductivity of the fluid.

Equation 35 is often rewritten in terms of the absolute temperature and the heat capacity at constant pressure per unit mass $(\hat{C}_p)$ by using the following thermodynamic relation:

$$d\hat{U} = \hat{C}_p dT - PdV - T\left(\frac{\partial(1/\rho)}{\partial T}\right)_P dP \qquad (Eq\ 38)$$

to give:

$$\rho \hat{C}_p \frac{DT}{Dt} = -\nabla \cdot \mathbf{q} - \mathbf{\tau} : \nabla \mathbf{v} + \left(\frac{\partial \ln(1/\rho)}{\partial \ln T}\right)_P \frac{DP}{Dt} \qquad (Eq\ 39)$$

For ideal gases, the right-side term of Eq 39 reduces to DP/Dt, and the double-dot product, $\mathbf{\tau} : \nabla \mathbf{v}$, which represents the heat dissipation due to viscous forces, is generally negligible. For a monoatomic ideal gas, $\hat{C}_p$ is simply equal to $5R/2M$.

Note that neither radiant energy transfer nor heat sources (electrical or nuclear) have been considered in Eq 35 and 39. In most CVD and PVD problems, the gas may be assumed transparent to infrared radiation; hence, only radiation heat transfer between solid walls must be considered (through boundary conditions; see the section "Boundary Conditions" in this article), which in turn affects gas temperature and velocity fields.

Conservation Equations in Multicomponent Gaseous Mixtures. The mass and momentum conservation equations given by Eq 29 and 32 are valid for pure substances as well as for mixtures; in the latter case, however, an additional conservation equation must be considered for

each of the N species that may be involved in NR gas-phase reactions. For species i:

$$\frac{\partial}{\partial t}(\rho \omega_i) = -\nabla \cdot (\rho \mathbf{v}\omega_i) - \nabla \cdot \mathbf{j}_i + M_i \sum_{k=1}^{NR} \nu_{ik} R_k^g \qquad (Eq\ 40)$$

In this equation, ν_{ik} is the net stoichiometric coefficient of species i in gas-phase reaction k, the net molar reaction rate of which is R_k^g. The mass fraction (ω_i) and the mole fraction (x_i) of species i in the mixture are defined as:

$$\omega_i = \frac{\rho_i}{\rho} \qquad (Eq\ 41a)$$

$$x_i = \frac{C_i}{C} \qquad (Eq\ 41b)$$

For a mixture, the mean molar mass, M, used in Eq 27 can be simply obtained from:

$$\frac{1}{M} = \sum_{i=1}^{N} \frac{\omega_i}{M_i} = \frac{1}{\sum_{i=1}^{N} x_i M_i} \qquad (Eq\ 42)$$

Only $N - 1$ balance equations (Eq 40) are required (e.g., no equation needed for the dominant gas), because mass fractions and mole fractions satisfy:

$$\sum_{i=1}^{N} \omega_i = 1 \text{ and } \sum_{i=1}^{N} x_i = 1 \qquad (Eq\ 43)$$

The diffusive mass flux of species i ($\mathbf{j}_i$) is the sum of several contributions associated with different driving forces (concentration, pressure, and thermal gradients as well as forced diffusion). In most VPP, pressure diffusion is negligible, and forced diffusion must be considered only for charged species in the presence of an electric field (e.g., in gas discharges). Two contributions usually remain: ordinary (or concentration-driven) diffusion, $\mathbf{j}_i^C$, and thermal (or temperature-driven) diffusion, $\mathbf{j}_i^T$, which obey:

$$\mathbf{j}_i = \mathbf{j}_i^C + \mathbf{j}_i^T \left(\text{with } \sum_{i=1}^{N} \mathbf{j}_i^C = 0\right) \qquad (Eq\ 44)$$

The total mass flux of species i ($\mathbf{n}_i$) is related to the diffusion mass flux by the following relation:

$$\mathbf{n}_i = \rho_i \mathbf{v}_i = \rho_i \mathbf{v} + \mathbf{j}_i \qquad (Eq\ 45)$$

For ideal gas mixtures, mass (or mole) fractions are related to species fluxes by the Stefan-Maxwell relation (Ref 8):

$$\nabla \omega_i + \omega_i \frac{\nabla M}{M} = \frac{M}{\rho} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(\frac{\omega_i \mathbf{j}_j - \omega_j \mathbf{j}_i}{M_j \mathcal{D}_{ij}}\right) \qquad (Eq\ 46)$$

In Eq 46, a binary diffusion coefficient, $\mathcal{D}_{ij}$, independent of the mixture composition, is used instead of the generalized composition-

dependent diffusion coefficient, D_{ij}. From Eq 46, the ordinary diffusion flux, $\mathbf{j}_i^C$, is written as:

$$\mathbf{j}_i^C = -\rho\mathcal{D}_i^{\text{eff}}\nabla\omega_i - \rho\omega_i\mathcal{D}_i^{\text{eff}}\frac{\nabla M}{M}$$
$$+ M\omega_i\mathcal{D}_i^{\text{eff}}\sum_{\substack{j=1\\j\neq i}}^{N}\frac{\mathbf{j}_i^C}{M_j\mathcal{D}_{ij}}\left(\text{with }\mathcal{D}_i^{\text{eff}}=\frac{1}{\displaystyle\sum_{\substack{j=1\\j\neq i}}^{N}\frac{x_i}{\mathcal{D}_{ij}}}\right)$$

(Eq 47)

The thermodiffusion flux is expressed as:

$$\mathbf{j}_i^T = -D_i^T\frac{\nabla T}{T}, \text{ with } D_i^T = \sum_{\substack{j=1\\j\neq i}}^{N}\frac{C^2}{\rho}M_iM_jD_{ij}k_i^T$$

(Eq 48)

where D_i^T is the thermodiffusion coefficient, and k_i^T is the thermodiffusion ratio. In the particular case where species are present in small amounts relative to a carrier gas (dilute mixtures), simplified expressions can be used for ordinary and thermal diffusion fluxes:

$$\mathbf{j}_i = -\rho\mathcal{D}_{im}(\nabla x_i + k_{im}^T\nabla\ln T)$$

(Eq 49)

where $\mathcal{D}_{im}$ and k_{im}^T are, respectively, the binary coefficient and the thermodiffusion coefficient of species i in the mixture. Expressions for $\mathcal{D}_{im}$ are given in the section "Some Important Macroscopic Properties of Vapors from the Molecular Model" in this article.

Finally, the energy equation for a gas mixture becomes (Ref 10):

$$\rho\hat{C}_p\frac{DT}{Dt} = -\nabla\cdot\mathbf{q} - \boldsymbol{\tau}{:}\nabla\mathbf{v} + \left(\frac{\partial\ln(1/\rho)}{\partial\ln T}\right)_P\frac{DP}{Dt}$$
$$+ \sum_{i=1}^{N}\sum_{k=1}^{NR}\bar{H}_i\nu_{ik}R_k^g$$

(Eq 50)

where the energy flux, $\mathbf{q}$, relative to the mass average velocity has three different contributions:

$$\mathbf{q} = -k\nabla T + \sum_{i=1}^{N}\bar{H}_i\frac{\mathbf{j}_i}{M_i} - \nabla\cdot RT\sum_{i=1}^{N}\frac{D_i^T}{M_i}\frac{\nabla x_i}{x_i}$$

(Eq 51)

The second term of Eq 51 comprising $\bar{H}_i$ (the partial molar enthalpy of the ith species) is the heat flux caused by interdiffusion of species, and the third term is the Dufour effect or diffusion-thermo effect, which is most often negligible. The right-side term in Eq 50 is the heat flux due to chemical reactions.

Laminar and Turbulent Flows. The macroscopic differential equations of change presented previously are valid for laminar flows. The laminar regime is a stable regime characterized by regular streamlines and the absence of time fluctuations of local fluid velocities

when the flow is established. When the flow becomes turbulent, some important time fluctuations of local velocities (instabilities) are observed around their mean value, although the flow seems macroscopically established.

Several dimensionless numbers are used for identifying the flow regime, such as the Reynolds (Re), Grashof (Gr), and Rayleigh (Ra) numbers, defined as:

$$\text{Re} = \frac{\rho Lv}{\mu}$$

(Eq 52a)

$$\text{Gr} = \frac{L^3\rho^2 g\beta\Delta T}{\mu^2}$$

(Eq 52b)

$$\text{Ra} = \text{GrPr} \quad \left(\text{with Pr} = \frac{C_p\mu}{k}\right)$$

(Eq 52c)

where L is a characteristic length, and β is the coefficient of volume expansion ($\beta = 1/T$ for an ideal gas). The transport equations presented in the previous sections "Macroscopic Conservation Equations in Pure Gases" and "Conservation Equations in Multicomponent Gaseous Mixtures" are often rewritten in dimensionless form by introducing such numbers.

Laminar flows are typically observed for Re < 2000 and Gr or Ra < 10^8 to 10^9. The Reynolds number corresponds to the ratio of inertial forces to viscous forces (or momentum flux by convection to momentum flux by diffusion). The Grashof and Rayleigh numbers correspond to the ratios of buoyancy forces to viscous forces; they are used in free thermal convection problems.

Because the mean free path of the fluid remains low compared to the average size of the turbulent eddies, the conservation equations presented in the preceding sections are still valid in turbulent regime; in this case, however, the velocity components remain time-dependent and cannot be simplified. Time-smoothed equations of change are usually used to describe variables such as velocity, pressure, and so on in turbulent flows. These equations are similar to those used for laminar flows but contain both laminar and turbulent components, the latter being generally obtained from empirical formulae. Reynolds-averaged Navier-Stokes (RANS) formulations have been developed accordingly for modeling turbulent flows, such as k-ε models (Ref 9) in which additional balance equations are used for the turbulent kinetic energy (k) and viscous dissipation (ε). More sophisticated approaches, such as large-eddy simulations (LES) and direct numerical simulations (DNS), are also used for obtaining more accurate solutions of turbulent flow problems, but they usually require large or even huge computational resources. With a few exceptions, most flows encountered in VPP remain laminar; hence, the modeling of turbulent flows is not further discussed in this

section. There is abundant literature on turbulence modeling; the reader may see Ref 9 for an introduction.

Some Important Macroscopic Properties of Vapors from the Molecular Model. In continuum transport equations, the main flow variables (such as $\mathbf{v}$) and gas properties are macroscopic statistical quantities averaged over a large number of colliding molecules (ensemble average) and time much longer than the mean time between successive collisions (time average). Any kind of average can be used for describing an ergodic molecular motion. An important result from the analysis of molecular motions is the total flux of some quantity, $\mathbf{Q}$ (constant, scalar, or vector), across a small element of area (ΔS) in the gas:

$$\text{Total flux across }\Delta S = n\overline{\mathbf{Q}\cdot\mathbf{cn}_0}$$

(Eq 53)

where $\mathbf{c}$ is the molecular velocity, and $\mathbf{n}_0$ is the outwardly directed normal unit vector. By setting $Q = m$, the flux vector related to mass transport is obtained:

$$\text{Flux of mass transport} = nm\,\bar{\mathbf{c}} = \rho\bar{\mathbf{c}}$$

(Eq 54)

The gas density, ρ, has already been defined at the molecular level as:

$$\rho = nm$$

(Eq 55)

Flux vectors can also be defined for thermal velocities by replacing $\mathbf{c}$ by $\mathbf{c}'$ in Eq 54, that is, by using a coordinate system moving at velocity $\mathbf{c}$ with the fluid. Setting $Q = m\mathbf{c}'$ gives the flux of momentum transport by peculiar (thermal) motion, that is, the pressure tensor ($\boldsymbol{\pi}$):

$$\boldsymbol{\pi} = nm\,\overline{\mathbf{cc}'} = \rho\overline{\mathbf{cc}'}$$

(Eq 56)

The scalar pressure (P) is then defined as:

$$P = \frac{1}{3}\rho\overline{\mathbf{c}'^2}$$

(Eq 57)

and the components τ_{ij} of the viscous stress tensor ($\boldsymbol{\tau}$) in rectangular coordinates (x,y,z coordinates are identified with i or j = 1, 2, and 3):

$$\tau_{ij} = -\pi_{ij} + \delta_{ij}P = -(\rho\overline{c_i'c_j'} - \delta_{ij}P)$$

(Eq 58)

where δ_{ij} is the Kronecker delta.

In a gas mixture containing N species (i), Eq 55 and 57 must simply be replaced by:

$$\rho = \sum_{i=1}^{N}m_in_i = \sum_{i=1}^{N}\rho_i = n\bar{m}$$

(Eq 59)

$$P = \sum_{i=1}^{N}\frac{1}{3}m_in_i\,\overline{c_i'^2} = \frac{1}{3}n\overline{mc'^2}$$

(Eq 60)

Finally, the heat flux vector, q, is obtained by setting Q = molecular energy = $\frac{1}{2}mc'^2 + \varepsilon_{\text{int}}$:

$$\mathbf{q} = \frac{1}{2}\rho\overline{c'^2\mathbf{c'}} + n\overline{\epsilon_{int}\mathbf{c'}} \qquad \text{(Eq 61)}$$

where $e_{int} = \overline{\epsilon_{int}}/m$, with ϵ_{int} the internal energy of a molecule.

The macroscopic temperature, T, used in the macroscopic equations of change is also defined from a molecular level; for a molecule with thermal speed c', the translational kinetic temperature, T_{tr}, is:

$$\frac{3}{2}k_B T_{tr} = \frac{1}{2}\overline{c'^2} \qquad \text{(Eq 62)}$$

T_{tr} is also the macroscopic temperature for a monoatomic gas or an ideal dilute gas. In case of a polyatomic molecule with s internal degrees of freedom and internal energy ϵ_{int}, a temperature T_{int} is also defined for the internal modes, as in Eq 62 (replacing $3T_{tr}$ by sT_{int} and $\frac{1}{2}\overline{c'^2}$ by e_{int}). As a consequence, an overall kinetic temperature (T_{ov}) may be defined as the weighted mean of T_{int} and T_{tr}, according to:

$$T_{ov} = \frac{(3T_{tr} + sT_{int})}{3 + s} \qquad \text{(Eq 63)}$$

with $T_{tr} = T_{int} = T$ for an equilibrium gas. Equation 63 could also be extended to include rotational degrees of freedom in polyatomic molecules. However, note that the ideal gas equation of state does not apply to Eq 63 in a nonequilibrium situation, but Eq 62 can be used in this case.

The macroscopic transport coefficients used in the conservation equations can also be derived from molecular-level analysis. From the simplest case of a mixture composed of hard spheres of identical mass and diameter (Ref 10), the following expressions for μ, k, and D_{ij} are obtained:

$$\mu = \frac{1}{3}nm\lambda\overline{c'} \qquad \text{(Eq 64a)}$$

$$k = \frac{1}{2}nk_B\lambda\overline{c'} \qquad \text{(Eq 64b)}$$

$$\mathcal{D}_{ij} = \frac{1}{3}\lambda\overline{c'} \qquad \text{(Eq 64c)}$$

with $\overline{c'}$ given by Eq 21.

Rather simple expressions have also been established for more realistic situations involving spheres of different mass and diameter, although the derivation becomes more complex, but more accurate estimates are usually preferred. The Chapman-Enskog theory provides a solution of the Boltzmann equation for the particular case where the distribution function, f, is slightly perturbed from its equilibrium value (f_0) (Ref 7). To express f in the form of a power series:

$$f = f_0(1 + \Phi_1 + \Phi_2 + \Phi_3 + \dots) \qquad \text{(Eq 65)}$$

The Chapman-Enskog solution of the Boltzmann equation has been obtained for the second-order development of the distribution function, and the following expressions of the transport coefficients have been derived, which are widely used for low-density gases:

$$\mu = 2.6693 \ 10^{-6}\frac{\sqrt{MT}}{\sigma^2\Omega_\mu} \qquad \text{(Eq 66a)}$$

$$k = 8.3228 \ 10^{-2}\frac{\sqrt{M/T}}{\sigma^2\Omega_k} \qquad \text{(Eq 66b)}$$

$$\mathcal{D}_{ij} = 1.8583 \ 10^{-7}\frac{\sqrt{T^3\left(\frac{1}{M_i} + \frac{1}{M_j}\right)}}{P\sigma_{ij}^2\Omega_{\mathcal{D}ij}} \qquad \text{(Eq 66c)}$$

where σ is the Lennard-Jones collision diameter of the species considered, and Ω is the collision integral ($\Omega_\mu = \Omega_k \neq \Omega_{\mathcal{D}ij}$), which is a slowly varying function of the dimensionless temperature $k_B T/\epsilon$, where ϵ is the energy of interaction in the 6–12 Lennard-Jones potential, φ:

$$\phi(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6\right] \qquad \text{(Eq 67)}$$

with r the intermolecular distance. The Lennard-Jones parameters (σ, ϵ), which are either measured or estimated from critical properties (Ref 10), are available for a large number of gases. Values of Ω can be found in tables (Ref 8, 10) or calculated from correlations as a function of $k_B T/\epsilon$ (Ref 11). The Lennard-Jones parameters for a colliding pair of molecules (i,j) are simply obtained from:

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j) \qquad \text{(Eq 68a)}$$

$$\epsilon_{ij} = (\epsilon_i\epsilon_j)^{1/2} \qquad \text{(Eq 68b)}$$

Transport coefficients in Eq 66(a-c) are in SI units, that is in Po, W/m$^{-1}\cdot$ K^{-1}, and m$^2\cdot$ s^{-1} for μ, k, and $\mathcal{D}_{ij}$, respectively. T is in K, but M is in g$\cdot$ mol^{-1}, σ is in Å, and P is in atm. Although derived for monoatomic gases, these equations turn out to be very good for polyatomic gases as well. They can also be extended to N-component gas mixtures. For μ and k, the Wilke semiempirical formula is used:

$$\mu = \sum_{i=1}^N \frac{x_i\mu_i}{\sum_{j=1}^N x_j\Phi_{ij}} \qquad \text{(Eq 69a)}$$

$$k = \sum_{i=1}^N \frac{k_i\mu_i}{\sum_{j=1}^N x_j\Phi_{ij}} \qquad \text{(Eq 69b)}$$

with

$$\Phi_{ij} = \frac{1}{\sqrt{8}}\left(1 + \frac{M_i}{M_j}\right)^{-1/2}\left[1 + \left(\frac{\mu_i}{\mu_j}\right)^{1/2}\left(\frac{M_j}{M_i}\right)^{1/4}\right]^2$$

For the binary diffusion coefficient of i in the mixture ($\mathcal{D}_{im}$), the following expressions can be used, respectively, for very low and low concentrations of components $i = 1,\dots N - 1$ in dominant species N:

$$\mathcal{D}_{im} = \mathcal{D}_{iN} \qquad \text{(Eq 70a)}$$

$$\mathcal{D}_{im} = \frac{(1 - x_i)}{\sum_{\substack{j=1 \\ j\neq i}}^N \frac{x_i}{\mathcal{D}_{ij}}} \qquad \text{(Eq 70b)}$$

Thermal diffusion coefficients in gaseous mixtures are temperature- and composition-dependent parameters that can be derived from Lennard-Jones parameters and transport collision integrals defined in the Chapman-Enskog theory (Ref 8, 12). For ideal gas mixtures, the following formula can be used, for instance:

$$D_i^T = -2.5910^{-7}T^{0.659}\left(\frac{M_i^{0.511}x_i}{\sum_j M_j^{0.511}x_j} - \omega_i\right)\left(\frac{\sum_j M_j^{0.511}x_j}{\sum_j M_j^{0.489}x_j}\right) \qquad \text{(Eq 71)}$$

where D_i^T is in SI units (kg $\cdot$ m$^{-1}\cdot$ s^{-1}), T is in K, and M_i and M_j are in g $\cdot$ mol^{-1}.

The physical and chemical properties of pure substances can usually be found in handbooks or in databases such as CHEMKIN (Ref 13) or the National Institute of Standards and Technology (NIST) (Ref 14), although transport property data are available only for a restricted set of compounds. Additional data and estimation methods can be found in reference books, for example, by Hirshfelder et al. (Ref 8) and Prausnitz et al. (Ref 15). Routines for the estimation of transport properties in mixtures are included in most of today's (2009) commercial simulation software.

Modeling of Surface Interactions with the Vapor Phase

As illustrated in Fig. 1, modeling of VPP involves not only transport processes but also surface processes and gas-phase reactions. This section describes the modeling of vapor-surface interactions and kinetics of hetereogeneous (gas-surface) processes. The next section, "Gas-Phase Reactions in CVD," covers the modeling and kinetics of homogenous reactions.

Simulating surface processes requires an accurate model of the surface interacting with the vapor. A simple microscopic representation of a perfect crystal, according to Kossel (Ref 16), considers terraces (T), ledges (L). and kink (K) sites at the surface (Fig. 3), depending on the number of first-nearest neighbors of the atomic site considered. Ideal crystal surfaces have various densities of steps, depending on their Miller indices (hkl). Some surfaces with low Miller indices theoretically have no steps below the so-called roughening temperature (Ref 17); they
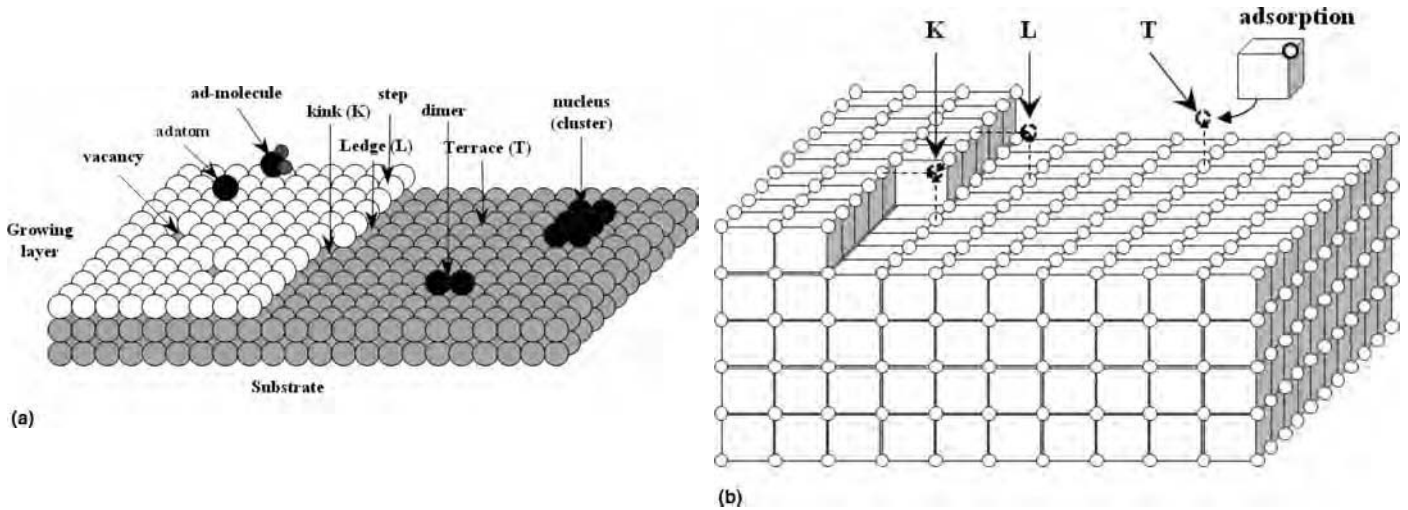
Fig. 3 Models of growing surface structures. (a) Conventional atomistic model. (b) Kossel model

are called singular faces and are flat on the atomic scale. Other faces offer a more-or-less important density of steps; they are called nonsingular faces and have a rough structure at the atomic scale. As a rule of thumb, the higher the density of steps, the stronger the surface reactivity.

Earlier simulation models of crystal growth considered these simple but useful representations of surfaces, but they have been encompassed by modern atomistic simulation models (see the section "Modeling and Computation of Particle-Surface Interactions" in this article), which use very accurate representations of surfaces. Most of today's (2009) commercial atomistic simulation software, for example, Materials Studio (Ref 18), have powerful graphical utilities for displaying surface structures of perfect crystals from crystallographic data; however, real surfaces are more complex than ideal ones because of crystallographic defects (stacking faults, dislocations, vacancies, etc.), reconstructions, or external contamination (water, oxygen, etc.). Modeling the real surfaces of growing films remains a difficult task in VPP simulation.

Particle-Surface Interactions

Accurate description of these interactions is crucial for the simulation of vapor deposition (or etching). The main surface processes in vapor deposition from a near-equilibrium gas are reflection or adsorption of gaseous particles, surface diffusion of adsorbed species, adsorbate incorporation into the film, and adsorbate desorption (Fig. 1). Chemical reactions of adsorbed species may lead to surface-atom removal instead of incorporation in the case of chemical etching processes.

Incoming Flux and Particle Reflection at a Surface. For a near-equilibrium gas of mean molecular weight M at temperature T and partial pressure P interacting with a solid surface,
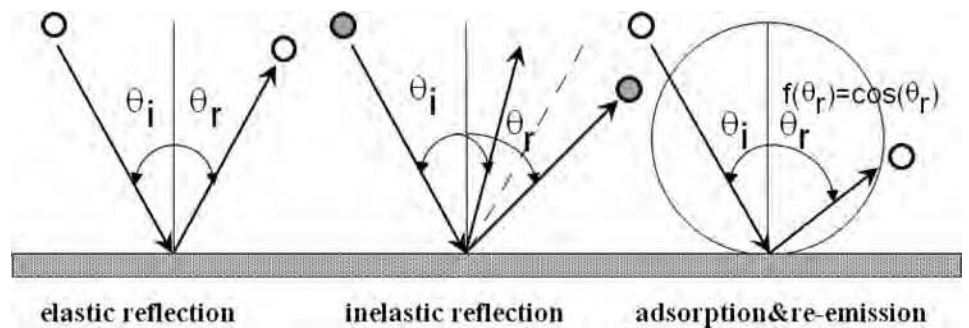
Fig. 4 Various possible reflection modes for a particle striking a surface

the macroscopic flux of particles impinging the surface is simply derived from gas kinetic theory, according to Eq 9:

$$F = \frac{P}{(2\pi MRT)^{0.5}} \qquad \text{(Eq 72)}$$

For a given species (i) in a gaseous mixture, the flux (F_i) is simply obtained by replacing P with the partial pressure P_i and M with M_i.

The thermal accommodation coefficient (α_c) is defined as:

$$\alpha_c = \frac{q_i - q_r}{q_i - q_s} \approx \frac{T_i - T_r}{T_i - T_s} \qquad \text{(Eq 73)}$$

where indexes i, r, and s denote, respectively, the incident particle, the reflected particle, and the solid surface; q and T denote the energy flux and the temperature. According to the value of α_c, the gaseous species incoming at the surface with an angle θ_i from the normal to the surface can be simply reflected without exchanging energy (specular elastic reflection: $\alpha_c = 0$, $T_r = T_i$, and $\theta_r = -\theta_i$) or being partly (inelastic reflection: $0 < \alpha_c < 1$, $T_r \neq T_i$, and $\theta_r \neq -\theta_i$) or completely thermalized (diffuse reflection: $\alpha_c = $

1, $T_r = T_s$, and cosine law for the space distribution of particles emitted from the surface, which is independent of θ_i). These various reflection modes are schematically illustrated in Fig. 4.

According to Maxwell, the distribution function of a gas in equilibrium with a solid surface remains a Maxwellian distribution either for specular or for diffuse reflection.

Some nonequilibrium processes, such as ion-beam-assisted deposition or ion beam etching (IBE), also involve energetic beams of particles that bombard the solid surfaces. For such line-of-sight nonequilibrium processes, the flux ($\mathbf{F}$) of particles emitted from a source of area, S_1, is calculated by integration of $dn\mathbf{c} \cdot \mathbf{ds_1}$ over the phase space domain, for example, from:

$$\mathbf{F} = \int_{c=0}^{\infty} \int_{\theta=0}^{\pi} \int_{\varphi=0}^{2\pi} n\mathbf{c} \cdot \mathbf{ds_1} f(\mathbf{c}, \theta) c^2 \sin\theta \, d\theta \, d\varphi \, dc$$

$$\text{(Eq 74)}$$

where $f(\mathbf{c}, \theta)$ is the particle distribution function in phase space, n is the particle density, and θ is the angle between the particle velocity and the

surface. $f(\mathbf{c},\theta)$ must be normalized to fulfill the condition:

$$\int_{c=0}^{\infty}\int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi} f(\mathbf{c},\theta)c^2 \sin\theta \, d\theta \, d\varphi \, dc = 1 \qquad \text{(Eq 75)}$$

In the case of IBE, for example, monoenergetic ion beams with narrow Gaussian angular distribution are usually considered:

$$f(\mathbf{c},\theta) = \delta(\mathbf{c}-\mathbf{c_0})g(\theta)$$
$$= \delta(\mathbf{c}-\mathbf{c_0})\frac{1}{\sigma_\theta\sqrt{2\pi}}\exp\left(-\frac{\theta^2}{2\sigma_\theta}\right) \qquad \text{(Eq 76)}$$

where δ is the Dirac delta function, $\mathbf{c_0}$ is the velocity of incident particles, and σ_θ is the angular divergence, which has a typical value of 5 to $10°$ for IBE.

Adsorption, Desorption, and Surface Diffusion. When its kinetic energy is well redistributed inside the solid, a particle impinging the surface can be temporarily captured as a result of van der Waals interactions. This process, leading to a weakly bounded adsorbed particle, is referred to as physisorption. The heat released in the physisorption process is Q_{ads}. The adsorbate-surface energy function, $\varphi_{ads\text{-}surf}$, can be written as a sum of Lennard-Jones potentials over the n surface atoms interacting with the adsorbate at a distance r_i:

$$\varphi_{ads\text{-}surf} = \sum_{i=1}^{n} 4\varepsilon\left[\left(\frac{\sigma}{r_i}\right)^{12} - \left(\frac{\sigma}{r_i}\right)^{6}\right] \qquad \text{(Eq 77)}$$

The molar rate of physisorption, $R\varphi_i$, of species i is proportional to the capture coefficient (β_i), which represents the fraction of incoming species flux (F_i) adsorbed:

$$R\varphi_i = \beta_i F_i = \frac{P_i}{(2\pi M_i RT)^{0.5}} \qquad \text{(Eq 78)}$$

Contrary to physisorption, chimisorption cannot occur in multiple layers. This adsorption process leads to formation of strong chemical bonds between the adatom or admolecule and the surface atoms; hence, Q_{ads} is much higher for chemisorption than for physisorption. Chimisorption may occur at monoatomic or polyatomic sites, sometimes with dissociation of the adsorbate. The adsorbate-surface interactions can be described by a sum of simple Morse potentials or more complex potentials. The molar rate of chimisorption, $R_{\chi}i$, of i is related to the sticking coefficient, s_i, by:

$$R_{\chi i} = \frac{s_i}{1-s_i/2}F_i = \frac{s_i}{1-s_i/2}\frac{P_i}{(2\pi M_i RT)^{0.5}} \qquad \text{(Eq 79)}$$

In the physisorbed state, a molecule moves almost freely at the surface (mobile diffusion), whereas in the chemisorbed state, it moves by jumping from one surface site to another neighbor surface site, breaking chemical bonds with the surface and forming new ones. Surface diffusion is a thermally activated process that proceeds with an activation energy of typically 0.1 to 0.3 Q_{ads}. Desorption is the reverse process of adsorption. An adsorbed molecule may desorb if its kinetic energy is greater than Q_{ads}.

The sticking coefficient defined in Eq 79 is a statistical average that represents the probability for a gaseous molecule to be adsorbed at a given surface. In the most general case, it depends on the particle and surface temperatures, on the surface structure, and also on the fraction of surface, θ, already covered by adsorbates. In the simplest situation, where adsorption of a given species A occurs on a single surface site, S, without dissociation, and neglecting adsorbate-adsorbate interactions, the rate of adsorption is $k_a[A](1-\theta_A)$, and the reverse rate of desorption is $k_{-a}\theta_A$, where k_a and k_{-a} are the adsorption and desorption rate constants, respectively. At equilibrium, both rates are equal, and θ can be simply related to the adsorption equilibrium constant, K_A ($= k_a/k_{-a}$), by the following equation:

$$\theta_A = \frac{K_A[A]}{1+K_A[A]} \qquad \text{(Eq 80)}$$

Equation 80 can be generalized to the case of adsorption of A in the presence of a nonreacting substance (poison or inhibitor, I) by simply adding $K_I[I]$ to the denominator. Expressions of the surface coverage can also be simply derived for dissociative adsorption or competitive adsorption of species A and B at the surface. In the last case, writing the rate equations and the equilibrium condition leads to (Ref 19):

$$\theta_A = \frac{K_A[A]}{1+K_A[A]+K_B[B]} \qquad \text{(Eq 81a)}$$

$$\theta_B = \frac{K_B[B]}{1+K_A[A]+K_B[B]} \qquad \text{(Eq 81b)}$$

Reactions. Gaseous or adsorbed species may eventually find favorable sites for reacting with the substrate surface (growth or etching). As discussed previously and illustrated in Fig. 3, real surfaces exhibit various types of sites with different reactivity. The global reaction rate of a given reactant species A at a surface can then be written as a sum of Arrhenius rate expressions:

$$v = [A]\sum_i A_i e^{-E_i/RT} \qquad \text{(Eq 82)}$$

where the A_i are the pre-exponential factors for the reaction of A at sites of type i (whose concentration is proportional to A_i), and the E_i are the corresponding activation energy. Activation energies are lower on those sites with higher activity; hence, assuming that the A_i have the same order of magnitude, and $E_1 << E_2 << \ldots E_i << \ldots << E_n$, the global rate of reaction reduces to:

$$v \approx A_1 e^{-E_1/RT}[A] \qquad \text{(Eq 83)}$$

Unimolecular and bimolecular reactions involving gaseous species (G) adsorbed at surface sites (S) are usually encountered. Unimolecular reactions simply involve a single gaseous molecule that first adsorbs at a monoatomic surface site:

$$G + S \underset{\rightarrow}{\leftarrow} (G-S)^{\ddagger} \rightarrow \text{Products} \qquad \text{(Eq 84)}$$
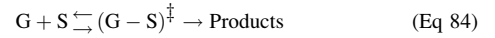
The symbol $\ddagger$ indicates that the adsorbate state is a transition state in the reaction. The corresponding rate equation is:

$$v = k\theta = \frac{kK_A[A]}{1+K_A[A]} \qquad \text{(Eq 85)}$$

When present in significant concentration in the gas phase, the reaction products (P) may behave as a poison and must also be considered in Eq 85 by adding $K_P[P]$ in the denominator (Ref 19). Bimolecular reactions involve either two gaseous molecules adsorbing at two adjacent surface sites (Langmuir-Hinshelwood mechanism):

$$G_1 + G_2 + -S-S- \underset{\rightarrow}{\leftarrow} \begin{pmatrix} G_1 & G_2 \\ -S- & S- \end{pmatrix}^{++} \rightarrow \text{Products}$$
$$\text{(Eq 86)}$$
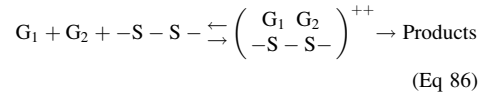
with the following reaction rate:

$$v = k\theta_{G_1}\theta_{G_2} = \frac{kK_{G_1}K_{G_2}[G_1][G_2]}{(1+K_{G_1}[G_1]+K_{G_2}[G_2])^2} \qquad \text{(Eq 87)}$$

or a gaseous molecule adsorbing at a single surface site (Langmuir-Rideal mechanism):

$$G_1 + G_2 + -S \underset{\rightarrow}{\leftarrow} (G_1-S+G_2)^{++} \rightarrow \text{Products}$$
$$\text{(Eq 88)}$$
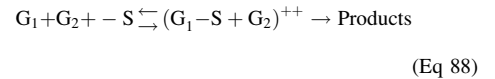
with the following reaction rate:

$$v = k\theta_{G_1}[G_2] = \frac{kK_{G_1}[G_1][G_2]}{1+K_{G_1}[G_1]+K_{G_2}[G_2]} \qquad \text{(Eq 89)}$$

The reaction rate coefficients and reaction orders are usually identified from experimental data. As an illustrative example, the following expression of the deposition rate (R_{Si}) of silicon from gaseous silane (SiH_4) has been empirically derived:

$$R_{Si} = \frac{kP_{SiH4}}{1+k_1 P_{H2}^{\alpha}+k_2 P_{SiH4}} \qquad \text{(Eq 90)}$$

Equation 90 corresponds to a unimolecular dissociation reaction of SiH_4 with inhibition by hydrogen:

$$SiH_4(g) \rightarrow Si(s) + 2H_2(g) \qquad \text{(Eq 91)}$$
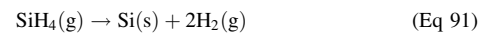
Various values of rate coefficients have been proposed for the heterogeneous decomposition of silane (Ref 20, 21). Alternatively, sticking

coefficient values have been proposed for this reaction (Ref 22) that do not account for the possible adsorption of H_2. Both kinds of rate equations have been used for modeling CVD of silicon from silane.

Interactions Involving Energetic Ions. As an energetic ion (e.g., in an ion beam) hits a solid surface, its energy in excess of the lattice binding energy may be transferred to an atom in the solid, which is removed from its original site; this atom can also transfer energy to other lattice atoms, giving rise to a so-called collision cascade. Some atoms can be ejected out from the lattice (sputtering), whereas the projectile is either adsorbed in the solid or reflected at the surface. This phenomenon is characterized by the sputtering yield, that is, the average number of atoms ejected from the solid per incident ion, which depends on the nature, energy, and angle of incidence of the incoming ion and on the solid target characteristics as well (composition, structure). The energy and angular distributions of sputtered atoms also depend on these parameters (Ref 23). Other effects, such as defect creation, may also result from the interaction of energetic particles with surfaces.

Growth and Etching from the Vapor Phase

Driving Force for Growth and Evaporation in Chemical Vapor Processes. At thermodynamic equilibrium, the chemical potentials of the gas phase and solid phase must be equal, which is written:

$$\mu^G = \mu^S \qquad \text{(Eq 92)}$$

In practical situations, μ^G and μ^S have different values, and the gas-solid system evolves for minimizing its energy; that is, it tends to move toward thermodynamic equilibrium. The chemical driving force for crystal growth from the vapor is the chemical potential excess in the gas phase ($\Delta\mu = \mu^G - \mu^S$). For deposition to be observed, $\Delta\mu$ must be positive; however, the supersaturation ratio, S, and the relative supersaturation, σ, are generally used instead of $\Delta\mu$ in thermodynamic studies of gas-solid systems:

$$\sigma = S - 1 = e^{\Delta\mu/RT} - 1 \qquad \text{(Eq 93)}$$

The value of σ at thermodynamic equilibrium is zero. Solid deposition occurs for $\sigma > 0$, whereas solid evaporation is observed for $\sigma < 0$, for example, in chemical etching processes. Consider a very simple situation, for example, deposition of silicon from SiH_4:

$$Si(s) + 2H_2 \rightleftharpoons SiH_{4(g)} \qquad \text{(Eq 94)}$$

with equilibrium constant K_1. The chemical potential excess in the gas phase is:

$$\Delta\mu = \mu_{SiH_{4(g)}} - \mu_{Si_{(s)}} - 2\mu_{H_2(g)}$$
$$= \mu^\circ_{SiH_{4(g)}} - \mu^\circ_{Si_{(s)}} - 2\mu^\circ_{H_2(g)} + RT \ln\left(\frac{P_{SiH_4(g)}}{P^2_{H_2(g)}}\right) \qquad \text{(Eq 95)}$$

where μ° is the standard chemical potential. At thermodynamic equilibrium, $\Delta\mu = 0$ and $P_{SiH_4} = P^\circ_{SiH_4}$ and $P_{H_2} = P^\circ_{H_2}$; the supersaturation ratio of the gas phase with respect to the silicon solid phase is:

$$S_{Si} = \frac{P_{SiH_4}}{P^2_{H_2} K_1} \qquad \text{(Eq 96)}$$

Because supersaturation strongly influences film growth, several previous studies of CVD processes have tried to correlate the value of σ and the resulting film microstructure (see Ref 5 and the references therein). However, obtaining a precise value of supersaturation in reacting multi-component systems often turns out to be a difficult task, because many species and reactions must be considered in this case (Ref 5, 24, 25).

Growth Mechanism. The Kossel model has been widely used in earlier simulations of crystal growth because it is particularly helpful for studying surface processes and growth phenomena. The species to be incorporated on a Kossel surface adsorb first on a terrace and then diffuse toward ledges and finally to kink sites. Kinks are the most energetically favorable sites for reactions (formation of a maximum number of bonds with the first-nearest neighbor atoms); hence, those surfaces with the highest density of steps and kinks are the most reactive. Nevertheless, growth also occurs on atomically flat surfaces; the source of repeatable steps and kinks can be either nucleated islands or screw dislocations (Ref 26) in this case.

Distinction is made between the very first stages of growth (heterogeneous nucleation) involving gas-substrate interactions and the subsequent steps involving interactions between the gas and the growing solid film. The final deposit structure mainly results from the starting substrate surface structure and from the competing surface interactions at the surface: interactions between adsorbed species and interactions between adsorbates and substrate atoms. Depending on the relative importance of the aforementioned interactions and on the film-substrate lattice mismatch, either two-dimensional (2-D) layer-by-layer growth (Frank-Van der Merwe mechanism) (Ref 27, 28) or three-dimensional (3-D) growth (Volmer-Weber mechanism) (Ref 29) is obtained. When the substrate-adsorbate interactions are stronger and the substrate film-mismatch is low, the substrate may impose its crystallographic structure to the layer-by-layer growing film. This process, which is referred to as epitaxial growth, occurs when the vapor is supersaturated with respect to the 2-D solid phase but remains undersaturated with respect to the 3-D phase. The 3-D nucleation or Volmer-Weber mechanism is favored when the free energy of interface formation is high, which

requires a large supersaturation to overcome the nucleation barrier. In this case, some adsorbed species diffusing on the surface are temporarily trapped in potential wells, where they may capture other moving adsorbates and form solid embryos (nuclei) that become stable beyond a critical size. Diffusing adsorbates attach to the growing clusters, which finally coalesce to form a continuous film. A third growth mode (Stranski-Krastanov mechanism) (Ref 30) is also observed when internal strains due to film-substrate lattice mismatch become too large; in this case, layer-by-layer growth stops and is followed by growth of 3-D islands. The aforementioned growth modes, however, were initially proposed for PVD, and these are not strictly valid for CVD, which is also characterized by an incubation time for nucleation. Kajikawa and Noda (Ref 31) discussed the growth mode during the initial stages of CVD, which were found to depend on the differences between precursor sticking coefficients.

After the birth stage of the solid film, subsequent growth will determine its final microstructure. At low or moderate temperature, supersaturation is generally high, but most adsorbates cannot overcome the energy barriers of the various surface processes. Because adsorbates only move over very short distances in this case, they have a low probability of finding the most favorable surface sites for incorporation, and nonequilibrium amorphous solid phases are usually obtained. At higher temperature, monocrystalline films may result from layer-by-layer epitaxial growth on a single-crystal substrate. In contrast, polycrystalline film microstructures may be grown from individual crystalline nuclei; when tiny crystals with different crystallographic orientations are obtained at the nucleation stage, the final film microstructure may result from an evolutionary selection process (Ref 32). During the selection process, the differently oriented crystals compete to survive. The crystals that are less favorably oriented are gradually buried by those having their fastest direction of growth nearly perpendicular to the substrate. As a result, the grain size and orientation sharpness of the film increases with increasing film thickness.

Deposition and Etch Rates. Consider the case of CVD or etching of a polyatomic solid film, f, of density ρ_f. Let N_k be the number of chemical elements contained in the film and N_s the number of surface reactions with molar reaction rate R_s, and N the total number of gaseous species. R_s is negative for etching and positive for deposition reactions. The local net deposition rate (G_R) or etch rate (E_R) can be calculated from:

$$G_R \text{ or } E_R (kg \cdot m^{-2} s^{-1}) = \sum_{k=1}^{Nk} \sum_{i=1}^{N} \sum_{j=1}^{Ns} n_i^k \nu_{ij} R_{s_j} M_k$$
$$(kg \cdot m^{-2} s^{-1}) \qquad \text{(Eq 97)}$$

where n_i^k is the number of atoms of kth element deposited or etched from species i, and ν_{ij} is

the stoichiometric coefficient of species i in reaction j. Dividing Eq 97 by ρ_f gives G_R or E_R in $m \cdot s^{-1}$.

Maintaining polyatomic film stoichiometry is often desirable for modeling; in this case, the N_s surface reaction terms used in Eq 97 must explicitly include the density and coverage of the N_k types of surface sites, and additional balance equations must then be solved for these sites.

More complex expressions of the deposition or surface reaction rates must be considered when energy transfer between film surface and impinging atoms is large. Indeed, previous studies have shown that deposition rate or nucleation rate can be enhanced by ion bombardment (Ref 33). Energetic ion bombardment is also used for increasing the etch rate due to chemical reactions in reactive ion etching (RIE) processes.

Gas-Phase Reactions in CVD

Gas-phase reactions are of major importance in CVD or hybrid CVD- PVD processes. During their transport from the reactor inlet to the substrate, the precursor molecules undergo chemical transformations that may drastically change the composition of the gas phase. As a rule of thumb, the higher the pressure and temperature, the higher the importance of homogeneous reactions. To perform reliable simulations of CVD processes, a chemical scheme must be included in the coupled transfer model, which at least captures the key gas-phase reactions with the corresponding rates.

Gas-Phase Reaction Categories

In conventional thermal VPP, homogeneous chemical transformations mainly result from the internal redistribution of the vibrational energy gained by molecules in collisions, which may lead to bond breaking and/or molecular rearrangement. A chemical reaction of known stoichiometry is usually written as:

$$n_1 R_1 + n_2 R_2 + \ldots \rightleftharpoons m_1 P_1 + m_2 P_2 + \ldots \tag{Eq 98}$$
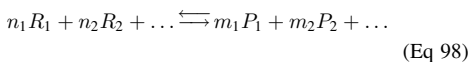
A single equation such as Eq 98 could be used to perform an overall balance over the chemical species present, but detailed mechanisms involving a succession of elementary reactions (that is, with a single mechanistic step) are used in today's (2009) CVD models. Distinction is usually made between two or three types of gas-phase reactions: unimolecular, bimolecular, and termolecular reactions, involving, respectively, one, two, or three reacting species plus eventually a third body, as further explained later. Depending on the degrees of freedom participating in the reaction and on the internal rearrangement in the reacting molecule, different types of unimolecular reactions are found, which also involve a third-body molecule. Dissociation reactions (e.g., $SiH_4 \rightarrow SiH_3 + H$) imply the breaking of a single bond in the reactant and proceed with a negligible energy barrier, contrary to elimination (e.g., $Si_2H_4 \rightarrow Si_2H_2 + H_2$) and isomerization reactions (e.g., $H_3SiSiH \rightarrow H_2SiSiH_2$) that involve internal rearrangement of the reacting molecule with and without bond breaking, as illustrated in Fig. 5.

Among bimolecular reactions, distinction is made between metathesis reactions, in which an atom or a group of atoms is transferred from one reactant to the other (e.g., $SiH_4 + H \rightarrow SiH_3 + H_2$); displacement reactions (e.g., $Si_2H_6 + H \rightarrow SiH_4 + SiH_3$); addition reactions (e.g., $SiH_4 + SiH \rightarrow Si_2H_5$); and recombination reactions (e.g., $SiH_3 + H \rightarrow SiH_4$), which are the reverse of unimolecular dissociation reactions. Termolecular reactions sometimes occur but are less frequent because they involve three reactant molecules plus a third body. They are not further discussed here.

Reaction schemes may become even more complex in VPP containing charged species and energetic particles. For instance, ion-ion and ion-neutral reactions are very important in ionized vapor phases such as plasma discharges (Ref 35). Reactions between ions of opposed charge usually lead to neutralization with or without bond formation/breaking (e.g., $Si_2H_5^- + SiH_3^+ \rightarrow Si_3H_6 + 2H$ or, alternatively, $Si_2H_5 + SiH_3$), whereas ion-neutral reactions produce a neutral and a charged species (e.g.,

$SiH^- + SiH_4 \rightarrow H_3SiSi^- + H_2$, or, alternatively, $Si_2H_5 + e^-$). Vibrationally or electronically excited products carrying an excess of energy are sometimes produced in highly exothermic reactions. Energetic ions or metastable species from discharge gas mixtures containing rare gases may dissociate molecules (e.g., $Ar^+ + SiH_4 \rightarrow SiH_3 + H + Ar$, and $Ar(^3P_2) + SiH_4 \rightarrow SiH_2 + 2H + Ar$). Electrons and photons also play key roles in ionized vapor-phase chemistry. Depending on their initial kinetic energy and reaction cross section, electron-molecule collisions may lead to excitation, ionization, neutral fragmentation, or dissociative ionization of molecules (e.g., $e^- + SiH_4 \rightarrow SiH_3 + H + e^-$ or $SiH_3^- + H$). Electrons may recombine with positive ions to give a neutral species, with the excess of energy released by emission of radiation, or attach to gas molecules with or without dissociation (e.g., $SiH_3 + e^- \rightarrow SiH_2^- + H$ or SiH_3^-). Depending on their energy, photons may also dissociate molecules (e.g., $SiH_4 + hv$ (147 nm) $\rightarrow SiH_3 + H$ and $SiH_2 + H_2$) (Ref 36). Either direct photolysis or mercury-photosensitized decomposition of reactants is used to prepare thin films in photo-CVD processes.

Gas-Phase Reaction Rates

One of the most difficult tasks in CVD modeling is to find accurate values of the rate constants for each elementary reaction considered, either from the information available in the literature (experimental or calculated data) or from estimates.

Estimation of Rate Coefficients. Consider the following gas-phase reaction (R_1), $WF_5 + F \rightleftharpoons WF_6$, that is involved in tungsten deposition from WF_6. At very high pressure, the reaction rate (R_{G1}) of the direct association reaction is written:

$$R_{G1} = \frac{\partial[WF_6]}{\partial t} = -\frac{\partial[WF_5]}{\partial t} = -\frac{\partial[F]}{\partial t}$$
$$= k_{R_{G1}}[WF_5]^1[F]^1 \tag{Eq 99}$$

According to Vant Hoff's law, the reaction order with respect to the reactants is equal to the stoichiometric coefficient of the reactants in the

Fig. 5 Potential energy along reaction coordinate for dissociation, elimination, and isomerization reactions. Source: Ref 34

reaction; hence, in the present case, the exponent is unity for both WF_5 and F concentrations. $k_{R_{G1}}$ is the reaction rate coefficient of reaction R_1. If R_{G1} is expressed in SI units (mol $\cdot$ m^{-3}s^{-1}), $k_{R_{G1}}$ is then in units of mol^{-1}m^{3}s^{-1}.

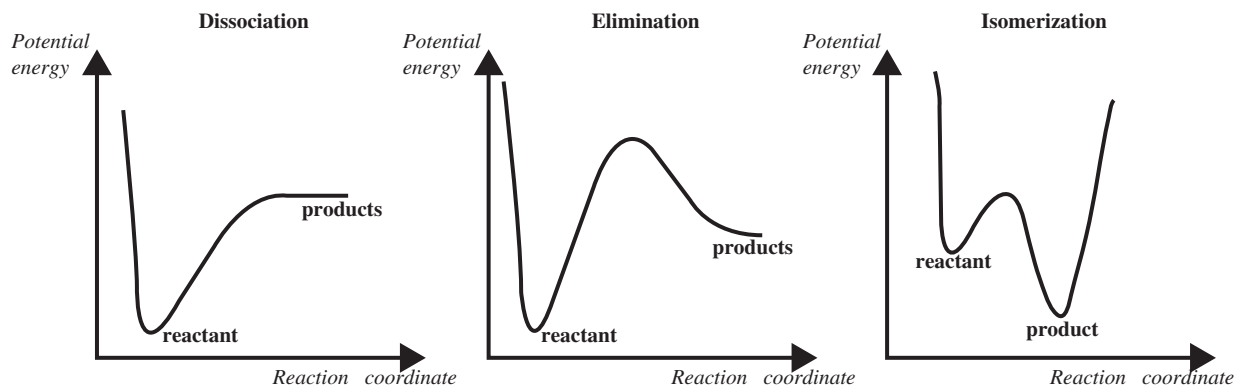For the reverse dissociation reaction (R_{-1}), the rate is:

$$R_{G-1} = -\frac{\partial[WF_6]}{\partial t} = \frac{\partial[WF_5]}{\partial t} = \frac{\partial[F]}{\partial t} = k_{R_{G-1}}[WF_6]$$
(Eq 100)

and $k_{R_{G-1}}$ has units of s^{-1} in this case.

R_{G1} and R_{G2} are equal at equilibrium, hence:

$$\frac{k_{R_{G1}}}{k_{R_{G-1}}} = \frac{[WF_6]}{[WF_5][F]} = Kc_1 = K_{P1}\left(\frac{P_{atm}}{RT}\right)^{\Delta n}$$
(Eq 101)

In this equation, the constants of the forward and reverse reactions are related to each other through the equilibrium constant, Kc_1, in concentration units. P_{atm} is the reference atmospheric pressure (= 1 in atm units), and Δn is the variation of the total number of moles during the reaction ($\Delta n = -1$ in the present example). The thermodynamic properties of gaseous species are related to the equilibrium constant in pressure units (Kp_1) through the following thermodynamic relation:

$$K_{P1} = \exp\left(\frac{\Delta S_1^\circ}{R} - \frac{\Delta H_1^\circ}{RT}\right)$$
(Eq 102)

where ΔS_1° and $\Delta H_1^\circ$are, respectively, the molar entropy change and the molar enthalpy change resulting from reaction R_1. It means that the rate constant of a given reversible reaction can be simply obtained from the reverse rate constant and thermodynamic properties of the compounds involved in the reaction.

The rate coefficients of reactions are usually assumed to have Arrhenius temperature dependence:

$$k_{R_1} = A_1 T^{\beta_1}\exp\left(-\frac{E_1}{R}\right)$$
(Eq 103)

with A_1 the pre-exponential factor, β_1 the temperature exponent, and E_1 the activation energy. Energy barriers of reactions are generally obtained from either experimental measurements or quantum mechanical calculations. As illustrated in Fig. 5, recombination reactions (reverse process of dissociation) usually proceed with no barrier and hence activation energies. Pre-exponential factors of reactions (and temperature exponents) are most often obtained from experimental measurements; however, simple considerations and formula can be used to obtain at least rough estimates in the absence of experimental data. Only simple methods for estimating upper bounds or orders of magnitude of rate coefficients are included as follows. The reader may refer to several books, in particular, a famous book by S. Benson (Ref 37), for further insight into estimation methods in chemical kinetics.

Neglecting possible dipole contributions, an upper bound of the bimolecular reaction rate coefficient between thermalized neutral species i and j is given by the Lennard-Jones collision rate:

$$k_{LJ} = \sqrt{\frac{8k_BT}{\pi m_{ij}}}\pi\sigma_{ij}^2\Omega_{Dij}$$
(Eq 104)

The order of magnitude of k_{LJ} is approximately 10^8 m^{3}mol^{-1}s^{-1}.

The Langevin theory gives the following equation for estimating the rate coefficient of ion-molecule reactions involving nonpolar molecules:

$$k_L = 2\pi(4\pi\varepsilon_0)^{-1/2}\left(\frac{\alpha q^2}{m_r}\right)^{1/2}$$
(Eq 105)

In this expression, α is the polarizability of the neutral molecule, q is the electrical charge of the ion, m_r is the reduced mass of the colliding pair, and ε_0 is the vacuum dielectric permittivity. The order of magnitude of k_L is 10^8 to 10^{10} m^{3}mol^{-1}s^{-1}, but the rate coefficients derived from the Langevin theory are usually higher than the experimental ones, often by 1 order of magnitude. Corrections to Eq 104 and 105 can easily be made for polar-neutral molecules (Ref 36).

As explained in the section "Gas-Phase Reaction Categories" in this article, electron-molecule interactions give rise to a large panel of processes (dissociation, ionization, etc.), depending on the electron-molecule collision cross sections and the kinetic energy of the electrons. Electron energy-dependent cross sections (σ_{ei}) are available for a variety of molecules or atoms, which can be used to compute the electron rate constants k_{ei} for a given collision process (i):

$$k_{ei} = \int_0^\infty \sigma_{ei}(E)\left(\frac{E}{2m_e}\right)^{1/2}f(E)dE, \text{ with } \int_0^\infty f(E)dE = 1$$
(Eq 106)

where m_e is the electron mass. The electron energy distribution function f(E) in gas discharges is known either from experimental measurements or from numerical modeling. Note that in most cold plasma discharges, f(E) is not exactly a Maxwellian distribution.

Pressure Dependence of Rate Coefficients. Unimolecular and bimolecular reactions, except metathesis and displacement reactions, depend not only on temperature but also strongly on pressure (Ref 34). The origin of this pressure dependence can be understood from the Lindemann-Christiansen theory (Ref 38, 39) for unimolecular decomposition reactions AB $\rightarrow$ A + B, which involve the following steps:

1. Collisional energy transfer (activation and deactivation steps) between AB and a third-body molecule (M): AB + M $\rightarrow$ AB* + M, rate constant k_1; AB* + M $\rightarrow$ AB +M, rate constant k_{-1}
2. Intramolecular rearrangement of AB* leading to dissociation (reaction step): AB* $\rightarrow$ A + B, rate constant k_2

Collisional energy transfers and intramolecular rearrangements proceed on different timescales (10^{-13} and 10^{-9} s, respectively) but may compete in a given pressure range, which results in a bath-gas pressure dependence of the rate coefficient k_{uni} for the overall reaction AB $\rightarrow$ A + B. The quasi-steady-state approximation applied to AB* leads to the following expression of the unimolecular rate constant:

$$k_{uni} = k_1[M]\left(\frac{k_2}{k_2 + k_{-1}[M]}\right)$$
(Eq 107)

At very high pressure (say at infinite pressure), gas-phase collisions between the reactant and third-body molecules are so frequent that the energy and angular momentum of particles retain their equilibrium distributions. As a consequence, the reaction step is rate determining, and k_{uni} does not depend on pressure (k_{uni} = Cte= k_∞ = k_1k_2/k_{-1}). Conversely, at very low pressure ($P \rightarrow 0$), the population of molecules having enough energy for reaction is strongly depleted, because collisions are no longer sufficiently frequent to maintain the equilibrium distribution. The result is that collisional processes become rate-limiting, and k_{uni} is proportional to pressure in the low-pressure limit (k_{uni} = k_1[M]). In the intermediate range of pressure, also referred to as the fall-off region, the pressure dependence of the unimolecular rate constant is no longer linear because the reaction step competes with collisional processes. This behavior is exemplified in Fig. 6(a), which shows the pressure dependence of the SiH$_4$ $\rightarrow$ SiH$_2$ + H$_2$ decomposition reaction

A similar analysis can be performed for isomerization and bimolecular association reactions, leading to a more-or-less similar pressure behavior (Ref 40). Bimolecular reactions (such as SiH$_3$ + H $\rightarrow$ Products) may have a chemical activation-energy distribution different from the thermal energy distribution because of the energy released by making the new bond. As seen in Fig. 6, the pressure dependence of the rate coefficient is similar for both the bimolecular recombination and unimolecular decomposition channels, in agreement with Eq 101. However, the bimolecular channel leading to dissociation of the complex into new products has a reverse pressure dependence as compared to association (Fig. 6b). Bimolecular or unimolecular reactions involving multiple isomerization may have a much more complex pressure dependence (Ref 41), but this case is not further discussed here.

Although explaining the pressure dependence for the rate constant of elementary gas-phase reactions, the Lindemann-Christiansen theory does not provide a very accurate representation of the
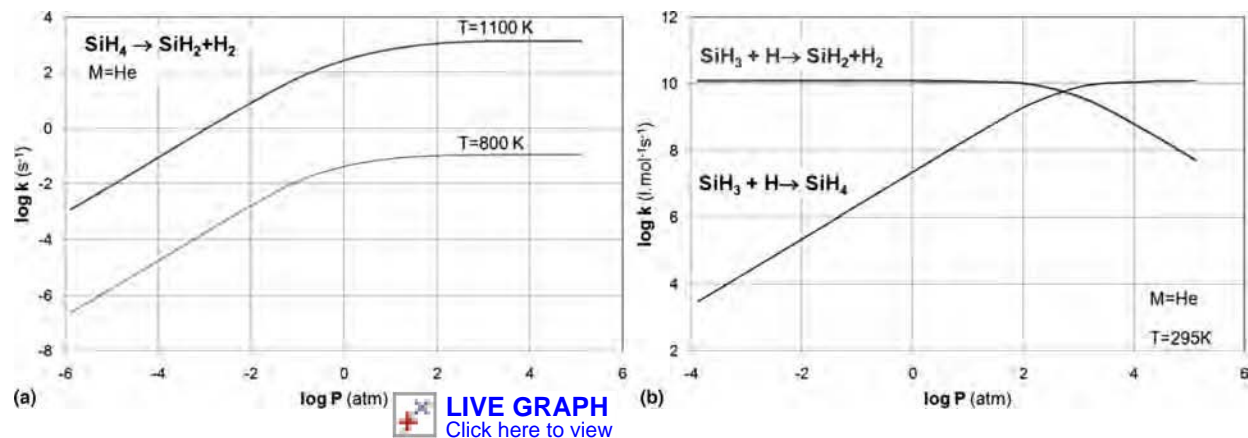
Fig. 6 Quantum Rice-Ramsperger-Kassel calculations showing the pressure dependence of unimolecular and bimolecular reactions involving silane. Source: Ref 40

rate constant. The Lindemann theory was first improved by Hinshelwood (Ref 42), who derived an expression of the energy dependence of k_1, but more sophisticated theories were derived by Rice and Ramsperger (Ref 43) and independently by Kassel (Ref 44, 45), which are today (2009) referred to as RRK theories. Even more sophisticated treatments were proposed later by Marcus and Rice for unimolecular reactions in the so-called RRKM theory (Ref 46, 47). The quantum version of RRK theory (QRRK), proposed by Kassel for unimolecular reactions, was later extended to bimolecular reactions by Dean (Ref 48). Currently, both QRRK and RRKM theories are widely used to derive pressure-dependent rate coefficients for the simulation of combustion or vapor deposition processes (Ref 40, 41, 48–50). The QRRK calculations are less sophisticated than RRKM, but they require less input data and can be conducted more rapidly (Ref 41).

Analytical expressions of the rate constant $k\,(T, P)$ must finally be derived for each homogeneous reaction and used as input data for the CVD process model. Several parameterization methods have been proposed for reproducing the shape of the curve over the whole pressure range, based on the Lindemann-Hinshelwood formalism. For unimolecular reactions:

$$k_{uni} = k_\infty F_{LH} = k_\infty \frac{P_r}{1 + P_r} \qquad \text{(Eq 108)}$$

The F_{LH} is the so-called Lindemann-Hinshelwood factor, and P_r is the reduced pressure, defined as:

$$P_r = \frac{k_0[M]}{k_\infty} \qquad \text{(Eq 109)}$$

The center of the fall-off domain is the pressure that corresponds to $k_\infty = k_0[M]$, that is, to $P = P_c$ when $P_r = 1$. The broadening of the real fall-off curve as compared to the Lindemann-Hinshelwood fitting can be simply taken into account by introducing a correction factor, F:

$$k_{uni} = k_\infty F_{LH} F \qquad \text{(Eq 110)}$$

Troe (Ref 51) proposed the following optimized expression of the broadening factor:

$$\log F = \frac{\log F_{cent}}{1 + \left[\frac{\log P_r + c}{N - d(\log P_r + c)}\right]^2} \qquad \text{(Eq 111)}$$

with

$$c = -0.4 - 0.67 \log F_{cent},$$
$$N = 0.75 - 1.27 \log F_{cent} \text{ and } d = 0.14 \qquad \text{(Eq 112)}$$

and

$$F_{cent} = (1 - a) \exp\left(-\frac{T}{T^{***}}\right) + a \exp\left(-\frac{T}{T^*}\right)$$
$$+ \exp\left(-\frac{T^{**}}{T}\right)$$
$$\text{(Eq 113)}$$

A total of ten parameters must be identified for each reaction rate constant when using the Troe formalism: a, T^*, T^{**}, T^{***}, A_∞, β_∞, Ea_∞, A_0, β_0, and Ea_0. The first four parameters are characteristic of the reactant and bath gas molecules. They are identified by fitting $k(T,P)$ at various temperatures for $P = P_c$ ($P_r = 1$) (Ref 40). The last six parameters are simply those used in the Arrhenius expressions of k_∞ and k_0. It is important to note that negative values can be found for the activation energies of gas-phase reactions in the low-pressure limit.

A different expression has been proposed by Larson (Ref 52), referred to as SRI or the Stewart expression, and improved by Kee et al. (Ref 53). A third approach has been proposed by Oref (Ref 54) for fitting pressure-dependent unimolecular rate constants, which is not based on the Lindemann-Hinshelwood formalism, but the Troe and SRI approaches have usually been preferred. Both methods remain valid for chemically activated bimolecular reactions, but in this case, the following

expressions must be used for the rate constant (Ref 40):

$$k^{dec} = k_0 \frac{1}{1 + P_r} F \qquad \text{(Eq 114)}$$

They can also be extended to reactions involving multiple isomerizations (Ref 41).

Construction, Analysis, and Reduction of Gas-Phase Reaction Mechanisms

Construction. Building a chemical reaction mechanism for VPP models requires three steps: identify the species involved, find out accurate thermodynamic data for these species, and gather the rate constants of all elementary reactions with their temperature and pressure dependence (from measurements or calculations). An exhaustive analysis of the literature is usually required for the construction step. Thermodynamic data are now available for several chemical systems of interest for vapor deposition, for example, for Si-H (Ref 55, 56), Si-N-H (Ref 57), Si-C-H (Ref 58), Si-O-H (Ref 59), and so on, which have most often been derived from ab initio calculations. Some selected temperature-dependent thermodynamic data can currently be found in databases included in simulation software, for example, in CHEMKIN (Ref 13), or simply provided as tables (Ref 60). When not available, thermodynamic data can usually be estimated by simple methods, as explained by Benson (Ref 37). Chemical kinetic databases are also available, such as NIST (Ref 61), which provide compilations of previously published data for a large number of reactions.

Most full chemical mechanisms, often involving hundreds of reactions, have been developed for combustion problems (e.g., Ref 49, 62–64). Early models of CVD processes used rudimentary chemical schemes, often including only a single overall heterogeneous reaction whose rate constant was simply estimated

by fitting experimental growth rate profiles to calculated profiles. Simulations performed by using such models with no or very few gas-phase reactions provided mitigated or even poorly reliable results. These rudimentary chemical schemes have now been encompassed by more accurate models, especially for gas-phase chemistry. One of the first detailed chemical mechanisms used in CVD was developed by Coltrin et al. (Ref 22) for silicon deposition from silane. Twenty-six gas-phase reactions involving 17 species were considered in this mechanism, however, with few considerations for the pressure dependence of the rate constants. Ho et al. (Ref 65) have proposed a chemical mechanism derived from previous experimental measurements and RRKM calculations on SiH_4 chemistry that includes a complete paramerization of the pressure-dependent rate coefficients. Subsequent reaction schemes have considered additional reactions and species (Ref 56), but very few have provided in-depth investigations of the pressure dependence of reactions (Ref 41). Other chemical systems of interest for CVD have been investigated over the past 15 years, and several gas-phase reaction schemes are currently available for modeling a variety of deposition processes. With a few exceptions (e.g., Ref 66), most of these schemes are simplified, including those of interest for deposition of metallic materials

Analysis. Analyzing very large reaction mechanisms without numerical tools is almost impossible. The analysis step consists of examining both the stoichiometric and the parametric information provided by each reaction; the first one depends on the chemical pathways, and the second one depends on the thermodynamic and kinetic parameters. Several types of analysis can be conducted:

- The sensitivity analysis identifies the rate-limiting steps (Ref 67, 68).

- The reaction flow analysis highlights the chemical pathways (Ref 69).
- The rate-of-reaction analysis helps in understanding the mechanisms (Ref 67).
- The eigenvalue-eigenvector analysis identifies the characteristic timescales and directions of the chemical reactions (Ref 69).

These methods, originally used for combustion problems, have also been used to study vapor deposition problems, for example, for TiC (Ref 70) or SiC (Ref 71) deposition. The computer package KINALC (Ref 72) has often been used for performing kinetic analysis of reaction mechanisms used in CVD problems.

Note that the time variations of the gas-phase composition must be simulated prior to analysis (and reduction) of a reaction mechanism. This can be done by using a one-dimensional plug-flow reactor model, because the time coordinate in a closed system is equivalent to the axial distance along a tube in an open system (assuming a constant total number of moles). Such simulation models are provided in chemical kinetics packages such as CHEMKIN (Ref 13).

Reduction. Two kinds of methods can be used for reducing chemical mechanisms: methods that involve time-scale analysis and methods that do not. Computational singular perturbation (Ref 73) and intrinsic low-dimensional manifold (Ref 74) methods rely on time-scale and Jacobian analysis; slow and fast processes are decoupled to reduce the stiffness of the system. With a few exceptions (Ref 75), these methods have almost never been used in vapor deposition problems. Among the methods that are not based on time-scale analysis, Turanyi's method (Ref 76) considers that all species and reactions are not equally important and reduces their number in the chemical scheme. According to Turanyi (Ref 76), important species are those for which the accurate reproduction of the concentration profiles

follows directly from the aim of the investigation. The decision of which species and/or feature are considered important depends on the objective of the modeling. Necessary species are those for which realistic concentrations are required to calculate accurate concentration profiles for the important species or to reproduce important features. Redundant species are those that may be omitted from the mechanism without jeopardizing the aim of the modeling. An important reaction is a reaction in which both stoichiometric information and parametric information are required. In a necessary reaction, only stoichiometric information is required. A redundant reaction can be eliminated from the mechanism.

The first step of Turanyi's method consists of testing each species to know whether it is significant or not under the selected conditions; then, the redundant species are deleted. Redundant species can be identified either by generating pseudoreduced models (one for each potentially redundant species), in which all reactions involving the species of interest have been eliminated, or by the investigation of the Jacobian. The first procedure is time-consuming but more powerful. When all the redundant species have been eliminated from the mechanism, information provided by the remaining reactions (involving important and necessary species) is tested, and the nonsignificant reactions are eventually eliminated. Classical analysis methods, such as the principal component analysis of the rate-sensitivity matrix, rate-of-production analysis, and so on, are used to test reaction information. The reduced chemical scheme obtained is equivalent to the full mechanism for the specific conditions considered, but only for those conditions. Figure 7 presents a comparison between time variations of mole fractions before and after reduction of a large chemical scheme of the CH system. The reduction was conducted for the following CVD conditions: 1573 K, 5330 Pa, and a C_3H_8-H_2 gas

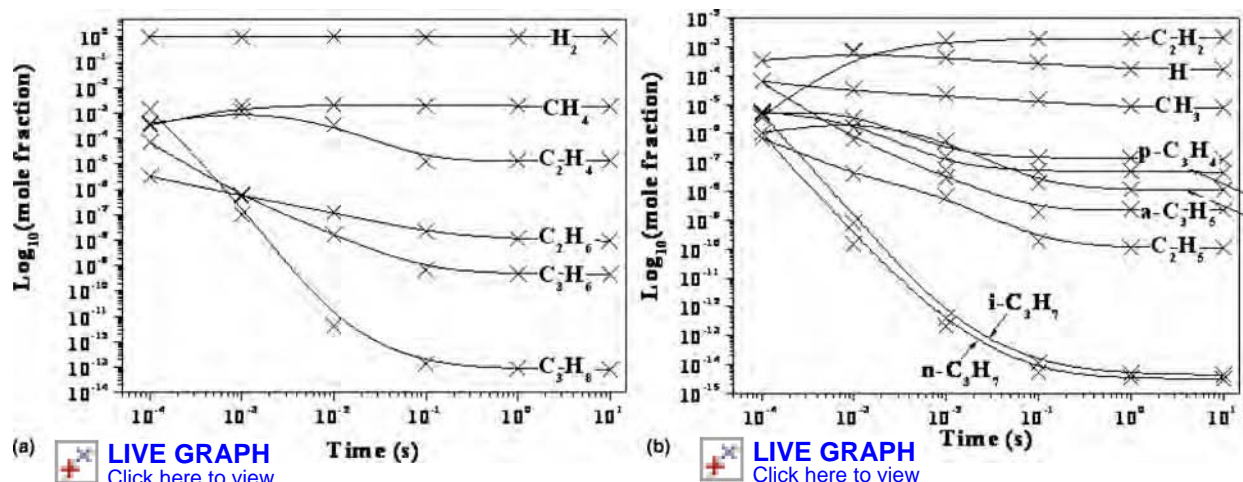**Fig. 7** Calculated time variations of the mole fractions of hydrocarbon species in a C_3H_8-H_2 mixture of (a) important species and (b) necessary species. Comparison between the full (plain lines) and reduced (crosses) mechanisms. See text for conditions. Source: Ref 71

mixture (Ref 71).The initial scheme, comprising 305 reversible reactions involving 111 species, has been reduced to 24 reversible reactions involving 16 species.

Thermodynamic Modeling of CVD

Near-Equilibrium Approximation. The derivation of accurate reaction schemes, including gas-phase and surface reactions, remains a tedious and difficult task when all the mechanistic and kinetic information is not available. If at least thermodynamic data are accessible, thermodynamic equilibrium calculations can be conducted for finding the final state of the gas-solid system. This approach, which has been widely used in CVD, may provide qualitative information on the nature and composition of the various phases and on the influence of operating parameters when the system is not too far from equilibrium (high temperature, long residence times), that is, when supersaturation is low. Equilibrium calculations are usually performed for nonflowing, closed CVD systems (e.g., Ref 77), thus neglecting transport phenomena, although open systems have sometimes been considered as well (Ref 78). Thermodynamic calculations have also been conducted for estimating the supersaturation of a given CVD system (Ref 24). The gas-phase equilibrium is first simulated, then the gas-solid equilibrium. The concentrations of depositing species at homogeneous equilibrium are then considered as the concentrations of the supersaturated vapor, whereas those calculated in heterogeneous equilibrium conditions are the true equilibrium concentrations. Although they provide valuable qualitative information, thermodynamic calculations lead to species concentrations that sometimes differ markedly from real concentrations, especially for radical species (Ref 77).

The preferred calculation method, which has been used in most commercial programs, consists of minimizing the Gibbs free energy (G) of the system. For a closed system at pressure P and temperature T, containing φ phases with N_k components per phase k, G can be written as:

$$G_{(T,P)} = \sum_{k=1}^{\varphi} \sum_{i=1}^{N_k} n_i^k \mu_{i(T,P)}^k \qquad \text{(Eq 115)}$$

The chemical potential of species i in phase k is:

$$\mu_{i(T,P)}^k = \mu_{i(T,P)}^{o^k} + RT \ln(a_i^k) \qquad \text{(Eq 116)}$$

where a_i^k is the activity of component i in phase k ($a_i = 1$, x_i or x_iP, respectively, for solids, ideal solutions, or ideal gases). The equilibrium condition is obtained by minimizing G:

$$dG_{(T,P)} = 0 \qquad \text{(Eq 117)}$$

with the constraint of conservation of the N_E elements present in the mixture:

$$b_j = \sum_{k=1}^{\varphi} \sum_{i=1}^{N_k} n_i^k c_{ij}^k \quad (j = 1, \dots . N_E) \qquad \text{(Eq 118)}$$

where b_j is the total number of gram atoms of element j in the mixture, and c_{ij}^k is the number of gram atoms of element j in species i and phase k. This set of $N_E + 1$ equations is usually solved by the Lagrange multiplier method. Today (2009), commercial packages such as Thermo-calc (Ref 79), Gemini (Ref 80), and FactSage (Ref 81) are routinely used for performing equilibrium calculations of multiphase systems (Ref 82) such as CVD systems. The CalPhad (Ref 83) approach (computer coupling of phase diagrams and thermochemistry) combines thermodynamic models, software, and databases as well as interface models.

Modeling and Computation of Transport Equations in Continuous Media

This section describes the various stages of developing models for numerical simulation of the transport phenomena in VPP. The modeling process involves the basic steps of defining boundary conditions and applying discrete numerical methods (e.g., finite element) for solution.

Currently, powerful commercial simulation packages capable of addressing complex multiphysics problems tend to supplant user's personal codes developed mainly in the 1980s for the study of transport phenomena. Among the most popular codes used for simulating VPP are those based on finite-volume methods such as FLUENT (Ref 84), PHOENICS-CVD (Ref 85), CFD- ACE (Ref 86), CFX (Ref 87), and those based on the finite-element method, such as MP-SALSA (Ref 88) and COMSOL-Multiphysics (Ref 89). Sometimes, CVD reactor simulation modules have also been introduced in thermochemistry simulation programs such as in CHEMKIN (Ref 13).

Boundary Conditions

These conditions are usually specific to the reactor configuration, but typical conditions are listed as follows.

Nonreacting Solid Walls. For velocity, a no-slip and no-penetration condition is usually considered at walls (w):

$$\mathbf{v} = \mathbf{0} \qquad \text{(Eq 119)}$$

For temperature, a prescribed temperature profile is considered in simplest cases:

$$T = T_w \qquad \text{(Eq 120)}$$

Alternative conditions are adiabatic wall:

$$\mathbf{n} \cdot \nabla T = 0 \qquad \text{(Eq 121)}$$

or constant heat flux:

$$\mathbf{n} \cdot \nabla T = h(T - T_{amb})/k \qquad \text{(Eq 122)}$$

A detailed heat-transfer balance may also be written for reactor walls in the so-called conjugate heat-transfer problem according to the three following equations:

$$\nabla \cdot [k_w \nabla T] = 0 \quad \text{Steady-state conduction in wall} \qquad \text{(Eq 123)}$$

$$k\mathbf{n} \cdot \nabla T = k_w \mathbf{n} \cdot \nabla T + \sigma \sum_{j=1}^{N_w}$$
$$(\mathbf{F}_{j-w} \varepsilon_j T_j^4 - \mathbf{F}_{w-j} \varepsilon_w T_w^4) \quad \text{Interior wall} \qquad \text{(Eq 124)}$$

$$-k_w \mathbf{n} \cdot \nabla T = h(T_w - T_{amb})$$
$$+ \sigma \varepsilon_{amb}(T_w^4 - T_{amb}^4) \quad \text{External wall} \qquad \text{(Eq 125)}$$

In these equations, $\mathbf{n}$ is the normal unit vector to the surface, h is the heat-transfer coefficient at the wall, N_w is the number of internal solid walls, ε_j (respectively, ε_{amb}) is the temperature-dependent emissivity of wall j (respectively of ambient), and $\mathbf{F}_{j-w}$ (respectively, $\mathbf{F}_{w-j}$) is the configuration factor for the jth wall element to the reactor wall w considered (respectively for wall w to the jth wall element). σ is the Stefan-Boltzmann constant. Formulas are usually available for configuration factors (Ref 90). Note that reflected radiation has not been included in Eq 124, which is generally a reasonable assumption for quartz walls.

For species:

$$(\rho \omega_i \mathbf{v} + \mathbf{j}_i) \cdot \mathbf{n} = 0 \ (i = 1, \dots N \text{ species}) \qquad \text{(Eq 126)}$$

Reacting Solid Walls. For velocity:

$$\rho \mathbf{v} \cdot \mathbf{n} = \sum_{i=1}^{N} M_i \sum_{j=1}^{N_s} v_{ij}^s R_{s_j} \qquad \text{(Eq 127)}$$

(N species, N_s surface reactions)

For temperature, prescribed temperature profiles can be considered such as for nonreacting walls; otherwise, an additional source term:

$$\left(\sum_{i=1}^{N} \sum_{j=1}^{N_s} \overline{H}_j v_{ik}^s R_{sj} \right)$$

may be added on the right side of Eq 123 to 125 to account for the heat released by surface reactions.

For species:

$$(\rho \omega_i \mathbf{v} + \mathbf{j}_i) \cdot \mathbf{n} = M_i \sum_{j=1}^{N_s} v_{ij}^s R_{s_j}$$
$$(i = 1, \dots N \text{ species}) \qquad \text{(Eq 128)}$$

In this equation, it has been written that the net mass flux of species i equals its total net consumption rate. Equation 128 may easily be generalized to establish that the sum of mass fluxes of depositing species is equal to the deposition rate, which may become limited either by gas-phase transport or by surface reactions. This is illustrated in Fig. 8 (Ref 91),
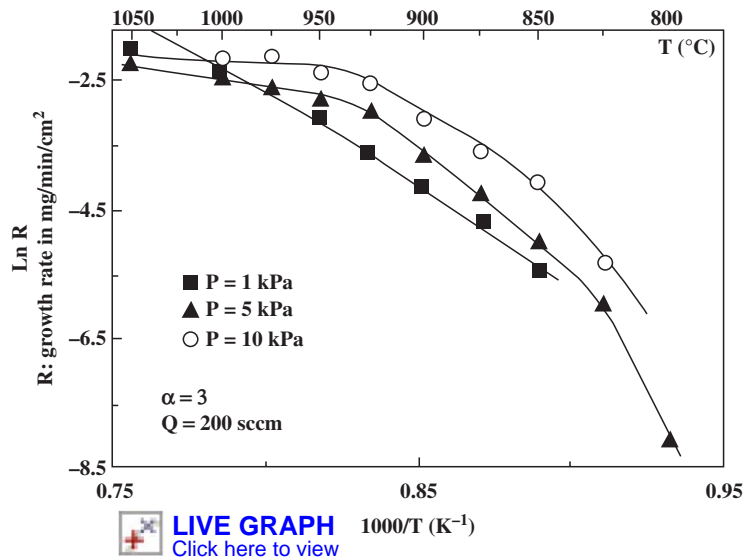
Fig. 8 Plot of log (deposition rate) versus $1/T$ in the case of deposition of SiC from CH_3SiCl_3/H_2 at various pressures. Source: Ref 91

showing a typical plot of the logarithm of the deposition rate versus $1/T$. In the lowest-temperature region (high values of $1/T$), deposition is limited by surface reactions and the corresponding activation energy is high; conversely, at sufficiently high temperature, surface reactions are fast, and deposition becomes limited by transport in the gas phase.

Reactor Inlet. For velocity, a prescribed velocity profile is most often selected:

$$\mathbf{v} = \mathbf{v_0} \qquad \text{(Eq 129)}$$

For temperature, a prescribed profile may be used as well:

$$T = T_0 \qquad \text{(Eq 130)}$$

Otherwise, for convection-dominated flows:

$$\mathbf{n}\cdot\nabla T = 0 \qquad \text{(Eq 131)}$$

and for flows with nonnegligible back-conduction effects:

$$\rho C_p(T_0 - T)\mathbf{v_0}\cdot\mathbf{n}/k = -\mathbf{n}\cdot\Delta T \qquad \text{(Eq 132)}$$

For species, mass fractions may be set to feed values in convection-dominated flows:

$$\omega_i = \omega_{i0} \quad (i = 1, \ldots N) \qquad \text{(Eq 133)}$$

or, alternatively, when back diffusion cannot be neglected:

$$-\mathcal{D}_{im}\mathbf{n}\cdot\nabla\omega_i = (\omega_{i0} - \omega_i)(\mathbf{v_0}\cdot\mathbf{n}) \qquad \text{(Eq 134)}$$

Reactor Outlet. For velocity, as well as for temperature and species, a zero axial gradient in the direction normal to the outflow opening is usually adequate:

$$\mathbf{n}\cdot(\nabla\rho\mathbf{v}) = 0 \qquad \text{(Eq 135a)}$$

$$\mathbf{n}\cdot\nabla T = 0 \qquad \text{(Eq 135b)}$$

$$\mathbf{n}\cdot\nabla\omega_i = 0 (i = 1, \ldots N) \qquad \text{(Eq 135c)}$$

Methods for the Discretization of Transport Equations

The complete transport model consists of a set of nonlinear partial differential equations (PDEs), defined in the previous sections "Macroscopic Conservation Equations in Pure Gases" and "Conservation Equations in Multicomponent Gaseous Mixtures," that are closed by boundary conditions (see the section "Boundary Conditions" in this article). These PDEs have the general form of a convection-diffusion equation for the transported variable ϕ:

$$\frac{\partial}{\partial t}(\rho\phi) = \nabla\cdot(\rho\mathbf{v}\phi) + \nabla\cdot(\Gamma_\phi\nabla\phi) + S_\phi \qquad \text{(Eq 136)}$$

where Γ_ϕ and S_ϕ are the generalized diffusion coefficient and generalized source term, respectively. Analytical solutions can be found only for very simple reactor geometries when numerous and strong simplifying assumptions are formulated. Although these approaches provided interesting qualitative results, they have been encompassed for at least 25 years by more and more accurate numerical treatments. Three main classes of numerical methods are used in computational fluid dynamics (CFD) for the discretization of transport equations:

- The finite-difference (FD) method can be used in quite simple situations, for example, 2-D axisymmetric and/or isothermal reactors (Ref 22, 92, 93). The derivatives in the

differential equations are simply approximated by a truncated Taylor-series development. The values of the variable (ϕ) are calculated at grid points, implicitly assuming a polynomial variation of ϕ with the coordinates.

- The finite-volume (FV) method is based on an FD approach, but it uses a conservative formulation of the transport equations. The FV or control-volume method, which can also be regarded as a simple variant of the weighted-residual method (Ref 94), has been widely used for solving coupled transfer problems, including vapor deposition (Ref 95–97), and is subsequently presented in more detail.

- The finite-element (FE) method is based on the transformation of the original transport equations into an equivalent integral formulation using a variational principle, then performing a functional minimization. Piecewise profiles are used for interpolating the variable between grid nodes defined by the intersections of small volume elements. The Galerkin weighted-residual method is generally used for minimizing the equation residual (that is, the difference between the exact solution and its approximation). The FE method has been especially used for modeling complex CVD problems (Ref 98–101).

This article gives a very brief description of the control-volume method. The goal is not to present numerical methods for solving PDEs or other numerical methods in detail; the reader will find many excellent books addressing that topic (Ref 9, 94, 102).

The Finite-Volume (FV) Method

The FV method, or control-volume FD method, which has been described in detail in a famous book by Patankar (Ref 94) or in Ref 9, has been widely used for solving coupled transfer problems.

Derivation of Linear Algebraic Equations. As explained previously, the FV method is based on an FD formulation, but the PDEs are integrated over nonoverlapping control volumes using piecewise profiles expressing the variation of ϕ between the grid points. All the scalar variables (T, P, ω_1, ...ω_n, fluid properties) are calculated at grid points, "P," each of them being surrounded by small cells (control volumes). Vector quantities ($\mathbf{v}$ and $\mathbf{j}$) are calculated using a staggered grid, in points located halfway between the scalar grid points, that is, at the cell walls in the scalar grid (Fig. 9).

In the original form of the FV method, only regular grids were considered, and it was not possible to use triangular or more complex forms for the cells, as in the FE method; however, control-volume-based approaches have been developed that allow irregular grid geometries such as triangular grids (Ref 84, 94).
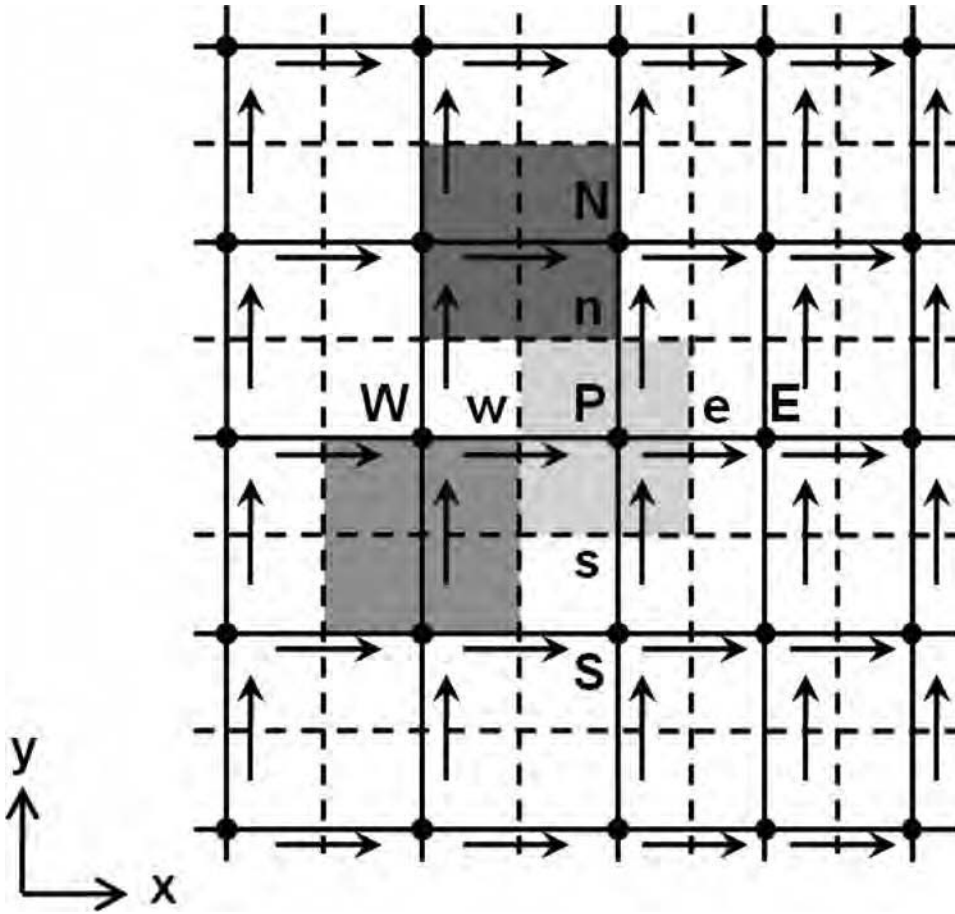
Fig. 9 Location of a control volume (light gray) for scalar variables around point "P" in a two-dimensional rectangular grid. A particular grid point "P" has four neighbor grid points ("W," "E," "N," "S"), and the control volume intercepts the scalar grid at points ("w," "e," "n," "s"). Control volumes for the x (dark gray) and y (medium gray) components of fluid velocity are staggered.

After discretization of Eq 136 and integration over the control volume surrounding "P," as shown in Fig. 9 (which is a particular case), a linearized equation is obtained:

$$a_P \phi_P = a_E \phi_E + a_W \phi_W + a_N \phi_N + a_S \phi_S + b \tag{Eq 137}$$

where the discretization coefficients a_P, a_E, a_W, a_N, and a_S depend on the interpolation formula stemming from the profile assumed for ϕ between grid points. Different interpolation schemes can be used: upwind, power law, exponential, and so on, as described in Ref 94. The source term has been linearized in ϕ:

$$S = S_c + S_p \phi_P \tag{Eq 138}$$

A fully implicit discretization leads to the following relationships for the discretization coefficients:

$$a_P = a_E + a_W + a_N + a_S + a_P^\circ - S_p \tag{Eq 139}$$

$$b = S_c + a_P^\circ \phi_P^\circ \quad \text{with } a_P^\circ = \frac{\rho_P^\circ}{\Delta t} \tag{Eq 140}$$

where Δt is the discretized time step. Exponent ($^\circ$) refers to known values at time t, while unknown values calculated at $t + \Delta t$ have no exponent.

Numerical Considerations. The discretized transport equations obtained are usually solved iteratively in a segregated way, that is, by sequentially solving the governing equation for each variable at each iteration step; however, a coupled nonsegregated approach with an appropriate linearization of the source term is sometimes preferred when stiff equations must be handled, for example, for species-conservation equations with source terms including fast reactions, to avoid a huge number of iterations to reach steady state. A nonsegregated approach can then be used in regions where chemical reactions prevail, in combination with a segregated approach in regions where convection and diffusion dominate (Ref 103). The pressure field is not known a priori, and a pressure-correction equation must be added to ensure that the calculated velocity field always satisfies the continuity equation. A Semi Implicit Method for Pressure Linked Equations

(SIMPLE algorithm) has accordingly been proposed by Patankar for solving the momentum and continuity equations. Iterations start from guessed values of the variables at each grid point. At any given iteration step, each point of the grid must be visited to calculate the new values of the considered variable. The well-known Gauss-Seidel point-by-point method (Ref 9, 102) can be used for this purpose, but its convergence is slow because the boundary-condition information is transmitted at a rate of only one grid point per iteration. Faster line-by-line algorithms, such as alternating direction implicit or the Thomas algorithm of tridiagonal-matrix algorithm (Ref 9) are most often preferred (Ref 94). When necessary, underrelaxation coefficients (α) can be used to stabilize the iterative procedure:

$$\phi = \alpha \phi_{\text{new}} + (1 - \alpha) \phi_{\text{old}} \tag{Eq 141}$$

The solution of a steady-state problem can also be obtained by iterating over time, starting from the known initial ϕ-values. In this case, the time step will be limited by stability criteria and not by accuracy; this false time step may even be different from one grid point to another or for each variable. The accuracy of the solution is closely related to the number of grid points, which depends on both the geometry and the nature of the problem. As a rule of thumb, the steeper the gradients in a given zone, the higher the number of grid points and the lower the value of grid spacing in this zone. However, the number of grid points should always be kept as low as possible to avoid the use of huge computational resources and prevent prohibitive calculation times. A variety of numerical methods are currently available to increase the computation speed, such as multigrid algorithms or parallelization strategies (Ref 9, 84, 102).

Finally, criteria must be used to check the convergence of the iterative procedure; this can be done by calculating the value of the scaled residual for each variable, ϕ (occasionally or at each iteration), and comparing it to an accepted maximum value, ε_ϕ, for example:

$$R\phi = \frac{\sum_P |a_E \phi_E + a_W \phi_W + a_N \phi_N + a_S \phi_S + b - a_P \phi_P|}{\sum_P |a_P \phi_P|} \leq \varepsilon_\phi \tag{Eq 142}$$

Other criteria may also be considered, for example, chemical balances over the chemical elements, relative change of variables at key points, and so on.

Modeling and Computation of Transport Equations in Transition Regime Flows

As for transport equations in continuous media, analytical solutions of the Boltzmann equation may sometimes be obtained for

rarefied transport problems in very simple situations, when strong assumptions are considered, but numerical methods are better suited for the simulation of real cases.

Direct Boltzmann CFD Method. One approach that seems straightforward is to discretize the Boltzmann equation by using the FV or FE methods introduced in the section "Modeling and Computation of Transport Equations in Continuous Media" in this article. Although this can be done in very simple situations (monoatomic gas, simple geometry, etc.), this direct approach becomes intractable when the geometry becomes more complex and the number of particles becomes important. Indeed, the evaluation of the collision term involves a huge number of operations in this case, and when the number of grid intervals exceeds a rather modest value in each dimension, the size of the required matrix becomes immediately prohibitive (Ref 7). Consider an unsteady 3-D flow with only 100 grid intervals in each of the 7 dimensions of the phase space; the grid has 10^{14} points! For these reasons, physically based simulation methods, as explained subsequently, are usually preferred to the direct Boltzmann CFD method.

Molecular Dynamics Simulations. Molecular dynamics (MD) is a deterministic, direct simulation method. The real gas is represented by a sufficiently large number of simulated molecules (a few hundreds to 10^8 to 10^9 or even more). Starting from a selected initial configuration of the system at $t = 0$ (for positions and velocities), molecules are allowed to move and interact with their counterparts and with the system boundaries. The positions, velocities, and internal state of the individual simulated molecules are stored and modified after each time step (typically only 10^{-14} to 10^{-15} s, corresponding to short molecular displacements that do not significantly affect the molecular interaction potential). Classical MD simulations do not consider quantum effects; in the simplest case, molecular movements are governed by Newton's equations of motion:

$$m_i \frac{\partial \mathbf{c}_i}{\partial t} = m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i = \frac{\partial E(\mathbf{r}_1, \dots \mathbf{r}_N)}{\partial \mathbf{r}_i} \qquad \text{(Eq 143)}$$
$$(i = 1, \dots N)$$

where $\mathbf{F}_i$ is the force exerted on particle i by the $N - 1$ other particles at a distance $\mathbf{r}_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. The corresponding potential E can usually be calculated by summing classical atomic pair potentials, ψ_{ij}:

$$E = \frac{1}{2} \sum_{i=1 \neq j}^{N} \sum_{j=1 \neq i}^{N} \psi_{ij}(\mathbf{r}_{ij}) \qquad \text{(Eq 144)}$$

The simplest pair potential without any cohesive interaction is given by the hard sphere model:

$$\psi_{ij}(\mathbf{r}_{ij}) = \begin{cases} = \infty & \text{for } \mathbf{r}_{ij} \leq \mathbf{r}_0 \\ = 0 & \text{for } \mathbf{r}_{ij} > \mathbf{r}_0 \end{cases} \qquad \text{(Eq 145)}$$

Another simple but more realistic potential that can be used for van der Walls interactions is the L-J potential defined in Eq 15.

The N equations of motion defined in Eq 143 are usually solved by using an FD algorithm, such as the popular Verlet or Gear predictor-corrector algorithms (Ref 104). Due to the large computational power required for calculating the force field and the very small time steps used in the iterative procedure, most MD simulations cover time periods less than 1 ns. Moreover, for flow in gases, a minimum number of molecules in a cubic mean free path must be considered, which is proportional to $\lambda^{-1/3}$ (Ref 7). As a consequence, MD simulations are usually restricted to high-density gases; hence, they are not well suited for solving transport problems in dilute or rarefied flows.

Monte Carlo Simulations. The basic idea of the test particle Monte Carlo (MC) method is to generate a large number of molecular trajectories from an initially estimated configuration of the system. A corresponding distribution function is selected and stored at each point in the phase space. From the assumed distribution, a large number of test trajectories and molecular collisions are computed on a statistical basis, and an updated distribution is obtained that is used to perform another series of test trajectory calculations, and so on. The process is repeated until the distribution no longer evolves between calculations. However, there is no real time variable in this iterative process.

Conversely, as in MD simulations, the real-time displacements of a large number of simulated molecules are tracked in the direct-simulation Monte Carlo (DSMC) method, which uses probabilistic procedures derived for the dilute gas assumption only. The DSMC method described in the reference book by Bird (Ref 7) is based on the approximation that the molecular motion and the intermolecular collisions can be decoupled over a small time step. As in MD simulations, particles are moved over small distances over a time step that must itself be kept very small compared to the mean collision time, and representative molecular collisions are calculated that involve near neighbors. Cells in physical space are required for selecting the collision partners and sampling flow properties, and the simulation domain must be meshed similar to CFD simulations. Superparticles or pseudoparticles representing a fixed number (F_N) of averaged real molecules can be used when the number of real molecules becomes too large. The representative collisions are selected on a probabilistic basis by using relations derived for the kinetic theory of gases (Ref 7), which was briefly introduced in the section "Transport Regimes and Transport Equations" in this article. For instance, in a homogeneous gas, the probability $(\mathcal{P})$ of collision between two molecules over the time interval Δt can be calculated as (Ref 7):

$$\mathcal{P} = F_N \sigma_T c_r \Delta t / V_c \qquad \text{(Eq 146)}$$

In other words, $\mathcal{P}$ is equal to the ratio of the volume swept out by their total cross section moving at relative speed, c_r, to the volume of the cell (V_c).

The main limitations of the DSMC method are the assumptions of molecular chaos and dilute gas; furthermore, it cannot be used for the simulation of long-range-interaction-dominated processes such as highly ionized (hot) plasmas, but it can be applied to weakly ionized (cold) plasmas and to reacting gas mixtures as well.

Other simulation methods can be used for the simulation of transition regime flows (Ref 7), for example, the lattice gas cellular automata method (Ref 104), that have computational efficiency higher than the DSMC method but give a poor representation of the physical reality.

Modeling and Computation of Particle-Surface Interactions

This section very briefly introduces a large family of methods used for molecular modeling in computational materials science (CMS) (Ref 105). It is out of the scope of this article to discuss these methods in details; however, they are increasingly introduced in simulation of VPP, and this subsection may be helpful to the reader who is not familiar with CMS. A list of popular molecular modeling software can be found in Ref 106. It is worth noting that a few commercial packages are now becoming available for the simulation of multiple problems in materials science, such as Materials Studio (Ref 18), which offers not only ab initio, semiempirical calculation software and MD simulators but also model builders and force fields as well as powerful graphical postprocessing utilities for displaying simulation results.

Ab Initio and Related Quantum Methods. The fundamental equation to be solved for the simulation of materials microstructure and gas-surface interactions is the famous Schrödinger wave equation. Exact solutions of this equation cannot be obtained for complex real molecular systems such as gas-solid systems; however, more-or-less approximate solutions can be found using various methods. Ab initio methods are the most sophisticated ones; they use very few approximations but require a huge computational power (calculations typically scale with N^4) and are restricted to the simulation of very small-sized systems (typically, a few tens to a few hundreds of molecules). Hartree-Fock (HF) or density functional theory (DFT) methods based on electronic calculations for a molecular system with heavy nuclei at fixed positions are currently the most popular ab initio methods. For example, they can be used to calculate transition-state structures to

find the most probable surface reaction pathways in CVD (Ref 107). Commercial packages such as Gaussian 03 (Ref 108) can be used for performing ab initio calculations, but many other software programs are available. Semiempirical methods are also based on quantum mechanics but use stronger approximations for the Hamiltonian and empirically derived parameters as well. They are less accurate than ab initio methods but demand less computational resource (typically scale with N^3), and medium-sized systems of a few hundreds to a few thousands of atoms can be modeled. The MOPAC (Molecular Orbital PACkage) software (Ref 109) has also been used for performing semiempirical calculations of interest for CVD problems (Ref 110).

Molecular Mechanics and Molecular Dynamics Simulations. Models based on molecular mechanics (MM) use purely classical potentials for describing atomic interactions (Ref 111). The potential energy functions for the series of atoms (treated as spheres) and the related empirical parameters constitute a force field (FF). A rather large number of FFs have been developed (Ref 112) that are most often used in MM for energy minimization but also in MD, which aims at simulating molecular motions (see the section "Molecular Dynamics Simulations" in this article). Less accurate but faster calculations (typically scale with N^2) can be conducted using such classical empirical potentials, but semiempirical or ab initio potentials can alternatively be considered for more accurate MD simulations. Although not particularly well suited for the simulation of gas flows, MD methods have been widely used for the simulation of gas-surface interaction dynamics.

Monte Carlo Simulations. The MC methods discussed in this section are slightly different from the DSMC method introduced previously in "Monte Carlo Simulations" in the section "Modeling and Computation of Transport Equations in Transition Regime Flows." They can often be used as alternatives to energy minimization or MD methods. In the first case, the new configurations of the system containing N particles are sampled by importance; they are assigned a statistical weight, and each new configuration is either accepted or rejected after comparing it with a random number, $\mathcal{R}$, in the range [0,1[. For instance, during system energy minimization, the new configuration $n + 1$ can only be accepted if its energy E_{n+1} is lower than E_n and if the selected random number satisfies:

$$\exp\left(-\Delta E/kT\right) > \mathcal{R} \qquad \text{(Eq 147)}$$

Using this kind of classical MC algorithm, however, the evolution of the system is not the true dynamic evolution as in MD simulations. Kinetic Monte Carlo (KMC) methods can be used for simulating the system dynamics. In KMC simulations, a list of n possible events with rate R_i ($i = 1...n$) is updated at each time step, Δt, of which the total rate for the system in configuration, C, at t is:

$$R_T(C) = \sum_{i=1}^{n} R_i \qquad \text{(Eq 148)}$$

The probability for each event is then:

$$\mathcal{P}_i = \frac{R_i}{R_T(C)} = \frac{R_i}{\sum\limits_{i=1}^{n} R_i} \quad (i = 1,...n) \qquad \text{(Eq 149)}$$

One event is selected among the n possible events after comparing its probability to a random number in a uniform distribution in the [0,1[interval, and the time is also incremented on a statistical basis after choosing a new random number (ξ) in the [0,1[interval:

$$\Delta t = -\frac{\ln(\xi)}{R_T(C)} = -\frac{\ln(\xi)}{\sum\limits_{i=1}^{n} R_i} \qquad \text{(Eq 150)}$$

The KMC are usually less accurate than MD simulations, but calculations are faster and much larger periods of time can be explored.

Simulation of CVD Processes

Significant progress has been accomplished since the development of the first realistic numerical simulations of transport phenomena in CVD reactors in the 1980s (Ref 20, 113). In the 1990s, the effects of the reactor geometry, operating conditions, and so forth on the reactor performance were investigated in depth, and chemistry models were considered in reactor simulations (Ref 103 and the references therein). Complex coupled heat-transfer and fluid-flow problems were solved, and detailed models of CVD gas-phase and surface chemistry were also developed. Nowadays, multidimensional treatments of the coupled transport equations, including detailed chemistry, are performed in most reactor simulations. Figure 10 shows the typical structure of a CVD simulation. It can be seen that even if commercial software has now become available (as explained in the section "Modeling and Computation of Transport Equations in Continuous Media" in this article), numerous input data, which are not always readily available, must be gathered, which usually remains a tedious and difficult task.

A few representative examples are now briefly presented to illustrate the interest and capability of CVD reactor modeling. Several very good review articles on transport phenomena modeling in CVD reactors are of great interest for further reading on the subject (Ref 103, 114, and 115).

Fluid Flow and Heat Transfer in Conventional CVD Reactors

Simple Models. The transport equations presented previously in the sections "Macroscopic Conservation Equations in Pure Gases" and "Conservation Equations in Multicomponent Gaseous Mixtures" are usually strongly coupled, and their resolution may become very difficult if complex reactor geometries are considered. However, several situations are encountered for which strong simplifying assumptions can be made. Hot-wall CVD reactors have often been considered as isothermal (this is a reasonable assumption in the hot zone of horizontal multiwafer low-pressure CVD reactors) (Ref 116), and in this case, there is no need to solve the energy equation. The axisymmetric flow approximation has often been considered in vertical reactors also; furthermore, when reactants are highly diluted in a carrier gas, species conservation and momentum-transfer equations can be decoupled, which considerably simplifies the numerical treatment. Two-dimensional models based on the FV and
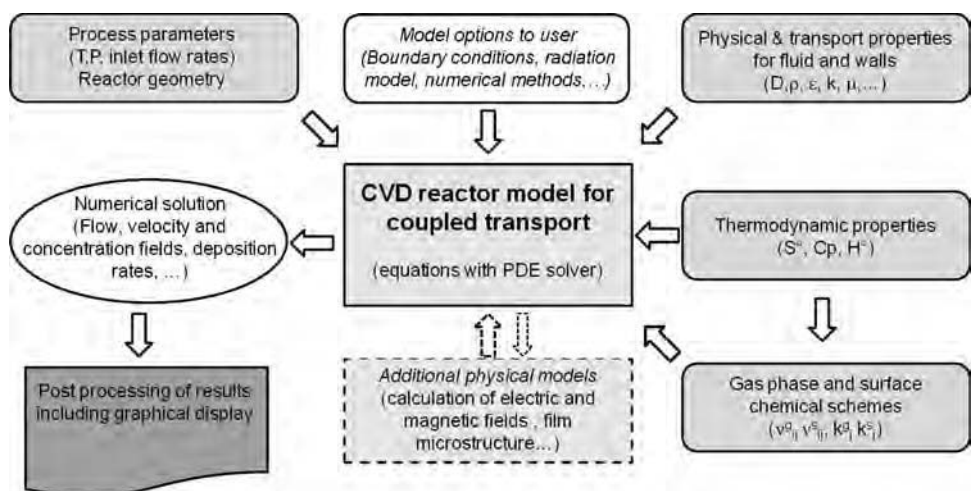

Fig. 10 Structure of a chemical vapor deposition (CVD) reactor simulation. PDE, partial differential equation

FD methods have often been used to simulate nonisothermal-axisymmetric flows (Ref 92, 95) and/or isothermal flows (Ref 93, 116). Even simpler one-dimensional models have also been considered in a few cases that usually consider more detailed chemical schemes (Ref 114 and references therein).

Effects of Reactor Geometry and Operating Conditions in Mixed Convection Problems. One of the key issues in CVD operations is to obtain films of desired composition with a good uniformity and, if possible, at high deposition rate. Operating conditions (T, P, inlet flow composition and rates) and reactor geometry both affect film characteristics, but the latter cannot be modified, and it is highly desirable to optimize both the reactor shape and operating conditions prior to building the reactor. Due to the strong coupling between transport phenomena, it is difficult to draw universal conclusions on CVD process behavior; most experimental or simulation results are specific to the selected operating conditions and reactor geometry. To derive more universal results and conclusions, the transport equations are often scaled, and dimensionless variables and groups (see the section "Laminar and Turbulent Flows" in this article) can be considered in numerical simulation studies (Ref 114, 115). In a review article, Holstein (Ref 115) discussed the influence of the design of reactor entrance regions as well as the effect of buoyancy forces on the flow field and deposition rate in various types of CVD reactors: horizontal, barrel, and vertical reactors (chimney and rotating-disk reactors). The Rayleigh number (Ra) and the Grashof/Reynolds numbers (Gr/Re^2) ratio (ratio of buoyancy to inertial forces) are the critical dimensionless numbers in horizontal reactors; the formation of longitudinal roll waves and transverse roll waves are observed at high values of Ra and Gr/Re^2, respectively, and even more complex or time-dependent flows are predicted at very high Ra and Gr/Re^2 (Ref 115 and references therein). Operating a CVD reactor under such conditions is detrimental to growth-rate uniformity. In vertical chimney reactors, the flow is determined by the ratio of buoyancy to viscous forces, that is, by the ratio Gr/Re. The same ratio governs flow instabilities in barrel reactors, but in this case, even more complex flows may result from the susceptor rotation, for example, formation of Taylor vortices at high rotation speed. In stagnation-point flow and rotating-disk reactors, Ra and Gr/Re^2 are the critical dimensionless groups. Downward-forced flows are usually employed for convenience, but inverted upward-flow reactors do not exhibit strong recirculation instabilities because forced and natural convection have the same direction in this case.

Fotiadis et al. (Ref 100, 117) have performed an in-depth analysis of the effects of operating conditions, reactor geometry, and heat-transfer characteristics on growth-rate uniformity and flow patterns in vertical nonisothermal
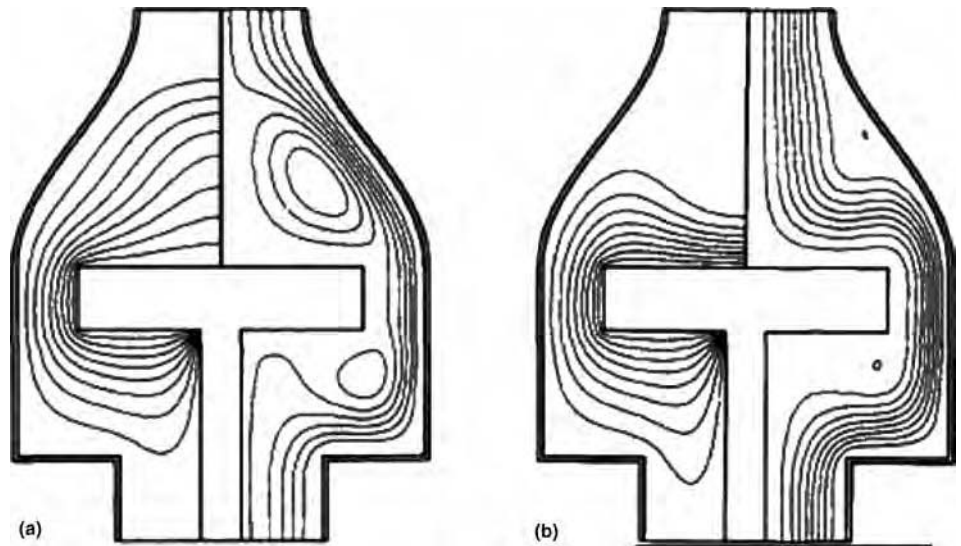
Fig. 11 Effect of the inlet flow rate on streamlines (right) and isotherms (left). (a) Inlet flow rate = 4 SLM (standard liters per minute). (b) Inlet flow rate = 8 SLM. H_2 pressure, 0.13 atm; $T_{susceptor}$, ~900 K. Source: Ref 100

axisymmetric metal-organic CVD (MOCVD) reactors used for GaAs deposition. They have developed 2-D and 3-D models based on the Galerkin FE method to simulate the complex behavior of such reactors, in which forced and natural convection effects often superimpose in mixed convection flows. They have evidenced recirculation cells due either to natural convection induced by large temperature gradients or to flow separation associated with sudden area expansions. These undesirable effects, which are detrimental to uniformity, can be reduced by increasing the inlet flow rate (Fig. 11), rotating the susceptor, reducing the pressure, inverting the reactor, shortening the inlet-susceptor distance, introducing baffles, and reshaping the reactor.

In their well-documented analysis, these authors also evidenced flow-symmetry breaking in axisymmetric geometries due to highly nonlinear effects in mixed-flow problems (Ref 118). In addition, time-varying or multiple stable flow fields may exist for the same reactor configuration and operating conditions in a restricted range. These nonlinear effects were further investigated in subsequent studies by Van Santen et al. (Ref 119,120). They used a 3-D FV model to study the origin of symmetry breaking in a cold-wall stagnation-flow CVD reactor (Ref 119). For a reactor height-to-diameter aspect ratio (H/D) larger than unity and Rayleigh numbers between 2×10^3 and 10^5, multiple axisymmetric and nonaxisymmetric flows (Fig. 12) were found. Symmetry breaking can be suppressed by selecting low inlet flow and low rotation speed of the susceptor. In such configurations, however, it is difficult to know if a solution is unique or not. In a subsequent paper (Ref 120), the same authors implemented arc-length continuation techniques to compute a solution branch as a function of one selected operating parameter and to track

solutions through turning points. They have shown that the nonlinear interaction between the convection term in the energy equation and the gravity term in the momentum equations is responsible for turning-point instabilities resulting in multiple stable flows.

Thermal instabilities in axisymmetric forced radial flow between horizontal plates were also studied by Van Santen et al. (Ref 96). In this case, mixed convection leads to 3-D rolls in conditions of low Reynolds (Re) and Prandtl (Pr) numbers and high Rayleigh (Ra) numbers, similar to the well-known Rayleigh-Bénard problem of convection between two infinite parallel plates at different temperatures. Of course, such effects cannot be captured by 2-D models using axisymmetry assumption; full 3-D numerical treatments are required. At sufficiently high operating pressure and large reactor size, buoyancy effects may become so important that flow may turn turbulent. In another paper (Ref 121), the same authors studied the effect of turbulence on the CVD process in the same kind of stagnation-flow cold-wall CVD reactor. They performed large-eddy simulations (LES) rather than Reynolds-averaged Navier-Stokes simulations because the latter are not capable of correctly resolving the characteristic features of such turbulent Rayleigh-Bénard convection problems. They used a parallel algorithm to reduce the computation time inherent to LES. The heat flux was increased by turbulence but became nonuniform. Nevertheless, turbulent mixing provides good uniformity conditions in a large part of the wafer and therefore may favor higher and more uniform growth rate.

Fotiadis et al. also underscored the importance of the heat-transfer treatment on the accuracy of model prediction, which must include appropriate boundary conditions as well as radiation heat transfer in cold-wall configurations

rate constants of adsorption/desorption processes and surface reactions were estimated from statistical mechanism, transition-state theory, and Benson's method of bond-dissociation enthalpies (Ref 19, 37) by considering different surface sites (ledges, terraces) according to the Kossel representation of the surface (see the section "Modeling of Surface Interactions with the Vapor Phase" in this article). Quantum Rice-Ramsperger-Kassel (QRRK) calculations (see the section "Pressure Dependence of Rate Coefficients") were performed to account for the pressure dependence of homogeneous reactions. The most important sp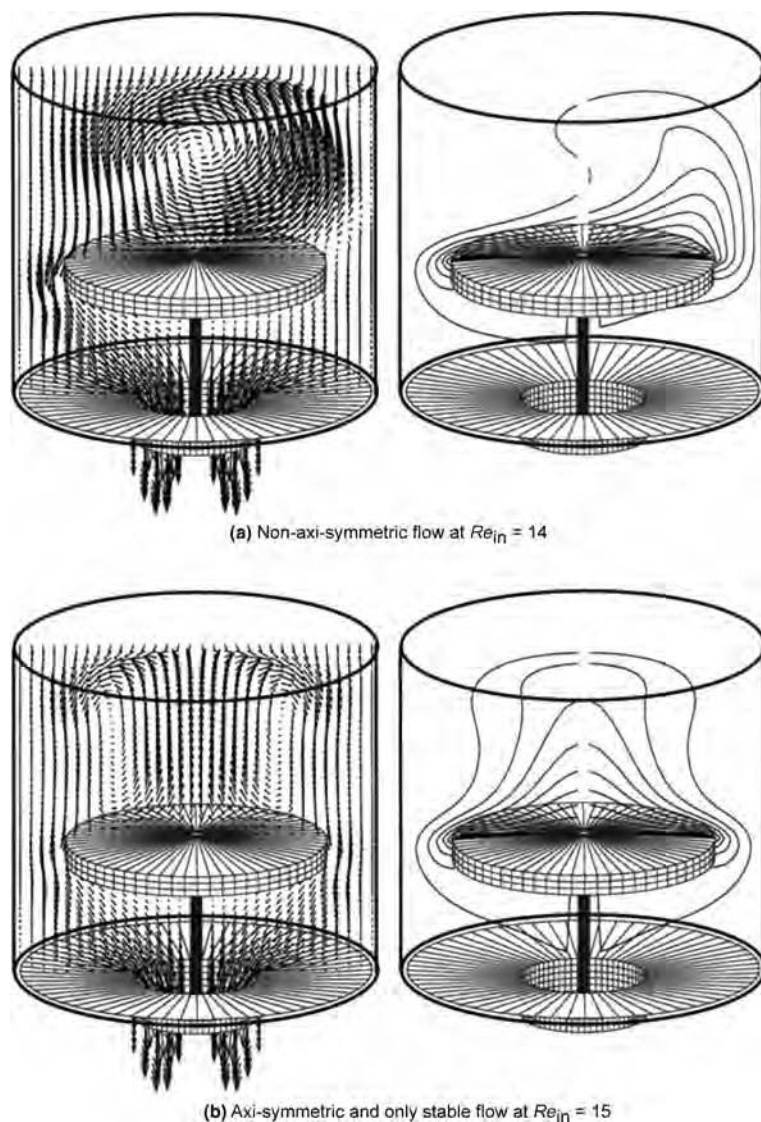ecies and reactions (7 gas-phase and 18 surface reactions) were finally identified. After the two previously mentioned landmark papers, several chemical mechanisms were developed, and more and more simulations of CVD processes have included detailed chemistries; however, most works have focused on only two to three chemical systems of interest, such as silicon deposition from silane and carbon deposition from hydrocarbon species. In the latter case, many kinetic data were available from combustion works (as seen in the section "Construction"), which allow considering very large reaction mechanisms. Recently, Norinaga et al. (Ref 66) developed a 227 species, 827 reactions mechanism. A reduced version of their mechanism (Ref 128) has been considered for simulating transport phenomena in chemical vapor infiltration of pyrolytic carbon by using COMSOL (Ref 89), a FE method-based commercial software.

However, except for systems of interest for combustion and a few other systems, such as Si-H (Ref 40, 41, 65), Si-C-Cl-H (Ref 129), Ti-Cl (Ref 50), GaAs (Ref 127), and Si-O-H (Ref 130, 131), and so on, most chemical mechanisms developed so far do not account for the pressure dependence of gas-phase reactions. Furthermore, the handling of numerous stiff chemical equations remains quite difficult even today (2009) (Ref 132), and it is usually desirable to reduce the chemical scheme to a reasonable size by using the rigorous methods presented in the section "Construction, Analysis, and Reduction of Gas-Phase Reaction Mechanisms" in this article. Another major interest of using mechanisms of reasonable size in complex CVD process simulations is that it allows a better and easier identification of the key deposition pathways (Ref 71, 133), provided that the simplified mechanism has been obtained from a rigorous reduction procedure. As an illustration, Fig. 14(a) shows a comparison between the calculated contributions of species using two simplified mechanisms for silicon-carbide deposition from C_3H_8/ SiH_4/H_2 mixtures (Ref 134). Predictions from the oversimplified (10 species, 13 reactions) mechanism, which do not include organo-silicon species (Ref 134 and references therein), overestimate CH_3 contribution to deposition and do not capture the important participation of the Si_2C species. Furthermore, the simulated

(a) Non-axi-symmetric flow at Re_{in} = 14

(b) Axi-symmetric and only stable flow at Re_{in} = 15

Fig. 12 Transition from nonaxisymmetric to axisymmetric flow when increasing the inlet Reynolds number (Re_{in}) from 14 to 15. Rayleigh number = 10,000; Ga = 0.5 (Ga is the dimensionless Gay-Lussac number, defined as ($T_{wafer} - T_{wall}$)/T_{ref}); Prandtl number = 0.71; no susceptor rotation. Velocity and temperature fields in a cross section are plotted on the left and right sides, respectively. Source: Ref 119

(Ref 100). Such a rigorous treatment is essential to obtain an accurate solution of the so-called "cold-finger" phenomenon experimentally evidenced in horizontal cold-wall CVD reactors (Ref 122). The cold finger arises from radiative heating of the reactor walls in front of the heated susceptor and the subsequent heat transfer from these walls to the inlet cold gas (Ref 123). Three-dimensional buoyancy-driven secondary flows in horizontal cold-wall MOCVD reactors were first simulated by Moffat and Jensen (Ref 124), who used a 3-D Galerkin FE method to study the conditions leading to the development of convection rolls (Fig. 13). These authors considered various heat-transfer conditions at reactor walls (adiabatic or cooled walls), and they also showed that the flow structures are very sensitive to the reactor cross-sectional aspect ratios and to the nature of carrier

gas as well. Convection rolls were found to be detrimental to film thickness and composition uniformity.

Complex Reacting Flows. Earlier simulations of the CVD process considered only lumped chemistry. In their 2-D simulation of silicon deposition from silane, Coltrin et al. (Ref 22, 125, 126) considered the first detailed gas-phase chemical scheme, which involved 17 species and 27 elementary reactions. Tirtowidjojo and Pollard (Ref 127) developed the first detailed chemical scheme including both surface and gas-phase reactions. They first considered 60 species and 232 gas-phase reactions, together with 19 species and 115 processes at the surface. This mechanism was used to simulate GaAs deposition from a Ga(CH$_3$)$_3$-AsH$_3$ mixture in a MOCVD reactor. The most important species and reactions were identified. The

Fig. 13 Transverse velocities at three axial positions, illustrating the formation of convection rolls inside an adiabatic-wall horizontal metal-organic chemical vapor deposition reactor. Source: Ref 124

and experimental deposition-rate profiles are not in very good agreement when considering the oversimplified chemistry (Fig. 14b).

The simulation of GaAs, carbon, or silicon-containing film deposition received far more attention, but several other chemical mechanisms have also been developed for the simulation of metallic film CVD, for example, for tungsten (Ref 136, 137), titanium (Ref 50), and copper (Ref 138) deposition, which are more or less detailed.

Beyond the difficulties inherent to the resolution of a large number of coupled stiff equations when considering large reaction mechanisms, one additional difficulty arises from the strong coupling between species conservation equations and momentum + energy conservation equations. Most mixed-convection problems have focused on thermally driven natural convection, but buoyancy effects resulting

from solutal convection must also be considered when large changes in gas-phase composition occur due to chemical reactions (Ref 114). Particular attention must also be paid for describing diffusion phenomena in complex multicomponent mixtures (see the sections "Conservation Equations in Multicomponent Gaseous Mixtures" and "Some Important Macroscopic Properties of Vapors from the Molecular Model" in this article). Compositional uniformity of films may be affected by thermal and ordinary diffusion factors in diffusion-limited situations, because deposition species with the highest diffusion coefficients are transported faster to the substrate, and species with the highest thermal diffusion factors are transported preferentially toward the colder walls (Ref 115). This is particularly important in the presence of large temperature and concentration gradients and also when there are large

differences in molar masses and diffusion coefficients. Kuijlaars et al. (Ref 139) have shown that in this case, solving the Stefan-Maxwell equations (Eq 46) is much better than simply using an effective diffusivity (Eq 47 or 70b). The Damköhler number (Da), which can be defined as the ratio of reaction rate to transport rate, can be used as an indicator of growth-rate uniformity in some cases; when surface reactions are fast compared to the diffusion of deposition precursors, Da is low, which means that depletion effects may induce large nonuniformities in film thickness and composition along the flow direction (Ref 114). Thermal diffusion effects have also been found to have a nonnegligible influence on the deposition process behavior in horizontal cold-wall reactors, in particular by decreasing the deposition rate as a result of decreasing the overall mass transfer rate to the hot susceptor surface (Ref 99,134).
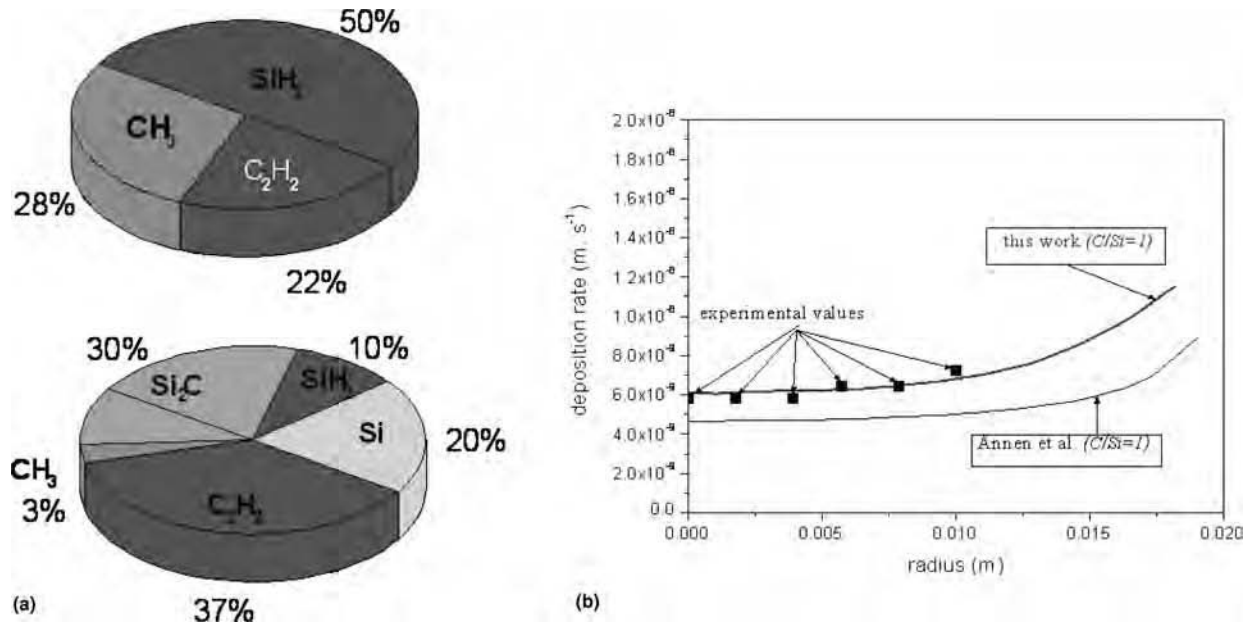
Fig. 14 (a) Theoretical molar contributions of the most important reacting species to the deposition mechanism. Top: 10 species and 13 reactions (Source: Ref 135). Bottom: 18 species and 33 reactions. (b) Calculated deposition-rate profiles obtained from the two different reaction mechanisms using a stoichiometric constraint compared with the experimental data. Source: Ref 134

Finally, thermophoretic transport effects are also of key importance in nonisothermal CVD processes involving particle formation in the gas phase, such as aerosol CVD (Ref 131).

Transport Phenomena at High Knudsen Numbers

A large number of papers have been published on simulation of transport phenomena in CVD reactors; however, the literature dedicated to simulation of transport in the transition regime is less abundant.

Transport in Low-Pressure CVD Reactors. Several previous works have used a continuum approximation for modeling industrial multiwafer low-pressure CVD (LPCVD) reactors (Ref 114, 140). However, these reactors operate at approximately 10 Pa, in the transition regime; hence, a correct treatment of transport should be based on the resolution of the Boltzmann equation (Ref 140). DSMC simulations were performed to simulate transition regime flows in a multiple-wafer LPCVD reactor (Ref 140), which demonstrated that important diffusion effects cannot be captured by continuum models. Coronell and Jensen (Ref 141) used Monte Carlo techniques to model both heat transfer and fluid flow in a multiwafer ultrahigh-vacuum CVD reactor used for epitaxial deposition of silicon or germanium. This reactor is similar to a conventional industrial LPCVD reactor but operates in the 0.1 to 1 Pa pressure range. The heat-transfer calculations have been separated from the flow simulations because the gas is transparent to radiation at very low pressure, and experiments show that the wafer temperature is independent of both gas flow rate and composition. Heat-transfer calculations include a
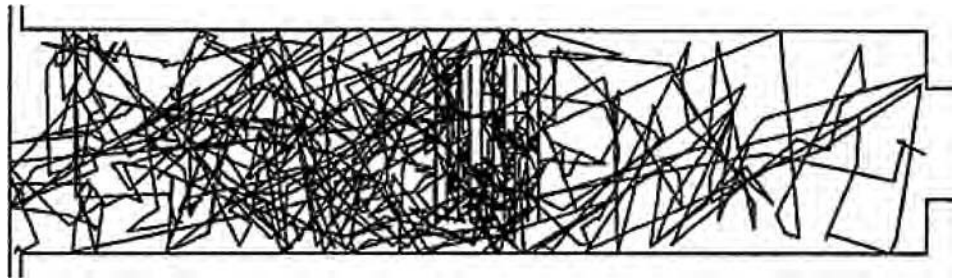
Fig. 15 Typical SiH₄ molecule trajectory plot for an ultrahigh-vacuum chemical vapor deposition gas flow. Source: Ref 141

conventional FD treatment of heat conduction at quartz walls and an MC treatment of radiation. The 2-D MC simulation allows the inclusion of detailed radiation properties for walls and wafers, that is, depending on wavelength, direction, temperature, and nature of materials used in the different reactor zones. The simulation method consists of generating a large number of representative photons from each surface, which are then tracked until they interact with another surface of the reactor enclosure. Simulation is performed by using unbiased probability distribution functions for emission of photons from surfaces, as well as probabilities for absorption, reflection, or transmission of photons incoming on surfaces; hence, it is a DSMC. The DSMC method (Ref 7), briefly described in "Monte Carlo Simulations" in the section "Modeling and Computation of Transport Equations in Transition Regime Flows," has been adapted to simulate the flow in the UHV-CVD reactor in two dimensions. An arbitrary initial state is specified for performing a

transient simulation that achieves a steady-state flow after a sufficient number of time increments. Gas-phase collisions are treated as explained in "Monte Carlo Simulations" in the section "Modeling and Computation of Transport Equations in Transition Regime Flows," and gas-phase reactions are neglected. Molecules impinging upon the surface are either reflected or incorporated by selecting a random number in a uniform distribution in the [0-1[interval and comparing it to the reactive sticking coefficient of the molecule. Diffuse reflection is assumed at the surface, with complete thermal accommodation to the surface temperature. The theoretical partial pressure and temperature profiles of species were calculated, as well as the deposition rate profiles. Figure 15 shows a simulated trajectory plot for a SiH₄ molecule. Simulations revealed a longitudinal depletion of the most reactive species and a strong separation effect between species of different molecular weight, due not only to thermal diffusion but also to pressure diffusion (Ref 141).

Other examples of the use of the DSMC method for CVD problems can be found in Ref 142.

Transport on the Feature Scale. Some CVD models have been specifically developed for predicting deposition profiles in microscopic trenches or via holes. Even if the main flow inside the reactor is a continuous flow ($Kn < 0.01$), transport becomes rarefied inside features of microscopic size. Distinction can be made between models that consider gas-phase collisions and models that do not, the latter being more adapted for simulation of line-of-sight processes, which are most often PVD or etching processes. In a pioneer work by Ikegawa and Kobayashi (Ref 143), the DSMC method was used to simulate deposition profiles in small trenches, not only in atmospheric CVD situations but also in collision-less plasma (PACVD) and sputtering (PVD) deposition conditions. They considered various values of the surface-reactive coefficients of incoming precursor molecules that were treated as hard spheres of a simple gas, and distinction was made between reactant molecules and products. A Boltzmann distribution was assumed for incoming molecules entering the domain. The cell size in DSMC is a fraction of the mean free path of molecules, and the time increment is smaller than the mean collision time. A string algorithm is used, and the moving film surface is approximated by a series of nodes joined by straight-line segments, which individually move in proportion to their flux of incoming molecules but in the opposite direction; hence, cells are distorted during the simulation of growth. Better step coverage is obtained for the lower values of sticking coefficients, in agreement with experimental results. In a subsequent publication, Coronell and Jensen (Ref 144) also used the DSMC method to simulate deposition in microscopic features, but a pileup method was preferred to the string algorithm used by Ikegawa and Kobayashi because of its ability to capture microstructural details.

Beyond molecular collisions and species incorporation at surfaces, adsorption, desorption, diffuse reflection, as well as surface diffusion were considered. Two- and even 3-D simulations were performed. Figures 16(a and b) show an example of simulated profiles obtained for two different sticking coefficients of the depositing species, whereas Fig. 16(c) illustrates the deviation of the velocity distribution function from the Maxwellian distribution at the vicinity of the growing surface, which is due to the rarefaction of the gas. Continuum approaches have also been proposed for the simulation of CVD or PVD over topography at the feature scale (Ref 145), which have interesting predicting capabilities. More sophisticated models were developed later that allow consideration of several reacting species as well as multiscale effects; these models are presented in the section "Multiscale Modeling" in this article.

Simulation of PVD and Etching Processes

As explained at the beginning of this article, most PVD processes operate at low or very low pressure; hence, transport phenomena are usually (although not always) of less importance in PVD compared to CVD. Many PVD processes are line-of-sight processes in which gas-phase reactions can be ignored, but in turn, as seen earlier in the section "Physical Vapor Deposition and Related Processes," they often involve energetic particles in ionized media, thus giving rise to various and complex physical phenomena. A large number of PVD techniques are currently available, and more-or-less sophisticated models, which are usually multiphysics models, have been developed for simulating these processes. Simulation of particle-surface or beam-surface interactions is a key issue in PVD modeling as well as modeling of plasma discharges. In this section, examples focusing on sputtering deposition and reactive

or ion beam etching are discussed. Other examples of PVD processes are discussed in the section "Multiscale Modeling" in this article.

Reactive Sputtering Deposition

Berg (Ref 146, 147) proposed a simplified but elegant approach to model reactive sputtering (RS) processes. The derivation of Berg's model is based on strong assumptions (constant ratio between the electron and ion currents, uniform sputtering of the target, and uniform collection of sputtered particles at the substrate, etc.) and on several balances that lead to simple expressions of the sputtering and deposition rates. This simple model is capable of predicting the hysteresis classically observed in RS when plotting the sputtering rate of the target versus the reactive-gas flow rate. The hysteresis phenomenon, which is due to target poisoning, is particularly important in titanium and TiN RS deposition (Ref 4). Berg's model can be used to study the influence of target material, reactive gas, target-to-substrate distance, and so on and can be extended to simulate more complicated configurations, such as co-sputtering and sputtering from an alloy target or processing with several reactive gases, pulsed dc, and so on. Other authors have included this very simple zero-dimensional model in more sophisticated treatments, such as Kobayashi (Ref 148), who developed a 2-D model of a large-area TiN sputtering deposition system based on Berg's model and the DSMC method to simulate rarefied gas transport. Four gaseous neutral species were included in his analysis (N_2 and argon as inlet gases; titanium and TiN as sputtered species that were tracked during the simulation). The space was meshed for DSMC, but the target and the walls were also divided into cells, with boundary conditions based on Berg's model in each of these cells. Simulations predicted a nonuniform composition of the TiN film resulting from depletion of N_2. Dynamic Monte Carlo simulations based on the TRIDYN computer program (Ref 149) were performed to study implantation and ion bombardment effects that may decrease the sputter erosion rate (target poisoning) (Ref 150, 151). In TRIDYN, the target composition and the binding energies between the different elements of the target must be defined, and particle transport inside the target during ion bombardment can be dynamically simulated. Various energies and incidence angles can be considered for species in the beam. Depth profiles of atomic species in the target, as well as sputtering yields and re-emitted amounts, can be predicted. It is worth noting that plasma effects were not considered in the aforementioned analyses. To account for electron interactions, test-electron Monte Carlo simulations were performed (Ref 152 and references therein), but self-consistent simulations of plasmas must be preferred. Clenet et al. (Ref 153) developed a 2-D model of a circular

LIVE GRAPH
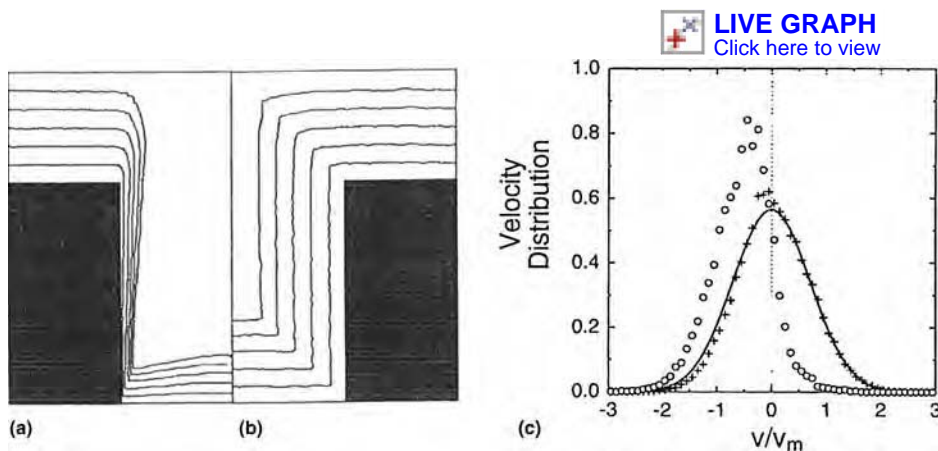Click here to view

Fig. 16 (a, b) Simulated profiles at various stages of growth for an infinitely long trench, with Knudsen number = 1 and a sticking coefficient of (a) s = 1 and (b) s = 0.01. (c) Direct-simulation Monte Carlo-computed perpendicular velocity distribution. $\lambda/3$ (o) and 3λ (+) above the growth surface and the Maxwellian distribution (−). Source: Ref 144

planar radio-frequency magnetron sputtering system used for titanium-tungsten film deposition. The model is based on a particle-in-cell Monte Carlo collision (PIC-MCC) approach for simulating the transport of sputtered species in the argon plasma. The SRIM code (Ref 23, 154), which is open access (Ref 155), was used to calculate the sputtering yields of titanium and tungsten target atoms. PIC-MCC models as well as hybrid fluid/particle models have been developed for more than 15 years to simulate radio-frequency magnetron sputtering discharges (Ref 156). Approaches based on PIC-DSMC methods have been proposed to self-consistently simulate plasma and neutral gas as well as sputtering and deposition, as explained in Ref 156. Magnetron sputtering models are reviewed in Ref 152 and 156.

Some models have also been developed for simulating film-growth dynamics and microstructure formation in magnetron sputtering processes, either on flat surfaces or in microscopic trenches (see the section "Modeling Line-of-Sight Deposition or Etching over Topography" in this article). Kinetic Monte Carlo (KMC) models are most often used for this purpose, as well as MD models. The KMC models require detailed energies of the possible atomic configurations in order to calculate the corresponding probabilities. Good interatomic potentials (Ref 157, 158) are currently available from first-principles (ab initio) for a variety of materials, and MD calculations can also provide data for KMC simulations. A representative example of KMC simulation of growth in a magnetron sputtering process can be found in a paper by Gilmer et al. (Ref 159), who simulated the film morphology of aluminum films obtained by magnetron sputtering. The MD simulations were first performed to obtain the kinetic energies and the angle of the sputtered atoms as well as the potential energies of the surface atoms. Figure 17 shows KMC simulation results, illustrating the competitive growth process leading to a polycrystalline film structure.

These simulations provide considerable insight into the growth mechanism and the influence of parameters and suggest possible routes to improve the film structure. More sophisticated multiscale models (see the section "Multiscale Modeling" in this article) of magnetron sputtering deposition have now been developed that have the capability of simulating both plasma generation and film deposition at different scales (see Ref 160 and the references in Ref 152).

Modeling Line-of-Sight Deposition or Etching over Topography

Ultralow-pressure PVD processes, such as evaporation, molecular beam epitaxy, or ion beam etching, typically involve no collisions in the gas phase; hence, there is theoretically no need for real transport modeling for those processes, but usually, both the trajectories of particles traveling from the source to the substrate and the particle energy distribution functions must be computed. Most simulations performed to date aimed at predicting the microstructure or the uniformity of deposited films, not only on flat surfaces but also on microscopic features. Line-of-sight models have also been used for simulating magnetron sputtering processes, neglecting collisions in the gas phase (Ref 159, 161). Simple ballistic aggregation models were first proposed to simulate columnar microstructures and deposition profiles in microscopic trenches (Ref 162–164). The orientation angle (β) of the columns with respect to a vector perpendicular to the substrate surface differs from the incidence angle (α) of incoming particles and obeys the following experimentally verified relation: 2 tan (β) = tan (α) known as the tangent rule (Ref 165). This is mainly due to a self-shadowing effect. These simple but useful models were mainly used to simulate deposition in microscopic features. Integro-differential equations can also be derived for describing free molecular flows in trenches (Ref 166). The difficulty in this case arises from the multiple reflections of particles inside the features, which complicate the calculations of the shape or view factors (Ref 166, 167). As discussed in Ref 159 and 161, continuum models can be very efficient in simulating the step coverage and topographic evolution for moving interfaces, either for deposition or etching over topography. Approaches based on the level-set method (Ref 168) have been introduced as an alternative to front-tracking and segment-based methods (Ref 161).

Much more sophisticated models allowing the prediction of both topographic evolution and microstructure formation are based on KMC methods and often make use of first-principles calculations and MD simulations. Figure 19(c) shows the sputtered aluminum deposition profile in a submicron-sized trench obtained from KMC simulations performed in Ref 159.

Advanced Topics

Modeling of vapor processes includes advanced methods, such as when additional physical processes are coupled with the vapor process. Examples of multiphysical simulation models are given subsequently on vapor deposition processes coupled with cold plasma discharges and fluidized beds. Another advanced topic is the cross-disciplinary approach of multiscale modeling, which is illustrated through several examples of vapor deposition (CVD and PVD) over micron-sized features.

Modeling Complex/Multiphysics Processes Involving Vapors

Example 1: Plasma-Enhanced CVD (PECVD). Plasmas are often encountered in PVD processes such as magnetron sputtering, but they are also widely used in combination with CVD in the so-called plasma-enhanced (or plasma-assisted) CVD processes. In plasma-assisted processes, reactions are initiated by charged species formed in the cold gas discharge (mainly electrons) rather than by thermal activation of the reactant molecules. The presence of charged species in an electric field induces many new physical or chemical phenomena (complex energy transfers between particles, ion-molecule reactions, substrate bombardment by ions, etc.), which are strongly coupled, and seriously complicates the modeling task (Ref 169). One of the most fundamental steps in PECVD simulation is to model the plasma discharge itself. Many review papers have been published on plasma modeling (Ref 170), which is a wide field of research by itself. Several types of models have been developed for simulating plasmas: fluid models, which are based on a continuum approximation and are not valid in rarefied flow conditions, and particulate models such as particle-in-cell Monte Carlo collision (PIC-MCC) models, which are more accurate (but require time-consuming calculations) and can also be used in transition regime flows, and finally hybrid (particle-fluid) models (Ref 170). The development of 3-D or even 2-D plasma models in complex geometries remains a difficult task; furthermore, even if collision-impact cross sections and transport parameters are available for several simple gases (rare gases and some diatomic gases) and a few polyatomic molecules of interest for PECVD, such as SiH_4 (Ref 35), data are still lacking for many important reactant molecules. When possible, for example, when reactants are highly diluted in a carrier gas and/or

Fig. 17 Growth of aluminum onto a flat substrate (dark gray) with a cosine angular distribution for the impinging aluminum atoms. (001) clusters (light gray) and (111) clusters (intermediate gray) are permitted to nucleate and grow simultaneously using a multilattice model. Configurations from left to right correspond to film at various times. Source: Ref 159

when depletion effects are not important, the plasma discharge and the CVD reactor models can be decoupled. In this case, the rates of dissociation of reactants by electron impact are obtained from the electrical plasma discharge model and used as input data to the CVD reactor model (Ref 169, Ref 170, Ref 171); otherwise, the transport of both charged and neutral species must be simulated in a single but more complicated model. In spite of these difficulties and limitations, sophisticated 1- and 2-D PECVD reactor models involving complex chemistries have been developed (Ref 172).

Example 2: Fluidized-Bed CVD (FB-CVD) consists of depositing films over particles forming a bed that expands when swept by a gas flow above a critical value. The FB-CVD and related processes, such as spouted-bed CVD, ensure very good gas-solid contact due to particle circulation (for a review of FB-CVD, see Ref 173). However, the simulation of FB-CVD processes is a very difficult task due to the complexity of the coupled-transfer phenomena problem in a two-phase unsteady medium. Simplified engineering models were first developed for the simulation of FB-CVD or spouted-bed CVD that consist of dividing the reactor in several regions where plug-flow or perfectly mixed conditions can be assumed and deriving simple mass and energy balances for the calculation of transfer rates (see the references in Ref 173). These models have now been encompassed by 2-D and 3-D two-fluid Eulerian models and discrete-element Lagrangian models. Simulations based on Lagrangian models demand huge computational resources and are most often restricted to the

study of unreactive flows. A two-fluid approach that considers the two phases of the fluidized bed as continuous and interpenetrating media has been used for the simulation of silicon deposition from silane (Ref 174, 175), using the open-access software MFIX (Ref 176).

A few commercial CFD multiphysics software, such as COMSOL (Ref 89), CFX (Ref 87), CFD-ACE (Ref 86), and FLUENT (Ref 84), have integrated modules for plasma simulation or/and fluidized-bed reactor simulation.

Multiscale Modeling

Multiscale modeling is a very interesting cross-disciplinary approach for the simulation of VPP. This section briefly reviews the basic ideas of multiscale simulation in materials science before illustrating its application to VPP through a few representative examples. Several books or comprehensive review papers are cited throughout this section for further reading on the subject.

Basic Ideas of Multiscale Modeling. Tools and methods are developed in computational materials science that aim at understanding the mechanisms underlying the formation of materials microstructure from various processes as well as predicting and optimizing materials properties. As discussed earlier in the section "Modeling and Computation of Particle-Surface Interactions," a large panel of numerical models is currently used in computational materials science (Ref 104, 105) for the simulation of various phenomena at different length and time scales (Fig. 18). Ab initio methods such as

DFT that model electronic interactions between a few hundreds of atoms or less are used to calculate surface or transition-state structures as well as reaction pathways (Ref 105). Ab initio MD calculations can be used to simulate atomic motions; they are restricted to a small number of atoms and to very short time durations (typically less than a picosecond), but the use of semiempirical or empirical potentials allows consideration of 10^7 to 10^8 moving particles over time scales of typically 1 ns. Dynamic simulations based on KMC models also consider up to 10^8 atoms or more; they are usually lattice based and less accurate than MD simulations, but atomic motions can be tracked over much longer periods, up to 10^3 s, and they cover both the atomic and microscopic scales. Other models, such as cellular automata, level set, geometrical models, and so on, aim at simulating phenomena occurring at the microscopic to mesoscopic scale, whereas continuous models based on FE and FV methods are used at the mesoscopic or macroscopic length scales, for example, for the simulation of transport phenomena in VPP.

As soon as a single type of model cannot capture all the time and length scales required for the simulation of a particular problem, a multiscale approach must be considered. Multiscale methods should always be used for full simulations of VPP processes, because the length scales of interest range from 10^{-12} to 10^{-10} m for electronic and atomic interactions, up to $\sim 10^{-6}$ m for film thickness or microstructure, and to approximately 1 m for the size of the vessel. The time scales range from 10^{-15} to 10^{-12} s for atomic relaxations or vibrations to several minutes or hours for completion of film growth.

Strategies for Multiscale Modeling. Theoretically, first-principles calculations based on the resolution of the Schrödinger equation could be performed to accurately simulate the dynamics of materials characteristics in a given elaboration process by considering the electronic interactions between all the atoms present in the system and tracking the atomic motions over the whole time scale of the process. From a practical sense, of course, such a simulation is totally intractable, but alternatively, multiscale strategies that consist of wisely combining several of the models presented in Fig. 18 can be considered. The development of a multiscale model is based on very few general guiding principles. In fact, there are two main multiscale approaches:

Hierarchical (or sequential) modeling consists of establishing a hierarchy of length and time scales, defining the elementary objects (atoms, clusters, etc.) to be handled on the selected scales, and identifying the irreducible and independent processes for these scales. Processes and objects can be averaged at a given scale to be transferred at the immediate upper scale. A bottom-up approach consists of starting simulation from the lowest scale and then calculating and transferring averages
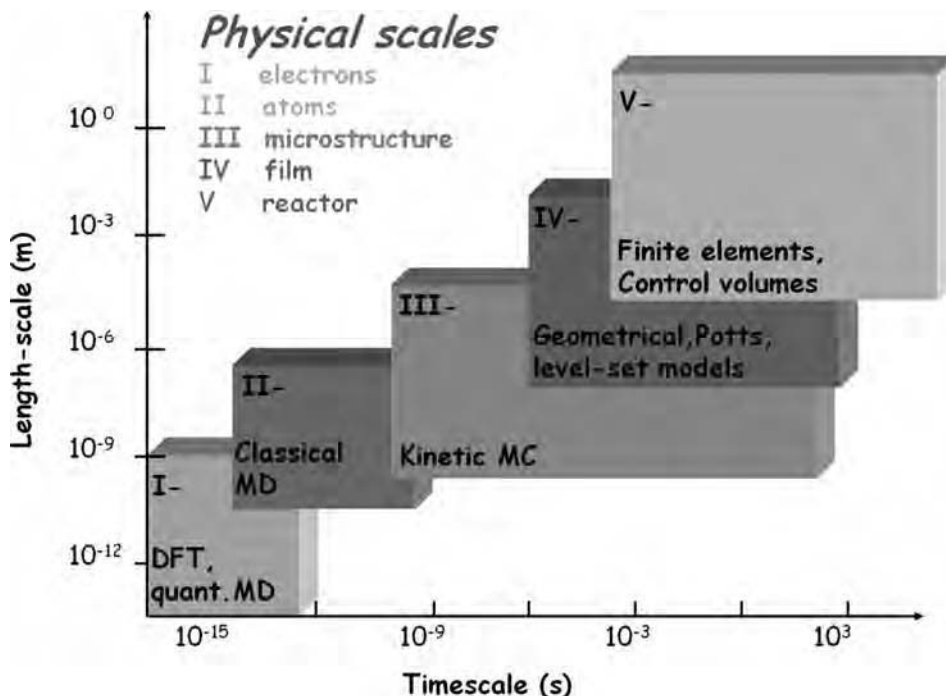

Fig. 18 Main simulation models used in materials science and related length and time scales. DFT, density functional theory; MD, molecular dynamics; MC, Monte Carlo. Source: Ref 177

sequentially to the upper scales. The reverse process (from upper scales to lower scales) is referred to as the top-down approach.

Concurrent analysis consists of using simulation methods that allow consideration of several scales simultaneously.

Except for the guiding principles summarized previously, there is no concrete universal method or algorithm available for performing multiscale simulations. Several approaches combining various models can often be used for solving a particular problem.

Examples of Multiscale Simulations of VPP. This section discusses a few examples of multiscale simulations of CVD or PVD processes. Several interesting papers have also addressed or reviewed this topic (Ref 177–181). There is abundant literature on the subject; examples can be found (nonexhaustive list) for ionized PVD (Ref 182–185), pulsed-laser ablation (Ref 186, 187), reactive magnetron sputtering (Ref 160), CVD (references in Ref 177), PECVD (Ref 188), or for related topics such as process engineering (Ref 189), fluidized beds (Ref 190), and plasma treatments (Ref 191). The examples discussed in this article focus on deposition over microscopic features; it is one of the most typical multiscale problems and has already generated many interesting multiscale simulation studies. As discussed in Ref 192, several algorithms for flux distribution and surface representation have been used for simulating deposition in micron or submicron features in 3-D:

- MC methods, visible solid angle, and ballistic transport and reaction models for the flux distribution algorithm
- Facet motion model, building-block model, equi-volume rate model, cellular model and level-set method

The earlier section "Transport on the Feature Scale" discussed the simulation of deposition inside submicron features based on ballistic or DSMC models to account for rarefied transport conditions. This section extends the discussion and examples to simulations based on hybrid models and to fully integrated simulations capable of bridging the reactor scale, the feature scale, as well as the nanostructure scale.

PVD on Micron- or Submicron-Sized Features. A sputtering process is referred to as ionized PVD (IPVD) when the deposition flux consists of more ions than neutrals. It is particularly used to deposit metal layers and diffusion barriers into micron trenches or vias of high aspect ratio. Several papers on multiscale modeling of IPVD inside features have already been published (Ref 182–184).

Hansen et al. (Ref 182) used a multiscale approach for simulating aluminum IPVD on microscopic features, which includes a detailed treatment of phenomena inside the feature but no plasma model. The phase-space distribution function of atoms and ions sputtered from the target is simply approximated by superposing a fraction of Thompson distribution (for ions)

and a fraction of Maxwellian distribution (for atoms). The MD simulations using appropriate potentials for aluminum-aluminum and argon-aluminum interactions are performed for a large number of trajectories to obtain good statistics for processes occurring at the feature surface (adsorption, reflection, surface sputtering). To obtain the net deposition rate at various locations on the surface, a first MD simulation based only on incoming fluxes from the target is performed, and the distributions of re-emitted particles at different locations are derived. Iterations are then performed to obtain the contribution of all the particles re-emitted from the various surface points to a particular location. Green functions are used to ensure self-consistency of simulations. Finally, a level-set method is used to move the gas-solid interface and simulate trench filling. Simulations have confirmed the superiority of IPVD compared to PVD for obtaining more conformal coverage on micron features.

One of the most interesting examples was presented by Arunachalam et al. (Ref 184), who developed a fully integrated multiscale model for IPVD of copper in submicron trenches. A 2-D hybrid model of the plasma region is used to compute the flux and energy of species incoming in the feature zone, which are passed to a 3-D MC model to simulate the deposition inside the feature. Cu, Cu*, and Cu+, are the copper species considered in addition to Ar, Ar*, and Ar+. The sputter yields are derived from this model as well as the fluxes of species. Ion energy and angular distributions are obtained after postprocessing through a plasma sheath model. The MC model accounts for adsorption, sputtering, energy loss, reflection, and surface diffusion; the kinetic parameters for all of these surface processes except the latter are obtained from MD simulations. The influence of the process parameters (substrate bias, coil and target powers, and nature of buffer gas) has been carefully examined and provides considerable insight into the complex deposition processes. For instance, it has been shown that increasing the substrate bias causes an increase in the mean ion energy and the amount of sputtering inside the feature, which finally results in both transferring material from the bottom of the trench to the sidewalls and removing deposited material from the upper corners of the trench. The analysis described previously for the simulation of IPVD over topography is rather similar to that represented in Fig. 19(a) for TiN IPVD (Ref 185). A global plasma model is used to compute the densities of species and sputtering yield (Ref 193), then a second sheath model allows the calculation of the angular and energy distribution functions of incoming ions inside the features, which are used as input data to a third feature model that simulates film evolution inside the feature. Finally, a fourth MD-type model can be used to simulate particle-surface interactions. However, in the analysis presented in Ref 185, no MD simulation results were reported, and a simple global plasma

model was used, whereas much more detailed hybrid PIC-MCC 2-D models would have been preferred, such as in Ref 160 or 184. Figure 19 (d) shows an example of a MD trench-filling simulation in the case of copper IPVD (Ref 194).

CVD on Microscopic Features. In the hierarchical top-down multiscale approach depicted in Fig. 19(a), information is simply transferred sequentially from the reactor scale to the feature and nanostructure scales with no feedback to upper scales. However, when the substrate surface is not flat at the microscopic scale (e.g., when the surface is rough or when the density of submicron features is high), microscopic effects that cannot be captured by macroscopic models may drastically affect the reactor behavior; hence, self-consistent approaches are needed for performing accurate simulations in such cases.

In a pioneer analysis of tungsten CVD from WF_6, Rodgers and Jensen (Ref 195) introduced the concept of effective reactivity function (ε) for coupling a continuum CVD reactor model to a feature-scale ballistic MC model. The $\varepsilon(x, y, t)$ integrates the effects of surface microscopic heterogeneity and microscale transport resistance. The MC simulations are performed first; particles are launched from a source point (S) at a sufficient height above a particular feature and undergo a sequence of events, such as multiple reflections at feature walls, reaction, etc., that are recorded over a large number of trajectories to obtain a mean value of the effective sticking coefficient, s', which satisfies, for species i:

$$s'_i = \varepsilon_i(S, t)s_i \qquad (Eq\ 151)$$

where s_i is the real sticking coefficient of species i (i.e., on a flat surface). The process is repeated for a large number of source points to obtain $\varepsilon_i(t)$ for the selected feature, then these individual reactivity functions are linearly superimposed to obtain the effective reactivity map of the whole surface for species i, which is used to formulate the surface-boundary condition for the macroscopic reactor model. At each time step, both models are sequentially iterated until self-consistent results are obtained; hence, long calculation times are required for a complete transient simulation of deposition. The concentration profiles calculated using this approach differ markedly from those obtained when neglecting the effects of microscopic heterogeneity.

Another multiscale approach was used by Merchant et al. (Ref 196) for simulating silicon dioxide deposition from tetraethoxysilane. They coupled together a transient continuum reactor model that solves the coupled transport equations at the macroscopic scale by using an FE method, a deterministic ballistic transport and reaction model at the feature scale, and a continuum model at the mesoscale (scale of a small patterned area consisting of several features), which is used to transfer information between
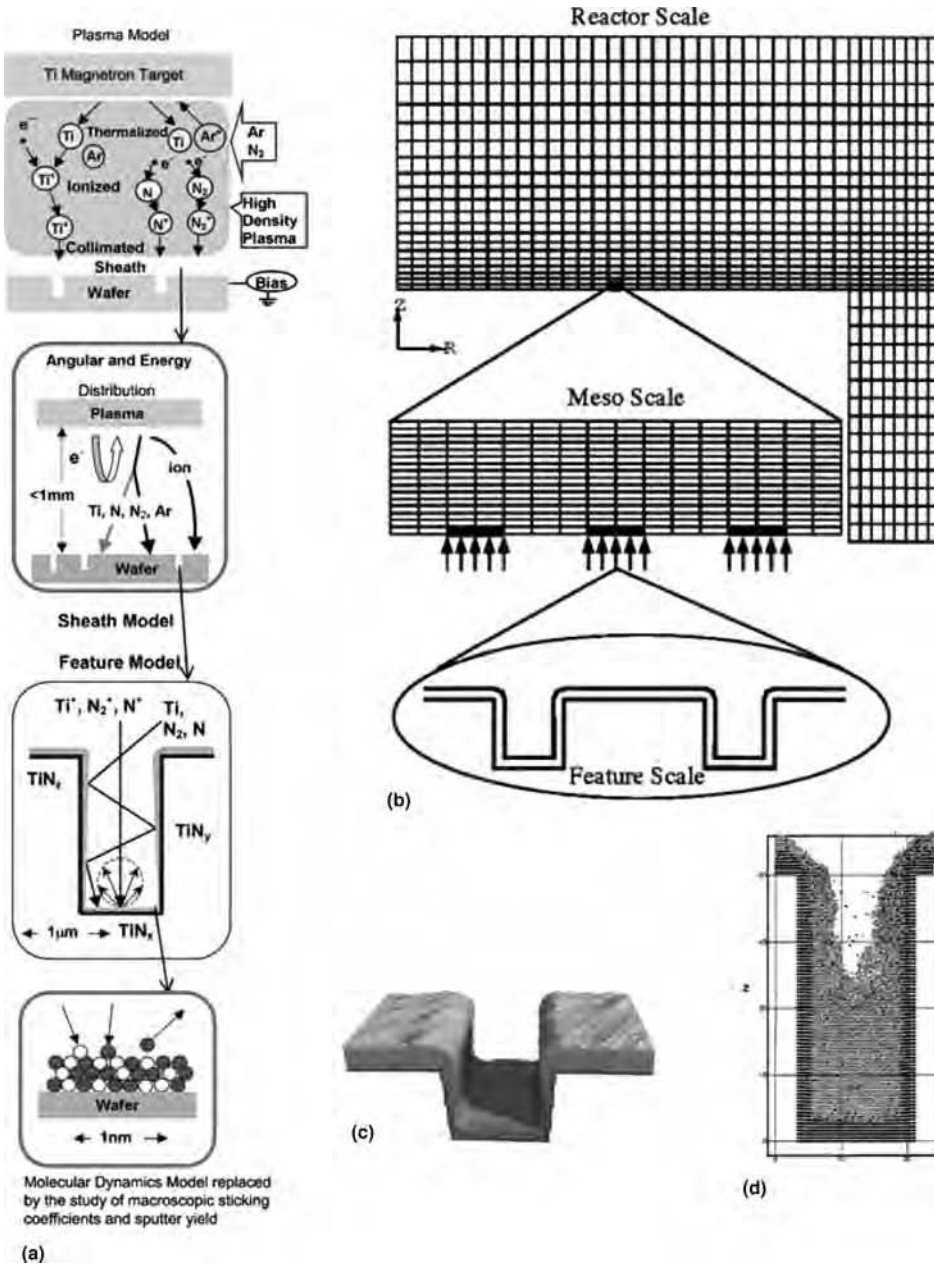
Fig. 19 Procedures and models used for the multiscale simulation of chemical vapor deposition (CVD) and physical vapor deposition (PVD) over topography. (a) Sequence of models proposed for simulation of TiN ionized PVD (IPVD) over topography. Source: Ref 185. (b) Multilength approach proposed for SiO$_2$ CVD, combining a macroscale finite-element model (FEM) and a ballistic transport and reaction model at the feature scale. A third FEM mesoscale model is used to link both scales. Source: Ref 186. (c) Monte Carlo simulation of sputtered-aluminum deposition on a 0.025 mm trench. Source: Ref 159. (d) Molecular dynamics (MD) simulation of trench filling in copper IPVD. Source: Ref 194

feature-scale models (Ref 195) but cannot capture microstructural details. Gobbert et al. (Ref 197) have extended the capability of the multiscale model described previously by replacing their deterministic model of ballistic transport at the feature scale by an FE-based model that solves the transient Boltzmann equations inside the feature by using a discontinuous Galerkin method. The capability of this transport model was further extended to a wide range of transport regimes (from collision-less to transition regime flows) and to multiple species in Ref 198 and 199.

Conclusions and Outlook

Early studies and optimization of VPP were mainly conducted by experimental empirical trial-and-error procedures, but the tremendous increase in computing power at the beginning of the 1980s allowed the development of more and more sophisticated simulation models, which were briefly introduced and discussed in this article. These models provided considerable insight into the phenomena involved during deposition and etching from various VPP. The contribution of numerical modeling and simulation was particularly important to elucidate the complex coupled-transport phenomena involved in CVD reactors. In the 1990s, more and more complex situations were considered; full 3-D simulations of coupled transport phenomena in CVD processes, including unsteady flows and complex chemistries, were performed, and complex PVD processes were also simulated. Commercial CFD software are now available, which allows the simulation of multiphysics problems in complex geometries as well as process optimization.

Nevertheless, there are still several important limitations to VPP modeling. First, physical data (which cannot always been measured) are still lacking for many systems of interest in vapor deposition or etching. This is particularly true for thermochemical data, gas-phase, and surface-reaction kinetics in CVD or electron-molecule collision cross sections in ionized systems (PVD, PECVD). Second, although highly desirable, the development of single simulation models capable of predicting all the features of VPP, from very small to very large length and time scales, clearly remains an elusive goal. Over the last decade, the ever-increasing power of computers allowed a strong development of first-principles calculations, which are used more and more for providing fundamental data to VPP simulators. However, despite their growing interest and increasing performance, accurate simulation models based on quantum mechanics cannot be used for simulating real-size vapor-solid systems that include a huge number of particles. Alternatively, ab initio calculations can be integrated into multiscale models.

Multiscale simulation has now become a central problem in materials science and engineering as well as many other fields of modern

the macroscopic reactor model and the feature-scale model (Fig. 19b). A top-down procedure is first used to pass information from the macro- to the microscale. An initial guess of species concentration over a given grid element of the reactor model is interpolated onto the mesoscale grid, then the guess concentrations are transferred to the feature scale at each node of the mesoscale grid at the surface boundary. The feature model and the mesoscale models are iteratively solved to ensure self-consistency,

and the net fluxes of species at each node of the mesoscale model are fed back into the reactor model. Information from nonpatterned areas (Fig. 19b) is also fed back into the reactor model. Both FE-based models (macro and meso) are iteratively solved until convergence is obtained, then the time is incremented and the procedure is repeated. Six gas-phase species were considered in these simulations, which require less computation time than simulations based on effective reactivity functions and MC

physics and chemistry. Some of the multiscale VPP models developed to date have astonishing predicting capabilities, but undoubtedly, even more sophisticated multiscale models will be developed in the future that will become invaluable tools for the study and optimization of VPP.

REFERENCES

1. J.M. Blocher, *Vapor Deposition,* C.F. Powell, J.H. Oxley, and J.M. Blocher, Ed., John Wiley & Sons, New York, 1967

2. D.T.J. Hurle, *Handbook of Crystal Growth,* Elsevier, Amsterdam, 1994

3. K.L. Choy, *Prog. Mater. Sci.,* Vol 48, 2003, p 57–170

4. J.S. Colligon, Physical Vapor Deposition, *Non-Equilibrium Processing of Materials,* Pergamon Materials Series, R.W. Cahn, Ed., Elsevier, Oxford, U.K., 1999, p 225–253

5. F. Teyssandier and A. Dollet, Chemical Vapor Deposition, *Non-Equilibrium Processing of Materials,* Pergamon Materials Series, R.W. Cahn, Ed., Elsevier, Oxford, U.K., 1999 p 255–285

6. A.C. Jones and M.L. Hitchman, *Chemical Vapour Deposition: Precursors, Processes and Applications,* The Royal Society of Chemistry, Thomas Graham House, Cambridge, U.K., 2009

7. G.A. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows,* Oxford Science Publications, 1994

8. J. Hirschfelder, C. Curtiss, and R.B. Bird, *Molecular Theory of Gases and Liquids,* John Wiley & Sons, 1954

9. H. Versteeg and W. Malalasekera, *An Introduction to Computational Fluid Dynamics: The Finite Volume Method,* 2nd ed., Pearson, 2007

10. R.B. Bird, W.E. Stewart, and E.N. Lightfoot, *Transport Phenomena,* Wiley Int. Ed., New York, 1960

11. P.D. Neufield, A.R. Janzen, and A. Aziz, *J. Chem. Phys.,* Vol 57, 1972, p 1100–1102

12. W.L. Holstein, *J. Electrochem. Soc.,* Vol 135 (No. 7), 1988, p 1788–1793

13. R.J. Kee, F.M. Rupley, J.A. Miller, M.E. Coltrin, F.F. Grcar, E. Meeks, H.K. Moffat, A.E. Lutz, G. Dixon-Lewis, M.D. Smooke, J. Warnatz, G.H. Evans, R.S. Larson, R.E. Mitchell, L.R. Petzold, W.C. Reynolds, M. Caracotsios, W.E. Stewart, P. Glarborg, C. Wang, O. Adigun, W.G. Houf, C.P. Chou, S.F. Miller, P. Ho, and D.J. Young, *Chemkin Release 4.0,* Reaction Design Inc., San Diego, CA, 2004, www.reactiondesign.com/

14. NIST Reference Fluid Thermodynamic and Transport Properties Database (REFPROP): Version 8.0, http://www.nist.gov/

15. B. Poling, J. Prausnitz, and J. O'Connell, *The Properties of Gases and Liquids,* 5th ed., McGraw Hill, 2007

16. W. Kossel, *Nachr. Ges. Wiss. Göttingen,* math-phys. klasse, 1927, p 135–143

17. A.C. Levi and M. Kotrla, *J. Phys.: Condens. Matter,* Vol 9, 1997, p 299–344

18. Materials Studio, http://accelrys.com/products/materials-studio/

19. K.J. Laidler, *Chemical Kinetics,* 3rd ed., Harper Collins Publishers, New York, 1987

20. K.F. Jensen and D.B. Graves, *J. Electrochem. Soc.,* Vol 130 (No. 9), 1983, p 1950–1957

21. T.E. Wilke, K.A. Turner, and C.J. Takoudis, *Chem. Eng. Sci.,* Vol 41, 1986, p 643–650

22. M.E. Coltrin, R.J. Kee, and G.H. Evans, *J. Electrochem. Soc.,* Vol 136 (No. 3), 1989, p 819–829

23. J.F. Ziegler et al., *The Stopping and Range of Ions in Matter,* Vol 2–6, Pergamon Press, New York, 1977–1985

24. F. Langlais, F. Hottier, and R. Cadoret, *J. Cryst. Growth,* Vol 56, 1982, p 659–672

25. L. Vandenbulcke and G. Vuillard, *Proc. Sixth Int. Conf. on CVD,* L.F. Donagey, P. Rai-Choudhury, and R.N. Tauber, Ed., The Electrochemical Society, Pennington, 1977, p 318–329

26. F.C. Franck, *Disc. Faraday Soc.,* Vol 5, 1949, p 48–54

27. F.C. Franck and J.H. Van der Merwe, *Proc. R. Soc. Lond. A,* Vol 198, 1949, p 205–216

28. F.C. Franck and J.H. Van der Merwe, *Proc. R. Soc. Lond. A,* Vol 200, 1949, p 125–134

29. M. Volmer and A. Weber, *Z. Physik. Chem.,* Vol 119, 1927, p 277–301

30. I.N. Stranski and L. Krastanov, *Ber. Akad. Wiss. Wien.,* Vol 146, 1938, p 797–801

31. Y. Kajikawa and S. Noda, *Appl. Surf. Sci.,* Vol 245, 2005, p 281–289

32. A. van der Drift, *Philips Res. Reports,* Vol 22, 1967, p 267–288

33. R. Stöckel, M. Stammler, K. Janischowsky, L. Ley, M. Albrecht, and H.P. Strunk, *J. Appl. Phys.,* Vol 83, 1998, p 531–539

34. F. Teyssandier, S. De Persis, and A. Dollet, *J. Chem. Edu.,* Vol 81 (No. 6), 2004, p 832–833

35. J. Perrin, O. Leroy, and M.C. Bordage, *Contrib. Plasma Phys.,* Vol 36 (No. 1), 1996, p 3–49

36. C.G.A. Perkins, E.R. Austin, and F.W. Lampe, *J. Am. Chem. Soc.,* Vol 101, 1979, p 1109–1115

37. S.W. Benson, *Thermochemical Kinetics,* 2nd ed., Wiley-Interscience, New York, London, Sydney, Toronto, 1976

38. F.A. Lindemann, *Trans. Faraday Soc.,* Vol 17, 1922, p 598–599

39. R.G. Gilbert and S.C. Smith, *Theory of Unimolecular and Recombination Reactions,* Blackwell Scientific Publications, Oxford, London, Edinburgh, Boston, Melbourne, 1990

40. A. Dollet, S. De Persis, and F. Teyssandier, *Phys. Chem. Chem. Phys.,* Vol 6, 2004, p 1203–1212

41. A. Dollet and S. de Persis, *J. Anal. Appl. Pyrolysis,* Vol 80, 2007, p 460–470

42. C.N. Hinshelwood, *Proc. R. Soc. A,* Vol 113, 1927, p 230–233

43. O.K. Rice and J. Ramsperger, *J. Amer. Chem. Soc.,* Vol 49 (No. 7), 1927, p 1617–1629

44. L.S. Kassel, *J. Phys. Chem.,* Vol 32, 1928, p 225–242

45. L.S. Kassel, *J. Phys. Chem.,* Vol 32, 1928, p 1065–1079

46. R.A. Marcus and O.K. Rice, *J. Phys. Colloid Chem.,* Vol 55, 1951, p 894–908

47. R.A. Marcus, *J. Chem. Phys.,* Vol 20, 1952, p 359–364

48. A.M Dean, *J. Phys. Chem.,* Vol 89 (No. 21), 1985, p 4600–4608

49. D.R. Burgess, M. Zachariah, W. Tsang, and P.R. Westmoreland, *Prog. Energy. Combust. Sci.,* Vol 21, 1996, p 453–529

50. F. Teyssandier and M.D. Allendorf, *J. Electrochem. Soc.,* Vol 145 (No. 6), 1998, p 2167–2178

51. J. Troe, *Ber. Bunsen-Ges. Phys. Chem.,* Vol 87, 1983, p 161–169

52. C.W. Larson, R. Patrick, and D.M. Golden, *Combust. Flame,* Vol 58, 1984, p 229–237

53. R.J. Kee, F.M. Rupley, and J.A. Miller, SAND89-8009B.UC-706, 1989

54. I. Oref, *J. Phys. Chem.,* Vol 93, 1989, p 3465–3469

55. G. Katzer, M.C. Ernst, A.F. Sax, and J. Kalcher, *J. Phys. Chem. A,* Vol 101, 1997, p 3942–3958

56. M.T. Swihart and S.L. Girshick, *J. Phys. Chem. B,* Vol 103, 1999, p 64–76

57. C.F. Melius and P. Ho, *J. Phys. Chem.,* Vol 95, 1991, p 1410–1419

58. M.D. Allendorf and C.F. Melius, *J. Phys. Chem.,* Vol 96, 1992, p 428–437

59. M.R. Zachariah and W. Tsang, *J. Phys. Chem.,* Vol 99, 1995, p 5308–5318

60. M.W. Chase, *NIST-JANAF Thermochemical Tables,* 4th ed., Monograph No. 9 (Part I and Part II), American Institute of Physics, Melville, NY

61. NIST Chemical Kinetic Database, http://kinetics.nist.gov/kinetics/index.jsp

62. N.M. Marinov, W.J. Pitz, C.K. Westbrook, M.J. Castaldi, and S.M. Senkan, *Combust. Sci. Technol.,* Vol 116, 1996, p 211–287

63. N.M. Marinov, M.J. Castaldi, C.R. Melius, and W. Tsang, *Combust. Sci. Technol.,* Vol 128, 1997, p 295–342

64. J.M. Simmie, *Prog. Energy Combust. Sci.,* Vol 29, 2003, p 599–634

65. P. Ho, M.E. Coltrin, and W.G. Breiland, *J. Phys. Chem.,* Vol 98 (No. 40), 1994, p 10138–10147

66. K. Noringa and O. Deutschmann, *Ind. Eng. Chem. Res.,* Vol 46, 2007, p 3547–3557

67. A.J. Tomlin, T. Turanyi, and M.J. Pilling, Mathematical Tools for the Construction, Investigation and Reduction of

Combustion Mechanisms, *Comprehensive Chemical Kinetics,* Vol 35, *Low Temperature Combustion and Autoignition,* M.J. Piling, Ed., Elsevier Science, Amsterdam, The Netherlands, 1997, p 293–437

68. D. Edelson and D.L. Allara, *Int. J. Chem. Kinet.,* Vol 12, 1980, p 605–621

69. J. Warnatz, U. Maas, and R.W. Dibble, *Combustion: Physical and Chemical Fundamentals, Modeling and Simulation, Experiments, Pollutant Formation,* 3rd ed., Springer-Verlag, New York, NY, 2001

70. S. de Persis, F. Teyssandier, A.H. Mc Daniel, and M.D. Allendorf, *Adv. Mater.,* CVD 8 (No. 2), 2001, p 63–73

71. S. de Persis, A. Dollet, and F. Teyssandier, *J. Anal. Appl. Pyrol.,* Vol 70, 2003, p 55–71

72. T. Turanyi, KINALC, http://www.chem.leeds.ac.uk/Combustion/Combustion.html

73. S.H. Lam and D.A. Goussis, *Int. J. Chem. Kinet.,* Vol 26, 1994, p 461–486

74. U. Maas and S.B. Pope, *Combust. Flame,* Vol 88, 1992, p 239–264

75. S. de Persis, F. Teyssandier, D. Thévenin, and N. Darabiha, *J. Electrochem. Soc.,* Vol 151 (No. 4), 2004, p 236–244

76. T. Turanyi, *New J. Chem.,* Vol 14, 1990, p 795–803

77. M.D. Allendorf, *J. Electrochem. Soc.,* Vol 140 (No. 3), 1993, p 747–753

78. M. Pons, C. Bernard, E. Blanquet, and R. Madar, *Thin Solid Films,* Vol 365, 2000, p 264–274

79. THERMOCALC, http://www.thermocalc.com/

80. GEMINI, http://gmini.org/

81. FACTSAGE, http://www.factsage.com

82. B. Hallstedt and Z.-K. Liu, *Calphad: Computer Coupling of Phase Diagrams and Thermochemistry,* Vol 33, 2009, p 265

83. P. Shi, A. Engström, L. Höglund, Q. Chen, B. Sundman, J. Agren, and M. Hillert, *J. Iron Steel Res. Int.,* Vol 14 (No. 5), Suppl. 1, 2007, p 210–215

84. FLUENT, http://www.fluent.com/

85. PHOENICS-CVD, http://www.cham.co.uk/

86. CFD ACE, http://www.esi-group.com/products/Fluid-Dynamics/cfd-ace

87. CFX, http://www.ansys.com/products/fluid-dynamics/cfx/

88. MPSALSA, http://www.cs.sandia.gov/CRF/MPSalsa/

89. COMSOL Multiphysics, http://www.comsol.com/products/multiphysics/

90. J.R. Howell, *A Catalog of Radiation Configuration Factors,* McGraw-Hill, New York, 1982

91. F. Loumagne, F. Langlais, and R. Naslain, *J. Cryst. Growth,* Vol 155, 1995, p 198–204

92. H. Chehouani, B. Armas, S. Benet, and S. Brunet, *J. Phys., Coll.,* C5 (Suppl.) (No. 5), 1989, p 47–56

93. A. Dollet, Y. Casaux, G. Chaix, and C. Dupuy, *Thin Solid Films,* Vol 406 (No. 1–2), 2002, p 1–16

94. S.V. Patankar, Numerical Heat Transfer and Fluid Flow, *Computational Methods in Mechanics and Thermal Sciences,* Routledge, Taylor & Francis Group, New York, 1980

95. Y.B. Wang, F. Teyssandier, J. Simon, and R. Feurer, *J. Electrochem. Soc.,* Vol 141, 1994, p 824–842

96. H. Van Santen, C.R. Kleijn, and H.E.A. Van den Akker, *Int. J. Heat Mass Transf.,* Vol 43, 2000, p 1523–1535

97. J. Ouazzani and F. Rosenberger, *J. Cryst. Growth,* Vol 100, 1990, p 545–576

98. H. Moffat and K.F. Jensen, *J. Cryst. Growth,* Vol 77, 1986, p 108–119

99. W.L. Holstein, J.L. Fitzjohn, E.J. Fahi, P.W. Gilmour, and E.R. Schmelzer, *J. Cryst. Growth,* Vol 94, 1989, p 131–144

100. D.I. Fotiadis, S. Kieada, and K.F. Jensen, *J. Cryst. Growth,* Vol 102, 1990, p 441–470

101. A. Li, K. Norinaga, W. Zhang, and O. Deutschmann, *Compos. Sci. Technol.,* Vol 68, 2008, p 1097–1104

102. S.C. Chapra and R.P. Canale, *Numerical Methods for Engineers,* 5th ed., McGraw-Hill Int. Ed., 2006

103. C.R. Kleijn, *Thin Solid Films,* Vol 365, 2000, p 294–306

104. D. Raabe, *Computational Materials Science,* Wiley-Vch, Weinheim, Germany, 1998

105. K. Ohno, K. Esfarjani, and Y. Kawazoe, *Computational Materials Science,* Springer Series in Solid-State Sciences, Springer-Verlag, Berlin, Germany, 1999

106. "Molecular Modelling," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/wiki/Molecular_modelling

107. I. Zahi, H. Vergnes, B. Caussat, A. Estève, M. Djafari Rouhani, P. Mur, P. Blaise, and E. Scheid, *Surf. Coat. Technol.,* Vol 201, 2007, p 8854–8858

108. GAUSSIAN Program, Gaussian Inc., http://www.gaussian.com/

109. MOPAC Program, Stewart J.J.P., *J. Comput.-Aided Molec. Des.,* Vol 4, 1990, p 1–105

110. M. Frenklach, *Pure Appl. Chem.,* Vol 70 (No. 2), 1998, p 417–484

111. N.L. Allinger and U. Burkert, *Molecular Mechanics*, American Chemical Society Publication, 1982

112. "Force Field (Chemistry)," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/wiki/Force_field_(chemistry)

113. G. Wahl, *Thin Solid Films,* Vol 40, 1977, p 13–26

114. K.F. Jensen, Transport Phenomena in Vapor Phase Epitaxy Reactors, *Handbook of Crystal Growth,* Vol 3b, D.T.J. Hurle, Ed., Elsevier, Amsterdam, 1994, p 543–599

115. W.L. Holstein, *Prog. Cryst. Growth Charact.,* Vol 24, 1992, p 111–211

116. C. Vinante, P. Duverneuil, and J.P. Couderc, *J. Phys., Coll.,* C5 (Suppl.) (No. 5), 1989, p 35–43

117. D.I. Fotiadis, A.M. Kremer, D.R. McKenna, and K.F. Jensen, *J. Cryst. Growth,* Vol 85, 1987, p 154–164

118. M. de Keijser, C. Opdorp, and C. Weber, *J. Cryst. Growth,* Vol 92, 1988, p 33

119. H. Van Santen, C.R. Kleijn, and H.E.A. Van den Akker, *J. Cryst. Growth,* Vol 212, 2000, p 311–323

120. H. Van Santen, C.R. Kleijn, and H.E.A. Van den Akker, *Int. J. Heat Mass Transf.,* Vol 44, 2001, p 659–672

121. H. Van Santen, C.R. Kleijn, and H.E.A. Van den Akker, *J. Cryst. Growth,* Vol 212, 2000, p 299–310

122. L.J. Giling, *J. Electrochem. Soc.,* Vol 129, 1982, p 634

123. K.F. Jensen, D.I. Fotiadis, and T.J. Mountziaris, *J. Cryst. Growth,* Vol 107, 1991, p 1–11

124. H. Moffat and K.F. Jensen, *J. Cryst. Growth,* Vol 77, 1986, p 108–119

125. M.E. Coltrin, R.J. Kee, and R.J. Miller, *J. Electrochem. Soc.,* Vol 131 (No. 2), 1984, p 425–434

126. M.E. Coltrin, R.J. Kee, and J.A. Miller, *J. Electrochem. Soc.,* Vol 133 (No. 6), 1986, p 1206–1213

127. M. Tirtowidjojo and R. Pollard, *J. Cryst. Growth,* Vol 93, 1988, p 108–114

128. A. Li, K. Norinaga, W. Zhang, and O. Deutschmann, *Compos. Sci. Technol.,* Vol 68, 2008, p 1097–1104

129. T.H. Osterheld, M.D. Allendorf, and C.F. Melius, *J. Phys. Chem.,* Vol 98, 1994, p 6995–7003

130. S.M. Suh, M.R. Zachariah, and S.L. Girshick, *J. Vac. Sci. Technol. A,* Vol 19 (No. 3), 2001, p 940–951

131. S.M. Suh, M.R. Zachariah, and S.L. Girshick, *Aerosol Sci.,* 2002, p 943–959

132. S. Van Veldhuizen, C. Vuik, and C.R. Kleijn, *Surf. Coat. Technol.,* Vol 201, 2007, p 8859–8862

133. S. de Persis, A. Dollet, and F. Teyssandier, *Proc. 16th Int. Conf. CVD,* April 27–May 2, 2003 (Paris), M. Allendorf, F. Maury, and F. Teyssandier, Ed., The Electrochemical Society, Pennington, NJ, PV2003-8, 2003, p 163–170

134. A. Dollet, S. de Persis, M. Pons, and M. Matecki, *Surf. Coat. Technol.,* Vol 177–178, 2004, p 382–388

135. K.D. Annen, C.D. Stinespring, M.A. Kuczmarski, and J.A. Powell, *J. Vac. Sci. Technol., A,* Vol 8, 1990, p 2970–2975

136. K.J. Kuijlars, C.R. Kleijn, and H.E.A. van den Akker, *Thin Solid Films,* Vol 270, 1995, p 456–461

137. R. Arora and R. Pollard, *J. Electrochem. Soc.,* Vol 138, 1991, p 1523–1537

138. H. Wolf, J. Röber, S. Riedel, R. Streiter, and T. Gessner, *Microelec. Eng.,* Vol 45, 1999, p 15–27

139. K.J. Kuijlars, C.R. Kleijn, and H.E.A. van den Akker, *Chem. Eng. J.,* Vol 57, 1995, p 127–136

140. D.G. Coronell and K.F. Jensen, *J. Electro-chem. Soc.,* Vol 139 (No. 8), 1992, p 2264–2272

141. D.G. Coronell and K.F. Jensen, *J. Comp.-Aided Mater. Des.,* Vol 1, 1993, p 3–26

142. R. Sinkovits and C.R. De Vore, *J. Appl. Phys.,* Vol 80 (No. 11), 1996, p 6474–6488

143. M. Ikegawa and J. Kobayashi, *J. Electro-chem. Soc.,* Vol 136 (No. 10), 1989, p 2982–2986

144. D.G. Coronell and K.F. Jensen, *J. Electro-chem. Soc.,* Vol 141 (No. 9), 1992, p 2545–2551

145. T.S. Cale, B.R. Rogers, T.P. Merchant, and L.J. Borucki, *Comp. Mater. Sci.,* Vol 12, 1998, p 333–353

146. S. Berg, H.-O. Blom, T. Larsson, and C. Nender, *J. Vac. Sci. Technol.,* Vol A5, 1987, p 202–207

147. S. Berg and T. Nyberg, *Thin Solid Films,* Vol 476, 2005, p 215–230

148. T. Kobayashi, *Vacuum,* Vol 74, 2004, p 379–385

149. W. Möller, W. Eckstein, and J.B. Biersack, *Comp. Phys. Comm.,* Vol 51, 1988, p 355

150. Z.Y. Chen, A. Bogaerts, D. Depla, and V. Ignatova, *Nucl. Instr. Meth. Phys. Res. B,* Vol 207, 2003, p 415–423

151. D. Rosen, I. Katardjiev, S. Berg, and W. Möller, *Nucl. Instr. Meth. Phys. Res. B,* Vol 228, 2005, p 193–197

152. S. Kadlec, *Surf. Coat. Technol.,* Vol 202, 2007, p 895–903

153. F. Clénet, P. Briaud, G. Lempèrière, and G. Turban, *Thin Solid Films,* Vol 336, 1999, p 290–298

154. J.P. Biersack and L. Haggmark, *Nucl. Instr. Meth.,* Vol 174 (No. 1–2), 1980, p 257–269

155. J.F. Ziegler, "Particle Interactions with Matter: Stopping and Range of Ions in Matter," http://www.srim.org/

156. C.H. Shon and J.K. Lee, *Appl. Surf. Sci.,* Vol 192, 2002, p 258–269

157. S. Yip, *Handbook of Materials Modeling,* Springer, 2005

158. Materials Research by Means of Multi-scale Computer Simulation, *MRS Bull.,* Vol 26 (No. 3), 2001

159. G.H. Gilmer, H. Huang, T. Diaz de la Rubia, J. Dalla Torre, and F. Baumann, *Thin Solid Films,* Vol 365, 2000, p 189–200

160. U.H. Kwon, S.H. Choi, Y.H. Park, and W. J. Lee, *Thin Solid Films,* Vol 475, 2005, p 17– 23

161. P. L. O'sullivan, F.H. Baumann, and G.H. Gilmer, *J. Appl. Phys.,* Vol 88 (No. 7), 2000, p 4061–4068

162. D. Henderson, M.H. Bodsky, and P. Chaudhari, *Appl. Phys. Lett.,* Vol 25 (No. 11), 1974, p 641–643

163. M.J. Brett, *J. Mater. Sci. Lett.,* Vol 8, 1989, p 415–417

164. R.N. Tait, T. Smy, and M.J. Brett, *Thin Solid Films,* Vol 187, 1990, p 375–384

165. H.J. Leamy, G.H. Gilmer, and A.G. Dirks, *Current Topics in Materials Science,* Vol 6, E. Kaldis, Ed., North-Holland, 1980, p 309–344

166. T.S. Cale and G.B. Raupp, *J. Vac. Sci. Technol. B,* Vol 8, 1990, p 1242–1248

167. M.M. IslamRaja, M.A. Capelli, J.P. McVittie, and K.C. Saraswat, *J. Appl. Phys.,* Vol 70, 1991, p 7137–7140

168. S. Osher and J.A. Sethian, *J. Comput. Phys.,* Vol 79, 1988, p 12–49

169. C. Cavallotti, M. Di Stanislao, and S. Carra, *Prog. Cryst. Growth Charact. Mater.,* Vol 48–49, 2004, p 123–165

170. H.C. Kim, F. Iza, S.S. Yang, M. Radmilo-vic-Radjenovic, and J.K. Lee, *J. Phys. D: Appl. Phys.,* Vol 38, 2005, p R283–R301

171. A. Dollet, J.P. Couderc, and B. Despax, *Plasma Sources Sci. Technol.,* Vol 4, 1995, p 94–106

172. K. Hassouni, G. Lombardi, X. Duten, G. Haagelar, F. Silva, A. Gicquel, T.A. Grotjohn, M. Capitelli, and J. Ropcke, *Plasma Sources Sci. Technol.,* Vol 15, 2006, p 117–125

173. C. Vahlas, B. Caussat, P. Serp, and G.N. Angelopoulos, *Mater. Sci. Eng. R,* Vol 53, 2006, p 1–72

174. L. Cadoret, N. Reuge, S. Pannala, M. Syam-lal, C. Coufort, and B. Caussat, *Surf. Coat. Technol.,* Vol 201, 2007, p 8919–8923

175. N. Reuge, L. Cadoret, and B. Caussat, *Chem. Eng. J.,* Vol 148, 2009, p 506–516

176. MFIX, www.mfix.org

177. A. Dollet, *Surf. Coat. Technol.,* Vol 177–178, 2004, p 245–251

178. K.F. Jensen, S.T. Rodgers, and R. Venka-taramani, *Curr. Op. Solid State Mater. Sci.,* Vol 3, 1998, p 562–569

179. H.N.G. Wadley, X. Zhou, R.A. Johnson, and N. Neurock, *Prog. Mater. Sci.,* Vol 46, 2001, p 329–377

180. C. Cavallotti, M. Di Stanislao, D. Mosca-telli, and A.Veneroni, *Electrochim. Acta,* Vol 50, 2005, p 4566–4575

181. C.R. Kleijn, R. Dorsman, K.J. Kuijlaars, M. Okkerse, and H. van Santen, *J. Cryst. Growth,* Vol 303, 2007, p 362–380

182. U. Hansen, S. Rodgers, and K.F. Jensen, *Phys. Rev. B,* Vol 62 (No. 4), 2000, p 2869–2878

183. P. Vogl, U. Hansen, and V. Fiorentini, *Comp. Mater. Sci.,* Vol 24, 2002, p 58–65

184. V. Arunachalam, S. Rauf, D.G. Coronell, and P.L.G. Ventzek, *J. Appl. Phys.,* Vol 90 (No. 1), 2001, p 64–73

185. D. Mao and J. Hopwood, *J. Appl. Phys.,* Vol 96 (No. 1), 2004, p 820–828

186. J.N. Leboeuf, K.R. Chen, J.M. Donato, D.B. Geohegan, C.L. Liu, A.A. Puretzky, and R.F. Wood, *Appl. Surf. Sci.,* Vol 96–98, 1996, p 14–23

187. M.I. Zeifman, B.J. Garrison, and L.V. Zhigilei, *Appl. Surf. Sci.,* Vol 197–198, 2002, p 27–34

188. M. Rondanini, S. Cereda, F. Montalenti, L. Miglio, and C. Cavallotti, *Surf. Coat. Technol.,* Vol 201, 2007, p 8863–8867

189. D.G. Vlachos, *Adv. Chem. Eng.,* Vol 30, 2005, p 1–61

190. W. Wang, B. Lu, N. Zhang, Z. Shi, and J. Li, *Int. J. Multiphase Flow,* 2009 (in press)

191. A.N. Bhoj and M.J. Kushner, *J. Phys. D: Appl. Phys.,* Vol 39 (No. 8), 2006, p 1594–1598

192. U. Kwon and W.-J. Lee, *Thin Solid Films,* Vol 445, 2003, p 80–89

193. K. Tao, D. Mao, and J. Hopwood, *J. Appl. Phys.,* Vol 91 (No. 7), 2003, p 4040–4048

194. C.-C. Hwang, G.-J. Huang, J.-G. Chang, and S.-P. Ju, *J. Appl. Phys.,* Vol 91 (No. 6), 2002, p 3569–3578

195. S.T. Rodgers and K.F. Jensen, *J. Appl. Phys.,* Vol 83, 1998, p 524–530

196. T.P. Merchant, M.K. Gobbert, T.S. Cale, and L.J. Borucki, *Thin Solid Films,* Vol 365, 2000, p 368–375

197. M.K. Gobbert, V. Prasad, and T.S. Cale, *Thin Solid Films,* Vol 410, 2002, p 129–141

198. M.K. Gobbert and T.S. Cale, *J. Comp. Phys.,* Vol 213, 2006, p 591–612

199. M.K. Gobbert and T.S. Cale, *Surf. Coat. Technol.,* Vol 201, 2007, p 8830–8837

Determination of Heat Transfer Coefficients for Thermal Modeling

D. Scott MacKenzie, Houghton International, Inc.
Andrew L. Banka, Airflow Sciences Corporation

HEAT TREATING AND QUENCHING are a complex business. The configuration of parts is endless, as is the types of furnaces available for heat treating. Numerous variables in the quenching process alone govern the ability of a part to meet distortion requirements. Heat treating is a constant balancing process. It is important to balance the ability of the material to achieve properties while at the same time control distortion. Because of the complexity of the heat treating process, it is rather difficult to understand the interaction of fluid flow and parts on part distortion and properties. Often, understanding is achieved only by experience, which comes from making mistakes and learning from those mistakes. There is less tolerance for "trial and error," and the emphasis is on "doing it right the first time." Unfortunately, there are few design rules that dictate the racking of a part in a given furnace. The application of computer modeling allows these mistakes to be made on the computer instead of on the manufacturing shop floor.

During the quenching process, when a hot component comes in contact with the liquid quenchant, there are normally three stages of quenching:

- Vapor stage (stage A or vapor blanket stage)
- Boiling stage (stage B or nucleate boiling stage)
- Convection stage (stage C)

An example showing the three stages of quenching is shown in Fig. 1.

The vapor stage is encountered when the hot surface of the heated component first comes in contact with the liquid quenchant. The component becomes surrounded with a blanket of vapor.

In this stage, heat transfer is very slow and occurs primarily by radiation through the vapor blanket. Some conduction also occurs through the vapor phase. This blanket is very stable, and its removal can only be enhanced by agitation or speed-improving additives. This stage is responsible for many of the surface soft spots

encountered in quenching. High-pressure sprays and strong agitation eliminate this stage. If the vapor phase is allowed to persist, undesirable microconstituents can form.

The second stage encountered in quenching is the boiling stage. This is where the vapor stage starts to collapse, and all liquid in contact with the component surface erupts into boiling bubbles. This is the fastest stage of quenching. The high heat extraction rates are due to carrying away heat from the hot surface and transferring it further into the liquid quenchant, which allows cooled liquid to replace it at the surface. In many quenchants, additives have been added to enhance the maximum cooling rates obtained by a given fluid. The boiling stage stops when the temperature of the component surface reaches a temperature below the boiling point of the liquid. For many distortion-prone components, high-boiling-temperature oils or liquid salts are used if the medium is fast enough to harden the steel, but both of these quenchants see relatively little use in induction hardening.

The final stage of quenching is the convection stage. This occurs when the component has reached a point below that of the quenchant boiling temperature. Heat is removed by

convection and is controlled by the quenchant specific heat and thermal conductivity and the temperature differential between the component temperature and that of the quenchant. The convection stage is usually the slowest of the three stages. Typically, it is this stage where most distortion occurs. Figure 2 shows the progression of the three phases of quenching during the immersion of a solid cylinder.

Obtaining specified properties and low distortion is usually a balancing act. Often, optimal properties are obtained at the expense of high residual stresses or high distortion. Low distortion or residual stresses are usually obtained at a sacrifice in properties. Therefore, the optimal quench rate is one where properties are just met. This usually provides the minimum distortion.

Sources of Distortion

From Fig. 3, it can be seen that there are many sources of residual stress and distortion. In quenching, the primary source of distortion and residual stresses is differential temperatures from the center of the part to the surface or
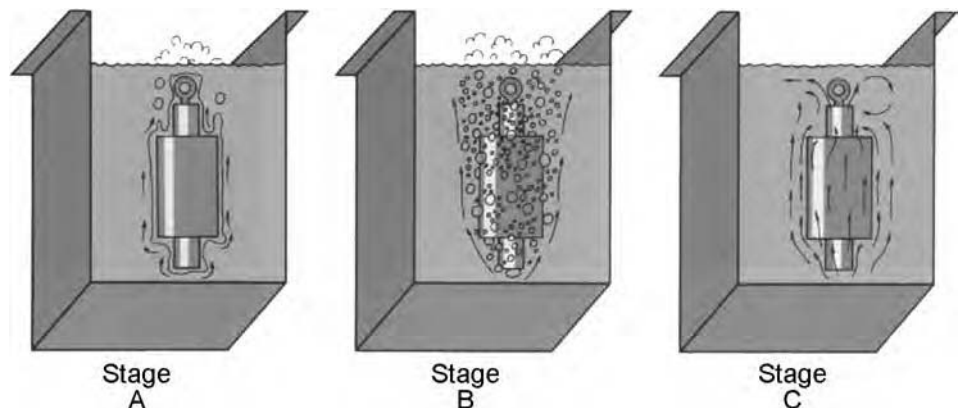

Fig. 1 Schematic of the three stages of quenching. A, vapor stage; B, nucleate boiling stage; and C, convection stage
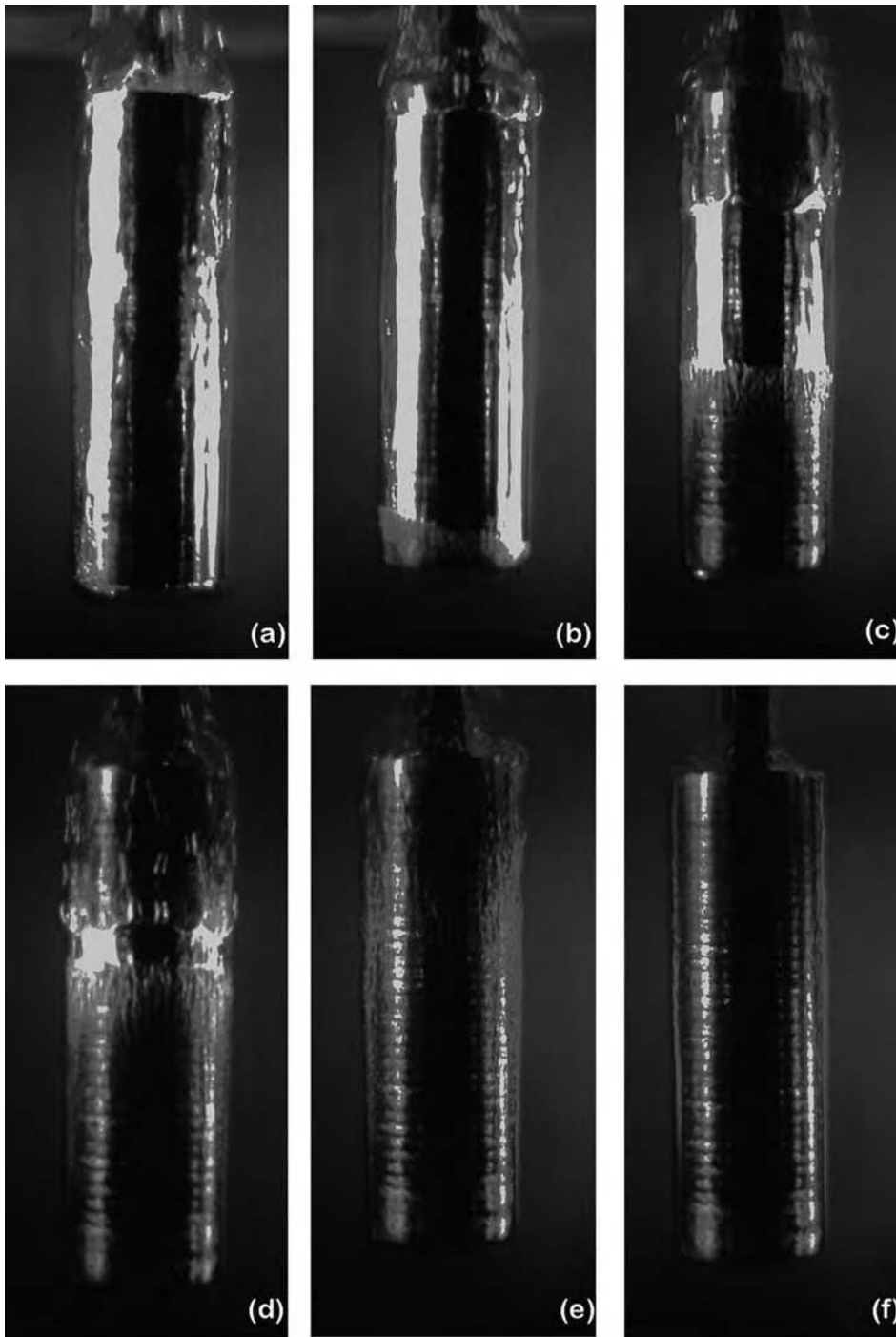
Fig. 2 Progression of the three phases of quenching during the immersion of a solid cylinder

Agitation. Distortion occurs because of differential temperature gradients, whether from the center to the surface or from surface to surface. As can be seen in Fig. 4, all three phases of cooling can be present at the same time, which means that some areas are cooled very slowly, while other parts are cooled rapidly. This has the effect of creating thermal gradients on the surface of the part, which can cause distortion (Fig. 4). The purpose of agitation is to minimize these surface gradients.

It can be seen that increasing the degree of agitation reduces the stability of the vapor phase and increases the maximum rate of cooling. This also has the benefit of minimizing any vapor pockets that can occur and ensuring that the part has a more uniform heat transfer across the surface of a part. Effective agitation is essential to ensure that optimum properties are obtained, to maintain circulation of the quenchant around parts, and to obtain uniform temperature in the bath.

Agitation can be provided by load oscillation, pumped circulation, or motor-driven propellers. Regardless if pumps or agitators are used, baffle or manifold arrangements are necessary to direct the flow of quenchant upward around the parts.

While propeller agitation is easy to design, install, and maintain, it is sometimes difficult to add to an existing quench tank, due to space limitations. In these circumstances, the introduction of pumped circulation through manifolds may be necessary. The use of small, submerged agitators may also be required.

Compressed air is not recommended under any circumstances, because it introduces a nonuniform quench and accelerates oxidation and aging of quench oil. Further, it is likely to introduce water contamination and substantially increase fire hazards.

Gears are the workpiece in this discussion, and oil is the quenchant, but the principles can be generalized.

Quenching characteristics are influenced significantly by the degree of agitation, as shown in Fig. 5 for a normal-speed quench oil under varying degrees of propeller agitation. It can be seen that increasing the degree of agitation reduces the stability of the vapor phase and increases the maximum rate of cooling. This also has the benefit of minimizing any vapor pockets that can occur and ensuring that the part has a more uniform heat transfer across its surface.

Racking of gears is critical to minimize distortion. Parts must be located so that uniform heat transfer will occur on all surfaces of the gear. Uniformity of heat transfer will minimize the formation of thermal gradients on the surface of the parts. The parts must be located so as not to create hot spots from adjacent parts or create mechanical damage from part-to-part interactions.

There are two primary methods for quenching parts. The first method is the use of a press quench (Fig. 6). This specialized technique

from different locations on the surface. By reducing the thermal gradients and differential temperatures, large reductions in residual stresses and distortion can be achieved. The largest factors that affect the creation of large thermal gradients in parts during quenching are temperature, agitation, parts racking, the quenchant chosen, and contamination of the quenchant.

Temperature. Increasing the oil temperature can reduce the distortion and residual stresses in a heat treated component. As the temperature of oil is increased, the temperature gradients in the part are decreased. This is the basic principle of martempering. Using increased temperature can also reduce thermal gradients in cold oils, up to the recommended use temperature of cold oil (typically 180 to 200 °F). Interestingly, increasing the temperature of the oil to approximately 170 °F can increase the speed of cold oil.

Material
- Alloy
- Quench sensitivity
- Prior condition and microstructure
- Alloy segregation
- Decarburization
- Transformation induced stresses

Design
- Alloy procurement
- Alloy selection
- Part geometry

Heat Treat
- Furnace temperature
- Preheat
- Heat-up rate
- Temperature uniformity
- Non-uniform heating
- Racking
- Load density
- Part-to-part interactions
- Carburizing
- Atmosphere control

Prior to Heat Treat
- Machining stresses
- Cold work
- Prior condition and microstructure
- Grinding stresses
- Shot and grit blasting
- Plating for decarburization control

Quenching
- Quench temperature
- Quench agitation
- Type of quenchant
- Contamination
- Racking
- Load density
- Part-to-part interactions
- Handling during quench
- Temperature at withdrawal

Post-Quench
- Part handling
- Delay before tempering
- Washing temperature
- Wash velocity
- Uneven cooling
- Refrigeration

Tempering
- Furnace temperature
- Preheat
- Heat-up rate
- Temperature uniformity
- Non-uniform heating
- Racking
- Load density
- Part-to-part interactions
- Fixturing

Finishing
- Machining
- Grinding
- Pickling
- Shot and grit blasting
- Straightening
- Plating
- Baking after painting
- Stress relief

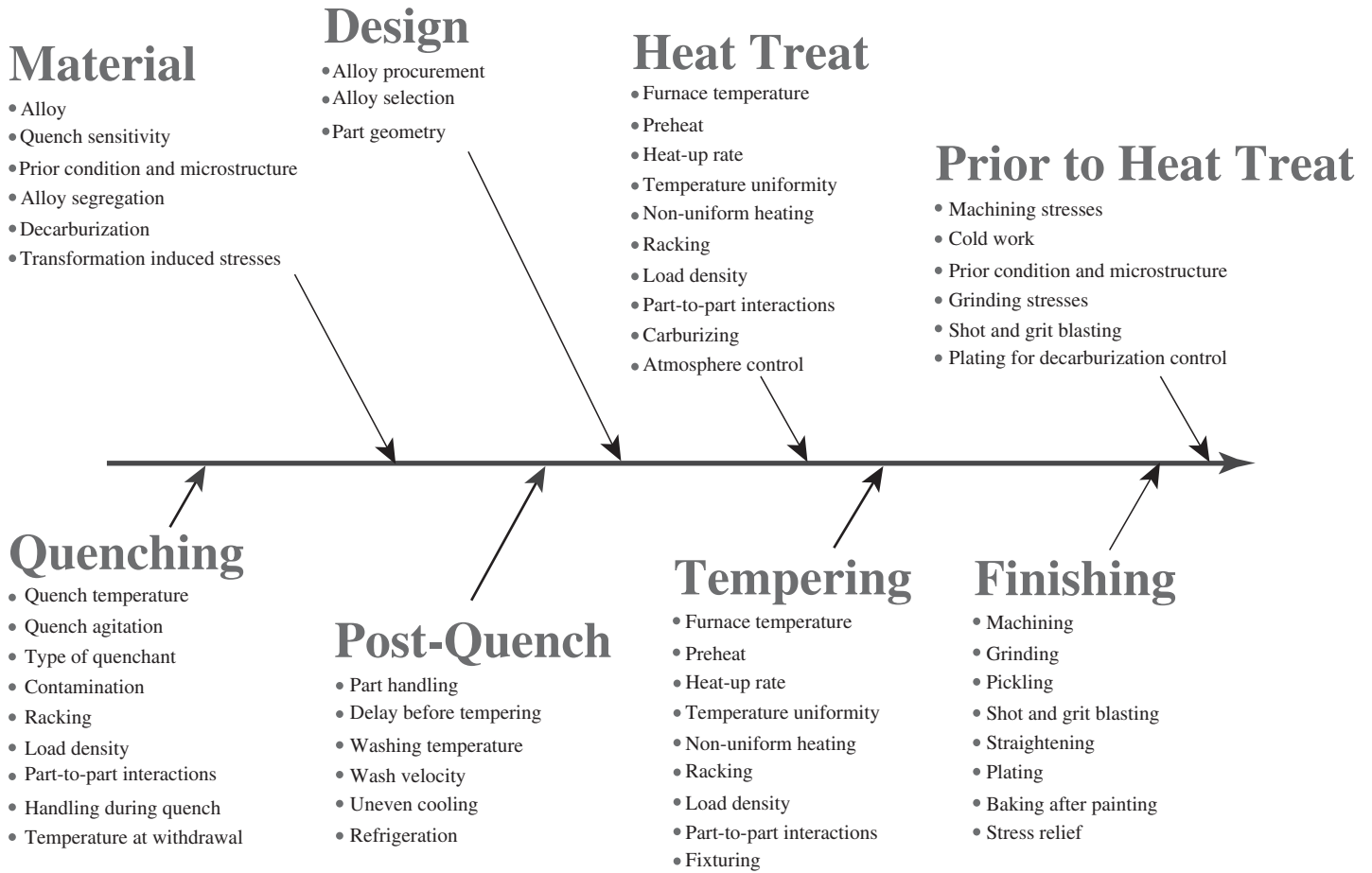**Fig. 3**   Fishbone diagram showing the many sources of distortion or residual stress in a heat treated part
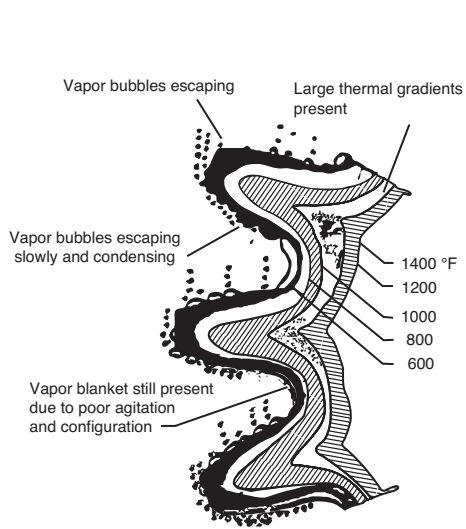

Fig. 4   Schematic representation of quenching a gear, showing the presence of three phases of quenching at the same time. Source: Ref 1
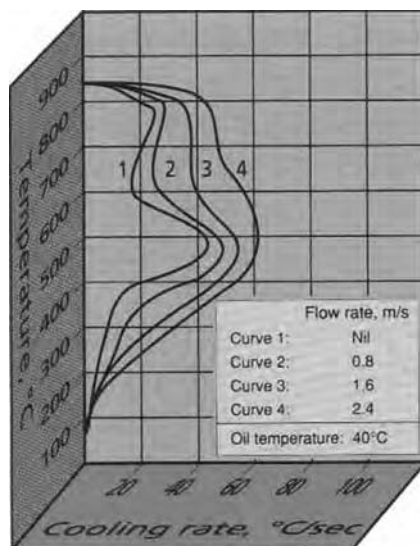

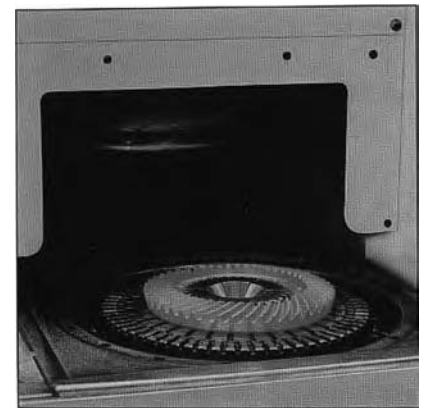**Fig. 5**   Effect of agitation on the cooling curve of an oil quenchant

Fig. 6   Example of a press quench for controlling the distortion of a gear

involves the physical restraint of distortion-prone parts on close-tolerance fixtures during the quenching operation. It minimizes distortion and is used mainly during the quenching of bearing rings and automotive transmission ring gears. It is a manually intensive operation, because each gear must be manually removed from the furnace and placed on a quench fixture. The press is actuated, and a large flow of quenchant is passed through the fixture. Highly accurate and low-distortion parts can be achieved in this manner.

There are several disadvantages to this technique. As indicated previously, it is manually intensive, although some robotized applications have been implemented. Because hydraulic fluids are used to actuate the dies, contamination of the quenchant is a problem. This can cause a change in the cooling rate and quenching characteristics of the quenchant, which can cause cracking or fires. If fire-resistant hydraulic fluids are used and leaking occurs, cracking can occur on the part or the close-tolerance fixture. The quenchant must be routinely checked for contamination and water content. The close-tolerance fixtures used in quench pressing are expensive to manufacture and must be

designed for each gear configuration. Should the gear dimensions change, it is necessary that a new fixture be designed. Further, the life of the dies is finite because of the thermal stresses experienced by the fixture. Distortion and cracking of the fixture can also cause premature replacement of the fixture. As a general rule, cold oils are used to harden the parts. This technique is generally limited to flat and symmetrical parts, such as ring gears.

The second method of quenching gears is to place them on a grid or fixture. Many gears can be heat treated in this fashion, greatly improving production rates. However, there are many ways to rack a gear that often depend on the type of furnace, quenchant, and the preference of the metallurgist.

Typically, ring rears are laid flat on a grid and stacked several high (Fig. 7). They can be offset or stacked directly on top of each other. They are often hung, with supports under the gear. Either method has benefits that depend on the configuration of the gear.

If gears are laid flat, they will tend to bend, or "potato chip," with gears on the top and bottom of the load most prone to this type of distortion. This is due to differential cooling of the gears.

In this case, the thermal mass of the grid retains heat, while the upper surface of the gear experiences the full quenching effect of the oil. The upper surface contracts due to rapid cooling, while the lower surface cools slower and does not experience as much thermal contraction. As the upper surface cools to a point where the martensitic transformation occurs, a volume change occurs, placing the upper surface in tension. When the lower surface cools and the martensitic transformation occurs, a stress reversal occurs, placing the upper surface in tension and the lower surface in compression. This is complicated by the round shape of the part, so that some areas bow up while other areas bow down, resulting in the "potato-chip" shape. The degree of distortion is often dependent on how stiff the section is (polar moment of inertia). This can be overcome by the proper design of racking fixtures. Figure 8 shows examples of properly racked parts to minimize distortion during quenching and to allow the parts to be evenly heated and cooled.

When parts are hung, the weight of the gear often causes the gear to distort, with the gear becoming the shape of an oval. The degree of ovality often depends on the quality of support and the weight of the part. Smaller parts, fully supported, will tend to distort less. Properly designed supports minimize distortion and provide for uniform heat transfer. One advantage of hanging gears is that all sides will experience similar heat transfer, assuming no hot spots or proximity of other parts (creating hot oil spots).

Pinion gears are racked vertically. It is preferred that the heavy section is up and is the last to quench. This is to prevent the shaft from being in the wake of the head. Often, the pinions are offset to allow uniform heat transfer and the minimization of hot spots. Spacers are usually used to maintain the pinions vertically and to prevent movement of the parts.

Contamination and Oxidation. The condition of the quench oil can also contribute to distortion of gears. Contamination of quenching oils with water must be avoided at all cost. As little as 0.05% of water in quenching oil influences quenching characteristics significantly and may cause soft spots, distortion, or cracking (Fig. 9). At concentrations greater than 0.05% or more, foaming during quenching is likely, and this can give rise to fires and explosions. Other contaminants, such as hydraulic oil and fire-resistant hydraulic fluids, can also alter the quenching characteristics, resulting in increased distortion and residual stresses.

The oxidation of quenching oil, measured by the precipitation number or total acid number, is an indication of the level of oxidation of the quenching oil. As the oil oxidizes, it forms organic acids. The formation of oxidized constituents decreases the stability of the vapor phase and increases the maximum cooling rate. This increases the risk of distortion and cracking. The use of stable, high-quality quench oils will reduce the possibility of this occurring.

The life of quench oil is also dependent on the level of oxidation. If the oxidation of the
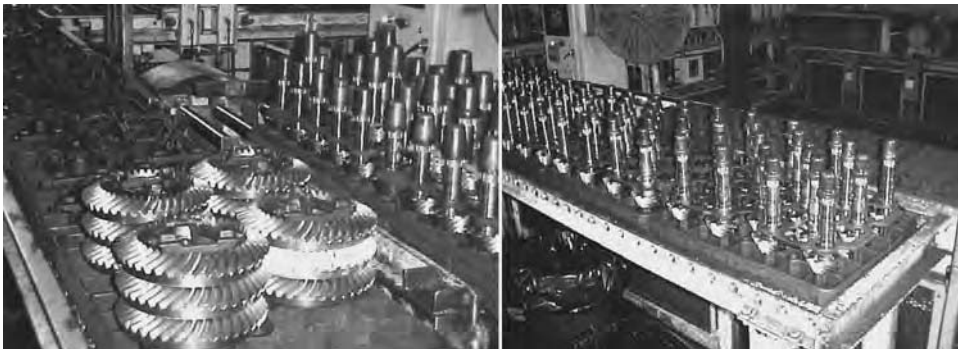

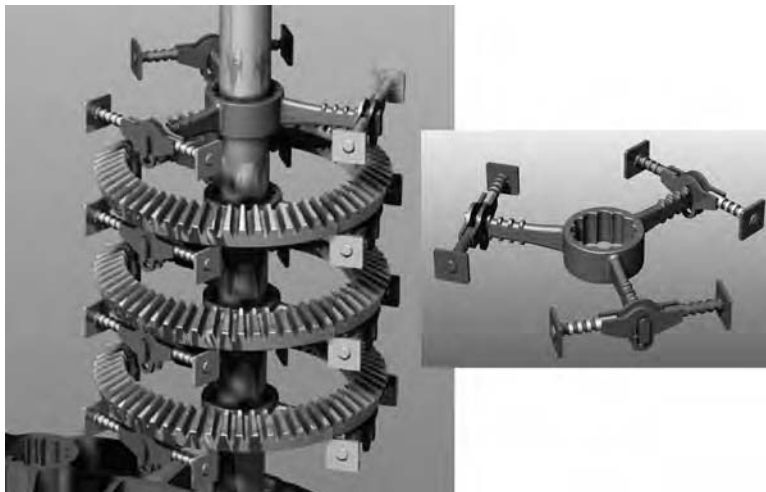**Fig. 7** Examples of ring gears and pinions in a typical heat treating application

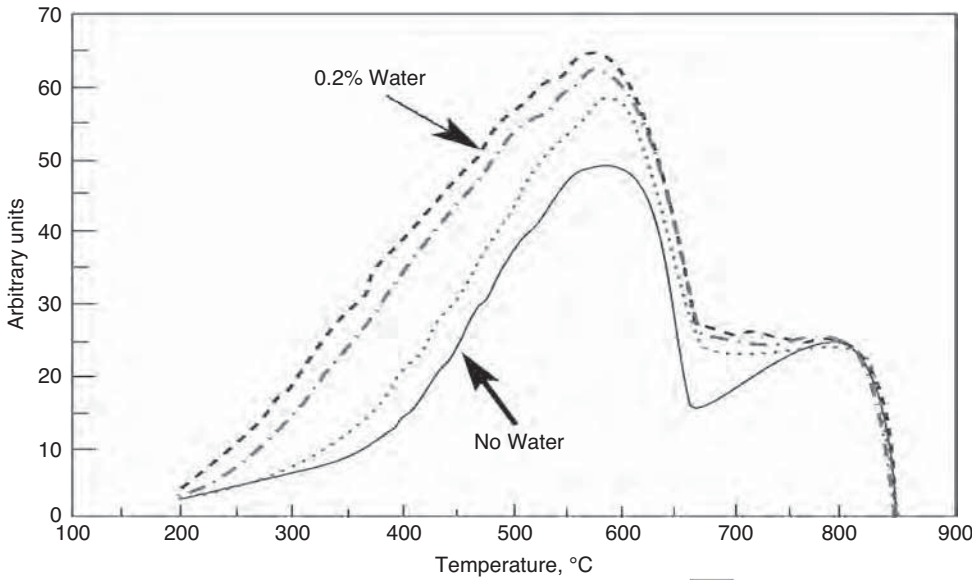**Fig. 8** Advanced racking fixture for proper distortion control

Fig. 9 Effect of water content on the cooling curve of a cold quench oil
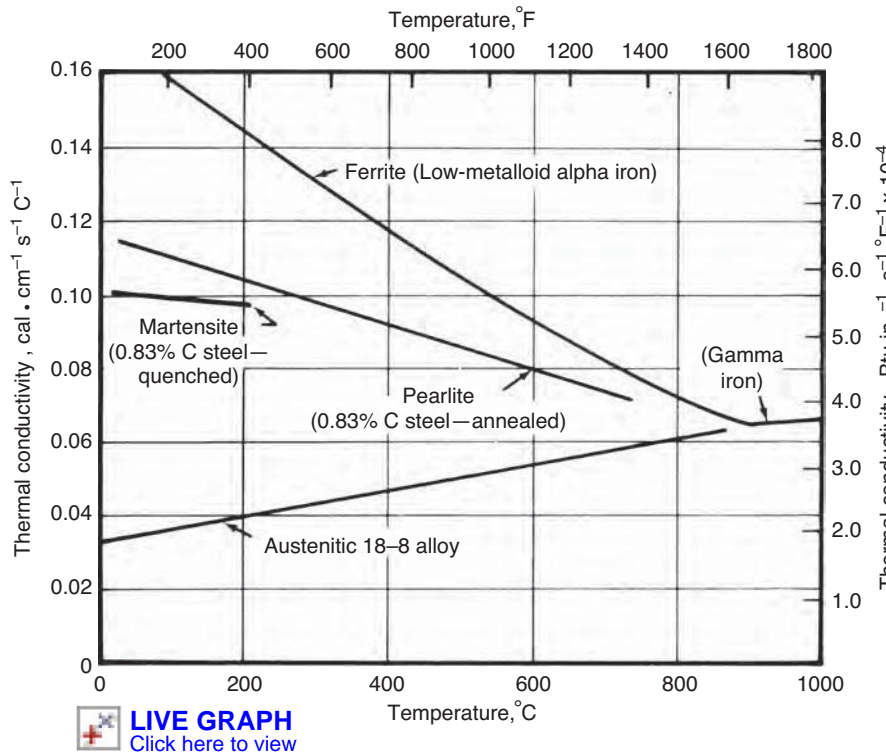
LIVE GRAPH
Click here to view

Fig. 10 Thermal conductivity of various steels

LIVE GRAPH
Click here to view

oil is high, then the oil is prone to staining, resulting in shortened quench oil life. The use of a proactive maintenance program of monthly or quarterly checks for contamination and oxidation is necessary to reduce staining and prolong oil life. Regular filtration of the quench oil is an effective method to reduce contamination, extend the life of the oil, and control the formation of organic acids.

Determination of Heat-Transfer Coefficients

Historical Perspective

Hardening steel by first austenitizing and then quenching involves a process of unsteady-state heat transfer in the steel and removal of the heat from the part by the quenchant. This is coupled with the phase transformations from austenite to martensite or, in the event of a slow quench, the transformation of austenite to bainite and pearlite. The heat-transfer problem has been studied, with investigators making the same assumptions for ease of calculation:

- The thermal conductivity of the steel, k, is a constant over the range considered.
- The thermal diffusivity, α, of the metal is considered to be constant over the temperature range considered.
- The quenching power, H, or the heat transfer remains constant during the entire cooling period.

Conductivity and Diffusivity. As can be seen in Fig. 10 and 11, the thermal conductivity and the thermal diffusivity show considerable variation over the range typically found during heat treatment. The curves for thermal conductivity and thermal diffusivity look very similar. This is not surprising, because the thermal diffusivity is defined as:

$$\alpha = \frac{k}{C_p \rho} \qquad \text{(Eq 1)}$$

where k is the thermal conductivity, α is the thermal diffusivity, C_p is the mass specific heat, and ρ is the density. The symbols used in this article are defined in Table 1.

The quenching power of the quenchant (and agitation), H, is defined as:

$$H = \frac{h_s}{2k} \qquad \text{(Eq 2)}$$

where h_s is the heat-transfer coefficient at the metal-quenchant interface, and k is the thermal conductivity. Interestingly, this equation for the hardening power of a quenchant is very similar to the dimensionless Biot number, Bi, described later in this text.

Grossmann, Asimov, and Urban (Ref 2) placed all steels on a common basis for hardenability. They made the following assumptions:

- Thermal diffusivity was constant.
- The quenching power of the bath, H, was constant.
- If the time to the half-temperature (halfway from the quenching temperature and the bulk temperature of the quenchant) was the same in two different positions in two different diameter bars of the same steel, the microstructure would be the same, and therefore, the hardness would be the same.

Diameters. Grossmann, Asimov, and Urban (Ref 2) defined the term D_u, which is the diameter of the unhardened core, as the diameter where the microstructure was 50% martensite. The larger the bar diameter, D, the greater the unhardened diameter, D_u. This is shown schematically in Fig. 12. The cross-hatched areas represent the unhardened region of the
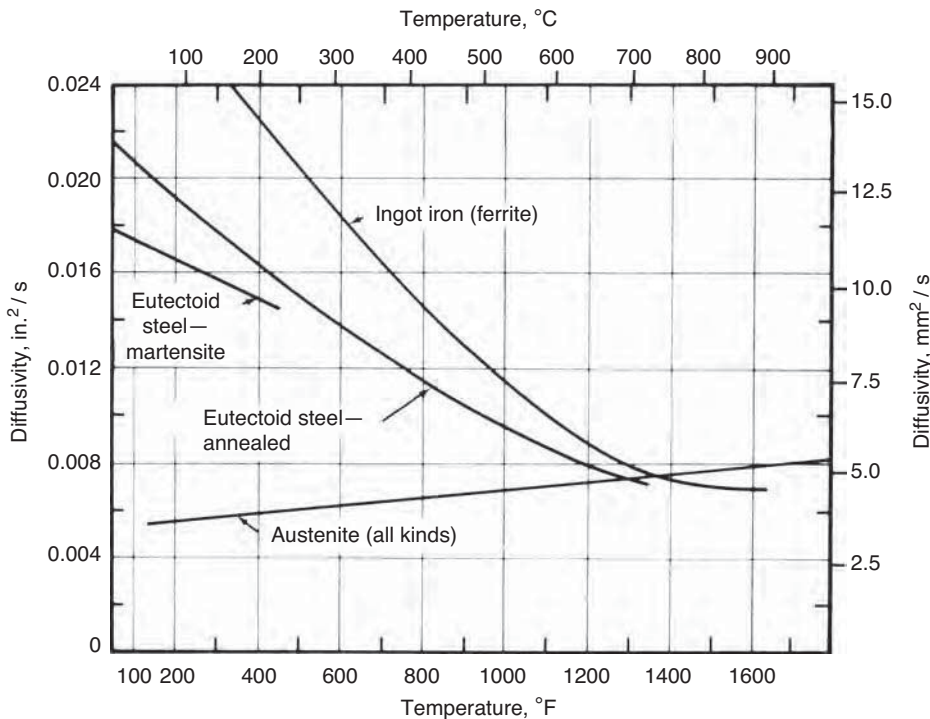
Fig. 11 Thermal diffusivity of various steels

LIVE GRAPH
Click here to view

Table 1 Symbols

Symbol	Definition
A	Area
B	Correction factor
Bi	Biot number
c	Heat capacity
C_p	Mass specific heat
C_{pG}	Heat capacity gas
C_{S_f}	Coefficient liquid-surface combination. See Table 3
D	Diameter
D_c	Critical diameter
D_I	Ideal critical size
D_0	Initial diameter
D_u	Unhardened diameter
$F(L')$	Correction factor
g	Acceleration due to gravity
h	Heat-transfer coefficient
H	Quenching power
h_{Al}	Heat-transfer coefficient, aluminum
h_b	Heat-transfer coefficient, boiling
h_b	Average film-boiling heat-transfer coefficient
h_{Cu}	Heat-transfer coefficient, copper
h_{fG}	Heat-transfer coefficient at metal-film interface
h_{LG}	Heat-transfer coefficient at liquid-gas interface
h_{Ni}	Heat-transfer coefficient, nickel
h_s	Heat-transfer coefficient at metal-quench interface
k	Thermal conductivity
k_g	Thermal conductivity gas
k_G	Conductivity of gas
k_s	Thermal conductivity of solid
L	Vertical height
L'	Characteristic length
L_{crit}	Critical length
L_S	Critical length (different than L_c)
Pr_L^S	Prandtl number
q/A	Heat flux
q_{boil}	Heat flux, boiling
q_{conv}	Heat flux, convection
q_{net}	Net heat flux
T_S	Surface temperature
T_{sat}	Saturation temperature
T_L	Leidenfrost temperature
T_W	Wall temperature
ΔT_x	Excess temperature difference
V	Velocity, volume
V_∞	Velocity at long distance away from object, m/s
α	Thermal diffusivity
μ_G	Viscosity of gas
μ_L	Viscosity of liquid
ν	Kinematic viscosity
ρ	Density
ρ_g, ρ_G	Density gas
ρ_L	Density liquid
σ^*	Surface tension of liquid-gas interface
σ_g	Surface tension of the liquid-vapor interface
∂T	Partial differential of temperature

various-diameter bars. Using various quenching media and different-sized rounds, the half-temperature was calculated. The result is Fig. 13.

This chart is used for determining the H-value of any quenching bath by quenching various-sized rounds of the same steel and determining D_u for each bar. By placing a transparent piece of paper over Fig. 13 and plotting the various values of D_u/D versus D, a curve can be drawn through the points. The tracing paper is then moved horizontally until the curve matches one of the calculated curves. This curve then represents the H-value × diameter curve. Dividing each value of HD by D of the corresponding bar results in a value for the hardening power, H.

Grossmann et al. (Ref 2) also define a quantity D_c as the critical diameter of a steel that, when quenched in a bath of known H, would harden to 50% martensite at its center. An ideal diameter, D_I was also defined as the diameter of a steel that would harden to 50% martensite at its center when quenched in a quenchant where $H = \infty$. In this case, H is the hypothetical quench that reduces the surface temperature of a part to the bulk bath temperature instantaneously.

Grossmann et al. also constructed the chart shown in Fig. 14 that makes it possible to translate D_c with a given H-value into a D_I for the steel. It is also possible to determine the H-value with a known D_I and known D_c.

The variation of H with D for agitated water and oil quenches, determined by Carney (Ref 4), is shown in Fig. 15 and 16. From these data, it is seen that there cannot be a single H-value

for a given quenching bath. The size of the bars quenched should be taken into account when determining the H-value for a specific quenchant system.

Table 2 lists the severity of quench for various quenching media. Note that the agitation or circulation of the quenchant is vaguely defined. This is a problem that is common in the heat treating industry, and it is only with the advent of modern computational fluid dynamics that the agitation rates around parts can be accurately defined.

Determination of Heat-Transfer Coefficients—Analytical and Empirical Results

Quenching is a complex heat-transfer mechanism controlled by the agitation experienced by the workpiece (Ref 6, Ref 7, Ref 8), the geometry of the workpiece (Ref 9, 10), how the workpiece is racked and immersed (Ref 11, Ref 12, Ref 13, Ref 14), the design of the quenching system (Ref 15, Ref 16, Ref 17, Ref 18, Ref 19Ref 20), the workpiece thermal properties, and fluid-specific properties. Most investigations have concentrated empirically on the motion of the quench medium in the tank, with the assumption that if uniform flow is obtained across the part being quenched, then uniform quenching is obtained. Heat transfer is not a steady-state condition. It requires the determination of heat-transfer coefficients as a function of fluid properties, geometry, surface condition, and agitation. It is time- and location-dependent.

Considering the previous transformational and residual-stress discussions, it can be seen that quenching is not path-independent. The quenching process (and the heat-transfer coefficient) must follow a certain "route" if properties are to be maximized and distortion is to be minimized. Anything that will affect the cooling curve shape will impact the as-quenched results.

From the previous discussion, it is seen that there are three distinct regions of heat transfer that occur during quenching: convection, nucleate boiling, and film boiling, which were illustrated in previous figures. Within these primary heat-transfer modes, there are subregimes of heat transfer that must be adequately

explained if accurate prediction of quenched properties and distortion are to be obtained.

Film Boiling. Because quenching occurs at temperatures that are significantly above the boiling temperature of the quenchant, the heat transfer occurring during pool boiling provides a convenient starting point. This heat-transfer system has been extensively studied, and the mechanisms are reasonably well understood. However, it has not been possible to theoretically predict the heat-transfer coefficients during boiling as accurately as single-phase
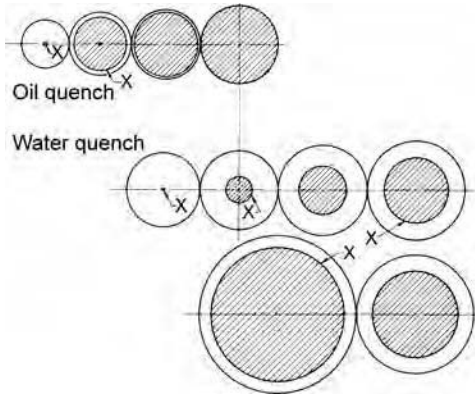

Fig. 12 Schematic representation of the hardening of oil-quenched and water-quenched bars of various diameters. The cross-hatched areas represent the unhardened core. Source: Ref 3

systems. In film boiling, the surface temperature is very high, vaporizing the quenchant. A stable film is formed around the quenched part. This vapor film is insulating, allowing very little heat transfer across the film. In general, heat transfer across this phase is constant and governed by the thermal conductivity of the vapor phase. As the part cools, the temperature at which stable film formation is no longer possible is called the Leidenfrost temperature, T_L. This temperature is influenced by many factors, such as cooling medium, quenched material and geometry, surface state, and agitation. Increasing the Leidenfrost temperature shortens the film boiling regime and increases the heat transfer during nucleate and transition boiling. As soon as the surface of the part reaches the Leidenfrost temperature, the vapor barrier collapses, and the transition boiling phase begins. The transition boiling phase shares characteristics of both the vapor barrier and nucleate boiling heat-transfer regimes. In these two stages, liquid is brought into contact with the hot part and immediately becomes vaporized. Bubbles form at the part surface and carry increasingly large amounts of heat. The temperature of maximum heat flux is called the burnout temperature, T_{BC}.

Tensi, Stich, and Totten (Ref 21) indicted that increasing the surface roughness and the thermal diffusivity of the material results in an increased heat-transfer coefficient. Increasing

sample diameter and oxidation decrease the measured heat-transfer coefficient, while increasing the quenching bath temperature decreases the heat-transfer coefficient. Increasing the agitation rate increases the Leidenfrost temperature and decreases the stability of the vapor phase.

The Leidenfrost temperature, indicating the change from film boiling to nucleate boiling, generates a clearly defined wetting front. In cylindrical specimens, wetting starts at the bottom of the cylinder and propagates upward in an annular fashion. This wetting front influences the distribution of temperature and resultant residual stresses. A moving wave front would generate an axial heat flux in addition to the radial heat flux. This rapidly moving wave front decreases the opportunity for axial heat flow and axial temperature gradients. Using hydrodynamic instability of the liquid vapor interface, the following equation for the minimum heat flux for a flat horizontal surface can be obtained (Ref 22):

$$\left(\frac{q}{A}\right)_{min} = ch_{fG}\rho_g\left[\frac{\sigma_g g(\rho_L - \rho_G)}{(\rho_L - \rho_G)^2}\right] \tag{Eq 3}$$

Stable film boiling has been studied by Bromley (Ref 23), who arrived at an equation describing the film boiling heat-transfer coefficient on a horizontal cylinder:
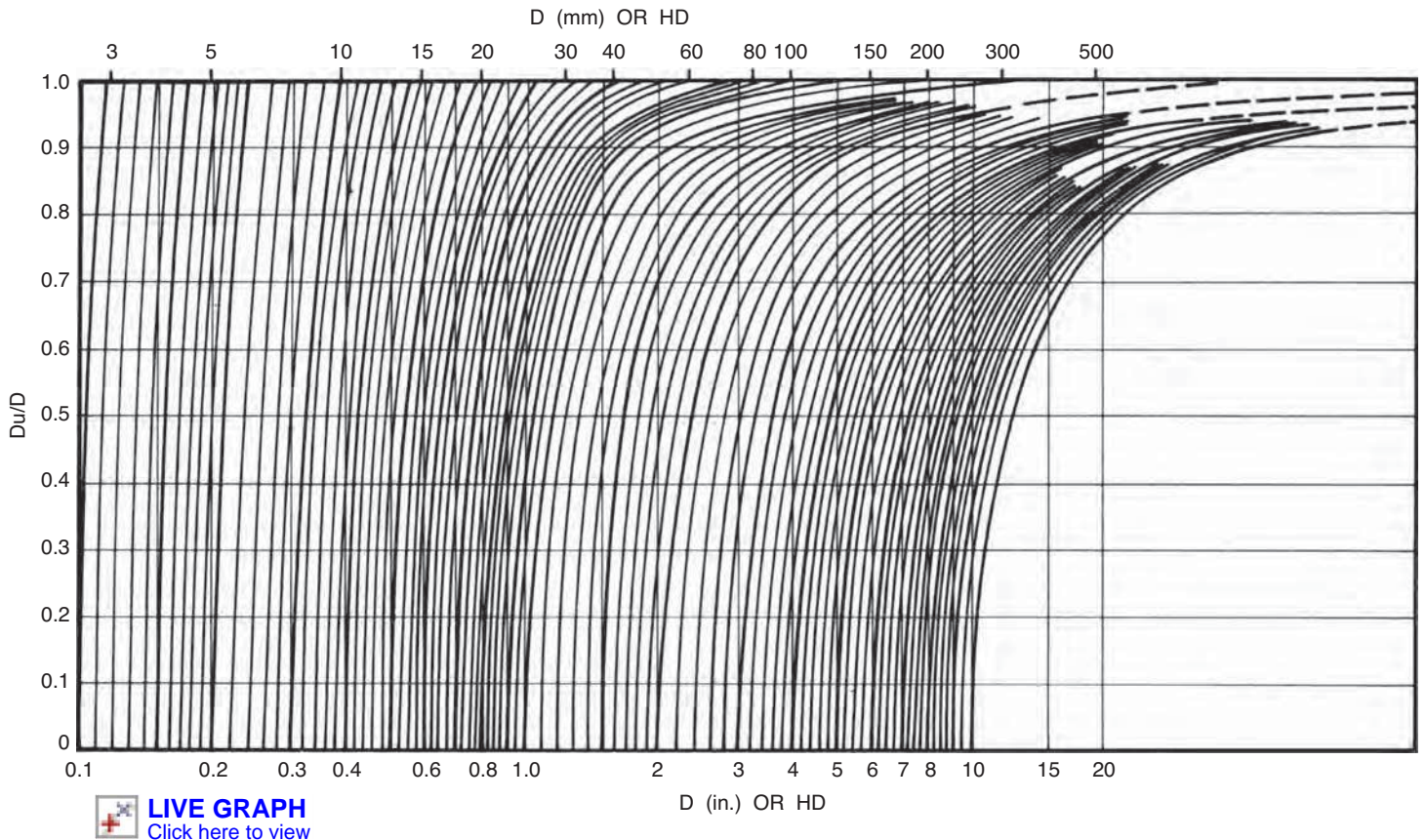
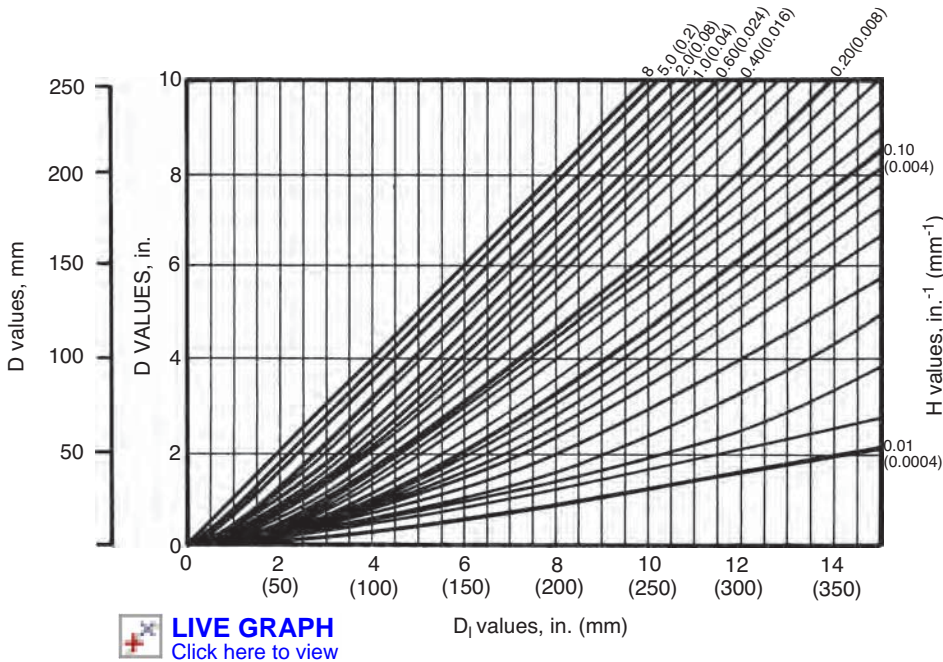**Fig. 13** Curves of D_u/D versus HD for estimating the severity of quench, H, of quenching baths

Fig. 14 Relationship between the actual critical size, D_c, the ideal critical size, D_I, and the severity of quench, H

Fig. 15 Variation of H-value at half-temperature time in round bars water quenched from 845 °C (1550 °F)

Fig. 16 Variation of H-value at half-temperature time in round bars oil quenched from 845 °C (1550 °F)

Table 2 Severity of quench, H, for various quenching media

Agitation	Air	Oil	Water	Brine
No circulation	0.02	0.25–0.30	0.9–1.0	2
Mild circulation	...	0.30–0.35	1.0–1.1	2.0–2.2
Moderate circulation	...	0.35–0.40	1.2–1.3	...
Good circulation	...	0.40–0.50	1.4–1.5	...
Strong circulation	0.05	0.5–0.8	1.6–2.0	...
Violent circulation	...	0.8–1.1	4	5

Source: Ref 3, 5

$$h_b = 0.62\left[\frac{k_G^3\rho_G g(\rho_L - \rho_G)\left[h_{fG} + 0.4C_{p_G}\right](T_W - T_{sat})}{D\mu_G(T_W - T_{sat})}\right]^{1/4}$$
(Eq 4)

To validate Eq 4, Bromley correlated film boiling data for a variety of fluids on horizontal carbon cylinders of various diameters over a wide range of ΔT up to l400 °C. He found good correlation except in the case of small-diameter wires, where the agreement between experiment and theory deviated by as much as 100%.

The effect of fluid flowing past the surface of a tube, similar to quenchant flow across a series of shafts, was suggested by Bromley (Ref 24):

$$\overline{h_b} = 2.7\sqrt{\frac{V_\infty k_g \rho_g \lambda'}{D_0 \Delta T_x}}$$
(Eq 5)

where:

$$\lambda' = h_{LG}\left(1 + \frac{0.4C_{p_G}\Delta T_x}{h_{LG}}\right)$$
(Eq 6)

for:

$$V_\infty > 2\sqrt{gD_0}$$
(Eq 7)

At bulk velocities less than:

$$2\sqrt{gD_0}$$

the flow is not fully developed, and the heat-transfer coefficient can be evaluated according to Eq 4. For vertical surfaces of height L, the following equation has been suggested (Ref 25):

$$\overline{h_b} = 0.943\left[\frac{k_G^3\rho_G(\rho_L - \rho_G)gh_{fG}g\lambda'}{L\mu_G\Delta T_s}\right]$$
(Eq 8)

This equation is satisfactory as long as the vapor film remains laminar. As the vapor rises, there is a critical Reynolds number that is reached at some critical point. This critical length, L_{crit}, can be estimated by (Ref 26):

$$L_{crit} = \frac{100\mu_G\lambda'}{2k_G\Delta T_x}\left[\frac{200\mu_G^2}{g\rho_G(\rho_L - \rho_G)}\right]^{1/3}$$
(Eq 9)

For surfaces that have a height greater than the critical length, Bankoff (Ref 27) showed that the equation:

$$\overline{h_b} = 0.20\rho_G C_{p_G}\left(\frac{g^2L\mu_G}{\rho_G}\right)^{1/3}\left(\frac{\rho_L - \rho_G}{\rho_G}\right)^{2/3}\left(\frac{C_{p_G}\Delta T_x}{h_{fG}}\right)^{1/3}$$
(Eq 10)

exhibits good correlation for predicting the average heat-transfer coefficient. The validity of Eq 10 is limited to vertical heights up to 6.5 in. Increasing agitation will facilitate the rupture of the vapor blanket, subsequently reducing the time for nucleate boiling formation. The role of cooling during both nucleate boiling and convective heat transfer will increase with increasing agitation.
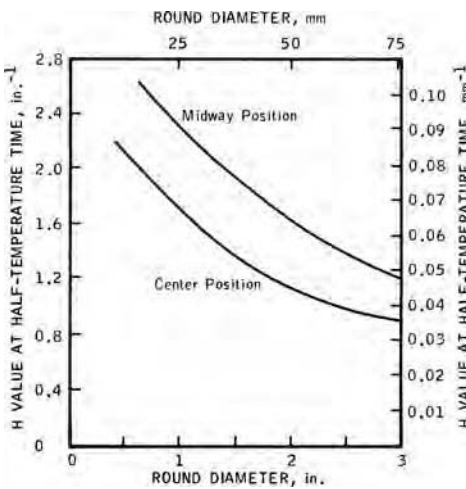
Nucleate Boiling Regime. The nucleate boiling regime involves two separate processes: the nucleation of bubbles on the surface, and the growth and escape of these bubbles. The heat transfer in this regime is affected by the nucleation process, the distribution of nucleation sites on the surface, and the growth and escape of the bubbles. As the number of active sites increases, the interaction of these bubbles becomes important. In addition to these variables, the fluid properties, surface condition (including mechanical and thermal properties), and part geometry are also important.

There have been numerous investigations into the process of nucleate boiling without applied agitation. Discussion of the empirical correlations between the heat flux, q, and the difference in the surface temperature and bulk fluid temperature, ΔT, are given in Ref 28 and 29. The most commonly used correlation of the heat flux and ΔT_x was developed by Rohsenow (Ref 30):

$$\frac{q}{A} = \mu_L h_{fG} \left[\frac{g(\rho_L - \rho_G)}{g_c \sigma^*} \right]^{1/2} \left[\frac{C_{p_L}(T_w - T_{sat})}{C_{sf} h_{fG} \Pr_L^S} \right]^3 \quad \text{(Eq 11)}$$

In Eq 11, the coefficients S and C_{sf} are provided to adjust the correlation to satisfy the liquid-surface combination. The coefficient S should be taken as 1.0 for water and 1.7 for all other liquids. Table 3 provides for values of C_{sf} to be used in Eq 11. A statistical correlation of the published experimental data of water in the nucleate boiling regime was conducted by Stephon and Abdelsalam (Ref 31) and found to be in excellent agreement with Eq 11. The values of the surface-liquid tension can be found in standard heat-transfer texts (Ref 32, 33).

The most important variables affecting the empirical constant C_{sf} are the surface roughness of the part, which determines the number of nucleation sites on the part at a given temperature (Ref 34), and the angle of contact between the bubble and the heated surface, which is a measure of the wettability of the surface and the fluid. Because of the lack of definitive information regarding surface condition, material, and fluid combinations, the constant C_{sf} must be determined experimentally for each material, surface, and fluid combination.

An advantage of the Rohsenow correlation is that the performance of any liquid-surface combination in nucleate boiling can be determined with a single test. Only one value of the heat flux (q/A) and the corresponding value of the excess temperature difference, ΔT_x, is necessary to evaluate C_{sf}, in Eq 11.

The geometrical shape of the heated surface has been found to have little effect on the nucleate boiling mechanism (Ref 35, 36). The reason is that the action of bubbles is localized to a region near the heated surface. However, other investigators have found substantial systematic variation of the heat-transfer coefficient in cylinders and rings. In the study by Segerberg and Bodin (Ref 37), the difference between the upper and lower cylinder heat-transfer coefficients was thought to be due to natural convection currents carrying the hotter quenchant up to the upper portion of the cylinder. If this were the case, then it would be expected that the upper portion of the cylinder would cool slower because of the decreased difference in temperature between the quenchant and the quenched part. It is more likely that a critical length was exceeded, and a transition between laminar-to-turbulent heat transfer occurred, resulting in increased heat transfer.

Flat rings were also quenched vertically, with the resulting heat-transfer coefficients approximating the relative magnitude of the quenched cylinders. In this case, the overall height was less, minimizing the effects of the transition from laminar to turbulent heat transfer. In this case, the upper surface of the ring showed a decrease in the heat-transfer coefficient, while the lower surface showed a heat-transfer coefficient twice the size of the upper surface. In all cases, the peak heat-transfer coefficient occurred at approximately the same temperature. When a wide, flat ring was quenched horizontally, the upper surface showed a heat-transfer coefficient of approximately the same magnitude as the cylinder, while the lower surface displayed a significantly reduced (approximately 20% of the upper surface) heat-transfer coefficient. This behavior was attributed to a layer of film forming on the bottom surface of the wide, flat ring, acting as an insulating layer. The escape of bubbles from the underside was more difficult, hence impeding heat transfer. This could be explained quantitatively if the differences in the net heat flux on the top and bottom surfaces were the result of the additive effects of natural convection and nucleate boiling:

$$q_{net} = q_{conv} + q_{boil} \quad \text{(Eq 12)}$$

In this case, the heat transfer due to convection is much larger on the upper surface than on the bottom surface, resulting in the differences in the heat-transfer coefficient. When the wide, flat ring was quenched vertically, the measured heat-transfer coefficient was nearly identical on each of the sides, giving further credence to the idea that the net heat flux is additive between convection and boiling. In this experiment, the heat-transfer coefficient of the inner upper surface of the ring was also measured. The relative magnitude was the same; however,

transition from film to nucleate boiling was at a much higher temperature. This could be due to separation of the vapor film at the inner edge of the ring or to the corner contributing to a much greater amount of turbulence, which, in turn, increases the Leidenfrost temperature, causing nucleate boiling to occur at a much higher temperature. This has been qualitatively confirmed by Maass and Jeschar (Ref 38). In this investigation, the transition from film boiling to nucleate boiling occurred at different temperatures for spheres and cylinders. This was consistent for all quenching temperatures investigated. Cylinders reached the transition temperature sooner than the spheres. Again, this is thought to be due to the increased turbulence along the length of the cylinder, with the edge of the cylinder acting as a "trip wire."

Segerberg and Bodin (Ref 37) also quenched similar-geometry, vertical cylinders of identical lengths and material but of different diameters, instrumented with thermocouples, near the surface of the cylinder at the midpoint of the length. Here, the measured heat-transfer coefficients were nearly identical, with very little scatter throughout the quenching temperatures. The peak flux and the peak heat-transfer coefficient were nearly identical in magnitude. While Eq 11 describes the heat flux during nucleate boiling, it does not predict the peak heat flux. An expression was developed (Ref 39, 40) for the peak heat flux based on the stability requirement of the liquid-vapor interface. This correlation is of the form (Ref 41):

$$\left(\frac{q}{A}\right)_{max} = 0.131 F(L') \rho_L^{1/2} h_{fG} [\sigma^* g g_c (\rho_L - \rho_G)]^{1/4} \quad \text{(Eq 13)}$$

where $F(L')$ is a correction factor for heated surface geometry and size. This factor depends on the dimensionless characteristic length, L', of the heated surface:

$$L' = L \sqrt{\frac{g(\rho_L - \rho_G)}{\sigma^*}} \quad \text{(Eq 14)}$$

Recommended values of L' are shown in Table 4 for various surface geometries. These relations are correlations for increases in heat flux for liquids at the saturated temperature. Subcooling, where the bulk fluid is at a much lower temperature than the saturation temperature, will substantially increase the heat flux and the rate of quenching (Ref 42–44). McAdams et al. (Ref 45) investigated the effect of subcooling the bulk fluid on the surface heat flux. They found that increasing the amount of subcooling increased the heat flux dramatically. This also caused the excess temperature difference, ΔT_x to increase with subcooling. This would indicate that as the quenchant temperature is lowered, the maximum heat flux temperature would move upward, increasing the overall heat-transfer coefficient. When a quenchant is subcooled, as is normally the case in heat treating, the maximum heat flux due to

Table 3   Values of the coefficient C_{sf} of Eq 11 for various liquid-surface combinations

Liquid-surface combination	C_{sf}
Water/copper	0.0130
Water/scored copper	0.0068
Water/emery-polished copper	0.0128
Water/chemically etched stainless steel	0.0133
Water/mechanically polished stainless steel	0.0132
Water/ground and polished stainless steel	0.0080

Table 4 Correction factor $F(L')$ for use in Eq 13

Part geometry	$F(L')$	Remarks
Infinite flat plate, facing up	1.14	$L' \geq 2.7$; L is part width or diameter
Horizontal cylinder	$0.89 + 2.27e^{-3.44\sqrt{L'}}$	$L' \geq 0.15$; L is the cylinder radius
Small sphere	$1.734L'^{-1/2}$	$0.15 \leq L' \leq 4.26$; L is the sphere radius
Large sphere	0.84	$L' \geq 4.26$; L is the sphere radius
Large finite body	0.90	$L' \geq 4$; $L = V/A$

subcooling can be estimated from the equation (Ref 46):

$$\left(\frac{q}{A}\right)_{\text{max}} = \left(\frac{q}{A}\right)_{\text{max sat}} \left[1 + \frac{24}{\pi h_{f_G} \rho_G} \left[\frac{2k_L(T_{\text{sat}} - T_\infty)}{\sqrt{\pi \alpha_L \tau}}\right]\right]$$
$$\left[\frac{\rho_L^2}{\sigma^* g(\rho_L - \rho_G)}\right]^{1/4}$$

(Eq 15)

where

$$\tau = \frac{\pi}{3}\sqrt{2\pi}\left[\frac{\sigma^*}{g(\rho_L - \rho_G)}\right]^{1/2}\left[\frac{\rho_G^2}{\sigma^* g(\rho_L - \rho_G)}\right]^{1/4}$$

(Eq 16)

The ratio of the critical heat flux for the subcooled and saturated boiling can be expressed as:

$$\frac{q_{\text{subcooled}}}{q_{\text{max sat}}} = 1 + B(T_{\text{sat}} - T_\infty)$$

(Eq 17)

where Kutateladze (Ref 47) found an empirical relationship for B:

$$B = 0.065\frac{C_{p_G}}{h_{f_G}}\left(\frac{\rho_L}{\rho_G}\right)^{0.8}$$

(Eq 18)

In general, the heat-transfer coefficient increases with the metal thermal conductivity. Bamberger and Prinz (Ref 48) determined heat-transfer coefficients for several materials and found that the heat-transfer coefficient changed with thermal conductivity (in the nucleate boiling regime). While the resultant slopes of the quenched materials were similar, the absolute values at specific surface temperatures could be ranked according to thermal conductivity.

For copper, aluminum, and nickel, the heat-transfer coefficient as a function of surface temperature, T_s (°C) is given (Ref 46) as:

$$\text{Log}(h_{\text{Cu}}) = 4.9788 - 0.0017T_S$$
$$\text{Log}(h_{\text{Al}}) = 4.8574 - 0.0020T_S$$
$$\text{Log}(h_{\text{Ni}}) = 4.6023 - 0.0017T_S$$

(Eq 19)

This is not apparent from Eq 11, except in the correction factor C_{sf}.

McAdams et al. (Ref 45) also investigated the effect of fluid agitation on the surface heat flux, (q/A), and the excess temperature difference, ΔT_x. They found that increasing the agitation or velocity of the fluid increased the surface heat flux and the excess temperature difference at which the maximum heat flux

occurred. This has been confirmed by several other investigators (Ref 49, Ref 50, Ref 51). Kreith (Ref 33) stated that the total net heat flux due to forced convection and pool boiling can be summarized by:

$$\left(\frac{q}{A}\right)_{\text{net}} = \left(\frac{q}{A}\right)_{\text{boil}} + h_c(T_s - T_\infty)$$

(Eq 20)

where h_c can be determined using standard forced convection relationships found in many heat-transfer texts. The boiling heat flux can be evaluated using Eq 11.

These fundamental studies show that quenchants mediate heat transfer by the formation of fluid films around the hot metal parts. Fluid flow rates and turbulence (these are two components of the general term *agitation*) enhance the rupture of these films, thus accelerating cooling rates. In addition to the mass flow rate of the quenchant by the metal part, wetting kinematics may dramatically affect the uniformity of the cooling process. Finally, anything that affects the cooling profile on the surface of the metal part may potentially affect the formation, or reduction, of thermal and transformational stresses.

Based on the aforementioned relationships, it should be possible to determine the net surface heat flux or the heat-transfer coefficient during quenching. The use of the additive relationships found in Eq 12 and 20, the basic relationships for film boiling in Eq 4, and the correlation for nucleate boiling in Eq 11 will be of substantial use in determining the boundary conditions for modeling the distortion and residual stresses in quenched structures.

Determination of Heat-Transfer Coefficients—Application of Cooling Curves

Probes, used to determine the cooling rate of quenchants, have been produced in a large variety of sizes and shapes. These are generally stand-alone systems used to determine either the cooling curve of a quenchant for quality-control purposes or the behavior of a part. The probes have been made in a variety of sizes and shapes, including cylinders, plates, and rings (Ref 47, Ref 52–54). Various materials, such as silver (Ref 55), gold (Ref 56), nickel-base alloys (Ref 57), and stainless steel (Ref 58) have been used in the construction of probes. There are generally two types of probes: those that are Newtonian (lumped system) and those that are non-Newtonian. Because of a lack of phase transformations, and latent heat from these phase

transformations, single-phase materials, such as silver and nickel-base alloys, are preferred. Latent heat generation during phase transformations generally complicates cooling-curve behavior. Because of this, alloy steel probes are generally not used except in specific cases.

Newtonian probes are sized in such a way that the temperature gradients within the probe are minimal. To determine if the temperature distribution can be considered uniform within the solid (and a lumped-system analysis appropriate), it is first necessary to establish a critical length, L_s, as:

$$L_S = \frac{V}{A}$$

(Eq 21)

and the dimensionless quantity, the Biot number, Bi:

$$Bi = \frac{hL_S}{k_s}$$

(Eq 22)

where k_s is the thermal conductivity of the solid. For solids in the shape of a cylinder or sphere, the temperature distribution during transients within the solid is typically uniform, with a temperature gradient of less than 5% if:

$$Bi = \frac{h_s L_S}{k_s} \ll 0.1$$

(Eq 23)

The assumption of uniform temperature within the solid is valid if the specific conductance of the solid is much larger than the heat-transfer coefficient.

Remember that for the aforementioned, the quenching power of the quenchant (and agitation), H, was defined as:

$$H = \frac{h_s}{2k}$$

The product HD is commonly used with Fig. 13 to calculate the severity of quench, H, with D being the diameter of a bar. Using the definition of the Biot number, Bi, then:

$$H = \frac{h_s}{2k}$$
$$Bi = \frac{h_s L_S}{k_s}$$
$$Bi = 2HD$$

(Eq 24)

This basic relationship can provide an approximate value for the severity of quench, H, which can be used to estimate steel properties and microstructure.

Calculation of the heat-transfer coefficient is simple and requires the solution of the basic equation:

$$-h_c A(T - T_\infty) = \rho V C_p(T)\frac{dT}{dt}$$

(Eq 25)

Rearranging this equation:

$$h_c = \rho\frac{V}{A}C_p(T)\frac{dT}{dt}(T - T_\infty)$$

(Eq 26)

This lumped-system analysis is only valid when the Bi number is much less than 0.1.

One system, using the lumped-system approach, is the Center for Heat Treating Excellence (CHTE) probe (Ref 59). This system consists of a computer-driven data-acquisition device, pneumatic cylinder with air valve, a small box furnace, 1 L container for the quenchant, and the probe (Fig. 17). The pneumatic cylinder moves the probe up and down and provides for a consistent and repeatable quench. One primary feature of this device is the ability to quickly change the probe tip. This enables probe tips to be fabricated from any material and to investigate different oxides or surface roughnesses.

This probe is easy to use and is capable of obtaining heat-transfer coefficients quickly. However, because of its small size, the probe does not have enough thermal mass to observe a stable vapor phase that would be observed on a part. This has the net effect of calculating a heat-transfer coefficient that is significantly faster than probes of larger size.

The cylindrical silver probe adopted for use in Japan (Ref 56) has the advantages of lumped-system analysis and direct determination of heat-transfer coefficients. It also has no coupled phase transformations or issues with surface oxidation. The Japanese Industrial Standards (JIS) probe is a solid cylinder of 10 mm diameter and has a length of 30 mm (Fig. 18). These probes are polished with 500-grit emery paper prior to use. Several issues have prevented its use in the United States. The cost of silver is prohibitive, and the mechanical strength of silver is low.

Other temperature probes used in the measurement of cooling curves are of sufficient size or of materials that have lower thermal conductivity; that is, the $Bi > 0.1$. In this case, significant thermal gradients exist within the probe. Determination of the heat-transfer coefficients is much more difficult and requires more intensive numerical methods.

One such probe is described in ISO 9950 and ASTM 6200 (Fig. 19). It is a probe fabricated from Inconel 600, 12.5 mm in diameter, and 60 mm long. Inconel 600 is a nickel-base superalloy that is heat resistant, has high resistance to oxidation, and does not have any coupled phase transformations. Because of the relative low cost of this probe and acceptance by international standards organizations, the ISO probe is the preferred quenchant characterization method. Because of its relatively small size (but still larger than the CHTE probe), ISO 9950 probes are not directly applicable to predicting as-quenched parts under industrial conditions. This method provides a precise laboratory method for the characterization of quenchants (Ref 37). When modified to provide agitation (Ref 60, 61), different quenchants can be compared and the data used in quenchant maintenance. Advantages of the ISO 9950 probe include:

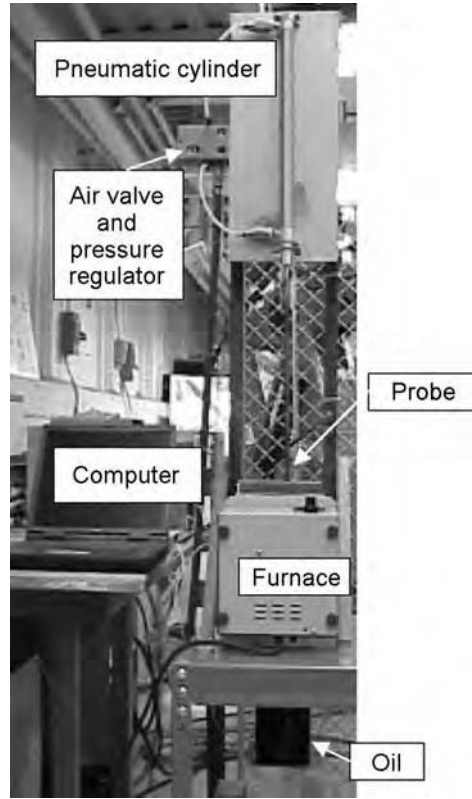- Interlaboratory reproducibility, demonstrated by round-robin studies (Ref 62)

Fig. 17 CHTE probe, Center for Heat Treating Excellence

- Thermal conductivity close to steel (much closer to steel than silver)
- Uniform surface oxide that is stable through the ranges of interest in steel heat treating
- Small probe size that is convenient for routine use in monitoring quenchants
- Small size of quenchant needed for testing (approximately 2 L)
- No phase transformation in Inconel 600 during quenching

Another larger probe that has a following is the Liscic/NANMAC probe (Ref 63). This is a cylindrical probe that is 200 mm long, 50 mm in diameter, and made from AISI 304 stainless steel. This probe contains three thermocouples, all located in the same radial plane along the geometrical center of the probe length. One thermocouple is located on the surface of the probe, using a proprietary thermocouple design that is self-renewing. The length of the probe allows the probe to be modeled as an infinite cylinder, and to that end, effects can be neglected. This symmetry of heat flow in only the radial direction greatly simplifies calculations. It has been reported (Ref 64) that the response time of the thermocouple is on the order of 10^{-5} s, making it very sensitive to very small surface heat-transfer conditions. This probe is shown in Fig. 20.

Consider a probe with a thermocouple at the center of the probe. The temperature of the
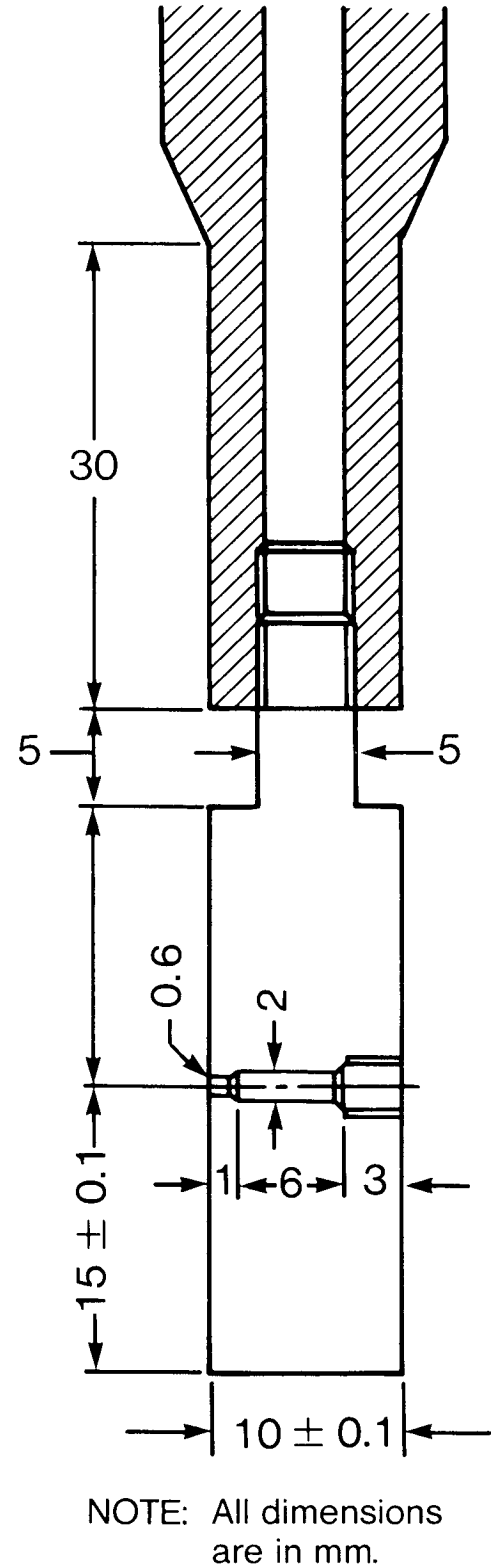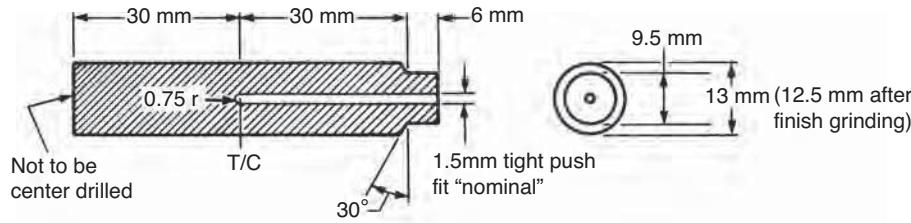

NOTE: All dimensions are in mm.

Fig. 18 JIS probe, Japanese Industrial Standards

probe is recorded at a discrete time, depending on the data-acquisition rate. It is desired to determine the heat flux at the surface ($x = L$) of the probe. The temperature, T, at any
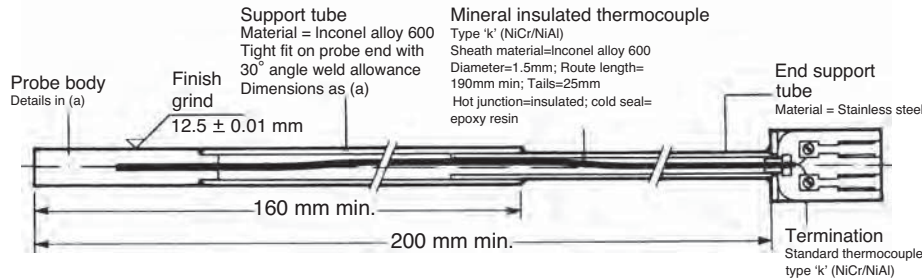
(a) Probe details

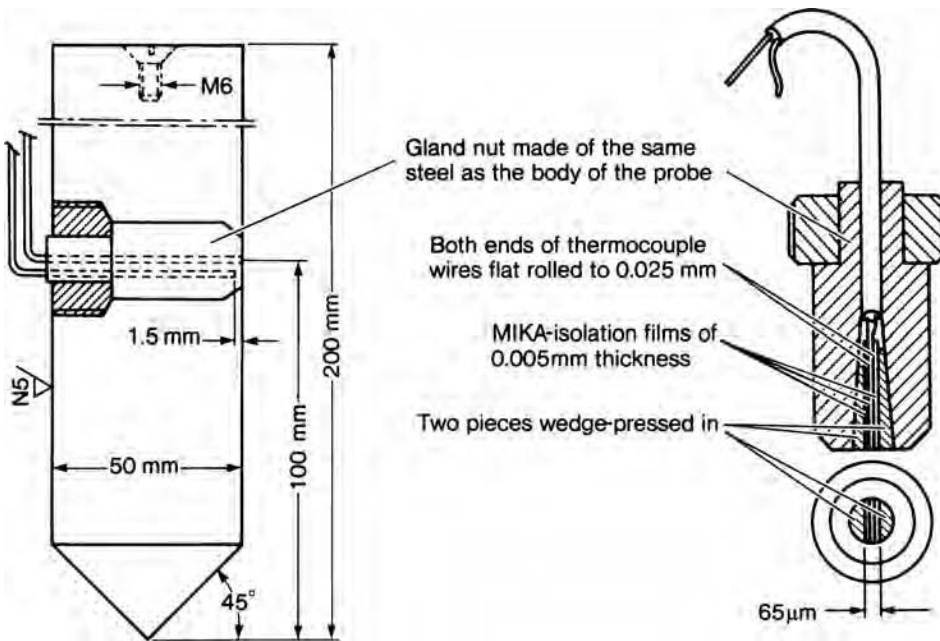**Fig. 19** ISO 9950 probe, International Organization for Standardization

Fig. 20 Liscic/NANMAC probe manufactured by NANMAC Corp.

time, t, and position, x, is governed by the standard conduction equation:

$$\rho c \frac{\partial T(x,t)}{\partial t} = k \frac{\partial^2 T(x,t)}{\partial x^2} \text{ for } 0 < x < L \text{ and}$$
$$0 < t < t_{max}$$
(Eq 27)

where L is the dimension perpendicular to the quenched surface. To make things simple, the temperature effects on density, ρ, thermal conductivity, k, and specific heat, c, are considered to remain constant. For a conventional heat conduction problem, the previous equation would be solved using the following boundary conditions:

$$\frac{\partial T}{\partial x}(0,t) = 0 \text{ and} - k\frac{\partial T}{\partial x}(L,t) = q(t)$$
(Eq 28)

where $q(t)$ is the surface heat flux and the initial condition $T(x,0) = f(x)$ for $0 \le x \le L$ to determine the temperature field. In other words, the surface-boundary conditions are known. In the case of $Bi < 0.1$, the assumption is made that the thermal gradients are negligible and that

the surface-boundary conditions will be nearly identical.

For a probe where $Bi > 0.1$, the thermal gradients cannot be ignored. In this case, the temperature at one location is known at discrete times, t_i:,

$$T(x_0, t_i) \text{ for } i = 1, 2, 3, \cdots i_{max}$$
(Eq 29)

where the thermocouple is located at the center of the probe, at x_0. If the thermocouple were located at $x_0 = L$, then the heat-transfer coefficient and the heat flux could be determined directly (Ref 65). If $x_0 = 0$, then the data needed and the measurements are on opposite surfaces. This problem type is an example of an inverse heat conduction problem (Ref 66).

In the inverse heat conduction problem, the temperature as a function of time at some location is known at some point within the part. The problem is determining the boundary conditions at the surface that produced the measured temperature. In other words, $T(x_0,t)$ is known, and it is desired to calculate the heat flux, $q(t)$, that produces $T(x_0,t)$. The calculation must take into account that any surface heat flux variations will be damped by the delay time before it reaches the thermocouple. Many finite-element software packages are capable of solving the partial differential equations (PDEs) necessary to determine the surface flux and heat-transfer coefficient. The algorithms used to solve the inverse conduction problem, specifically for quenching, are shown in the literature (Ref 67, Ref 68, Ref 69, Ref 70, Ref 71, Ref 72). These PDEs can be solved using several numerical techniques, such as finite element or finite difference. The probe supplier can direct the user to the appropriate software to be used for their probe. An example of the heat-transfer coefficient determined for the ISO 9950 probe is shown in Fig. 21.

Cooling Curves. For the most part, the cooling curve of an oil or polymer depends on the bulk temperature of the oil and the agitation, or:

$$h \approx f(T_\infty, V)$$
(Eq 30)

Figure 22 shows the changes in cooling-curve behavior as a function of temperature on cold oils. For convenience, most oils are tested with no agitation. Agitation plays a key role in determining the heat-transfer coefficients but also can control the residual stress and distortion.

The cooling-curve behavior is also dependent on the condition of the oil. As oil is oxidized, there is a buildup of organic acids and soot. This contamination of the oil, measured by the total acid number, creates shellac on the surface of the part. This shellac changes the surface condition of the part and has the tendency of causing the vapor phase to become unstable, with boiling being initiated earlier. This causes faster heat transfer and increases the heat-transfer coefficient. Figure 23 shows the effect of oxidation on the cooling curves of oil.
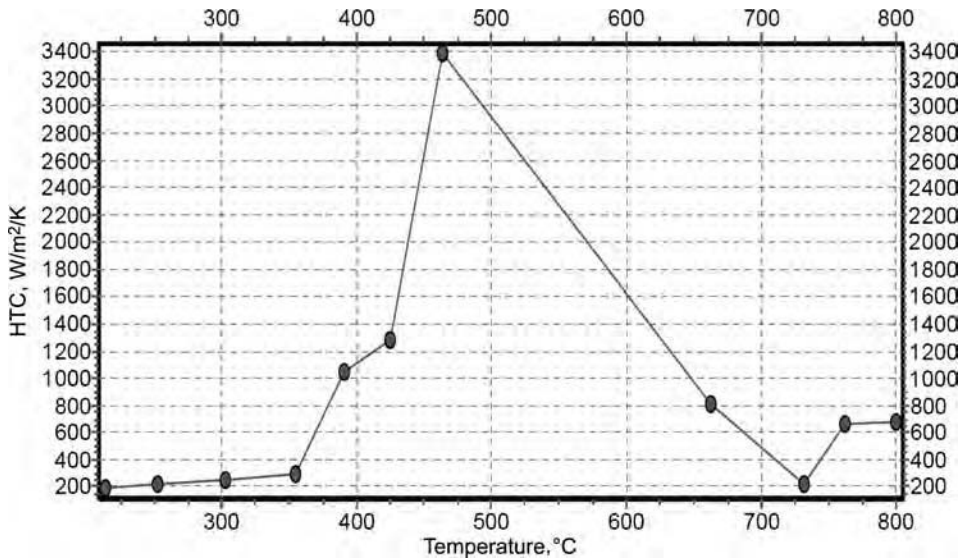
Fig. 21 Heat-transfer coefficient (HTC) determined by the inverse conduction method for a typical cold oil and the ISO 9950 probe
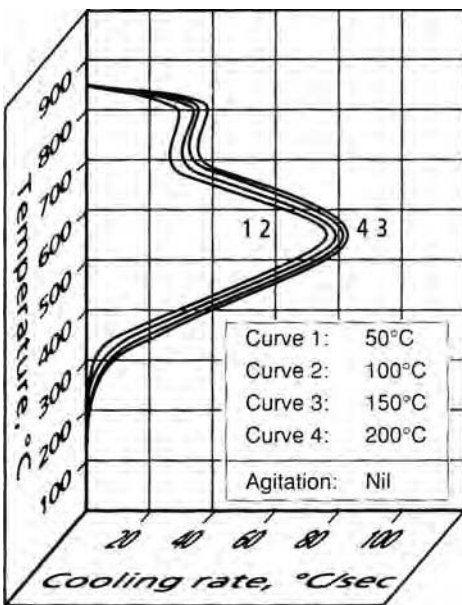
LIVE GRAPH
Click here to view

Fig. 22 Effect of temperature on the cooling-curve behavior of a cold oil

Fig. 23 Effect of oxidation on the cooling curve of a cold oil

LIVE GRAPH
Click here to view

LIVE GRAPH
Click here to view

When modeling or determining the heat transfer of oil or a quenchant, it is important to perform sensitivity studies to determine the effects of racking (the effect of parts located nearby or the effects of the rack itself), the effects of agitation, and the effects of the quenchant condition.

Determination of Heat-Transfer Coefficients via Computational Fluid Dynamics

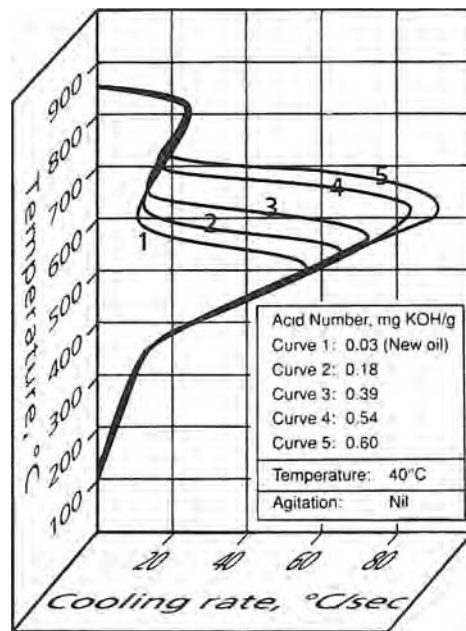Computational fluid dynamics (CFD) is a computational framework for simulating a wide array of situations that involve fluid flow, heat transfer, mass transfer, and species transport. Applications to heat treating operations include heating cycles in atmosphere and vacuum furnaces, carburizing, and quenching operations. For quenching processes, CFD can be used to develop improved quench system designs and to determine the heat flux rates and heat-transfer coefficients that will result when parts are quenched in a specific quench system.

A brief overview is provided of CFD methods, followed by a discussion of three topics that are important for simulation of quench processes: turbulence, near-wall treatment, and boiling heat transfer. The effectiveness of these methods is then discussed, along with the use of the resultant heat-transfer values in finite-element analysis simulations. An example of CFD as applied to a high-pressure gas quench is provided.

Overview of CFD Methods. A brief overview of CFD methods follows. A more complete treatment of this subject can be found in Ref 73, while advanced algorithms and solution techniques are discussed in Ref 74.

Flow, heat transfer, and species transport in fluids are governed by the Navier-Stokes equations, a set of PDEs that cannot be solved except for the simplest of scenarios (Ref 75, 76). The Navier-Stokes equations, named after Claude Navier and George Stokes, apply Newton's laws to fluid motion. CFD is a method of obtaining solutions to these equations for situations of practical interest through a series of approximations, mathematical techniques, and computer algorithms.

The Navier-Stokes equations are reduced to a solvable form by:

- Ignoring terms that are not important for the case of interest (e.g., compressibility)
- Selecting points of interest within the domain (nodes) where the field values will be calculated (i.e., applying a numerical grid to the domain)
- Developing linearized versions of the PDEs that are applied between nearby nodes (i.e., discretization)

Due to the nonlinear form of the Navier-Stokes equations, the linearized forms of the equations contain coefficients that are dependent on the current values of the field variables (velocity, temperature, etc.). These coefficients are treated as constants for the purpose of solving the set of linear equations. The coefficients are then updated, and the linear equations are solved again. For a properly constructed solution algorithm and a well-posed case, these iterative solutions will progress in an orderly fashion and will ultimately converge on the final solution.

Several methods have been used to define the relationship between the nodal values (discretization methods), including finite volume, finite element, and finite difference. Many commercial and research CFD codes use the finite-volume approach, while some codes use the finite-element method.

Turbulence. Turbulent flow is characterized by chaotic, stochastic variations in the flow that result when the kinetic energy in the flow field that leads to instabilities overwhelms the viscous forces that tend to damp out those instabilities. Extensive research has focused on understanding the nature of turbulence and developing accurate methods of representing turbulence in CFD models. This is of prime importance, because most flows of engineering interest are turbulent. More complete discussions of turbulence can be found elsewhere (Ref 77, 78).

Turbulent flow produces eddies of many different length scales that are superimposed on the bulk fluid motion. Directly simulating all of these eddy scales (direct numerical simulation of turbulence) requires an extremely fine grid that is not practical for problems of engineering interest. Direct numerical simulations are largely used for research purposes for the understanding of turbulence and the development of improved turbulence models for general-purpose CFD simulations.

Most practical turbulent flow methods directly simulate the bulk fluid motion and some of the larger eddy scales, while a "model" is used to account for the energy contained in the smaller-scale motions. An eddy viscosity approach is typically used, in which the fluid viscosity is augmented to account for the additional dissipative action of the modeled scales of motion. Turbulence models used for engineering applications include the k-ϵ model, the k-ω model, and the Reynolds stress model. Many variations exist for each of these turbulence models.

Wall Treatment. The region where the fluid flow is adjacent to the wall presents additional difficulty for the simulation process. Gradients of temperature, velocity, and other fields are often very steep near the wall. Turbulence, which is often isotropic in the bulk flow, becomes anisotropic near the wall and then rapidly dies out.

Historically, it was not possible to concentrate sufficient grid points near the wall to resolve these flow-field changes and sharp gradients. Instead, functions were developed that represented the overall effect of the boundary-layer region on the flow at the first near-wall node, which was typically placed outside the boundary layer. These "law of the wall" functions were based on theoretical and experimental data for special cases that were not always representative of the flow being simulated.

Extensive research has been devoted to developing improved wall functions that include the effect of pressure gradients, surface curvature, impinging flow, wall roughness, heat transfer, and so on. Many of the advanced wall functions require that additional computational nodes be placed within the boundary-layer region.

Boiling heat transfer is difficult to predict with CFD methods due to the presence of two phases in the fluid domain and the small scale of bubbles relative to the domain of interest. Detailed CFD studies have been performed in which the formations of individual bubbles are predicted, but these studies are of little practical interest to heat-transfer predictions for industrial processes.

Other approaches have been used to represent boiling heat transfer on a macroscopic scale. One approach is to divide the total wall heat flux into convective and boiling components. The convective component computed by standard CFD techniques may be modified to account for the presence of the bubbles.

Likewise, correlations used to predict the generation of vapor bubbles are affected by the fluid velocity near the part. Lahey and Drew (Ref 79) developed a four-field, two-fluid model that has been applied to a wide range of cases, with good success. Ho, Yeoh, and Tu (Ref 80) used population balance techniques to account for the presence of the bubbles on the flow field.

Application of CFD to Quench Systems and Determination of Heat-Transfer Coefficients. Gas quenching processes are dominated by convective heat transfer, and the heat-transfer coefficients predicted by CFD methods are expected to be very good, provided that best practices related to grid construction and wall treatment are observed. Similarly, intensive quenching operations, which exhibit very little boiling due to the high liquid flow rates, are dominated by convection and are expected to be well represented by current CFD techniques.

For liquid quench tanks, the critical form of heat transfer is boiling, including the effect of film boiling during the early stages of the quench process. While there is much current work in this area, the methods available in commercial CFD codes do not appear to be advanced enough to properly predict the heat-transfer coefficients for these processes. Given the rapid rate of change in the CFD field, however, it is likely that this capability will soon be available.

One advantage of determining the heat flux rates and heat-transfer coefficients via CFD is the amount of information that is provided. While a well-thermocoupled part can yield heat-transfer coefficients at a few dozen locations, a CFD simulation will provide flux rates and coefficients at every wall node in the grid, which will typically number in the thousands.

Application to Finite-Element Analyses. One application of CFD-derived heat-transfer coefficients is to the analysis of phase transformation, part distortion, and residual stress through finite-element analysis (FEA) software. Due to the different goals and analysis techniques, the grid element size for FEA will typically be larger than the corresponding grid for a well-resolved CFD simulation. In those cases, a relatively simple technique may be used to map the CFD results to the FEA boundary (e. g., nearest node). In cases where the grids are on the same length scale, or if the CFD cell size is larger than the FEA element size, it may be necessary to employ a more complex interpolation scheme to achieve smooth results on the FEA model boundary.

Due to changes in the fluid properties as the part surface changes temperature, it cannot be necessarily assumed that the heat-transfer coefficient remains constant at a given point on a part during the quench process. For best accuracy, a method is needed to represent the changing heat flux rates throughout the quench cycle (Ref 81).

Example of CFD Simulation of Gas Quenching Operation. A CFD simulation

was made of a high-pressure gas quench process applied to a production automotive ring gear. The load consisted of two stacks of ten fixtures each, with eight gears/fixture, for a total of 160 gears. An overall model of the quench cell was used to define the flow field in the vicinity of one of the gears on the top rack. A detailed model was then made for that gear.

The grid for the detailed model consisted of 13 million cells, with most of those cells concentrated in the region adjacent to the gear surface. The distance between the wall and the first near-wall node was 0.043 mm on average. A view of the grid detail near the wall is shown in Fig. 24 and 25.

The flow field for this model was simulated with FLUENT 6 using the k-ω SST model, which uses the fine grid near the wall to provide a better representation of the boundary-layer region. Heat-transfer coefficients for this gear are shown in Fig. 26. The average heat-transfer coefficients predicted from this simulation are in agreement with values derived from thermocouple tests. In those tests, only three thermocouples were used, such that much of the detail shown in the CFD results was lost.

Example of CFD Simulation of Oil-Quenched Automotive Pinions. A CFD simulation (Ref 82) was made of an oil-quenched
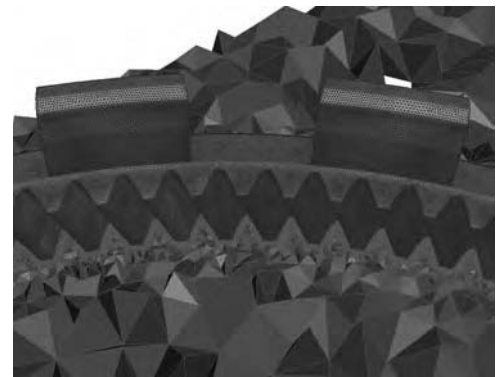

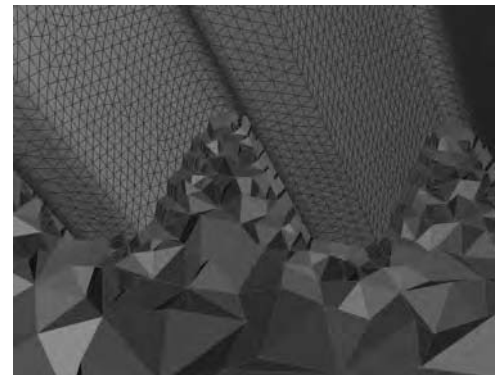**Fig. 24** Overall grid detail near the wall of a large gas-quenched ring gear

Fig. 25 Detail of grid close to the wall of a large gas-quenched ring gear shown in Fig. 24
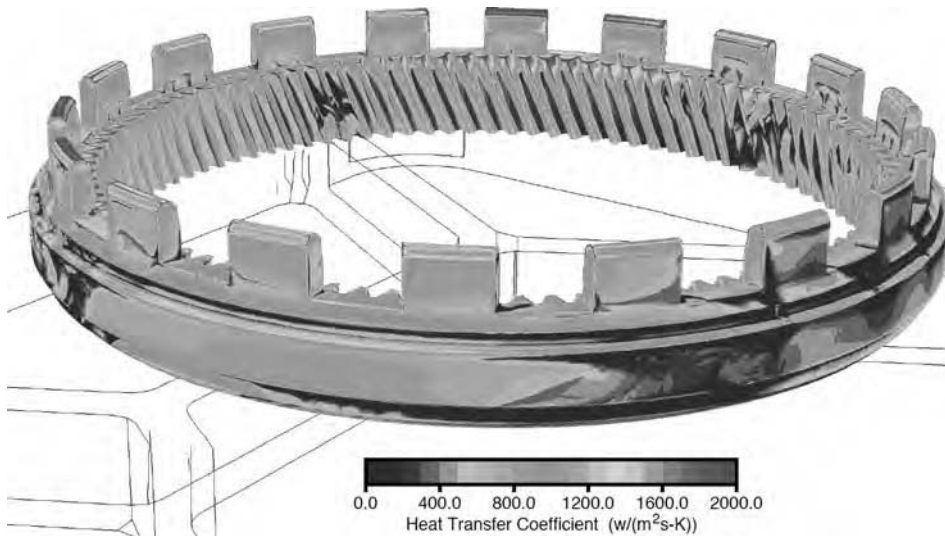
Fig. 26 Heat-transfer coefficients determined using computational fluid dynamics for a large ring gear that was gas quenched
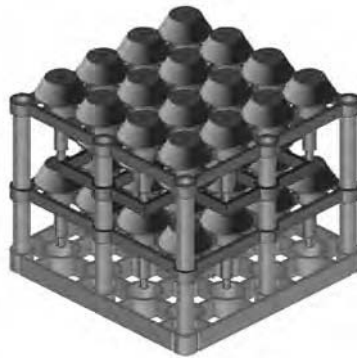

Fig. 27 Comparison of existing and proposed racking of automotive pinions

process applied to automotive pinion gears. In this instance, the customer was experiencing distortion of the pinion shafts, causing excessive rework and scrap. The customer was racking the parts pinion down, in a pyramid fashion. It was recommended to rack the parts in a new fixture, with the pinions spread out and racked shaft down. The load consisted of 40 pinion gears (Fig. 27) that were carburized in a pusher furnace, then quenched at 121 °C (250 °F).

Computer-aided design three-dimensional solid models of the quench tank and agitation system were obtained. The result was a very complicated model, with over 3000 entities. This model had to be simplified for proper analysis and to eliminate any geometry mismatches and any "leaks." Entities exterior to the tank were ignored, and only wetted surfaces were considered. The entire computational domain is shown in Fig. 28.

The parts were modeled as simple solids, without teeth. This drastically simplified the meshing and solution of the model. If the teeth

were included, the computation time would be increased. The primary goal was to examine the pinion shafts, because this was where most distortion was occurring.

The agitators were not explicitly modeled. Instead, the exit of the agitators and the inlet to the tank was modeled as a velocity inlet. On that face, an average speed was provided to yield the net mass flow rate supplied by the agitator. The inlet face of the agitator and outlet was represented by a pressure-outlet boundary condition.

The grid for the detailed model consisted of 7.3 million cells, with most of those cells concentrated in the region adjacent to the gear surface. A view of the grid detail is shown in Fig. 29. The resulting flow fields were determined and are shown in Fig. 30.

The oil flow rates around the pinions from static CFD analysis were exported in a spreadsheet that contained (x,y,z) positions of the centroids of the CFD cells surrounding the pinions and the magnitude of the oil velocity. The data from the spreadsheet file were imported and

mapped into the DANTE model to calculate the local heat-transfer coefficients for the pinion surface element faces. The relationship between oil flow rate and the heat-transfer coefficient is described using Eq 31 and is shown in Fig. 31:

$$h_c = h_{oil}{}^* v^n \qquad \text{(Eq 31)}$$

where h_c is the oil flow-rate-dependent heat-transfer coefficient, h_{oil} is the average oil heat-transfer coefficient, v is the oil flow rate, and the exponential n equals 0.466 in this study. The heat-transfer coefficient h_{oil} in Eq 31 is a function of part surface temperature, which also represents the vapor blanket, nucleate boiling, and convection phenomena during oil quenching. The heat-transfer coefficient h_{oil} was determined from inverse calculation from cooling curves of the martempering oil used in this application. The results of the FEA model (Ref 83) and distortion prediction were verified, using 60 heat treated loads of 40 parts, for a total of 2400 parts. The distortion of the shaft was measured and tabulated. The results are shown in Table 5.

Modeling tended to overpredict the distortion but was the proper magnitude and direction. Sources of error in this analysis were thought to be due to mismatch between the CFD and FEA meshes, with the CFD mesh being much coarser than the FEA mesh. A finer CFD mesh would have mapped better to the FEA mesh but would have greatly increased computational time. The correlation of $h_{oil}(v,T)$ only approximated real behavior. Because the vapour phase was not modeled directly, the stability of the vapor phase was not modeled accurately. The results of applying CFD as boundary conditions to an FEA model were shown to be sufficiently accurate to solve practical industrial distortion problems.

Determination of Heat-Transfer Coefficients by Inverse Conduction Calculation and Measurement of Parts

In a manner similar to determining heat-transfer coefficients from the cooling curves of instrumented probes of simple geometry, this method uses parts instrumented with multiple thermocouples and then uses solutions to the inverse conduction problem to determine the heat fluxes at the surface of the parts. Because of the complex geometry and large size of parts, a single thermocouple is inadequate to determine the heat flux at the surface of a part. Many thermocouples are needed and provide additional resolution. However, the solution of the problem requires even greater computation time as the number of thermocouples increases, because of the dramatically increased complexity.

Historically, this method is the prevalent method for determining heat-transfer coefficients. This method sounds straightforward; however, it requires extensive testing in the
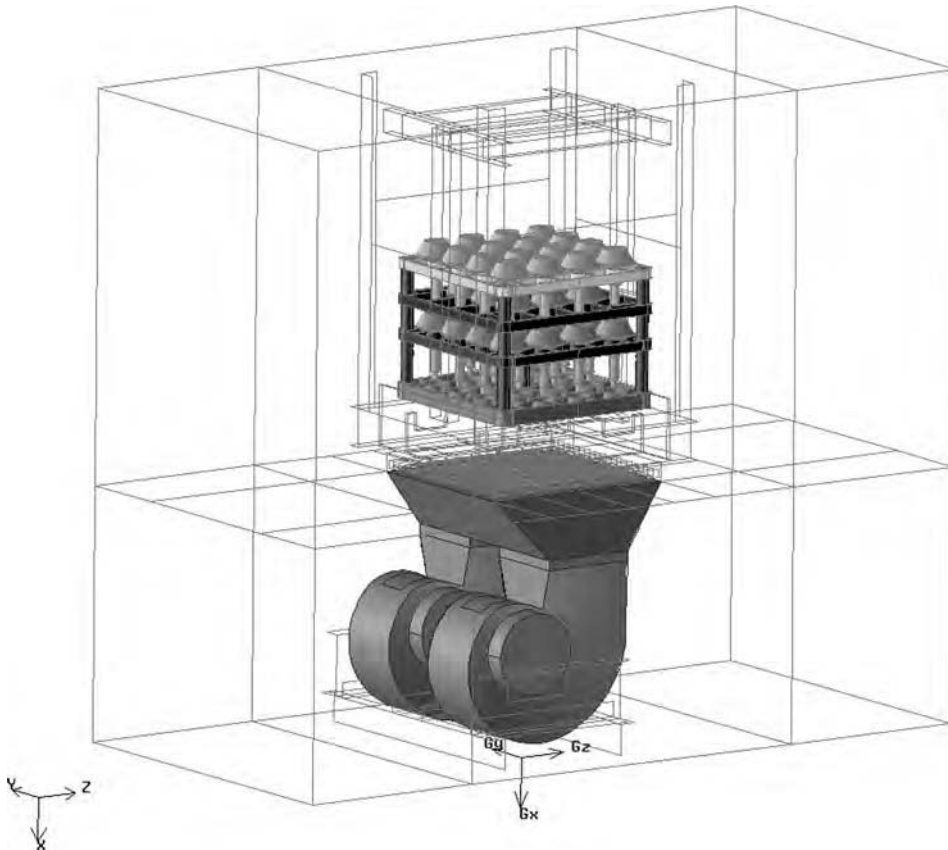
Fig. 28 Entire computational domain used for the computational fluid dynamics model of the quench tank and pinions
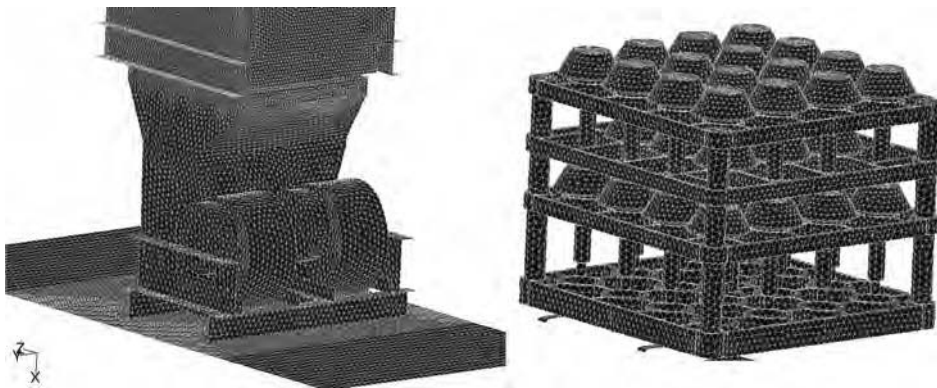

Fig. 29 Grid detail around the agitator system and parts

- What is the best way to determine the heat-transfer values from the center of each thermocouple zone to the surface boundary?
- How many thermocouples are necessary to determine the part thermal history and adequately calculate the surface-boundary conditions with necessary precision?
- Where should the thermocouples be located? These thermocouples can be located close to the surface or in the interior of the part. How many should be at the surface (within 12 mm), and how many are located inside the part?
- What is the best way to calculate the heat-transfer coefficients? Should an inverse method be used, and if so, which is the best method (Beck's second method, Ref 84; function decomposition method, Ref 85; or other method) to solve the inverse problem?
- What is the availability of the physical property data (specific heat, thermal diffusivity, phase transformation latent heats, etc.) as a function of temperature? Are the data well enough characterized to provide accurate results?
- Are the generated data generic enough to be applicable to entire classes of parts, or are they only applicable to a specific part, racked in a specific manner, in a specific quench tank?
- Will the number and placement of the thermocouples influence the conditions of flow within the quench tank and act as a "trip wire," causing early nucleate boiling initiation?

Many of these issues are practical ones, such as the availability of a limited number of data-acquisition channels or the difficulty of instrumenting a part. It may also be limited by the difficulty of loading the furnace with the thermocoupled part and moving it into the quench without losing data channels. Answers to other questions may have to rely on the judgment of the engineer regarding the flows in the quench tank and the behavior of the quenchant. If the quenchant flow is unknown or complex because of part shape or agitator or quench tank design, the number of thermocouples used may not adequately capture the thermal history of the part. In addition, the solution calculated may provide a set of results that satisfy the thermal data but is not the proper answer. Often, there is no "reality check" on the calculated data, and they are accepted without question. As a general rule, most thermocouples are located near the surface of the part, because these thermocouples are the most influenced by the local heat-transfer conditions. It will also provide better accuracy of match between the calculated and measured results.

The necessary constitutive equations governing the physical properties of an alloy are either missing or sparse. Microstructural effects, such as latent heat during transformation during quenching, are often missing from the constitutive equations. These effects and the various

shop environment to provide valid results. Excellent agreement is usually found between thermocouple measurements and the fitted inverse conduction solution. Planning the necessary experiments requires a number of decisions that can drastically impact the accuracy and precision of the resulting heat flux results, regardless of the quality of fit between the measured and calculated results. To plan and conduct such an experiment, the following questions must be asked:

- What is the best way to discretize or break up the surface of the part into small heat-transfer coefficient zones? That is, over how large or small of an area will the determined heat-transfer coefficient be valid? Too large an area, and the result is a bulk-average heat-transfer coefficient. Too small an area, and the dataset becomes too large and noisy and requires a large number of thermocouples and large computation time to determine the heat transfer over such a small area.
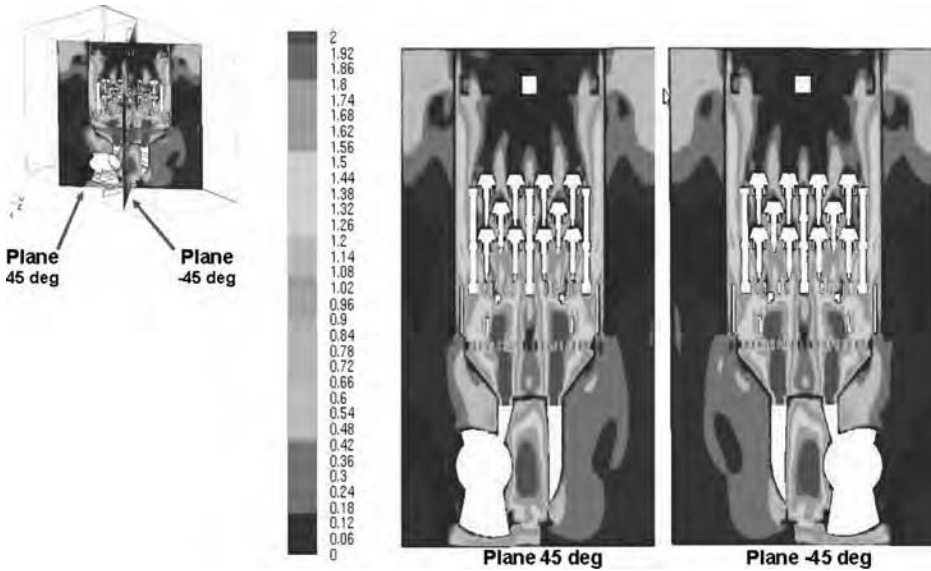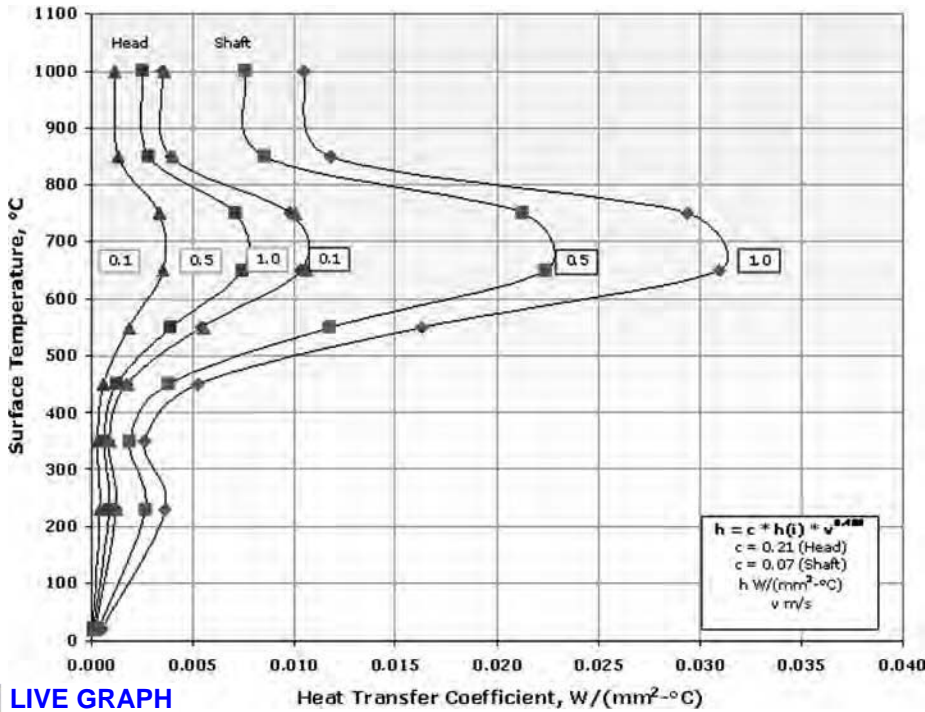
Fig. 30 Flow field determined for pinions. Shown are the diagonal cuts and the resulting flow velocities in meters per second.

Fig. 31 Heat-transfer coefficients as a function of velocity and wall temperature for typical martempering oil

Table 5   Predicted versus actual distortion of pinion gears modeled using computational fluid dynamics and finite-element analysis

Position	Predicted, mm	Actual, mm		
		Maximum	Average	Minimum
1	0.025	0.020	0.011	0.008
2	0.015	0.017	0.009	0.005

Source: Ref 83

phase mixtures can drastically impact the specific heat and thermal conductivity data.

Production variability can drastically influence the results. The location of parts (Fig. 32) and the position of adjacent parts can impact the resulting flows and heat transfer (Fig. 33). Because of the cost and difficulty of conducting experiments in a production environment, it is extremely difficult to determine a priori the sources of variability on the production floor.

Example of Direct Measurement. Ramakrishnan (Ref 86) and Bass et al. (Ref 87) calculated the heat-transfer coefficients as a function of surface location for a large turbine disk. The superalloy parts were thermocoupled using 13 type-K thermocouples sheathed in AISI 304 stainless steel (Fig. 34). For the oil quench experiment, the single disk was placed on a special heat treatment tray, loaded in a rotary furnace, and heated until the temperature in the turbine disk stabilized at 1150 °C. When the temperature in the part stabilized, the part was removed from the furnace, placed on a vertical elevator, and quenched into a cold-type oil at 25 °C. The oil was kept agitated and maintained at temperature. The disk was quenched horizontally into the quench tank. No details about the amount or direction of the flows in the quench tank were provided. Data acquisition rate was at 20 Hz, and the data were recorded until the turbine disk maximum temperature was below 100 °C.

An inversed heat conduction code, QUENCH2D, developed at Michigan State University, was used to process the data. This code was developed to solve nonlinear, two-dimensional inverse heat-transfer problems. After examining the data, the disk was divided into four distinct subsurfaces, that is, inner diameter, top, outer diameter, and the bottom surface. Using the temperature data from all 13 thermocouples, the heat fluxes were determined for each surface. The resulting heat-transfer coefficients as a function of surface temperature are shown in Fig. 35. The data showed that there was a substantial difference in the heat transfer on the top and bottom surfaces, likely caused by either the adjacent heat treating tray or the possible trapping of the vapor phase beneath the part. These results are similar to those of Segerberg and Bodin (Ref 37) when they quenched a small plate specimen instrumented with thermocouples (Fig. 36).

Determination of Heat-Transfer Coefficients for Die Blocks. In this example, Thuvander (Ref 88) thermocoupled small tool steel die blocks to determine heat-transfer coefficients to be used in FEA to estimate distortion. Three different dies of different configuration were used (Fig. 37). Because these tool steel dies were quenched in a fluidized bed, it was assumed that the heat-transfer coefficients would be uniform across each surface. It was also assumed that the heat-transfer coefficients would have no temperature dependence. Inverse conduction calculation was done using ABAQUS. While the author thought that the results would be improved by incorporating heat-transfer coefficients that varied with temperature, excellent correspondence with measured and calculated results was observed (Fig. 38).

The calculations to determine distortion were performed in three steps. The first step was to calculate the thermal history of the part. Second, the phase-transformation history was
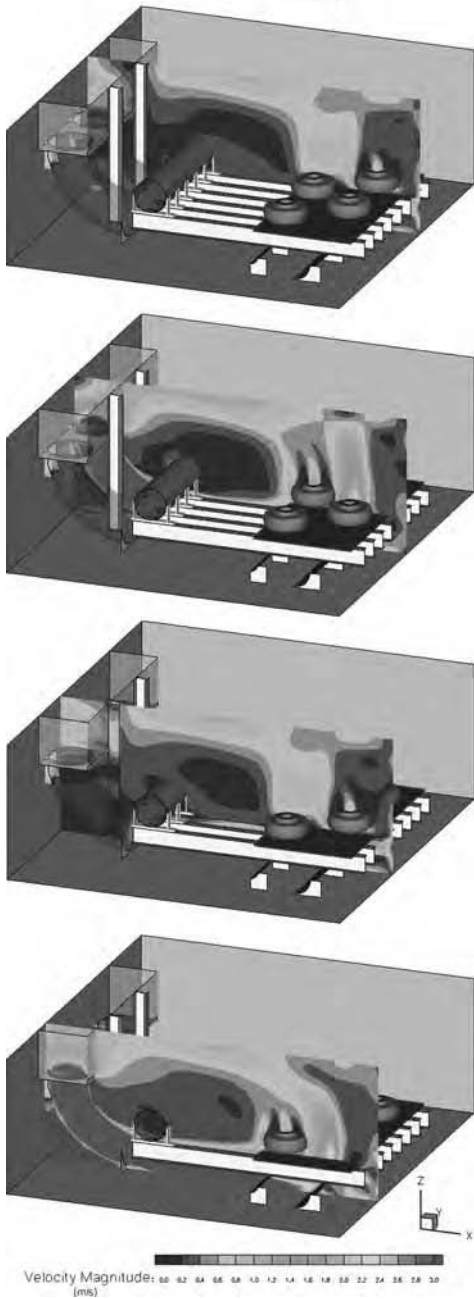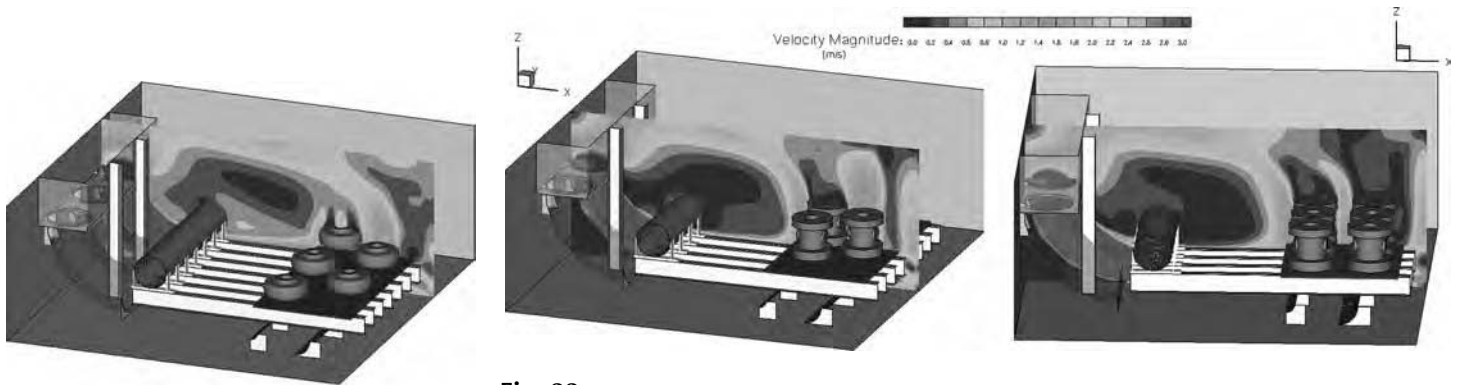
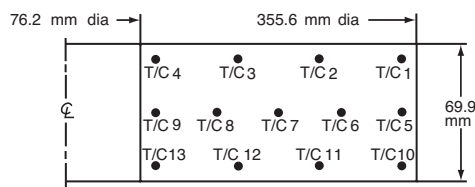**Fig. 33** Load density and number of parts affecting quenchant flow around parts in a production quench tank

Fig. 34 Instrumented turbine disk showing locations of thermocouples (T/C) used by Ramakrishnan. Source: Ref 86
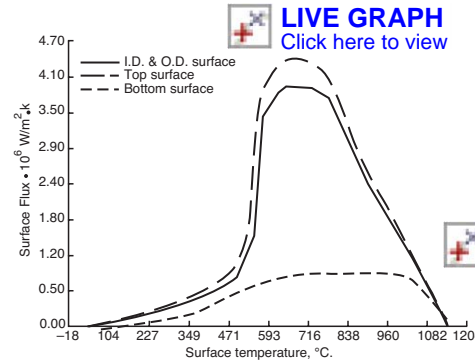
Fig. 35 Resultant heat-transfer coefficients showing substantial difference in upper and lower surfaces of a turbine disk quenching horizontally into an oil quench. Part was heated to 1150 °C and quenched into oil at 25 °C. No information was provided regarding the oil type, speed of the oil, or the flow rates and agitation in the quench tank. ID, inside diameter; OD, outside diameter
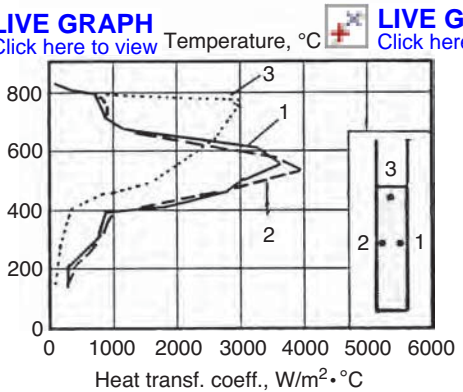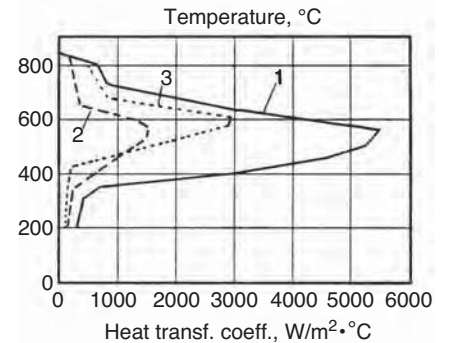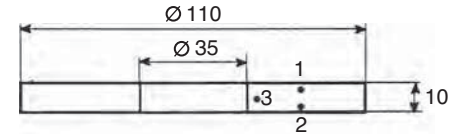
Fig. 36 Heat transfer showing differences in the top and bottom surfaces of a thin plate quenched horizontally in oil. Source: Ref 37

Fig. 32 Change in flow patterns and resultant heat-transfer coefficients as a function of position in a production quench tank

determined. Lastly, the mechanical property response to the temperature changes and phase transformations was performed. Only bainite and martensite transformations were included in this simulation. The mechanical properties were based on a separate finite-element method calculation that assumed that the properties would be a linear sum of the contributions from each phase present.

The calculated distortion of the dies is shown in Fig. 39. The numerical calculations gave close agreement between the calculated and measured distortion. The calculated distortion was of the same direction and magnitude as the actual parts. The ring-shaped die exhibited conical distortion on the inner surface of the bore, and the upper surface showed a concave surface, with the outer edge of the ring higher than the center. Some

out-of-roundness was exhibited. The box-shaped dies showed excellent agreement between experiment and calculations. The convex distortion of the bottom surface was very well reproduced by the simulations.

Conclusions

In the previous discussion, a review of the analytical methods for determining heat-transfer coefficients was discussed. Several methods

Fig. 37 Finite meshes of dies used. (a) Ring-shaped die. (b) Small box-shaped die. (c) Large box-shaped die

Fig. 38 Measured and calculated temperature history during quenching of small box-shaped dies in a fluidized bed

Fig. 39 Calculation of distortion in a ring-shaped die (top) and a small box-shaped die (bottom) after quenching in a fluidized bed. Distortion results are enlarged 100 times for the ring-shaped die and 50 times for the small box-shaped die.

of determining heat-transfer coefficients in an industrial setting were reviewed with examples. Each method has its advantages and drawbacks. Each method relies on certain assumptions that should be verified by experiment.

REFERENCES

1. *Heat Treating,* Vol 2, 8th ed., American Society for Metals, 1968 p 16
2. M.A. Grossmann, M. Asimov, and S.F. Urban, Hardenability, Its Relation to Quenching, and Some Quantitative Data, *Hardenability of Alloy Steels,* American Society for Metals, 1939, p 124
3. P.G. Winchell and M. Cohen, The Strength of Martensite, *Trans. ASM,* Vol 55, 1962, p 347–361
4. D.J. Carney, Another Look at Quenchants, Cooling Rates and Hardenability, *Trans. ASM,* Vol 46, 1951, p 480–496
5. M.A. Grossmann and E.C. Bain, *Principles of Heat Treatment,* 5th ed., American Society for Metals, 1964
6. "Improved Quenching of Steel by Agitation," U.S. Steel, 1954
7. G.E. Totten and K.S. Lally, *Heat Treat.,* Sept 1992, p 12
8. K.S. Lally and G.E. Totten, *Heat Treat.,* Oct 1992, p 28
9. D.R. Garwood, J.D. Lucas, R.A. Wallis, and J. Ward, *J. Mater. Eng. Perform.,* Vol 1, 1992, p 781
10. R.A. Wallis, D.R. Garwood, and J. Ward, The Use of Modeling Techniques to Improve the Quenching of Components, *Heat Treating: Equipment and Processes, 1994 Conference Proceedings,* G.E. Totten and R.A. Wallis, Ed., ASM International, 1994, p 105
11. K. Illgner, *Hart.-Tech. Mitt.,* Vol 42, 1987, p 113
12. R. Kern, *Heat Treat.,* April 1985, p 38
13. R. Kern, *Heat Treat.,* Feb 1971, p 1
14. R. Kern, *Heat Treat.,* March 1985, p 41
15. A. Terrier, *Trait. Therm.,* Vol 4, 1976, p 3
16. R.T. Von Bergen, The Effects of Quenchant Media Selection and Control on the

Distortion of Engineered Steel Parts, *Quenching and Distortion Control,* G.E. Totten, Ed., ASM International, 1992, p 275

17. E. Guan, J. Zhang, and H. Shen, *Heat Treat. Met. (China),* Vol 10, 1990, p 20

18. S. Segerberg, *Heat Treat.,* May 1988, p 26

19. C. Oldfather, Evaluating Quench Tanks Using Thermocouples in Parts, *Heat Treating: Equipment and Processes, 1994 Conference Proceedings,* G.E. Totten and R. A. Wallis, Ed., ASM International, 1994, p 155

20. N. Bogh, Quench Tank Design Using Flow Modeling, *Heat Treating: Equipment and Processes, 1994 Conference Proceedings,* G.E. Totten and R.A. Wallis, Ed., ASM International, 1994, p 51

21. H.M. Tensi, A. Stich, and G.E. Totten, *Proc. International Heat Treating Conference—Equipment and Process,* April 12–20, 1994 (Schaumberg, IL)

22. E.C. Guyer and D.L. Brownell, *Handbook of Applied Thermal Design,* McGraw-Hill, St. Louis, 1985

23. L.A. Bromley, Heat Transfer in Stable Film Boiling, *Chem. Eng. Prog.,* Vol 46, 1950, p 221

24. L.A. Bromley, Heat Transfer in Forced Convection Film Boiling, *Ind. Eng. Chem.,* Vol 46, 1953, p 221

25. J.W. Westwater and B.P. Green, Effect of Diameter of Horizontal Tubes on Film Boiling Heat Transfer, *Chem. Eng. Prog.,* Vol 58, 1962, p 67

26. Y.Y. Hsu and J.W. Westwater, An Approximate Theory for Film Boiling on Vertical Surfaces, *Chem. Eng. Prog. Symp. Series,* Vol 56, AIChE Heat Transfer Conference (Storrs, CT), 1959, p 15

27. S.G. Bankoff, Discussion of Approximate Theory for Film Boiling on Vertical Surfaces, *Chem. Eng. Prog. Symp. Series,* Vol 55, AIChE Heat Transfer Conference (Storrs, CT), 1959, p 22

28. L.S. Tong, *Boiling, Heat Transfer and Two Phase Flow,* Wiley, New York, 1966

29. W.M. Rohsenow and H.Y. Choi, *Heat, Mass and Momentum Transfer,* Prentice Hall, Englewood Cliffs, NJ, 1961

30. W.M. Rohsenow, A Method for Correlating Heat Transfer Data for Surface Boiling Liquids, *Trans. ASME,* Vol 3, 1952, p 969

31. K. Stephon and M. Abdelsalam, Heat Transfer Correlations for Natural Convection Boiling, *Int. J. Heat Mass Transf.,* Vol 25, 1980, p 73

32. M.N. Ozisik, *Heat Transfer—A Basic Approach,* McGraw-Hill, New York, 1985

33. F. Kreith, *Principles of Heat Transfer,* Intext Educational Publishers, New York, 1976

34. P.J. Berenson, Experiments on Pool-Boiling Heat Transfer, *Int. J. Heat Mass Transf.,* Vol 5, 1962, p 985–999

35. W.H. McAdams, Heat Transfer from Single Horizontal Wires to Boiling Water, *Chem. Eng. Prog. A,* Vol 4, 1948, p 639

36. W.H. McAdams, *Heat Transmission,* 3rd ed., McGraw-Hill, New York, 1954

37. S. Segerberg and J. Bodin, Variation in the Heat Transfer Coefficient Around Components of Different Shapes During Quenching, *Proc. First Int. Conf. on Quenching and Control of Distortion,* Sept 22–25, 1992 (Chicago, IL), ASM International, 1992, p 165

38. R. Maass and R. Jeschar, *Gas Warme Int.,* Vol 38, 1989, p 142–150

39. N. Zuber, On the Stability of Boiling Heat Transfer, *J. Heat Transf.,* Vol 80C, 1958, p 711

40. N. Zuber and M. Tribus, "Further Remarks on the Stability of Boiling Heat Transfer," UCLA Dept. Eng. Report 58-5, 1958

41. J.H. Lienhard and V.K. Dhir, Hydrodynamic Prediction of Peak Pool-Boiling Heat Fluxes from Finite Bodies, *J. Heat Transf.,* Vol 95C, 1973, p 152

42. A. Rose, *Arch. Eisenhuttenwes.,* Vol 13, 1940, p 345–354

43. C. Bates, G. Totten, and R. Brennen, *Heat Treating,* Vol 4, *ASM Handbook,* ASM International, 1991, p 91

44. N. Kobasko, *Metalloved. Term. Obrab. Met.,* Vol 3, 1968, p 2–6

45. W.M. McAdams, W.E. Kennel, C.S. Minden, R. Carl, M. Picarnell, and J.E. Drew, Heat Transfer at High Rates to Water with Surface Bolling, *Ind. Eng. Chem.,* Vol 41, 1945

46. N. Zuber, M. Tribus, and J. Westwater, The Hydrodynamic Crisis in Pool Boiling of Saturated and Subcooled Liquids, *Proceeding of the International Conference on the Developments in Heat Transfer,* ASME, New York, 1962, p 230–236

47. V. Paschkis and G. Stoltz, *Iron Age,* Vol 22, 1956, p 95–97

48. M. Bamberger and B. Prinz, *Mater. Sci. Technol.,* Vol 2, 1986, p 410–415

49. C.E. Bates and G.E. Totten, Application of Quench Factor Analysis to Predict Hardness Under Laboratory and Production Conditions, *Proc. First Int. Conti Quenching and Control of Distortion,* Sept 22–25, 1992 (Chicago, IL)

50. G.E. Totten, C.E. Bates, and N.A. Clinton, *Handbook of Quenchants and Quenching Technology,* ASM International, 1993

51. D.S. MacKenzie, G.E. Totten, and G.M. Webster, Quenching Fundamentals, *Proc. Second International Conf. on Quenching and Control of Distortion,* Nov 4–7, 1996 (Cleveland, OH)

52. "Heat Treating Oil," JIS K 2242, 1980

53. R. Price and A. Fletcher, *Met. Technol.,* Vol 5, 1980, p 201–203

54. R. Allen, A. Fletcher, and A. Mills, *Steel Res.,* Vol 60, 1989, p 552–530

55. B. Liscic, *Hart-Tech. Mitt.,* Vol 33, 1978, p 179–191

56. M. Tajima, T. Maki, and K. Katayama, *Nippon Kikai Gakkai Ronbunshu (B-ken),* Vol 54, 1988, p 3491–3496

57. "Laboratory Test for Assessing the Cooling Curve Characteristics of Industrial Quenching Media," ISO 9950

58. R.W. Hines and E.R. Mueller, *Met. Prog.,* Vol 122, 1982, p 33–39

59. S. Ma, "Characterization of the Performance of Mineral Oil Based Quenchants Using the CHTE Quench Probe System," M.S. thesis, Worcester Polytechnic Institute, June 27, 2002

60. "Standard Test Method for Determination of Cooling Characteristics of Aqueous Polymer Quenchants by Cooling Curve Analysis with Agitation (Tensi Method)," D 6482-06, ASTM

61. "Standard Test Method for Determination of Cooling Characteristics of Quenchants by Cooling Curve Analysis with Agitation (Drayton Unit)," D 6549-06, ASTM

62. D.A. Guisbert and L. Jarvis, Influence of Test Conditions on the Cooling Curve Response of Polymer Quenchants (Tensi Agitation Device), *Proc. 22nd Heat Treat Soc. Conference,* Sept 15–17, 2003 (Indianapolis, IN), p 218

63. B. Liscic and T. Filetin, *J. Heat Treat.,* Vol 5 (No. 2), 1988, p 115–124

64. J. Nanigian, Thermocouples: In-Wall, Wall Surface and In-Stream, *Proc. 18th Heat Treat Soc. Conference,* R. Wallis and H. Walton, Ed., Oct 12–15, 1998 (Indianapolis, IN), p 79

65. J.V. Beck and K.J. Arnold, *Parameter Estimation in Engineering and Science,* Wiley, New York, 1977

66. J.V. Beck, B. Litouhi, and C.R. St. Clair, Efficient Sequential Solution of the Nonlinear Inverse Heat Conduction Problem, *Numer. Heat Transf.,* Vol 5, 1982, p 275–286

67. B. Hernadez-Morales, J.K. Brimacombe, E. B. Hawbolt, and S.M. Gupta, Determination of Quench Heat-Transfer Coefficients Using Inverse Techniques, *Proc. First Int. Conf. on Quenching and Control of Distortion,* Sept 22–25, 1992 (Chicago, IL), ASM International, 1992, p 165

68. G. Sarmiento, A. Gaston, and J. Vega, Inverse Heat Conduction Coupled with Phase Transformation Problems in Heat Treating Processes, *Computational Mechanics,* S. Idelsohn, E. Onate, and E. Dvorkin, Ed., CCIMNE, Barcelona, Spain, 1998

69. K.J. Dowding, Multi-Dimensional Analysis of Quenching: Comparison of Techniques, *Proc. 18th Heat Treat Soc. Conference,* R. Wallis and H. Walton, Ed., Oct 12–15, 1998 (Indianapolis, IN), p 525

70. X. Chen, L. Meekisho, and G.E. Totten, Computer-Aided Analysis of the Quenching Probe Test, *Proc. 18th Heat Treat Soc. Conference,* R. Wallis and H. Walton, Ed., Oct 12–15, 1998 (Indianapolis, IN), p 545

71. T. Inoue, D.-Y. Ju, and K. Arimoto, Metallo-Thermo-Mechanical Simulation of

Quenching Process—Theory and Implementation of Computer Code "HEARTS," *Proc. First Int. Conf. on Quenching and Control of Distortion,* Sept 22–25, 1992 (Chicago, IL), ASM International, 1992, p 205

72. R. Fabrik and J. Kliber, Inverse Analysis Calculation of Heat Transfer Coefficient for FEM Simulation of Rails Hardening, *Acta Metall. Slovaca,* Vol 11 (No. 3), 2005, p 341–350

73. S.V. Patankar, *Numerical Heat Transfer and Fluid Flow,* Hemisphere Publishing Corporation, 1980

74. M. Peric and J.H. Ferziger, *Computational Methods for Fluid Dynamics,* Springer-Verlag, 1997

75. B.R. Munson, D.F. Young, T.H. Okiishi, and W.W. Huebsch, *Fundamentals of Fluid Mechanics,* Wiley, 2009

76. R.W. Fox and A.T. McDonald, *Introduction to Fluid Mechanics,* John Wiley & Sons, Inc., 1998

77. J.O. Hinze, *Turbulence,* McGraw-Hill Book Company, 1987

78. D.C. Wilcox, "Turbulence Modeling for CFD," DCW Industries, 1993

79. R.T. Lahey, Jr. and D.A. Drew, The Analysis of Two-Phase Flow and Heat Transfer Using a Multidimensional, Four Field, Two-Fluid Model, *Nucl. Eng. Des.,* Vol 204, 2001, p 29–44

80. M.K.M. Ho, G.H. Yeoh, and J.Y. Tu, Population Balance Models for Subcooled Boiling Flows, *Int. J. Numer. Meth. Heat Fluid Flow,* Vol 18 (No. 2), 2008, p 160–172

81. A.L. Banka, J.D. Franklin, Z. Li, B.L. Ferguson, and M. Aronov, Applying CFD to Characterize Gear Response during Intensive Quenching Process, *Proceedings of the 24th ASM Heat Treating Society Conference,* Sept 17–19, 2007

82. A. Kumar, H. Metwally, and D.S. MacKenzie, "Evaluation of Flow Uniformity Around Automotive Pinion Gears During Quenching," Fifth International Conference on Quenching and Control of Distortion, European Conference on Heat Treatment 2007, April 25–27, 2007 (Berlin), International Federation for Heat Treatment and Surface Engineering (IFHTSE)

83. D.S. MacKenzie, Z. Li, and B.L. Ferguson, "Effect of Quenchant Flow on the Distortion of Carburized Automotive Pinion Gears," Fifth International Conference on Quenching and Control of Distortion, European Conference on Heat Treatment 2007, April 25–27, 2007 (Berlin), International Federation for Heat Treatment and Surface Engineering (IFHTSE)

84. J.V. Beck, Nonlinear Estimation Applied to the Nonlinear Inverse Heat Conduction Problem, *Int. J. Heat Mass Transf.,* Vol 13, 1970, p 703–716

85. J. Frankel and M. Keyhani, A New Approach for Solving the Inverse Heat Conduction Problems, *J. Heat Transf.,* Vol 119 (No. 4), 1997, p 673–683

86. R.I. Ramakrishnan, Quench Analysis of Aerospace Components Using FEM, *Proc. First Int. Conf. on Quenching and Control of Distortion,* Sept 22–25, 1992 (Chicago, IL), ASM International, 1992, p 235

87. R. Bass, D. Leonard, M. Allen, J. Bennett, M. Cross, J. Morral, and K. Brown, Heat Transfer of Turbine Disks in a Liquid Quench: Part III—Experimental Results for a Disk with Bore, *Proc. 18th Heat Treat Soc. Conference,* R. Wallis and H. Walton, Ed., Oct 12–15, 1998 (Indianapolis, IN), p 552

88. A. Thuvander, Calculation of Distortion of Tool Steel Dies During Hardening, *Proc. Second International Conf. on Quenching and Control of Distortion,* Nov 4–7, 1996 (Cleveland, OH), p 297

Interface Effects for Deformation Processes

M. Krzyzanowski, The University of Sheffield, United Kingdom
J.H. Beynon, Swinburne University of Technology, Australia

KNOWLEDGE OF FRICTION AND HEAT TRANSFER is vital to the understanding and operation of metal forming. Modeling and numerical simulations are now extensively used for optimization of the deformation during processing of metallic materials. This article examines traditional deformation processes and gives consideration to the effects introduced by scale factors when microforming.

Whether it is modeling of turning or forging, deep drawing or rolling, in cold or hot conditions, realistic and precisely determined boundary conditions are essential for prediction accuracy in order to decrease costs and time-to-market. The contact conditions in metal-forming processes are, to a certain extent, specific to the particular operation and certainly very diverse. The tool-metal contact often involves large sliding lengths on which high pressures, sliding speeds, and temperatures may be maintained together with plastic deformation. This plastic deformation, coupled with lubrication, failure of oxide layers, creation of nascent, very reactive metallic surfaces, and the constant renewal of one of the partners of the contact, the workpiece, is typical for metal forming. Contact pressures can range from 1 to a few MPa in sheet-forming processes to a few GPa in hard metal rolling or wire drawing. Speeds vary between a few μm/s (superplastic processing) to tens of m/s (high-speed drawing, turning, or thin-strip rolling). Surface temperatures are in the range between 0 and 300 °C in cold forming, due to self-heating by plastic deformation, and up to 500, 1300, or even 2000 °C in hot forming, depending on whether aluminum alloys, steels, or refractory metals are considered (Ref 1). The effect of these severe contact conditions is emphasized when there are high values of the aspect ratio of contact surface area to deformed volume (e.g., contact length/characteristic thickness in the rolling of flat products). As a consequence, the usual friction test methods, which are in great favor in bearing and motor oil industries, are often inappropriate for the metal-forming industries. Instead, test methods adapted to each class of forming operation are used, such as the ring compression test (forging) (Ref 2), the plane-drawing test or the drawing-under-bending test (drawing) (Ref 3), and the plane-strain compression test (flat rolling) (Ref 4).

Process Parameters

There are many parameters and variables affecting surface interactions (Ref 5). Process parameters include the temperature, speed, reduction, stiffness, and dynamic response of the equipment. In rolling, for instance, among the parameters that influence the interactions are the work roll and the back-up roll dimensions; their hardness, magnitude, and direction of the surface roughness; cooling systems; lubricant delivery; and the location of nozzles. The mechanical properties of the rolls and the workpiece, including their resistance to deformation, all contribute here. The contributions of surface parameters, such as the chemical reactivity, the tendency to adsorb molecules from the environment, the adsorption of water vapor and oxygen, as well as surface energy, must be taken into consideration. The nature of oxide scale formation and failure during deformation, the chemical composition of the scale and underlying metal, and the adhesion between the scale and the metal surfaces in direct contact must be taken into account. Lubrication also significantly affects the interface interactions. The chemical composition of the lubricant, the additives and their concentration in the base oil, and the molecular chain length, density, viscosity, and its dependence on both temperature and pressure should be precisely described. If emulsions are used, the composition, the emulsifier, and the droplet dimensions are important.

Boundary Conditions

Despite all this complexity in the behavior at the interface, it is common to represent friction and heat transfer as simple coefficients. This is done because only a limited accuracy of prediction is needed, or the calculation is not overly sensitive to friction and heat transfer (e.g., because of a relatively small contact area or contact time), or because the details of the interface are not well understood, so more detailed information is simply not available. However, there are many circumstances where the success of a mathematical model depends on the appropriate formulation of the boundary conditions, which could be as sophisticated as the model itself. If the boundary conditions are expressed simply in terms of coefficients of friction and heat transfer, the inclusion of all of the mentioned complexities into a single mathematical model describing dependence of these two parameters is highly impractical. Instead, relatively simple formulae for friction and heat transfer are used for general applications. Of course, reasonable choices are necessary to achieve desirable precision; they should take into consideration the most important dependencies that affect the tribological system. The most important process parameters are the surface temperature, the relative velocity between the tool and the workpiece, and the amount of plastic deformation. The attributes of the deformed material are normally limited to the model that describes its resistance to deformation. For example, a true stress/true strain relationship is used in cold rolling, while in hot rolling, the true stress is related to the strain, strain rate, temperature, and, sometimes, to a metallurgical parameter. This parameter can be the activation energy for plastic deformation or the carbon equivalent in steels. The parameters of the contact surfaces can include the strip and roll roughness and the thickness of the oxide scale. When lubricants are used, the viscosity and its dependence on the temperature and the pressure are taken into consideration. The concentration of the oil in the water and the droplet dimensions are considered when emulsions are applied, as in the cold rolling of steel strips.

The following sections consider friction and heat transfer in metal-forming operations, providing guidance on the choice of simple coefficients while also providing insight into how

more sophisticated formulae for friction and heat transfer can be developed where more accurate representation of actual interface conditions is needed.

Friction Coefficient

In 1997, Roberts commented that of all the variables associated with rolling, none is more important than friction in the roll bite (Ref 6). Because the trend in modern strip rolling is to produce thinner strips of higher-strength metals, the control of friction in the roll bite is vitally important (Ref 7). The coefficient of friction arises from the fundamental definition of the Coulomb-Amonton coefficient, $\mu = \tau/p$, as the ratio of the interfacial shear stress, τ, to the normal pressure, p. There is a view that the coefficient of friction may not be the best description of interfacial phenomena between the roll and the roll metal (Ref 8). For example, in the flat rolling process, the normal pressure, p, may increase significantly beyond the material flow strength. The interfacial shear stress, τ, may also increase, but it cannot rise above the metal shear yield strength. This problem is overcome by the use of the Tresca friction factor instead. The friction factor, m, is defined as the ratio of the interfacial shear stress to the metal flow strength, k, in pure shear, $m = \tau/k$. Nevertheless, the coefficient of friction is widely used and understood by engineers in the metal-forming and the flat-rolling industry and is also often used in mathematical modeling of forming operations. Symbols used in this article are found in Table 1.

Empirical Formulae for Coefficient of Friction. There are different formulae for the friction coefficient in hot rolling proposed at different times by a variety of authors. Roberts gives an increasing relationship between the coefficient of friction and the temperature in °F, T_F (Ref 9):

$$\mu = 2.7 \times 10^{-4} T_F - 0.08 \qquad \text{(Eq 1)}$$

which can be rewritten for temperature in °C as:

$$\mu = 4.86 \times 10^{-4} T - 0.07136 \qquad \text{(Eq 2)}$$

Roberts combined the data from experimental 2.1 and 3.4 m (84 and 132 in.) two-high hot strip mills obtained for well-descaled strips. Geleji's formula indicates the opposite trend (Ref 10):

$$\mu = 1.05 - 0.0005T - 0.056v \qquad \text{(Eq 3)}$$

where T is the temperature in °C, and v is the rolling velocity in m/s. The relationship was obtained for steel rolls by applying the inverse method matching the measured and calculated roll forces. For doubled poured and cast rolls, the formula for the friction coefficient is slightly different:

$$\mu = 0.94 - 0.0005T - 0.056v \qquad \text{(Eq 4)}$$

and is changed again for the ground steel rolls:

$$\mu = 0.82 - 0.0005T - 0.056v \qquad \text{(Eq 5)}$$

These relationships, indicating decreasing coefficient of friction with increasing temperature and rolling speed, confirm the experimental results obtained elsewhere (Ref 11):

$$\mu = 0.84 - 0.0004T \qquad \text{(Eq 6)}$$

Equation 6 was obtained for rolling steel at temperatures above 700 °C. A comparison of the friction coefficients obtained using Eq 1 to 6 indicates large differences for the different rolling temperatures and thus may not be very reliable. In other words, typical of empirical formulae, their precision applies only to the specific circumstances of their measurements; different conditions will lead to variations. This is commonly the case when attempting to transfer results obtained in a laboratory to industrial operations. The surfaces of tools and rolls in laboratories are often not in as good condition as those in industry, because laboratories often use very different workpieces and lubrication conditions, resulting in variable surface roughness and a wide range of surface layers—often a mix of oxide, lubricant traces, and debris—unevenly distributed over the tool or roll.

Oxide Scale. In an attempt to provide a more physically-based determination of friction, one should consider the evolution of secondary oxide scale and its failure during hot rolling and interpass cooling with respect to its thickness, composition, ductile/brittle behavior, and thermal properties, all of which play a significant role in the tribological behavior. In 1984, Felder characterized the oxide scale behavior during hot rolling as highly influenced by the temperature (Ref 12). He defined the ratio H between the scale thickness, δ_{ox}, and the scale thickness thermally affected by the contact with the tool, δ_{oxt}, as:

$$H = \frac{\delta_{ox}}{\delta_{oxt}} = \delta_{ox} \times (6a_c \Delta t)^{-0.5} \qquad \text{(Eq 7)}$$

where a_c is the thermal diffusivity of the oxide scale, and Δt is the contact duration. According to Felder, there are three different tribological regimes related to the ratio. The first one is for $H > 2$ when the oxide scale is insignificantly cooled by the cold roll surface. For this regime, the scale can be characterized as ductile, softer than the metal, and strongly adherent to the metal surface. The friction is described by the Tresca friction factor, which is not sensitive to the pressure and the contact time. For the second regime, when $H < 0.05$, the oxide scale is considered to be significantly cooled due to contact with the roll; it is harder than the metal and can be considered quasi-rigid. The scale is brittle, has a low adherence, and can be considered abrasive. The Coulomb-type friction coefficient, proportional to the shearing and not very sensitive to the contact time, is applied for this regime. In between those two extreme cases, for $0.05 < H < 2$, the friction

Table 1 List of symbols

Symbol	Meaning
A	Empirical constant
A_a	Overall apparent contact area
A_{ox}	Apparent area occupied by the oxide scale
A_s	Apparent area occupied by the extruded fresh steel
a	Negative empirical constant
a_c	Thermal diffusivity
a_{tens}	Function of the front and back tension
B	Empirical constant
C_{b2}	Heat-transfer coefficient for partial contact at the "two-layer" zone
C_e	Effective interface heat-transfer coefficient
C_{e1}	Heat-transfer coefficient for contact at the "one-layer" zone
C_{e2}	Heat-transfer coefficient for contact at the "two-layer" zone
C_{ox}	Heat-transfer coefficient through the oxide scale
f_{Normal}	Contact force
H	Ratio between the scale thickness and the scale thickness thermally affected by contact with the tool
H_o	Strip entry thickness
H_{sc}	Factor depending on the state of the secondary scale at the roll gap
HV	Vickers hardness of the contacting material
HV_{ox}	Vickers hardness of the oxide scale
h_{exit}	Scale thickness at the exit from the roll gap
h_{min}	Oil film thickness
h_o	Interface heat-transfer coefficient
$h_{o\ slab}$	Initial slab thickness
Δh_{slab}	Absolute reduction in the slab thickness
k	Metal flow strength
k_1, k_2, k_3	Empirically established constants
k_{ox}	Scale thermal conductivity
k_{shear}	Shear strength
m	Friction factor
m_c	Empirical constant within the range of 0 to 1
p	Normal pressure
p_a	Apparent contact pressure
q	Heat flux per unit area across the interface
R	Roll radius
R_a	Surface roughness
R_e	Total thermal resistance over the entire apparent contact area
R_{e1}	Thermal resistance over the extruded fresh steel
R_{e2}	Thermal resistance over the oxide scale
S	Modified Sommerfeld number
T	Temperature in °C
T_1, T_2	Temperatures on either side of the interface
T_F	Temperature in °F
T_{oxs}	Surface temperature of the oxide scale
Δt	Contact time
v	Rolling velocity
v_{rel}	Relative velocity
α_s	Area fraction of the gaps formed from the through-thickness cracks at the interface and filled with fresh metal
γ	Pressure viscosity coefficient
δ	Temperature viscosity coefficient
δ_{ox}	Scale thickness
δ_{oxt}	Scale thickness thermally affected by contact with the tool
ε	Strain
ε_A	Fraction of the total area that is in direct contact
η	Viscosity coefficient
η_o	Uncorrected viscosity coefficient
θ	Cross coefficient accounting for the interaction between both the pressure and the temperature viscosity coefficients
λ_{h1}	Harmonic mean of the thermal conductivity of the roll material and the specimen material
λ_s	Thermal conductivity of the specimen material
μ	Coefficient of friction
σ^*	Combined roughness of the two surfaces
τ	Interface shear stress

behavior can become a complex function of the contact time and pressure.

In 1997, Munther and Lenard combined the data from rolling samples in a laboratory rolling mill with different scale thicknesses at various temperatures (Ref 13). Experimentally measured data, such as the roll-separating forces, torques, and forward slip, coupled with finite-element analysis led to the determination of the friction coefficient. They found that the friction coefficient increases with increasing reduction and decreasing temperature; it also increased with decreasing velocity and decreasing oxide scale thickness (Fig. 1). They put the experimental evidence of the effect of scale thickness on the friction coefficient into the following formula:

$$\mu = 0.369 - 0.0006 h_{exit} \qquad \text{(Eq 8)}$$

where h_{exit} is the scale thickness at the exit from the roll gap.

Li and Sellars reported that forward slip in the hot rolling of steel increases significantly with oxide scale thickness for the same reduction (Ref 14). The forward slip was measured for a relatively wide range of scale thickness, 20 to 670 μm. They attributed the change in the forward slip to the variations of the scale temperature and, as a result, to changes in the roll/scale contact conditions. The real contact area between the roll and the oxide scale will be less for a thick scale than for a thin one under similar contact pressure. This is because the thicker oxide scale, being cooler on its outer surface, is less able to fill the valleys of the roll surface asperities during a rolling pass. A smaller contact area means an easier relative movement between the roll and the scale layer, which, coupled with a lower oxide surface temperature, should lead to a larger forward slip. For a similar scale thickness, the measured forward slip for the higher reduction was higher than that measured for the lower reduction. The lubrication behavior of the thin oxide scales described previously is in agreement with the load and torque measurements made by El-Kalay and Sparling during hot rolling of mild steels (Ref 15). The decrease of the friction coefficient in the roll gap followed by the temperature increase was noticed by Ekelund in 1933 during hot rolling of carbon steels (Ref 11, 16). That effect can also be related to the lubrication behavior of the soft oxide scale.

Determination of Friction Coefficient

Because friction is highly sensitive to contact parameters, friction measurements must be gathered under conditions very similar to those of the process under consideration. Pressure, sliding speed and length, temperature, plastic deformation pattern, lubricant and mode of lubrication, tool and workpiece nature, structure and texture, degree of work hardening, surface chemistry, and roughness may all need to be taken into account. The best similarity can be achieved in the industrial process itself. However, such industrial tests are difficult and expensive, so that simpler and smaller laboratory tests using experimental rigs are often preferable. However, the further away from actual conditions the laboratory test is, the more interpretation is required to apply the laboratory results to the industrial operation.

Laboratory Monitoring Methods. Direct measurement of friction stresses requires embedding transducers into the tools, measuring both the normal and the tangential contact stresses. These methods have been applied to measure interfacial stresses in several bulk forming processes (Ref 17–22). A strain-gaged cantilever with its tip in the contact zone and its subsequent refinements are described elsewhere (Ref 23–25). An important concern with this method is the possibility of some metal intruding into the clearance between the pin and its housing, particularly if the metal is soft, such as when at high temperature. It is therefore necessary to substantiate the results by independent means (Ref 26). An alternative to this localized determination of friction conditions is to obtain an average frictional shear stress and/or an average coefficient of friction over the whole interface. Methods in this group include plane-strain drawing (Ref 27) and plane-strain compression (Ref 28). A twist compression technique has been reported to measure the friction factor when deforming aluminum (Ref 29), and caustics have also been used to evaluate the roll pressure distribution and thus to obtain the coefficient of friction during flat rolling (Ref 30, 31). These tests are very informative about interfacial conditions and can be confirmed by mechanical modeling followed by numerical simulation. However, these techniques are restricted to laboratory forming equipment, where operating conditions may not be quite the same as the industrial counterpart.

Indirect measurements of friction involve either load or simple geometrical measurements. A mechanical model is necessary to extract the relevant tribological information from such measurements, and any inaccuracy in the model will result in a potentially large error on the friction coefficient. For this reason, very simple sample and tool geometry and kinematics are often preferable for deriving simple but relatively accurate models for the friction coefficient, either as formulae (e.g., plane-strain compression test) or diagrams (e.g., ring compression test). The methods based on the deformation load measurement may be applied to uniaxial compression, extrusion, drawing, and rolling (Ref 33). The methods involving measurement of deformation include uniaxial compression with a tapered punch (Ref 34), measuring the forward slip or the bite angle in rolling (Ref 35), monitoring the fold-over in plane-strain compression (Ref 36), or the extrusion forging test (Ref 37). The most popular technique for forging conditions is the ring compression test (Ref 38, 39). Concave and convex shapes can be used in addition to the cylindrical ones (Ref 40). The shear factor in cold forward bar/backward cup extrusion was obtained by the upper-bound method and the measured extruded cup height (Ref 41, 42).

Simple Relationships for Coefficient of Friction. Several simple formulae have been published relating the coefficient of friction to various parameters, some of which are mentioned in the previous section. In cold rolling, for instance, those traditionally used most often are due to Hill, Roberts, and Ekelund (Ref 9–11). A comparison of the predicted coefficients of friction is shown in Fig. 2. The plotted friction coefficients were calculated using data obtained for cold rolling of the steel strips lubricated with mineral seal oil (Ref 43). All three formulae show the expected trend of lower friction with increased speed. The coefficient also tends to decrease as the reduction increases, demonstrating the combined result of the increasing number of contact points, the increasing temperature, and normal pressure.
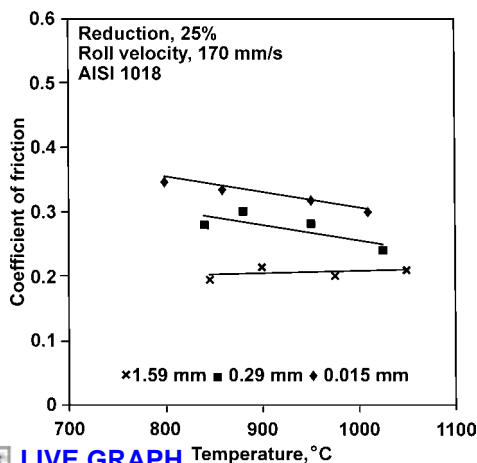
Fig. 1 Influence of rolling temperature on coefficient of friction for 25% reduction, 170 mm/s rolling velocity, and different scale thicknesses of 1.59, 0.29, and 0.015 mm. Source: Ref 13
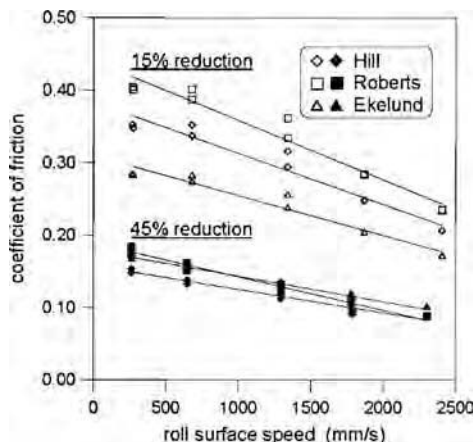

Fig. 2 Coefficient of friction according to the formulae for cold rolling of carbon steel by Hill, Roberts, and Ekelund. Source: Ref 43

A simple relationship for the minimum coefficient of friction in flat rolling, based on Bland and Ford's equations, has been reported in terms of the strain, ε, applied strip tension, and the entry thickness, H_o:

$$\mu = \frac{(\varepsilon - a_{tens})}{H_o} \qquad \text{(Eq 9)}$$

where μ is the friction coefficient, and a_{tens} is a function of the front and back tension (Ref 44).

Inverse Method. An increasingly popular means of calculation is the inverse method (Ref 45). In this technique, parameters of the process are determined experimentally. A mathematical model is used to calculate these parameters, and the friction coefficient is adjusted until a desirable level of accuracy is achieved in matching the measurements and predictions. In the inverse method, a finite-element simulation can be combined with the measurements of overall parameters to achieve better accuracy (Ref 46–48). A coefficient of friction has been obtained by examining the bulge on a conical, hollow cylinder, compressed between rigid platens. The emissivity is determined in the first turn, followed by the coefficient of heat transfer and the metal resistance to deformation, creating the bulge (Ref 49).

Some of the measured friction coefficients are presented in Table 2.

Importance of an Appropriate Model and Accurate Mechanical Properties

Nowadays, significantly increased performance and accuracy of numerical computations has allowed accurate simulation of complex processes to such an extent that indirect measurements using the real process have become possible (Ref 50). Computer-assisted tribology by trial-and-error or automated inverse modeling of the operations can provide accurate friction coefficients on the basis of measured force, torque, and/or geometrical workpiece evolution, taking into account that all other process parameters are precisely known.

For instance, a pilot cold rolling mill and associated mathematical model of the strip rolling process was developed through the 1980s (Ref 57, 58). The measurements recorded by a computer include roll load and torque, front and back tension forces, entry, roll and exit speeds (from which the reduction and forward slip are determined), and strip exit temperature. The plane-strain compression test is part of the procedure, being used both to measure the mechanical properties of the strips to be rolled and to obtain a first approximation of the tribological properties of the lubricants to be tested. The mode of lubrication in the compression test is quite different from the constant feeding of lubricant into the roll bite, making the friction coefficients obtained from the test unreliable for the rolling. The rolling model couples a slab method for strip deformation with an isotropic elastic finite-element model for roll deformation.

An inverse version of this model minimizes a quadratic error function, allowing the friction coefficient to be obtained. The importance of using a reliably accurate model in indirect methods is illustrated in Fig. 3, where the friction coefficient is an increasing, roughly constant, or strongly decreasing function of reduction, depending on the chosen model. The first five curves, Roberts I, Roberts II, Roberts III, Ekelund, and Ford, are based on formulae derived from simplified models of rolling, while the last three, Sims, Alexander (force) and Alexander (torque), are based on more elaborate models and give quite different friction coefficients.

The impact of the assumed mechanical properties on the computed friction coefficient must also be taken into consideration. The rolled strip is elastic-plastic, with some degree of strain-rate sensitivity (viscoplasticity), depending on the degree of work hardening. Moreover, it is anisotropic, sometimes significantly so. Some simplification can be assumed by using mechanical tests that ensure a deformation pattern that is close to that of the forming process. Friction and mechanical properties should ideally be measured by one and the same test, bearing a strong resemblance to the rolling process, providing that enough independent observables (preferably of different natures: force, geometry, kinematics) are made available. Modern computation capabilities allow this double identification to be performed on the industrial process itself. These nonindependent measurements are not completely sufficient, unless the same constitutive equations provide a good fit for a wide variety of tests performed on a given metal composition.

Interface Heat-Transfer Coefficient

The complexity of the interface between tool and stock makes measurements very difficult. The direct measurement of friction and heat transfer is impractical for most industrial hot metal-forming operations, and even for many conducted in the laboratory. The interface heat-transfer coefficient (IHTC), h_o, is used to quantify the resistance of an interface to the transfer of heat, usually from workpiece to tool. It is commonly defined by the equation $q = h_o(T_2 - T_1)$, in which T_1 and T_2 are the temperatures on either side of the interface, and q is

Table 2 Measured friction coefficients between workpiece and tool in metal forming

Metal	Conditions	Value(a)	Reference
Aluminum	Cold rolling and compression	$\mu = 0.05–0.41$	50
Aluminum 1100	Cold rolling, unlubricated	$\mu = 0.025–0.053$	51
Steel AISI 1010	Range of labs and techniques in round robin, lubricated	$m = 0.05–0.6$	52
Steel	Formulae for industrial rolling	$\mu = 0.1–0.4$	43
Stainless steel 304	High-temperature lab compression, lubricated	$m = 0.2$	53
Stainless steel 304	High-temperature lab compression, unlubricated	$m = 0.8$	53
Ti-6Al-4V	High-temperature lab compression, lubricated	$m = 0.2$	53
Ti-6Al-4V	High-temperature lab compression, unlubricated	$m = 0.8$	53
General	Microstrip drawing	$\mu = 0.08–0.49$	54
General	Microforming	$m = 0.02–0.4$	55
General	Microforming	$m = 0.15–0.85$	56

(a) Coefficient of friction, μ; friction factor, m. See text for definitions.

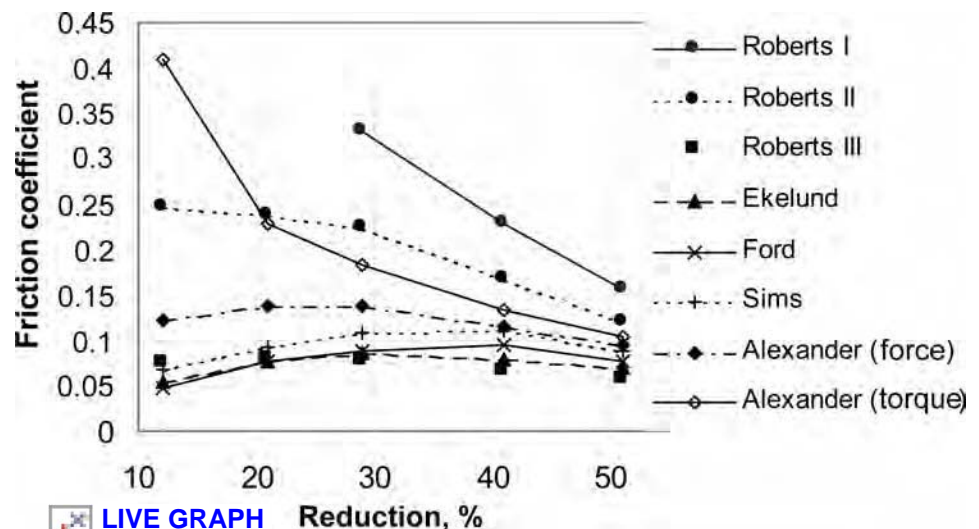LIVE GRAPH
Click here to view

Fig. 3 Application of various friction coefficient determination methods to five tests performed on an experimental rolling mill: Al 5052-H19, emulsion lubrication, SAE 52100 roll steel grade, speed 1 m/s, and variable reduction. Source: Ref 50

the heat flux per unit area across the interface. The difficulties of making laboratory measurements, combined with the complexity of events at the tool-stock interface, result in a wide range of reported values for the IHTC (Table 3) (Ref 59).

In the absence of detailed insight and a lack of fundamental understanding about the mechanism of heat transfer at a moving interface, most modelers assume a simple description, or an average value, of the heat-transfer coefficient. The contacting points between two surfaces serve as paths of lower resistance for heat flow in comparison to adjacent regions where heat transfer occurs by conduction through air gaps (Ref 72). Thus, it has been assumed that the link between friction and heat transfer at the interface is the fraction of the total area, ε_A, the area that

Table 3 **Measured interface heat-transfer coefficient (IHTC) between roll and stock for the hot rolling of steel and aluminum**

Material	IHTC, kW/m^{2}K	Reference
Steel		
	10–50	60
	15	61
	15–20	62
	19–22	63
	100–350	64
	200–450	65
Aluminum		
	2–20	66
	5–50	67
	10–260	68
	18–38	69
	23–81	70
	200	71

Source: Ref 59

Fig. 4 Influence of the mean roll pressure on the mean heat-transfer coefficient during hot rolling of low-carbon, stainless, and microalloyed steels. Source: Ref 75

is in direct contact. It has been postulated that the real contact area depends on both the interfacial pressure, p, and the shear strength, k_{shear}, in the real contact zone:

$$\varepsilon_A = \frac{p\mu}{m_c k_{shear}} \qquad (Eq\ 10)$$

where m_c is an empirical constant within the range of 0 to 1, and μ is the friction coefficient at the interface (Ref 73). Based on experimental results (Ref 74), it was pointed out that the variation in IHTC with reduction, rolling speed, and lubrication observed through pilot mill tests on a 316L austenitic stainless steel could be explained on the basis of the influence of these rolling parameters on fractional contact area. The observed increase in the IHTC during rolling was related to the increase in pressure at the roll-stock interface, leading to an increase in the real area of contact between two surfaces. The influence of other factors, such as rolling reduction, rolling temperature, roll speed, roll and workpiece mechanical properties, and surface roughness, can be related to their effect on roll pressure. It has been found that the mean IHTC is linearly related to mean pressure (Fig. 4) (Ref 75). This relationship can be used to determine the magnitude of the IHTC in industrial rolling from an estimate of the rolling load. According to the estimation, the heat losses to the work rolls during the early passes in hot rolling can be more than 30% of the total. This shows the importance of accurately characterizing the IHTC in the roll bite.

Thermal Effect of Oxide Scale. The application of lubricant or the presence of oxide scale introduces an additional thermal resistance between the roll surface and the workpiece. During strip rolling, for example, the oxide scale layer adhered to the surface of the strip attempts to elongate in the rolling direction with the deformation of the rolling stock. In many cases, with large reduction and low rolling temperature, the oxide will be unable to deform plastically, and through-thickness cracks will appear orientated mostly perpendicular to the rolling direction. These cracks will allow extrusion of fresh hot metal through the gaps within the scale under the pressure in the roll gap. As a result, direct contact between the relatively cold roll and the hot metal extrusions can occur. This type of scale behavior has been observed in the hot rolling of both aluminum (Ref 76) and steel (Ref 77). Based on the experimental observations of oxide scale behavior, analysis of real contact area and thermal resistance, combined with experimentally derived IHTC values, a physical model has been developed to represent heat transfer during hot steel rolling (Ref 78).

According to the model assumptions, the interface heat transfer within the roll gap consists of two parallel heat-flow systems: through the oxide scale, called a "two-layer" zone, and directly between the roll/fresh metal interface, called a "one-layer" zone. Thus, the total thermal resistance over the entire apparent contact area can be expressed from the following equation:

$$\frac{A_a}{R_e} = \frac{A_s}{R_{e1}} + \frac{A_{ox}}{R_{e2}} \qquad (Eq\ 11)$$

where A_a, A_s, and A_{ox} are the overall apparent contact area and the apparent areas occupied by the extruded fresh steel and by the oxide scales in the roll gap, respectively. The effective IHTC, C_e, can be derived from Eq 11 as:

$$C_e = C_{e1}\alpha_s + C_{e2}(1 - \alpha_s) \qquad (Eq\ 12)$$

where C_{e1} and C_{e2} are the heat-transfer coefficients for the one-layer and two-layer zones, correspondently. The term α_s is the area fraction of the gaps formed from the through-thickness cracks at the interface and filled with fresh metal and is defined as $\alpha_s = A_s/A_a$. To obtain the effective IHTC for the entire rolling pass, it is therefore necessary to obtain not only the heat-transfer coefficient components for the individual contact zones and thermal barriers but also to know the mean area fraction of the fresh steel in the roll gap. The mean area fraction of the fresh steel extruded through the gaps within the oxide scale can be estimated using:

$$\alpha_s = \Delta h_{slab}\left(\frac{2}{3h_{o\ slab}} + \frac{1}{8R}\right) \qquad (Eq\ 13)$$

where Δh_{slab} is the absolute reduction in the thickness, $h_{o\ slab}$ is the initial slab thickness, and R is the roll radius (Ref 78). The equation for the effective IHTC (Eq 12) can be rewritten, depending on heat-transfer coefficients for the individual contact zones and thermal barriers, as:

$$C_e = C_{e1}\alpha_s + \frac{C_{ox}C_{b2}}{C_{ox} + C_{b2}}(1 - \alpha_s) \qquad \text{(Eq 14)}$$

where C_{b2} is the heat-transfer coefficient for the partial contact at the two-layer zone, usually called contact conductance; and C_{ox} is the heat-transfer coefficient through the oxide scale. The coefficient C_{ox} can be approximately obtained for the given oxide scale thickness, δ_{ox}, and the scale thermal conductivity, k_{ox}, using:

$$C_{ox} = \frac{k_{ox}}{\delta_{ox}}. \qquad \text{(Eq 15)}$$

No systematic analysis appears to have been conducted for quantitative variations of the contact conductance with surface, interface, and deformation conditions during metal-forming operations. However, it has been shown that the contact conductance is related to the apparent contact pressure, p_a, and the Vickers hardness of the softer contacting material, HV, in addition to the surface roughness and thermal conductivity of two contacting solids under normal static contact conditions (Ref 79–81). Assuming this relationship and also the relationship between the degree of real contact and the dimensionless contact pressure obtained on the basis of experimental measurements and mathematical analysis (Ref 82, 83), an exponential relationship between the contact conductance and the contact pressure during hot rolling has been established (Ref 78). The same contact and heat-transfer states at the scale layer/tool interface were assumed for forging and rolling. Accordingly, the contact conductance for a two-layer zone, C_{b2}, during hot steel rolling can be calculated by using the following same equation developed for hot forging of steel:

$$C_{b2} = A\frac{\lambda_{h2}}{R_a}\left[1 - \exp\left(-0.3\frac{p_a}{HV_{ox}}\right)\right]^B \qquad \text{(Eq 16)}$$

where A and B are empirical constants whose values for plain carbon steel are 0.4×10^{-3} and 0.392, respectively; R_a is the roll surface roughness, and λ_{h2} is the harmonic mean of the thermal conductivity of the oxide scale, λ_{ox}, and the steel roll, λ_r, and is determined by:

$$\frac{1}{\lambda_{h2}} = \left(\frac{1}{\lambda_r} + \frac{1}{\lambda_{ox}}\right)\Big/2 \qquad \text{(Eq 17)}$$

The Vickers hardness of the oxide scale, HV_{ox}, is considered to vary with the surface temperature of the oxide scale, T_{oxs}, according to the following equation developed on the basis of available experimental data (Ref 84):

$$HV_{ox} = 7075 - 538T_{oxs}(293K \leq T_{oxs} \leq 1273K) \qquad \text{(Eq 18)}$$

Equation 16 can be replaced by the following simpler version for low pressure only:

$$C_{b2} = A\frac{\lambda_{h2}}{R_a}\left(0.3\frac{p_a}{HV_{ox}}\right)^B \qquad \text{(Eq 19)}$$

For a one-layer zone and for the rolling conditions where the initial rolling temperature is approximately 1000 °C, the scale thickness is within the 25 to 700 µm range, the rolling reduction is between 10 and 50%, and the corresponding average rolling pressure is between 130 and 200 MPa, the contact conductance can be calculated using the same equations (16 and 19) where the constants A and B are set to 0.405 and 1.5, respectively, and the parameter λ_{h2} is replaced by λ_{h1}. The parameter λ_{h1} is the harmonic mean of the thermal conductivity of the roll material, λ_r, and the specimen material, λ_s, and is determined by $1/\lambda_{h1} = (1/\lambda_r + 1/\lambda_s)/2$, while HV_s is the Vickers hardness of the fresh steel specimen.

Figure 5 illustrates changes of the IHTC derived for the different scale thicknesses and rolling reductions. As can be seen, the IHTC decreases dramatically when the scale thickness increases because of the relatively poor thermal conductivity of the oxide scale. At the same time, the IHTC increases rapidly with rolling reduction. This is physically consistent with the variation of the real contact area and the high contact conductance in the fresh steel zone that dominates the overall high values of heat-transfer coefficient at the interface during steel rolling, even though the area fraction of the fresh steel zone is less than that of oxide scale for rolling passes with reduction less than 50%.

Determination of Interface Heat-Transfer Coefficient

There are many models of heat-transfer conditions that have been reported (Ref 66, 85–90). Verification of these models with experimental work, wherever it is possible, is essential for achieving reasonably accurate predictions. Normally, these techniques are based on temperature measurements around the tool-workpiece interface, although they only give estimates of interface temperatures and hence the contact resistance. For the forging problem, for instance, one of the reasons for the lack of

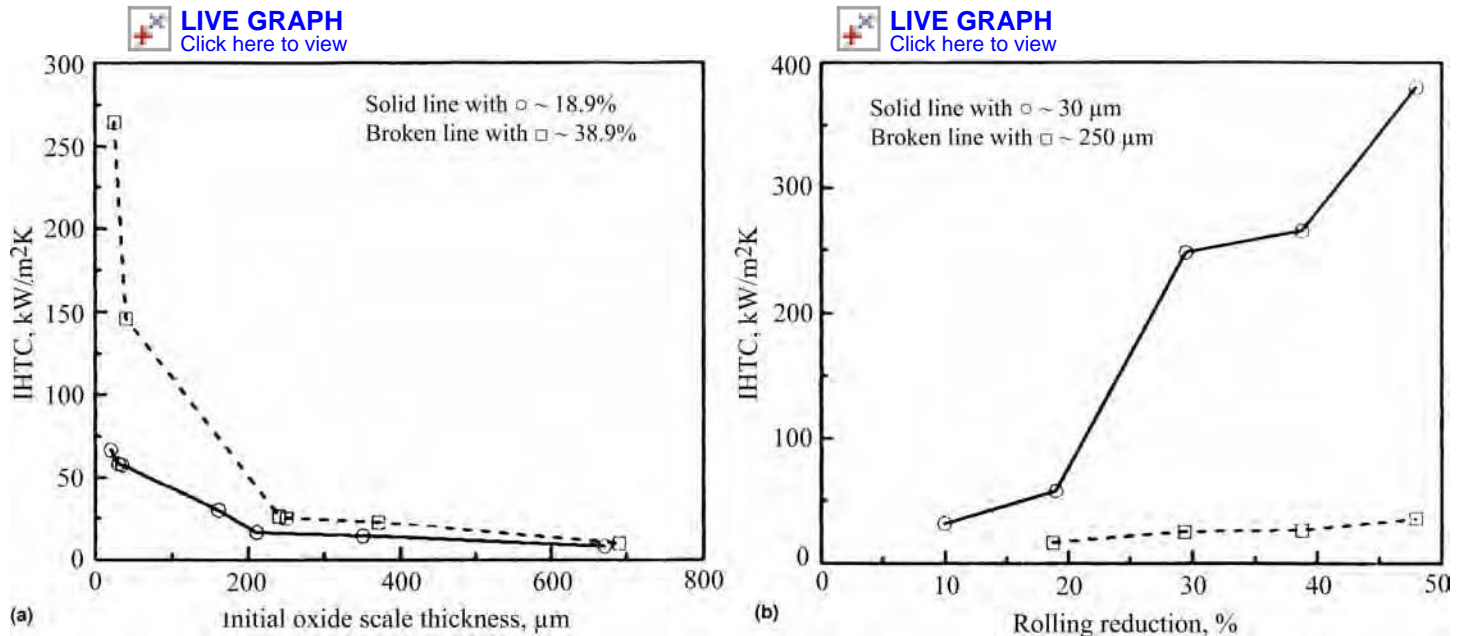**Fig. 5** Interfacial heat-transfer coefficient (IHTC) during steel hot rolling with initial temperature of approximately 1000 °C. (a) Derived for different scale thicknesses. Solid line with open circles, reduction ~18.9%; broken line with open squares, reduction ~38.9%. (b) Derived for different rolling reductions. Solid line with open circles, scale ~30 µm; broken line with open squares, scale ~250 µm. Source: Ref 78

accurate temperature measurement directly at the die-billet interface is the severe conditions both in the bulk metal and at the interface. There is always an integrity problem with the wire leads for thermocouples mounted within the billet. Moreover, the tool-workpiece interface conditions are never constant. To avoid these difficulties, the variation of temperature in the die and billet is recorded using thermocouples mounted away from the interface, and the results are compared with an analytical or numerical analysis in which both the lubricant thickness and the value of conductivity are chosen for the best fit to the obtained experimental data (Ref 91).

The ring test applied for studying friction conditions in hot metal forming can also be used for measurement and analysis of the temperature at the contact zone (Ref 89). A technique for evaluation of the heat-transfer coefficient has been developed for a nonisothermal bulk-forming process based on both measurements of the internal die temperatures for two dies brought together under different pressures and using a one-dimensional numerical analysis. This technique was further developed for use in a hot forging process to derive the IHTCs (Ref 62, 92). To achieve better accuracy, the forging temperatures can be measured directly and simultaneously at the billet surface, the die surface, and also inside the die during upsetting (Ref 93). The IHTC is then derived

by applying finite-element or inverse numerical methods for analysis of the temperature jump across the interface. This measurement technique was also successfully applied for hot rolling purposes by measuring three rolling temperatures simultaneously: at the billet surface, the roll surface, and inside the roll (Ref 94). The IHTC was varied in the simulation while other parameters, such as the friction factor obtained from a ring test, were assumed to be constant. The value of IHTC obtained for cold upsetting was close to 100 kW/m^{2}K and was 50 kW/m^{2}K for the hot billet on a cold die, both for unlubricated conditions.

Some of the IHTCs determined by different authors are presented in Table 4.

Computer simulations have become widely used in the development of the boundary-condition models for deformation processes. The effect of the dynamically changing friction and heat-transfer coefficients is essential for prediction accuracy. In most commercially available metal-forming analysis software, the coefficients are implemented as constants, some allow for implementation of varying coefficients, and seldom, generating the changing values based on contact conditions is possible. An important ground for more sophisticated friction and interface heat-transfer "coefficients" occurs in hot forming when an oxide on the metal surface plays a crucial role in interface behavior. The next section considers this case.

Oxide Scale Mechanical Behavior

The behavior of oxide scale on the surface of hot metal undergoing thermomechanical processing presents a rich variety of phenomena of great technological importance. To break the primary scale in readiness for hot rolling of flat products, the slab is passed through a slab descaler. However, a secondary scale is formed between successive rolling passes, which has to be removed by high-pressure water jets before the subsequent passes during reversing rolling or before the strip enters the tandem finishing mill. This secondary scale grown after passing the first slab descaler significantly influences the chain of microscale events at the roll/workpiece interface. Despite considerable complexity, a combination of careful experiments and detailed finite-element analysis has been successfully applied to represent a wide range of the observed physical phenomena (Ref 102). Although the results presented as follows are mainly for carbon steels, the method can be applied to other technologically important alloys, such as stainless steels and aluminum alloys.

The friction forces drawing the stock into the roll gap exert a tensile pull that can deform and even fracture the oxide scale. The simple uniaxial tensile test can thus provide much valuable information on the behavior of oxide scale that is relevant to thermomechanical processing (Ref 103, 104). The tests reveal two types of accommodation by the oxide scale of the deformation of the underlying steel substrate. At lower temperatures, the oxide scale fractures, usually in a brittle manner, with the through-thickness cracks triggering spallation of the oxide scale from the steel surface. At higher temperatures, the oxide scale does not fracture; rather, it slides over the steel surface, eventually producing delamination of the scale. The temperature of transition between these two types of failure is sharp and sensitive to steel chemical composition (Ref 105).

Because the oxide scale conducts heat at a much lower rate than the underlying metal, steep temperature gradients can be developed across the oxide scale thickness. This leads to thicker scales having a cooler outer surface than thin scales. Thin scales can thus remain hot and deform in a ductile manner along with the steel substrate as the stock is drawn into the roll gap (Ref 106). Thicker scales have cooler outer surfaces, where fracture can initiate more easily, even well ahead of roll contact. Much thicker scales can withstand higher forces and may not crack until subject to the additional force due to bending as the stock first meets the roll. These observations are significant, because the presence of cracks will affect subsequent interface behavior, and furthermore, cracks that have formed well ahead of roll contact will have opened more than those appearing just as the roll grips the surface of the stock.

An open gap in the oxide scale may enable the steel underneath to extrude up under the roll

Table 4 Determined interface heat-transfer coefficients (IHTCs) between workpiece and tool in metal forming

Metal	Conditions	IHTC, kW/m^{2}K	Reference
Steel	General test, laboratory, no scale, no lubricant, direct measurement	29–35	70
	General test, laboratory, no scale, water lubricant, direct measurement	23–81	70
	General test, laboratory, no scale, oil lubricant, direct measurement	70–465	70
	General test, laboratory, 10 μm scale, no lubricant, direct measurement	7–10.5	70
	General test, laboratory, 10 μm scale, water lubricant, direct measurement	10.5	70
	General test, laboratory, 10 μm scale, oil lubricant, direct measurement	5.8–10.5	70
Low-carbon steel	Laboratory conditions	13	85
	Laboratory conditions	10–20	86
	Industry conditions	40–80	86
	Industrial conditions, water cooled	38.7	95
Steel	Industrial conditions, no lubrication	18–37.6	95
	Industrial conditions, no lubrication	30	96
	Laboratory conditions, no lubrication	17–57	74
	Laboratory conditions, oil lubrication	17–31	74
	Laboratory conditions, no lubrication	12–29	20
	Compression, laboratory	10.7–41.7	87
Steel M40/EN8	Compression, laboratory	7.79	97
Steel	Extrusion, laboratory	4–10	71
Steel Ma8	Laboratory compression	20–40	98
Stainless steel 304	Laboratory compression	20–40	53
Aluminum	Laboratory conditions	30	85
		15–30	86
	Laboratory conditions	30	99
	Laboratory conditions, emulsion lubricant	10–54	60
	Upsetting, laboratory, slow test, no lubricant, direct measurement	7.5–10	62
	Upsetting, laboratory, fast test, Renite 28 lubricant, direct measurement	15–20	62
	General test, laboratory, no lubricant, direct measurement	20–40	100
	General test, laboratory, MoS$_2$ lubricant, direct measurement	20–80	100
Aluminum 6061	Laboratory compression	50	93
Ti-6Al-4V	Laboratory compression	20–40	53
Steel	Influence of oxide scale thickness, forging	1.1–11.6	101
	Influence of oxide scale thickness, rolling	11.6–116.3	101

contact pressure (Fig. 6) (Ref 107). When such hot steel makes direct contact with the roll, the local friction and heat-transfer conditions can be expected to change dramatically.

Pickup on Roll. An important surface-quality defect stems from pickup by the roll of oxide scale from the steel surface, usually in small patches that can then come back around on the roll surface and indent into the following metal. This pickup can be replicated in a simple compression experiment under controlled atmospheric conditions. This experiment has been successfully modeled using finite-element analysis and the appropriate material properties transferred to the rolling model to demonstrate pickup of oxide scale during rolling (Fig. 7) (Ref 102, 108).

Descaling. A further surface defect may arise when the oxide scale is removed after hot rolling. If the scale had fractured and the metal had extruded up through the gaps during rolling, such extrusions become protrusions when the oxide has been removed and will need to be cold rolled to smooth the surface. Because the oxide scale may be severely damaged by rolling (Fig. 6), it is appropriate to consider whether this damage makes subsequent descaling easier or not. By including the effect of differential thermal contraction under the influence of water jets, as well as the impacting force of those jets, the finite-element model has been extended to demonstrate that, indeed, partially detached oxide scale is the first to be removed by hydraulic descaling (Fig. 8) (Ref 109).

Mechanical descaling, by reverse bending, is a popular if not always completely successful method for removing the oxide scale at room temperature. Under ambient conditions, the scale is much more brittle than the underlying steel and can be expected to be removed relatively easily. Several types of oxide scale failure have been observed in a parallel investigation on the mechanical descaling of wire, and the finite-element model was able to replicate each of these failure modes (Ref 110).

During multipass hot rolling of long products, the magnitude of the coefficient of friction within the roll bite varies due to the complex pressure-slip variations. A new Coulomb-Norton-type friction model for long products and bar sections has been developed (Ref 111). Among other assumptions, the model takes into consideration some of the complex interactions at the stock-roll interface due to the presence of secondary oxide scale. Hence, the different modes of scale failure, such as the through-thickness cracking and the sliding mode, depending on the temperature and steel composition, have been implemented. The friction force occurs either between the roll surface and the stock oxide scale or between the roll surface, the oxide scale fragments, and eventually fresh steel extruded through the scale gaps, depending on the relative magnitude of the shear stresses inside the scale layer and at the oxide scale/stock interface. The coefficient
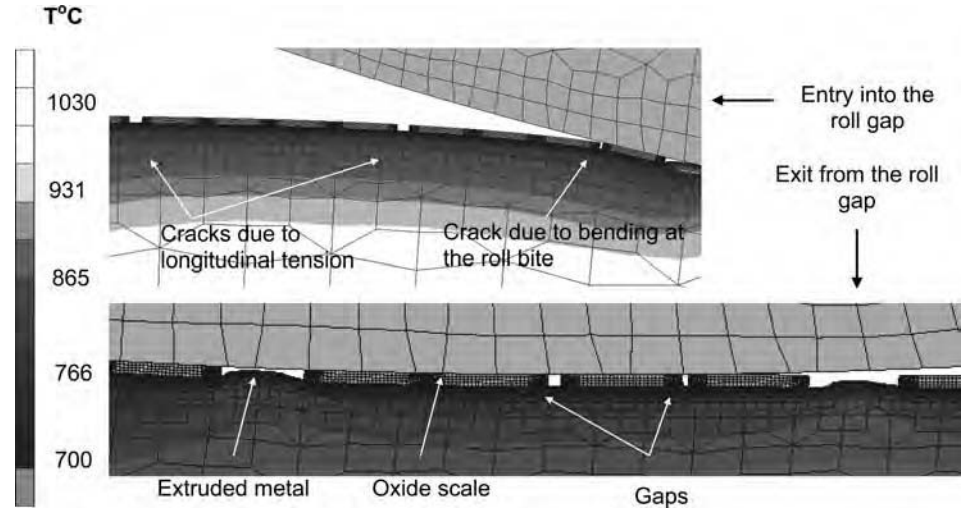

Fig. 6 Temperature and crack distribution at the oxidized stock/roll interface during hot rolling. Source: Ref 107
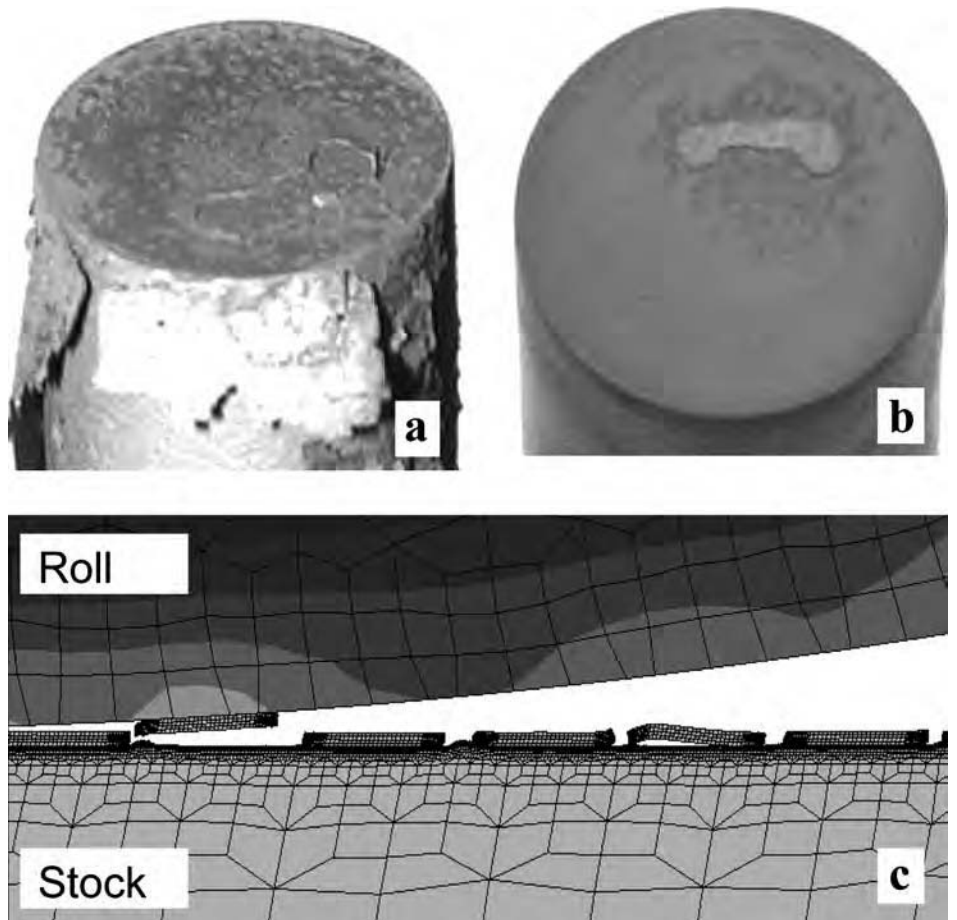

Fig. 7 Photographs illustrating the oxide scale (a) partly separated from the specimen and (b) transferred to the tool after being in contact during compression test at 870 °C. (c) Oxide scale pattern predicted at exit from the roll gap. Note the pick-up effect. Source: Ref 102

of friction is calculated as a function of the contact force, f_{Normal}, the sliding velocity, v_{rel}, the stock temperature, T, the roll surface roughness, R_a, and the factor H_{sc}, which depends on the state of the secondary oxide scale at the roll gap:
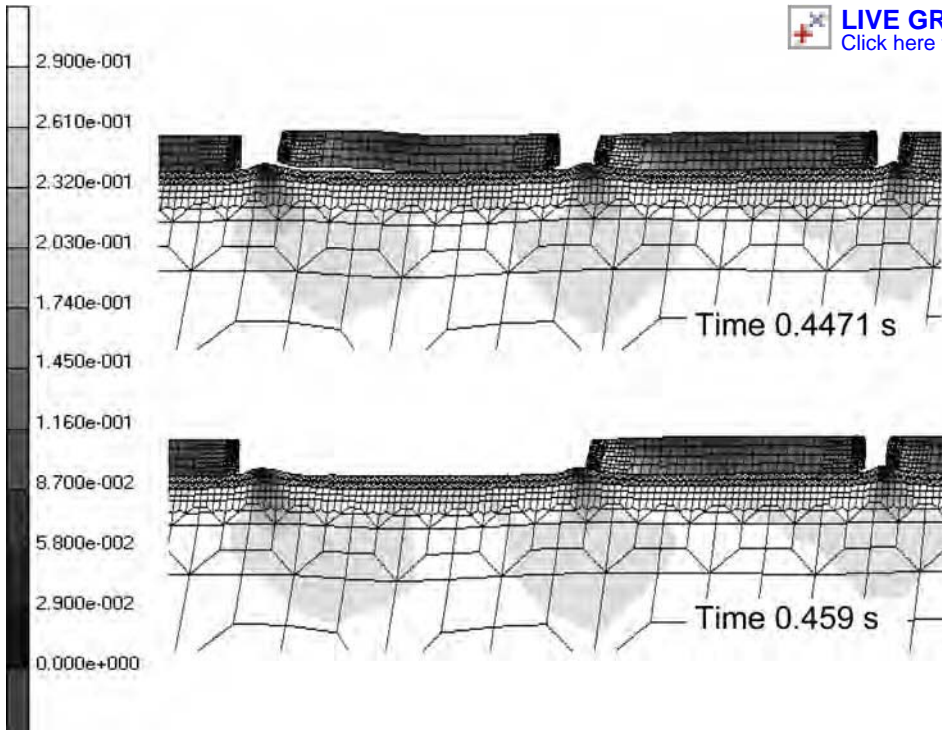
Fig. 8 Equivalent total strain distribution at the homogeneous scale for the different time moments during subsequent descaling. Note the peeling off of the partly spalled scale as first in the sequence. Source: Ref 109
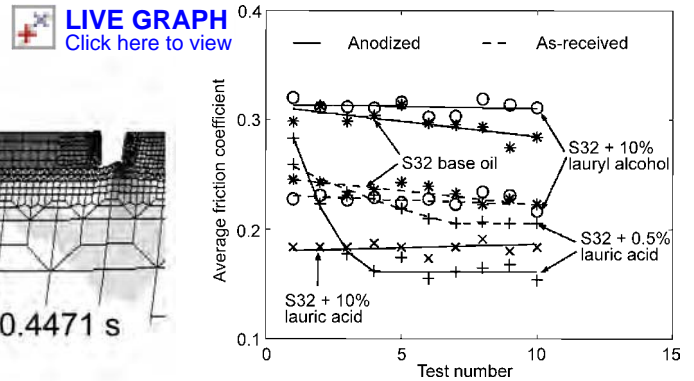

Fig. 9 Effect of additives in S32 base oil on the evolution of friction with test number. Solid curves correspond to anodized samples, while dashed curves represent as-received samples. Source: Ref 115

$$\mu = k_1 H_{sc}^{-\log\left(\frac{1200}{T}\right)} a \tan\left(R_a^{\frac{Tk_3}{1200}}\right) \frac{\log(1 + f_{Normal})}{\log(k_2 + |v_{rel}|)}$$

$$\text{(Eq 20)}$$

where k_1, k_2, and k_3 are constants established experimentally. The factor H_{sc} is the following function of the thickness of the secondary scale, δ_{ox}, the thermal diffusivity, a_c, and the contact time, Δt:

$$H_{sc} = \delta_{ox}(6a_c\Delta t)^{-0.5} \qquad \text{(Eq 21)}$$

The model has been implemented as a subroutine in a commercial finite-element code and used to represent the effect of each variable on the coefficient of friction (Ref 112).

Rolling Aluminum with Lubrication. It has been shown that surface bending and stretching, similar to those observed during hot rolling of steel, are also responsible for the fracture of the surface oxide layer present on aluminum strip in rolling. Fresh metal then extrudes through the microcracks formed in the oxide layer (Ref 113). This cracking process is essential to allow the boundary additive in the lubricant to react with the exposed fresh metal surface and produce organometallic compounds. An innovative technique based on the measurement of contact electrical resistance in plane-strain compression tests has been reported, looking at how metal-to-metal contact is established, whether the oil can enter the microcracks in the oxide film, and how fast the transfer film develops (Ref 114, 115). It has been shown that metal-to-metal contact is established after the oxide breaks and metal extrudes through the oxide. More viscous base oil inhibits the establishment of direct metal-to-metal contact and the corresponding fall of electrical resistance, perhaps due to oil trapped in microcracks in the oxide layer. The addition of lauric acid in a low-viscosity base oil led to the development of a low-friction transfer film, affecting the friction conditions in the area, giving rise to metal-to-metal contact at a smaller reduction and fluctuations in the electrical resistance during the tests (Fig. 9).

Effect of Lubrication

The introduction of lubrication has several objectives. Lubricants help to reduce the loads, resulting in lower energy expenditure. They aid in the production of high-quality surfaces, resulting in a more valuable product by reducing the incidence of surface defects. These are achieved by controlling the friction and heat transfer. The nature of the lubrication in the contact zone also affects the dimensional consistency of the product.

Introducing lubricants into the contact zone increases the complexity of the events at the interface. The viscosity of the lubricant and its pressure and temperature sensitivity will affect the interactions as will the volume of lubricant entering the contact zone. As the relative velocity is increased, more lubricant is dragged in to coat the surfaces, and the coefficient of friction decreases under most circumstances with most lubricants. As the pressure increases, so does the viscosity, and the coefficient of friction falls. A power law relating the viscosity and the coefficient of friction was suggested in the mid-1980s (Ref 5):

$$\mu = f(\eta^a) \qquad \text{(Eq 22)}$$

where η is the viscosity coefficient, and a is a negative number in all cases, varying from -0.2 to -0.5. Many authors have since reported corrected relationships, some of which are summarized elsewhere (Ref 43). The viscosity depends on the process parameters and is affected mostly by the temperature and the pressure. A popular relationship, for instance, is:

$$\eta = \eta_o \exp(\gamma p - \vartheta pT - \delta T) \qquad \text{(Eq 23)}$$

where η_o is the uncorrected viscosity, γ is the pressure and δ the temperature viscosity coefficients, while τ is the cross coefficient accounting for the interaction between γ and δ (Ref 116). Correcting the viscosity for these effects is a necessary but not easy step. A good collection of specific values of the pressure-viscosity coefficient for some automotive lubricants is presented elsewhere (Ref 117, 118). For mineral oils, the viscosity-pressure coefficient is given as:

$$\gamma \approx (0.6 + 0.965\log_{10}\eta_o) \times 10^{-8} \qquad \text{(Eq 24)}$$

where the viscosity at zero pressure, η_o, is in centipoises, and the viscosity-pressure coefficient is in Pa^{-1} (Ref 119).

Emulsions. Use of an emulsion brings with it further parameters. Emulsions are composed of oil and water in which the droplets of oil, with diameters ranging from 1 to 100 pm, are dispersed along with an emulsifier to prevent the droplets from coalescing. Emulsifiers are composed of a molecular structure with two distinct ends. The hydrophilic end is made of polar covalent bonds and is soluble in water. The lipophilic end is soluble in oils. When the emulsion is formed, the hydrophilic groups orient toward the water phase, and the lipophilic hydrocarbon groups orient toward the oil phase.

The droplet size, its standard variation within the spray, its concentration, and the emulsifier and its concentration are all involved in the process and determination of the nature of the mechanisms. These mechanisms are the droplet capture, which depends on the adhesion between the droplets and the moving surfaces; plate-out, which forms a uniform oil pool at the entrance to the contact zone; and the dynamic concentration theory. It has been shown that an increase in oil concentration leads to an increase of the film thickness and that the lubrication mechanisms are very sensitive to speed (Ref 120). The fundamental problem in the use of emulsions is the behavior of the oil particles, the capture of which by the entering strip or by the roll surfaces is not yet fully understood (Ref 121).

Lubrication Regimes. Several lubrication regimes are observed. In rolling, for instance, it is possible to determine the nature of the lubricating regime by comparing the thickness of the oil film and the combined asperity heights of the rolls and the rolled metal. Their ratio can be defined as:

$$\lambda_{\text{ratio}} = \frac{h_{\min}}{\sigma*} \qquad \text{(Eq 25)}$$

where $h_{\min}$ is the oil film thickness, and $\sigma*$ is the combined roughness of the two surfaces (Ref 122). When the oil-film-thickness-to-surface-roughness ratio is less than unity, boundary lubrication is present. When $1 \leq \lambda_{\text{ratio}} \leq 3$, mixed lubrication prevails, while, for a ratio over three, hydrodynamic conditions and full separation of the contacting surfaces prevail.

The lubrication regimes may be illustrated by making reference to the Stribeck curve, first plotted for describing friction conditions in the bearings of rail car wheels. In the Stribeck curve, the coefficient of friction is plotted against the modified Sommerfeld number, S, defined as:

$$S = \frac{\eta \nu_{\text{rel}}}{p} \qquad \text{(Eq 26)}$$

where η is the dynamic viscosity, $\Delta\nu$ is the relative velocity, and p is the pressure. A schematic diagram of the Stribeck curve is shown in Fig. 10, with the identified boundary, mixed, and hydrodynamic mode of lubrication.

Boundary lubrication is characterized by significant amounts of metal-to-metal contact and some lubricating pockets. The local friction and metal-to-metal contacts depend on the quality of the molecular bonding to the surfaces by the lubricant. This is temperature sensitive, meaning that boundary lubrication is usually ineffective in hot rolling, even for aluminum, which is usually hot rolled in the 300 to 500 °C range. The thickness of the film is low, and the asperities pierce through the lubricant. As the viscosity and/or the relative velocity increase, a mixed mode of lubrication is observed in which more lubricant pockets and

fewer asperity contacts are found. In these regimes, the surface roughness of the resulting product decreases as a consequence of the contact and approaches that of the die. The hydrodynamic regime follows, with complete separation of the surfaces. The lubricant film is thicker than the combined surface roughness of the die and plastically deformed workpiece. Plastic deformation of the surface under a fluid film causes roughening. Further subdivision of the hydrodynamic regime is possible by identifying elastohydrodynamic or plastohydrodynamic lubrication, depending on the deformation of the asperities and their resistance to deformation. The events in the contact zone are largely influenced by entrainment of liquid lubricants. In general, more lubricant is carried into the roll gap as the relative speed increases (Ref 123). Increasing viscosity results in hydrodynamic escape, joining increasing speed and larger frictional resistance to increase the tendency toward hydrostatic behavior (Ref 124). When the lubricant is within the deformation zone, its effect on the rolling process is dependent on the directionality of the roll roughness. Randomly oriented roughness leads to the lowest loads on the rolling mill because less lubricant leaks away (Ref 125). Lubricant composition has a significant impact on the tribological conditions; hence, different additives are used for different materials. For instance, mineral seal oil with appropriate additives is useful when cold rolling aluminum. Palm oil in an emulsion is used when steel is cold rolled. Mineral seal oil in water emulsion may be used when hot rolling aluminum. Synthetic oil in water emulsion may be used when hot rolling steel, although environmental concerns often limit its use, and plain water is commonly used instead. Emulsifiers and additives are used as required (Ref 5, 117, 118, 126). Control of lubrication is the main barrier to improving productivity and surface quality (Ref 127).

Sheet metal forming of tribologically difficult materials, such as stainless steel, aluminum alloys and titanium alloys, or forming in tribologically difficult operations, such as ironing, punching, or deep drawing of thick plate, often requires the use of environmentally hazardous lubricants such as chlorinated paraffin oils to avoid galling. A system of tests has been developed for laboratory testing of friction and lubrication in sheet metal forming that especially focuses on appropriate simulation of the actual conditions in production (Fig. 11) (Ref 128).

This test system has proven appropriate to evaluate a large number of lubricants in a broad variety of sheet-forming processes, thus allowing a speedy selection of appropriate lubricants for production. Production tests of the same lubrication systems have been run in parallel to prove that the laboratory tests work satisfactorily in simulating the production conditions. This work resulted in more environmentally friendly lubricants being found that could replace the hazardous ones in sheet-forming
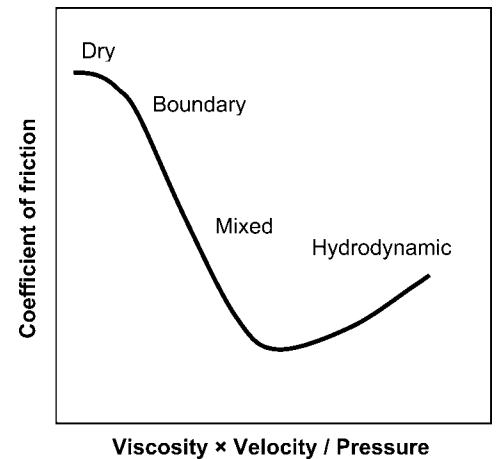

Fig. 10 Schematic representation of the Stribeck curve

operations on mild steel as well as stainless steel.

To create ultrafine-grained steels, hot rolling with a high reduction and a high rolling force is necessary, leading to the development of innovative hot rolling oils. Data for the coefficient of friction under the wide range of tribological conditions have been obtained in order to understand the lubrication mechanism in hot rolling of these materials (Ref 129).

Effect of Process and Material Parameters on Interfacial Phenomena

Every tribological system in metal forming involves a large number of parameters, the interaction of which determines the success of a deformation process. For instance, in flat rolling the most important parameters are the reduction, speed, temperature, oxide scaling, and surface roughness, while the material parameters are the resistance to deformation, surface hardness, and anisotropy. The effects of some of these on the roll/strip interface phenomena are considered next.

The coefficient of friction is dependent on the loads acting normal to the surface. The exact nature of that dependence is not clear but is connected to the attributes of the contacting materials, their elastic and plastic strength, roughness, relative velocity, and so on. It is the interaction of these parameters that will determine the frictional behavior of the contacting materials. In rolling, for instance, as the reduction is increased, the loads on the rolled metal and thus the roll pressures increase. As a consequence, the asperities are flattened, and the real area of contact approaches the apparent area at a rate that depends on the elastic and plastic strengths of the contacting metals. The number of adhesive bonds formed between the two surfaces also increases. The strength of these bonds depends on the two materials, including their chemical affinity for each other.
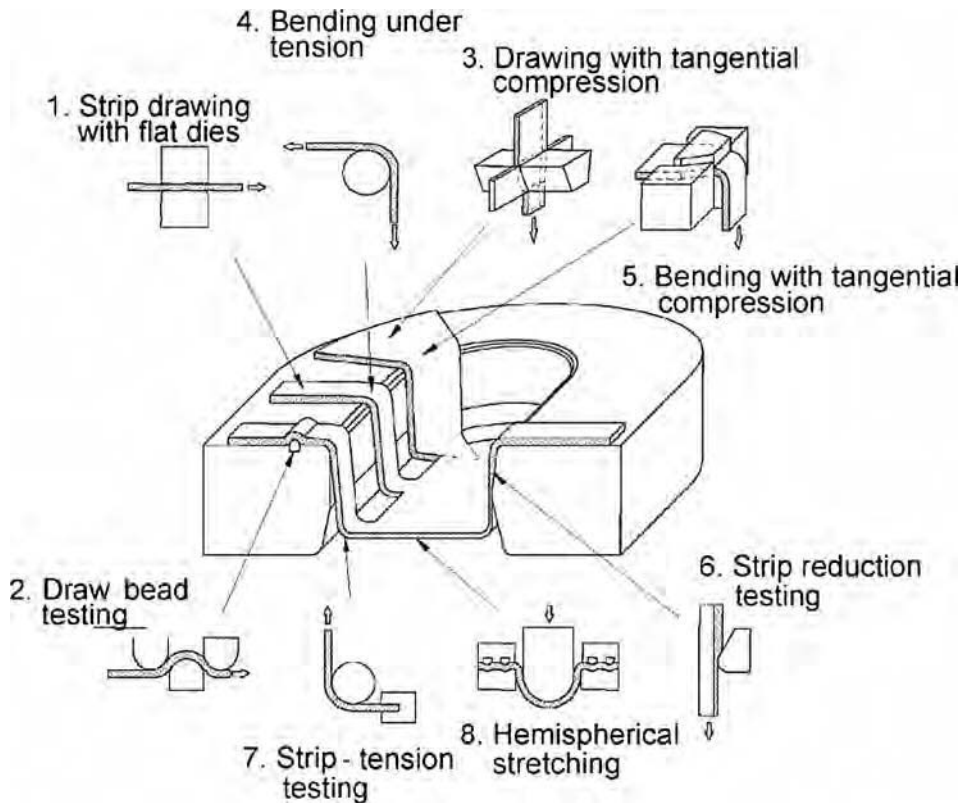
Fig. 11 Schematic representation of tribological simulative tests for sheet metal forming. Source: Ref 128

The frictional resistance is likely to be increased for dry and clean surfaces. The roughness of the rolled surface is reduced with the roll surface imprinted on the metal. Introducing lubricants into the contact surface changes the reaction of the rolled metal to the reduction, and the number of operating mechanisms also increases. These involve the composition of the oil, the presence of antifriction and extreme-pressure additives, the oil viscosity, and its viscosity-pressure and viscosity-temperature coefficients. The main mechanisms and their potential interconnections are presented in Fig. 12 (Ref 43).

The coefficient of friction drops as the velocity increases under the boundary- and mixed-lubrication regimes, while frictional resistance increases with speed under hydrodynamic lubrication caused by the increasing frictional resistance within the layer of oil separating the surfaces (Ref 130). There are several mechanisms responsible for the velocity dependence of frictional resistance. The first is the increase of the metal resistance to deformation when the strain rate is increased. Another is related to the time available for the adhesion of the contacting asperities. The increase of oil volume, drawn into the deformation zone, also affects the frictional phenomena. The roughness of the roll and its direction determine the amount of the oil entering the contact zone and its distribution.

The effect on interface phenomena of the temperature at the contacting surfaces is the least researched area because of the difficulties associated with its measurement. Temperature measurement at the center of a rolled sample using embedded thermocouples is the common technique, and a mathematical model is then required to estimate the temperature at the surface. The model would also require an accurate heat-transfer coefficient in the contact zone, which introduces another level of complexity. Comprehensive data are obtained using the ring compression test for determining the friction factor for a number of materials and bulk temperatures (Ref 38). Care must be taken to distinguish the influence of varying heat-transfer conditions on the ring deformation from the effect of friction, because both are of similar magnitude (Ref 131). The effect of the temperature on the magnitude of the coefficient of friction in hot rolling should not be separated from other phenomena. Chemical composition, scale breakers, time in the furnace, and so on should be taken into account when the coefficient is chosen for modeling.

Although general tribology opinion is that, except in the case of very low and very high values, friction is independent of the surface roughness (Ref 117, 132), in metal forming the surface roughness has an effect on friction. Evidence that the magnitude of the roughness of the surface and especially its direction affect frictional resistance is evident (Ref 125), with Fig. 13 illustrating that the lowest roll forces during Al-Mg-Cu strip rolling, lubricated with a mineral oil, are obtained when the roughness direction is random.

Adverse lubricating conditions were found while cold rolling thin steel and aluminum strips when the roughness direction was parallel to the direction of rolling (Ref 133). There were also observations that friction is defined by surface conditions, and the load-carrying ability is increased when the roughness decreases (Ref 134, 135). It was also found that increasing the oil film thickness or decreasing the roughness lead to lower friction (Ref 136).

Microforming and Size Effects Related to the Tool-Workpiece Interface

Microforming is an appropriate technology to manufacture very small metal parts, in particular for bulk production, because they are required in many industrial products resulting from micro-technology. The manufacture of microformed metal parts always involves size effects. These size effects make the downscaling of macroscale processes even more difficult, because the production parameters cannot simply be changed according to the rules of similarity. The sources of the size effects can be divided into physical and structural sources (Ref 137). Among physical sources directly related to the tool-workpiece interface is the surface-to-volume size ratio (SVS). This ratio is relevant for all effects that are dominated by surface effects. As the part size decreases, the ratio of surface to volume increases, which makes the surface effects more pronounced. The occurrence of size effects due to the SVS depends on whether the whole or only a fraction of the surface participates in the process. Another physical source is the forces-relation size effect. There are different forces that act on a part during deformation processing besides the processing force, namely, van der Waals force, surface tension, and gravitation. These forces are relatively small and can be neglected in conventional forming of macroscale parts. However, in microforming these forces must be taken into account, because they are relatively large when compared with the process forces. The relationship between these forces and process force, or between two of these forces, can lead to size effects. The surface structure scalability size effect should be named among structural sources of the size effects related to the interface. Like the grain size of the metal, the surface structure is the result of the history of the part. It is influenced by cutting processes and surface roughening, flattening, or intended structuring during forming processes. The result of these treatments is also influenced by the grain size, because it affects elementary processes such as grain tilting in forming or elastic springback in cutting operations. In principle, the surface roughness is independent of the part size. It is often not possible to reduce the surface roughness along with the part dimension, thereby increasing the influence of surface roughness. The microscale contact remains unchanged, while the macroscale
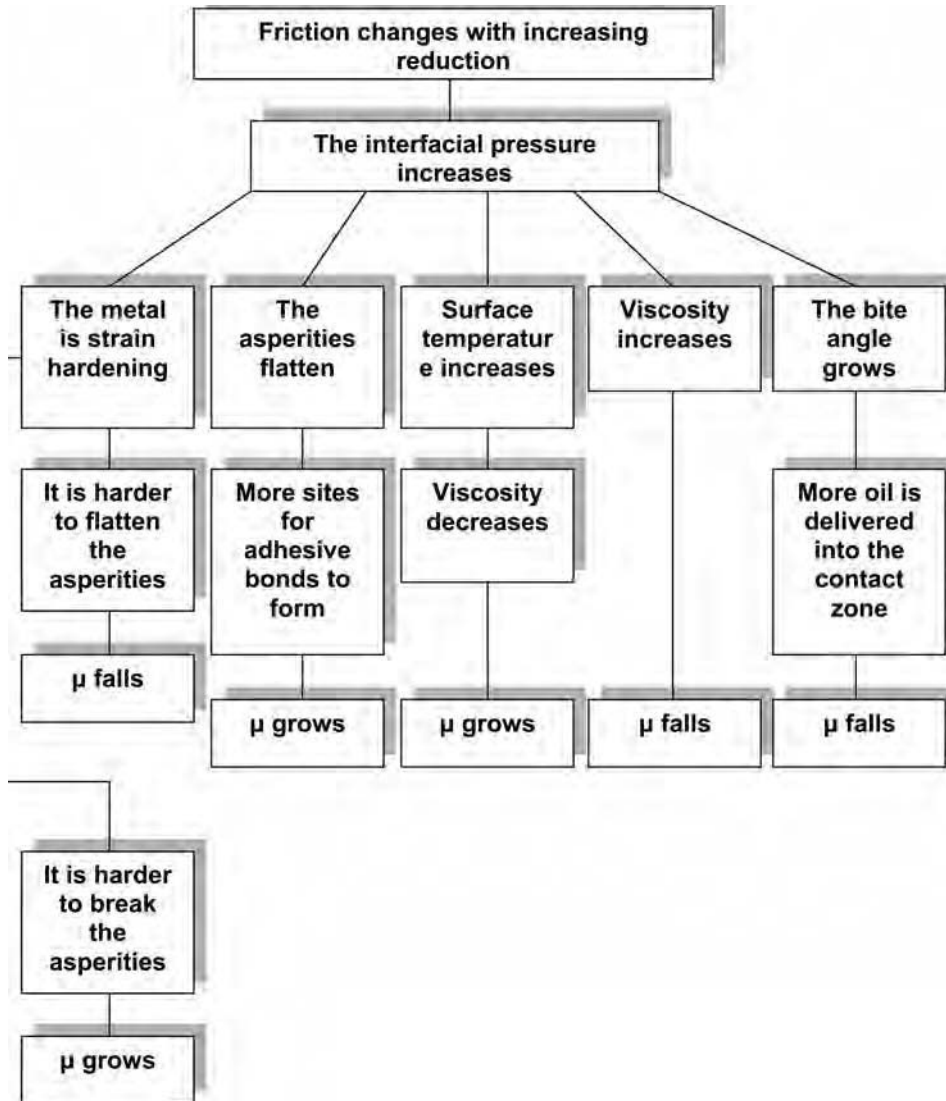
Fig. 12 Schematic representation of the competing mechanisms affecting friction in the flat rolling process. Source: Adapted from Ref 43

contact is changed due to miniaturization. Particularly under lubricated conditions, this leads to size-dependent friction behavior.

Deep Drawing. There are several parameters that affect the deep drawing process. The most important is friction at the flange and at the radius of the die. This friction is also affected by size effects when transferring the forming technology from macro- to microforming. A new friction test method has been developed to investigate the friction behavior within deep drawing (Ref 138, 139). The friction coefficients at the flange and at the die radius were calculated from the punch force and the punch stroke of the deep drawing process. Instead of blank deep drawing, strip drawing was applied for this investigation, so that the tangential pressure at the flange area is avoided. The results of the friction coefficient calculation for microscale strip drawing are presented in Fig. 14. The results consistently show that friction is strongly dependent on pressure, being highest at low pressure.

Experimental results show a significant increase in friction, up to a factor of 20, with decreasing specimen size when using extrusion oil as the lubricant (Ref 140). The amount of changes in friction conditions was determined during double-cup extrusion testing supported by numerical identification and finite-element simulation. It was concluded that the so-called open and closed lubricant pockets are responsible for the size effect (Fig. 15). The peaks at the rough surface are flattened, and the oil is held in the valleys between them during the extrusion. When the valley is located at the border of the part, the oil can escape and no lubrication is given.

The existence of open and closed lubricant pockets could be inferred from measurement of the roughness of the macroscopic part after the extrusion. The roughness at the border areas was significantly less than the roughness of the middle area (Ref 141, 142).

When scaling down forming processes, the contact area is reduced significantly, while the size of single-topography features remains roughly constant. The ratio of closed lubricant pockets is also reduced, and the external load will be transmitted more and more by the real contact area, which will increase friction. Additionally, the total number of asperities is reduced. At present, nanotopography is being considered for characterizing friction in microforming applications. The impact of the measured nanotopography on the friction conditions, in combination with liquid lubrication, has been investigated during flattening experiments with idealized asperities coupled with finite-element simulation of the loaded contact state (Ref 143). The next step is detailed investigation of the influence of the microstructure of the workpiece material on the formation of the nanotopography. Detailed knowledge about the evolution of surface topography is essential for characterizing the workpiece surface in terms of its frictional behavior in microforming applications.

An alternative approach for modeling friction on the tool surface involves the creation of an empirical relationship between the friction coefficient and the tool roughness profile (Ref 144). The effects of load and metal flow have been considered, and a geometric description of the tool interface with zero friction has been developed, based on surface measurements (Ref 56). The tool surface during the ring compression test was modeled with a sinusoidal profile and zero friction in this approach (Fig. 16).

In the ring compression test, the same change of inner radius with reduction in height between the conventional friction factor with a flat surface profile and the given geometric model under conditions of zero friction is used to determine the equivalent roughness geometry for a particular friction factor. It can be used for correlation with the real surface of a tool by direct measurement of the die surface. This model can be incorporated into simulation software for a potentially more realistic interpretation of the friction effect in terms of flow and load.

Application of the atomic force microscope as a research tool to study the adhesion, friction, and lubrication phenomena is becoming important. The tip of the microscope can be considered as a single asperity and is able to measure forces as low as 10^{-9} N. The data obtained by the atomic force microscope for the mechanical properties of a thin film of gold were reported as early as 1993 (Ref 145). The wear properties of organic lubricants have been measured (Ref 146), and friction forces at the atomic level have been shown to depend partly on sound energy (Ref 147). However, the potential of the atomic force microscope for characterizing friction on the microscale and in microforming applications is still largely unexploited.
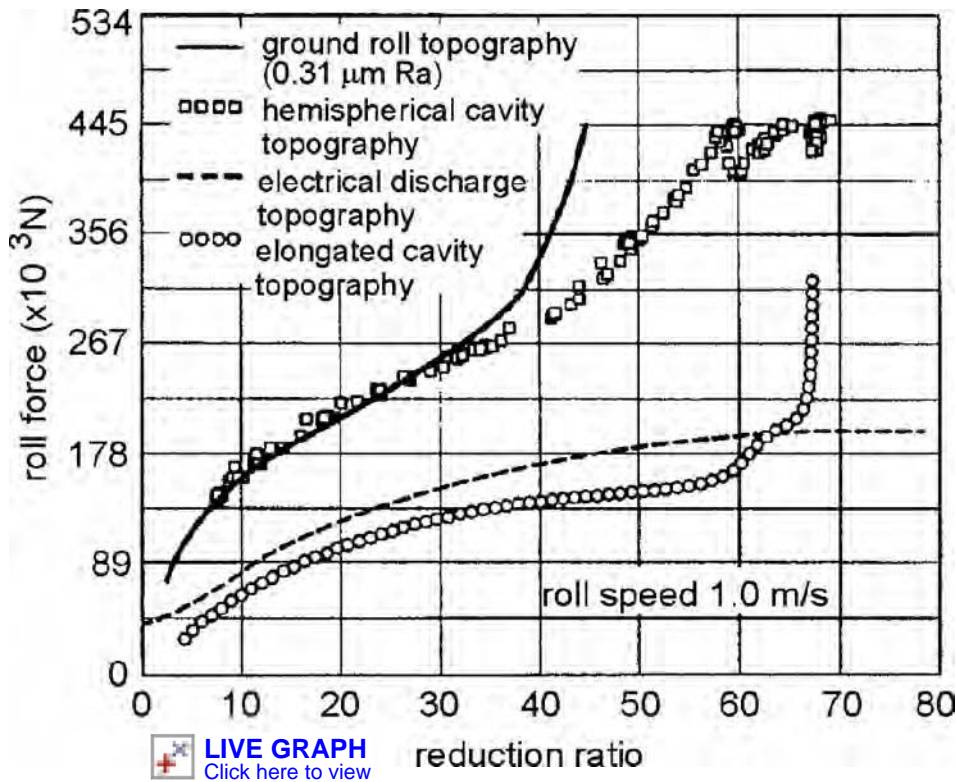
LIVE GRAPH
Click here to view

Fig. 13 Influence of surface roughness on roll forces for different reductions. Note that the roll separating force increases with reduction at a much lower rate when the roll surface direction is random. Source: Ref 125
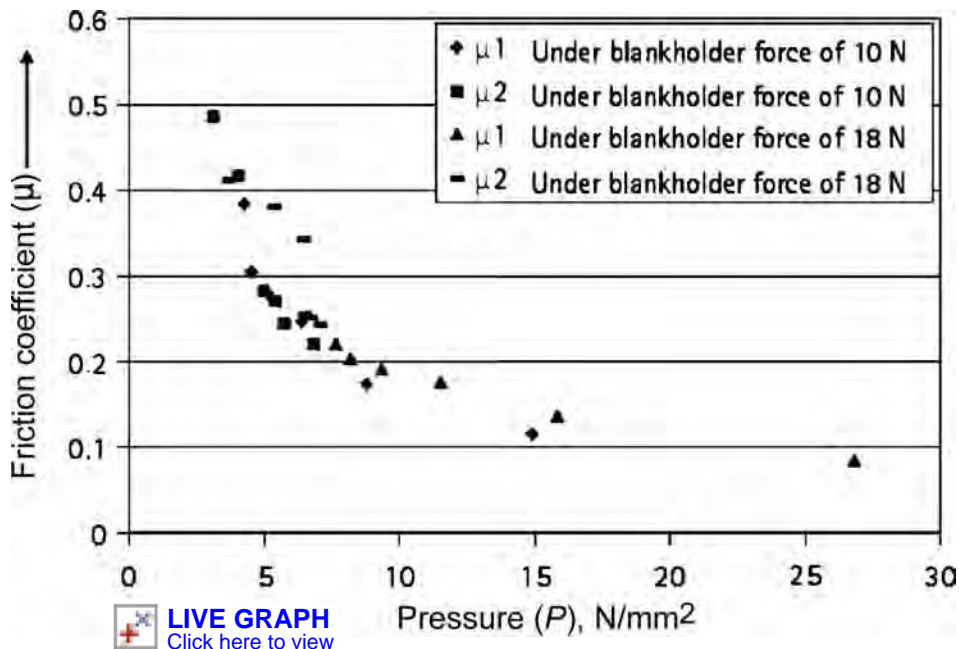

LIVE GRAPH
Click here to view

Fig. 14 Friction coefficients calculated for microscale strip drawing of E-Cu58 alloy with 0.02 mm thickness, 10 mm width, and 1.8 mm length; the punch length is 1 mm. Source: Ref 139

REFERENCES

1. J.A. Schey, *Metal Deformation Processes: Friction and Lubrication*, Marcel Dekker, New York, 1970

2. A.T. Male and M.C. Cockcroft, A Method for the Determination of the Coefficient of Friction under Conditions of Bulk Plastic Deformation, *J. Inst. Met.*, Vol 93, 1964, p 38–46

3. R. Wagoner, W. Wang, and S. Sriram, Development of OSU Formability Test and OSU Friction Test, *J. Mater. Process. Technol.*, Vol 45, 1994, p 13–18

4. F. Delamare, M. de Vathaire, and J. Kubie, An Evaluation of the PSCT. II. Experimental Study of the Friction Test: Role of the Transfer Layer in Boundary Lubrication, *J. Lub. Technol. (Trans. ASME)*, Vol 104, 1982, p 545–551

5. J.A. Schey, *Tribology in Metalworking, Friction, Lubrication and Wear*, The American Society for Metals, 1983

6. C.D. Roberts, Mechanical Principles of Rolling, *Iron Steelmaker*, Vol 24, 1997, p 113–114

7. W.Y.D. Yuen, Y. Popelianski, and M. Prouten, Variations of Friction in the Roll Bite and Their Effects on Cold Strip Rolling, *Iron Steelmaker*, Vol 23, 1996, p 33–39

8. J.G. Lenard, An Examination of the Coefficient of Friction, *Metal Forming Science and Practice*, J.G. Lenard, Ed., Elsevier, 2002, p 85–114

9. W.L. Roberts, *Hot Rolling of Steel*, Marcel Dekker, Inc., New York, 1983

10. Z. Wusatowski, *Fundamentals of Rolling*, Pergamon Press, Oxford, 1969

11. G.W. Rowe, *Principles of Industrial Metalworking Processes*, Edward Arnold, London, 1977

12. E. Felder, Modes D'Usure et D'Endommagement des Cylindres de Laminage a Chaud. Relations Avec les Caracteristiques Thermomecaniques des Calamines, *Rev. Métall.*, Vol 12, 1984, p 931–942

13. A. Munther and J.G. Lenard, A Study of Friction during Hot Rolling of Steels, *Scand. J. Metall.*, Vol 26, 1997, p 231–240

14. Y.H. Li and C.M. Sellars, Modelling Deformation Behaviour of Oxide Scales and Their Effects on Interfacial Heat Transfer and Friction during Hot Steel Rolling, *Second Int. Conf. on Modelling of Metal Rolling Processes, Proceedings*, J.H. Beynon, P. Ingham, H. Teichert, and K. Waterson, Ed., The Institute of Materials, London, U.K., 1996, p 192–206

15. A.K.E.H.A. El-Kalay and L.G.M. Sparling, Factors Affecting Friction and Their Effect upon Load, Torque, and Spread in Hot Flat Rolling, *J. Iron Steel Inst.*, Vol 206 (No. 2), 1968, p 152–163

16. N. Rudkins and P. Evans, Mathematical Modelling of Mill Set-Up in Hot Strip Rolling of High Strength Steels, *J. Mater. Process. Technol.*, Vol 80–81, 1998, p 320–324

17. A.N. Karagiozis and J.G. Lenard, The Effect of Material Properties on the Coefficient of Friction in Cold Rolling, *EURO-TRIB'95, Proceedings* (Lyon), 1985, p 1–7

18. J.G. Lenard and Z. Malinowski, Measurements of Friction during Warm Rolling of Aluminium, *J. Mater. Process. Technol.*, Vol 39, 1993, p 357–371
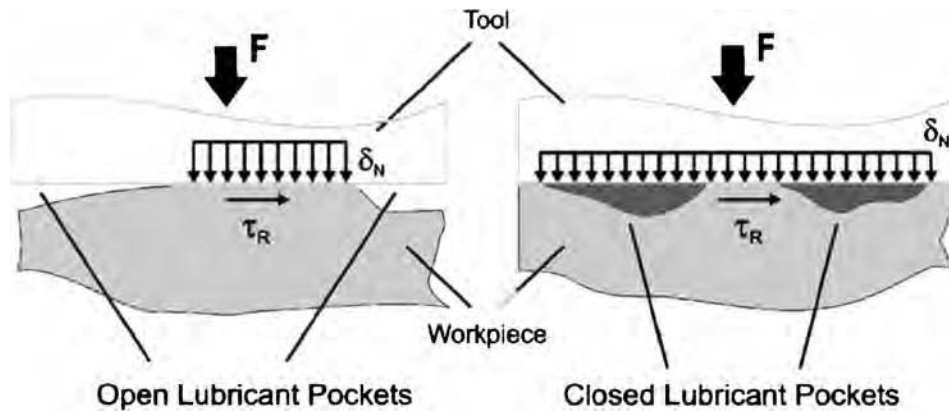
Fig. 15 Schematic representation of the open and closed lubricant pockets formed at the tool-workpiece interface during the double-cup extrusion test. Source: Ref 142
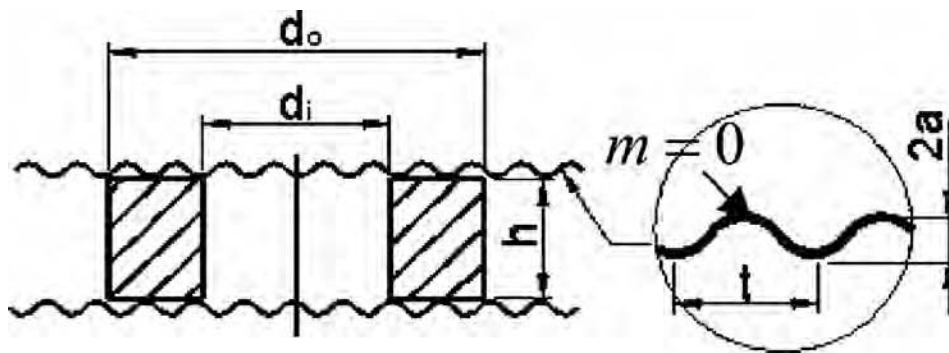

Fig. 16 Schematic representation of the tool surface model using an elliptical profile. Source: Ref 145

19. L.S. Lim and J.G. Lenard, Study of Friction in Cold Strip Rolling, *ASME J. Eng. Mater. Technol.*, Vol 106, 1984, p 139–146
20. J.G. Lenard, An Experimental Study of Boundary Conditions in Hot and Cold Flat Rolling, *Annals of the CIRP*, Vol 39, 1990, p 279–282
21. J.G. Lenard, Measurements of Friction and in Cold Flat Rolling, *J. Mater. Shaping Technol.*, Vol 9, 1991, p 171–180
22. T. Yoneyama and Y. Hatamura, Measurement of Actual Stress and Temperature on the Roll Surface during Rolling, *JSME Int.*, Vol 32, 1989, p 113–117
23. A. Banerji and W.B. Rice, Experimental Determination of Normal Pressure and Friction Stress in the Rolling Gap during Cold Rolling, *Ann. CIRP*, Vol 21, 1972, p 53–54
24. J. Jeswiet, A Friction Sensor for Sheet Metal Rolling, *Ann. CIRP*, Vol 40, 1991, p 231–234
25. J. Jeswiet, P. Wild, and H. Sefton, Sensing Friction: Methods and Devices, *International Workshop on Friction and Flowstress in Cutting and Forming* (Paris), 2000, p 111–118
26. B. Hum, H.W. Colquhoun, and J.G. Lenard, Measurements of Friction during Hot Rolling of Aluminium Alloys, *J. Mater. Process. Technol.*, Vol 60, 1996, p 331–338

27. O. Pawelski, Ein Neues Gerät zum Messen des Reibungsbeiwertes bei Plastischen Formänderungen, *Stahl Eisen*, Vol 84 (Part 2), 1964, p 1233–1243
28. A. Nagamatsu, T. Murota, and T. Jimma, On the Non-Uniform Deformation of Block in Plane-Strain Compression Caused by Friction, *JSME Bull.*, Vol 13, 1970, p 1396–1405
29. B.G. Hansen and N. Bay, Two New Methods for Testing Lubricants for Cold Forging, *J. Mech. Work. Technol.*, Vol 13, 1986, p 189–204
30. P.S. Theocaris, C.A. Stassinakis, and A.G. Mamalis, Roll-Pressure Distribution and Coefficient of Friction in Hot Rolling by Caustics, *Int. J. Mech. Sci.*, Vol 11, 1983, p 833–844
31. P.S. Theocaris, A.G. Mamalis, and C.A. Stassinakis, On the Pressure Distribution and Coefficient of Friction Between Stock and Rolls in Ingot Hot Rolling, *Adv. Technol. Plasticity*, Vol 1, 1984, p 321–326
32. W. Schroeder and D.A. Webster, Press-Forging Thin Sections: Effect of Friction, Area and Thickness on Pressures Required, *ASME J. Appl. Mech.*, Vol 16, 1949, p 289–294
33. W. Evans and B. Avitzur, Measurement of Friction in Drawing, Extrusion and Rolling, *ASME J. Lub. Technol.*, Vol 90, 1968, p 72–88
34. D. Wang, Methods for Measurement of the Coefficient of Friction in Plastic Deformation, *The Third Forging Annual Meeting* (Xi'an, China), Chinese Society of Mechanical Engineering, 1983
35. W.L. Roberts, *Hot Rolling of Steel*, Marcel Dekker, Inc., New York, 1983
36. B. Avitzur and R. Kohser, Disk and Strip Forging for the Determination of Friction and Flow Strength Values, *ASLE Trans.*, Vol 21, 1978, p 143–151
37. J.S. Gunasekera and S. Mahadeva, An Alternative Method to Predict the Effects of Friction in Metal Forming, *Friction and Material Characterisation*, MD-Vol 10, ASME, 1988, p 55–62
38. A.T. Male and M.G. Cockroft, A Method for Determination of the Coefficient of Friction of Metals under Conditions of Bulk Plastic Deformation, *J. Inst. Met.*, Vol 93, 1964, p 38–46
39. V. DePierre and F. Gurney, A Method for Determination of Constant and Varying Friction Factors during Ring Compression Tests, *ASME J. Lubr. Technol.*, Vol 96, 1974, p 482–488
40. X. Tan, P.A.F. Martins, N. Bay, and W. Zhang, Friction Studies at Different Normal Pressures with Alternative Ring Compression Tests, *Seventh Int. Metal Forming 98, Birmingham, Proceedings*, P. Hartley, I. Pillinger, F.R. Hall, C.E. Neal-Sturgess, M. Pietrzyk, and J. Kusiak, Ed., 1998, p 292–297
41. L.R. Sanchez, K.J. Weinmann, and J.M. Story, A Friction Test for Extrusion Based on Combined Forward and Backward Flow, *NAMRC XIII, Proceedings*, 1985, p 110–117
42. M.E. Popilek, K.J. Weinmann, and S.A. Majlessi, A Friction Test, Based on Combined Backward Can–Forward Bar Extrusion with Emphasis on Backward Flow, Trans. *NAMRI*, Vol XX, 1992, p 25–31
43. J.G. Lenard, Tribology in Metal Rolling, *CIRP Annals—Manuf. Technol.*, Vol 49 (No. 2), 2000, p 567–590
44. G.-Y. Li, Analytical Equations for On-Line Modelling of Friction in Cold Rolling, *Modelling of Metal Rolling Processes 3, Proceedings*, J.H. Beynon, M.T. Clark, P. Ingham, P. Kern, and K. Waterson, Ed. (London), IOM Communications, 1999, p 370–379
45. J.F. Lin, T.K. Huang, and C.T. Hsu, Evaluation of Lubricants in Cold Strip Rolling, *Wear*, Vol 147, 1991, p 79–91
46. J. Kusiak, M. Pietrzyk, and J.G. Lenard, Application of FE Simulation of the Compression Test to the Evaluation of Constitutive Equations for Steels at Elevated Temperatures, *NUMIFORM' 95, Proceedings*, S.F. Shen and P.R. Dawson, Ed. (Ithaca), 1995, p 277–282

47. Z. Malinowski, J. Kusiak, and M. Pietrzyk, Application of the Inverse Techniques to the Experimental Tests, *Application of Mathematical Methods in Science and Technique, Proceedings*, Vol 2 (Krakow), 1995, p 43–52

48. S. Khoddam, Y.C. Lam, and P.F. Thomson, Numerical Results Achieved with an Inverse Computational Method for Determining the Constitutive Parameters Using the Hot Torsion Test Results, *Steel Res.*, Vol 67, 1996, p 39–43

49. R. Kopp and F.-D. Philipp, Physical Parameters and Boundary Conditions for the Numerical Solution of Hot Forming Processes, *Steel Res.*, Vol 63, 1992, p 392–398

50. P. Montmitonnet, Measurement and Modelling of Friction in Cold Rolling Processes, *Second Int. Symp. on Fuels and Lubricants—ISFL 2000, Proceedings*, Vol II, S.P. Srivastava, Ed., March 10–12, 2000 (New Delhi), Allied Publishers Ltd., New Delhi, 2000, p 503–508

51. J. Jeswiet, A Comparison of Friction Coefficients in Cold Rolling, *J. Mater. Process. Technol.*, Vol 80–81, 1998, p 239–244

52. Q. Zhang, E. Felder, M. Arentoft, N. Paldan, L. Dubar, F. Borsetto, and S. Bruschi, Results of Friction Benchmark in Cold Forging Conditions, *12th Int. Conf. Metal Forming 2008, Proceedings*, Special Edition, Vol 2, Sept 21–24, 2008 (Krakow, Poland), Steel Research International, 2008, p 781–788

53. P.R. Burte, S.L. Semiatin, and T. Altan, An Investigation of the Heat Transfer and Friction in Hot Forging of 304 Stainless and Ti-6A1-4V, *Trans. NAMRI/SME*, 1990, p 59–66

54. F. Vollertsen, H. Schulze Niehoff, and Z. Hu, State of the Art in Micro Forming, *Int. J. Mach. Tools Manuf.*, Vol 46, 2006, p 1172–1179

55. M. Geiger, M. Kleiner, R. Eckstein, N. Tiesler, and U. Engel, Microforming, *Ann. CIRP*, Vol 50 (No. 2), 2001, p 445–462

56. J. Jeon and A. Bramley, A Friction Model for Microforming, *Int. J. Adv. Manuf. Technol.*, Vol 33, 2007, p 125–129

57. P. Montmitonnet, E. Wey, F. Delamare, J.L. Chenot, C. Fromholz, and M. de Vathaire, A Mechanical Model of Cold Rolling. Influence of the Friction Law on Roll Flattening Calculated by a Finite Element Method, *Fourth Int. Steel Rolling Conf., Proceedings*, June 1987 (Deauville), IRSID/ATS, 1987

58. P. Montmitonnet and P. Buessler, A Review on Theoretical Analyses of Rolling in Europe, *ISIJ Int.*, Vol 31 (No. 6), 1991, p 525–538

59. J.H. Beynon, Modelling of Friction and Heat Transfer during Metal Forming, *KomPlasTech'99*, Jan 17–20, 1999 (Szczyrk, Poland), 1999, p 47–54

60. B.K. Chen, P.F. Thomson, and S.K. Choi, Temperature Distribution in the Roll-Gap during Hot Flat Rolling, *J. Mater. Process. Technol.*, Vol 30, 1992, p 115–130

61. S.P. Timothy, H.L. Yiu, J.M. Fine, and R.A. Ricks, Simulations of Single Pass of Hot Rolling Deformation of Aluminium Alloy by Plane Strain Compression, *Mater. Sci. Technol.*, Vol 7, 1991, p 255–261

62. S.L. Semiatin, E.W. Collings, V.E. Wood, and T. Altan, Determination of the Interface Heat Transfer Coefficient for Non-Isothermal Bulk Forming Process, *Trans. ASME, J. Eng. Ind.*, Vol 109, 1987, p 49–57

63. M. Pietrzyk and J.G. Lenard, A Study of Thermal Mechanical Modelling of Hot Flat Rolling, *J. Mater. Shaping Technol.*, Vol 7, 1989, p 117–126

64. C.O. Hlady, I.V. Samarasekera, E.B. Hawbolt, and J.K. Brimacombe, Heat Transfer in the Hot Rolling of Aluminium Alloys, *Int. Symp. on Light Metals Processing and Applications; 32nd Annual Conf. of Metallurgists CIMM*, C. Bickert et al., Ed., 1993, p 511–522

65. C.O. Hlady, J.K. Brimacombe, I.V. Samarasekera, and E.B. Hawbolt, Heat Transfer in the Hot Rolling of Metals, *Met. Mater. Trans. B.*, Vol 26B, 1995, p 1019–1027

66. Z. Malinowski, J.G. Lenard, and M.E. Davies, A Study of Heat Transfer Coefficient as a Function of Temperature and Pressure, *J. Mater. Process. Technol.*, Vol 41, 1994, p 125–142

67. M. Pietrzyk and J.G. Lenard, A Study of Boundary Conditions in Hot/Cold Flat Rolling, *Int. Conf. Computational Plasticity: Models, Software and Applications, Proceedings*, D.R.J. Owen et al., Ed. (Wales), Pineridge Press, 1989, p 947–958

68. W.C. Chen, I.V. Samarasekera, and E.B. Hawbolt, Characterisation of the Thermal Field during Rolling of Microalloyed Steels, *33rd Mechanical Working and Steel Processing Conf., Proceedings*, Vol XXIX (USA), Iron and Steel Soc., 1992, p 349–357

69. P.G. Stevens, K.P. Ivens, and P. Harper, Increasing Work-Roll Life by Improved Roll-Cooling Practice, *J. Iron Steel Inst.*, Vol 209, 1971, p 1–11

70. K. Murata, H. Morise, M. Mitsutsuka, H. Haito, T. Kumatsu, and S. Shida, Heat Transfer between Metals in Contact and Its Application to Protection of Rolls, *Trans. ISIJ*, Vol 24, 1984, p B309

71. C.M. Sellars, Computer Modelling of Hot Working Processes, *Mater. Sci. Technol.*, Vol 1, 1985, p 325–332

72. W.C. Chen, I.V. Samarasekera, A. Kumar, and E.B. Hawbolt, Mathematical Modeling of Heat Flow and Deformation during Rough Rolling, *Ironmaking Steelmaking*, Vol 20 (No. 2), p 113–125

73. T. Wanheim and N. Bay, A Model for Friction in Metal Forming Processes, *Ann. CIRP*, Vol 27 (No. 1), 1978, p 189–193

74. C. Devadas, I.V. Samarasekera, and E.B. Hawbolt, The Thermal and Metallurgical State of Steel Strip during Hot Rolling: Part 1, Characterization of Heat Transfer, *Metall. Trans. A.*, Vol 22A. 1991, p 307–319

75. W.C. Chen, I.V. Samarasekera, and E.B. Hawbolt, Fundamental Phenomena Governing Heat Transfer during Rolling, *Metall. Trans. A.* Vol 24A. 1993, p 1307–1320

76. J. Ball, J.A. Treverton, and M.C. Thornton, Evaluation of the Effects of Stresses in Hot Rolling Mills on Oxide Films on Aluminium, *J. Soc. Trib. Lub. Eng.*, Vol 50, 1994, p 89–93

77. Y.H. Li and C.M. Sellars, Comparative Investigations of Interfacial Heat Transfer Behaviour during Hot Forging and Rolling of Steel with Oxide Scale Formation, *J. Mater. Process. Technol.*, Vol 80–81, 1998, p 282–286

78. Y.H. Li and C.M. Sellars, Modelling Deformation Behaviour of Oxide Scales and Their Effects on Interfacial Heat Transfer and Friction during Hot Steel Rolling, *Second Int. Conf. on Modelling of Metal Rolling Processes, Proceedings*, J.H. Beynon, P. Ingham, H. Teichert, and K. Waterson, Ed. (London, U.K.), The Institute of Materials, 1996, p 192–206

79. M.G. Cooper, B.B. Mikic, and M.M. Yovanovich, Thermal Contact Resistance, *Int. Heat Mass Transf.*, Vol 12, 1969, p 279–300

80. M. Bahrami, J.R. Culham, M.M. Yovanovich, and G.E. Schneider, Thermal Contact Resistance of Nonconforming Rough Surfaces, Part 1: Contact Mechanics Model, *J. Thermophys. Heat Transf.*, Vol 18 (No. 2), 2004, p 209–217

81. M. Williamson and A. Majumdar, Effect of Surface Deformations on Contact Conductance, *Trans. ASME. J. Heat Transf.*, Vol 114, Nov 1992, p 802–809

82. J. Pullen and J.B.P. Williamson, O. the Plastic Contact of Rough Surfaces, *Proc. R. Soc. London*, Ser. A. Vol A327, 1972, p 159–173

83. B.B. Mikic, Thermal Contact Conductance: Theoretical Considerations, *Int. J. Heat Mass Transf.*, Vol 17, 1974, p 205–214

84. G.V. Samsonov, *The Oxide Handbook*, IFI/Plenum, New York, 1973

85. M. Pietrzyk and J.G. Lenard, The Effect of the Temperature Rise of the Roll on the Simulation of the Flat Rolling Process, *J. Mater. Process. Technol.*, Vol 22, 1990, p 177–190

86. M. Pietrzyk, H. Kusiak, J.G. Lenard, and Z. Malinowski, Heat Exchange between the Workpiece and the Tool in Metal Forming Processes, *Conf. Formability'94,*

Proceedings (Ostrava, Czech Republic), 1994, p 329–338

87. P. Dadras and W.R. Wells, Heat Transfer Aspects of Non-Isothermal Axisymmetric UPSE. Forging, *ASM. J. Eng. Ind.*, Vol 106, 1984, p 187–195

88. C.R. Boër and G. Schröder, Temperature in the Die-Billet Zone in Forging, *Ann. CIRP*, Vol 30 (No. 1), 1991, p 152–157

89. O. Pawelski, W. Rasp, and C. Hoerster, Ring Compression Test as Simulation Test for the Investigation of Friction in Hot Metal Forming, *Steel Res.*, Vol 60 (No. 9), 1989, p 395–402

90. J. Jeswiet, W. Nshama, and P.H. Oosthuizen, Evaluation of Temperature and Heat Transfer Conditions in Metal Forming, *11th Int. Heat Transfer Conference: Heat Transfer 1998, Proceedings*, Vol 7, Aug 23–28, 1998 (Kyongju, Korea), 1998, p 205–209

91. C.R. Boër, H. Rydstad, and G. Schröder, Choosing Optimal Forging Conditions in Isothermal and Hot-Die Forging, *J. Appl. Metalwork.*, Vol 3 (No. 4), 1985, p 421–431

92. P.R. Burte, Y. Im, T. Altan, and S.L. Semiatin, Measurement and Analysis of Heat Transfer and Friction during Hot Forging, *Trans. ASME*, Vol 112, Nov 1990, p 332–339

93. W. Nshama and J. Jeswiet, Evaluation of Temperature and Heat Transfer Conditions at the Metal-Forming Interface, *Ann. CIRP*, Vol 44 (No.1), 1995, p 201–204

94. J. Jeswiet and S. Zou, A Multi-Point Temperature Sensor for Metal Rolling, *Ann. CIRP*, Vol 41 (No. 1), 1992, p 299–302

95. P.G. Stevens, K.P. Ivens, and P. Harper, Increasing Work-Roll Life by Improving Roll Cooling Practice, *J. Iron Steel Inst.*, Vol 209, 1971, p 1–11

96. A. Silvonen, A. Malinen, and A.S. Korhonen, A Finite Element Study of Plane Strain Hot Rolling, *Scand. J. Metal.*, Vol 16, 1987, p 103–108

97. C.C. Chang and A.N. Bramley, Determination of the Heat Transfer Coefficient at the Workpiece-Die Interface for the Forging Process, *Proc. Inst. Mech. Eng., Part B—J. Eng. Manuf.*, Vol 216 (No. 8), 2002, p 1179–1186

98. M. Rosochowska, K. Chodnikiewicz, and R. Balendra, A New Method of Measuring Thermal Contact Conductance, *J. Mater. Process. Technol.*, Vol 145, 2004, p 207–214

99. R.E. Smelser and E.G. Thompson, Advances in Inelastic Analysis, *Winter Annual Meeting of the ASME*, S. Nakazawa, K. Willam, and N. Rebelo, Ed., AMP-88, PED-28, Boston, MA. 1987, p 273–278

100. V.K. Jain, Determination of Heat Transfer Coefficient for Forging Applications, *J. Mater. Shaping Technol.*, Vol 8, 1990, p 193–202

101. O. Pawelski, Berechnung der Wärmedurchgangsztahl für das Warmwalzen und Schmieden, *Arch. Eisenhüttenwes.*, Vol 40, 1969, p 821–827

102. M. Krzyzanowski and J.H. Beynon, Modelling the Behaviour of Oxide Scale in Hot Rolling, *ISIJ. Int.*, Vol 46 (No. 11), 2006, p 1533–1547

103. M. Krzyzanowski and J.H. Beynon, The Tensile Failure of Mild Steel Oxides under Hot Rolling Conditions, *Steel Res.*, Vol 70 (Issue 1/99), 1999, p 22–27

104. K.S. Tan, M. Krzyzanowski, and J.H. Beynon, Effect of Steel Composition on Failure of Oxide Scales in Tension under Hot Rolling Conditions, *Steel Res.*, Vol 72, (No. 7), 2001, p 250–258

105. M. Krzyzanowski and J.H. Beynon, Modelling the Boundary Conditions for Thermo-Mechanical Processing—Oxide Scale Behaviour and Composition Effects, *Model. Simul. Mater. Sci. Eng.*, Vol 8, 2000, p 927–945

106. M. Krzyzanowski, J.H. Beynon, and C.M. Sellars, Analysis of Secondary Oxide-Scale Failure at Entry into the Roll Gap, *Metall. Mater. Trans.*, Vol 31B. 2000, p 1483–1490

107. M. Krzyzanowski, P. Suwanpinij, and J.H. Beynon, Analysis of Crack Development, Both Growth and Closure, in Steel Oxide Scale under Hot Compression, *Materials Processing and Design: Modelling, Simulation and Applications, Numiform 2004, Proceedings*, Vol 712, S. Ghosh, J.C. Castro, and J.K. Lee, Ed., American Institute of Physics, Melville, NY. 2004, p 1961–1966

108. M. Trull and J.H. Beynon, High Temperature Tension Tests and Oxide Scale Failure, *Mater. Sci. Technol.*, Vol 19, 2003, p 749–755.

109. M. Krzyzanowski and J.H. Beynon, Effect of Oxide Scale Failure in Hot Steel Rolling on Subsequent Hydraulic Descaling: Numerical Simulation, *Third Int. Conf. Hydraulic Descaling, Proceedings*, Sept 14–15, 2000 (London, U.K.), IO. Communications Ltd., 2000, p 77–86

110. M. Krzyzanowski, W. Yang, C.M. Sellars, and J.H. Beynon, Analysis of Mechanical Descaling: Experimental and Modelling Approach, *Mater. Sci. Technol.*, Vol 19, 2003, p 109–116

111. C. Fedorciuc-Onisa and D.C.J. Farrugia, Simulation of Frictional Conditions during Long Product Hot Rolling, *Sixth ESAFOR. Conference on Material Forming, Proceedings*, V. Brucato, Ed. (Salermo, Italy), 2003, p 763–766

112. C. Fedorciuc-Onisa and D.C.J. Farrugia, Through Process Characterisation of Frictional Conditions for Long Product Hot Rolling, *Steel Grips*, Vol 2, 2004, p 331–336

113. H.R. Le, M.P.F. Sutcliffe, P.Z. Wang, and G.T. Burstein, Surface Oxide Fracture in Cold Aluminium Rolling, *Acta Mater.*, Vol 52, 2004, p 911–920

114. H.R. Le. and M.P.F. Sutcliffe, Finite Element Modelling of the Evolution of Surface Pits in Metal Forming, *J. Mater. Process. Technol.*, Vol 145 (No. 3), 2004, p 391–396

115. H.R. Le, M.P.F. Sutcliffe, P. Wang, and G.T. Burstein, Surface Generation and Boundary Lubrication in Bulk Forming of Aluminium Alloy, *Wear*, Vol 258, 2005, p 1567–1576

116. C.-Y. Sa and W.R.D. Wilson, Full Film Lubrication of Strip Rolling, *ASM. J. Tribol.*, Vol 116, 1994, p 569–576

117. E.R. Booser, *Handbook of Lubrication*, Vol II. CR. Press, Boca Raton, FL. 1984

118. E.R. Booser, *Tribology Data Handbook*, CR. Press, Boca Raton, FL. 1997

119. I.M. Hutchings, *Tribology: Friction and Wear of Engineering Materials*, CR. Press, Boca Raton, FL. 1992

120. S.R. Schmid and W.R.D. Wilson, Lubrication of Aluminum Rolling by Oil-in-Water Emulsions, *Tribol. Trans.*, Vol 38, 1995, p 452–458

121. A. Kumar, S.R. Schmid, and W.R.D. Wilson, Particle Behaviour in Two-Phased Lubrication, *Wear*, Vol 206, 1997, p 130–135

122. I.M. Hutchings, *Tribology: Friction and Wear of Engineering Materials*, CR. Press, Boca Raton, FL. 1992

123. O. Pawelski, W. Rasp, and S. Draese, Influence of Hydrodynamic Lubricant Entrainment on Friction Effects in Cold Rolling, *Steel Res.*, Vol 65, 1994, p 488–494

124. J. Bech, N. Bay, and M. Eriksen, Entrapment and Escape of Liquid Lubricant in Metal Forming, *Eighth Int. Conf. Tribology, Proceedings*, S.S. Eskildsen, D.S. Larsen, H. Reitz, E.J. Bienk, and C.A. Straede, Ed. (Aarhus), 1998, p 713–719

125. S. Sheu, L.G. Hector, and O. Richmond, Tool Surface Topographies for Controlling Friction and Wear in Metal Forming Processes, *ASM. J. Tribol.*, Vol 120, 1998, p 517–527

126. J. Laemmle, Metalworking Lubricants, *Friction, Lubrication, and Wear Technology*, Vol 18, *ASM Handbook*, ASM International, 1992

127. W.R.D. Wilson, Lubrication in Metal Forming, *Lubricants and Lubrication*, D. Dowson, Ed., Elsevier Science BV. 1995

128. N. Bay, D.D. Olsson, and J.L. Andreasen, Lubricant Test Methods for Sheet Metal Forming, *Tribol. Int.*, Vol 41, 2008, p 844–853

129. A. Azushima, W.D. Xue, and Y. Yoshida, Lubrication Mechanism in Hot Rolling by Newly Developed Simulation Testing Machine, *Ann. CIRP*, Vol 56 (No. 1), 2007, p 297–300

130. S. Zhang and J.G. Lenard, Reduction of the Roll Force during Lubricated Cold

Rolling of Aluminum Strips, *J. Synth. Lubr.*, Vol 12, 1996, p 303–321

131. J.D. Fletcher, Y.H. Li, J.H. Beynon, and C.M. Sellars, The Influence of Surface Conditions in Hot Forming Determined by Ring Upsetting: A Numerical and Experimental Investigation, *Proc. Inst. Mech. Eng., Part J. J. Eng. Tribol.*, Vol 212, 1998, p 453–465

132. E. Rabinowicz, *Friction and Wear of Materials*, 2nd ed., John Wiley and Sons, Inc., New York, 1995

133. S.S. Lu and Y.H. Chuang, Effects of Surface Roughness on Lubrication in the Cold Rolling of Metals, *ASM. J. Tribol.*, Vol 107, 1985, p 522–526

134. A. Azarkhin and O. Richmond, Limits to Adhesive Friction, *Numerical Methods in Industrial Forming Processes, Proceedings*, J.-L. Chenot, R.D. Wood, and O.C. Zienkiewicz, Ed., Balkema, 1992, p 143–148

135. S. Andersson and E. Salas-Russo, The Influence of Surface Roughness and Oil Viscosity on the Transition in Mixed Lubricated Sliding Steel Contacts, *Wear*, Vol 174, 1994, p 71–79

136. P.E. Tabary, M.P.F. Sutcliffe, F. Porral, and P. Deneuville, Measurements of Friction in Cold Rolling of Metals, *ASM. J. Tribol.*, Vol 118, 1996, p 629–636

137. F. Vollertsen, Size Effects in Manufacturing, *Process Scaling*, Vol 24, F. Vollertsen and F. Hollmann, Ed., BIA. Bremen, Verlag, 2003, p 1–9

138. F. Vollertsen, Z. Hu, H. Schulze Niehoff, and C. Theiler, State of the Art in Microforming and Investigations in Micro Deep Drawing, *J. Mater. Process. Technol.*, Vol 757, 2004, p 70–79

139. Z. Hu and F. Vollertsen, Friction Test for Deep Drawing with Respect to Size-Effects, *First International Conference on New Forming Technology, Proceedings* (China), Harbin Institute of Technology Press, 2004, p 153–158

140. N. Tiesler and U. Engel, Microforming—Effects of Miniaturisation, *Eighth International Conference on Metal Forming—Metal Forming 2000, Proceedings*, M. Pietrzyk et al., Ed., Sept 3–7, 2000 (Krakow, Poland), Balkema, Rotterdam, 2000, p 355–360

141. M. Geiger, M. Kleiner, R. Eckstein, N. Tiesler, and U. Engel, Microforming, *51st General Assembly of CIRP*, Vol 50 (No. 2), Aug 19–25, 2001 (Nancy, France), 2001, p 445–462

142. N. Tiesler, U. Engel, and M. Geiger, Basic Research on Cold Forging of Microparts, *Seventh ICTP. Proceedings*, Vol 1, Oct 27–Nov 1, 2002 (Yokohama, Japan), 2002, p 379–384

143. S. Weidel and U. Engel, Characterisation of the Flattening Behaviour of Modelled Asperities, *Wear*, Vol 266, 2009, p 596–599

144. P. Becker, H.J. Jeon, C.C. Chang, and A.N. Bramley, A Geometric Approach to Modelling Friction in Metal Forming, *Ann. CIRP*, Vol 52 (No. 1), 2003, p 209–212

145. M. Salmeron, Use of the Atomic Force Microscope to Study Mechanical Properties of Lubricant Layers, *MR. Bull.*, Vol XVIII. 1993, p 20–25

146. R. Overney and E. Meyer, Tribological Investigations Using Friction Force Microscopy, *MR. Bull.*, Vol XVIII. 1993, p 26–34

147. J. Krim, Friction at the Atomic Scale, *Lubr. Eng.*, Vol 53, 1997, p 8–13

Heat-Transfer Interface Effects for Solidification Processes

P.A. Kobryn, *Air Force Research Laboratory*

A KEY ASPECT OF SOLIDIFICATION process modeling is the treatment of the interface between the solidifying casting and the mold in which it is contained. The solidification rate of a casting is governed by the rate of heat extraction, which in turn is dominated by the rate of heat transfer across the casting-mold interface (although heat flux through the casting and the mold and heat transfer from the mold to the environment also play a role). Hence, solidification process modeling tools must be able to treat the casting-mold interface in a manner that appropriately captures the relevant interface heat-transfer phenomena.

Solidification process modeling tools typically model heat transfer within the casting and within the mold by numerically solving the transient, nonlinear heat conduction energy equation as a function of time, given user-specified thermophysical material properties, initial conditions, and boundary conditions (Ref 1). Heat transfer across the casting-mold interface is generally treated by applying a heat-flux boundary condition to the interface. The heat flux, q, across the interface is defined as:

$$q = h(T_2 - T_1)$$

in which h is the interface heat-transfer coefficient, and $(T_2 - T_1)$ is the temperature drop across the interface (Fig. 1). Hence, the interface region is not explicitly modeled but instead is treated as a zero-volume surface that behaves as a resistor to heat flux.

This interface heat-flux boundary condition is an empirical simplification of the actual phenomena of interface heat transfer. It is employed because the physical mechanisms of interface heat transfer, which are described in the following section, are too complex to be simulated in an efficient manner, given the current state of modeling and simulation technology. Use of this boundary condition requires the interface heat-transfer coefficient to be specified by the user during preprocessing. Typical solidification process modeling tools allow h to be defined as a constant value or as a function of time, temperature, and/or interface pressure/gap size (Ref 2).

The proper selection of h is critical for obtaining accurate simulation results. Unfortunately, the true value of h is highly dependent on the specific characteristics of the interface, which vary with time and with casting configuration. Additionally, the value of h cannot be calculated a priori and cannot be explicitly measured experimentally. Instead, analytical or numerical techniques must be used to infer the value of h for a given casting arrangement, using temperature measurements from the casting and the mold during solidification.

An in-depth discussion of casting-mold interface heat-transfer phenomena is provided in the following section. The subsequent sections contain discussions of practical considerations and methods for incorporating h into models, quantifying h experimentally, and selecting h for a given casting configuration. Finally, select results from the literature are described to guide the solidification modeler in determining appropriate values of h for specific casting configurations.

Casting-Mold Interface Heat-Transfer Phenomena

Although the casting-mold interface commonly is assumed to be a zero-volume surface for process modeling, it is actually an interface region with finite volume and three-dimensional characteristics. The specific behavior of the interface region is governed by a combination of complex phenomena, including heat transfer, fluid flow, solidification, deformation, solid-state phase transformation, and chemical reaction. To understand the evolution of heat-transfer conditions across the interface as solidification progresses, it is helpful to consider the physical characteristics of this interface region over time.

When the casting is in the liquid state, the extent of physical contact between the casting and the mold is high and the thickness of the interface region is small, which allows a high rate of heat transfer via conduction through those portions of the casting and the mold that
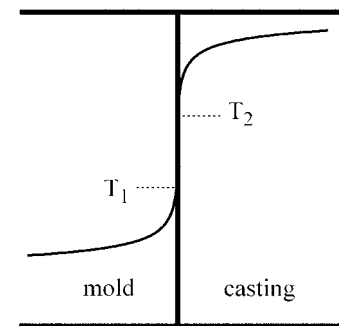

Fig. 1 Temperature drop across the casting-mold interface

comprise the interface. As the liquid cools and solidification begins, asperities form on the surface of the casting, and the contact between the casting and the mold degrades. Heat transfer still occurs through conduction at the points of contact, but the contact area is smaller and the conduction distance is greater; concomitantly, the rate of heat transfer is lower.

As solidification proceeds, local conditions often cause the casting to shrink away from the mold surface, thereby forming a physical gap between the casting and the mold, for example, when the casting contracts as it cools (Fig. 2) (Ref 3, 4). This gap is likely to be filled with a gas either from the ambient environment, from outgassing of the mold, or from a chemical reaction between the casting and the mold surface. At this point, heat transfer occurs via conduction through and/or radiation across the gap when the gap is small and conduction, convection, and radiation as the gap becomes larger. The rate of interface heat transfer in this gap-forming scenario is usually much lower than the initial rate due to the typically low values of thermal conductivity exhibited by gases, the increased thickness of the interface region, the decreasing temperature of the casting surface (which greatly reduces radiation heat transfer), and the limited space available for convection (Fig. 3a).
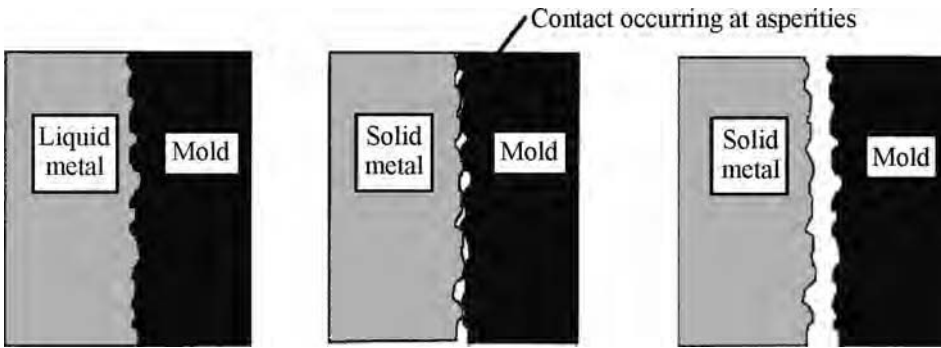
Fig. 2 Formation of a gap as the casting shrinks away from the mold. Source: Ref 4
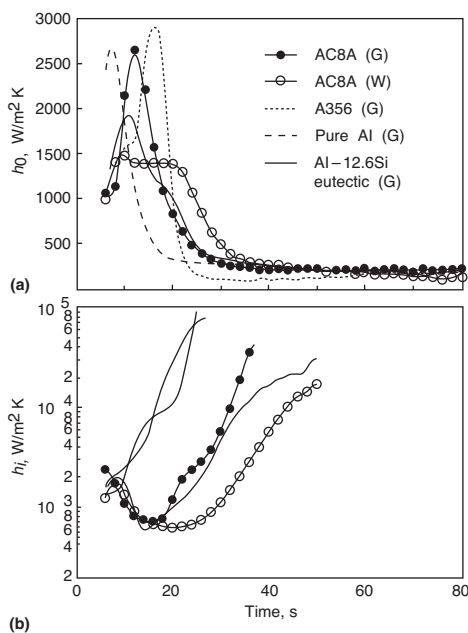

Fig. 3 Experimentally determined values of $h(t)$ for interfaces in a tube-shaped casting of various aluminum alloys. AC8A is an aluminum-silicon (plus copper or magnesium) alloy, JIS H5205, similar to UNS A03360. (a) Outer interface. (b) Inner casting-mold interface. Source: Ref 5

LIVE GRAPH
Click here to view

Alternatively, local conditions can cause the casting to move toward the mold surface as solidification proceeds, for example, if there is a mold core in the casting (Ref 5, 6). In this case, the contact pressure between the casting and the mold at the interface increases, and the thickness of the interface region becomes smaller while the contact area becomes larger; hence, the rate of heat transfer is increased. In certain cases, the effect of this phenomenon on the value of the heat-transfer coefficient can be quite large (Fig. 3b).

The behavior of the casting-mold interface is influenced by many factors related to the casting and mold materials, the casting and mold geometries, and the selected casting process (Ref 7). For example, interface contact is impacted by gravity, metallostatic head, volumetric expansion/contraction during liquid-solid and solid-state phase transformations, thermal expansion/contraction, viscosity/flow stress of the casting, mold-wetting characteristics, and the surface finish of the mold. Additionally, the presence of reaction products, entrained gases, mold washes, lubricants or other films or coatings, and the composition of the ambient atmosphere can impact the behavior of the interface. The thermal properties of the materials present at the interface also play a role in determining the rate of heat transfer across the interface.

Incorporating the Interface Heat-Transfer Coefficient in Models

As discussed earlier, the complexity of the interface heat-transfer process makes explicit modeling of the interface region generally infeasible and therefore forces modelers to employ simplified methods for treating heat transfer across the interface. The use of a heat-flux boundary condition at the interface is widely accepted by the solidification modeling community and requires the user to define the value of the interface heat-transfer coefficient, h. The extent to which h is defined as a function of time, temperature, and/or interface pressure is governed by the capability of the selected solidification modeling software, the required fidelity of model predictions, the complexity of the casting process and casting geometry, and the availability of appropriate experimental input data.

Most commercial solidification modeling packages allow the user to define h as a constant, a function of time, a function of casting temperature at the interface, or a function of mold temperature at the interface. Although contact condition (including contact pressure) is a key factor in determining the rate of heat transfer across the interface, most commercial codes are either incapable of modeling contact stresses/gap formation, lack appropriate material property data to accurately model interface contact conditions, or add an unacceptable amount of computation time if mechanical stress/strain is modeled. Hence, modelers generally choose to define h as a function of casting surface temperature alone, $h(T)$. Various researchers have shown this choice to be more accurate than the other commonly available options (Ref 8–11). This approach works reasonably well when the required model fidelity is not overly high, the casting geometry is relatively simple, the casting process itself is not overly sensitive to the value of h, and experimentally determined values of h as a function of casting surface temperature are available for the selected casting system. That is, the systems have the same or very similar casting alloy, mold material, general casting shape/size, and casting environment.

As the required model fidelity becomes greater and/or the geometric complexity of the casting increases, the appropriateness of defining h as a function of casting surface temperature alone becomes questionable. In such cases, local interface conditions drive local variations in h at a given casting surface temperature and therefore drive error into the model results. This error can be reduced either by modeling contact conditions/gap formation and defining h as a function of both casting surface temperature and contact pressure/gap size (or of gap size alone), or by intelligently preselecting appropriate values of h as a function of casting surface temperature for geometrically differing portions (segments) of the interface. The capability of commercial casting simulation packages to couple thermal and mechanical phenomena has greatly improved over the last decade, making the modeling of gap formation increasingly feasible. However, the availability of appropriate mechanical input data is still quite limited.

The availability of experimentally determined values of h also drives the modeler's choice of how to define h. As described subsequently, many researchers have developed experimental and analytical/numerical procedures for determining h as a function of time, interface contact pressure, and/or gap size for various casting geometries, casting alloys, mold materials, mold coatings, and casting processes. (Reference 7 and the Appendix of Ref 12 provide a reasonably comprehensive list of these publications.) However, few researchers have published results that show the variation of h with casting surface temperature, and no readily accessible publication has provided values of h as a function of both casting surface temperature and contact condition/gap size (Ref 8–11, 13, 14). Additionally, a limited number of casting systems have been studied, resulting in a limited availability of data. Hence, the modeler is often left to either perform system-specific experiments and analyses to determine appropriate values of h or to estimate values of h from nonrepresentative published data.

Quantifying the Interface Heat-Transfer Coefficient Experimentally

Researchers face many challenges in attempting to quantify the value of h. As stated earlier,

h cannot be calculated a priori from material properties and geometric parameters, and *h* cannot be measured directly via physical experiment. Instead, *h* generally is inferred by measuring the transient temperature distributions (temperature as a function of time, $T(t)$, at multiple locations) within the mold and the casting experimentally and then using the resultant $T(t)$ data to determine *h* using one or more analytical or numerical modeling/estimation/optimization techniques. Accuracy can be impacted by various factors related to the numerical analysis techniques, the casting system being modeled, and the experimental apparatus, but the ability to accurately obtain $T(t)$ data from experiment is the most challenging factor to overcome.

Experimental Challenges. The process of collecting $T(t)$ data from points within the casting and the mold through experiment is not trivial. Ideally, one would measure the temperatures precisely at the casting and mold surfaces that comprise the interface as well as at multiple locations near the interface in both the casting and the mold throughout the casting process. Such measurement would require the precise placement of near-zero latency (that is, very small time lag between a real change in temperature and the detection of that change by the sensor); near-zero thermal mass; and high-accuracy temperature sensors at locations in the mold, at the mold surface, and within the mold cavity prior to initiating the experiment. Unfortunately,

such sensors do not exist, so most experimentalists resort to using thermocouples. Figure 4 shows a typical test arrangement.

The use of thermocouples is problematic because they have non-zero latency, non-zero thermal mass, and non-zero innate error, and they require lead wires (and often a cavity through which the lead wires can pass) to connect to the data acquisition system. Hence, they cannot capture extremely rapid transients, and they often have a nonnegligible effect on the thermal field in which they are embedded. They are also difficult to embed at precise locations within the mold (particularly in sand molds) and are even more difficult to locate precisely within the mold cavity (the volume into which the molten alloy will be poured to make the casting). Furthermore, it is difficult to ensure that the thermocouples remain in place as the casting is poured; as the casting solidifies, cools, and contracts; and as the mold heats up and expands. In fact, it is possible for a gap to form between the casting or mold and the thermocouple, resulting in inaccurate temperature measurements (Ref 15). Finally, the limited temperature ranges and survivability of thermocouples in high-melting-point, reactive alloys impede the ability to collect complete temperature data inside the casting. Titanium castings are an example (Ref 15). A detailed review of experimental considerations related to the use of thermocouples to capture

temperature histories during casting can be found in Ref 10 and 14.

Characterizing and controlling variability in the casting experiments is also a concern. Variations in parameters such as initial mold temperature, pouring temperature, and melt volume and composition can be characterized and/or controlled fairly accurately, but variations in pouring conditions, thermocouple placement/movement, and mold composition and geometry (in the case of sand or investment casting, for instance) can be very difficult to characterize and control. The level of variability will depend greatly on the choices made during the design and execution of the experiment. Hence, it is very important to anticipate, minimize, and understand experimental variability when planning and conducting interface heat-transfer experiments and to adjust $T(t)$ results (and/or error bars) accordingly.

In extreme casting conditions (such as in reactive-alloy die casting), the reliability of thermocouple data can be extremely low. In such cases, certain researchers have complemented thermocouple readings with the use of mold microstructure signature analysis (Ref 15). The mold microstructure signature analysis relies on destructive inspection of the microstructure of the mold to infer local temperature maxima based on well-characterized liquid- and solid-state reactions, phase transformations, and annealing/aging phenomena (Fig. 5–7). The timing and shape of the temperature transients are captured by the thermocouples, while the local temperature maxima are obtained from the postcasting mold microstructure characterization.

Gap Size. In addition to measuring temperature, many researchers have also chosen to measure gap width in an attempt to correlate *h* to gap size (Ref 3, 4, 16). In recent times, these researchers have selected linear variable differential transformers (LVDT) to measure gap size over time (Fig. 8, 9). Other techniques have also been employed. The gap-size measurement complicates the experiments but enables a much more informed analysis. As with thermocouples, consideration must be given to errors associated with the use of LVDTs (or whatever gap-measurement sensor is employed) when planning the experiment and interpreting the results.

Analytical Challenges. One approach to determining *h* from experimental $T(t)$ data is to solve the inverse heat conduction problem for *h* (*t*). Various methods for solving this ill-posed, nonlinear inverse problem can be found in the literature, with the nonlinear estimation method proposed by Beck being the method most commonly applied (Ref 17). Alternatively, one can assume a functional form of *h* as a function of time, casting surface temperature, and other parameters and iteratively determine the best-fit values of *h* using a direct calibration-curve method, that is, by matching the temperatures predicted via solidification modeling to those determined experimentally (Fig. 10, 11) (Ref
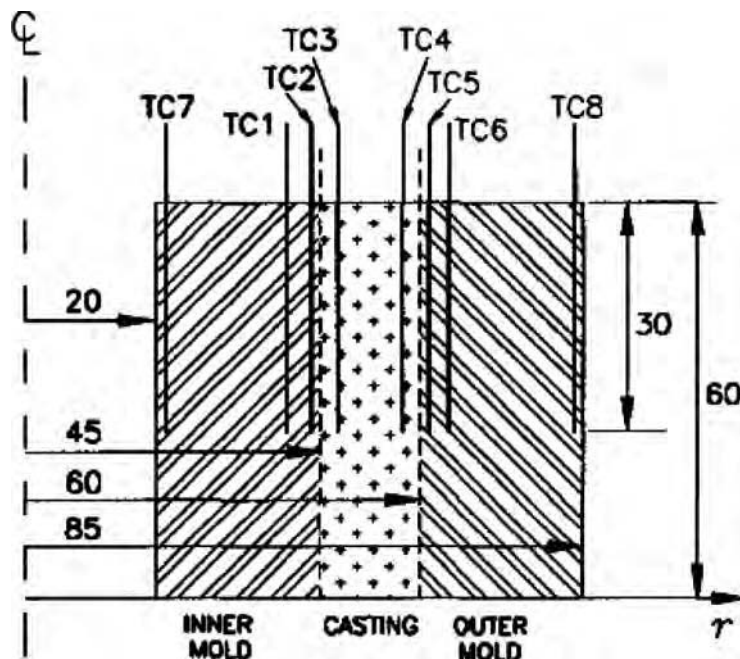
TC	Distance from center	TC	Distance from center
TC7	22 mm	TC4	56 mm
TC1	40 mm	TC5	62 mm
TC2	43 mm	TC6	65 mm
TC3	49 mm	TC8	83 mm

Fig. 4 Selection of thermocouple (TC) locations in the casting and mold for determining *h* at outer and inner casting-mold interfaces in a tube-shaped casting. Source: Ref 5
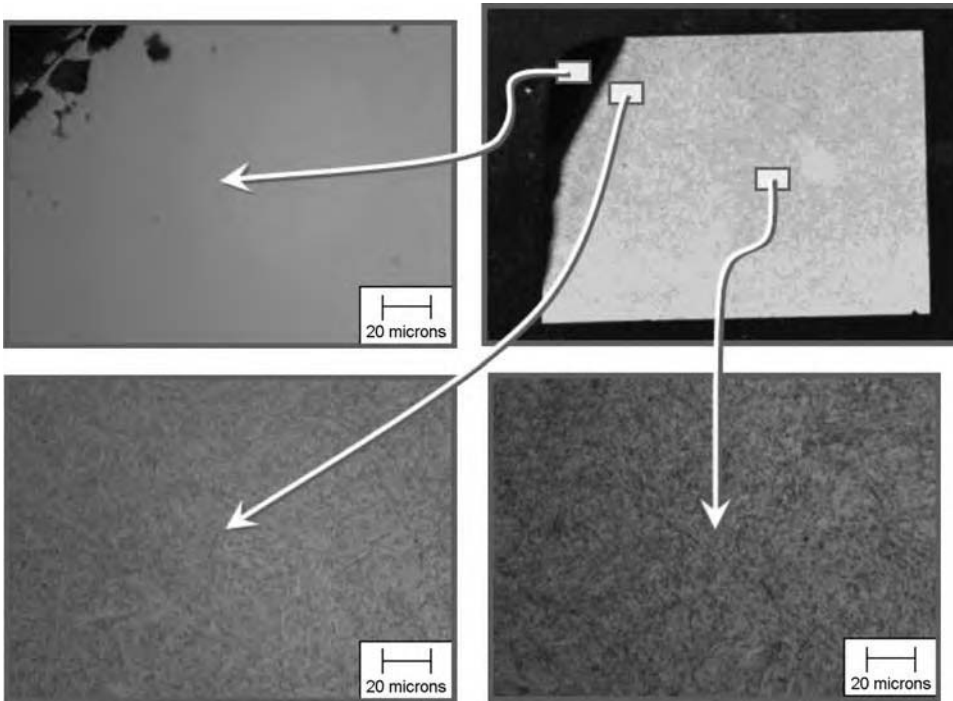
Fig. 5 Macrograph and associated micrographs illustrating the microstructure variations observed in a tool-steel mold insert after multiple titanium alloy castings. Source: Ref 15
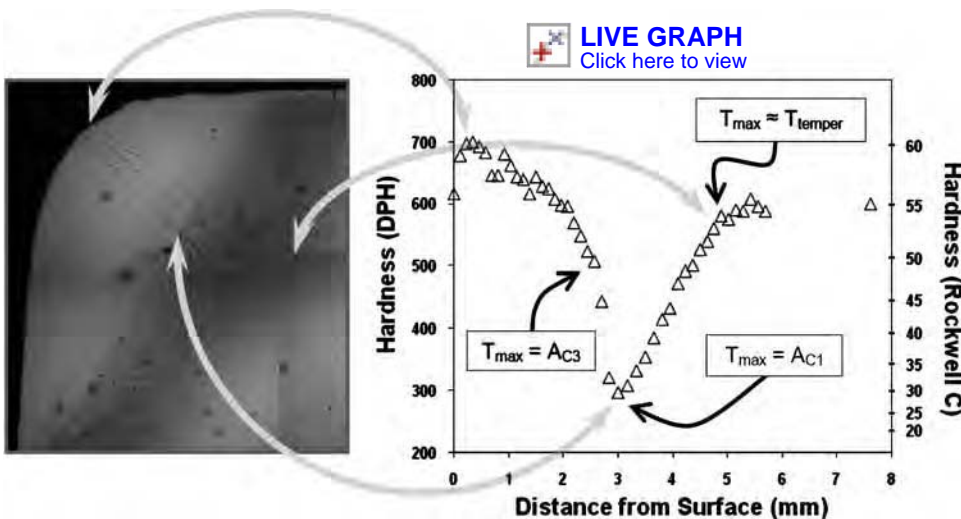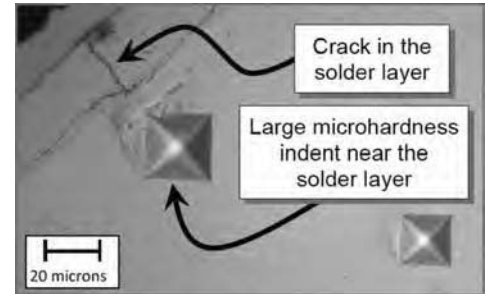

Fig. 6 Micrograph illustrating the formation of a solder layer at the surface of a tool steel mold insert after multiple titanium alloy castings. Source: Ref 15

Fig. 7 Macrograph and associated plot of microhardness illustrate the correlation of hardness, microstructure, and known steel phase transformation and tempering temperatures. Diamond pyramid hardness (DPH, Vickers) and Rockwell C values are given. Source: Ref 15

shapes/sizes also abound. One major concern is that the output of the inverse analysis is usually a set of values for h as a function of time, $h(t)$. Clearly, such values are tied to the specific casting configuration employed in the experiment, because the timeframe for cooling and solidification is closely tied to the size and shape of the casting. Given that it is cooling and solidification that drive the contact condition at the interface, which, in turn, drives the value of h, $h(t)$ results are not universally applicable to different casting configurations. Instead, the $h(t)$ values must be used to develop values of h as a function of some parameter that correlates more generally to changes in the interface as casting progresses, such as casting surface temperature, contact pressure, or gap size.

For example, modelers often choose to express h as a function of casting surface temperature. Values of h as a function of casting surface temperature can be derived readily from $h(t)$ data by modeling the experimental casting using the $h(t)$ results from the inverse analysis, extracting model results for $T(t)$ at the surface of the casting, and manually converting the $h(t)$ data to $h(T)$ data.

Correlations between h and interface contact pressure and/or gap size can also be developed directly from $h(t)$ results if applied contact stresses and/or gap sizes have been recorded/measured as a function of time in the experimental casting. Alternatively, the same model-based procedure described in the preceding paragraph for deriving $h(T)$ using results extracted from the model can be employed to derive h as a function of contact condition, provided that the chosen modeling software can predict interface contact condition.

Selecting the Interface Heat-Transfer Coefficient for a Given Casting Configuration

Because the goal of solidification modeling is to predict how a casting solidifies without performing physical experiments, modelers must select appropriate values of h from available data. Large, technologically advanced foundries

11, 14). Such calibration-curve techniques rely on the selection of a reasonably accurate form of the $h(t)$ or $h(T)$ function, and their efficiency is greatly aided by using experience- or literature-based "educated guesses" for the initial values of h. However, they avoid many of the numerical issues related to data conditioning and solution convergence often encountered during the application of inverse methods.

Concerns regarding error in and the conditioning of temperature data, model oversimplification, solution convergence, and computational

intensity remain for many of these methods, making them somewhat difficult to apply in general practice. Modern solidification modeling packages (such as ProCAST or MAGMASOFT) and optimization software packages (such as iSIGHT) can greatly simplify and improve the process of determining h from experimental $T(t)$ data, but the lack of precise physical experiments still hampers the proliferation of heat-transfer coefficient data (Ref 14, 18).

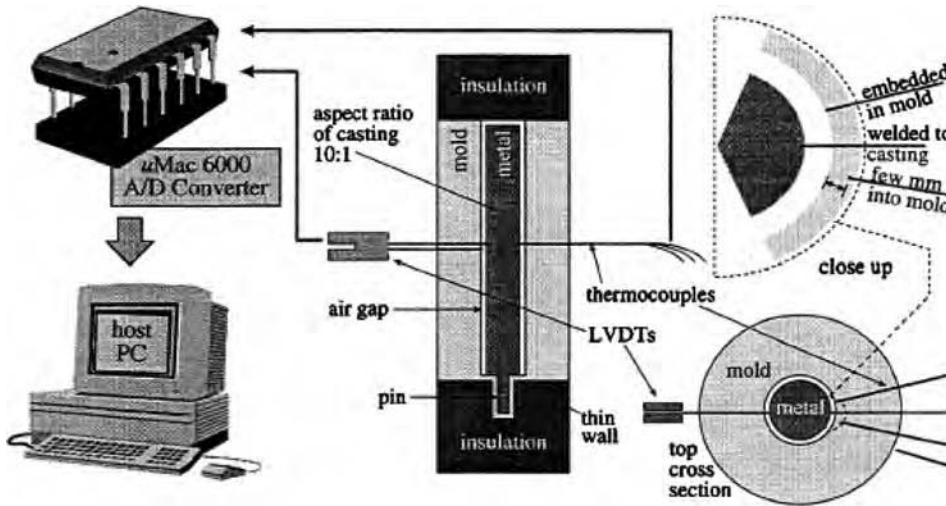Concerns regarding the applicability of experimentally developed values of h to other casting

Fig. 8 Schematic diagram of an experimental setup for measuring gap size, using linear variable differential transformers (LVDTs), and temperature, using thermocouples, in a cylindrical casting. Source: Ref 5
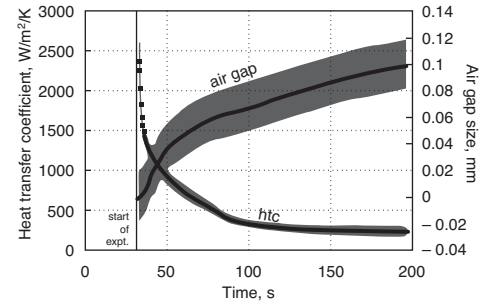

Fig. 9 Measured gap size and calculated h as a function of time for the experimental setup in Fig. 8. htc, heat-transfer coefficient. Source: Ref 5

LIVE GRAPH
Click here to view

Fig. 10 Variation of h as a function of casting surface temperature derived via the calibration-curve method. Source: Ref 14

Fig. 11 Comparison of the temperature history predicted using the $h(t)$ function in Fig. 10 with the temperature history measured experimentally. Source: Ref 14

often generate a set of heat-transfer coefficient data that spans the range of casting alloys, mold materials, and generic casting shapes encountered during the course of their business, but they are reluctant to share their databases. Hence, smaller foundries and those engaged in independent solidification modeling research must seek data from the open literature.

A multitude of publications related to interface heat transfer in castings are available in the scientific and technical literature. Over the last several decades, numerous experimentalists have studied the effect of select casting process variables on the value of h by conducting controlled experiments using a variety of casting configurations. A rather comprehensive listing of such studies was compiled by Papai and Mobley in 1987 and more recently by Pehlke in 2001 (Ref 7, 12). The effect of factors such as casting alloy, mold material, mold surface finish, mold wash, superheat of the melt, interface contact pressure, and casting shape on h has been investigated in great detail.

Unfortunately for the modeler, the vast majority of published work regarding interface heat transfer has been focused on determining physical mechanisms rather than on providing modelers with appropriate input data. Data that are available are often in the form of $h(t)$ for the specific experimental casting, and complete information regarding the experimental configuration is not always provided. Databases included with commercial software packages are encumbered by similar issues. Hence, the modeler is forced to apply engineering expertise in choosing the best selection of h for a given casting model.

A selected summary of h data from the open literature for various types of aluminum casting is provided in Table 1. Publicly available data for other materials, such as cast iron, steel, magnesium, and copper, are quite scarce and therefore are not included in the summary table.
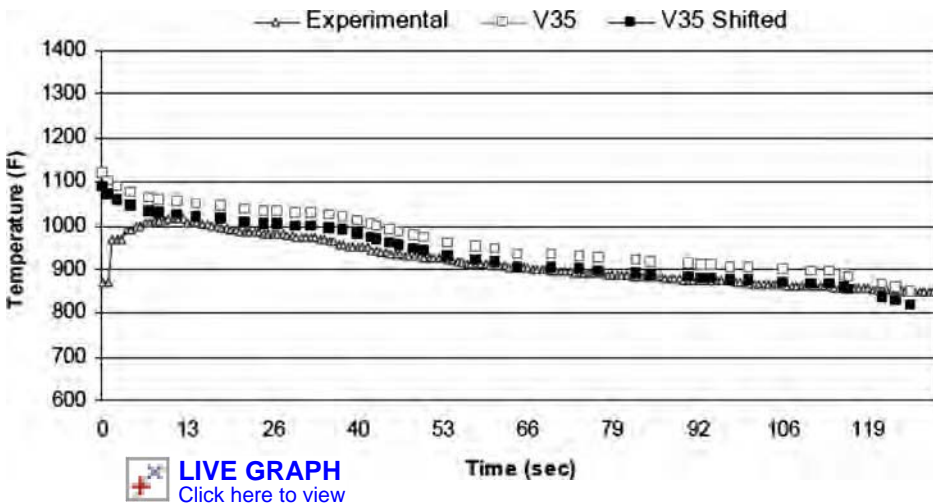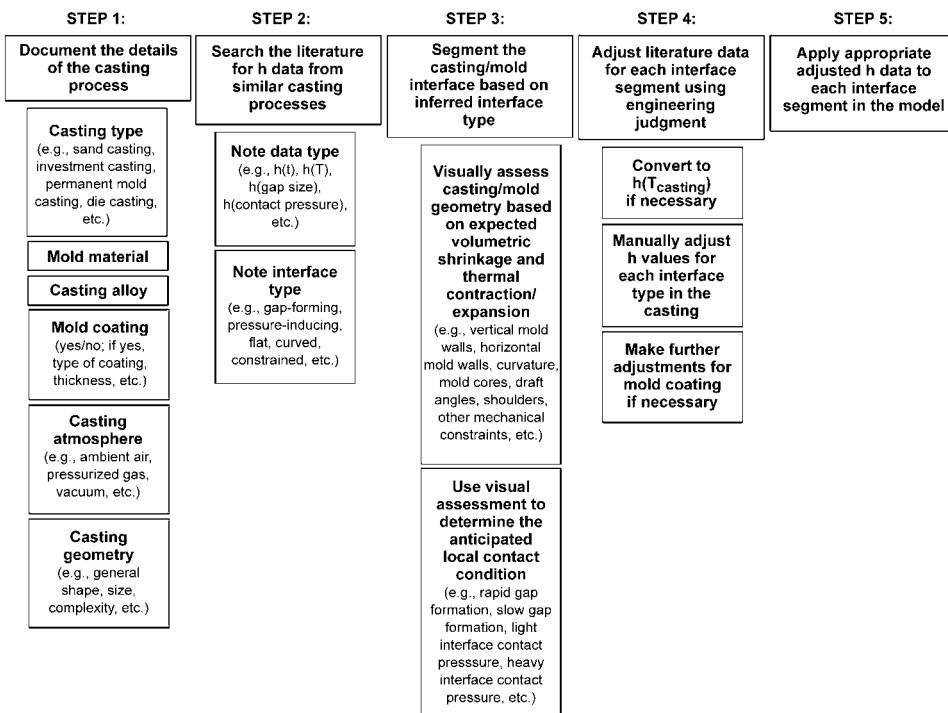
A general flowchart for selecting appropriate values of h from data in the literature for a selected casting configuration is presented in Fig. 12. This flowchart is based on an assimilation of various methodologies suggested by multiple researchers and is intended to serve as a guide for the casting modeler.
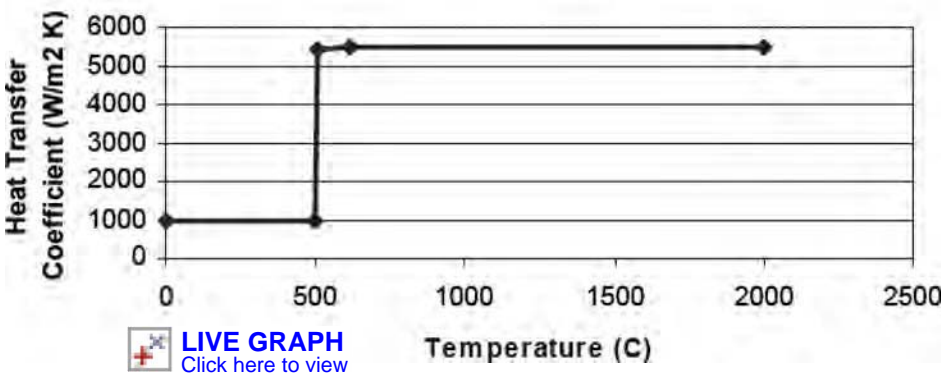
Table 1 Maximum reported _h_ values for various casting configurations during permanent mold casting, die casting, and sand casting of aluminum alloys

Casting aluminum alloy	Material Mold	Mold coating	Casting thickness, mm	_h_, W/m²·K	Comments	Ref
Pure	(a)	Graphite	34	~2250–3250	Cylinder in a permanent mold	16
Pure	(a)	Graphite	22	~3000–5000	Horizontal plate in a permanent mold	16
Al-13.2Si	(a)	Graphite	34	~3000	Cylinder in a permanent mold	16
Al-13.2Si	(a)	Graphite	22	~3000–5000	Horizontal plate in a permanent mold	16
Pure	Cast iron	(a)	19	~1150	Vertical plate in a permanent mold	8
Pure	Steel	(a)	19	~1900	Vertical plate in a permanent mold	8
AC8A(b)	SKD61 die steel	TOYOCA-ACE GR-851 graphite	15	~2700 (outer wall) ~40,000 (inner wall)	Tube in a permanent mold	5
AC8A(b)	SKD61 die steel	FOSECO-DYCOTE 39 white coating	15	~1500 (outer wall) ~20,000 (inner wall)	Tube in a permanent mold	5
A356	SKD61 die steel	(a)	15	~3000 (outer wall) ~40,000 (inner wall)	Tube in a permanent mold	5
Pure	SKD61 die steel	(a)	15	~3000 (outer wall) ~100,000 (inner wall)	Tube in a permanent mold	5
Al-Si eutectic	SKD61 die steel	(a)	15	~2000 (outer wall) ~100,000 (inner wall)	Tube in a permanent mold	5
A356	(a)	(a)	(a)	~5500	Commercial hubcap permanent mold casting	14
A380	H13 tool steel	(a)	0.8	78,000–87,000	Die casting; _P_ = 12,700–25,400 psi	19
Unspecified	Green sand	N/A	30	~1000	Cylinder sand casting	20
Pure	Sand	N/A	(a)	~150	Sand casting	21

(a) Not reported. (b) JIS H5202, similar to A336

Fig. 12 Flowchart for selecting appropriate values of the interface heat-transfer coefficient for a complex casting from published interface heat-transfer coefficient data

Examples

To illustrate how a modeler may experimentally determine and apply appropriate values of _h_ for a selected casting model, two examples from the literature are provided. The first example illustrates a method for characterizing the difference in _h_ as a function of location in bar and plate castings, while the second example illustrates a method for characterizing _h_ values for gap-forming and contact-inducing interfaces and then selectively applying them to a complex casting.

Example 1: Bar and Plate. Prasana Kumar and Kamath recognized the need to characterize _h_ for different boundary types (Ref 22). Hence, they designed an experiment to assess the variation of _h_ in different segments of the casting-mold interface for a square bar and rectangular plate casting. The Al-3Cu-4.5Si castings were made in thermocoupled cast iron molds. Four thermocouples were placed at distinct locations along the horizontal axis in each mold at the midheight of the casting. Analysis was conducted to determine the values of _h_ as a function of time, considering one, two, and three interface segments, respectively. The results showed that the bar casting has little variation in _h_ as a function of horizontal location, while the plate casting has a significant variation in _h_ as a function of horizontal location. Information from experiments such as these can be used to apply engineering judgment to identifying appropriate interface segments and associated _h_ values for a real casting.

Example 2: Titanium Cylinder and Tube. Kobryn and Semiatin also recognized the need to consider local interface characteristics when defining _h_ for a given casting (Ref 11). In this case, the variation of _h_ with two interface types, gap-forming, or "shrink-off," and contact-inducing, or "shrink-on," were explored for cylinder- and tube-shaped Ti-6Al-4V castings in H13 tool steel molds. Two cylinder diameters and two core diameters were considered, resulting in a total of four different casting geometries. Thermocouples were embedded in the molds and castings at various locations, and multiple casting trials were performed for each casting geometry (Fig. 13).

The calibration-curve method was used to determine the best-fit values of _h_ as a function of casting surface temperature for one gap-forming casting and one contact-inducing casting, based on an assumed functional form of the _h_(_T_) curve for each interface type. For the gap-forming case, _h_ was assumed to have its maximum value in the liquid state, decrease linearly from the liquidus to the solidus, then decrease linearly from the solidus to a presumed gap-forming temperature, and finally decrease linearly from the gap-forming temperature to room temperature. For the contact-inducing case, the _h_ was assumed to have the same initial value in the liquid as for the gap-forming case, but then to increase linearly from the liquidus to the solidus and from the solidus to the same assumed gap-forming temperature and remain constant thereafter.

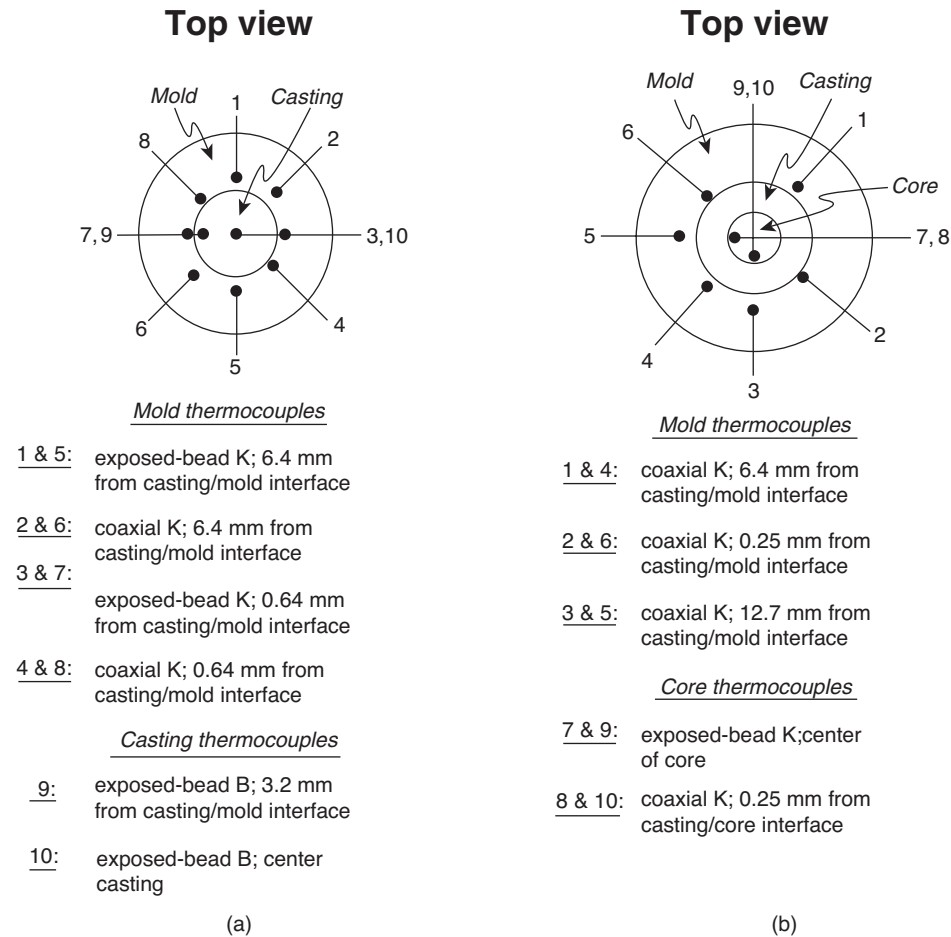The _h_ values for each interface type that were obtained via the calibration-curve method

Top view

Mold — 1 — Casting
8
2
7,9 — — 3,10
6
4
5

Mold thermocouples

1 & 5: exposed-bead K; 6.4 mm from casting/mold interface

2 & 6: coaxial K; 6.4 mm from casting/mold interface

3 & 7: exposed-bead K; 0.64 mm from casting/mold interface

4 & 8: coaxial K; 0.64 mm from casting/mold interface

Casting thermocouples

9: exposed-bead B; 3.2 mm from casting/mold interface

10: exposed-bead B; center casting

(a)

Top view

Mold — 9,10 — Casting
6 — — 1
5 — — Core — 7,8
4 — — 2
3

Mold thermocouples

1 & 4: coaxial K; 6.4 mm from casting/mold interface

2 & 6: coaxial K; 0.25 mm from casting/mold interface

3 & 5: coaxial K; 12.7 mm from casting/mold interface

Core thermocouples

7 & 9: exposed-bead K; center of core

8 & 10: coaxial K; 0.25 mm from casting/core interface

(b)

Fig. 13 Schematic diagram of the experimental setup discussed in "Example 2: Titanium Cylinder and Tube." (a) "Shrink-off" casting. (b) Inner core "shrink-on" casting. Source: Ref 11
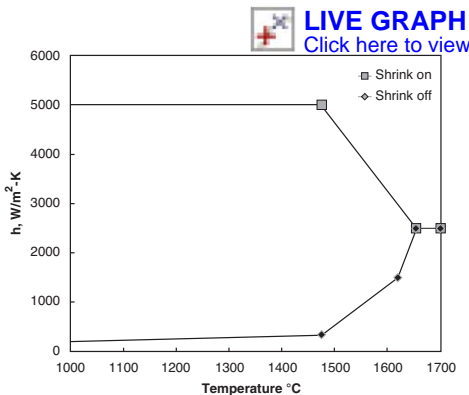
Fig. 14 Values of $h(T)$ from "Example 2: Titanium Cylinder and Tube." Source: Ref 11

(Fig. 14) were then applied to the remaining castings for validation purposes. Reasonable agreement between predicted and measured temperature profiles was obtained for these castings using the best-fit $h(T)$ values from the other castings, indicating that the proposed functional form and associated best-fit values of h as a function of casting surface temperature can be applied to similar interface segments in different castings.

The validation experiment was taken one step further by applying the $h(T)$ values to a completely different casting geometry (Ref 15). In this case, a rib-web casting, comprised of a weblike plate with protruding ribs that form discrete pockets, was modeled by applying the gap-forming $h(T)$ curve to the flat areas of the interface and the contact-inducing $h(T)$ to the edge and corner radii of the interface. Again, reasonable agreement was obtained between predicted and simulated thermal histories, indicating the engineering-expertise approach to selecting $h(T)$ is quite reasonable.

Summary

The proper selection of interface heat-transfer coefficients is critical to successful simulation of casting processes. Many casting variables impact the value of h as casting proceeds. Hence, it is necessary to choose experimentally determined values of h from experimental casting configurations that most closely match the casting to be modeled. Specifying h as a function of casting surface temperature provides more robust results than

specifying h a as function of time. However, specifying h as a function of interface contact condition (e.g., gap size or contact pressure) can provide more accurate results if appropriate input data are available and the added computation time is acceptable.

The application of engineering judgment to segment the casting-mold interface into gap-forming and contact-inducing sections is useful for applying $h(T)$ values obtained from simple experiments to the modeling of complex castings. Best results are obtained when the modeler can conduct his/her own simple experiments to determine h for the selected casting process, casting alloy, and generic geometric features likely to be encountered prior to simulating complex castings. However, extreme care must be taken to design the casting experiments to minimize error in the temperature measurement and variability in the casting trials. Once thermocouple data are collected, the calibration-curve method is a simple, robust method for determining values of $h(T)$ from thermocouple data for different interface types.

REFERENCES

1. "ProCAST Users Manual and Technical Reference," Version 3.1.0, UES Software, Inc., Dayton, OH, 1998, p 282–308
2. T. Midea, Solidifying Casting's Future: Process Simulation Software Round-Up, *Mod. Cast.,* Vol 92 (No. 8), Aug 2002, p 32
3. K. Ho and R.D. Pehlke, Mechanisms of Heat Transfer at a Metal-Mold Interface, *AFS Trans.,* Vol 61, 1984, p 587–598
4. M. Trovant and S. Argyropoulos, Finding Boundary Conditions: A Coupling Strategy for the Modeling of Metal Casting Processes: Part I. Experimental Study and Correlation Development, *Metall. Trans. B,* Vol 31, Feb 2000, p 75–86
5. T.-G. Kim and Z.-H. Lee, Time-Varying Heat Transfer Coefficients Between Tube-Shaped Casting and Metal Mold, *Int. J. Heat Mass Transf.,* Vol 40 (No. 15), 1997, p 3513–3525
6. M. Aloe and M. Gremaud, Advanced Metal Casting Simulations Save Time and Cost, *Cast Met. Die Cast. Times,* April/May 2007, p 35–38
7. J. Papai and C. Mobley, "Heat Transfer Coefficients for Solidifying Systems," Report ERC/NSM-887-13, Engineering Research Center for Net Shape Manufacturing, Aug 1987
8. L.J.D. Sully, The Thermal Interface Between Castings and Chill Molds, *AFS Trans.,* Vol 84, 1976, p 735–744
9. M. Rappaz et al., Application of Inverse Methods to the Estimation of Boundary Conditions and Properties, *Modeling of Casting, Welding and Advanced Solidification Processes VII, Proceedings,* Sept 1995 (London), TMS, 1995, p 449

10. K.A. Woodbury et al., "The Relationship between Casting Distortion, Mold Filling, and Interfacial Heat Transfer in Sand Molds," Report DOE/ID/1336-F, Sept 1999
11. P.A. Kobryn and S.L. Semiatin, Determination of Interface Heat-Transfer Coefficients for Permanent-Mold Casting of Ti-6Al-4V, *Metall. Trans. B,* Vol 32, Aug 2001, p 685–695
12. R.D. Pehlke et al., "Heat Transfer at the Metal-Mold Interface in Permanent Mold Casting of Aluminum Alloys Project—Final Report," U.S. Dept. of Energy Award No. DE-FC07-97ID159, Dec 14, 2001
13. D. O'Mahoney and D.J. Browne, Use of Experiment and an Inverse Method to Study Interface Heat Transfer during Solidification in the Investment Casting Process, *Exp. Therm. Fluid Sci.,* Vol 22 (No. 3–4), Sept 2000, p 111–122
14. R.D. Pehlke et al., "Investigation of Heat Transfer at the Mold/Metal Interface in Permanent Mold Casting of Light Alloys—Final Technical Report," U.S. Dept. of Energy Award No. DE-FC36-02ID14236, Oct 27, 2005
15. P. Kobryn, "The Effect of Interface Heat Transfer on Solidification, Microstructure Evolution, and Mold Wear in Permanent Mold Casting of Ti-6Al-4V," Ph.D. Dissertation, The Ohio State University, 1999
16. Y. Nishida, W. Droste, and S. Engler, The Air-Gap Formation Process at the Casting-Mold Interface and the Heat Transfer Mechanism through the Gap, *Metall. Trans. B,* Vol 17, 1986, p 833–844
17. J.V. Beck, B. Blackwell, and C.R. St. Clair, Jr., *Inverse Heat Conduction—Ill-Posed Problems,* Wiley, New York, 1985
18. M. Li, R. Vijayaraghavan, and J.E. Allison, Development and Applications of Opt-Cast—A Thermal Boundary Condition and Casting Process Optimization Tool, *Simulation of Aluminum Shape Casting Processing: From Alloy Design to Mechanical Properties, Proceedings,* March 2006 (San Antonio, TX), TMS, 2006, p 91–100
19. S. Hong, D.G. Backman, and R. Mehrabian, Heat Transfer Coefficient in Aluminum Alloy Die Casting, *Metall. Trans. B,* Vol 10, 1979, p 299–301
20. R. Rajaraman and R. Velraj, Comparison of Interfacial Heat Transfer Coefficient Estimated by Two Different Techniques During Solidification of Cylindrical Aluminum Alloy Casting, *Heat Mass Transf.,* Vol 44, 2008, p 1025–1034
21. M. Trovant and S. Argyropoulos, Finding Boundary Conditions: A Coupling Strategy for the Modeling of Metal Casting Processes: Part II. Numerical Study and Analysis, *Metall. Trans. B,* Vol 31, 2000, p 87–96
22. T.S. Prasanna Kumar and H.C. Kamath, Estimation of Multiple Heat-Flux Components at the Metal/Mold Interface in Bar and Plate Aluminum Alloy Castings, *Metall. Trans. B,* Vol 35, 2004, p 575–585

Fundamentals of the Modeling of Microstructure and Texture Evolution

Modeling Diffusion in Binary and Multicomponent Alloys

John Morral, The Ohio State University
Frederick Meisenkothen, UES, Inc./Air Force Research Laboratory

MODELING DIFFUSION in alloys can be divided into two activities. One activity is to model the changes in composition and phases that take place due to diffusion. The other is to determine the kinetic constants that are contained in the model. Determining the kinetic constants is an important part of diffusion modeling, and an article in this Handbook ("Diffusivity and Mobility Data") has been devoted to that topic. The article includes information on how the constants are related to the atomic mechanisms of diffusion, the atom mobilities, and the thermodynamic properties of alloys. In the present article, it is assumed that the kinetic constants are known, and the focus is on models that can be used to solve various diffusion problems.

The objective of this article is to provide the reader with information needed to model diffusion. To this end, the fundamental laws of diffusion and the equations derived from them are given as necessary background. Then, well-known solutions to these equations are presented to provide a sense of how concentrations of individual elements and the overall composition of an alloy can change during a diffusion-controlled process. A series of applications give examples of how various diffusional processes can be modeled with the equations. Also, the applications are intended to further extend the reader's understanding of conditions when diffusion is important and how it leads to composition variations as a function of time and temperature.

Diffusion in Technology

Diffusion can play an important role in both materials processing and materials degradation during the service life of a part. For example, nearly all heat treatments involve diffusion. Annealing to soften or homogenize, precipitation treatments to strengthen, or spheroidizing or tempering treatments to increase ductility all require the redistribution of atoms in a material. Gas-solid reactions such as carburizing and nitriding to increase the wear resistance of gears and shafts, as well as pack aluminizing to increase oxidation resistance of turbine blades, are other examples where engineers use the concepts of diffusion to design new heat treatments or modify existing ones. Other processing technologies in which diffusion plays an important role include metal joining by diffusion bonding, material densification by hot isostatic pressing, and low-pressure deformation by superplastic forming.

With regard to material degradation, there are a variety of ways that diffusion can limit the service life of a part. For example, high-temperature coatings interdiffuse with parts they are meant to protect. By this process, critical elements in the coating diffuse into the underlying part, thereby reducing their oxidation resistance and often reducing their mechanical properties. Also, oxidation of high-temperature materials is a diffusion-affected process that reduces the cross section of a part and may create stress raisers for crack initiation. Metals that are body-centered cubic, such as ferritic steels and refractory metals, may pick up solute atoms from the environment, such as hydrogen, carbon, and nitrogen. In time, these elements can embrittle a material and make it difficult to deform during processing or be the cause of unexpected and catastrophic failures, as in the case of hydrogen embrittlement.

Sources of Diffusion Data. The most complete compilation of diffusion data is *Diffusion in Metals and Alloys,* edited by Mehrer (Ref 1). It provides data on diffusion in pure, binary, and selected ternary systems. In addition, it has special chapters on diffusion in amorphous metals, diffusion of interstitial atoms, and diffusion in high-diffusivity paths such as grain and interphase boundaries, dislocations, and on surfaces. Another key reference on high-diffusivity paths is *Handbook of Grain Boundary and Interphase Boundary Diffusion Data* by I. Kaur et al. (Ref 2).

A less complete but more readily available reference is *Smithells Metal Reference Book* (Ref 3).

Most editions contain diffusion data for pure and binary alloys. Another resource is *Diffusion and Defect Data,* a 92-volume set of diffusion data, abstracts, and articles that was published from 1967 to 1992. Back issues are available online. It has since become an online journal called *Diffusion and Defect Forum* (Ref 4).

Because of the large number of alloy systems of industrial importance and the limited amount of diffusion data, there are many systems for which diffusion data are not available in these references. In that case, a computer search using a general online search engine or technical journal search engine can provide data from recent journal articles. In some cases, people make their own measurements of diffusion data to properly model or understand a diffusion problem. Also, there are methods to estimate the diffusivity from existing mobility and thermodynamic data. These topics are treated in detail in the article "Diffusivity and Mobility Data" in this Volume.

Fundamentals of Diffusion

Models of diffusional processes are typically designed to simulate the concentration profile of one or more components as a function of time. A concentration profile is a plot of concentration versus distance that fits the boundary and initial conditions of the process and satisfies the laws of diffusion given subsequently. Concentration profiles can be measured experimentally by using an electron microprobe to make concentration measurements incrementally along the length of a sample. Experimental concentration profiles taken on a ternary diffusion couple are compared with a model simulation in Fig. 1 taken from Ref 5.

Comparing simulated and experimental concentration profiles is one way to evaluate the accuracy of a diffusion model and the diffusion database employed. As explained in the article "Diffusivity and Mobility Data" in this Volume, concentration profiles can be used to create diffusion databases, too.
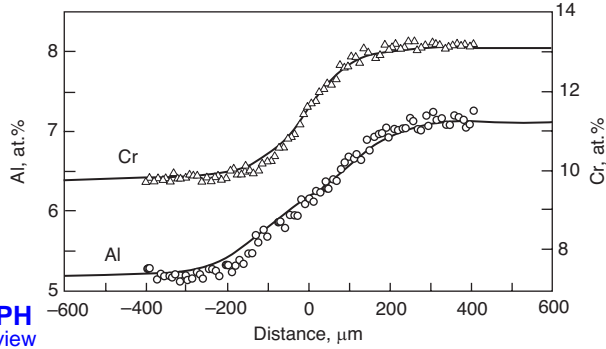
Fig. 1 Concentration profiles measured on a Ni-Cr-Al diffusion couple using an electron microprobe (points) and the prediction of an error function model (lines). Source: Ref 5

Fig. 2 A small-volume element in a crystal, with a flux entering the left side of the volume and a different flux leaving the right side. As a result, the concentration of the volume element will change with time.

Figure 1 is for diffusion in a single phase region. Diffusion can occur in multiphase regions as well, but only in ternary and higher-order systems. In multiphase regions, it is possible to plot profiles both for concentrations in each phase and for the local average concentration. In this article, the modeling of both single-phase and multiphase profiles is covered.

Diffusion Flux and Fick's First Law

Two fundamental laws of diffusion were formulated in the 19th century by Adolph Fick. These laws can be used to model a variety of diffusion problems in binary systems, as long as the system is either isotropic or has a cubic crystal structure (e.g., nickel-copper alloys). These laws can be extended to more components and other crystal structures, as discussed in a later section.

Fick's first law gives the flux as:

$$J = -D\frac{\partial C}{\partial x} \tag{Eq 1}$$

Equation 1 indicates that a flux of atoms results from a variation of concentration, C, with distance, x (i.e., the concentration gradient). The proportionality constant, D, is the diffusivity or diffusion coefficient of the alloy. As explained in the article "Diffusivity and Mobility Data" in this Volume, this is the chemical diffusivity and can be used to calculate the flux for either component. The chemical diffusivity should not be confused with the tracer or intrinsic diffusivity, which are different for different components and are not appropriate for the modeling described in this article.

The units of the flux depend on the units of C, which in SI units is normally given as either kg/m^2 or mol/m^2, although other units are used as well. With mass units, the flux is given as $kg/m^2/s$, and with mole units as $mol/m^2/s$ or mole fraction. However, regardless of the units of C, the diffusivity has units of m^2/s.

The flux in Fick's first law is sometimes referred to as the flux density, because it refers to the flux through a unit area (e.g., per square meter). It follows that if one were calculating the total flux passing through a membrane, it would be the flux density times the area.

It is important to realize that J is a vector quantity. When J is positive, mass is flowing in the plus x-direction. When it is negative, it flows in the negative x-direction. A key concept of diffusion in binary alloys is that solute diffuses down its concentration gradient; that is, atoms move from regions of high concentration to regions of lower concentration. However, for ternary and higher-order systems, this may not be the case, as shown later.

For isotropic or cubic crystals, Fick's first law applies when the concentration gradient is in any crystal direction. However, in crystals with less symmetry, the crystal direction is important, as indicated by the subsequent equations, which show that a concentration gradient in one direction can lead to a flux in another direction:

$$J_x = -D_{xx}\frac{\partial C}{\partial x} - D_{xy}\frac{\partial C}{\partial y} - D_{xz}\frac{\partial C}{\partial z}$$
$$J_y = -D_{yx}\frac{\partial C}{\partial x} - D_{yy}\frac{\partial C}{\partial y} - D_{yz}\frac{\partial C}{\partial z}$$
$$J_z = -D_{zx}\frac{\partial C}{\partial x} - D_{zy}\frac{\partial C}{\partial y} - D_{zz}\frac{\partial C}{\partial z} \tag{Eq 2}$$

In this notation x, y, and z refer to three orthogonal axes in a single crystal. The diffusion coefficients in Eq 2 can be zero or have finite values, depending on the symmetry of the crystal. For a hexagonal crystal, there are two diffusivities. One is for concentration gradients in the basal plane, and the other is for concentration gradients perpendicular to the basal plane. However, because most available data are for isotropic or cubic materials, the focus of this article is on those materials for which there is one D that is independent of direction, and therefore, Fick's first law applies. Additional information on this topic is available in Chapter 2 of Glicksman's book, *Diffusion in Solids* (Ref 6).

Polycrystalline materials may act in an isotropic way, regardless of the crystal structure, if the grains have a random orientation. Then, Fick's law, with one effective diffusivity, would apply.
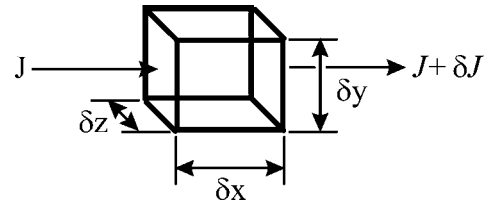
However, polycrystalline materials contain grain boundaries that are high-diffusivity paths for diffusion. This can lead to surprising amounts of diffusion but normally only at a temperature below half the melting point, which is when grain-boundary diffusion tends to dominate mass transport. More details on high-diffusivity paths are in the article "Diffusivity and Mobility Data" in this Volume.

The Diffusion Equation and Fick's Second Law

Fick's first law is a useful equation for calculating the flux when the concentration gradient is known. However, of more general interest is how the concentration of a specific element in a material varies with respect to distance and time, $C(x,t)$. This is obtained from the diffusion equation, which is given subsequently.

The diffusion equation is derived by performing a mass balance on a small-volume element in an alloy. As illustrated in Fig 2, the flux into the volume element, J, may be different than the flux out of the volume element, $(J + \delta J)$. It follows that the change in concentration of the volume element at a given position, x, depends on how rapidly the flux is changing with distance at a given time, t, according to:

$$\left.\frac{\partial C}{\partial t}\right|_x = -\left.\frac{\partial J}{\partial x}\right|_t \tag{Eq 3}$$

Note that if the flux is decreasing with distance, the concentration will increase, because more is entering the volume element than is leaving. Also, if there is no difference between the fluxes in and out, then the concentration will stay the same.

When Fick's first law is substituted into Eq 3, the diffusion equation is obtained:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial C}{\partial x}\right) \tag{Eq 4}$$

Equation 4 is the basic equation used by computer programs to solve for $C(x,t)$ and is quite general for the diffusion of either component in a binary alloy. The only exception to Eq 4

is when there are sources or sinks contained in a volume element. Sources and sinks can be a factor when considering the diffusion of vacancies, because they can be created or annihilated at dislocations or other defects. However, sinks and sources are not a factor for alloying elements, because they cannot be created or destroyed.

If the diffusivity is a function of concentration, there are no simple analytical solutions to Eq 4. However, if D is a constant, Eq 4 simplifies to Fick's second law:

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} \qquad \text{(Eq 5)}$$

This is the equation proposed by Fick in the 19th century. It is a linear differential equation for which there are several well-known solutions.

Diffusion in Ternary and Higher-Order Multicomponent Alloys

Binary alloys have only one independent variable, because the mole fraction or weight fraction of one component can always be used to calculate the fraction of the other component. That is why Fick's first and second laws are written in terms of one variable. However, ternary alloys have two independent variables, and as a result, Fick's laws must be extended to consider the additional variable.

Extending Fick's first law to more components yields a series of linear equations. For ternary alloys, the expanded form is:

$$J_1 = -D_{11} \frac{\partial C_1}{\partial x} - D_{12} \frac{\partial C_2}{\partial x}$$
$$J_2 = -D_{21} \frac{\partial C_1}{\partial x} - D_{22} \frac{\partial C_2}{\partial x} \qquad \text{(Eq 6)}$$

Here, four diffusivities are needed. These diffusivities are for isotropic or cubic materials and should not be confused with the directional diffusivities given in Eq 2. Equation 6 can be used to calculate the flux of the third component, J_3, with the conservation equation:

$$J_1 + J_2 + J_3 = 0 \qquad \text{(Eq 7)}$$

Equation 6 shows that, unlike binary alloys, the flux of atoms may go up a concentration gradient. For example, if $\partial C_2/\partial x$ is sufficiently greater than $\partial C_1/\partial x$, then solute 1 will diffuse down the concentration gradient of component 2, even if, at the same time, it is diffusing up the concentration gradient of component 1. An atom moving up its own concentration gradient is known as up-hill diffusion and is one of several multicomponent effects that are unique to systems that have three or more components. Figure 3 from Ref 5 gives an example of up-hill diffusion in which chromium moves from a low concentration on the right to a higher concentration on the left.

These equations can be extended to alloy systems with four components by recognizing that,
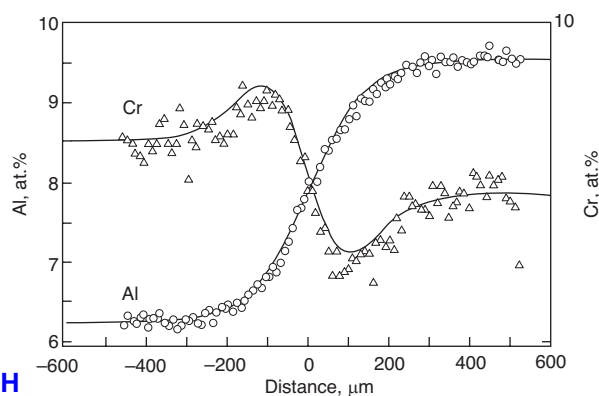
Fig. 3 Concentration profiles measured on a Ni-Cr-Al diffusion couple using an electron microprobe (points) and the prediction of an error function model (lines). This example illustrates the up-hill diffusion of chromium. Source: Ref 5

with three independent variables, there must be three flux equations containing gradients in C_1, C_2, and C_3. Accordingly, a five-component system would require four flux equations with gradients in four independent variables, C_1, C_2, C_3, and C_4. Writing out these equations for the flux, diffusion equation, and the extended form of Fick's second law offers little additional insight into diffusion in multicomponent systems but does indicate the need for a more compact notation. In the following, a matrix notation is described that is used throughout, in hopes of reaching the largest number of readers. However, a more advanced treatment would benefit from using tensor notation.

Rounded-Bracket Notation. Note: The symbols "A," "B," "C," "D," and so on in this section should not be confused with similar symbols in other sections of this article.

Matrices play an important role in the derivation, application, and understanding of multicomponent diffusion equations. As indicated previously, the diffusion equations are laborious to write out and difficult to manipulate for systems that contain more than two components. Here, a specific type of matrix notation, known as rounded-bracket notation (Ref 7), is used that allows the reader to quickly recognize whether a matrix is a column, a row, or a square matrix. Calculations involving matrices that are written in rounded-bracket notation are straightforward and can be performed on most spreadsheets. A summary of rounded-bracket notation is given as follows, along with several examples of how to apply the notation.

A square matrix is designated by using two square brackets, [ ], to enclose the matrix:

$$[A] = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad \text{(Eq 8)}$$

Column and row matrices are represented by using one rounded bracket with one square bracket. A column matrix is indicated by [ ), while a row matrix is written as ( ], as shown in the following examples:

$$[B) = \text{column matrix} = \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} \qquad \text{(Eq 9)}$$

$$(C] = \text{row matrix} = \begin{bmatrix} c_{11} & c_{12} \end{bmatrix} \qquad \text{(Eq 10)}$$

The product of a row and column matrix is a scalar, which is just a number, and is represented by using two rounded brackets, ( ), or by using no brackets:

$$(C][B) = (d) = d \qquad \text{(Eq 11)}$$

Rows and columns of a matrix [A] are indicated by using a subscript for the number of the row or column:

$$(A_1] = \text{first row of } [A] = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix} \qquad \text{(Eq 12)}$$

$$[A_2) = \text{second column of } [A] = \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} \qquad \text{(Eq 13)}$$

A single element of a matrix is designated by indicating its row and column with rounded brackets:

$$(A_{12}) = a_{12} \qquad \text{(Eq 14)}$$

Calculations are simplified through the use of rounded-bracket notation. For example, suppose that the following equation needs a solution:

$$? = [H][G][L) \qquad \text{(Eq 15)}$$

From the principles of matrix algebra, it follows that the answer will have the same first and last brackets of the equation. Therefore, the solution is a column vector:

$$[E) = [H][G][L) \qquad \text{(Eq 16)}$$

In a more complex example:

$$[A) = [H](I][G][K][L) \qquad \text{(Eq 17)}$$

The rounded-bracket notation shows that the product $(I][G][K)$ is a scalar equal to some number N:

$$N = (I)[G][K] \qquad \text{(Eq 18)}$$

Therefore, Eq 17 can be simplified to:

$$[A] = N[H][L] \qquad \text{(Eq 19)}$$

In another example, the first element of the column matrix in Eq 16 can be obtained by substituting the first row of $[H]$, (H_1) in the equation to obtain:

$$(E_1) = (H_1)[G][L] \qquad \text{(Eq 20)}$$

Fick's First Law for Single-Phase Multicomponent Systems. Fick's first law can be extended to multicomponent systems in the rounded-bracket notation by the expression:

$$[J] = -[D]\left[\frac{\partial C}{\partial x}\right) \qquad \text{(Eq 21)}$$

For a ternary system:

$$[D] = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \qquad \text{(Eq 22)}$$

As explained in the section on rounded-bracket notation, the flux of component 2 can be written as:

$$J_2 = (J_2) = -(D_2)\left[\frac{\partial C}{\partial x}\right) \qquad \text{(Eq 23)}$$

Equation 23 is equivalent to the extended form of Fick's first law for ternaries given by Eq 6.

Fick's Second Law for Single-Phase Multicomponent Systems. Fick's second law can be extended to ternary systems with the equations:

$$\frac{\partial C_1}{\partial t} = D_{11}\frac{\partial^2 C_1}{\partial x^2} + D_{12}\frac{\partial^2 C_2}{\partial x^2}$$
$$\frac{\partial C_2}{\partial t} = D_{21}\frac{\partial^2 C_1}{\partial x^2} + D_{22}\frac{\partial^2 C_2}{\partial x^2} \qquad \text{(Eq 24)}$$

The equations take into account that there are two independent concentration variables in a ternary system. In rounded-bracket notation, Fick's second law for any number of components is:

$$\left[\frac{\partial C}{\partial t}\right) = [\mathrm{D}]\left[\frac{\partial^2 C}{\partial x^2}\right) \qquad \text{(Eq 25)}$$

As in Fick's first law, $[D]$ for a ternary system is a 2×2 matrix, while $[D]$ for an n-component system would be an $(n-1) \times (n-1)$ matrix. Although measured diffusivities are available only for selected ternary systems and even fewer quaternary systems, it is possible to estimate these higher-order diffusivities using methods described in the section "Modeling Multicomponent Diffusivity Data" and in Appendix 1 of the article "Diffusivity and Mobility Data" in this Volume.

Equations 24 and 25 are two coupled linear differential equations. They are coupled because C_1 and C_2 appear in both equations. In the 1940s, Onsager explained that these equations could be readily solved by using a mathematical

procedure to diagonalize the diffusivity matrix, $[D]$. This procedure uncouples the equations and makes their solution straightforward.

The procedure involves writing the concentrations in a new concentration coordinate system in which the diffusivity is a diagonal matrix. For a ternary system, the diagonal matrix has the form:

$$[\tilde{D}] = \begin{bmatrix} \tilde{D}_1 & 0 \\ 0 & \tilde{D}_2 \end{bmatrix} \qquad \text{(Eq 26)}$$

The diagonal matrix is obtained from the equation:

$$[\tilde{D}] = [\alpha][D][\alpha]^{-1} \qquad \text{(Eq 27)}$$

The $[\alpha]$ matrix and its inverse, $[\alpha]^{-1}$, as well as the elements of $[\tilde{D}]$ can be obtained from spreadsheets, which use numerical methods to obtain them. In the parlance of matrix algebra, the diagonal elements of Eq 26 are eigenvalues of $[D]$, while the column vectors of $[\alpha]^{-1}$ are eigenvectors of $[D]$. The largest eigenvalue, $\tilde{D}_1$, is called the major eigenvalue, while the smaller eigenvalues are designated as minor eigenvalues. The two α matrices are related by the equation:

$$[I] = [\alpha]^{-1}[\alpha] = [\alpha][\alpha]^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{(Eq 28)}$$

In which $[I]$ is the identity matrix. Multiplying a matrix by the identity matrix is the matrix equivalent of multiplying by one.

Concentrations in the new coordinate system are related to concentrations in the normal coordinate system by the equations:

$$(\tilde{C}_1) = \alpha_{11}C_1 + \alpha_{12}C_2 = (\alpha_1)[C]$$
$$(\tilde{C}_2) = \alpha_{21}C_1 + \alpha_{22}C_2 = (\alpha_2)[C] \qquad \text{(Eq 29)}$$

Equation 29 is written in regular algebraic notation for ternaries and for n-component systems in the rounded-bracket notation.

Equation 25 in the new coordinate system is:

$$\left[\frac{\partial \tilde{C}}{\partial t}\right) = [\tilde{D}]\left[\frac{\partial^2 \tilde{C}}{\partial x^2}\right) \qquad \text{(Eq 30)}$$

The equations are now uncoupled and have the form:

$$\frac{\partial \tilde{C}_1}{\partial t} = \tilde{D}_1\frac{\partial^2 \tilde{C}_1}{\partial x^2}$$
$$\frac{\partial \tilde{C}_2}{\partial t} = \tilde{D}_2\frac{\partial^2 \tilde{C}_2}{\partial x^2} \qquad \text{(Eq 31)}$$

Model solutions to Eq 31 are well known for binary systems, as shown in the section "Modeling Diffusion with Constant D Equations" in this article. These solutions for $\tilde{C}_1(x,t)$ and $\tilde{C}_2(x,t)$ can be substituted into the reciprocal relations:

$$C_1(x,t) = \alpha_{11}^{-1}\tilde{C}_1 + \alpha_{12}^{-1}\tilde{C}_2 = (\alpha_1)^{-1}[\tilde{C}]$$
$$C_2(x,t) = \alpha_{21}^{-1}\tilde{C}_1 + \alpha_{22}^{-1}\tilde{C}_2 = (\alpha_2)^{-1}[\tilde{C}] \qquad \text{(Eq 32)}$$

to obtain models for diffusion in terms of normal concentrations. In Eq 32, the constant

parameters are elements of the matrix $[\alpha]^{-1}$ and are not to be confused with reciprocals of individual elements of the matrix $[\alpha]$.

A key generalization one learns from Eq 32 is that solutions to multicomponent diffusion problems are linear sums of solutions to binary diffusion problems. Also, one learns that the kinetic constants for multicomponent diffusion problems are eigenvalues of the diffusivity matrix, $[D]$, and not individual elements of the diffusivity matrix. The eigenvector directions are important, too, as seen in the discussion of models in "Modeling Diffusion with Constant D Equations" in this article.

Diffusion in Ternary and Multicomponent, Multiphase Regions

A multiphase region in an alloy is a region that contains more than one phase. For example, a multiphase region that consists of a matrix phase and a precipitate is illustrated in Fig. 4. As is the case of single-phase regions, there can be fluxes in and out of the volume element, and these will determine how the overall or average composition of the volume element changes with time.

For a binary alloy at constant temperature, all concentrations of the matrix and precipitate phase are fixed by the ends of a tie line on the binary-phase diagram. Because the concentrations are fixed, there can be no concentration gradients and therefore no diffusion. However, the concentrations can vary with distance if the volume element is in a temperature gradient, and in that case, diffusion will occur. If the temperature gradients are sufficiently large, as occurs, for example, in nuclear reactors, then an additional term must be added to Fick's first law to account for thermotransport (Ref 8). However, for this article, isothermal conditions are assumed, and diffusion in multiphase systems can occur only if ternary or higher-order systems are involved.

The change in average concentration of a volume element with time is given by the mass balance equation for each component as (Ref 7):

$$\left.\frac{\partial \bar{C}_i}{\partial t}\right|_x = -\left.\frac{\partial J_i}{\partial x}\right|_t \qquad \text{(Eq 33)}$$
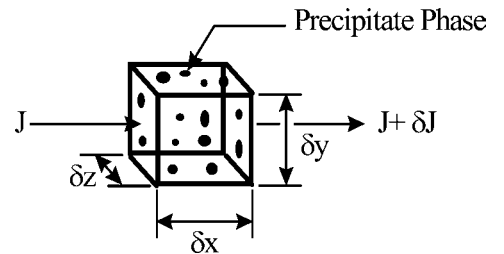

Fig. 4 Flux in and out of a volume element containing more than one phase. In this case, the variation of flux with distance changes the average concentration of the volume element.

or for the composition $[\bar{C}]$ as:

$$\frac{[\partial \bar{C}]}{\partial t} = -\frac{\partial[J]}{\partial x} \qquad \text{(Eq 34)}$$

The concentration $\bar{C}_i$ is the overall concentration of component i that includes i in both the matrix and precipitate phases:

$$\bar{C}_i = f^m C_i^m + f^{ppt} C_i^{ppt} \qquad \text{(Eq 35)}$$

In Eq 35, f^m and f^{ppt} refer to the fraction, while C_i^m and C_i^{ppt} refer to concentrations of the matrix and precipitate phase, respectively. As Eq 35 implies, measurements required to obtain the average concentration are the concentration in each phase and the fraction of one of the phases.

Equations 33 and 34 are completely general and require no assumptions beyond those needed for the binary diffusion equation. However, simplifying assumptions are needed to expresses the flux in these equations. In general, the flux will depend on the precipitate morphology and volume fraction. Here, the assumption that leads to the simplest equations is made. It is assumed that the volume fraction is so small that precipitates will act like point sources or sinks of solute and thus will not block or mediate the matrix flux in any way. Therefore, the flux will be given by Eq 21, because it can be assumed that all diffusion is taking place in the matrix phase. It follows that for ternary, two-phase systems, Eq 33 becomes:

$$\frac{\partial \bar{C}_1}{\partial t} = \frac{\partial}{\partial x}\left(D_{11}^\gamma \frac{\partial C_1^\gamma}{\partial x} + D_{12}^\gamma \frac{\partial C_2^\gamma}{\partial x}\right)$$
$$\frac{\partial \bar{C}_2}{\partial t} = \frac{\partial}{\partial x}\left(D_{21}^\gamma \frac{\partial C_1^\gamma}{\partial x} + D_{22}^\gamma \frac{\partial C_2^\gamma}{\partial x}\right) \qquad \text{(Eq 36)}$$

in which the superscript γ designates concentrations or diffusivities associated with the matrix phase.

Equation 36 depends on four different concentrations: the average and γ concentrations of solute 1 and the average and γ concentrations of solute 2. However, Eq 36 can be rewritten entirely in terms of average concentrations by using the chain rule to change variables (Ref 7):

$$\frac{\partial \bar{C}_i}{\partial t} = \frac{\partial}{\partial x}\left(D_{i1}^\gamma \left(\frac{\partial C_1^\gamma}{\partial \bar{C}_1}\frac{\partial \bar{C}_1}{\partial x}\right)_{\bar{C}_2} + D_{i1}^\gamma \left(\frac{\partial C_1^\gamma}{\partial \bar{C}_2}\frac{\partial \bar{C}_2}{\partial x}\right)_{\bar{C}_1}\right.$$
$$\left. + D_{i2}^\gamma \left(\frac{\partial C_2^\gamma}{\partial \bar{C}_1}\frac{\partial \bar{C}_1}{\partial x}\right)_{\bar{C}_2} + D_{i2}^\gamma \left(\frac{\partial C_2^\gamma}{\partial \bar{C}_2}\frac{\partial \bar{C}_2}{\partial x}\right)_{\bar{C}_1}\right) \qquad \text{(Eq 37)}$$

Gathering like terms in Eq 37 yields:

$$\frac{\partial \bar{C}_i}{\partial t} = \frac{\partial}{\partial x}\left(\left(D_{i1}^\gamma \frac{\partial C_1^\gamma}{\partial \bar{C}_1} + D_{i2}^\gamma \frac{\partial C_2^\gamma}{\partial \bar{C}_1}\right)\frac{\partial \bar{C}_1}{\partial x}\right.$$
$$\left. + \left(D_{i1}^\gamma \frac{\partial C_1^\gamma}{\partial \bar{C}_2} + D_{i2}^\gamma \frac{\partial C_2^\gamma}{\partial \bar{C}_2}\right)\frac{\partial \bar{C}_2}{\partial x}\right) \qquad \text{(Eq 38)}$$

Equation 38 shows that diffusion in a multiphase region can be treated like diffusion in a single-phase region according to the equation:

$$\left[\frac{\partial \bar{C}}{\partial t}\right] = \frac{\partial}{\partial x}\left[D^{\text{eff}}\right]\left[\frac{\partial \bar{C}}{\partial x}\right] \qquad \text{(Eq 39)}$$

in which $[D^{\text{eff}}]$ is an effective diffusivity defined by:

$$\left[D^{\text{eff}}\right] = \left[D^\gamma\right]\left[C^{\text{TM}}\right] \qquad \text{(Eq 40)}$$

in which $[C^{\text{TM}}]$ is a matrix that transforms the γ-phase diffusivity into an effective diffusivity for the two-phase mixture. The transformation matrix is given by:

$$\left[C^{\text{TM}}\right] = \begin{bmatrix} \dfrac{\partial C_1^\gamma}{\partial \bar{C}_1}\Big|_{\bar{C}_2} & \dfrac{\partial C_1^\gamma}{\partial \bar{C}_2}\Big|_{\bar{C}_1} \\ \dfrac{\partial C_2^\gamma}{\partial \bar{C}_1}\Big|_{\bar{C}_2} & \dfrac{\partial C_2^\gamma}{\partial \bar{C}_2}\Big|_{\bar{C}_1} \end{bmatrix} \qquad \text{(Eq 41)}$$

Figure 5 shows that for every pair of $\bar{C}_1$ and $\bar{C}_2$, there is a corresponding value of $C_1^{\gamma/\beta}$ and $C_2^{\gamma/\beta}$, which shows that the partial derivatives can be defined. However, the determinant of Eq 41 is zero, which indicates that the minor eigenvalue of Eq 40 must be zero also. The physical significance of the zero minor eigenvalue is related to the minor eigenvector direction. It lies along a tie line where the composition of each phase is fixed. It follows that without concentration gradients in the individual phases, there can be no diffusion in this composition direction. The lack of diffusion is reflected in the zero eigenvalue, which is the rate constant for gradients in average concentration along a tie line.

Diffusion Paths, Diffusion Couples, and Composition Vectors

Diffusion paths can provide insights about the microstructure that forms in an interdiffusion region. The paths are plotted on concentration axes and show how the local composition changes on moving from one side of an interdiffusion region to another. Diffusion paths are particularly helpful when analyzing diffusion couples, and this section is devoted to that analysis.

Diffusion couples consist of two single or multiphase alloys that are placed in intimate contact with one another to share a common interface through which diffusion can occur. The initial composition of a diffusion couple can be characterized by a list of concentrations for each alloy, as shown in Fig. 6, or by the composition of one alloy and the composition vector. The composition vector connects the two initial compositions, as shown in Fig. 7, and by convention (Ref 9) for a ternary is given by:

$$\left[\Delta C^0\right] = \begin{pmatrix} C_1^R - C_1^L \\ C_2^R - C_2^L \end{pmatrix} \qquad \text{(Eq 42)}$$

In which (C_1^R, C_2^R) is the initial composition of the alloy on the right in the positive x-direction,
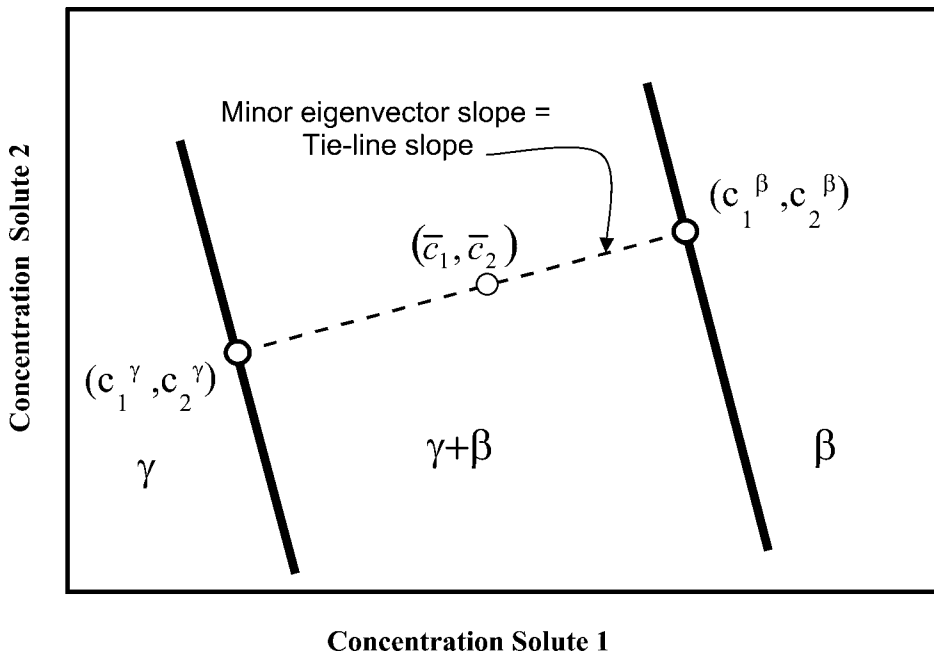

Fig. 5 Illustration that the average composition in a two-phase region determines the composition of each phase via the tie-line ends

Fig. 6 Illustration of a diffusion couple with alloy compositions given by a list of concentrations
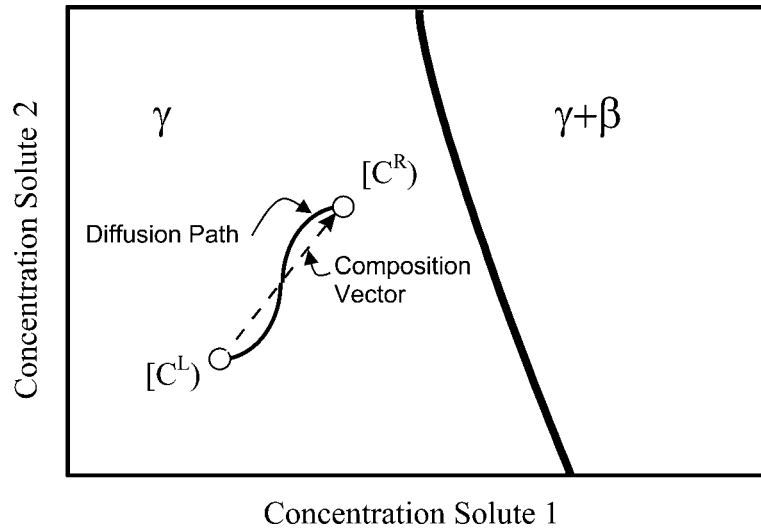
Fig. 7 Diffusion path and composition vector for a single-phase diffusion couple plotted on a phase diagram. Open circles are the initial diffusion-couple alloys.

and $(C_1^{\mathrm{L}}, C_2^{\mathrm{L}})$ is the alloy on the left in the negative x-direction. The vector is important because it appears in equations used to model diffusion couples.

If either experimental or simulation data are available in tabular form, with one column for position and other columns for each independent concentration, for example, C_1, C_2, C_3, and so on, and rows are concentrations at a particular x, then diffusion paths are obtained by cross plotting concentrations. For a ternary system, one would plot C_2 versus C_1 and would likely obtain a figure similar to Fig. 7. The ends of the path for a diffusion couple are the initial compositions of the alloys and represent compositions that are outside the interdiffusion region. Moving along the path from one end to other is equivalent to moving across the interdiffusion region, although the relationship between distance on the path and distance on the diffusion couple is highly nonlinear, except near $x = 0$, the initial interface.

The slope of the diffusion path as it leaves either initial composition is given by $[\alpha_1]^{-1}$, the major eigenvector direction of $[D]$. The major eigenvector tends to dominate here because it is associated with the largest kinetic constant. Therefore, this term in the solution to Fick's second law reaches deeper into each alloy than other terms in Eq 32.

In single-phase regions, the diffusion path of a diffusion couple will not change with time as long as the interdiffusion region is smaller than the diffusion couple. When concentration profiles reach the ends of the couple, the end compositions will drift and change the diffusion path. In diffusion couples with regions that are multiphase, the diffusion path may change with time if the effective diffusivity changes as a result of changes in precipitate size and morphology.

Diffusion paths are two-dimensional for ternary alloys, because there are two independent concentration variables. Therefore, they can be plotted on related ternary phase diagram isotherms, as in Fig. 7. Here, one can see why the diffusion couple is single phase and what changes would be needed to introduce a two-phase region. Also, the diffusion path has a serpentine shape, which is expected for single-phase ternary couples unless the composition vector is in or nearly in an eigenvector direction, where the diffusion path is a straight line. If one models diffusion paths as a function of the composition vector angle and assumes constant $[D]$, one finds that the path starts as a straight line along an eigenvector direction. As the angle changes, it becomes more and more serpentine. Then, as it approaches the next eigenvector direction, it collapses back into a line (Ref 9).

In four-component and higher-order systems, the diffusion paths and isothermal phase diagrams can be plotted as two-dimensional projections, but these phase diagrams are complex and may provide only qualitative information.

Zero-Flux Planes

In diffusion couples, it is possible to determine the local flux of an element at time t with the general equation:

$$J_i(x', t) = \frac{1}{2t} \int_{C(\infty, t)}^{C(x', t)} x\, dC_i \qquad \text{(Eq 43)}$$

in which x is the distance from the Matano plane, and x' is where the flux is evaluated. Kim and Dayananda (Ref 10) discovered in a study of Ni-Cu-Zn diffusion couples that at specific locations, the flux could be zero. Due to the one-dimensional nature of diffusion couples, this was named a zero-flux plane, or ZFP for short. Figure 8 is from their classic paper (Ref 10) and shows that there are two ZFPs for copper, one for nickel, and none for zinc.

In addition to being an unexpected phenomenon, ZFPs have applications to alloy design. Constant $[D]$ models for single-phase diffusion couples indicate that alloys can be designed so there is a ZFP at the initial interface where two alloys are joined (Ref 11, 12). Therefore, a coating that is high in a protective element such as aluminum could be designed so that aluminum would not be lost to the underlying matrix in the early stages of interdiffusion and would be minimized over the life of the coating. This a multicomponent effect that is not seen in binary diffusion couples. The effect is possible only in ternary or higher-order systems. More information about modeling ZFPs using constant D equations is provided in the section "Modeling Diffusion with Constant D Equations" in this article.

Kirkendall Porosity

The Kirkendall effect is described in some detail in the article "Diffusivity and Mobility Data" in this Volume. The effect is associated with various atomic elements having different intrinsic diffusivities. As a result, porosity in a diffusion couple can develop in regions where atoms with a higher intrinsic diffusivity leave a region faster than they can be replaced by atoms with a smaller intrinsic diffusivity.

Eliminating Kirkendall porosity in a binary diffusion couple is a nearly intractable problem for the following reason. The porosity is related to the movement of markers placed at the initial interface, which in turn is proportional to the flux of vacancies. The flux of vacancies, ${}^{\mathrm{L}}J_{\mathrm{V}}$, makes up for the difference between the individual intrinsic fluxes, ${}^{\mathrm{L}}J_1$ and ${}^{\mathrm{L}}J_2$, via the relationship:

$$0 = {}^{\mathrm{L}}J_{\mathrm{V}} + {}^{\mathrm{L}}J_1 + {}^{\mathrm{L}}J_2 \qquad \text{(Eq 44)}$$

Because:

$$\begin{aligned} {}^{\mathrm{L}}J_1 &= -{}^{\mathrm{L}}D_{11}\frac{\partial C_1}{\partial x} \\ {}^{\mathrm{L}}J_2 &= -{}^{\mathrm{L}}D_{21}\frac{\partial C_1}{\partial x} \end{aligned} \qquad \text{(Eq 45)}$$

The flux of vacancies is:

$$ {}^{\mathrm{L}}J_{\mathrm{V}} = \left({}^{\mathrm{L}}D_{11} + {}^{\mathrm{L}}D_{21}\right)\frac{\partial C_1}{\partial x} \qquad \text{(Eq 46)}$$

Therefore, the only way to reduce the vacancy flux is to reduce the concentration gradients in the couple by reducing the initial concentration differences of both components.

However, in multicomponent systems, the flux of vacancies at the initial interface can be reduced to zero by alloy design (Ref 13). For example, in a ternary system, Eq 46 becomes:

$$ {}^{\mathrm{L}}J_{\mathrm{V}}(0,t) = \left({}^{\mathrm{L}}D_{11} + {}^{\mathrm{L}}D_{21} + {}^{\mathrm{L}}D_{31}\right)\frac{\partial C_1}{\partial x}\bigg|_{x=0} $$
$$ + \left({}^{\mathrm{L}}D_{12} + {}^{\mathrm{L}}D_{22} + {}^{\mathrm{L}}D_{32}\right)\frac{\partial C_2}{\partial x}\bigg|_{x=0} $$
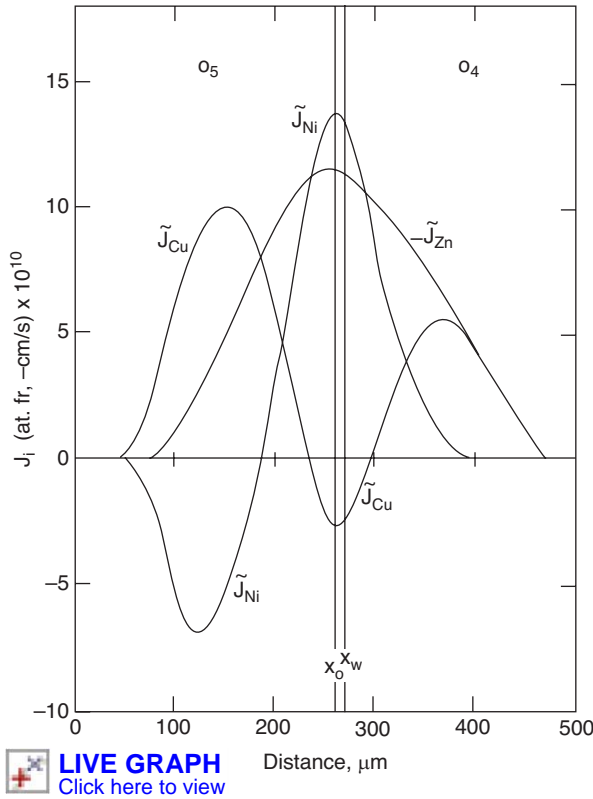$$ \text{(Eq 47)} $$

Fig. 8 Plots of flux versus distance for a Ni-Cu-Zn diffusion couple. In this case, there are two zero-flux planes for copper atoms and one zero-flux plane for nickel atoms. Source: Ref 10

Fig. 9 Concentration profiles around a diffusion-controlled moving interface. The phases are in equilibrium at the interface.

Setting Eq 47 equal to zero yields the relationship:

$$\frac{dC_2}{dC_1}\bigg|_{x=0} = -\frac{(^LD_{11} + {}^LD_{21} + {}^LD_{31})}{(^LD_{12} + {}^LD_{22} + {}^LD_{32})} \qquad \text{(Eq 48)}$$

Equation 48 is the slope of the diffusion path at $x = 0$ and does not vary with time for a diffusion couple. With this direction, there should be neither Kirkendall porosity or marker movement.

Diffusion at Moving Boundaries

The simplest type of moving boundary to describe is an interface between two phases that are in local equilibrium at the interface. Local equilibrium means that concentrations at the interface are given by the ends of a tie line on a phase diagram. In this case, the movement of the boundary is said to be diffusion controlled.

The velocity of the interface is governed by a mass balance at the interface. For the boundary in Fig. 9 to move a distance $d\eta$, the difference between the flux entering and leaving the interface in time dt must equal $\left(C_i^{\beta/\alpha} - C_i^{\alpha/\beta}\right)d\eta$, which is the amount of solute that must be displaced for the interface to move. Writing:

$$\left(C_i^{\beta/\alpha} - C_i^{\alpha/\beta}\right) = \Delta C_i^{\alpha\beta} \qquad \text{(Eq 49)}$$

the mass balance can be written as:

$$\Delta C_i^{\alpha\beta} d\eta = \left(J_i^{\beta} - J_i^{\alpha}\right) dt \qquad \text{(Eq 50)}$$

For a binary system, Eq 50 can be written as:

$$\Delta C^{\alpha\beta} \frac{d\eta}{dt} = D^{\alpha}\frac{\partial C^{\alpha}}{\partial x}\bigg|_{\eta} - D^{\beta}\frac{\partial C^{\beta}}{\partial x}\bigg|_{\eta} \qquad \text{(Eq 51)}$$

Equation 50 and Fig. 9 show that the flux in α away from the interface tends to drive the interface in the positive x-direction, while the flux in β into the interface drives the interface in the negative x direction.

Equation 51 must be satisfied for each component in a multicomponent system. The equation can be generalized to any number of components using the rounded-bracket notation by simply writing each term as its matrix equivalent:

$$\left[\Delta C^{\alpha\beta}\right]\frac{d\eta}{dt} = [D^{\alpha}]\left[\frac{\partial C^{\alpha}}{\partial x}\right]_{\eta} - [D^{\beta}]\left[\frac{\partial C^{\beta}}{\partial x}\right]_{\eta} \qquad \text{(Eq 52)}$$

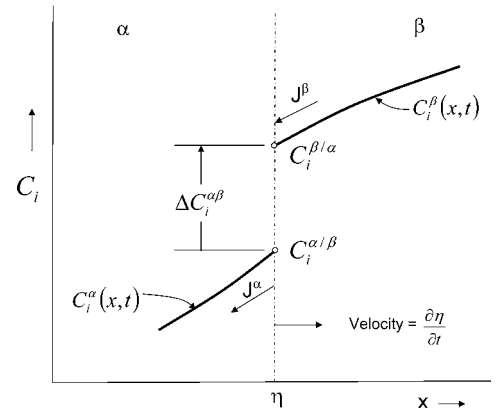However, there is a problem. In Eq 52, the value of $\Delta C^{\alpha\beta}$ can be read off a binary phase diagram from the tie line associated with the isothermal temperature. However, in ternary and higher-order systems, all the equations in Eq 52 as well as equations for possible tie lines in the multicomponent system must be solved simultaneously. To avoid writing equations for tie lines, it is necessary to use computer models that link to thermodynamic databases in order to solve multicomponent moving-boundary problems.

In addition to interfaces between single-phase regions, there are other types of moving boundaries that involve diffusion. For example, there can be a boundary between a single-phase and two-phase region in a binary system. Boundaries such as this have been categorized according to the number of phases that change when crossing the boundary (Ref 14). Figure 10 shows three types of boundaries that can be found in binary diffusion couples. The type 0 boundary has no changes in phase, just a change in volume fraction at $x = 0$. No diffusion is associated with this type of binary (except for coarsening), but there is diffusion at type 0 boundaries in ternary and higher-order systems. In any case, type 0 boundaries are stationary unless they are displaced by the Kirkendall effect. The type 1 boundary does move in a binary, but only in the direction to extend the single-phase region. Type 2 boundaries are an interface between two single-phase regions, and they can move in either direction, depending on Eq 52.

Another multicomponent effect is that type 1 boundaries can move in either direction in ternary and higher-order systems (Ref 15, 16). Also, in principle, there can be as many types as there are components in the alloys. However, the most common types are 0, 1, and 2 (Ref 17), with an occasional type 3 boundary reported, even in a ternary system (Ref 18, 19). There may be type 4 boundaries in quaternary and higher-order systems as well, but no reference to this type of boundary has been found by the authors.

Modeling Diffusion with Constant D Equations

Solutions to the diffusion equation and Fick's second law are functions that give concentrations as a function of distance and time, $C_i(x,t)$. The diffusion equation is most accurate because it takes into account the variation of diffusivities with composition, while Fick's second law assumes they are constants. Solutions to Fick's second law have the advantage that they often provide simple equations that are useful when designing alloys and diffusional processes. However, a disadvantage of solutions to Fick's second law is that they only apply to certain boundary and initial conditions. Another disadvantage is the constant diffusivity assumption. However, the constant D solutions tend to be robust, and variations in D of 10% would likely be insignificant compared to scatter in experimental data.

In the following, four standard solutions to Fick's second law for binary single-phase systems are given. The derivation of these and additional solutions can be found in the well-known text by Crank, *The Mathematics of Diffusion* (Ref 20). The only disadvantage of that text is that most examples pertain to polymers.
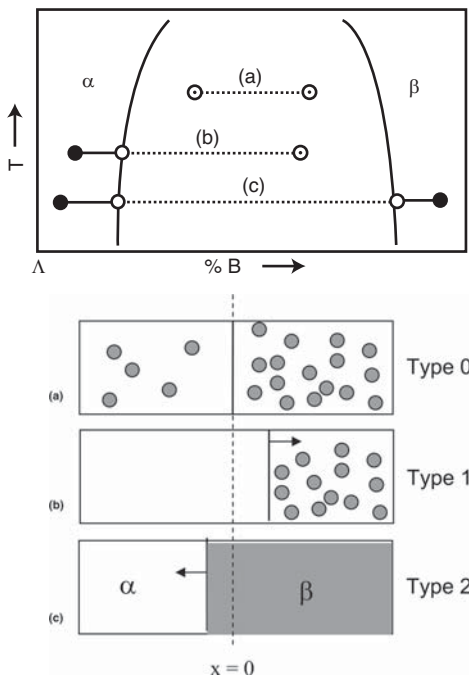

Fig. 10 Illustration of three types of boundaries in binary diffusion couples: (a) type 0, (b) type 1, and (c) type 2. In the phase diagram, closed circles and circles containing a dot are initial alloy concentrations of diffusion couples. Open circles indicate single-phase α concentrations at the moving boundaries. Arrows indicate boundary direction. In (a), the boundary is stationary; in (b), the boundary can only move to the right, while in (c), the boundary can move either to the left or right, depending on Eq 52. Source: Ref 14

Some cases show how the four solutions can be modified to apply them to systems containing more than two components and more than one phase. As seen, the change is often no more than a change from scalar to matrix notation.

The Steady-State Solution

The steady-state solution applies when the concentration is a function of distance only and therefore must satisfy the equation:

$$\frac{\partial C}{\partial t} = 0 \tag{Eq 53}$$

For some problems, steady-state solutions are applied when concentrations vary so slowly with time that useful models are obtained. In such cases, they are termed quasi-steady-state solutions. In the following, it is assumed that the systems being modeled are single phase. When multiphase systems are modeled, it is indicated in the text.

One-Dimensional Solution. Fick's second law, given by Eq 5, is for diffusion in one dimension (1-D). For problems with cylindrical or spherical symmetry, the two-dimensional (2-D) and three-dimensional (3-D) solutions apply. They are given in later sections.

To agree with Eq 53, Fick's second law requires that:

$$\frac{\partial^2 C}{\partial x^2} = 0 \tag{Eq 54}$$

By integrating Eq 54 twice, one obtains the solution:

Fig. 11 Model for diffusion of hydrogen through a flat wall. Equation 68 for the flux also applies to a curved wall if the wall is sufficiently thin.

$$C(x) = A + Bx \tag{Eq 55}$$

in which A and B are constants that can be calculated from any two concentrations along the concentration profile. Equation 55 is the 1-D solution. It applies when the diffusivity is a constant.

Application—Diffusion of Gas Through a Flat Wall. The loss of gas from a container by diffusion through the container walls, as well as the contamination of a vacuum by gas diffusing into a vacuum chamber, are problems that can be modeled by the 1-D steady-state solution. In either case, it is a problem of gas diffusing through a wall or other thin section that separates a region where gas is at a high pressure from a region where it is at a lower pressure.

Substituting Eq 55 into Fick's first law for the flux yields:

$$J = -D\frac{d(A + Bx)}{dx} = -DB \tag{Eq 56}$$

in which D is the diffusivity of the diffusing species. The value of B can be obtained by knowing the solute concentration on either side of the wall. As shown in Fig. 11, the concentration on the left side of the wall is C^L, the concentration on the right side of the wall is C^R, and the thickness of the wall is h_0. Substituting these values into Eq 55 yields:

$$C^L = A + B(0)$$
$$C^R = A + B(h_0) \tag{Eq 57}$$

Solving Eq 57 for B and substituting it into Eq 56 gives the flux as:

$$J = D\frac{C^L - C^R}{h_0} \tag{Eq 58}$$

The values of C^L and C^R for a diatomic gas can be obtained from the gas partial pressures with Sievert's law. For the diffusion of hydrogen, Sievert's law is:

$$C_H = k\sqrt{p_{H_2}} \tag{Eq 59}$$

in which C_H is the concentration of atomic hydrogen in the metal, k is the Sievert's law constant for the dissociation of hydrogen, and p_{H_2} is the partial pressure of hydrogen in the gas phase. If there is a vacuum on the right side of the wall, the flux of hydrogen through the wall is:

$$J_H = D_H\frac{k\sqrt{p_{H_2}}}{h_0} \tag{Eq 60}$$

If the total area of the wall (the area across which the diffusion is occurring) is A_T, the time over which diffusion occurs is t_T, and the gas pressure is constant, then the total mass loss is $J_H A_T t_T$. However, if the gas pressure is changing with time, as it would from a closed
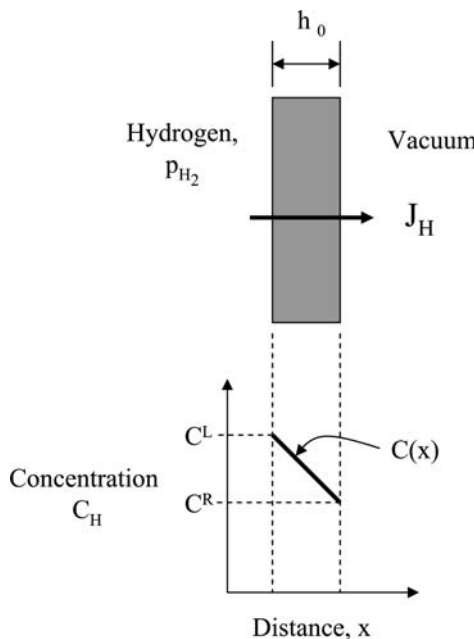
container, then Eq 60 can be treated as a quasi-steady solution, and the mass loss is:

$$\text{Mass } H_2 \text{ lost} = A_T \int_0^{t_T} J_H dt \qquad \text{(Eq 61)}$$

Although Eq 58 and 60 were derived for a flat wall, the next section shows that they apply to curved walls, too, as long as the radius of curvature is much larger than the wall thickness.

Two-Dimensional Solution. The 2-D version of Fick's second law for systems with cylindrical symmetry is:

$$\frac{\partial C}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial C}{\partial r}\right) \qquad \text{(Eq 62)}$$

Setting Eq 62 equal to zero and integrating twice yields the equation:

$$C(x) = A + B\ln r \qquad \text{(Eq 63)}$$

In these equations, r is the radial distance from the cylindrical axis.

Application—Diffusion of Gas Across a Pipe Wall. Equation 63 can be applied to the problem of calculating the flux of gas escaping from a pipe through the pipe walls. The approach is much like that for diffusion through a thin wall. Using Sievert's law, one obtains the concentrations of the gas species on both the inside diameter of the pipe, C^{id} at r^{id}, and the outside diameter of the pipe, C^{od} at r^{od}. With these two concentrations at two different radii, one can calculate both A and B. However, as in the case of a flat wall, only B is needed to calculate the flux because:

$$J = -D\frac{\partial C}{\partial r} = -D\frac{\partial(A + B\ln r)}{\partial r}$$
$$J = -D\frac{B}{r} \qquad \text{(Eq 64)}$$

Solving for B using equations for concentrations at the inside and outside radius gives:

$$B = \frac{C^{id} - C^{od}}{\ln(R^{id}/R^{od})} \qquad \text{(Eq 65)}$$

Therefore:

$$J = -\frac{D}{R^{id}}\left(\frac{C^{id} - C^{od}}{\ln(R^{id}/R^{od})}\right) \qquad \text{(Eq 66)}$$

Equation 66 can be simplified by taking the thickness of the wall as h_0, assuming that $h_0 << r^{id}$, and making use of the relationship that:

$$\ln\left(\frac{R^{id}}{R^{od}}\right) = \ln\left(\frac{R^{id}}{R^{id} + h_0}\right) = \ln\left(\frac{1}{1 + (h_0/R^{id})}\right)$$
$$\cong -\frac{h_0}{R^{id}} \qquad \text{(Eq 67)}$$

The percentage error in making this approximation is half the percentage difference between h_0 and R^{id}. For example, if h_0 is 10% of R^{id}, then the approximation error will be 5%.

Substituting Eq 67 into Eq 66 yields the thin-wall approximation for cylinders, which is equivalent to the flat-wall Eq 58:

$$J = D\left(\frac{C^{id} - C^{od}}{h_0}\right) \qquad \text{(Eq 68)}$$

The total flux leaving the pipe is the flux times the area of the pipe wall. For a pipe of length L, the area is equal to $2\pi R^{id}L$.

Three-Dimensional Solution. The 3-D version of Fick's second law for systems with spherical symmetry is:

$$\frac{\partial C}{\partial t} = \frac{D}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial C}{\partial r}\right) \qquad \text{(Eq 69)}$$

Setting Eq 54 equal to zero and integrating twice yields the equation:

$$C(x) = A + B/r \qquad \text{(Eq 70)}$$

In Eq 70, r is the radial distance from the origin of the sphere.

Application—Growth of a Spherical Precipitate. When precipitates grow by a diffusional process, they either extract solute from the surrounding matrix phase or reject solute back into the matrix, depending on their phase diagram. In either case, when diffusion is the rate-limiting step, the growth is said to be diffusion controlled or diffusion limited. In this section, an equation is derived for the diffusion-controlled growth of a spherical, solute-rich precipitate. It is assumed that the precipitate size is large enough so that the Gibbs-Thomson effect on the interface concentrations is negligible. Therefore, concentrations at the curved precipitate/matrix interface are nearly the same as equilibrium concentrations at a flat interface. Also, the model assumes that the precipitate is isolated from other growing precipitates. Therefore, the solute-depleted region surrounding the precipitate does not overlap that of other precipitates.

As shown in Fig. 12, the concentration of solute in the precipitate, β, is given by $C^{\beta/\alpha}$, and the concentration of the matrix, α, in equilibrium with the precipitate is $C^{\alpha/\beta}$. Far from the precipitate the solute concentration is the initial alloy concentration, C^0. The growth rate of the precipitate is a moving boundary problem, and therefore, its equation is similar to Eq 51, except that the concentration gradient in the β phase is zero, and here, the coordinates are spherical. Therefore, a mass balance at the moving interface is:

$$\left(C^{\beta/\alpha} - C^{\alpha/\beta}\right)\frac{dR}{dt} = D^\alpha \frac{\partial C^\alpha}{\partial r}\bigg|_R$$
$$= D^\alpha \frac{\partial(A + B/r)}{\partial r}\bigg|_R = -D^\alpha \frac{B}{R^2} \qquad \text{(Eq 71)}$$

Because the interface is moving, this is a quasi-steady-state approximation. As before, A and B

are obtained from two positions on the concentration profile. In this case, the positions are at the moving interface, R, and at infinity:

$$C^{\alpha/\beta} = A + B/R$$
$$C^0 = A + B/\infty \qquad \text{(Eq 72)}$$

Solving for B and inserting it in Eq 71 yields:

$$\left(C^{\beta/\alpha} - C^{\alpha/\beta}\right)\frac{dR}{dt} = D^\alpha \frac{(C^0 - C^{\alpha/\beta})}{R} \qquad \text{(Eq 73)}$$

The notation in Eq 73 can be simplified by setting the concentration difference between the α and β phases at their interface as $\Delta C^{\alpha\beta}$ and the concentration difference in the α phase as ΔC^α. Then, Eq 73 can be written as:

$$\Delta C^{\alpha\beta}\frac{dR}{dt} = D^\alpha \frac{\Delta C^\alpha}{R} \qquad \text{(Eq 74)}$$

Integrating Eq 74 gives an expression for R as a function of time:

$$\Delta C^{\alpha\beta}\frac{1}{2}R^2 = tD^\alpha \Delta C^\alpha \qquad \text{(Eq 75)}$$

or

$$R = \sqrt{2tD^\alpha \frac{\Delta C^\alpha}{\Delta C^{\alpha\beta}}} \qquad \text{(Eq 76)}$$

Equation 76 shows that precipitates grow proportional to the square root of time. However, when the solute-depleted regions of nearby precipitates start to overlap, the growth slows and the square root dependence no longer applies. Also, the interface must move slowly enough so that the quasi-steady-state solution applies. This requires that:

$$\frac{\Delta C^\alpha}{\Delta C^{\alpha\beta}} \langle\langle 1.0 \qquad \text{(Eq 77)}$$

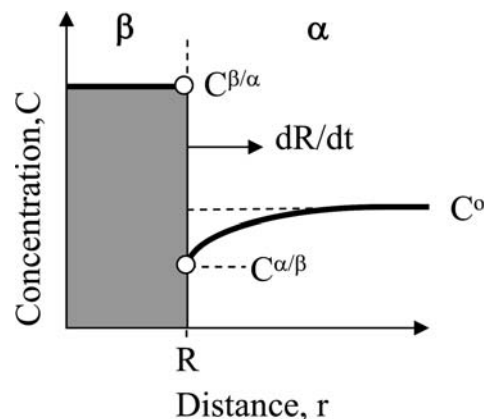A full explanation of this problem has been given by Zener (Ref 21).

Fig. 12 Concentration profile for a growing spherical precipitate modeled by Eq 74 to 77

Application—Precipitate Coarsening. When the β precipitates described in the previous application have grown and consumed most of the excess solute in the matrix phase, the alloy microstructure will coarsen to reduce the amount of interface energy in the system. The microstructure coarsens by larger precipitates growing at the expense of smaller precipitates that, with time, dissolve back into the α-matrix phase.

The classical theory of precipitate coarsening was developed independently by Lifschitz and Slyozov and by Wagner (the LSW theory). Although the mechanism of coarsening had been known before, they provided analytical equations for how the average particle size and the size distribution vary with time. Their temporal equation for the change in average precipitate size with time is:

$$\bar{R}^3 = \bar{R}_0^3 + \frac{4}{9}Kt \tag{Eq 78}$$

in which $\bar{R}_0$ is the average size of a spherical precipitate at time $t = 0$. Note that $\bar{R}_0$ and $t = 0$ can refer to any time during the coarsening process. The constant K was derived for a dilute α solution and for β, a pure-solute precipitate. In this case, K is given by:

$$K = \frac{2D^\alpha \sigma V_M^\beta C^{\alpha/\beta}}{RT} \tag{Eq 79}$$

in which D^α is the diffusivity in α, σ is the α/β interfacial tension, V_M^β is the molar volume of the β-phase, $C^{\alpha/\beta}$ is the α-concentration in equilibrium with a flat β-precipitate interface, and RT is the gas constant times absolute temperature.

The LSW theory uses a mean-field approach in which it is assumed that the matrix phase has a uniform concentration except for the vicinity around a precipitate, as shown in Fig. 13. The derivation makes use of the 3-D steady-state solution by using it to model the concentration profiles around each precipitate. Then, the growth or dissolution of each precipitate is given by a formula similar to Eq 74 in the previous section, except in this case, ΔC^α is a function of the precipitate radius:

$$\Delta C^{\alpha\beta} \frac{dR}{dt} = D^\alpha \frac{\Delta C^\alpha(R)}{R} \tag{Eq 80}$$

The value for $\Delta C^\alpha(R)$ depends on the Gibbs-Thomson equation, which gives the concentration of solute in the α-phase that is in equilibrium with a β-precipitate phase of a given radius. A general form of the Gibbs-Thomson equation that does not require assuming a dilute solution for α or pure solute for β is (Ref 22, 23):

$$\Delta C^{\alpha\beta} \frac{\partial^2 G^\alpha}{\partial C^2} \left(C^{\alpha/\beta} - C^\alpha(R) \right) = \frac{2\sigma V_M^\beta}{R} \tag{Eq 81}$$

in which $\Delta C^{\alpha\beta}$ is the concentration of β minus the concentration of α when the two phases are in equilibrium across a flat interface (i.e., the tie-line length from a phase diagram), and $\partial^2 G^\alpha/\partial C^2$ is

the second derivative of the Gibbs energy of the α-phase with respect to concentration.

The mechanism of coarsening makes use of Eq 81 by recognizing that larger precipitates will be able to draw solute away from smaller precipitates, because they maintain a lower solute concentration in their vicinity than do smaller precipitates. Smaller precipitates must dissolve to maintain a higher solute concentration in their vicinity.

The mean-field concentration, $\bar{C}^\alpha$, can be associated with a radius $\bar{R}$ via Eq 77 in:

$$\Delta C^{\alpha\beta} \frac{\partial^2 G^\alpha}{\partial C^2} \left(C^{\alpha/\beta} - \bar{C}^\alpha \right) = \frac{2\sigma V_M^\beta}{\bar{R}} \tag{Eq 82}$$

Subtracting Eq 81 from Eq 82 yields:

$$\Delta C^{\alpha\beta} \frac{\partial^2 G^\alpha}{\partial C^2} \left(\bar{C}^\alpha - C^\alpha(R) \right) = 2\sigma V_M^\beta \left(\frac{1}{\bar{R}} - \frac{1}{R} \right) \tag{Eq 83}$$

The bracketed term on the left side of the equation is $\Delta C^\alpha(R)$. Solving Eq 83 for $\Delta C^\alpha(R)$ and substituting it into Eq 80 yields one of the key equations in coarsening theory:

$$\left(\Delta C^{\alpha\beta} \right)^2 \frac{\partial^2 G^\alpha}{\partial C^2} \frac{dR}{dt} = \frac{2D^\alpha \sigma V_M^\beta}{R} \left(\frac{1}{\bar{R}} - \frac{1}{R} \right) \tag{Eq 84}$$

This, on rearrangement, is:

$$\frac{dR}{dt} = \left(\frac{2D^\alpha \sigma V_M^\beta}{\left(\Delta C^{\alpha\beta} \right)^2 \frac{\partial^2 G^\alpha}{\partial C^2}} \right) \left(\frac{1}{\bar{R}} - \frac{1}{R} \right) \tag{Eq 85}$$

The first bracket on the right side of Eq 85 is K in the LSW theory (Ref 24):

$$K = \left(\frac{2D^\alpha \sigma V_M^\beta}{\left(\Delta C^{\alpha\beta} \right)^2 \frac{\partial^2 G^\alpha}{\partial C^2}} \right) \tag{Eq 86}$$

Equation 86 reduces to the LSW form for dilute solutions by letting $\Delta C^{\alpha\beta} \cong 1$ and $\partial^2 G/\partial C^2 \cong RT/C^{\alpha/\beta}$. (Note: In these equations, C has units of mole fraction.)

A simplification of both the general and LSW theory equation for K is made by recognizing that D can be separated into a kinetic mobility term, M, and a thermodynamic term with the equation:

$$D = M \left(\partial^2 G/\partial^2 C \right) \tag{Eq 87}$$

It follows that:

$$K = \left(\frac{2M^\alpha \sigma V_M^\beta}{\left(\Delta C^{\alpha\beta} \right)^2} \right) \tag{Eq 88}$$

which shows that coarsening is a function of atomic mobilities and not free energy functions (Ref 24).

For multicomponent systems, K can be expressed in the rounded-bracket notation by a similar formula (Ref 24) :

$$K = \left(\frac{2\sigma V_M^\beta}{\left(\Delta C^{\alpha\beta} \right] [M^\alpha]^{-1} [\Delta C^{\alpha\beta}]} \right) \tag{Eq 89}$$

In Eq 89, the terms $\left(\Delta C^{\alpha\beta} \right]$ and $\left[\Delta C^{\alpha\beta} \right)$ are vectors that coincide with a tie line in the two-phase region of a multicomponent system. The average composition of the alloy lies on this tie line.

It can be seen that Eq 89 reduces to Eq 88 for a binary system when all the matrices are replaced with scalar quantities.

The Error Function Solution

The error function solution is the most useful of the constant D solutions, because it applies to a number of types of problems and can be
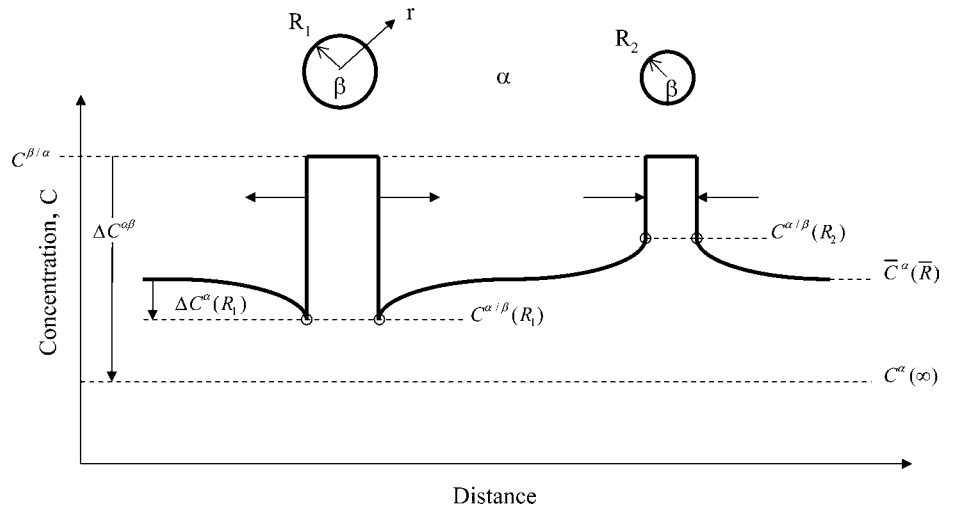

Fig. 13 Concentration profiles during coarsening. The larger β-precipitate grows, while the smaller precipitate dissolves. The scale of this diagram is distorted because the concentrations at the α/β interfaces are typically much smaller than $\Delta C^{\alpha\beta}$.

readily extended to multicomponent systems. Normally, it is restricted to 1-D problems, but, as is shown for carburizing, it can be used for certain problems by superimposing the 1-D solution in multiple directions.

Error Function Equations. The solution to Fick's second law can be written as:

$$C(x) = A + B\,\text{erf}(z) \tag{Eq 90}$$

in which A and B are constants that can be calculated from any two concentrations along the concentration profile, and z is a dimensionless variable defined by:

$$z = \frac{x}{2\sqrt{Dt}} \tag{Eq 91}$$

and $\text{erf}(z)$ is the error function that is defined by:

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-\eta^2)\,d\eta \tag{Eq 92}$$

The error function integral has no simple algebraic solution, but tables obtained by numerical methods are available in handbooks and on the internet. Also, there is an approximate equation that is accurate to within 1% for values of z up to 1.7. It is given by:

$$\text{erf}(z) \cong \frac{2}{\sqrt{\pi}} \frac{3z}{z^2 + 3} \tag{Eq 93}$$

Equation 93 is plotted as a line in Fig. 14 and is compared there with the more exact numerical values that are plotted as points. As shown in the figure, the error function becomes asymptotic to $\text{erf}(z) = -1$ and $+1$ for $z < -2$ and $z > 2.0$, respectively.

Exact expressions for both the differential of the error function and a definite integral of the complementary error function, $\text{erfc}(z) = 1 - \text{erf}(z)$, exist. The differential of $\text{erf}(z)$ is:

$$\left.\frac{d\,\text{erf}(z)}{dz}\right|_{z'} = \frac{2}{\sqrt{\pi}} \exp(-z'^2) \tag{Eq 94}$$

The definite integral of $\text{erfc}(z)$ is:

$$\int_0^\infty \text{erfc}(z)\,dz = \frac{1}{\sqrt{\pi}} \tag{Eq 95}$$

These are helpful relationships when deriving equations for applications.

Application—Interdiffusion of Diffusion Couples. Most experimental research work on diffusion is done through the use of diffusion couples. A diffusion couple is a tool that allows the engineer in a laboratory to study and measure the effect of interdiffusion between two alloys that are in intimate contact. The 1-D geometry simplifies the analysis, because concentration profiles can be measured and modeled with error functions as long as the diffusivity is approximately constant.

Concentration Profile Equation for Binary Systems. The error function solution to Fick's second law can be expressed as:

$$C(x,t) = A + B\,\text{erf}(z) \tag{Eq 96}$$

Evaluating A and B from concentrations at the left, C^L, and right, C^R, ends of the diffusion couple, yields:

$$C^L = A + B\,\text{erf}(-\infty) = A - B$$
$$C^R = A + B\,\text{erf}(+\infty) = A + B \tag{Eq 97}$$

And therefore:

$$A = \frac{C^R + C^L}{2} \tag{Eq 98}$$

$$B = \frac{C^R - C^L}{2} \tag{Eq 99}$$

It follows that the error function solution for a binary diffusion couple is:

$$C(x,t) = \left(\frac{C^R + C^L}{2}\right) + \left(\frac{C^R - C^L}{2}\right)\text{erf}(\frac{x}{2\sqrt{Dt}}) \tag{Eq 100}$$

The notation in Eq 100 can be simplified by recognizing that A is the average concentration of the diffusion couple, C^{Ave}, and $C^R - C^L$ is the initial concentration difference of the couple, ΔC^0. In this notation:

$$C(x,t) = C^{Ave} + \left(\frac{\Delta C^0}{2}\right)\text{erf}\left(\frac{x}{2\sqrt{Dt}}\right) \tag{Eq 101}$$

Amount of Interdiffusion Equation for Binary Systems. One measure of the extent of interdiffusion is given by the total amount of solute

that has crossed from one diffusion-couple alloy to the other. This amount is illustrated graphically in Fig. 15 and is obtained from the integral:

$$S = \int_0^\infty \left(C(x,t) - C^R\right)dx$$
$$= -\int_0^\infty \frac{\Delta C^0}{2}\text{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)dx \tag{Eq 102}$$

$$S = -\Delta C^0 \sqrt{\frac{Dt}{\pi}} \tag{Eq 103}$$

The minus sign in Eq 103 indicates that when ΔC^0 is positive, S is negative, and solute is leaving the right side of the diffusion couple. However, when S is positive, solute is entering the right side. Equation 103 shows that to reduce the amount of interdiffusion, one must reduce the initial concentration difference, reduce the diffusivity (e.g., by reducing the temperature), and/or reduce the time. As seen in the section "Concentration Profile Equations for Multiphase Systems" in this article, the amount of interdiffusion can be reduced in ternary and higher-order systems by adjusting the alloy compositions.

Concentration Profile Equations for Single-Phase Ternary Systems. Concentration profile equations for multicomponent systems are similar to that for binary systems, except that a concentration profile for an n-component system contains $n - 1$ error functions, and there are $n - 1$ concentration profiles to be
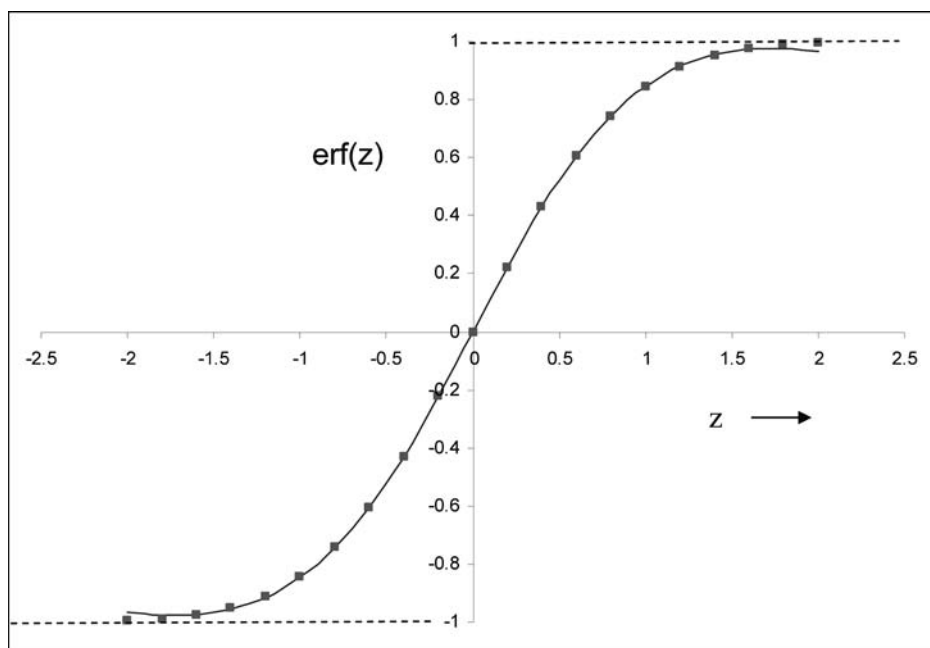

Fig. 14 Approximate error function (line) compared with the numerical solution (points). The numerical solution for erf(z) becomes asymptotic to ± 1.0, while the approximate solution does not. However, the error in the approximate solution is less than 0.3% up to $z = 1.5$.
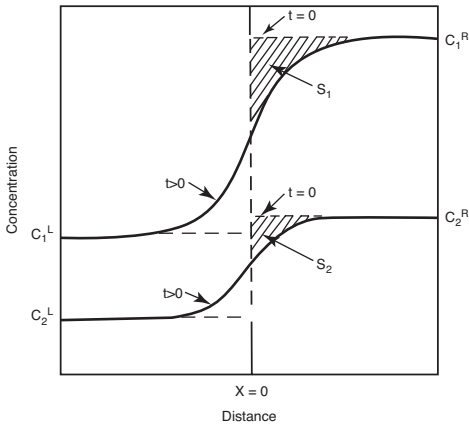
LIVE GRAPH
Click here to view

Fig. 15 Illustration of the amount of interdiffusion as defined by the solute, S, that has entered the right side of the couple from the left side. In both cases shown in the figure, the value of S is negative because solute has left the right side. Source: Adapted from Ref 5

considered. Any one of the profiles has an equation with the form:

$$
\begin{aligned}
C_i(x,t) = C_i^{\text{Ave}} + B_{i1}\text{erf}(z_1) + B_{i2}\text{erf}(z_2) \\
+ B_{i3}\text{erf}(z_3).....B_{in-1}\text{erf}(z_{n-1})
\end{aligned}
$$
(Eq 104)

As in binary systems, the constants are functions of the initial alloy compositions, but unlike binaries, they are functions also of the diffusivity matrix via the $[\alpha]$ and $[\alpha]^{-1}$ matrices.

As indicated previously, there are two concentration equations for a ternary system and two error functions in each equation. Written together in matrix notation, they are given by:

$$
\begin{aligned}
[C(x,t)) = [C^{\text{Ave}}) \\
+ \frac{1}{2}[\alpha_1)^{-1}(\alpha_1)[\Delta C^0]\text{erf}\left(\frac{x}{2\sqrt{\tilde{D}_1 t}}\right) \\
+ [\alpha_2)^{-1}(\alpha_2)[\Delta C^0]\text{erf}\left(\frac{x}{2\sqrt{\tilde{D}_2 t}}\right)
\end{aligned}
$$
(Eq 105)

The two diffusivities associated with the two error functions are eigenvalues of the diffusivity matrix $[D]$. For short annealing times, the first error function dominates because it contains the major eigenvalue, which is the largest of the kinetic constants. Equations similar to these were derived by Cooper and Gupta (Ref 25).

Equation 105 illustrates the advantage of the rounded-bracket notation. For example, terms such as $(\alpha_1)[\Delta C^0]$ are quickly identified as scalar quantities, and one sees that the first error function contributes concentration profiles along the major eigenvector direction, $[\alpha_1)^{-1}$, while the second error function contributes concentration profiles along the minor eigenvector direction.

Whenever the composition vector $[\Delta C^0]$ falls in one of the eigenvector directions, for

example, when $[\Delta C^0] \propto [\alpha_1)^{-1}$, then $(\alpha_2)[\Delta C^0]$ goes to zero. Equation 105 reduces to a single error function expression along the other eigenvector direction, $[\alpha_1)^{-1}$. In this case, the concentration profile for each solute is similar to that for a binary diffusion couple, and the diffusion path plotted on a ternary phase diagram is a straight line. When diffusion couples have composition vectors in other directions, the concentration profiles are more varied and can have extrema and zero slope at $x = 0$. Example ternary concentration profiles have already been given in Fig. 1 and 3.

When the composition vector is in other directions, the diffusion path will be serpentine. The ends of the diffusion path will have an initial slope given by the major eigenvector. Because the major eigenvector is associated with the largest diffusion constant, it dominates the diffusion path at the outer fringes of the interdiffusion region.

Concentration Profile Equations for Multiphase Systems. The diffusion equation for multiphase systems given by Eq 39 reduces to an equation similar to the extended form of Fick's second law (Eq 25) by assuming that the effective diffusivity is constant. The result is (Ref 7):

$$
\left[\frac{\partial \bar{C}}{\partial t}\right] = \left[D^{\text{eff}}\right]\left[\frac{\partial^2 \bar{C}}{\partial x^2}\right]
$$
(Eq 106)

It follows that the solution to this equation is the same as for the single-phase equation, except that concentrations will be averaged over two phases, and $[D^{\text{eff}}]$ will have one positive and one zero eigenvalue. The result is (Ref 7):

$$
\begin{aligned}
[\bar{C}(x,t)) = [\bar{C}^{\text{Ave}}) \\
+ \frac{1}{2}[\alpha_1)^{-1}(\alpha_1)[\Delta\bar{C}^0]\text{erf}\left(\frac{x}{2\sqrt{\tilde{D}_1 t}}\right) \\
+ [\alpha_2)^{-1}(\alpha_2)[\Delta\bar{C}^0]\text{erf}\left(\frac{x}{2\sqrt{0}}\right)
\end{aligned}
$$
(Eq 107)

The last term in Eq 103 is a step function that changes on crossing $x = 0$, because:

$$
\begin{aligned}
\text{erf}\left(\frac{x}{2\sqrt{0}}\right) = +1 \text{ for } x > 0 \\
= -1 \text{ for } x < 0
\end{aligned}
$$
(Eq 108)

It follows that a diffusion couple prepared from alloys in the two-phase region of a ternary phase diagram will have a step in average composition at the initial interface. The only exception will be when the composition vector is in the major eigenvector direction. Then, the coefficient of the step term is zero. However, if the composition vector is in the minor eigenvector direction, then the coefficient of the first error function term will be zero. As a result, there will be a step but no interdiffusion. The reason there is no interdiffusion is because both alloys will be on the same tie

line. Therefore, all phases in the alloys will be in equilibrium, and there will be no driving force for change except for that due to capillarity.

Amount of Interdiffusion Equation for a Ternary System and the Square Root Diffusivity. Equations for the amount of interdiffusion in a ternary system are similar to that for binaries, except that the term:

$$
\sqrt{D}
$$

is replaced by the square root of the diffusivity matrix:

$$
\sqrt{[D]}
$$

The notation used here for the square root of the diffusivity matrix is $[r]$. It is defined by the equation:

$$
[D] = [r][r]
$$
(Eq 109)

For a multicomponent system, the equation for the amount of interdiffusion written in matrix notation is (Ref 26):

$$
[S] = -\sqrt{\frac{t}{\pi}}[r][\Delta C^0]
$$
(Eq 110)

Written out for a ternary system, the equations are:

$$
S_1 = -\sqrt{\frac{t}{\pi}}\left(r_{11}\Delta C_1^0 + r_{12}\Delta C_2^0\right)
$$
(Eq 111)

$$
S_2 = -\sqrt{\frac{t}{\pi}}\left(r_{21}\Delta C_1^0 + r_{22}\Delta C_2^0\right)
$$
(Eq 112)

Reducing Interdiffusion and Zero-Flux Planes. According to Eq 111 and 112, the amount of solute either leaving or entering the right side of a diffusion couple can be manipulated by varying concentration differences between the alloys in a diffusion couple. The amount can be reduced to zero by setting Eq 111 or 112 equal to zero. For example, S_1 will be zero when:

$$
\Delta C_2^0 = -\Delta C_1^0 \frac{r_{11}}{r_{12}}
$$
(Eq 113)

Under these conditions, $x = 0$ will be a zero-flux plane for component 1. Although there will be flux of component 1 on either side of the initial interface, the total amount of solute 1 on either side will stay the same, and therefore, no flux will cross the initial interface between the diffusion-couple alloys.

Application—Interdiffusion of Coatings. For short annealing times, the interdiffusion of a coating with its substrate and the interdiffusion of a diffusion couple are the same. The equations for the amount of solute crossing the initial diffusion-couple interface, Eq 103 and 110 to 112, are the same equations for the amount of solute diffusing from a coating into its substrate. In this case:

$$\Delta C^0 = C^{\text{coat}} - C^0 \qquad \text{(Eq 114)}$$

in which C^0 is the initial solute concentration of the substrate, and C^{coat} is the initial solute concentration of the coating. As in the case of diffusion couples, it is possible to reduce or eliminate the loss of a component due to interdiffusion by making the coating/substrate interface a zero-flux plane. However, once the interdiffusion region extends to the free surface of the coating, a flux will develop across that interface, but it will be from the substrate into the coating, at least initially.

For binary systems, there is an error function equation that includes the thickness of the coating. It involves superimposing two error functions. For a coating with thickness $2h_0$ (Ref 20):

$$C(x,t) = C^0 + \frac{\Delta C^0}{2}\left[\text{erf}\left(\frac{x+h_0}{2\sqrt{Dt}}\right) - \text{erf}\left(\frac{x-h_0}{2\sqrt{Dt}}\right)\right] \qquad \text{(Eq 115)}$$

In this equation, $x = 0$ is the free surface of the coating, as indicated in Fig. 16. In addition, the boundary condition that the concentration gradient at the surface must be zero at all times is satisfied by Eq 115. The flux must be zero there, because, presumably, no solute leaves the coating at the free surface.

In addition to the amount of solute that interdiffuses between the coating and substrate for early times, another measure of interdiffusion is the concentration of solute at the coating surface. Evaluating Eq 115 at $x = 0$ gives:

$$C(0,t) = C^0 + \Delta C^0 \text{erf}\left(\frac{h_0}{2\sqrt{Dt}}\right) \qquad \text{(Eq 116)}$$

Equation 108 indicates that there is little change in the surface concentration until the argument of the error function is less than 1.5, which occurs when $t > h_0^2/9D$, as shown in Fig. 14. According to Eq 116 and Fig. 14, it is not until $t \cong h^2/D$ that the solute excess in the coating $(C^{\text{coat}} - C^0)$ has dropped to half of its original value.

Application—Carburizing, Decarburizing, and the Corner Effect. Modeling carburizing is much like modeling interdiffusion. However, to apply the error function solution, both the carbon concentration at the surface and the diffusivity must be a constant. Unfortunately, when carburizing steels, neither of these conditions are met. Carburizing normally starts with a transient period as the carbon content at the surface increases to approach the equilibrium value. Also, the diffusivity of carbon steels can vary by a factor of 2 across a carburized region (Ref 27).

To minimize the error when modeling carburizing, it is best to use a numerical method in which the transient period and variable diffusivity are modeled (Ref 28). If a rough estimate is sufficient, then the constant D solution can be used. However, the diffusivity at the surface concentration should be used. That means that

a high diffusivity corresponding to the high carbon content at the surface should be used when modeling carburizing, and a low diffusivity corresponding to a low carbon concentration should be used when modeling decarburizing.

The equation for carburizing and decarburizing is readily derived by solving for A and B in Eq 96 in terms of the surface concentration, $C_{\text{C}}^{\text{surface}}$, and the initial carbon concentration of the sample, C_{C}^0. The result for either carburizing or decarburizing is:

$$C_{\text{C}}(x,t) = C_{\text{C}}^{\text{surface}} - (C_{\text{C}}^{\text{surface}} - C_{\text{C}}^0)\text{erf}\left(\frac{x}{2\sqrt{D_c t}}\right) \qquad \text{(Eq 117)}$$

Equation 117 gives the carbon concentration versus distance along a line perpendicular to a flat surface.

The aforementioned is a solution to the 1-D carburizing or decarburizing problem. The error function can be used for 2-D or 3-D problems as well if the 1-D solution can be superimposed in different directions (Ref 20). One example is to calculate the carbon concentration at an external corner of a part. The deeper amount of carburization there is called the corner effect. The 3-D equation for carbon concentration versus distance at and near the corner is:

$$C_{\text{C}}(x,y,z,t) = C_{\text{C}}^{\text{surface}} - (C_{\text{C}}^{\text{surface}} - C_{\text{C}}^0)$$
$$\left[\text{erf}\left(\frac{x}{2\sqrt{D_c t}}\right) + \text{erf}\left(\frac{y}{2\sqrt{D_c t}}\right) + \text{erf}\left(\frac{z}{2\sqrt{D_c t}}\right)\right] \qquad \text{(Eq 118)}$$

The corner in this case is at $x = y = z = 0$, and it is assumed that the carbon concentration at the surface is a constant. However, in practice, the carbon concentration at the surface can be higher at the corner (Ref 29), in which case Eq 118 would underestimate the corner effect. The corner effect also refers to edges, which is a 2-D problem that requires only two error functions for x- and y-directions.

Layer Growth between Bonded Alloys. When bonded alloys are multiphase and multicomponent, there are special difficulties when modeling interdiffusion. The extension of single-phase problems from binary to multicomponent systems is straightforward by using matrix algebra. However, when more than one single or multiphase alloy is involved and when layers form between the initial alloys, then equations become considerably more complex.

Technical examples where layers form include at coating/substrate interfaces, soldered and braised joint interfaces, and internal composite interfaces. The layers can be predicted in binary systems from a phase diagram whenever all intermediate phases form. An example is with the diffusion couple at the lowest temperature in Fig. 17. Here, a Ni-4%Al/Ni-61% Al diffusion couple is depicted with the initial alloys (circles) joined by a straight line. It can be shown by free-energy-versus-concentration curves that each single-phase region is expected to form a separate layer in the same sequence as

on the phase diagram. In the example, the phases present will be in the sequence Ni$_2$Al$_3$/NiAl/Ni$_3$Al/(Ni). No two-phase regions can form, because the concentrations of the phases in a two-phase region are constant, and therefore, two-phase regions can only shrink with time. In ternary and higher-order systems, the concentration of phases does vary in two-phase regions, and such regions do grow and form intermediate layers.

If one or more phases are missing for kinetic reasons, then the sequence will stay the same, except without the missing phases. However, then it is possible for additional metastable phases to form and appear in the sequence where stable phases did not form.

A number of interdiffusion problems involving multiphase regions and single-phase layers have been solved for binary systems (Ref 22, 30). The additional diffusion couples in Fig. 17 illustrate the types of diffusion couples that were treated in that reference. The solution to the single-phase couple in this group has already been given by Eq 101. Compare the simplicity of that equation with the solution for the diffusion couple in the reference that involves a single-phase layer. The solution to that problem is given by the following three equations. The solution considers that there is an α + β alloy on the left, a β + γ on the right, and a single-phase β-layer growing between them. The first equation for the concentration profile is obtained by solving the general error function equation for the constants A and B by using the concentrations at the two moving boundaries:

$$C^\beta(x,t) = C^{\beta/\gamma}\left(\frac{\text{erf}\left(\frac{\eta_1}{2\sqrt{D^\beta t}}\right) - \text{erf}\left(\frac{x}{2\sqrt{D^\beta t}}\right)}{\text{erf}\left(\frac{\eta_1}{2\sqrt{D^\beta t}}\right) - \text{erf}\left(\frac{\eta_2}{2\sqrt{D^\beta t}}\right)}\right)$$
$$+ C^{\beta/\gamma}\left(\frac{\text{erf}\left(\frac{x}{2\sqrt{D^\beta t}}\right) - \text{erf}\left(\frac{\eta_2}{2\sqrt{D^\beta t}}\right)}{\text{erf}\left(\frac{\eta_1}{2\sqrt{D^\beta t}}\right) - \text{erf}\left(\frac{\eta_2}{2\sqrt{D^\beta t}}\right)}\right) \qquad \text{(Eq 119)}$$

The following two equations give positions of the moving interfaces. These are obtained by solving Eq 51 for the moving-boundary
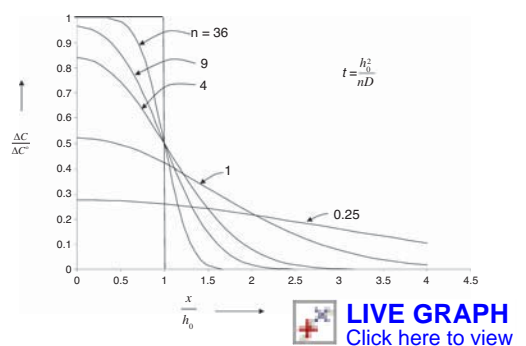

Fig. 16 Variation of concentration profiles versus time for a coating modeled with Eq 115. On the y-axis, ΔC^0 is the initial coating concentration, C^{Coat}, minus the initial substrate concentration, C^0.

velocity and then integrating with respect to time to determine the position. When solving Eq 51, there is only a flux in the β-layer, because the concentrations in the two-phase alloys are constant. The results are:

$$\eta_1 = \sqrt{\frac{4D^\beta t}{\pi}} \frac{\Delta C^\beta}{\Delta C^{(\alpha+\beta)\beta}} \frac{\exp\left(-\frac{\eta_1^2}{4D^\beta t}\right)}{\text{erf}\left(\frac{\eta_1^2}{4D^\beta t}\right) - \text{erf}\left(\frac{\eta_2^2}{4D^\beta t}\right)} \quad \text{(Eq 120)}$$

$$\eta_2 = \sqrt{\frac{4D^\beta t}{\pi}} \frac{\Delta C^\beta}{\Delta C^{\beta(\beta+\gamma)}} \frac{\exp\left(-\frac{\eta_2^2}{4D^\beta t}\right)}{\text{erf}\left(\frac{\eta_1^2}{4D^\beta t}\right) - \text{erf}\left(\frac{\eta_2^2}{4D^\beta t}\right)} \quad \text{(Eq 121)}$$

in which ΔC^β is the concentration difference across the single-phase β-region, while $\Delta C^{(\alpha+\beta)\beta}$ and $\Delta C^{\beta(\beta+\gamma)}$ are concentration differences across the interfaces.

The extension of this problem to ternaries is straightforward, but there are additional equations that must be added to account for the concentration profiles in the two-phase regions and for the equilibrium phase diagram boundaries. These types of problems do not lend themselves to finding solutions to simultaneous equations.

Instantaneous Source

The instantaneous source (Ref 20), sometimes referred to as the limited source, is a source of solute that is initially confined to a localized region that is a point, a line, or a plane. For example, the region for a plane source can be envisioned as a thin plate with thickness δx. However, the plane source of solute refers to the limit as $\delta x \to 0$. As a result, equations for an instantaneous source have the property that the concentration goes to infinity at zero time (this type of plane source is also known as a Dirac delta function). Such behavior may seem nonphysical, but the instantaneous source equations are not used at zero time. Instead, they are applied to longer time, when $t > h_0^2/(4D)$ and the plane source equation approaches the solution for a source of finite thickness, h_0. Its appeal is that the instantaneous source equation is far simpler than the finite thickness source equation, as seen in the following. The general equation for instantaneous sources is:

$$C(x,t) = A + B\exp\left(-\frac{r^2}{4Dt}\right) \quad \text{(Eq 122)}$$

in which r is the distance from the instantaneous source. For a point source, r is the radius of a sphere; for a line source, it is the radius of a cylinder; and for a plane source, it is the distance, x, from the plane. Substituting this equation into the 1-, 2-, and 3-D versions of Fick's second law, one finds that the value of A is the concentration of the surrounding material, and B is obtained by solving the equation for the total solute, S_T:

$$S_T = \int_{-\infty}^{\infty} \left(C(x,t) - C^0\right)dx \quad \text{(Eq 123)}$$

which is:

$$B = \frac{S_T}{2^\phi(\pi Dt)^{\phi/2}} \quad \text{(Eq 124)}$$

In Eq 124, ϕ is the dimension of the version of Fick's second law that applies, and S_T is a constant equal to the total solute in the source.

One-Dimensional Thin-Film Solution. The thin-film solution gets its name from the initial condition, which can be viewed as a thin film of solute caught between two alloys or pure metals. The equation models the concentration profile as the solute diffuses into its surroundings with time. In this case, $\phi = 1$ corresponds to 1-D, and the equation is:

$$C(x,t) = C^0 + \frac{S_T}{2\sqrt{\pi Dt}}\exp\left(-\frac{x^2}{4Dt}\right) \quad \text{(Eq 125)}$$

Application—Diffusion Bonding. The bonding of two alloys to each other is often facilitated by placing a thin sheet or film of metal between them, as is done in braising. However, unlike braising, in which the braise material remains in the joint, diffusion-bonded parts are heat treated to diffuse the thin film into the metal, to where it no longer has an appreciable concentration and effect on properties. Accordingly, the interdiffusion model applies to long times, which is when the thin-film solution is most accurate.

For a film of thickness $2h_0$ and a concentration of $C^{\text{thin-film}}$, the thin-film solution is:

$$C(x,t) = C^0 + \frac{2h_0\left(C^{\text{thin-film}} - C^0\right)}{2\sqrt{\pi Dt}}\exp\left(-\frac{x^2}{4Dt}\right) \quad \text{(Eq 126)}$$

An example of the thin-film solution is shown in Fig. 18. The maximum concentration occurs at $x = 0$, where the concentration is:

$$C(0,t) = C^0 + \frac{h_0\left(C^{\text{thin-film}} - C^0\right)}{\sqrt{\pi Dt}} \quad \text{(Eq 127)}$$

It is apparent that at $t = 0$, the equation goes to infinity and does not apply to the physical problem. However, after a time of $t = h_0^2/4D$, Eq 127 for the maximum concentration is
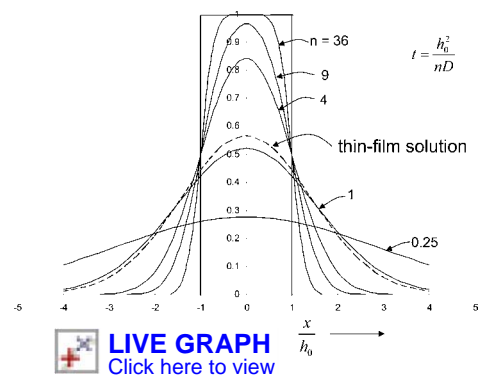

Fig. 17 Various types of diffusion couples, each of which is modeled with different equations

Fig. 18 Diffusion-bonding concentration profiles predicted by the error function Eq 115. For $n = 1$, the thin-film solution and the error function solution are nearly the same. For $n = 0.25$, the profiles appear to overlap.
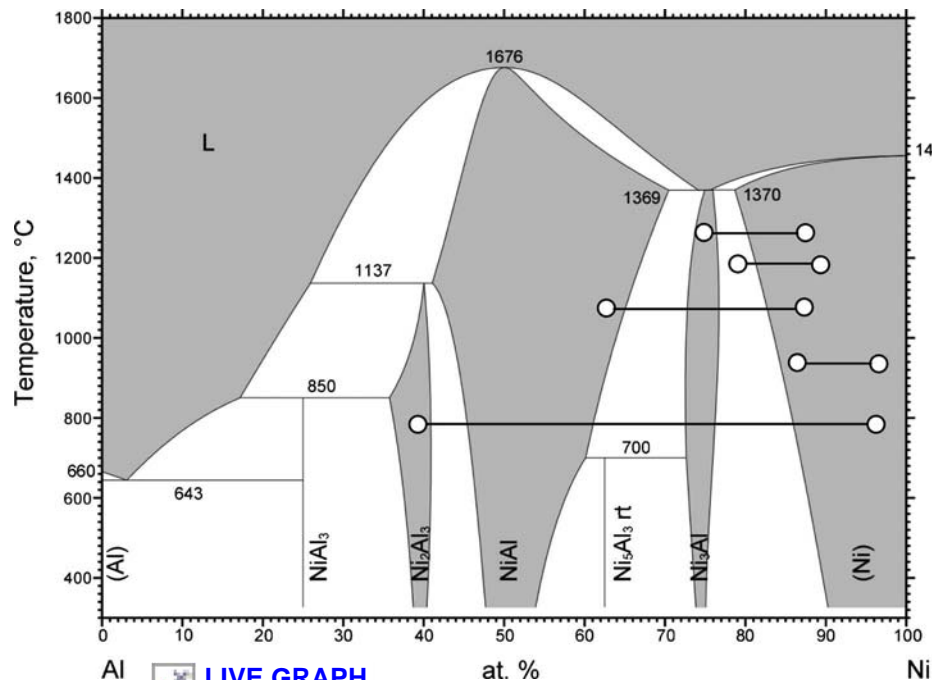
equivalent to that of the more exact error function equation for the maximum given by Eq 116. Figure 18 plots the error function solution Eq 115 as it applies to diffusion bonding and shows that when the maximum concentration reaches approximately half its initial value, Eq 116 is nearly the same as Eq 127 and, for all practical purposes, is identical for longer times.

If solute concentrations are different in the two alloys, a solution can be obtained by superimposing the error function solution for a diffusion couple, Eq 101, with the thin-film solution given previously to obtain:

$$C(x,t) = C^{\text{Ave}} + \frac{\Delta C^0}{2}\text{erf}\left(\frac{x}{2\sqrt{Dt}}\right) + \frac{h_0\left(C^{\text{thin-film}} - C^{\text{Ave}}\right)}{\sqrt{\pi Dt}}\exp\left(\frac{x^2}{4Dt}\right) \tag{Eq 128}$$

in which C^{Ave} is the average of the two alloy concentrations, and ΔC^0 is the concentration difference between the alloys.

Series Solution

The series solution, or trigonometric solution, as it is sometimes called, is similar to the thin-film solution in that it provides simple solutions to diffusion problems during the later stages of a diffusional process. It is particularly helpful in 1-D problems in systems involving periodic changes in concentration and in thin plates that have equal and constant boundary conditions on each side of the plate.

The general equation is an infinite series of the type:

$$C(x,t) = \sum_{m=1}^{\infty}\left(A_m \sin \beta_m x + B_m \cos \beta_m x\right)\exp\left(-\beta_m^2 Dt\right) \tag{Eq 129}$$

The coefficients A_m, B_m, and β_m are determined by performing a Fourier analysis so that the coefficients fit the initial concentration profile in the part. The most common initial condition is a square wave, as shown in Fig. 19. It can be modeled with the series:

$$C(x,t) = C^{\text{Ave}} + \frac{4\Delta C^0}{\pi}\sum_{n=0}^{\infty}\frac{1}{2n+1}$$
$$\times \exp\left(-\frac{(2n+1)^2\pi^2}{\lambda^2}Dt\right)\sin\frac{(2n+1)\pi x}{\lambda} \tag{Eq 130}$$

In Eq 130, ΔC^0 is the initial deviation of the square wave from the average concentration, C^{Ave}, and $2\lambda/(2n+1)$ is the wavelength of the nth sine wave in the series. Equation 130 can be greatly simplified by applying it to times when the terms associated with $n \geq 1$ have become insignificant with respect to the primary wave associated with $n = 0$. This occurs when the exponential term is less than 0.05, which occurs when:

$$\left(\frac{3\pi}{\lambda}\right)^2 Dt \geq 3 \tag{Eq 131}$$

At that time, the exponential term of the primary wave is still greater than 0.70 and therefore still significant.

Ignoring $n \geq 1$ terms, Eq130 simplifies to a single sine wave:

$$C(x,t) = C^{\text{Ave}} + \frac{4\Delta C^0}{\pi}\exp\left(-\frac{\pi^2}{\lambda^2}Dt\right)\sin\frac{\pi x}{\lambda} \tag{Eq 132}$$

The result is a sine wave with amplitude $A(t)$, given by:

$$A(t) = \frac{4\Delta C^0}{\pi}\exp\left(-\frac{\pi^2}{\lambda^2}Dt\right) \tag{Eq 133}$$

Application—Homogenization of Casting Segregation. The segregation in castings tends to be periodic because of the shape of dendrites. During the later stages of a homogenizing heat treatment, the concentration profile between secondary dendrite arms becomes sinusoidal, and the segregation is reduced to a fraction of its original value. The fraction according to Eq 133 will be:

$$\frac{A(t)}{\Delta C^0} = \frac{4}{\pi}\exp\left(-\frac{(2\pi)^2}{d_{\text{sda}}^2}Dt\right) \tag{Eq 134}$$

In Eq 134, the half-wavelength, λ, has been replaced with half the distance between centers of secondary dendrite arms, $d^{\text{sda}}/2 = \lambda$. Setting the exponent in Eq 134 equal to -3.0 and solving for time yields an equation for when the segregation has been reduced to approximately 0.5% of its original value:

$$t_{0.5\%} = \frac{3d_{\text{sda}}^2}{4\pi^2 D} \cong \frac{d_{\text{sda}}^2}{4\pi D} \tag{Eq 135}$$

Application—Carburization of a Thin Foil. To determine the equilibrium concentration of a steel in a particular carburizing atmosphere, it is common to include a thin foil of metal with the parts. The objective is to saturate the foil with carbon to evaluate the carbon activity in the atmosphere. The series solution can be used to predict the maximum foil thickness that can be saturated in such a test.

The initial condition for carburizing is illustrated in Fig. 20. The initial carbon concentration of the steel foil is C^0. The concentrations on either side of the foil are assumed to be at the saturated concentrations, C^{sat}. This initial condition corresponds to a square wave in which the amplitude is negative. In Eq 132, C^{sat} is equivalent to C^{Ave}, while $\Delta C^0 = C^0 - C^{\text{sat}}$. Therefore, the concentration across the foil as it approaches saturation is given by:

$$C(x,t) = C^{\text{sat}} + \frac{4\Delta C^0}{\pi}\exp\left(-\frac{\pi^2}{h_0^2}Dt\right)\sin\frac{\pi x}{h_0} \tag{Eq 136}$$

in which h_0 is the thickness of the foil. The time needed to saturate the foil is given by the time to reduce the amplitude of the sine wave to a small value. As in the previous example, an exponent of -3.0 will reduce the amplitude to 0.5% of its original value. For a fixed carburizing time, t^{carb}, the equations predict that the maximum foil thickness should be:

$$h_0^{\max} \leq \sqrt{\pi Dt^{\text{carb}}} \tag{Eq 137}$$

For example, the thickness of a pure iron foil carburized at 950 °C for 1 h should be:

$$h_0^{\max} \leq \sqrt{\pi(0.1\text{ cm}^2/\text{s})\exp(-136{,}000/8.314/(950+273))(3600\text{ s})}$$
$$= 0.05\text{ cm} \tag{Eq 138}$$

Note that the result is independent of the level of carburizing. It only depends on the diffusivity and the carburizing time.
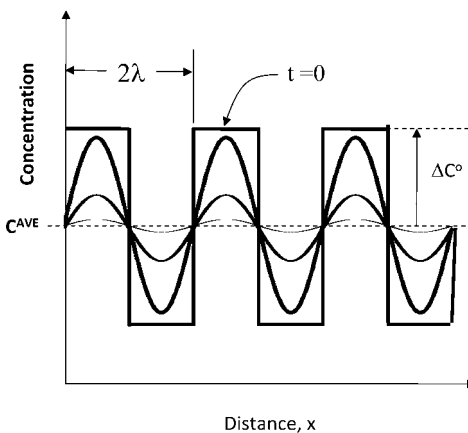

Fig. 19 Concentration profiles as a function of time for a profile that starts as a square wave. After a time of $t = \lambda/(3\pi^2 D)$, the concentration profile looks like a sine wave that decreases in amplitude with time.
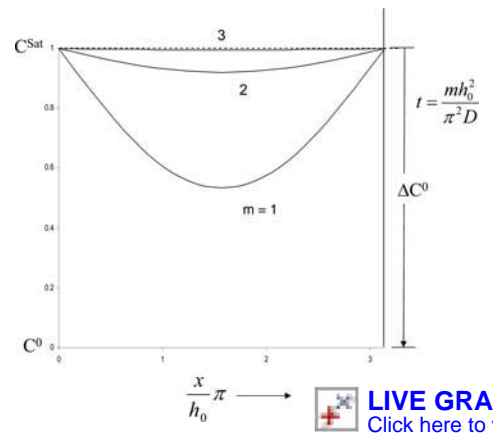

Fig. 20 Concentration profiles for carburizing a thin foil of metal as a function of time. At zero time, the film has a concentration of C^0, except at the surface where the atmosphere holds the concentration at C^{sat}.

For pure iron and for most steels, Eq 137 gives a conservative estimate. The reason is that the diffusivity of carbon in iron increases with carbon content by as much as a factor of 2 (Ref 27). Therefore, the small value of diffusivity will underestimate the maximum thickness.

Modeling Variable *D*, Multicomponent, and Multiphase Diffusion Problems

In the previous section, a number of examples were given in which solutions to diffusion problems were fashioned from four solutions to Fick's second law. For the most part, the solutions were applied to binary systems, although in several cases the binary solutions could be extended to multicomponent and multiphase systems without making the solution overly complex. All of the solutions had constant boundary conditions and assumed that the diffusivity was constant. In Crank's book (Ref 20), there are additional solutions that permit diffusivities to vary with composition and boundary conditions to vary with time. However, for industrial problems that need greater accuracy, the use of commercial codes is necessary.

The use of commercial codes often require extensive training. Many people familiar with the codes received their training in graduate school and used the codes as part of their doctoral work. Others train themselves by attending workshops and reading handbooks on the subject. Regardless of how training on the codes is received, a good foundation of thermodynamics and kinetics is essential for understanding the inputs to the programs and for assessing the validity of the outputs, because programs may produce outputs that contain artifacts, and it is important to recognize them.

Artifacts are most easily detected for ternary systems by plotting composition data as diffusion paths on phase diagrams. If the program assumes local equilibrium, then rules must be followed for how the diffusion path can appear (see the section "Zero-Flux Planes" in this article as well as Ref 15, 31, and 32). Special attention should be paid to predictions at boundaries between single-phase and multiphase regions.

The advantage of programs that assume local equilibrium is that there are definite rules that diffusion paths must follow, whereas more general assumptions are less restrictive, and violations may be more difficult to recognize. Local equilibrium is a simplification that avoids the need to include nucleation theory, because it assumes that precipitates form instantly when there is supersaturation. In addition, it assumes that interphase boundaries are infinitely mobile and that within a differential volume element the diffusion distances are so small that the time to equilibrate occurs between each time step. Without assuming local equilibrium, modelers are faced with having to include additional models for nucleation and interface mobility. Here, the most formidable problems may be to obtain viable databases to run the programs and to obtain the type of results that can be verified by experiment.

REFERENCES

1. H. Mehrer, Ed., *Diffusion in Metals and Alloys,* Vol 26, Landolt-Bornstein, New Series, Springer-Verlag, New York, 1990
2. I. Kaur, W. Gust, and L. Kozma, Ed., *Handbook of Grain Boundary and Interphase Boundary Diffusion Data,* Vol 1, Ziegler Press, Stuttgart, 1989
3. W.F. Gayle and T.C. Totemeir, Ed., *Smithells Metal Reference Book,* 8th ed., Elseveir, Burlington, MA, 2004
4. *Diffusion and Defect Forum,* Scientific.Net, Materials Science and Engineering, http://www.scientific.net/1012-0386/
5. M.S. Thompson, J.E. Morral, and A.D. Romig, *Metall. Trans. A,* Vol 21, 1990, p 2679
6. M.E. Glicksman, *Diffusion in Solids,* Wiley, New York, 2000
7. W.D. Hopfe and J.E. Morral, *Acta Metall. Mater.,* Vol 42 (No. 11), 1994, p 3887
8. F.A. Nichols, *J. Nucl. Mater.,* Vol 84, 1979, p 1
9. M.S. Thompson and J.E. Morral, *Acta Metall.,* Vol 34, 1986, p 339
10. M.A. Dayananda and C.W. Kim, *Metall. Trans. A,* Vol 10, 1979, p 1333
11. M.S. Thompson and J.E. Morral, *Metall. Trans. A,* Vol 18, 1987, p 1565
12. M.E. Glicksman and A.O. Lupulescu, *Acta Mater.,* Vol 51, 2003, p 1181
13. Y.H. Son and J.E. Morral, *Metall. Trans. A,* Vol 20, 1989, p 2299
14. J.E. Morral, C. Jin, A. Engstrom, and J. Agren, *Scr. Mater.,* Vol 34, 1996, p 1661
15. W.J. Boettinger et al., *Acta Mater.,* Vol 48, 2000, p 481
16. J.E. Morral and H. Chen, *Scr. Mater.,* Vol 43, 2000, p 699–703
17. C. Jin and J.E. Morral, *Scr. Mater.,* Vol 33, 1997, p 621–626
18. S.M. Merchant, M.R. Notis, and J.I. Goldstein, *Metall. Trans. A,* Vol 21, 1990, p 1901
19. L.A. Carol, M.S. Thesis, Michigan Technological University, 1984, and NASA Contractor Report 174852, 1985
20. J. Crank, Ed., *The Mathematics of Diffusion,* 2nd ed., Oxford Scientific Pub., 1975
21. C. Zener, *J. Appl. Phys.,* Vol 20, 1949, p 950
22. M. Hillert, *Lectures on the Theory of Phase Transformations,* 2nd ed., H.I. Aaronson, Ed., TMS, Warrendale, PA, 1999, p 1
23. J.E. Morral and G. Purdy, *J. Alloy. Compd.,* Vol 220, 1995, p 132
24. J.E. Morral and G. Purdy, *Scr. Metall. Mater.,* Vol 30, 1994, p 905
25. P.K. Gupta and A.R. Cooper, *Physica,* Vol 54, 1971, p 39
26. M.S. Thompson and J.E. Morral, *Acta Metall.,* Vol 34, 1986, p 220
27. J. Agren, *Scr. Metall.,* Vol 20, 1986, p 1507
28. J.E. Morral, B.M. Dupen, and C.C. Law, *Metall. Trans. A,* Vol 23, 1994, p 2069
29. K.D. Jones and G. Krauss, *Heat Treatment '79,* 1980, p 188–193
30. R.F. Sekerka and S.-L. Wang, *Lectures on the Theory of Phase Transformations,* 2nd ed., H.I. Aaronson, Ed., TMS, Warrendale, PA,1999, p 231
31. A. Engstrom, J.E. Morral, and J. Agren, *Acta Mater.,* Vol 45 (No. 3), 1997, p 1189
32. J.E. Morral and H. Chen, *Scr. Mater.,* Vol 43, 2000, p 699

Diffusivity and Mobility Data

Carelyn E. Campbell, National Institute of Standards and Technology

DIFFUSION is the process by which molecules, atoms, ions, point defects, or other particle types migrate from a region of higher concentration to one of lower concentration. The diffusivity of an atom or diffusion coefficient is the rate at which a particle migrates through a particular material and is dependent on the temperature, composition gradient, and pressure. Diffusivity determines how quickly equilibrium is reached from a nonequilibrium state. In solid-state materials, diffusion can occur by a variety of different mechanisms. Lattice diffusion (bulk or volume diffusion) occurs as a result of individual jumps of atoms or point defects, such as vacancies, divacancies, or interstitials, within the crystal. Diffusion may also occur along the surface or along line defects, such as grain boundaries or dislocations. These line, planar, and surface diffusion mechanisms are generally much faster than the lattice-diffusion-based-mechanisms and, as a result, are termed high-diffusivity paths (or short-circuit diffusivity). Depending on the temperature and/or microstructure of the material, these high-diffusivity paths or bulk diffusion may be the dominate diffusion mechanism. As a wide variety of microstructural processes are controlled by lattice diffusion mechanisms and much of the published diffusion data is for bulk diffusion processes, this article focuses primarily on the diffusivity data and modeling of lattice diffusion in solid-state materials. A list of symbols is in Table 1.

Diffusion Mechanisms

In a crystal, lattice vibrations cause atoms to oscillate around equilibrium positions with frequencies, ν_0 ($\sim 10^{12}$ to 10^{13} Hz); however, occasionally the oscillations are large enough to allow an atom to jump to a different lattice site, resulting in diffusion of the atom. In the absence of a driving force, the energy barrier that an atom must overcome to jump to an empty site is the free energy of migration, G_m. Figure 1 shows the G_m as a function of atomic

Table 1 List of symbols

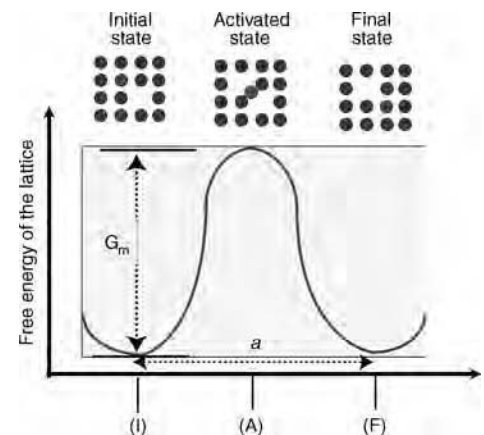|	
a	= nearest-neighbor atomic distance (m)
$^k A_i^{pj}$	= contribution to the diffusion-activation energy of component i, in a lattice occupied by p and j atoms in a given phase. k is the order of the interaction parameter (i.e., 0, 1, 2, ...) (Eq 27).
b	= thickness of thin layer (Eq 10) (m)
c_i	= volume concentration of component i (mol/m³)
c_s	= concentration at the surface of semi-infinite solid (Eq 11) (mol/m³)
c_0	= initial concentration of component at time = 0 s (Eq 10) (mol/m³)
c_1, c_2	= initial concentrations at the end of infinite couple (Eq 12) (mol/m³)
c_V^{eq}	= thermal equilibrium vacancy concentration (mol/m³)
c^-, c^+	= initial compositions of a given diffusion couple (Eq 18) (mol/m³)
c^*, z^*	= composition at a given distance (z) (Eq 18)
D	= diffusion rate in the absence of any driving force (m²/s)
D_0	= pre-exponential factor for diffusion (m²/s)
D_i^*	= tracer diffusion coefficient for component i (m²/s)
D_i^s	= self-diffusion coefficient for component i (m²/s)
$^L D_{kj}$	= diffusion coefficient for the diffusing component, j, with respect to the composition gradient of component k, in the lattice-fixed frame of reference (m²/s)
$^V D_{kj}$	= diffusion coefficient for the diffusing component, j, with respect to the composition gradient of component k, in the volume-fixed frame of reference (m²/s)
$\tilde{D}_{kj}^n$	= interdiffusion coefficient, where n is the dependent component (m²/s)
f	= correlation factor for self-diffusion, dependent on crystal structure
g	= geometric factor dependent on the lattice geometry and type of interstitial site
G_m	= free energy of vacancy migration (J/mol)
H_F	= enthalpy of formation of a vacancy (J/mol)
H_m	= enthalpy of vacancy migration (J/mol)
J_i	= flux of particles (number particles per second and unit area) (m⁻² s⁻¹)
$^L J_k$	= flux of particles in the lattice-fixed frame of reference (number particles per second and unit area) (m⁻² s⁻¹)
$^V J_k$	= flux of particles in the volume-fixed frame of reference (number particles per second and unit area) (m⁻² s⁻¹)
M_i	= atomic mobility of component i (m/N/s)
$^L M_{ki}$	= mobility of the diffusing component, k, with respect to the composition gradient of component i, in the lattice-fixed frame of reference
N_i^α	= number of i atoms on the α sublattice
N_{total}^α	= total number of atoms on the α sublattice
p	= probability of the next neighboring site being vacant
ΔQ	= diffusion-activation energy (J/mol)
ΔQ_i^*	= diffusion-activation energy of component i in a given phase (J/mol)
Q_i^j	= diffusion-activation energy of component i in a lattice occupied by pure j atoms in a given phase (J/mol)
$\Delta Q_i'$	= diffusion-activation energy assuming the pre-exponential term is included (J/mol)
ΔQ_k^{ord}	= ordered contribution to the diffusion-activation energy (J/mol)
ΔQ_k^{dis}	= disordered contribution to the diffusion-activation energy (J/mol)
R	= gas constant (J/mol K)
S_F	= entropy of formation of a vacancy (J/K)
S_m	= entropy of vacancy migration (J/K)
t	= time (s)
T	= temperature (K)
V_m	= molar volume of a phase (m³/mol)
x_i	= mole fraction of component i
y_i^α, y_i^β	= site fractions of component i on the α and β sublattices, respectively
Y_i	= normalized concentration variable (Eq 16)
z	= distance (m)
z_K	= position of the Kirkendall plane (m)
z_M	= position of the Matano plane (m)
δ_{ik}	= Kronecker delta symbol, equals 1 when $i = k$ and 0 when $i \neq k$
Φ_i	= pre-exponential factor defining the mobility of atom i
Γ	= jump frequency (s⁻¹)
λ	= diffusion length (m)
μ_i	= chemical potential of component i (J/mol)
ν_0	= equilibrium lattice frequency (Hz) (s⁻¹)
π	= constant, 3.14...
υ_K	= Kirkendall velocity (m/s)
ω	= jump rate for atom to a neighboring empty lattice site

Fig. 1 Assuming no driving force on an atom, a schematic of the energy barrier that must be overcome for a diffusion to occur. I, initial state; A, activated state, F, final state

position. The jump rate, ω, for an atom to jump to a neighboring empty site is given by:

$$\omega = \nu_0 \exp\left(-\Delta G_m / RT\right) \text{ where } \Delta G_m = H_m - TS_m$$

(Eq 1)

where ν_0 is the atomic vibration frequency, R is the gas constant, T is the temperature in Kelvin, and H_m and S_m are the enthalpy and entropy of migration. The jump frequency, Γ, of an atom is then defined as the jump rate times the probability, p, of the next neighboring site being vacant:

$$\Gamma = \omega p$$

(Eq 2)

Two of the most common mechanisms by which atoms diffuse are interstitial diffusion and vacancy diffusion. Interstitial diffusion occurs as small interstitial atoms, which are solute atoms that are considerably smaller than the solvent atoms, jump from one interstitial site to the next-nearest unoccupied interstitial site. Thus, for a dilute interstitial alloy, the probability defined in Eq 2 is approximately equal to 1, and the diffusion rate in the absence of any driving force (concentration gradient) is:

$$D = \left[ga^2\nu_0 \exp\left(\frac{S_m}{R}\right) \exp\left(\frac{-H_m}{RT}\right)\right]$$

(Eq 3)

where a is the nearest-neighbor atomic distance, and g is a geometric factor that depends on the lattice geometry and the type of interstitial site (i.e., octahedral or tetrahedral). This expression is commonly simplified to an Arrhenius-type equation:

$$D = D_0 \exp\left(\frac{-\Delta Q}{RT}\right) \text{ where } D_0$$
$$= ga^2\nu_0 \exp\left(\frac{S_m}{R}\right) \text{ and } H_m = \Delta Q$$

(Eq 4)

Vacancy-driven diffusion occurs when a nearest-neighbor atom (substitutional solute atom) jumps onto an unoccupied lattice site. The probability of the nearest-neighbor site being vacant, p from Eq 2, is defined by the thermal equilibrium vacancy concentration, c_V^{eq}, which is given by:

$$c_V^{eq} = \exp\left(\frac{S_F}{R}\right) \exp\left(\frac{-H_F}{RT}\right)$$

(Eq 5)

where S_F and H_F are the formation entropy and enthalpy, respectively, of a vacancy. Thus, similar to Eq 3 for the interstitial diffusion rate, the substitutional diffusion rate in the absence of a driving force is given by:

$$D = \left[fa^2\nu_0 \exp\left(\frac{S_F + S_m}{R}\right) \exp\left(\frac{-(H_F + H_m)}{RT}\right)\right]$$

(Eq 6)

where f is the correlation factor that determined by the crystal structure. Like Eq 3, 6 can be written in terms of an Arrhenius relation, where:

$$D_0 = fa^2\nu_0 \exp\left(\frac{S_F + S_m}{R}\right)$$

and $\Delta Q = H_F + H_m$. It should be noted that while these Arrhenius-type relations (Eq 4 and 6) are common, they are not universal. Grain-boundary, impurities, or other microstructural features; temperature-dependent activation parameters; and other active diffusion mechanisms may all result in deviations from an Arrhenius relation.

Diffusion Equation

In a steady-state one-dimensional system, the flux of particles is proportional to the concentration gradient:

$$J_i = -D_i \frac{\partial c_i}{\partial z}$$

(Eq 7)

where J_i describes the amount of material that passes through a unit area of a plane per unit time (t) within a volume-fixed frame of reference (otherwise known as flux). The variable D_i is the diffusivity of component i for a given diffusion mechanism. The variable c_i is the concentration of particles i (Note: $c_i = x_i V_m$, where x_i is the mole fraction of component i, and V_m is the molar volume of the phase), and z is the diffusion distance. Equation 7, better known as Fick's law (Ref 1, 2), assumes there are no external forces or driving forces acting on the particles and is formally identical to Fourier's law of heat conduction and Ohm's law of current flow.

For a nonsteady-state one-dimensional system where the flux at each point varies with time, Fick's law must be combined with a mass balance or continuity equation (Eq 8) to determine the time-dependent concentration, where t is the time in seconds:

$$\frac{\partial J}{\partial z} = -\frac{\partial c}{\partial t}$$

(Eq 8)

Equations 7 and 8 are combined to form the general diffusion equation, which is a second-order linear partial differential equation and cannot be solved analytically:

$$\frac{\partial c_i}{\partial t} = \frac{\partial}{\partial z}\left(D_i \frac{\partial c_i}{\partial z}\right)$$

(Eq 9)

However, if D_i is assumed to be concentration independent, the diffusion equation can be solved for a variety of initial and boundary conditions, as demonstrated by Ref 3 and 4. Two simple examples of these solutions that are commonly used experimentally are the thin-film and error-function solutions.

The thin-film solution assumes that a thin layer, with a thickness, b, of the diffusing species A is concentrated at $z = 0$ of a semi-infinite sample, as seen in Fig. 2. Then, concentration profiles after time, t, are given by:

$$c(z,t) = \frac{bc_0}{\sqrt{\pi Dt}}\exp\left(-\frac{z^2}{4Dt}\right)$$

(Eq 10)

where c_0 is the initial concentration of the A layer. The diffusion length, λ, is represented by the $\sqrt{Dt}$ quantity and is a characteristic length used in solving diffusion equations. The thin-film solution is valid for applications where λ is much greater than the initial layer thickness. The geometry represented by the thin-film solution (Eq 10) is commonly used to measure tracer diffusion coefficients in substitutional alloys (defined in the section "Tracer Diffusivity" in this article).

Infinite and Semi-Infinite Solutions. Error-function solutions can be applied to semi-infinite and infinite samples. The concentration profiles in a semi-infinite solid with a constant concentration of a component at the surface, c_s, are defined by the following error functions and initial and boundary conditions:

Initial conditions: $t = 0$; $z > 0$; $c(z,0) = c_0$, where c_0 is the initial concentration in the solid

Boundary conditions: at all $t > 0$, $z = 0$; $c(0,t) = c_s$

The concentration profile for a given time is then given by:

$$\frac{c - c_s}{c_0 - c_s} = \text{erf}\left(\frac{z}{2\sqrt{Dt}}\right)$$

(Eq 11)

These types of solutions are applicable to modeling various carburization and coating problems.

For an infinite sample with a concentration profile defined by a step function, the composition profile at given time, t, is:

$$\frac{c - c_1}{c_2 - c_1} = \frac{1}{2}\text{erfc}\left[\frac{z}{2\sqrt{Dt}}\right]$$

(Eq 12)

where the initial boundaries at $t = 0$ are given by:

$$t = 0 \begin{cases} z \langle 0 & c = c_2 \\ z \rangle 0 & c = c_1 \end{cases}$$

and c_1 and c_2 are the compositions at either end of the infinite couple, as seen in Fig. 3. This type of solution is often used when the diffusion distances in the two materials are much smaller than the width of the samples. While analytical solutions to the diffusion equation assuming a constant diffusivity are useful in solving some practical problems, to solve most problems of interest, the composition and temperature dependence of the diffusivity must be considered, and the equation must be solved numerically.

Diffusion Data

Diffusivity in a material can be evaluated in variety of different ways, including measurement of diffusion coefficients, composition profiles, and layer growth widths. Tracer, intrinsic, and chemical diffusivities can all be extracted from various types of direct and indirect diffusion experiments. Table 2 reviews some of the common direct and indirect methods to measure these diffusion coefficients. Direct experiments are based on Fick's law (Eq 7) and the phenomenological definitions of the diffusion coefficients. Indirect experimental methods are not based on Fick's law and require a microscopic model of the atomic jump processes to deduce a diffusion coefficient.

Tracer Diffusivity. Tracer diffusion is the migration of a tagged atom through a material of which it is a component. As such, the tracer diffusivity is generally measured by introducing a radioactive isotope in dilute concentration into an otherwise homogeneous material. Thus, the only driving force in the system is that of the concentration gradient of the tracer. In a homogenous material, the mean square displacement of the tracer in the z-direction is defined by the Einstein formula for Brownian motion (random walk) as:

$$\langle z_i^2 \rangle = 2D_i^* t \text{ where } D_i^* = RTM_i \quad \text{(Eq 13)}$$

where D_i^* is the tracer diffusion coefficient, t is the diffusion time, and M_i is the atomic mobility of the i atoms or the movement of atoms in response to a given force, in this case, the result of continuous random movement. (Note that for a multicomponent alloy, the mobility is a function of each component in the alloy. Further description of the calculation of the composition-dependent mobility matrix is given in the section "Disordered Phase" in this article.) The tracer diffusion coefficient is equal to the self-diffusion coefficient, D_i^S, if diffusion takes place by uncorrelated atomic jumps; otherwise, the tracer diffusion coefficient is related to the self-diffusion coefficient by the correlation factor, f, $D_i^* = fD_i^S$. The correlation factor is dependent on crystal structure and introduces off-diagonal terms into the M_i matrix. For the body-centered cubic (bcc) and face-centered cubic (fcc) crystal structures assuming a vacancy-diffusion mechanism, the contribution of these off-diagonal terms is small. Thus, for simplicity, these terms are not included in the present discussion; however, calculation methods are discussed by Ref 5 and 6. When a tracer impurity is measured in a homogeneous material (i.e., solute C in a pure A alloy or homogeneous AB alloy), the measured tracer diffusivity is often referred to as the impurity diffusivity.

Chemical Diffusivity (Interdiffusion). In contrast to tracer diffusivity measurements that only consider negligible amounts of a tracer element in an otherwise homogeneous material, interdiffusion (chemical diffusion) and intrinsic diffusion coefficient measurements are performed in nonhomogeneous materials where the diffusion flux is proportional to the gradient of the chemical potential (e.g., in the presence of a driving force). The interdiffusion coefficient, $^V D_{kj}$, is defined in the volume-fixed frame of reference, where the sum of the fluxes equals zero, and is given by:
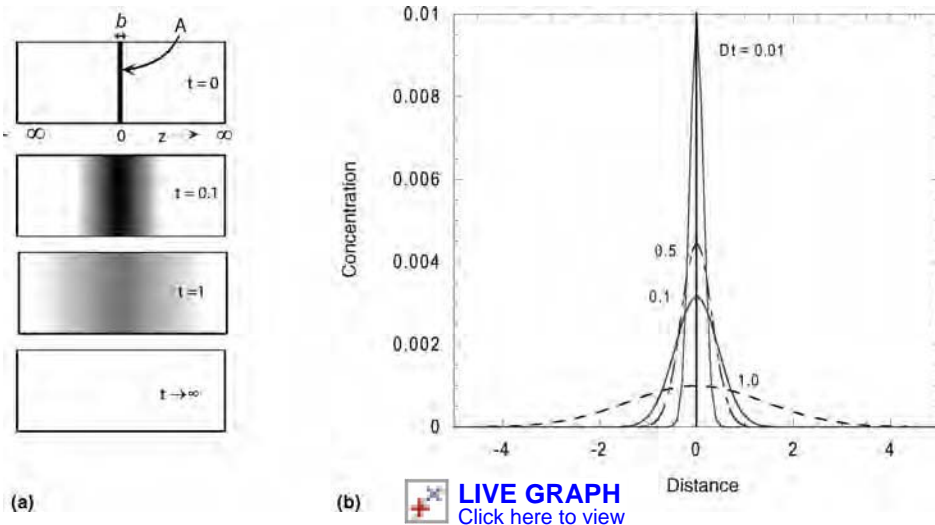

Fig. 2 Schematic of (a) thin-film geometry and (b) solution. (a) A thin layer of the diffusing species (A atoms per unit area) is concentrated at $z = 0$. As time increases, the A atoms diffuse such that concentration becomes negligible. (b) The concentration of A atoms as a function of distance for different times, where the time increments are given as $D*t$, where D is a constant.

Table 2 Direct and indirect methods for measuring diffusion coefficients

| Method	Δx	D, m²/s
Direct		
Lathe sectioning, grinding	0.1–250 µm	10^{-19} to 10^{-10}
Microtome	1–10 µm	10^{-17} to 10^{-12}
Chemical	10 µm	10^{-15} to 10^{-12}
Electrochemical	50 nm	10^{-20} to 10^{-17}
Sputtering	5–100 nm	10^{-22} to 10^{-17}
Modulated structures	0.5–5 nm	$\geq 10^{-26}$
Ion microprobe (secondary ion mass spectroscopy)	1–100 nm	10^{-23} to 10^{-17}
Electron microprobe	≥ 2 µm	10^{-16} to 10^{-12}
Rutherford backscattering	50 nm	10^{-20} to 10^{-17}
Nuclear reaction analysis	20–100 nm	5×10^{-21} to 5×10^{-16}
Indirect		
Nuclear magnetic resonance	...	10^{-20} to 10^{-9}
Neutron inelastic scattering	...	10^{-11} to 10^{-9}
Mossbauer effect	...	10^{-15} to 10^{-11}
Conductivity (ionic crystals)	...	10^{-17} to 10^{-10}
Resistivity (semiconductors)	...	10^{-20} to 10^{-12}
Elastic after-effect	...	10^{-25} to 10^{-21}
Internal friction	...	10^{-20} to 10^{-15}
Magnetic anisotropy	...	10^{-25} to 10^{-21}

Source: Ref 5
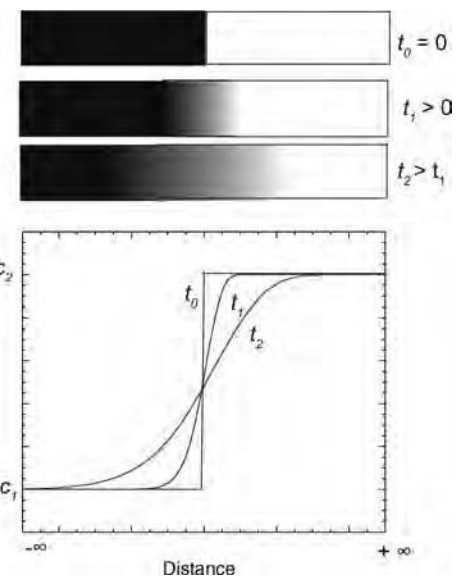
Fig. 3 Example of error-function solutions for an infinite couple with initial end-member compositions of c_1 and c_2. The composition profiles show the diffusion of the species with increasing time.

$$\sum_{k=1}^{n} {}^{\mathrm{V}}J_k = 0 \qquad \text{(Eq 14)}$$

$$^{\mathrm{V}}D_{kj} = \sum_{i=1}^{n} (\delta_{ik} - x_k) x_i M_i \frac{\partial \mu_i}{\partial x_j} V_{\mathrm{m}} \qquad \text{(Eq 15)}$$

The component j is the diffusing component, and k is the gradient component. The δ_{ik} is the Kronecker delta symbol and equals 1 when $i = k$ and 0 when $i \neq k$. The partial derivative of the chemical potential, μ_i, with respect to the mole fraction, x_i, corresponds to the thermodynamic factor. Note that the partial derivative of the chemical potential can be easily calculated using a functional representation of the Gibbs energy, for example, an appropriate multicomponent thermodynamic database. However, the thermodynamic factor must be evaluated in the form $\mu_k(x_1, x_2, \ldots x_n)$ because there are n-1 independent concentrations. As there are only $(n-1)$ independent components, diffusion couple experiments are only able to directly evaluate the interdiffusion coefficient, $\tilde{D}_{kj}^n$:

$$\tilde{D}_{kj}^n = D_{kj}^{\mathrm{V}} - D_{kn}^{\mathrm{V}} \qquad \text{(Eq 16)}$$

where n is the dependent variable. Using these interdiffusion coefficients, the flux equations (Eq 7) in the volume-fixed frame of reference, where the sum of the fluxes equals zero, can be written as:

$$J_k = -\sum_{j=1}^{n-1} \tilde{D}_{kj}^n \frac{\partial c_j}{\partial z} \qquad \text{(Eq 17)}$$

There are several methods for determining interdiffusion coefficients from measured composition profiles from diffusion couple experiments. Figure 4 shows that for diffusion couples that can be approximated as an infinite medium, the Boltzmann-Matano method (Ref 7, 8) can be used to determine the interdiffusion coefficients from experimental composition profiles at a given time, t:

$$\tilde{D}(c^*) = \left(2t \frac{dc}{dz}\bigg|_{z^*} \right)^{-1} \int_{c}^{c^*} (z_M - z)dc \text{ and }$$

$$\int_{c^-}^{c^+} (z_M - z)dc = 0 \qquad \text{(Eq 18)}$$

The variable z_M defines the Matano plane through which equal amounts of material have moved in the positive and negative directions. The concentrations c^- and c^+ represent the initial compositions of the diffusion couple. However, this commonly used method does not consider the change in molar volume across the diffusion couple. When significant molar volume changes are present (e.g., in the intermetallic NiAl-B2, Ref 9), the interdiffusion coefficients should be calculated using methods that include the composition dependence of the molar volume, such as that proposed by Ref 10 to 12:
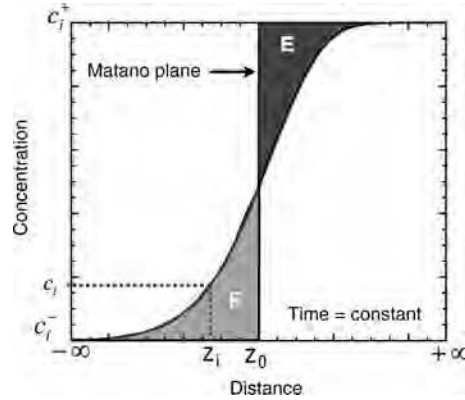

Fig. 4 Example of Boltzmann-Matano calculation for a single-phase interdiffusion experiment with end-member composition of c_i^- and c_i^+. The Matano plane is located at z_0 and is chosen such that the two shaded areas, E and F, are equal. The diffusion coefficient at given c_i concentration is then given by Eq 18.

$$\tilde{D} = \frac{V_{\mathrm{M}}}{2t} \left(\frac{dz}{dY_i} \right)_{Y^*} \left[(1 - Y_i^*) \int_{-\infty}^{z^*} \frac{Y_i}{V_{\mathrm{M}}} dz \right.$$

$$\left. + Y_i^* \int_{z^*}^{-\infty} \frac{(1 - Y_i)}{V_{\mathrm{M}}} dz \right] \text{ where } Y_i = \frac{(c_i - c_i^-)}{(c_i^+ - c_i^-)} \qquad \text{(Eq 19)}$$

where Y_i is a normalized concentration variable. This method also eliminates the need to determine a Matano interface, which is often a source of error. An example of the method is shown in Fig. 5 for a cobalt-nickel diffusion couple.

As the number of elements in the diffusion couples increases, determining the interdiffusion coefficients becomes more difficult; for each interdiffusion coefficient, $(n-1)$ composition profiles with different terminal compositions must intersect at one common intersection point. In an effort to overcome this complexity, Ref 13 derived a new analysis method that enables the determination of an average interdiffusion coefficient over a selected composition range from a single multicomponent diffusion couple by integrating the interdiffusion flux of a component over the diffusion distance for a selected range of compositions. This method has been implemented in the computational software program MultiDiFlux (Ref 14). (Commercial products are referenced in this paper as examples. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology.)

Intrinsic Diffusivity. The intrinsic diffusion coefficient defines the diffusion of a component relative to the lattice planes and is the product of the diffusion mobility and the thermodynamic factor in the lattice-fixed frame of reference, where the sum of the diffusion fluxes equals the vacancy flux:

$$\sum_{k=1}^{n} {}^{\mathrm{L}}J_k = -{}^{\mathrm{L}}J_{\mathrm{vacancy}} \qquad \text{(Eq 20)}$$
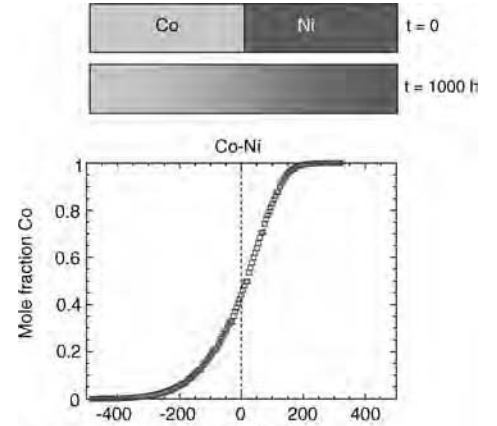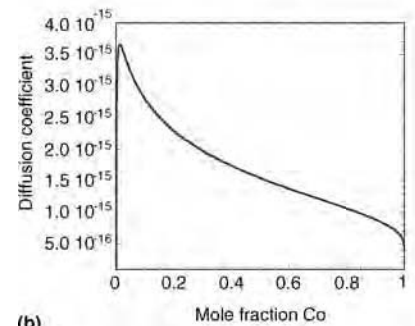

Fig. 5 Example of the Sauer-Freise method to calculate the interdiffusion coefficient for face-centered cubic cobalt-nickel at 1150 °C. (a) Measured composition profile after 1000 h at 1150 °C. (b) Calculated interdiffusion coefficient at 1150 °C. Source: Ref 10

$$^{\mathrm{L}}D_{kj} = \sum_{i=1}^{n} \delta_{xi} x_i M_k \frac{\partial \mu_k}{\partial x_j} \qquad \text{(Eq 21)}$$

A net flux of atoms across any lattice plane occurs during interdiffusion as the diffusion rates of the components in a material are different. Thus, there is a shift of lattice planes relative to a fixed lattice axis, which is known as the Kirkendall effect. This shift of lattice planes is observed by placing inert markers at the initial interface of a diffusion couple (Fig. 6) (Ref 15, 16). The velocity of the inert markers equals the Kirkendall velocity. Thus, the difference between Eq 15, $^{\mathrm{V}}D_{kj}$, and Eq 21, $^{\mathrm{L}}D_{kj}$, is the reference state. The $^{\mathrm{V}}D_{kj}$ represents the measurement of the diffusivity relative to a fixed position, while the $^{\mathrm{L}}D_{kj}$ represents the measurement of diffusivity relative to a fixed lattice plane. This difference in the frame of reference is similar to measuring the speed of a train when the observer is standing at a fixed position on the train platform (in the volume-fixed frame of reference) versus when the observer is sitting in one of the train cars (the lattice-fixed frame of reference).

For a substitutional binary alloy, the intrinsic and tracer diffusivities are related to the interdiffusion coefficient and the Kirkendall

velocity. For example, for a binary AB alloy, Darken (Ref 17) deduced the following approximate relations, which disregard the coupling between the fluxes of the two atoms and the vacancy flux:

$$^{V}\tilde{D}_{AB} = x_B{}^L D_A + x_A{}^L D_B = \left(x_A D_B^* + x_B D_A^*\right)\frac{\partial \mu_A}{\partial x_B}$$
(Eq 22)

The Kirkendall velocity, v_K, is given by:

$$v_K = \left(^L D_A - ^L D_B^L\right)\frac{\partial x_A}{\partial z}$$
(Eq 23)

For a more complete description of the relationship between the diffusivities in a "random" alloy during a vacancy-driven diffusion process, the Darken-Manning relations (Ref 18, 19) should be used.

As an example of the relationship between the diffusivities, consider the fcc diffusivities for the iron-nickel system at 1200 °C (Fig. 7). Figure 7 (a) shows the calculated thermodynamic factors, and Fig. 7(b) shows the mobilities for nickel and iron. The Darken relationship (Eq 22) is demonstrated in Fig. 7(c):

$$\tilde{D}_{FeNi} = x_{Fe}{}^L D_{Ni} + x_{Ni}{}^L D_{Fe}$$
(Eq 24)

where the intrinsic lattice diffusivities, Eq 21, $(^L D_{Ni}, ^L D_{Fe})$ are calculated by multiplying the thermodynamic factor (Fig. 7a) by the mobility (Fig. 7b).

First-Principles Data. In addition to the variety of experimental methods available for measuring diffusivity, first-principles calculations may be available to help estimate difficult-to-measure or metastable diffusion coefficients. Density-functional methods can be used to calculate the self-activation diffusion energies (Ref 20). Embedded-atom potentials can be used to evaluate diffusion mechanisms and determine activation energies (Ref 21–23). Diffusion coefficients can be extracted from kinetic Monte Carlo simulations using Kub-Green expressions (Ref 24, 25).

Modeling Multicomponent Diffusivity Data

While much of the diffusivity data for the pure elements and many binary alloys have been measured and are available in the literature, diffusivity data for multicomponent systems are scarce and difficult to measure. Experimentally determining all the needed diffusion coefficient matrices for a multicomponent diffusion simulation is simply not practical or efficient. However, these multicomponent data are critical for correctly predicting diffusion behavior in many industrial (commercial) applications and may be strongly dependent on composition. Thus, multicomponent diffusion mobility databases are developed to predict the needed bulk diffusion coefficients.
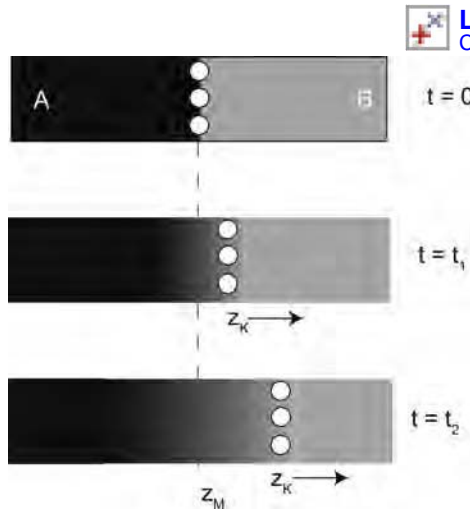

Fig. 6 Schematic of Kirkendall effect in an A-B diffusion couple where the B atoms diffuse faster than the A atoms $(D_B > D_A)$, and the interface moves to the right. The Matano plane is defined by z_M, and the Kirkendall plane is located at z_K.

Using Onsager's relations and a Calphad-based method (Ref 26–28), Ref 29 developed a formalism to describe diffusion mobilities in multicomponent systems and to develop multicomponent diffusion mobility databases. These diffusion mobilities can then be combined with the needed thermodynamic factors to calculate the multicomponent diffusion coefficients. Appendix 1 demonstrates how ternary tracer, intrinsic, interdiffusion coefficients are calculated for a given set of diffusion mobilities and chemical potentials.

Disordered Phase

Substitutional Diffusion. Assuming a vacancy diffusion mechanism in a crystalline phase, the mobility matrix in the lattice-fixed frame, $^L M_{ki}$, which is both composition and temperature dependent, can be written in terms of an Arrhenius-type relation similar to Eq 4 and 6:

$$^L M_{ki} = \delta_{ki} x_i M_i$$
(Eq 25)

$$M_i = \Theta_i \frac{1}{RT}\exp\left(\frac{\Delta Q_i^*}{RT}\right)$$
(Eq 26)

Following the work of Ref 29, the off-diagonal terms of the diffusion mobility matrix are assumed to be zero; that is, correlation effects are assumed to be negligible. M_i is the mobility of component i in a given phase (this is the same M_i as in Eq 13, 15, and 21), Θ_i represents the effects of the atomic jump distance (squared) and the jump frequency, and ΔQ_i^* (with units of J/mol) is the diffusion-activation energy of component i in a given phase. The partial molar volumes are assumed to be constant, and the composition and temperature dependence of each ΔQ_i^* are expressed in terms of a Redlich-Kister (Ref 30) polynomial (Ref 29, 31–35):
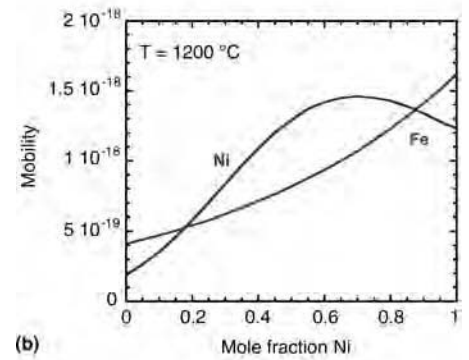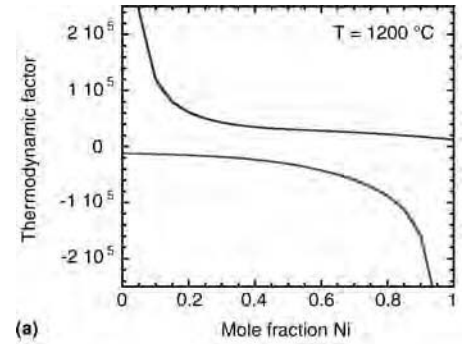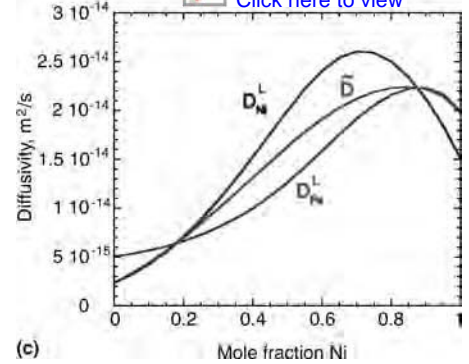

Fig. 7 Example of the calculation of the intrinsic diffusion and interdiffusion coefficients for the iron-nickel system at 1200 °C. Thermodynamic factors (a) and mobilities (b) for iron and nickel as functions of composition are multiplied to calculate the intrinsic diffusion coefficients shown in (c). The Darken relation (Eq 22) is used in (c) to calculate the interdiffusion coefficient.

$$\Delta Q_i^* = \sum_j x_j Q_i^j + \sum_p \sum_{j>p} x_p x_j \sum_k {}^k A_i^{pj}(x_p - x_j)^k$$
(Eq 27)

where Q_i^j and $^k A_i^{pj}$ are linear functions of temperature. The expansion of the composition dependence in terms of a Redlich-Kister polynomial is similar to the development of thermodynamic databases with the Calphad method (Ref 26–28). Note that for a given diffusing component, i, if all Q_i^j are equal and $^k A_i^{pj}$ equals zero, then ΔQ_i^* and the corresponding M_i are not concentration dependent. The composition dependence of Θ_i can also be

represented by Eq 27. If there is no ferromagnetic contribution, it is frequently assumed that Θ_i depends exponentially on composition (Ref 36), and it is included in the activation energy term. With this assumption, Eq 26 can be written as:

$$M_i = \frac{1}{RT}\exp\left(\frac{\Delta Q_i'}{RT}\right)$$

(Eq 28)

where $\Delta Q_i' = \Delta Q_i^* - RT \ln \Theta_i$

Interstitial Diffusion. Interstitial elements can be added to the database by using a sublattice description and by assuming the partial molar volume of the interstitial element is zero (Ref 37). An example of this model is the addition of carbon to the Fe-Ni-Cr fcc phase in the diffusion assessment by Ref 38.

Magnetic Transition. For substitutional elements in bcc alloys, such as transition metals, the effect of the transition between the para- and ferromagnetic states contributes to the diffusion. The effect of magnetic ordering on diffusion can be included (Ref 39, 40), using the model of Braun and Feller-Kniepmeier (Ref 41), which relates the change in diffusivity to the magnetic enthalpy. For interstitial elements, such as carbon or nitrogen, the effect of the magnetic transition is less well established. The magnetic transition has a strong effect on carbon diffusivity (Ref 42), which can be modeled by applying the same activation energy to both the paramagnetic and ferromagnetic states (Ref 43). However, no significant change in the nitrogen diffusivity is observed as the magnetic transition occurs.

Ordered Phases

For an ordered phase, the composition dependence of the diffusion mobilities must include the effect of chemical ordering. Based on the model by Girifalco (Ref 44), which assumes the activation energy from chemical ordering is dependent on a long-range order parameter, the effect of chemical ordering is included by dividing the activation energy into two terms (Ref 35). The first term represents the contribution from the disordered state, ΔQ_k^{dis}, and the other term represents the contribution from the ordered state, ΔQ_k^{ord}, which is based on a long-range order-type parameter, the site fraction of a given component i:

$$\Delta Q_k = \Delta Q_k^{dis} + \Delta Q_k^{ord}$$

(Eq 29)

where ΔQ_k^{ord} is defined as:

$$\Delta Q_k^{ord} = \sum_i \sum_{i \neq j} \Delta Q_{kij}^{ord}\left[y_i^\alpha y_j^\beta - x_i x_j\right]$$

(Eq 30)

and ΔQ_{kij}^{ord} are the contributions to the activation energy for component k as a result of the chemical ordering of the i and j atoms on the two sublattices; x_i is the mole fraction of component i; and y_i^α and y_i^β are the site fractions of component i on the given sublattices:

$$y_i^\alpha = \frac{N_i^\alpha}{N_{total}^\alpha}$$

(Eq 31)

where N_i^α equals the number of i atoms on the α sublattice, and N_{total}^α equals the total number of atoms on the α sublattice. This approach was developed for an AB (B2) alloy where diffusion occurs via jumps between two metal sublattices; however, the approach is also valid for A_3B (Fe_3Al) alloys ($D0_3$ ordering) (Ref 45), where diffusion occurs via a network of nearest-neighbor jumps and where the fcc or hexagonal close-packed crystal structure is the base for the ordered phase. This model has been successfully used to describe the Fe-Ni-Al diffusion in the B2 phase (Ref 34) and to describe the Ni-Al-Cr diffusion in the B2 and γ' phases (Ref 46).

Stoichiometric Phases

For binary stoichiometric phases, the diffusivity is assumed to be proportional to the difference in the chemical potentials at each end of the stoichiometric phase multiplied by the mobility for the component in the phase. Tracer diffusivity data for the component in the stoichiometric phase are used to assess the diffusion mobility functions. This type of model has been applied to the diffusivity of carbon in cementite (Ref 47).

Determination of Diffusion Mobility Coefficients

Similar to the Gibbs energy function coefficients used in multicomponent thermodynamic databases, the diffusion mobility parameters in Eq 26 and 30 are determined from experimental data for each system and can be evaluated using trial-and-error methods or mathematical methods that minimize the error between the calculated and experimental diffusion coefficients, as indicated in Fig. 8. The trial-and-error method is only feasible if a few different data types are available. This method becomes increasingly cumbersome as the number of components and/or number of data types increases. When this occurs, mathematical methods, such as the least-squares method of Gauss (Ref 48), the Marquardt method (Ref 49), or the Bayesian estimation method (Ref 50), are more efficient. The PARROT optimizer (Ref 51) within the DICTRA code (Ref 52, 53) allows direct optimization of diffusion mobility functions.

General Principles. The same principles guiding the assessment of thermodynamic data (Ref 28) also apply to diffusion data, with a few additional constraints. First, a thermodynamic database (or description) must be selected to calculate the needed thermodynamic factors for intrinsic and interdiffusion coefficients. In choosing a thermodynamic database, the phase models used for the thermodynamics must be the same as those used in the diffusion mobility database. For example, if a thermodynamic description uses a two-sublattice model of an fcc phase (one sublattice for the substitutional elements and a second sublattice for the interstitials), then the same two-sublattice model must be used in the diffusion mobility database. After selecting a thermodynamic database/description to use for developing the diffusion mobility database, a critical evaluation of all the available data must be performed. As the tracer diffusivity data are not dependent on the thermodynamics, these data are preferred and often weighted more heavily than other diffusivity data, which are dependent on the thermodynamic description used. The assessment process continues by optimizing the mobility parameters for each component in each phase separately and then optimizing the mobility parameters for all of the components in a given phase with all of the relevant, and appropriately weighted, diffusion data. Zero-order binary and ternary interaction parameters may be added as needed to fit the available diffusion data. Generally, ternary and higher-order binary interactions are rarely needed to fit the experimental data, or there are insufficient experimental data to justify such terms. After all of the needed mobility parameters are optimized, the assessment is verified using diffusion data not considered during optimization, such as a comparison of calculated and measured composition profiles from diffusion-couple experiments. The assessed parameters may also be evaluated by comparing activation energies with diffusion correlations published in the literature (Ref 54) or with first-principles calculations.

Binary Assessment Example. The assessment of the nickel-tungsten diffusion mobilities in the fcc phase is described here as an example. The thermodynamic description developed by Gustafson et al. (Ref 55) is used. The fcc phase is modeled using a two-sublattice description (nickel, tungsten: vacancies), where nickel and tungsten occupy the substitutional sublattice, and vacancies occupy the interstitial sublattice. For the nickel-tungsten system in the fcc phase, both tracer (Ref 56) (Fig. 9) and interdiffusion (Ref 57) (Fig. 10) data are available. The nickel and tungsten diffusion mobilities in the fcc phase are described as:

$$M_i = \frac{1}{RT}\exp\left(\frac{\Delta Q_i'}{RT}\right) \text{ where } i = \text{Ni, W}$$

(Eq 32)

$$\Delta Q_{Ni}' = x_{Ni}Q_{Ni}^{Ni} + x_W Q_{Ni}^W + x_{Ni}x_W {}^0A_{Ni}^{Ni,W}$$
$$\Delta Q_W' = x_{Ni}Q_W^{Ni} + x_W Q_W^W + x_{Ni}x_W {}^0A_W^{Ni,W}$$

(Eq 33)

(Note that because this phase has no ferromagnetic contribution, the mobilities are expressed using Eq 28.) If a ferromagnetic contribution were present, separate composition-dependent functions for the activation energy and preexponential terms would be needed (Eq 26).
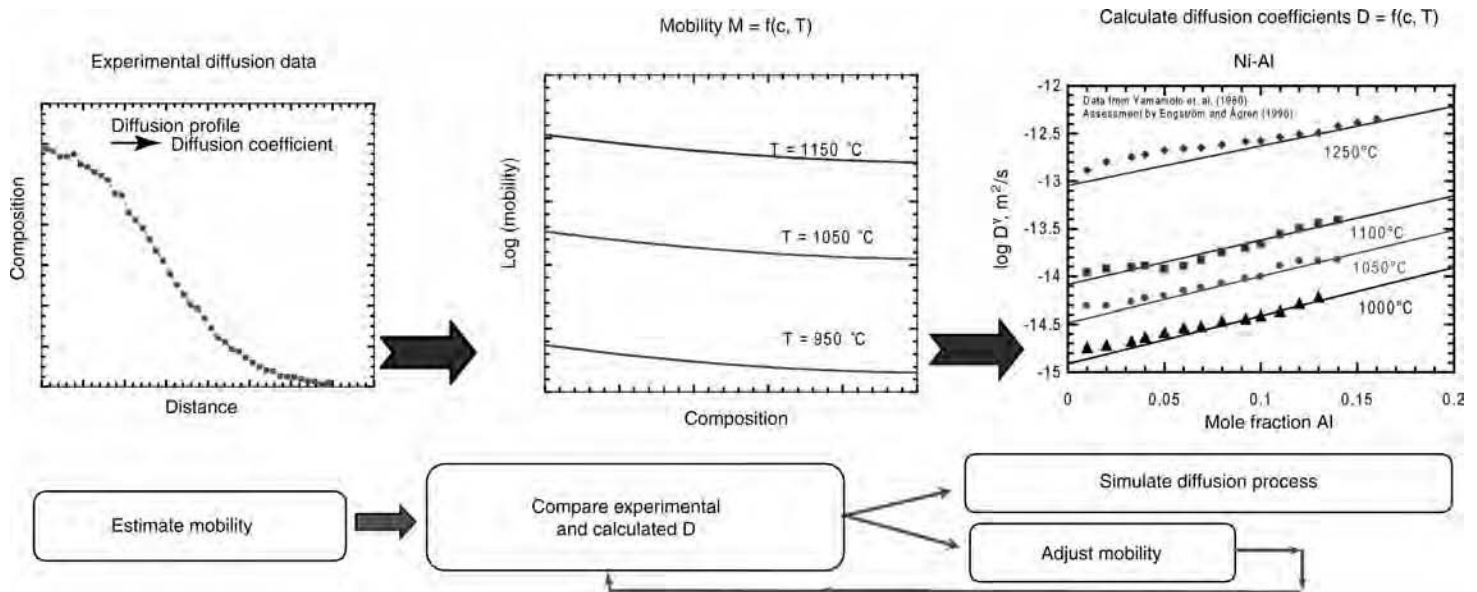
Fig. 8 Schematic of diffusion mobility parameter optimization procedure

$Q_{\text{Ni}}^{\text{Ni}}$ is the self-activation energy for diffusion of nickel in pure fcc nickel; it is well established experimentally, and the mobility parameters are previously determined in the assessment work by Ref 32. Two metastable end-member self-activation energies must be determined: one for the diffusion of tungsten in pure fcc tungsten, Q_{W}^{W}, and one for nickel in pure fcc tungsten, Q_{Ni}^{W}. Because these quantities cannot be measured experimentally (fcc tungsten is not stable), they are determined during the optimization, using diffusion correlations or first principles as initial estimates. The activation energy for tungsten diffusing in fcc nickel, Q_{W}^{Ni}, is determined using the available tracer and interdiffusion data. The three activation energies are modeled using the following form, where B and C are constants:

$$B + RT \ln(C) \qquad \text{(Eq 34)}$$

The binary interaction parameters, $^{0}A_{\text{Ni}}^{\text{Ni,W}}$ and $^{0}A_{\text{W}}^{\text{Ni,W}}$, are also considered in the optimization and are modeled using a constant value.

To start the assessment, the initial values for the activation energies are all set to equal the value of nickel in pure nickel: $-28,700 + 69.8*T$ (Ref 32) (Table 3). The binary interaction parameters are initially set to zero. Note that with this initial set of parameters, there is no concentration dependence for the tracer diffusivities. As seen in Fig. 9(a, b), the tracer diffusivities for all three alloy compositions are equal, using the initial parameters. The first parameters optimized are activation energies for the diffusion of tungsten in pure tungsten and nickel, Q_{W}^{W} and Q_{W}^{Ni}, using the tungsten tracer diffusivity data from Ref 56. The optimization is done by comparing the diffusivities calculated using the values defined for Eq 33 in Eq 13 and the experimental values:

$$D_{\text{W}}^{*}(x_{\text{W}}, T) =$$

$$\frac{1}{RT} \exp\left(\frac{\left(x_{\text{Ni}} Q_{\text{W}}^{\text{Ni}} + x_{\text{W}} Q_{\text{W}}^{\text{W}} + x_{\text{Ni}} x_{\text{W}}{}^{0} A_{\text{W}}^{\text{Ni,W}} \right)}{RT} \right)$$

$$\text{(Eq 35)}$$

After these parameters are optimized, the binary interaction term, $^{0}A_{\text{W}}^{\text{Ni,W}}$, is optimized, using a start value of -5000 J/mol. Once the tungsten mobility parameters have been optimized using tracer diffusivity data, the values for Q_{W}^{Ni}, Q_{W}^{W}, and $^{0}A_{\text{W}}^{\text{Ni,W}}$ are then fixed, and the nickel mobility parameters (Q_{Ni}^{W} and $^{0}A_{\text{Ni}}^{\text{Ni,W}}$) are optimized using the nickel tracer diffusivity data. After these nickel mobility values are optimized, all of the mobility values (excluding the values for $Q_{\text{Ni}}^{\text{Ni}}$) are optimized using both the tracer diffusivity data and the interdiffusion data. Again, the optimized parameters are used to calculate the tracer and interdiffusion diffusivities given by Eq 13 and 15 and then compared to the experimental values. The optimized mobility parameters are listed in Table 3.

The comparison of the diffusivities calculated with the optimized parameters and the experimental values are shown in Fig. 9 and 10. Good agreement between the measured and calculated tracer diffusivity and interdiffusion coefficients is achieved. Other recent diffusion mobility assessment examples are given in the literature by Ref 58 to 60.

In addition to optimizing the mobility functions using various composition-dependent diffusion coefficient data, diffusion mobility functions can be optimized directly from experimental composition profiles. Both Ref 61 and 62 developed methods that combine DICTRA with an optimization tool (MatLab or Mathematica) to assess the mobility parameters from

experimental composition profiles. For a given set of mobility parameters, the difference between the experimental composition and calculated composition is defined by a least-squares error function. The mobility parameters are optimized to minimize the error. This method has been successfully demonstrated for binary and ternary systems.

Strengths and Weaknesses of Assessment Method. The Calphad-based approach to modeling the diffusion mobilities provides an efficient representation of the composition dependence in multicomponent systems. The reduced number of parameters needed to describe diffusion in a multicomponent system occurs as a result of the assumption that the correlation factors are negligible in the lattice-fixed frame of reference, and only the diagonal terms of the mobility matrix must be evaluated. However, if the vacancy concentration is not in local equilibrium, the off-diagonal terms resulting from the correlation factors should be considered (Ref 5, 19). Using the Calphad method to describe the composition dependence of the mobility terms requires the determination of mobilities for fictive metastable end-member phases. Examples of such quantities are the mobility of tungsten in fcc tungsten and the mobility of tungsten in fcc aluminum at temperatures above the fcc aluminum melting temperature (e.g., 1300 °C). Determination of these end-member quantities may follow approaches similar to those used to determine the lattice stabilities of the metastable thermodynamic quantities of the elements (Ref 26–28). This determination of diffusion-activation energies for fictive end-member phases may appear to limit the Calphad method; however, it is these determinations that enable the extrapolation to higher-order systems where diffusion data are limited.

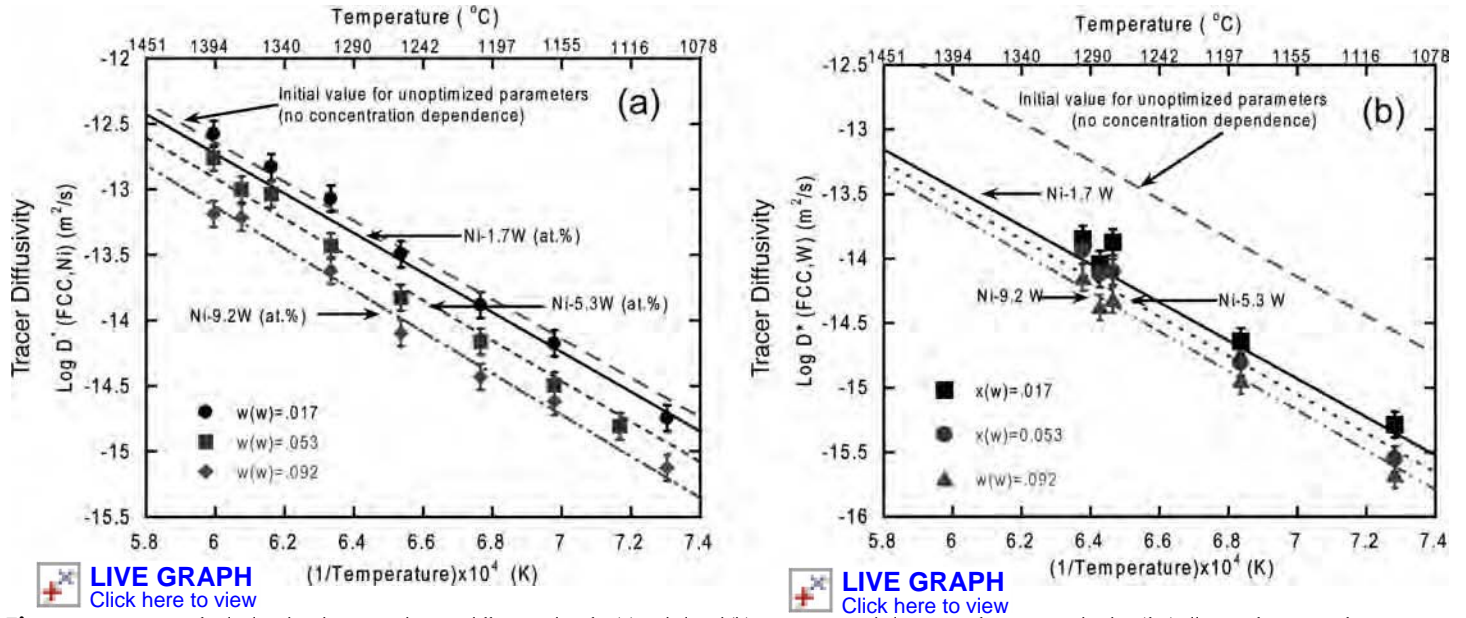These optimization methods have been employed to develop several commercial

Fig. 9 Comparison of calculated and measured tracer diffusivity data for (a) nickel and (b) tungsten in nickel-tungsten face-centered cubic (fcc) alloys as functions of temperature. The diffusivities are calculated before and after optimizing the nickel-tungsten system. Note: Before the system optimizes, there is no composition dependence in the mobility functions. Experimental data from Ref 56
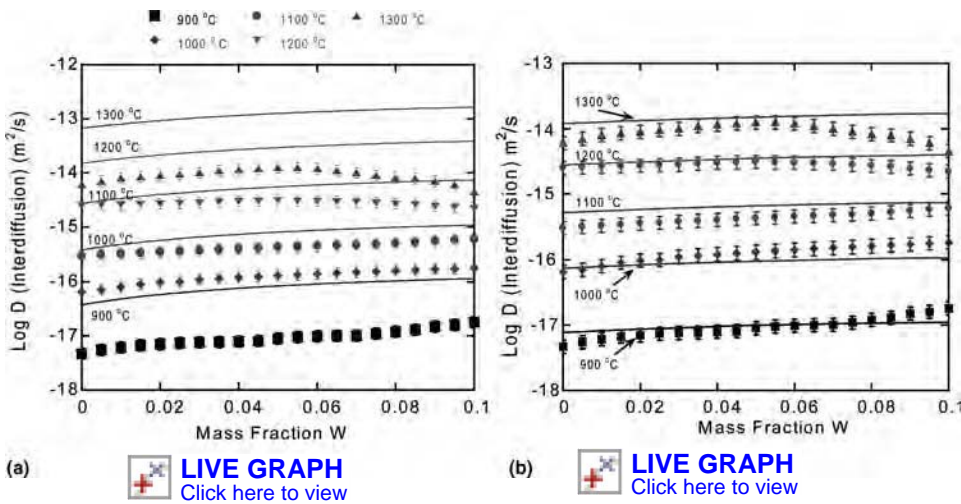

Fig. 10 Comparison of calculated and measured interdiffusion coefficients for the nickel-tungsten system as a function of composition for temperatures ranging from 900 to 1300 °C. (a) Calculated diffusivities before optimization; observed composition dependence is entirely from the composition dependence of the chemical potentials. (b) Calculated diffusivities after the nickel-tungsten system has been optimized. Source: Ref 57

Table 3 Optimized mobility parameters for the nickel-tungsten system

Parameter	Initial values, J/mol	Optimized value, J/mol
Q_{Ni}^{Ni}	$-28,700 - 69.8*T$ (Ref 32)	$-28,700 - 69.8*T$ (not optimized)
Q_{Ni}^{W}	$-287,000 - 69.8*T$	$-628,250 + RT$ $\ln(4.78 \times 10^{-4})$
Q_{W}^{W}	$-287,000 - 69.8*T$	$-411,423 + RT$ $\ln(2.18 \times 10^{-4})$
Q_{W}^{Ni}	$-287,000 - 69.8*T$	$-282,130 + RT$ $\ln(2.80 \times 10^{-5})$
$^{0}A_{Ni}^{Ni,W}$	$+ 0.0$	$+175,736$
$^{0}A_{W}^{Ni,W}$	$+ 0.0$	$-97,025$

Table 4 Available mobility databases

Database name/alloy system	Reference
MOB2 (general-purpose database, with emphasis on iron-base alloys	Ref 63
MOBAl1 (aluminum-base alloys)	Ref 63
MOBNi1	Ref 63
NIST-NiMob (face-centered cubic nickel-base alloys)	Ref 64
Cobalt-base alloys (face-centered cubic phase)	Ref 65
Solder alloys	Ref 66, 67
Zirconium-base alloys	Ref 68
Cemented carbides	Ref 69, 70

diffusion mobility databases, as well as many smaller databases for specific research applications. Table 4 lists several of the commercial and research databases available. All of these databases can be employed by a variety of diffusion codes, finite-difference codes that assume local equilibrium at each grid point (e.g., DICTRA), random walk methods (Ref 71), and phase-field codes.

Application

Single-Phase Diffusion. The most common diffusion simulation is the diffusion of one single-phase material into another at constant temperature and pressure. The results of these simulations are generally shown as composition profiles as a function of distance at a specified time. The complexity associated with an 11-component nickel-base superalloy diffusion couple is demonstrated in Fig. 11, where the interdiffusion between two single-phase γ(fcc) nickel-base superalloys (René-N4/René-N5) after 100 h at 1293 °C (Ref 72) is shown. The predictions were made using the DICTRA software in conjunction with the National Institute of Standards and Technology (NIST)-NiMob diffusion mobility database (Ref 64) and the Thermotech Ni-Data thermodynamic database (Ref 73). In addition to accurately predicting the composition profiles, the diffusion simulation also predicts the location of the maximum pore formation resulting from Kirkendall porosity. Figure 12(a) shows the predicted location of the maximum pore formation, given by the maximum of the negative derivative of the vacancy flux with respect to distance (Ref 74), and Fig. 12(b) reveals that the predicted location corresponds well to the location of Kirkendall porosity observed on the René-N4 side of the diffusion couple.
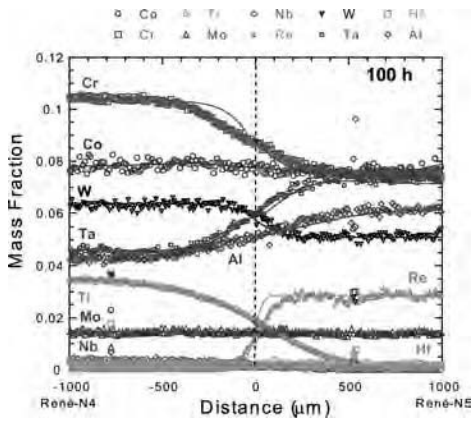
Fig. 11 Calculated (solid lines) and experimental (symbols) composition profiles for René-N4/ René-N5 diffusion couples after 100 h at 1293 °C. Source: Ref 72
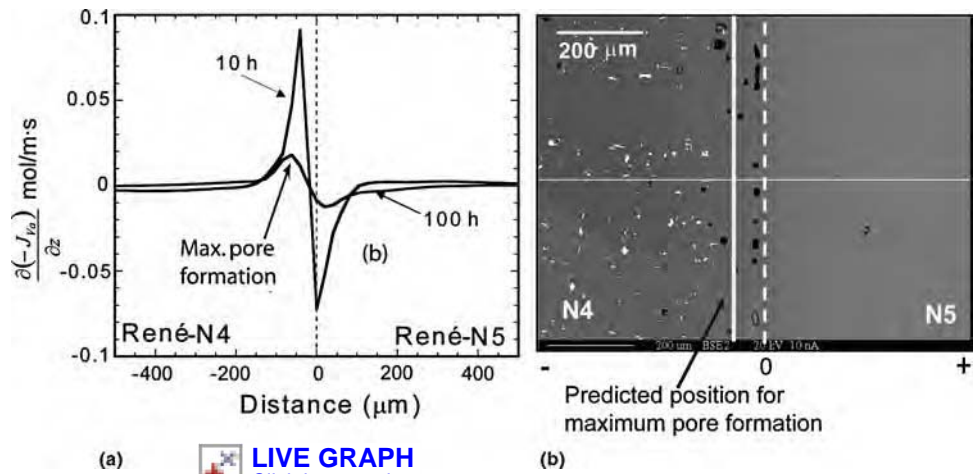
LIVE GRAPH
Click here to view

Fig. 12 (a) Predicted location of the maximum pore density for a René-N4/René-N5 diffusion couple at 1293 °C. (b) Backscattered image of René-N4/René-N5 diffusion couple after 100 h at 1293 °C. The thin white line indicates the position of the microprobe scan. The dashed white line corresponds to the Matano interface. The dashed line is the location of the predicted maximum porosity. Source: Ref 72
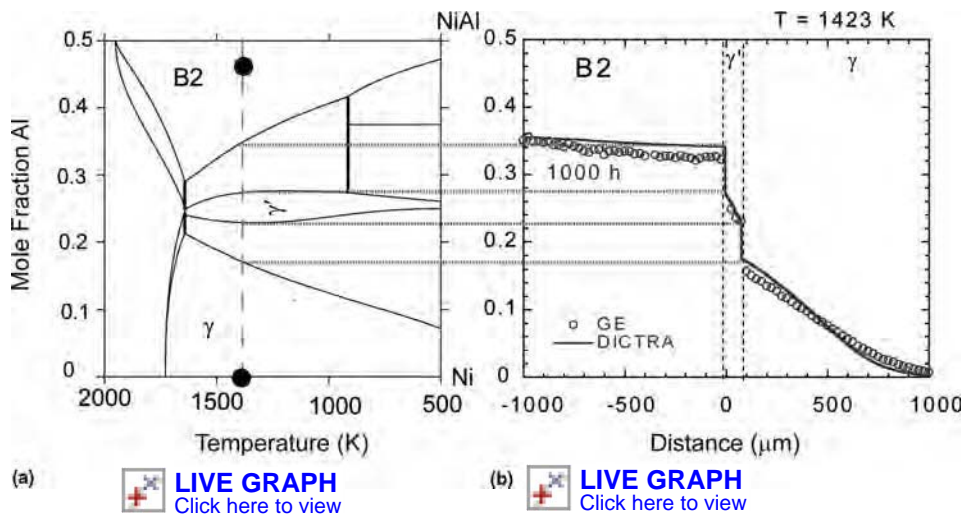
Fig. 13 Example of multiphase diffusion in the nickel-aluminum system at 1423 K. (a) The Ni-NiAl section of the nickel-aluminum phase diagram, where the solid circles indicate the initial end-member compositions of the diffusion couple. (b) The measured (open symbols) and calculated (solid line) composition profiles for the Ni-NiAl diffusion couple at 1423 K. Vertical dashed lines indicate the position of the Ni_3Al ($\gamma\gamma'$) region. The horizontal dashed lines extending between (a) and (b) demonstrate that the positions of the measured and calculated composition jumps corresponding to the different phase regions on the phase diagram. Source: Ref 46

Multiphase Diffusion. The formation of an additional phase is also a common occurrence during the diffusion process between two single-phase materials. Figure 13 shows the formation of a Ni_3Al (gamma prime, γ') layer between the NiAl-B2 and fcc nickel diffusion couple after heat treating for 1000 h at 1150 °C (1423 K). The Ni-NiAl section of the nickel-aluminum phase diagram is shown in Fig. 13(a). In Fig. 13(b), both the measured and predicted composition profiles show the formation of a Ni_3Al (gamma prime) layer between the initially present NiAl-B2 and nickel layers. The dashed lines between Fig. 13(a) and (b) show how the jumps in the composition profiles relate to the phase boundaries on the phase diagram.

In addition to the complexity of multicomponent single-phase diffusion couples, industrially relevant diffusion simulations often involve complicated time-temperature schedules and the precipitation and dissolution of a variety of different phases, both as planar layers and dispersed particles. These simulations are characterized by a variety of outputs, including the position of a moving phase boundary as a function of time, phase fraction profiles, particle-size diameters during coarsening, and locations of Kirkendall porosity. These outputs are essential in optimizing heat treating cycles and solidification schedules, predicting service lifetimes, and determining weldability. Examples of the complex uses of diffusion data to predict

microstructure evolution are found in other articles of this Handbook.

Appendix 1: Example of Diffusion Matrices for the Ni-0.05Al-0.10Cr fcc Composition at 1200 °C

The following is an example of the calculation of the various diffusion matrices for a ternary Ni-Al-Cr system, for a given set of diffusion mobilities and chemical potentials. The diffusion mobilities and chemical potentials are calculated using the data from Engström (Ref 33) and the SSOL4 substance database (Ref 75), respectively.

For the given Ni-0.05Al-0.1Cr (atomic fraction) composition at 1200 °C, the diffusion mobilities, in a lattice-fixed frame of reference, assuming the correlation effects are negligible, are given as:

$$
^{L}M_{ik} = \delta_{ik}x_iM_i =
\begin{bmatrix}
 & Al & Cr & Ni \\
Al & (0.05 \cdot 4.25) & 0 & 0 \\
Cr & 0 & (0.10 \cdot 2.35) & 0 \\
Ni & 0 & 0 & (0.85 \cdot 1.75)
\end{bmatrix} \cdot 10^{-18}\,m^2/s
$$

$$
=
\begin{bmatrix}
 & Al & Cr & Ni \\
Al & 2.12 & 0 & 0 \\
Cr & 0 & 2.35 & 0 \\
Ni & 0 & 0 & 14.9
\end{bmatrix} \cdot 10^{-19}\,m^2/s
$$

(Eq 1.1)

Assuming nickel is the dependent variable, the matrix of chemical potentials at 1200 °C is given by:

$$
\left[\frac{\partial\mu_k}{\partial x_j}\right] =
\begin{bmatrix}
 & Al & Cr & Ni \\
Al & +39.2 & +9.23 & +0 \\
Cr & +8.60 & +18.3 & +0 \\
Ni & -3.32 & -2.70 & +0
\end{bmatrix} \cdot 10^4\,(J/mol)
$$

(Eq 1.2)

The tracer diffusivities for the Ni-5Al-10Cr composition at 1200 °C are then calculated using Eq 13:

$$\mathbf{D}^* = \begin{bmatrix} \text{Al} & 5.20 \\ \text{Cr} & 2.87 \\ \text{Ni} & 2.14 \end{bmatrix} \cdot (\times 10^{-14} \, \text{m}^2/\text{s}) \qquad \text{(Eq 1.3)}$$

The diffusion matrix in the lattice frame of reference is the product of the mobility matrix (Eq 1.1) times the chemical potential matrix (Eq 1.2), as defined in Eq 21:

$$\mathbf{^L D} = (\delta_{xi} x_i [\mathbf{M}]) \times \left[\frac{\partial \boldsymbol{\mu_k}}{\partial \mathbf{x_j}} \right]$$

$$= \begin{bmatrix} & \text{Al} & \text{Cr} & \text{Ni} \\ \text{Al} & +8.33 & +1.96 & +0 \\ \text{Cr} & +2.09 & +4.30 & +0 \\ \text{Ni} & -4.95 & -4.02 & +0 \end{bmatrix} \cdot 10^{-14} \, (\text{m}^2/\text{s})$$

(Eq 1.4)

The diffusion matrix in the volume-fixed frame of reference, assuming constant molar volumes, is given by Eq 15:

$$\mathbf{^V D^n_{kj}} = (\delta_{ik} - x_k) x_i \mathbf{M} \times \frac{\partial \mu_i}{\partial \mathbf{x_j}}$$

$$= \begin{bmatrix} & \text{Al} & \text{Cr} & \text{Ni} \\ \text{Al} & 2.02 & -0.12 & -0.74 \\ \text{Cr} & -0.21 & 2.11 & -1.49 \\ \text{Ni} & -1.81 & -1.99 & 2.23 \end{bmatrix} \cdot 10^{-19}$$

$$= \begin{bmatrix} & \text{Al} & \text{Cr} & \text{Ni} \\ \text{Al} & +39.2 & +9.23 & +0 \\ \text{Cr} & +8.60 & +18.3 & +0 \\ \text{Ni} & -3.32 & -2.70 & +0 \end{bmatrix} \cdot 10^4$$

$$= \begin{bmatrix} & \text{Al} & \text{Cr} & \text{Ni} \\ \text{Al} & +8.06 & +1.85 & +0 \\ \text{Cr} & +1.48 & +4.07 & +0 \\ \text{Ni} & -9.54 & -5.92 & +0 \end{bmatrix} \cdot 10^{-14} \, (\text{m}^2/\text{s})$$

(Eq 1.5)

The reduced diffusion matrix in the volume-fixed frame of reference (Eq 15), which is commonly referred to as the interdiffusion coefficient matrix, is defined as the following, where nickel is the dependent variable:

$$\mathbf{^V D^{Ni}_{kj}} = \mathbf{^V D_{kj}} - \mathbf{^V D_{kn}}$$

$$^V \tilde{D}^{Ni}_{kj} = \begin{bmatrix} & \text{Al} & \text{Cr} \\ \text{Al} & ^V D^{Ni}_{AlAl} & ^V D^{Ni}_{AlNi} \\ \text{Cr} & ^V D^{Ni}_{CrAl} & ^V D^{Ni}_{CrCr} \end{bmatrix}$$

$$^V \tilde{D}^{Ni}_{kj} = \begin{bmatrix} & \text{Al} & \text{Cr} \\ \text{Al} & ^V D_{AlAl} - ^V D_{AlNi} & ^V D_{AlCr} - ^V D_{AlNi} \\ \text{Cr} & ^V D_{CrAl} - ^V D_{CrNi} & ^V D_{CrCr} - ^V D_{CrNi} \end{bmatrix}$$

$$= \begin{bmatrix} & \text{Al} & \text{Cr} \\ \text{Al} & +8.06 & +1.85 \\ \text{Cr} & 1.48 & +4.07 \end{bmatrix} \cdot 10^{-14} \, (\text{m}^2/\text{s})$$

(Eq 1.6)

Using this matrix of interdiffusion coefficients, the flux equations for the specified composition can be written in terms of Eq 16:

$$\tilde{J}_{Al} = -^V \tilde{D}^{Ni}_{AlAl} \frac{\partial c_{Al}}{\partial z} - ^V \tilde{D}^{Ni}_{AlCr} \frac{\partial c_{Cr}}{\partial z} \qquad \text{(Eq 1.7)}$$

$$\tilde{J}_{Cr} = -^V \tilde{D}^{Ni}_{CrAl} \frac{\partial c_{Al}}{\partial z} - ^V \tilde{D}^{Ni}_{CrCr} \frac{\partial c_{Cr}}{\partial z} \qquad \text{(Eq 1.8)}$$

Solution of these flux equations enables the calculation of composition profiles as a function of time in a single-phase region.

REFERENCES

1. A. Fick, Ueber Diffusion, *Poggendorff's Annalen,* Vol 94, 1855 p 59–86
2. A. Fick, *Philos. Mag. A*, Vol 10, S 4, 1855 p 30–39
3. H.S. Carslaw and J.C. Jaeger, *Conduction of Heat in Solids,* Clarendon Press, Oxford, 1959
4. J. Crank, *The Mathematics of Diffusion,* 2nd ed., Clarendon Press, Oxford, 1975
5. J. Philibert, *Atom Movements: Diffusion and Mass Transport in Solids,* Editions de Physique, Les Ulis, France, 1991
6. P. Shewmon, *Diffusion in Solids,* 2nd ed., TMS, Warrendale, PA, 1989
7. L. Boltzmann, *Ann. Physik,* Vol 53, 1894, p 960
8. C. Matano, *Jpn. J. Phys.,* Vol 8, 1933, p 109
9. A. Paul, A.A. Kodentsov, and F.J.J. van Loo, On Diffusion in the β-NiAl Phase, *J. Alloy. Compd.,* Vol 403, 2005, p 147–153
10. V.F. Sauer and V. Freise, Diffusion in binren Gemischen mit Volumennderung, *Z. Elektrochem.,* Vol 66 (No. 4), 1962, p 353–363
11. F.J.A. d. Broeder, A General Simplification and Improvement of the Matano-Boltzmann Method in the Determination of the Interdiffusion Coefficients in Binary Systems, *Scr. Metall.,* Vol 3, 1969, p 321–326
12. C. Wagner, The Evaluation of Data Obtained with Diffusion Couples of Binary Single-Phase and Multiphase System, *Acta Metall.,* Vol 17 (No. 2), 1969, p 99–107
13. M.A. Dayananda, Average Effective Interdiffusion Coefficients and the Matano Plane Composition, *Metall. Mater. Trans. A,* Vol 27 (No. 9), 1996, p 2504–2509
14. M.A. Dayananda, MultiDiFlux, Purdue University, West Lafayette, IN, 2005, https://engineering.purdue.edu/MSE/Research/MultiDiFlux/index.htm
15. E.O. Kirkendall, Diffusion of Zinc in Alpha Brass, *Trans. AIMME,* Vol 147, 1942, p 104–110
16. A.D. Smigelskas and E.O. Kirkendall, Zinc Diffusion in Alpha-Brass, *Trans. AIMME,* Vol 171, 1947, p 130–142
17. L.S. Darken, Diffusion, Mobility and Their Interrelation through Free Energy in Binary Metallic Systems, *Trans. AIMME,* Vol 175, 1948, p 184–201
18. J.R. Manning, Diffusion and the Kirkendall Shift in Binary Alloys, *Acta Metall.,* Vol 15 (No. 5), 1967, p 817–826
19. J.R. Manning, Cross Terms in the Thermodynamic Diffusion Equations for Multicomponent Alloys, *Metall. Trans.,* Vol 1 (No. 2), 1970, p 499–505
20. A. Janotti, M. Krcmar, C.L. Fu, and R.C. Reed, Solute Diffusion in Metals: Larger Atoms Can Move Faster, *Phys. Rev. Lett.,* Vol 92 (No. 8), 2004, p 085901-1
21. Y. Mishin and D. Farkas, Atomistic Simulation of Point Defects and Diffusion in B2 NiAl: Part I: Point Defect Energetics, *Philos. Mag. A,* Vol 75 (No. 1), 1997, p 169–185
22. Y. Mishin and D. Farkas, Atomistic Simulation of Point Defects and Diffusion in B2 NiAl: Part II: Diffusion Mechanisms, *Philos. Mag. A,* Vol 75 (No. 1), 1997, p 187–199
23. Y. Mishin, A. Lozovoi, and A. Alavi, Evaluation of Diffusion Mechanisms in NiAl by Embedded-Atom and First-Principles Calculations, *Phys. Rev. B,* Vol 67, 2003, p 014201
24. A. Van der Ven, G. Ceder, M. Asta, and P.D. Tepesch, First-Principles Theory of Ionic Diffusion with Nondilute Carriers, *Phys. Rev. B,* Vol 64, 2001, p 184307
25. H.-C. Yu, D.-H. Yeon, A. Van der Ven, and K. Thornton, Substitutional Diffusion and Kirkendall Effect in Binary Crystalline Solids Containing Discrete Sources and Sinks, *Acta Mater.,* Vol 55, 2007, p 6690–6704
26. L. Kaufman and H. Bernstein, *Computer Calculation of Phase Diagrams,* Academic Press, London, 1970
27. N. Saunders and A.P. Miodownik, *CALPHAD Calculation of Phase Diagrams: A Comprehensive Guide,* R.W. Cahn, Ed., Elsevier Science Inc., New York, 1998
28. H. Lukas, S.G. Fries, and B. Sundman, *Computational Thermodynamics: The CALPHAD Method,* Cambridge University Press, Cambridge, 2007
29. J.-O. Andersson and J. Ågren, Models for Numerical Treatment of Multicomponent Diffusion in Simple Phases, *J. Appl. Phys.,* Vol 72 (No. 4), 1992, p 1350–1355
30. O. Redlich and A. Kister, Algebraic Representation of Thermodynamic Properties and the Classification of Solutions, *Ind. Eng. Chem.,* Vol 40 (No. 2), 1948, p 345–348
31. B. Jönsson, Assessment of the Mobilities of Cr, Fe and Ni in bcc Cr-Fe-Ni Alloys, *ISIJ Int.,* Vol 35 (No. 11), 1995, p 1415–1421
32. B. Jönsson, Assessment of the Mobilities of Cr, Fe, and Ni in fcc Cr-Fe-Ni Alloys, *Z. Metallkd.,* Vol 86 (No. 10), 1995, p 686–692
33. A. Engström and J. Ågren, Assessment of Diffusional Mobilities in Face-Centered Cubic Ni-Cr-Al Alloys, *Z. Metallkd.,* Vol 87 (No. 2), 1996, p 92–97
34. T. Helander and J. Ågren, Diffusion in the B2-BCC Phase of the Al-Fe-Ni System—Application of a Phenomenological Model, *Acta Mater.,* Vol 47 (No. 11), 1999, p 3291–3300
35. T. Helander and J. Ågren, A Phenomenological Treatment of Diffusion in Al-Fe and Al-Ni Alloys Having B2-BCC Ordered Structure, *Acta Mater.,* Vol 47 (No. 4), 1999, p 1141–1152
36. J. Kucera and B. Million, *Metall. Trans. A,* Vol 1, 1970 p 2603–2606
37. J. Ågren, Diffusion in Phases with Several Components and Sublattices, *J. Phys. Chem. Solids,* Vol 43 (No. 5), 1982, p 421–430

38. B. Jönsson, Assessment of the Mobility of Carbon in fcc C-Cr-Fe-Ni Alloys, *Z. Metallkd.,* Vol 85 (No. 7), 1994, p 502–509

39. B. Jönsson, On Ferromagnetic Ordering and Lattice Diffusion—A Simple Model, *Z. Metallkd.,* Vol 83 (No. 5), 1992, p 349–355

40. B. Jönsson, Ferromagnetic Ordering and Diffusion of Carbon and Nitrogen in Bcc Cr-Fe-Ni Alloys, *Z. Metallkd.,* Vol 85 (No. 7), 1994, p 498–501

41. R. Braun and M. Fellerkniepmeier, On the Magnetic-Anomalies of Self-Diffusion and Heterodiffusion in Bcc-Iron, *Scr. Metall.,* Vol 20 (No. 1), 1986, p 7–11

42. J. Ågren, Computer Simulation of the Austenite/Ferrite Diffusional Transformation in Low Alloyed Steels, *Acta Metall.,* Vol 30, 1982, p 841–851

43. J. Ågren, Complex Simulations of Diffusional Reactions in Complex Steels, *ISIJ Int.,* Vol 32 (No. 3), 1992, p 291–296

44. L.A. Girifalco, Vacancy Concentration and Diffusion in Order-Disorder Alloys, *J. Phys. Chem. Solids,* Vol 24, 1964, p 323–333

45. Z. Tôkei, J. Bernardini, P. Gas, and D.L. Beke, Volume Diffusion of Iron in Fe_3Al: Influence of Ordering, *Acta Mater.,* Vol 45 (No. 2), 1997, p 541–546

46. C.E. Campbell, Assessment of the Diffusion Mobilities in the Ni-Al-Cr B2 and γ' phases, *Acta Mater.,* Vol 56 (No. 16), 2008, p 4277–4290

47. M. Hillert, L. Hoglund, and J. Agren, Diffusion in Interstitial Compounds with Thermal and Stoichiometric Defects, *J. Appl. Phys.,* Vol 98 (No. 5), 2005, p 053511

48. H.L. Lukas, E.T. Henig, and B. Zimmermann, Optimization of Phase-Diagrams by a Least-Squares Method Using Simultaneously Different Types of Data, *Calphad,* Vol 1 (No. 3), 1977, p 225–236

49. D.W. Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *J. Soc. Ind. Appl. Math.,* Vol 11 (No. 2), 1963, p 431–441

50. E. Konigsberger, Improvement of Excess Parameters from Thermodynamic and Phase-Diagram Data by a Sequential Bayes Algorithm, *Calphad,* Vol 15 (No. 1), 1991, p 69–78

51. B. Jansson, "PARRssOT, " Report TRITA-MAC-0234, Royal Institute of Technology, Stockholm Sweden,1984

52. J.O. Andersson, T. Helander, L. Höglund, P.F. Shi, and B. Sundman, Thermo-Calc and DICTRA, Computational Tools for Materials Science, *Calphad,* Vol 26, 2002, p 273–312

53. A. Borgenstam, A. Engström, L. Höglund, and J. Ågren, DICTRA, A Tool for Simulation of Diffusion Transformations in Alloys, *J. Phase Equil.,* Vol 21 (No. 3), 2000, p 269–280

54. A.M. Brown and M.F. Ashby, Correlations for Diffusion Constants, *Acta Metall.,* Vol 28, 1980, p 1085–1101

55. P. Gustafson, A. Gabriel, and I. Ansara, A Thermodynamic Evaluation of the C-Ni-W System, *Z. Metallkd.,* Vol 28, 1987, p 151–156

56. K. Monma, H. Suto, and H. Oikawa, Diffusion of Ni-63 and W-185 in Ni-W Alloys, *J. Jpn. Inst. Met.,* Vol 28, 1964, p 197–200

57. M.S.A. Karunaratne, P. Carter, and R.C. Reed, Interdiffusion in the fcc-Centred Cubic Phase of the Ni-Re, Ni-Ta, and Ni-W Systems between 900 and 1300 °C, *Mater. Sci. Eng. A,* Vol 281, 2000, p 229–233

58. Y.W. Cui, K. Oikawa, R. Kainuma, and K. Ishida, Study of Diffusion Mobility of Al-Zn Solid Solution, *J. Phase Equil. Dif.,* Vol 27 (No. 4), 2006, p 333–342

59. L. Zhang, Y. Du, Y. Ouyang, H. Xu, X.-G. Lu, Y. Liu, Y. Kong, and J. Wang, Atomic Mobilities, Diffusivities and Simulation of Diffusion Growth in the Co-Si System, *Acta Mater.,* Vol 56, 2008, p 3940–3950

60. J. Wang, L.B. Liu, H.S. Liu, and Z.P. Jin, Assessment of the Diffusional Mobilities in the Face-Centred Cubic Au-Ni Alloys, *Calphad,* Vol 31 (No. 2), 2007, p 249–255

61. C.E. Campbell, A New Technique for Evaluating Diffusion Mobility Parameters, *J. Phase Equil. Dif.,* Vol 26 (No. 5), 2005, p 435–440

62. L. Höglund, DICTRA ToolBox, Thermo-Calc AB, Stockholm, Sweden, 2004

63. Thermo-Calc Software, http://www.thermocalc.com/

64. C.E. Campbell, W.J. Boettinger, and U.R. Kattner, Development of a Diffusion Mobility Database for Ni-Base Superalloys, *Acta Mater.,* Vol 50, 2002, p 775–792

65. T. G¢mez-Acedbo, B. Navarcorena, and F. Castro, Interdiffusion in Multiphase, Al-Co-Cr-Ni-Ti Diffusion Couples, *J. Phase Equil. Dif.,* Vol 25 (No. 3), 2004, p 237–251

66. G. Ghosh, Dissolution and Interfacial Reactions of Thin-Film Ti/Ni/Ag Metallizations in Solder Joints, *Acta Mater.,* Vol 49 (No. 14), 2001, p 2609–2624

67. G. Ghosh and Z.K. Liu, Modeling the Atomic Transport Kinetics in High-Lead Solders, *J. Electron. Mater.,* Vol 27 (No. 12), 1998, p 1362–1366

68. C. Toffolon-Masclet, M. Mathon, A. Engström, and J.C. Brachet, in *14th International Symposium on Zirconium in the Nuclear Industry* (Stockholm, Sweden), 2004

69. Y.L. He, L. Li, S.G. Huang, J. Vleugels, and O. Van der Bies, Computer Simulation of W-C-Co-V System Diffusion Couples, *Rare Met.,* Vol 26 (No. 5), 2007, p 492–497

70. S. Haglund and J. Ågren, *Acta Mater.,* Vol 46 (No. 8), 1998, p 2801–2807

71. H. Larsson and J. Ågren, Combined Probability Distributions of Random-Walks: A New Method to Simulate Diffusion Processes, *Comput. Mater. Sci.,* Vol 34, 2005, p 254–263

72. C.E. Campbell, W.J. Boettinger, T. Hansen, P. Merewether, and B.A. Mueller, Examination of Multicomponent Diffusion Between Two Ni-Base Superalloys, *Third International Alloy Conference* (Lisbon, Portugal), Springer, 2002, p 241–250

73. N. Saunders, Phase Diagram Calculations for Ni-Based Superalloys, *Superalloys 1996* (Warrendale, PA),TMS, 1996, p 101–110

74. L. Höglund and J. Ågren, Analysis of the Kirkendall Effect, Marker Migration and Pore Formation, *Acta Mater.,* Vol 49, 2001, p 1311–1317

75. Scientific Group Thermodata Europe (SGTE), http://www.sgte.org

SELECTED REFERENCES

Diffusion texts

- M.E. Glicksman, *Diffusion in Solids: Field Theory, Solid-State Principles, and Applications,* Wiley-Interscience, 1999

- D. Gupta, *Diffusion Processes in Advanced Technological Materials,* Series William Andrew, Inc., Norwich, New York, 2005

- P. Heitjans and J. Kärger, *Diffusion in Condensed Matter,* Springer, New York, 2005

- J.S. Kirkaldy and D. Young, *Diffusion in the Condensed State,* Institute of Metals, London, 1987

- F.J.J. v. Loo, Multiphase Diffusion in Binary and Ternary Solid-State Systems, *Prog. Solid State Chem.,* Vol 20, 1990, p 47–99

- H. Mehrer, *Diffusion in Solids: Fundamentals, Methods, Materials, Diffusion-Controlled Processes,* Springer, 2007

- J. Philibert, *Atom Movements: Diffusion and Mass Transport in Solids,* Series Editions de Physique, Les Ulis, France, 1991

- P. Shewmon, *Diffusion in Solids,* 2nd ed., TMS, Warrendale, PA, 1989

Critical diffusion data collections

- H. Bakker, H.P. Bonzel, C.M. Bruff, M.A. Dayananda, W. Gust, J. Horvath, I. Kaur, G.V. Kidson, A.D. LeClaire, H. Mehrer, G.E. Murch, G. Neumann, N. Stolica, and N.A. Stolwijk, Ed., *Landolt-Börnstein: Diffusion in Solid Metals and Alloys,* New Series, Group III, Vol 25 and 26, Springer-Verlag, Berlin, 1990

- Diffusion and Defect Forum, http://www.scientific.net/DDF/

- W.F. Gale and T.C. Totemeier, Ed., *Smithells Metals Reference Book,* Elsevier, 2004

- G. Neumann and C. Tuijn, *Self-Diffusion and Impurity Diffusion in Pure Metals: Handbook of Experimental Data,* Elsevier, New York, 2009

- NIMS Diffusion Data, https://kuzuha.nims.go.jp/diffusion/diffusion/index_en.html

- NIST Diffusion Data Center, http://patapsco.nist.gov/diffusion/

Localization Parameter for the Prediction of Interface Structures and Reactions

Witold Lojkowski, Institute of High-Pressure Physics, Polish Academy of Sciences, Poland
Hans J. Fecht, Institute for Micro and Nanotechnology, University of Ulm, Germany

INTERFACE REACTIONS cannot be understood and modeled without knowing the structure of interfaces. However, there are an infinite number of possible interface structures, depending on the interface crystallography and physical parameters. Furthermore, interfaces may undergo phase transformations when the crystallographic parameters, temperature, pressure, or chemical composition are changed. The computation effort can be drastically reduced by focusing it on specific questions that result from the physical model.

The structure of interfaces can be described as being built of nanosized clusters of atoms interacting across the interface with each other. These clusters, or close-packed groups, interact with each other by means of elastic fields. However, if the interatomic interactions are relatively weak compared to the elastic interactions, the aforementioned model is not valid any more. Thus, the basic questions that must be answered to understand the interface structure are whether the energy of an interface can be minimized by formation of low-energy atomic clusters at the expense of elastic energy, what the structure is of such clusters, and whether a change of parameters leads to a change of structure. Transitions can occur between the aforementioned structures depending on independent variables such as the orientation relationships, temperature, pressure, and chemical composition.

A parameter called the localization parameter (p) was introduced that enables a simple method of predicting which structure is actually present under given conditions.

Internal interfaces form part of the microstructure of almost every natural or artificially produced material. An intercrystalline interface separates two crystals of different phases, while grain boundaries separate crystals of the same phase but with different spatial orientation. The arrangement of atoms in the interface is neither that of the crystal nor of a liquid (Ref 1). Depending on the mutual orientation relationship of the crystals and such parameters

as pressure, temperature, and chemical composition, there are an infinite number of interface structures. The energy structure and kinetics (Ref 2) as well as crucial properties of materials and devices, such as the mechanical strength, hardness, brittleness and ductility, corrosion and wetting behavior, transparency and optical properties, electrical resistivity, magnetic properties, reactivity, and so on, are determined by the properties of the interfaces (Ref 3, 4). The structure of interfaces plays an especially crucial role in nanotechnology, because in nanostructures a large fraction of atoms is situated at the interfaces (Ref 5). In microsystems technology, interfaces determine the service life and performance of devices (Ref 6).

Atomistic modeling is a powerful tool for understanding the structure of defects in a crystalline lattice. However, there is an infinite range of possible interface structures, and therefore, it is important to focus modeling efforts on specific interfaces; otherwise, the task of understanding the interface structure via modeling is almost impossible to be realized. A physical approximate model guides the direction of the modeling efforts and explains the structure of interfaces. With the help of such a model, the properties of all interfaces situated in the space of coordinates between some special interfaces can be estimated.

This article summarizes a physical model of the interface structure described in detail in Ref 2 and shows how this model may help to optimize atomistic modeling studies.

Interface Structure

A structure is defined as a set of elements and their interactions (Ref 2). What a structure element is depends on what kind of interactions are important in a given process. When the interaction is between individual atoms, the structure elements are atoms. When the interaction is between groups of atoms, the interactions are long-range fields, for example, elastic fields.

Interatomic interactions can be precisely calculated using various modeling techniques. In some cases, however, an approximate estimation can be made, starting from such data as cohesion energy, vaporization energy (showing the energy needed to break interatomic bonds in a material composed of one type of atoms), and free energy to form the given chemical compound. Furthermore, the chemical composition of the atmosphere around the material may influence the structure of the interface by changing the thermodynamic equilibrium and the strength of the interatomic bonds (Ref 2). The structure of the interfaces may also depend on the size of the crystals. In epitaxy or during the growth of precipitates, for instance, it depends on the amount of elastic energy the crystals can store (Ref 2).

The Orientation Relationship

The orientation relationship (OR) defines the mutual crystallographic position of the adjacent crystals. More about interface crystallography can be found in the book by Randle (Ref 7). To describe the OR of two crystals, there is a need for nine equilibrium and three nonequilibrium parameters (Ref 2). Imagine that two crystals are allowed to interpenetrate, forming a bicrystal lattice. Two parameters determine the orientation of the rotation axis, one determines the rotation around this axis (misorientation), and three determine the rigid body translation of one crystal relative to the other. This makes six parameters. Now, cut the system with a plane, and allow only atoms of one phase on each side of the plane. The orientation of the interface plane is defined by two parameters and its position by one parameter. At this point, nine parameters define the interface. Now, allow the system to relax to minimum energy. The spacing of the two crystals (in other terms, the interface thickness) and their mutual position (shift parallel to the interface) will adjust to a minimum energy value. At first glance,

the structure of the interface is now determined as a function of the aforementioned nine crystallographic parameters. However, some external stress may shift the interface out of such a local energy minimum: stress-induced change of thickness and shift parallel to the interface induced by interaction of the interface with dislocations (Ref 8–10), vacancies, segregated atoms, external or internal stresses, or high curvature (Ref 11). These conditions can be locally described by three parameters, which are the three principal stresses acting on a small segment of an interface, for example, the local distance between the two crystals and the local torsion or stretching (Ref 2). Thus, interfaces in a state of nonequilibrium are characterized by 12 parameters.

When the crystallographic parameters are varied, the interface may display a sharp or shallow energy minimum. One may imagine a nine-dimensional surface with energy valleys and energy minima or maxima. The set of crystallographic parameters of interfaces situated in the energy minima are called low-energy ORs (LEORs). The interfaces are called low-energy interfaces or special interfaces. Coincidence boundaries are those where the bicrystal lattice is periodic, with a relatively short period. The number of atoms in a periodic cell of such a bicrystal lattice are designated Σ. For instance, a coincidence boundary $\Sigma5$ means that for each atom in coincidence in the two lattices, there are atoms in each lattice that are not coincident. Much research was focused on symmetric grain boundaries, where the interface is a mirror plane, or on tilt grain boundaries, where the rotation axis is perpendicular to the plane. Less attention was given to interfaces parallel to index planes in at least one crystal, although, as shown subsequently, they are candidates to correspond to LEORs as well as the symmetric and periodic ones.

In real experiments or in atomistic modeling, one parameter is usually varied, and it is assumed the others take values of minimum energy by a relaxation process. In most cases, the variable is the misorientation angle or the inclination of the interface plane.

Model-Informed Atomistic Modeling of Interface Structures

The concept of model-informed atomistic modeling of interfaces is based on the analysis by Ashby (Ref 12) of the role of modeling and empiricism in reaching engineering solutions. A purely empirically based approach, as well as a purely atomistic modeling-based approach, lead to having too many parameters to be considered, for instance, the results of the calculations of atomic positions, or too many experimental variables. Performing atomistic modeling is very similar to an experiment and needs interpretation in terms of a physical model equally well as any experiment. A physical model is useful if it produces an analytical

expression describing important properties of the system as a function of independent variables. Atomistic modeling may help to calculate some parameters that can only be estimated using a physical model. Combining both saves computational time, experiment costs, and permits one to foresee results for interfaces that have not been modeled. The aim of this article is to provide for a physical model that may permit interpolating the results of atomistic modeling to predict the properties of interfaces.

Nanosized Structural Elements of the Interface

Atoms may interact with each other individually or as groups of atoms. In the latter case, interactions bonding an atom within the group must be stronger than those between the whole group. Structural elements or units (SUs) are atoms or groups of atoms. Information is sought about their internal structure, their relative position in space, and the interactions between them. For this assumption to be valid, the SUs cannot be a topological construction, where only mutual positions of atoms count but not their distances. The atoms being part of the SUs should preserve their relative positions to each other regardless of the OR of the interface in which the SU is embedded. If interactions between individual atoms across groups are equally as strong as within the group, the group reduces to a single atom, and the idea of structure groups becomes irrelevant for the interface structure. The original concept of SUs proposed by Bishop and Chalmers (Ref 13, 14) agrees well with this assumption, since their proposed SUs are built from close-packed groups of atoms in the size range of a nanometer.

Several computer modeling studies and transmission microscopy observations (Ref 13–18) have led to the conclusion that, in many cases, the interface structure can indeed be described in terms of SUs (Fig. 1). It was shown that low-energy grain boundaries are built from SUs of only one type, and grain boundaries of higher energy are built from a mixture of different SUs.

Because the two crystals must be strained to permit SUs to be built at their interface, there are dislocations between different kinds of SUs. Thus, the interface energy is lowered at the price of building elastic strain fields. Of course, if the atoms within the SUs are moved out of their positions, the localized strain field is relaxed due to interactions, and the SU model of the structure is invalid.

The dislocations separating the low-energy sectors of interfaces are called structural dislocations. For interface boundaries they are also called misfit dislocations; for grain boundaries, they are called displacement shift complete (DSC) dislocations. Figure 2 schematically illustrates a DSC dislocation for the $\Sigma5$ coincidence grain boundary.

Theories to Predict Low-Energy Orientation Relationships

It was soon realized that short-period interfaces with a large fraction of atoms belonging to vaguely defined "good matching areas" may display a relatively low energy (Ref 1, 19). The "good matching areas" are, in fact, regions of locally dense atomic packing. It was also soon realized that besides local atomic matching, low energy is connected with an average dense packing of atoms found in interfaces parallel to low-index crystallographic planes.
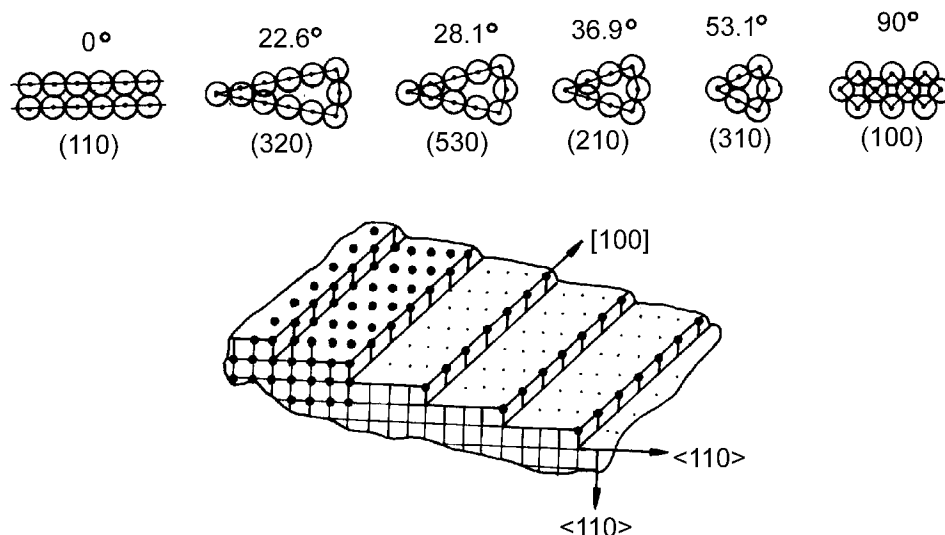

Fig. 1 Structural units (SUs) for grain boundaries in metals. Grain boundaries of low energy are composed of periodic arrangements of such units, and grain boundaries with different orientations are a mixture of SUs. Source: Ref 13
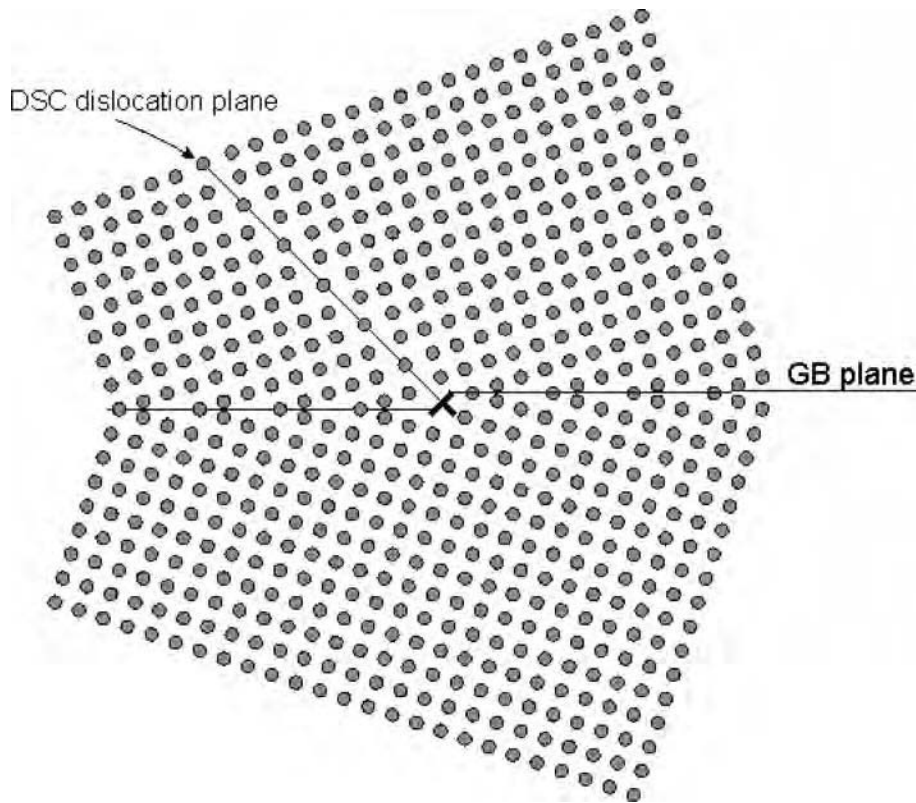
Fig. 2 Schematic illustration of a displacement shift complete (DSC) dislocation. The grain boundary was created by plotting two regular arrays of circles and approaching them in such a way that a periodic structure was created, without any relaxation. On each side of the dislocation, the grain-boundary structure is identical and composed of a regular arrangement of structural units. The Burgers vector of the DSC dislocation is smaller than for a lattice dislocation. Source: Ref 2
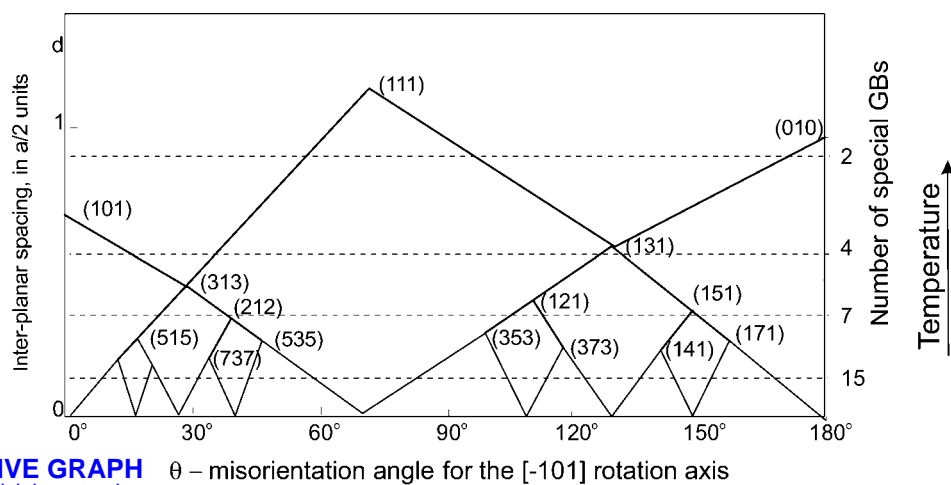

LIVE GRAPH
Click here to view

Fig. 3 Plot of grain boundary planes in the fcc lattice that correspond to low energy orientation relationships (LEORs). It is assumed that the LEORS are symmetric tilt grain boundaries with the tilt axis [−101]. When the angle of misorientation increases from 0° to 180°, some grain boundaries will be low energy ones corresponding to symmetric tilt with low index planes. The other ones will acquire SUs from the closest symmetrical grain boundaries with low index planes. For instance, the (373), grain boundary (GB) consists of 2 units of the (131) GB and one unit of the (111) or of 3 (111) units and 2 (010) units. The number of minima decreases as the temperature increases. At temperature close to melting point, only two minima are present. Source: Ref 20 and 2.

Wolf (Ref 20, 21), Paidar (Ref 22), Lojkowski et al. (Ref 14), and Fecht (Ref 23) have shown that special interfaces are parallel to the dense lattice planes in both crystals. For instance, Fecht (Ref 23) has shown that the pressure required to separate two crystals is:

$$\sigma(d^*) = \frac{E_0}{l}\exp(-d^*) \cdot (-1 - 0.15d^* + 0.05d^*)$$

(Eq 1)

where l is the characteristic screening length, and d^* is the interplanar spacing in dimensionless units. Because the strain to separate the two crystals increases when their distance decreases, the energy of such interfaces is also lower than for other interfaces. Close-packed interfaces with low interplanar spacing also have the lowest vibrational entropy, and thus, they are connected with energy minima at temperatures close to the melting point.

Figure 3, from Paidar's work (Ref 22), shows the hierarchy of grain boundaries in a face-centered cubic (fcc) material as sources of structural units to grain boundaries. Only symmetrical interfaces are considered, and their symmetry planes are marked in the diagram. The plane indices shown in Fig. 3 define special interfaces as follows: The rotation axis is <110>, and the plane (*xyz*) is the interface plane for a symmetric tilt boundary. Grain boundaries situated between two grain boundaries with index planes listed in the figure will "borrow" SUs from these grain boundaries. Crystalline planes with low crystallographic indices ({111}, {110}, and {100} in the fcc structure) have the densest packing and thus correspond to the lowest energy minima.

The temperature effect on the interface structure is now considered. It is known that with increasing temperature, the number of LEORs that are a source of SUs decrease (Ref 2), because the energy of thermal vibrations becomes comparable to the energy bonding the atoms to their SU, and with increasing temperature, the grain-boundary structure becomes gradually disordered. This process can be described as delocalization of the cores that separate the SUs. When the cores merge and the interfacial bonding is comparable to the strength of bonds inside the crystals, the interface will assume a disordered, amorphous-like structure. At temperatures close to the melting point, only the (111) and (010) planes are sources of SUs, and these are the close-packed atomic units described in the SU model by Bishop and Chalmers (Ref 14).

The earlier consideration concerned the tilt of grain boundaries. When two crystals are twisted relative to each other, the favored twist angle is where the dense atomic rows in one crystal are parallel to valleys between the atomic rows in the other crystal (Ref 14, 24), so that the two crystals can mutually lock (Fig. 4). The energy minima when the twist angle is varied are of less depth than in the case of tilt.

When factors that increase the density of atomic packing in the interface act simultaneously, as shown in Fig. 5 (Ref 2), the energy at the interface is minimized. These factors are:

- A, good local bonding (short distance between the good matching areas)
- B, parallelism of dense atomic planes on the two sides of the interface
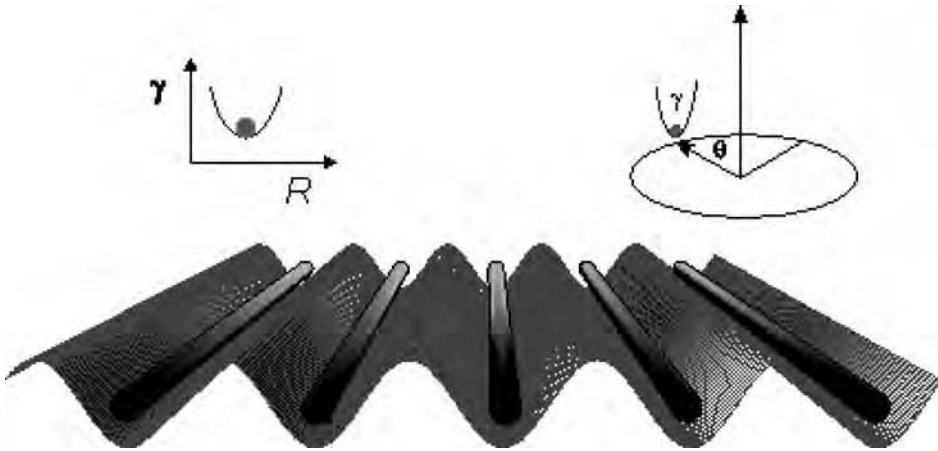- C, parallelism of the dense atomic rows

Fig. 4 Low-energy structure of an interface due to mutual locking of dense rows of atoms of one crystal into the surface of another. A twist, tilt, or rigid body shift of one of the crystals relative to the other is associated with an increase in energy. γ is the surface energy, θ is the tilt or twist angle, and R is the rigid body translation. Source: Ref 2
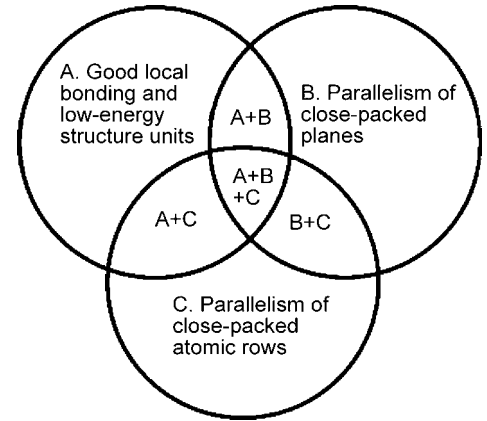

Fig. 5 Schematic illustration of factors contributing to energy minima of interfaces. Factors A, B, or C lower the energy, but their coincidence leads to the deepest energy minima. Source: Ref 2

(a) p = 0, no dislocations, non-coherent interface

(b) p = 0.3, delocalised core non-coherent interface

(c) p = 1, localised core semi-coherent interface

(d) p > 1, stand-off coherent interface
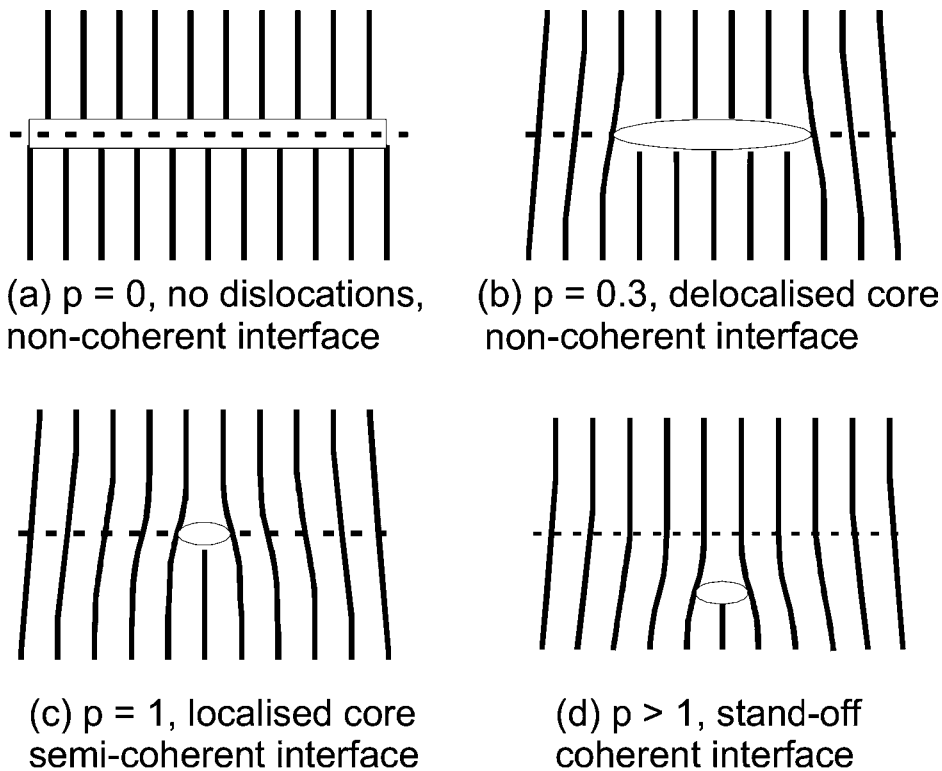
Fig. 6 Schematic of interface structures as a function of the localization parameter value, *p*. Source: Ref 2

Use of the Localization Parameter for Prediction of Interface Structures

The previous considerations concerned an interface between two phases, where the variables were the temperature and misorientation between the crystals. It was noted that a change of temperature may lead to a change of interfacial bonding strength, leading to a transformation in their structure. However, a similar effect can be caused by a change in thermodynamic conditions. The bonding strength across the interface also depends on the kind of bonded materials. Furthermore, an important role may be played by the thickness of the crystals, since the elastic energy depends on the thickness. Therefore, there is need for a theory permitting the prediction of the interface structure between two different crystalline phases. The presence of interfacial

dislocations is an indication that the SUs model well describes the interface structure. For such interfaces, one may expect high adhesion between the two crystals and the presence of sharp energy minima for LEORs (Ref 2). The principal question is: Under what conditions is this model of the interface structure applicable at all?

On one hand, there can be interfaces where adhesion is low, and their energy is not significantly lowered for the special LEORs; SUs do not form, and the atoms do not shift much from their position in each crystal. On the other hand, there can be interfaces with intermediate structures between these two, or amorphous interfaces. So, a basic question is whether energy optimization will lead to a coherent interface at the expense of elastic energy or whether it will lead to the formation of a fully incoherent interface. From an engineering point of view, the question is: Is it possible to predict which of the aforementioned structures will form in a real material?

It was recently shown that it can be predicted which interface forms by using one single parameter, called the localization parameter (*p*), expressed as:

$$p = G_i/G \qquad \text{(Eq 2)}$$

where G_i is the interface shear modulus, and G is the shear modulus of the less stiff of the two crystals. The notion of interface shear modulus is explained in detail in Ref 2.

Figure 6 (Ref 2) shows how the *p*-value determines the interface structure. The structure shown in Fig. 6(a) corresponds to the situation when $G_i << G$; stretching the bonds across the interface does not cost energy, and there is no gain in building SUs and separating them at dislocations. Some energy decrease is associated with the parallelism of atomic rows and/or planes. Figure 6(b) shows the intermediate case, where the structure of the interface is controlled both by the parallelism criterion

and by increasing atomic matching. Figure 6(c) shows schematically the case when the interface forms SUs, and dislocations help to minimize the energy. Figure 6(d) corresponds to $p > 1$, that is, where the shear stress in the interface is higher than in one of the crystals. In such a case, the misfit dislocations can be pushed away from the interface into the bulk material to conserve energy.

The interface shear modulus is defined in the same way as for the bulk structure: by the proportionality coefficient between applied shear and deformation. A detailed treatment of the interface shear modulus is given in Ref 2.

The equilibrium width of the structural dislocation core, s_m, is comparable to a dislocation in the bulk material:

$$s_m = aG/G_i = as/p \qquad \text{(Eq 3)}$$

This result is almost identical to the result for the core width of a dislocation given by the Perierls model. The only difference is that in the Perierls model, $G = G_i$.

In the Perierls model, the crystal is regarded as a continuum in which the elastic energy is stored, and its value can be calculated using the continuous elasticity theory. The equilibrium core width corresponds to the minimum of the sum of the elastic energy and stretched bonds energy:

$$\frac{d(E_{el} + E_{core})}{ds} = 0 \qquad \text{(Eq 4)}$$

The dislocation is represented as the infinite sum of dislocations with infinitely small Burgers vectors:

$$b = \int_{-\infty}^{\infty} \frac{du}{dx} dx \qquad \text{(Eq 5)}$$

The dislocation core width, s, is the diameter of the circuit enclosing half of the aforementioned dislocations. The stretched bonds energy in the plane of slip of the dislocation is calculated using an analytic expression and is proportional to the bulk shear modulus:

$$E_{el} = \text{const} \cdot G \qquad \text{(Eq 6)}$$

By simply replacing G with G_i, Eq 3 has been obtained. The constant a accounts for all the approximations made. Its exact value can be calculated only using atomistic modeling.

Equation 3 explains the meaning of the term *localization parameter*. For $p = 0$, the misfit is delocalized, and there is no atomic matching; for $p = 1$, the misfit is localized, and the misfit dislocation core has a width similar to a bulk dislocation. The previous equation was obtained by analytically solving simple equations for the energy of the interface, where the elastic part increases as the core is more localized, and the contribution made by broken bonds decreases as the dislocation core is more localized.

In the case of thin layers or small precipitates, the elastic energy stored in the material is less than for large crystals. Therefore, the interface structure for a thin layer or small precipitate may be coherent, while for large crystals, it may be amorphous or semicoherent. This is because the elastic energy increases with increasing material thickness, while the broken bond energy remains constant.

Estimating the Shear Modulus and Bonding Energy Across the Interface

To apply Eq 3 and predict the interface structure, it is necessary to know the interface shear modulus. It was assumed that:

$$G_i = E_{AB}/\Omega \qquad \text{(Eq 7)}$$

where E_{AB} is the bonding energy across the interface and Ω is the atomic volume. The bonding energy includes a contribution from the ionic or covalent bonding of atoms across the interface and van der Waals forces:

$$E_{AB} = C_B \frac{E_A + E_B}{2} + \Delta G_{AB} + E_{WAALS} \qquad \text{(Eq 8)}$$

where E_A and E_B are the cohesion energies of crystals A and B, respectively; ΔG_{AB} is the chemical bonding energy between the two crystals, and E_{WAALS} is the van der Waals bonding energy. In the case of grain boundaries in metals, it depends weakly on temperature and is expressed as:

$$E_{AB} = E_{COH} = \frac{E_A + E_B}{2} + \Delta G_{AB} \qquad \text{(Eq 9)}$$

where E_{COH} is cohesion energy.

The previous equations give a fairly good approximation for the localization parameter, and they enable the structure of interfaces to be predicted (Tables 1, 2). However, in some cases, it is not possible to predict which atom of one crystal forms a bond with which atom of another crystal. Furthermore, in ionic crystals, Madelung forces contribute to the bonding, which are not tabulated for interface atoms (Ref 25). Hence, it is important to carry out calculations assuming different kinds of atoms in front of each other in an interface structure and to understand whether the proper value of the interaction energy was taken into account in Eq 2.

Figure 7 shows how atomistic modeling can provide for proper application of the aforementioned model. An interface between two materials is shown: a chemical compound A-B and a chemical compound C-D. Atomistic modeling may help to determine what kind of chemical bonding is present and controls the adhesion of the crystals: A-C or A-D or B-C or B-D, and what is the interface shear modulus? With these data, it is possible to calculate the localization parameter, p. The figure also illustrates the concept of interface shear modulus, G_i.

Table 1 Correlation between the value of the localization parameter and properties of interfaces

Localization parameter (p)	Metal/oxide interface (p-value)	Interface description(a)
$p \gg 1$	Al/Al$_2$O$_3$ (2.1),	Standoff—the misfit dislocation is not in the interface but in the less stiff of the two crystals
	Nb/Al$_2$O$_3$ (1.5)	Coherence
		All features the same as for the 0.35–1 range
$0.35 < p < 1.1$	Al/Al (1.2), Al/GaAs (1.1),	Misfit dislocations separate structure units of low energy.
	NiCr (1.1), Au/Pd (0.9), MgO/CdO (0.9),	Periodic interfaces have low energy, but the depth of the energy minimum strongly depends on the orientation of the interface plane.
	MgO/MgO (0.8), Au/Au (0.8),	
	Pb/Sn (0.8), Au/Ag (0.7),	Asymmetric low-energy interfaces parallel to vicinal planes are observed
	Cu/Cu (0.7), Ag/Cu (0.7),	
	Cu/Fe (0.7), Ag/Ag (0.6),	
	Mo/Mo (0.6),	
	Al$_2$O$_3$/ZrO$_2$ (0.7),	
	Al$_2$O$_3$/Al$_2$O$_3$ (0.7),	
	Fe/Fe (0.5), Si/Si (0.5),	Combination of geometrical factors like short periodicity, parallelism to vicinal planes, and parallelism of close packed atomic rows leads to deep energy minima
	MgO/ZrO$_2$ (0.4),	
$0.3 < p < 0.35$	Fe/Al$_2$O$_3$ (0.3), Ni/MgO (0.3),	Parallelism
	NiO/NiO (0.3), ZrO$_2$/NiO (0.3),	Partially delocalized misfit dislocations
		LEORs of surface type
	MgO/Al$_2$O$_3$ (0.3),	
	Ag/Ni (0.35), SiC/SiC (0.3)	
$0.02 < p < 0.25$	Cu/NiO (0.2), Cu/MgO (0.2), Au/MgO (0.1),	Misfit dislocations are delocalized.
		Parallelism
	Al/MgO (<0.1), Ag/MgO (<0.1), Ag/NaCl (0.02),	Random twist angle at high temperatures
		All LEORs are surface type
	Au/NaCl (0.02), Au/LiF (0.01)	

(a) Characteristic features, type of low-energy orientation relationships (LEORs), degree of localization of misfit dislocations. Source: Ref 2

Table 2 Maximum disorientation angles for some groups of interfaces

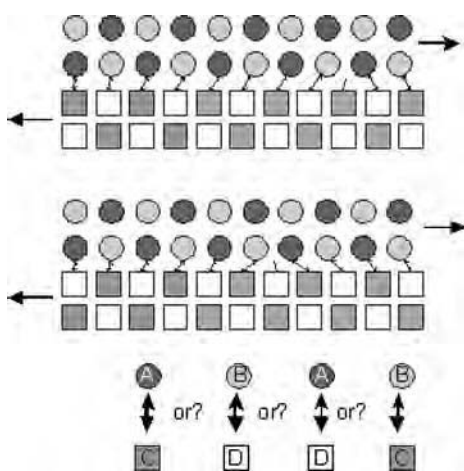| p-value	Maximum disorientation	Type of interface
$p < 1.0$	$\Delta\theta < 18°$	Metals and oxides of high cohesion energy
$p < 0.5$	$\Delta\theta < 9°$	Interfaces of medium bonding strength and/or highly stiff crystals
$p < 0.2$	$\Delta\theta < 4°$	SiC-like material and van der Waals bonding

Source: Ref 2

Fig. 7 Possible structures of an interface between two crystals. The top crystal (circles) has chemical components A-B, and the bottom (squares) C-D. Atomistic modeling helps to calculate which atoms are neighbors across the interface, what kind of chemical bonding is present, and what is the interface shear modulus. With these data, it is possible to calculate the localization parameter. The horizontal arrows at the interface represent an external stress applied to the interface. The small arrows linking atoms across the interface symbolize the stretched interatomic bonds across the interface. They are stretched due to misfit between the lattices of the two crystals and because of the applied stress. The interface shear modulus, G_i, describes the reaction of the interface to applied stress and therefore also the degree of localization of interface dislocation cores. Knowing its value and bulk elastic constants, it is possible to calculate the localization parameter, p.

Prediction of Interface Structure in Various Systems and Their Transformations

Table 1 lists the type of interface structure that can be expected, based on the aforementioned simplified calculations of the p-parameter.

It follows that the previous phase transformations in the interface structure (Fig. 8) will occur when moving in the space of coordinates of the crystallographic parameters, varying the misfit, temperature, pressure, and size of the crystal.

Table 2 shows the expected range of angles for the validity of the SU model, depending on the localization parameter, p.
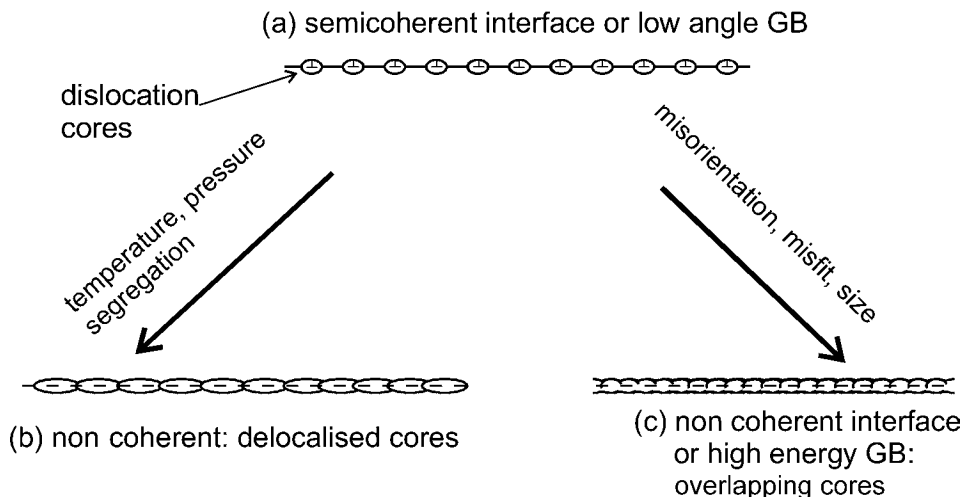
Fig. 8 Illustration of a change of interface structure, from that described in terms of structural units and interface dislocations to the one where maximizing atomic matching of the interface is not a criterion for low energy. (a) Structure with low-energy areas of the interface separated by interface dislocations. (b) Delocalization of a misfit dislocation due to change of bonding strength. (c) Overlap of dislocation cores due to their high density. Source: Ref 2

Figure 9 shows the phase diagram for various phases of an interface as a function of the misfit and thickness of the crystals.

Some maps showing the relaxation structures of interfaces depending on the degree of misfit and the adjoining metals are also given in the book by Howe based on the papers of Stoop and van der Merve (Ref 26) and Kato (Ref 27).

Therefore, changes of structure may occur because of:

- A change of bonding energy, which is a function of temperature and pressure or concentration of segregating atoms
- A change of thickness of the adjoining crystals
- A change of the misorientation of the crystals

However, the borders of the areas separating the various structures can hardly be estimated based only on the previous simplified calculations. The greatest challenge for analytical calculations is to estimate the bonding energy, E_{AB}, also including the temperature effects. Knowledge of the effect of interatomic spacing change on the energy of the broken bond is required. Similarly, the contribution of van der Waals forces is difficult to estimate, because they strongly depend on the crystal separation or separation of the interatomic rows. With an accurate value for the p-parameter, it may be possible to draw a precise phase diagram for the interface structure.

Nanoscale Effects. With the recent advent of nanotechnology, the importance of interfaces for the performance of materials and devices is crucial (Ref 28, 29), because for nanometric dimensions of the crystals, the atoms situated at interfaces form a large fraction of all the atoms in the material. It is therefore most important to consider how the effects of nanostructuring lead to a necessary modification of the localization parameter model of interfaces and include in the modeling efforts the effects of strong curvature, triple junctions, and small sizes of crystals.

Implications of Changes in Interface Structure for Interface Reactions

Because the interface structure is a function of the previously explained crystallographic parameters, strength of interfacial bonding, and stiffness of the crystals, the mechanism of interfacial reactions is a function of these parameters as well. Some examples of the application of the a forementioned theory of interface structure are presented to predict interface reaction mechanisms.

Growth of One Crystal at the Expense of Another. Two cases of interfaces between two crystals—a disordered interface and an ordered interface built from low-energy SUs separated by dislocations—are considered. In the case of a low-energy interface composed of one kind of SU, the reaction kinetics will be negligible, because there are no channels for transfer of atoms across the interface, except thermally activated vacancies. In the case of an interface built up from low-energy SUs separated by dislocations, the reaction kinetics will depend on the kinetics of glide and climb of the interfacial dislocations.

Figure 10 provides an example of a reaction driven by the movement of interface dislocations (Ref 30, 31). The high-resolution electron microscopy image shows misfit dislocations that accommodate the difference of the lattice constants of the two phases. This interface has a different structure for two different thicknesses of
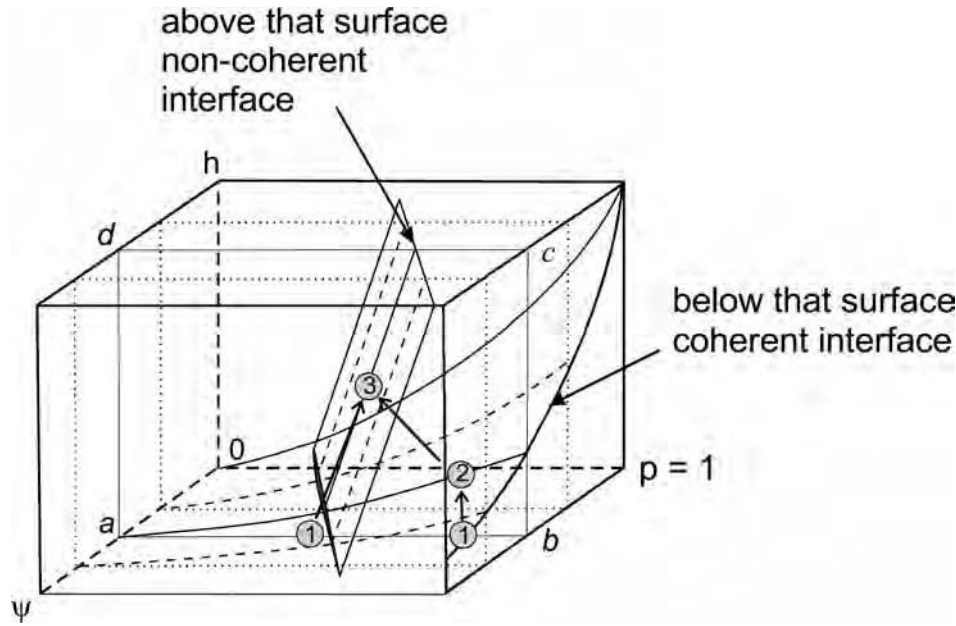
Fig. 9 Phase diagram with regions of stability of coherent, semicoherent, or noncoherent interfaces, depending on the value of the localization parameter value (p), misfit (ϕ), and crystal thickness (h). The plane ($abcd$) is a plane with a constant misfit. For low p-values, the path (1) $\Rightarrow$ (3) is the direct transformation from a coherent to a noncoherent interface. At high p-values, the sequence is (1) coherent, (2) semicoherent, and (3) noncoherent.
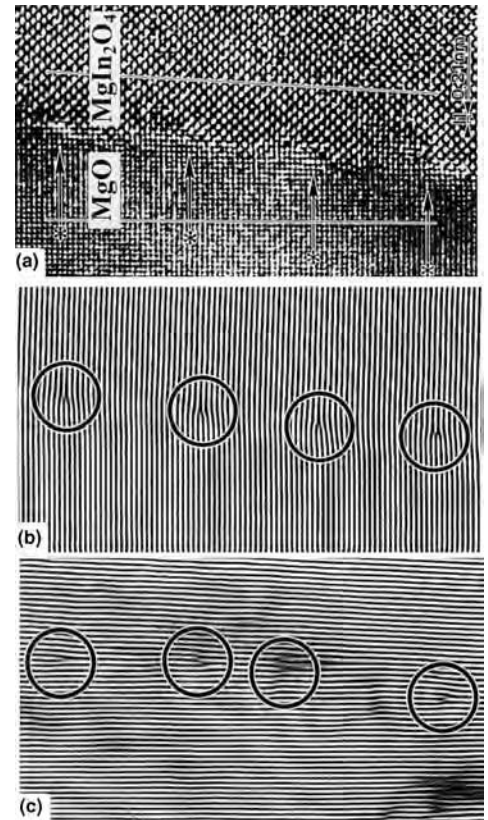

Fig. 10 High-resolution electron microscopy image of misfit dislocations that accommodate the difference of the lattice constants of the two phases MgO and $MgIn_2O_4$. In (a), the Burgers vectors of the misfit dislocations are parallel to the interface. The MgO layer grows slowly by a parabolic law, while the misfit dislocations climb. In (b), the Burgers vectors of the misfit dislocations are inclined to the interface plane. In (c), the MgO film grows by a combination of dislocation glide and climb, so that the rate of the film growth is faster and a linear growth law is followed as time elapses. Source: Ref 30

the growing spinel and different growth rates. In Fig. 10(a), the Burgers vectors of the misfit dislocations are parallel to the interface. The MgO layer grows slowly by a parabolic law, while the misfit dislocations climb. On the other hand, in Fig. 10(b), the Burgers vectors of the misfit dislocations are inclined to the interface plane. The MgO film grows by a combination of dislocation glide and climb, so that the rate of the film growth increases in a linear fashion. Hence, the dislocation structure of the interface is different for the two growth regimes. For the interfaces studied, misfit dislocation cores are the place where the actual reaction takes place that leads to new phase growth. It follows that for an interface composed of low-energy regions and structural dislocations, the kinetics of the reaction depend on the kind of dislocations present.

In the case of a disordered interface, one crystal will grow at the expense of the other along the whole interface, taking advantage of enhanced diffusion coefficients for mass transfer along and perpendicular to the interface. Grabski and Korski (Ref 8) proposed to regard interface kinetics as controlled by the number of channels that atoms can cross and the frequency of jumps. It is clear that for a disordered interface, anyplace can be such a channel for material transfer, while for interfaces built of SUs, such channels are most likely situated between the SUs, and their number depends on their spacing.

Surface Reactions in Thin Films. When a thin film grows on the surface of a large crystal, the first atomic layer must be fully coherent with the substrate, because the thin layer cannot sustain any stress. However, there will be a critical thickness when the elastic energy stored in the layer is already sufficient to drive one of two possible transformations:

- Growth in the form of islands separated by valleys, so that elastic energy is minimized
- Transformation into semicoherent and subsequently noncoherent interfaces

These stages are shown in Fig. 11 (Ref 2). Hence, the same interface may grow by different mechanisms, depending on the thickness of the growing layer.

The transition from coherent to semicoherent interface takes place when the grain-boundary structure with dislocations has less energy than the strained coherent structure (Ref 2):

$$h = \frac{bGp}{8E_Y\psi^2} \qquad \text{(Eq 10)}$$

where E_Y is Young's modulus, and G is shear modulus. It is clearly seen that h decreases as the p-value decreases, and ψ is the misfit between the lattices of the two crystals. One of the definitions of the misfit is as follows:

$$\psi = \frac{2|a_a - a_b|}{(a_a + a_b)} \qquad \text{(Eq 11)}$$

or:

$$\psi = \frac{\Delta}{a_o} \qquad \text{(Eq 12)}$$

where a_a and a_b are both lattice constants, Δ is $|a_a - a_b|$, and a_o is the average lattice constant.

However, if the dislocations have wide cores, they will overlap, and the semicoherent interface will transform into a noncoherent. The condition for a semicoherent interface is (Ref 2):

$$p > \frac{(a_a - a_b)}{b} \qquad \text{(Eq 13)}$$

where a_a and a_b are the lattice constants of the two phases. However, for delocalized MDs cores, this is a direct coherent-to-noncoherent transition instead of a coherent-to-semicoherent transition.

For $p = 0.1$ and $\psi = 2\%$, Eq 10 gives the value $h \approx 50\ b$, that is, approximately 50 atomic layers, which is a reasonable value.

Figure 9 (Ref 2) shows a phase diagram illustrating the regions of stability of each of the aforementioned structures. The plane *abcd* is a plane of constant misfit. For low-p-values, the path (1) => (3) is the direct transformation from a coherent to a noncoherent interface. At high-p-values, the sequence is (1) coherent => (2) semicoherent =>(3) noncoherent.
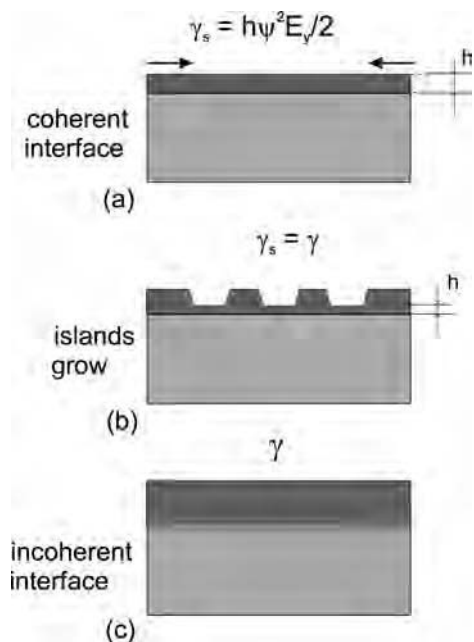
Fig. 11 Interface growth. (a) Coherent. (b) Semicoherent or coherent interface with formation of islands (so-called self-organization). (c) Noncoherent interface as the stored elastic energy caused by misfit increases. Source: Ref 2
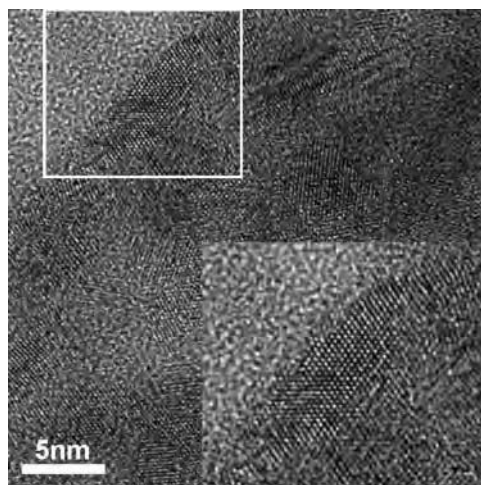

Fig. 12 High-resolution transmission electron microscopy image of the surface of a Raney nickel catalyst showing the small crystallite size and large number of corner and edge atoms. Courtesy of U. Hörmann, Ulm University, by personal communication

Some maps of interface relaxation structures that depend on misfit and adjoining metals are also given in the book by Howe based on the papers of Stoop and van der Merve (Ref 26) as well as Kato (Ref 27).

Surface Reactions with Nanoparticles. For nanosized particles on which a layer should grow due to interaction with the gas phase, similar to the manner for large crystals, SUs built from dense atomic blocks play a crucial role
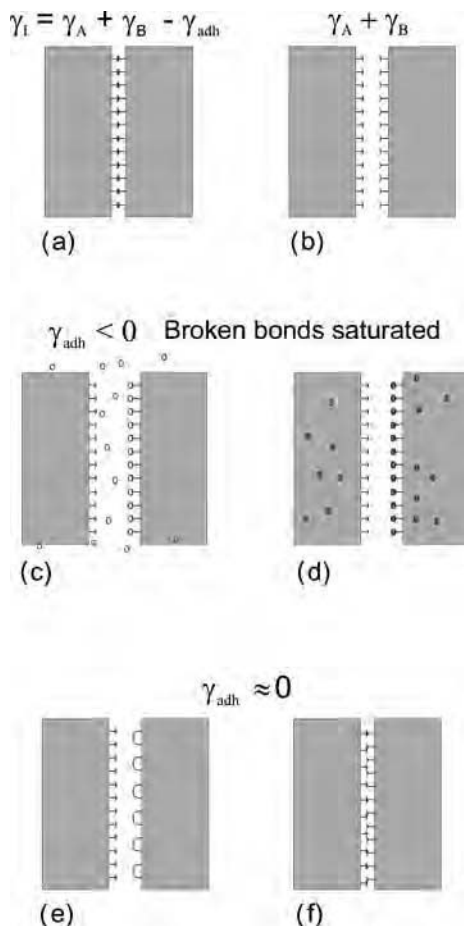
Fig. 13 The work of adhesion and factors decreasing adhesion. (a) Bonded interface. (b) Cleaved interface and dangling bonds. (c) Decrease of free surface energy and adhesion due to chemical reactions at the surfaces. (d) Decrease of free surface energy and adhesion due to segregation of impurities to the surfaces. (e) Decrease of surface energy and adhesion due to surface reconstruction. (f) Decrease of adhesion due to misfit. Source: Ref 2

in determining the process kinetics. As particles become smaller, they have a large number of atoms located at the edges and corners. The fraction of corners increases approximately d^{-3} as the diameter of the particle shrinks, while the total number of surface atoms changes only slightly when the particle size decreases from 10 to 2 nm. The number of corners and edges determines the rate of reactions between the particles and surrounding atmosphere, including catalytic reactions (Ref 32). An example of ledges on the surface of nanoparticles for the case of a Raney (sponge) nickel catalyst with a crystallite size of a few nanometers and a large number of corner and edge atoms is shown in Fig. 12. Thus, in the case of surfaces, similar SUs control interface kinetics, while in the case of interfaces, it is segments of close-packed planes and their ledges.

Bonding or Debonding Two Crystals. The crucial parameter for bonding or debonding two crystals is the work of adhesion:

$$\gamma_I = \gamma_A + \gamma_B - \gamma_{adh} \qquad (Eq\ 14)$$

where γ_A and γ_B are the surface energies of crystals A and B, respectively; γ_I is the interface surface energy; and γ_{adh} is the work of adhesion or adhesion energy. Figure 13 illustrates the concept of the work of adhesion. The work per unit surface to separate two crystals along the interface is the work of adhesion. Since the value of the localization parameter, p, depends on intercrystalline bonding energy, it depends on the adhesion energy as well. Actually, bonding energy and adhesion are equivalent (Fig. 13 a, b). Figure 13 illustrates qualitatively the various factors that may influence adhesion. Figures 13(c) and (d) show that saturation of the broken bonds by impurities from the gas atmosphere or atoms segregated to the interface may eliminate adhesion. In other words, the presence of such impurities or atoms added on purpose may completely change the bonding energy and interface structure. In real systems, that may be a kinetically controlled process. Figure 13(e) shows schematically how surface reconstruction may lead to a decrease of adhesion energy, since the broken bond energy would be partially relaxed by surface reconstruction. Figure 13(f) shows that high misfit leads to low adhesion as well. The aforementioned two types of reconstruction make it difficult to predict the interface energy based solely on the theoretical interaction energy between the atoms, as, for instance, in the Binder et al. (Ref 33), Miedema and den Broeder (Ref 34), Becker (Ref 35), or McDonnald-Eberhard (Ref 36) models.

In summary, the interface structure depends on factors completely independent of the crystallographic parameters, such as the presence of small quantities of foreign atoms or the possibility of surface reconstruction. Therefore, before any structural calculations are performed, the assessment of the interfacial bonding energy is of primary concern. This may impose a considerable challenge for atomistic modeling of the interface structure and reactions.

Conclusion

The kinetics of interface reactions depend on their structure, whether it is disordered and noncoherent, semicoherent, or with structural dislocations, or coherent/low energy composed of one kind of SUs. This, in turn, depends on the thermodynamic conditions, misfit of the crystals, size of the crystals, nature of the intercrystalline bonding, and nine crystallographic parameters that determine the mutual orientation of the crystals, without taking into account nonequilibrium structures. Phase transitions can occur between the various structures, depending on the variable parameters such as the orientation relationship, temperature, pressure, chemistry, and so on.

The physical model of the interfaces presented here may save computation time by:

- Indicating which interfaces play a crucial role in determining the structure of all other interfaces
- Asking crucial questions that advance understanding of the structure of interfaces and frequently cannot be answered without modeling.

The interface structures can be described as being built of nanosized clusters of atoms interacting with each other by means of elastic fields, or as dense atomic planes and rows interacting with each other by means of van der Waals forces and elastic fields, or as individual atoms interacting to locally form low-energy structures. The challenge is to develop phase diagrams for the interfacial structure based on precise calculations of the localization parameter, p. For that purpose, the interface shear modulus must be calculated. This, in turn, requires knowledge of the nature of interfacial bonding, particularly the Coulombic contribution, the covalent contribution, and the contribution of van der Waals forces. Knowledge of which atoms interact across the interface is also required. The localization parameter, p, can enable the prediction of the type of structure that is actually present in particular conditions.

ACKNOWLEDGMENT

REFERENCES

1. H. Gleiter, *Progr. Mater. Sci.*, Vol 10, p 125, 1981
2. W. Lojkowski and H.J. Fecht, The Structure of Intercrystalline Interfaces, *Progr. Mat. Sci.*, Vol 45, p 339, 2000 with permission from Elsevier
3. L.A. Murr, *Interfacial Phenomena in Metals and Alloys*, Addison-Wesley Publishing Company, London, 1975
4. H.-J. Fecht and G. Wilde, Processing Nanoscale Structures to Macro-Composites, *CRC Handbook*, 2007
5. *Nanostructure Science and Technology, A Worldwide Study*, R.W. Siegel, E. Hu, and M.C. Roco, Ed., WTEC, Loyola College in Maryland, 1999, http://www.wtec.org/loyola/nano/
6. J. Auersperg, B. Seiler, E. Cadalen, R. Dudek, and B. Michel, Simulation of Interface Cracks in Microelectronic Packaging, *16th European Conference of Fracture, Failure Analysis of Nano and Engineering Materials and Structures*, Alexandroupolis, Greece, July 3–7, 2006
7. V. Randle, The Measurement of Grain Boundary Geometry, *Electron Microscopy in Materials Science Series*, B. Cantor and M.J. Goringe, Ed., Institute of Physics Publishing, Bristol and Philadelphia, 1993
8. M.W. Grabski and R. Korski, *Phil. Mag.*, Vol 22, p 707, 1970
9. R.Z. Valiev, V.Yu. Gertsman, and O.A. Kaibyshev, *Phys. St. Sol.(a)*, Vol 97, p 11, 1986
10. E. Rabkin and L. Klinger, The Fascination of Grain Boundary Grooves, *Adv. Eng. Mater.*, Vol 3, p 277, 2001
11. P.H. Pumphrey and H. Gleiter, *Phil. Mag.*, Vol 30, p 593, 1974
12. M.F. Ashby, *Phil. Trans. R. Soc. Lond.*, Vol A322, p 393, 1987
13. G.H. Bishop and B. Chalmers, *Scr. Metall.*, Vol 79, p 133, 1968 with permission from Elsevier
14. W. Lojkowski, H. Gleiter, and R. Maurer, *Acta Metall.*, Vol 36, p 69, 1988
15. D. Wolf and S. Yip, Ed., *Materials Interfaces. Atomic-Level Structure and Properties*, Chapman and Hall, London, 1992
16. M. Weins, H. Gleiter, and B. Chalmers, *Scripta Metall.*, Vol 4, p 235, 1970
17. A.P. Sutton and V. Vitek, *Philos. Trans. Roy. Soc. Lond.*, Vol A309, p 1, 37, 55, 1983
18. S.J. Pennycook, D.E. Jesson, A.J. McGibbon, and P.D. Nellist, *J. Electron Micros.*, Vol 45, p 36, 1996
19. V. Vitek, *J. de Physique*, Colloque C5, Suppl. au No. 10, 49, C5–115, 1988
20. D. Wolf, *Acta Metall.*, Vol 32, p 245, 1984
21. D. Wolf, *Acta Metall.*, Vol 37, p 1983, 1989
22. V. Paidar, *Acta Metall.*, Vol 35, p 2035, 1987
23. H.J. Fecht, *Acta Metall. Mater.*, Vol 40, p s39, 1992
24. H.J. Fecht and H.Gleiter, *Acta Metall.*, Vol 33, p 557, 1985
25. G. Kimmel and J. Zabicky, Stability, Instability, Metastability and Grain Size in Nanocrystalline Ceramic Oxide Systems, *Solid State Phenomena*, Vol 140, 2008, p 29–36
26. L.C.A. Stoop and J.H. van der Merve, *"Thin Solid Films"*, Vol 94, p 341, 1982
27. M. Kato, *Mater. Sci. Eng.*, Vol A146, p 205, 1991
28. A. Daniszewska, W. Łojkowski, H.J. Fecht, K.J. Kurzydłowski, U. Narkiewicz, G. Salishchev, M.J. Zehetbauer, M. Kulczyk, M. Chmielecka, and D. Kuzmenko, *Metallic Nano-Materials and Nanostructures: Development of Technology Roadmap*, p 345
29. M. Winterer, Nanocrystalline Ceramics,—Synthesis and Structure, *Springer Series in Materials Science*, R. Hull, R.M. Osgood, and J. Parisi, Ed.
30. D. Hesse, S. Senz, P. Werner, and J. Heydenreich, *Interf. Sci.*, Vol 2, p 221, 1994
31. H. Sieber, D. Hesse, and P. Werner, *Phil. Mag.*, Vol A75, p 889, 1997
32. B. Hvolbæk, T.V.W. Janssens, B.S. Clausen, H. Falsig, C.H. Christensen, and J.K. Nørskov, *Nanotoday*, Vol 2 (No. 4), p 14, 2007
33. K. Binder, M. Bowker, J.E. Inglesfield, and P.J. Rous, "Cohesion and Structure of Surfaces", in the series *Cohesion and Structure*, Vol 4, F.R. De Boer and D.G. Pettifor, Ed., North-Holland, 1995
34. A. Miedema and J.O. den Broeder, *Z. Metalk.*, Vol 70, p 14, 1979
35. R. Becker, *Ann. Phys.*, Vol 32, p 128, 1938
36. J.E. McDonnald and J.G. Eberhard, *Trans. Metall. Soc. AIME*, Vol 233, p 512, 1965

Models for Martensitic Transformations

G.B. Olson, Northwestern University
A. Saxena, Los Alamos National Laboratory

MODERN NUMERICAL METHODS, paced by rapidly advancing computational power, have opened the way to a far more realistic theoretical treatment of the complexity of materials behavior. The resulting enhanced appreciation of natural complexity has fostered renewed interest in the systems view of materials proposed by the late C.S. Smith (Ref 1), who recognized the inevitable hierarchical nature of materials structure, with strong interactions amongst length scales and an interplay between perfection and imperfection involving duality, whereby structure can be equivalently regarded as a hierarchy of three-dimensional space-filling domains of relative perfection or a hierarchy of lower-dimensional defects/imperfections that bound them. Smith warned of a cultural bias toward perfection and emphasized the importance of defect hierarchy in the willful control of materials behavior. His view also acknowledged an intrinsic dynamic nature, whereby a spectrum of relaxation times associated with structural hierarchy assures an intrinsic path-dependent nonequilibrium structure of real materials. This inherent complexity sets limits on the degree of predictability of materials behavior from first-principles approaches. The late Morris Cohen (Ref 2) eloquently argued the importance of reciprocity between the deductive cause/effect logic of reductionist analysis and the inductive goal/means logic of systems synthesis, whereby useful insights such as structure/property relations can be distilled from natural complexity by probing structure from the viewpoint of specific desired properties.

This philosophy has formed the basis for a systems approach to computational materials design that has been successfully applied to martensitic steels over the past decade and is now extended to all classes of materials (Ref 3). The pioneering application in martensites stems from an early scientific appreciation of complexity in martensitic systems. Current design research aimed at acceleration of the full materials development cycle emphasizes:

- Integration of microstructural evolution in process models supporting a design-for-processability approach anticipating process scaleup

- A probabilistic science quantifying the role of distributed defect hierarchy, especially in structure/property relations, for predictive control of variation to accelerate materials qualification

This article assesses the evolution of martensite modeling in the context of this changing materials engineering environment.

Consonant with this view, Fig. 1 depicts the spatiotemporal hierarchy of first-order displacive transformations. Blocks at the left represent interactive levels of structure, both preceding and generated by the transformation, while blocks at the right denote a sequence of dynamic phenomena interacting with these structural levels. As denoted by the interconnections, initial nucleation is sensitive to interfacial defects generated by the processing history, determining the initial number density, N_i, of heterogeneous nucleation sites and

the critical driving force at which they operate. The latter determines the dynamics of unit growth events and influences the degree of dislocation substructure generation determining subunit morphology (plate/lath) and the initial average particle volume, $\overline{V}_o$. The character of growth in turn establishes the elastic or plastic fields governing the number density, N_A, and potency of autocatalytic nucleation sites, which evolve with phase fraction, f. The character of autocatalytic nucleation in turn establishes the initial geometry of variant pairs influencing the cooperative multivariant growth of higher-level groups (packet/sheaf) determining the evolution of average particle volume, $\overline{V}(f)$. These are also influenced by the transgranular spread of transformation in polycrystals, as well as other macroscopic constraints of component shape and external boundary conditions. On the finest scale denoted, solutes and other fine-scale defects exert an influence on several levels,
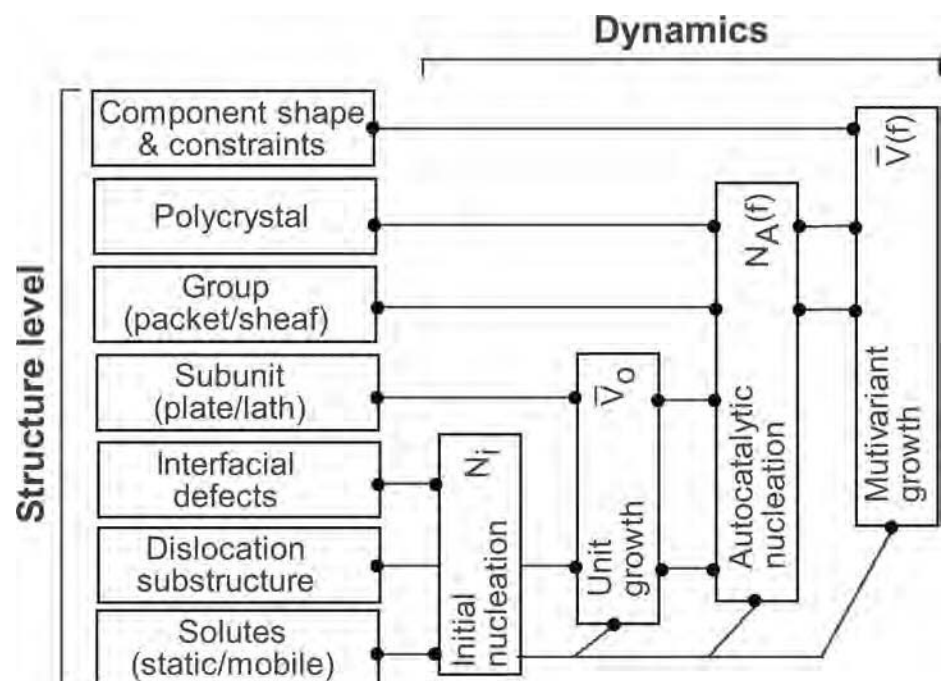

Fig. 1 Flow-block diagram of martensitic transformation as a multilevel dynamic system. See text for explanation of details.

primarily through interfacial friction. Progress in martensite theory has addressed each of these aspects, with due attention to both hierarchies of space-filling domains and space-dividing defects.

An overall assessment of the current state of computational materials science is that simulations have demonstrated remarkable success in reproducing many aspects of observed natural complexity. In the spirit of Cohen's reciprocity, the challenge for the near future is to purposefully exploit the capabilities of these techniques (allowing parametric studies not accessible by experiment) to yield new insights offering predictive control of specific desired behaviors.

Physics of Displacive Transformations

Ginzburg-Landau Theory. A general continuum framework for the representation of martensitic microstructures is now well established. To describe the thermodynamics of the phase transformation and the phase diagram, a free energy model of the transformation is needed. A Landau free energy (LFE) is a symmetry-allowed polynomial expansion in the order parameter that characterizes the transformation (Ref 4), for example, strain tensor components and/or (intraunit cell) shuffle modes, as represented schematically in Fig. 2. A minimization of this LFE with the order parameter components leads to conditions that give the phase diagram. Derivatives of the LFE with respect to temperature, pressure, and other relevant thermodynamic variables provide information about the specific heat, entropy, susceptibility, and so on. To study domain walls between different orientational variants (i.e., twin boundaries) or different shuffle states (i.e., antiphase boundaries), symmetry-allowed strain gradient terms or shuffle gradient terms must be added to the LFE. These gradient terms are called Ginzburg terms, and the augmented free energy is referred to as the Ginzburg-Landau free energy (GLFE). Variation of the GLFE with respect to the order parameter components leads to (Euler-Lagrange) equations (Ref 5, 6) whose solution leads to microstructure.

In two dimensions, the symmetry-adapted dilatation (area change), deviatoric, and shear strains (Ref 5–7), respectively, are defined as a function of the Lagrangian strain-tensor components, ε:

$$e_1 = \frac{1}{\sqrt{2}}(\varepsilon_{xx} + \varepsilon_{yy}), e_2 = \frac{1}{\sqrt{2}}(\varepsilon_{xx} - \varepsilon_{yy}), e_3 = \varepsilon_{xy}$$

(Eq 1)

As an example, for a square-to-rectangle transition, the LFE (Ref 8, 9) is given by:

$$F(e_2) = \frac{A}{2}e_2^2 + \frac{B}{4}e_2^4 + \frac{C}{6}e_2^6 + \frac{A_1}{2}e_1^2 + \frac{A_3}{2}e_3^2$$

(Eq 2)

where A_1, A, and A_3 are the bulk, deviatoric, and shear modulus, respectively, and B and C are higher-order elastic constants. The sixth-order term makes it a first-order transition. This free energy without the nonorder parameter strain (e_1, e_3) terms below a critical temperature (T_c) has two minima in e_2 corresponding to the two rectangular variants (of martensite). Above T_0, it has only one global minimum at $e_2 = 0$ associated with the stable square lattice (i.e., austenite; Fig. 2b). Because the shear modulus softens (partially) above T_0, $A = A_0(T - T_0)$.

As another example, the LFE for a triangular-to- (centered) rectangular transition is given by (Ref 8, 9):
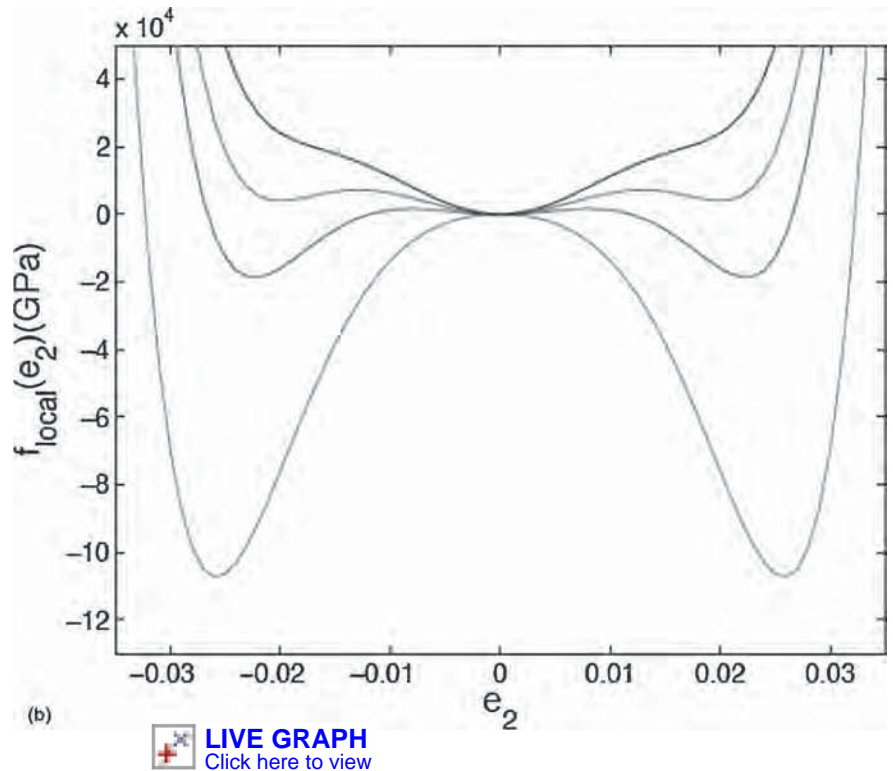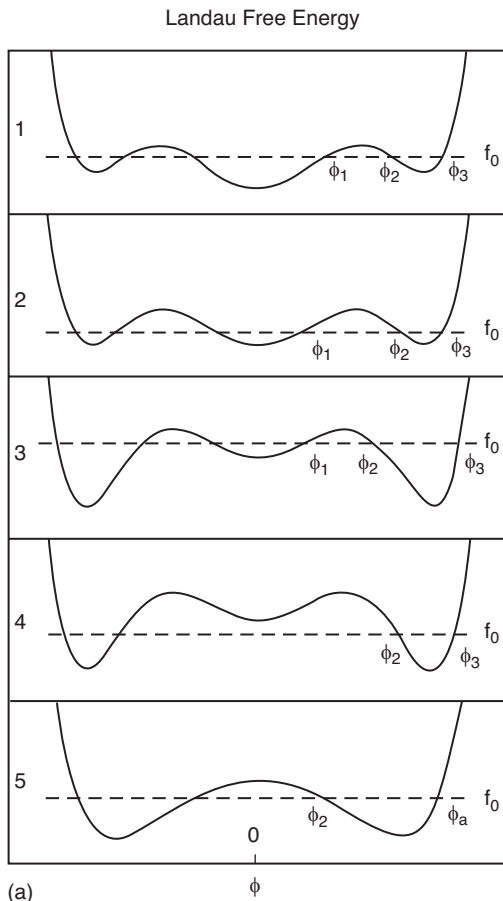
Landau Free Energy

Fig. 2 Schematic Landau free energy with three minima as a function of temperature. (a) Scaled free energy (f_0) vs scaled strain (ϕ) for different temperatures. (b) Eq 2 as a function of temperature for representative parameters.

$$F(e_2, e_3) = \frac{A}{2}(e_2^2 + e_3^2) + \frac{B}{3}(e_2^3 - 3e_2e_3^2)$$
$$+ \frac{C}{4}(e_2^2 + e_3^2)^2 + \frac{A_1}{2}e_1^2 \qquad \text{(Eq 3)}$$

where A is the shear modulus, A_1 is the bulk modulus, and B and C are third- and fourth-order elastic constants, respectively. The third-order term makes it a first-order transition. This free energy without the nonorder parameter strain (e_1) term below T_c has three minima in (e_2, e_3) corresponding to the three rectangular variants. Above T_c, it has only one global minimum at $e_2 = e_3 = 0$ associated with the stable triangular lattice. Again, above T_0, $A = A_0(T - T_c)$.

In three dimensions, symmetry-adapted strains are defined as (Ref 5):

$$e_1 = \frac{1}{\sqrt{3}}(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}), \quad e_2 = \frac{1}{\sqrt{2}}(\varepsilon_{xx} - \varepsilon_{yy}),$$
$$e_3 = \frac{1}{\sqrt{6}}(\varepsilon_{xx} + \varepsilon_{yy} - 2\varepsilon_{zz}),$$
$$e_4 = \varepsilon_{xy}, e_5 = \varepsilon_{yz}, e_6 = \varepsilon_{xz} \qquad \text{(Eq4)}$$

As an example, the Landau part of the elastic free energy for a cubic-to-tetragonal transition in terms of the symmetry-adapted strain components is given by (Ref 5, 6, 10, 11):

$$F(e_2, e_3) = \frac{A}{2}(e_2^2 + e_3^2) + \frac{B}{3}(e_2^3 - 3e_2e_3^2)$$
$$+ \frac{C}{4}(e_2^2 + e_3^2)^2 + \frac{A_1}{2}e_1^2$$
$$+ \frac{A_4}{2}(e_4^2 + e_5^2 + e_6^2) \qquad \text{(Eq 5)}$$

where A_1, A, and A_4 are bulk, deviatoric, and shear modulus, respectively, B and C denote third- and fourth-order elastic constants, and (e_2, e_3) are the order parameter deviatoric strain components. The nonorder parameter dilatation (e_1) and shear (e_4, e_5, e_6) strains are included to harmonic order. For studying domain walls (i.e., twinning) and microstructure, this free energy must be augmented (Ref 10) by symmetry-allowed gradients of (e_2, e_3). The plot of the free energy in Eq 5 without the nonorder parameter strain contributions (i.e., compression and shear terms) is identical to the two-dimensional triangle-to-rectangular case (Eq 3), except that the three minima in this case are associated with the three tetragonal correspondence variants.

While the crystal symmetry dictates the form of GLFE, the coefficients in the free energy are determined from a combination of experimental structural (lattice parameter variation as a function of temperature or pressure), vibrational (e.g., phonon dispersion curves along different high-symmetry directions from neutron scattering), and thermodynamic data (entropy, specific heat, elastic constants, etc.). Where sufficient experimental data are not available, electronic structure calculations and molecular dynamics simulations (using appropriate atomistic potentials) can provide the relevant information to determine some or all of the coefficients in the GLFE.

For simple phase transitions, for example, two-dimensional square-to-rectangle (Ref 6) in Eq 2 or those involving only one component

order parameter (Ref 12), the GLFE can be written down by inspection (from the symmetry of the parent phase). However, in general, the GLFE must be determined by group theoretic means that are now readily available for all 230 crystallographic space groups in three dimensions and (by projection) for all 17 space groups in two dimensions (Ref 12); see the computer program ISOTROPY by Stokes and Hatch (Ref 13).

Microstructure Representation. There are several different but related ways of modeling the microstructure in structural phase transformations:

- GLFE-based models, as described previously (Ref 5)
- Phase-field models in which strain variables are coupled in a symmetry-allowed manner to the morphological variables (Ref 14)
- Sharp interface models used by applied mathematicians (Ref 15, 16)

The natural order parameters in the GLFE are strain-tensor components. However, until recent years, researchers have simulated the microstructure in displacement variables by rewriting the free energy in displacement variables (Ref 8, 11). This procedure leads to the microstructure without providing direct physical insight into the evolution. A natural way to bring out the insight is to work in strain variables only. However, if the lattice integrity is maintained during the phase transformation, that is, no lattice dislocation (or topological defect) generation is allowed, then one must obey the St. Venant elastic compatibility constraints, because various strain-tensor components are derived from the displacement field and are not all independent. This can be achieved by minimizing the free energy with compatibility constraints treated with Lagrangian multipliers (Ref 7, 9). This procedure leads to an anisotropic long-range interaction between the order parameter strain components. The interaction (or compatibility potential) provides direct insight into the domain wall orientations and various aspects of the microstructure in general.

Mathematically, the elastic compatibility condition on the geometrically linear strain tensor, ε, is given by (Ref 17–19):

$$\vec{\nabla} \times (\vec{\nabla} \times \varepsilon) = 0 \qquad \text{(Eq 6)}$$

which is one equation in two dimensions connecting the three components of the symmetric strain tensor: $\varepsilon_{xx,yy} + \varepsilon_{yy,xx} = 2\varepsilon_{xy,xy}$. In three dimensions, it is two sets of three equations, each connecting the six components of the symmetric strain tensor ($\varepsilon_{yy,zz} + \varepsilon_{zz,yy} = 2\varepsilon_{yz,yz}$ and two permutations of x, y, z; $\varepsilon_{xx,yz} + \varepsilon_{yz,xx} = \varepsilon_{xy,xz} + \varepsilon_{xz,xy}$ and two permutations of x, y, z). For periodic boundary conditions in Fourier space, it becomes an algebraic equation that is then easy to incorporate as a constraint.

For the free energy in Eq 2, the Euler-Lagrange variation of $[F - \Lambda G]$ with respect to the nonorder parameter strains, e_1, e_3 is then (Ref 9, 12) $\delta(F^c -$

$\Sigma_r \Lambda G)/\delta \ e_i = 0$, where $i = 1,3$, and G denotes the constraint equation (Eq 6), Λ is a Lagrangian multiplier, and $F^c = (A_1/2)e_1^2 + (A_3/2)e_3^2$ is identically equal to: $\sum_k F^c(k)$.

The variation allows (in k-space, assuming periodic boundary conditions) nonorder parameter strains to be obtained in terms of $\Lambda(k)$.

$$e_1(k) = \frac{(k_x^2 - k_y^2)k^2}{k^4 + \left(8\frac{A_1}{A_3}\right)(k_xk_y)^2}, e_2(k),$$
$$e_3(k) = -\left(\frac{A_1}{A_3}\right)\frac{\sqrt{8}k_xk_y(k_x^2 - k_y^2)}{k^4 + \left(8\frac{A_1}{A_3}\right)(k_xk_y)^2}e_2(k) \qquad \text{(Eq 7)}$$

Then, $e_1(k)$, $e_3(k)$ are put back into the compatibility constraint condition (Eq 6) to solve for the Lagrange multiplier, $\Lambda(k)$. Thus, $e_1(k)$ and $e_3(k)$ are expressed in terms of $e_2(k)$.

$$U(\hat{k}) = \frac{(k_x^2 - k_y^2)^2}{k^4 + \left(8\frac{A_1}{A_3}\right)(k_xk_y)^2}, e_3(k)$$
$$= -\left(\frac{A_1}{A_3}\right)\frac{\sqrt{8}k_xk_y(k_x^2 - k_y^2)}{k^4 + \left(8\frac{A_1}{A_3}\right)(k_xk_y)^2} \qquad \text{(Eq 8)}$$

and

$$F^c(k) = \frac{A_1}{2}\sum_k U(\hat{k})|e_2(k)|^2$$
$$U(\hat{k}) = \frac{(k_x^2 - k_y^2)^2}{k^4 + \left(8\frac{A_1}{A_3}\right)(k_xk_y)^2} \qquad \text{(Eq 9)}$$

which is used in a (static) free energy variation of the order parameter strains. The (static) compatibility kernel, $U(\hat{k})$, is independent of $\left|\vec{k}\right|$ and therefore only orientationally dependent: $U(\vec{k}) \rightarrow U(\hat{k})$. In coordinate space, this is an anisotropic long-range ($\sim 1/r^2$) potential mediating the elastic interactions of the primary order parameter strain. From this compatibility kernel, one can obtain domain wall orientations, parent-product interface (i.e., habit plane) orientations, and local rotations (Ref 12) consistent with those obtained previously using macroscopic matching conditions and symmetry considerations (Ref 20, 21).

The concept of elastic compatibility in a single crystal can be readily generalized to polycrystals by defining the strain-tensor components in a global frame of reference (Ref 22, 23). By adding a stress term (bilinear in strain) to the free energy, one can compute the stress-strain constitutive response in the presence of microstructure for both single and polycrystals and compare the recoverable strain upon cycling. The grain rotation and grain boundaries play an important role when polycrystals are subject to external stress in the presence of a structural transition. Similarly, the calculation of the constitutive response can be generalized to improper ferroelastic materials, such as those driven by shuffle modes (intracell displacements or lattice-point relative displacements).

Dynamics and Simulations. The overdamped (or relaxational) dynamics can be used in simulations to obtain equilibrium microstructure:

$$\dot{e} = \frac{-1}{A'} \frac{\delta(F + F^c)}{\delta e}$$

where A' is a friction coefficient, and F^c is the long-range contribution to the free energy due to elastic compatibility. However, if the evolution of an initial nonequilibrium structure to the equilibrium state is important, one can use inertial strain dynamics with appropriate dissipation terms included in the free energy. The strain dynamics for the order parameter strain-tensor component, e_2, are given by (Ref 9):

$$\rho_0 \ddot{e}_2 = \frac{c^2}{4} \vec{\nabla}^2 \left(\frac{\delta(F + F^c)}{\delta e_2} + \frac{\delta(R + R^c)}{\delta \dot{e}_2} \right) \quad \text{(Eq 10)}$$

where ρ_0 is a scaled mass density, c is a symmetry-specific constant, $R = (A'/2)\dot{e}_2^2$ is the Rayleigh dissipation, R^c is a contribution to the dissipation due to the long-range elastic interaction, and $\ddot{\varepsilon}$ is the second time derivative of strain.

The compressional and shear free energy in Eq 2 is replaced with the corresponding long-range elastic energy in the order parameter strain and includes a gradient term, $F_G = (K/2)[(\nabla e_2)^2]$, where the gradient coefficient K determines the elastic domain wall energy and can be estimated from phonon dispersion curves. Simulations performed with the full underdamped dynamics for the square-to-rectangular transition are depicted in Fig. 3. The equilibrium microstructure is essentially the same as that found from the overdamped dynamics.

The three-dimensional cubic-to-tetragonal transition (free energy in Eq 5) can be simulated by either using the strain-based formalism outlined here (Ref 10) or directly using the displacements (Ref 11). Figure 4 depicts microstructure evolution for the cubic-to-tetragonal transition in FePd mimicked by a square-to-rectangle transition. To simulate mechanical loading of a polycrystal (Ref 22, 23), an external tensile stress, σ, is applied quasi-statically, that is, starting from the unstressed configuration. The applied stress, σ, is increased in steps of 5.13 MPa, after allowing the configurations to relax for t = 25 time steps after each increment. The loading is continued until a maximum stress of σ = 200 MPa is reached. Thereafter, the system is unloaded by decreasing σ to zero at the same rate at which it was loaded. The favored (rectangular) variants grow at the expense of the unfavored (differently oriented rectangular) variants. The orientation distribution does not change much. As the stress level is increased further, the favored variants grow. Even at the maximum stress of 200 MPa, some unfavored variants persist.

Note that the grains with large misorientation with the loading direction rotate. Grains with lower misorientation do not undergo significant rotation. The mechanism of this rotation is the tendency of the system to maximize the transformation strain in the direction of loading so that the total free energy is minimized (Ref 22, 23). Within the grains that rotate, subgrain bands are present that correspond to the unfavored strain variants that still survive. Upon removing the load, a domain structure is nucleated again, due to the local strains at the grain boundaries and the surviving unfavored variants in the loaded polycrystal configuration. This domain structure is not the same as that prior to loading, and thus, there is an underlying hysteresis. The unloaded configuration has nonzero average strain. This average strain is recovered by heating to the austenite phase, in accordance with the shape memory effect. The orientation distribution reverts to its preloading state because the grains rotate back when the load is removed.

The aforementioned mechanical behavior of the polycrystal in Fig. 4 is compared to the corresponding single crystal. The recoverable strain for the polycrystal is smaller than that for the single crystal, due to nucleation of domains at grain boundaries upon unloading. In addition, the transformation in the stress-strain curve for the polycrystal is not abrupt, because the response of the polycrystal is averaged over all grain orientations. Similarly, using the free energy given by Eq 3 for the triangular-to-rectangular transition, one can simulate the microstructure (Ref 9, 24) that has been observed in lead orthovanadate crystals (Ref 25, 26).

Comparison with Other Methods. The approach based on the work of Barsch and Krumhansl (Ref 5, 6) is compared with two other methods that make use of Landau theory to model structural transformations. Here, a brief outline of the differences is provided (except for models with dislocation (Ref 27)), and the methods are compared and reviewed in detail in Ref 28. Khatchachuryan and coworkers (Ref 14, 29, 30) used a free energy in which a structural or morphological order
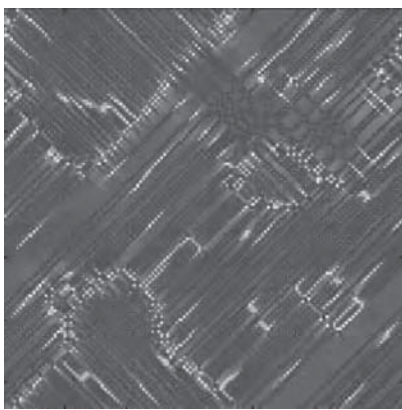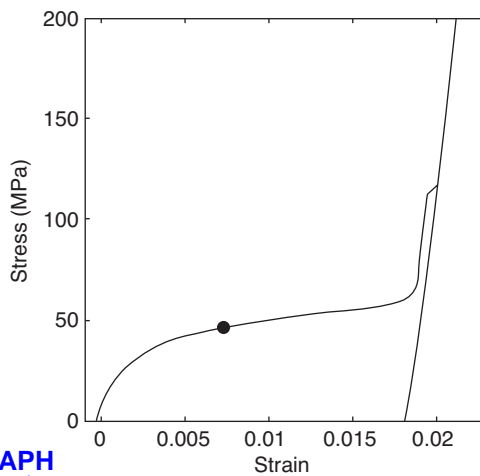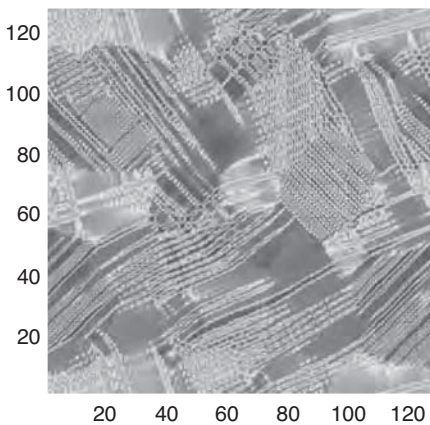

LIVE GRAPH
Click here to view

Fig. 3 Simulated microstructure and constitutive response for a single crystal at a certain applied stress below the transition temperature for the square-to-rectangle transition (FePd parameters). Source: Ref 24
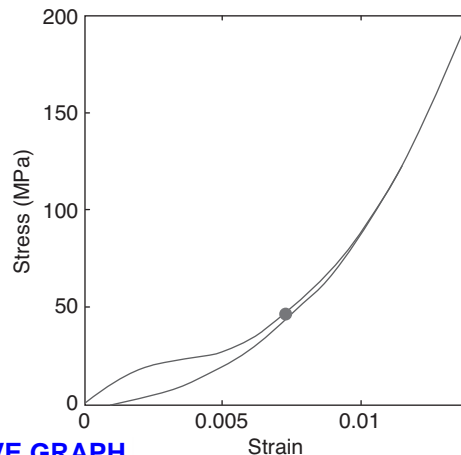

LIVE GRAPH
Click here to view

Fig. 4 Simulated microstructure and constitutive response for a polycrystal at a certain applied stress for the square-to-rectangle transition (FePd parameters). Source: Ref 23

parameter, η, is coupled to strains. This order parameter is akin to a shuffle order parameter (Ref 31), and the inhomogeneous strain contribution is evaluated using the method of Eshelby (Ref 14). The strains are then effectively removed in favor of the η's, and the minimization is carried out for these variables. This approach (sometimes referred to as phase field) applied to improper ferroelastics is essentially the same as the approach outlined in this article, with minor differences in the way the inhomogeneous strain contribution is evaluated. However, for the proper ferroelastics that are driven by strain, rather than shuffle, essentially the same procedure is used with phase field; that is, the minimization (through relaxation methods) is ultimately for the η's rather than the strains. In the approach outlined in this article, the nonlinear free energy is written upfront in terms of the relevant strain order parameters, with the discrete symmetry of the transformation taken into account. Here, terms that are gradients in strains, which provide the costs of creating domain walls, are also added according to the symmetries. The free energy is then minimized with respect to the strains. That the microstructure for proper ferroelastics obtained from either method would appear qualitatively similar is not surprising. Although the free energy minima or equilibrium states are the same from either procedure, differences in the details of the free energy landscape would be expected to exist. These could affect, for example, the microstructure associated with metastable states.

The method outlined in this article and that developed by the applied mechanics community (Ref 15, 16) share the common feature of minimizing a free energy written in terms of strains. The method is ideally suited for laminate microstructures with domain walls that are atomistically sharp. This sharp interface limit means that transformation strains are incorporated through the use of the Hadamard jump condition (Ref 15, 16). The method takes into account finite deformation and has served as an optimization procedure for obtaining static equilibrium structures, given certain volume fractions of variants. The approach outlined in this article differs in that it uses a continuum formulation with interfaces that have finite width, and therefore, the transformation strains are taken into account through the compatibility relation (Ref 7, 9). In addition, the full evolution equations are solved so that kinetics and the effects of inertia can be studied.

Density Functional Theory (DFT). Many of the martensitic phase transformations are driven by a shuffle mechanism that is reflected in the softening of a phonon mode in the phonon dispersion curves (energy versus wave vector). Inelastic neutron scattering provides a direct means of measuring the dispersion curves along various high-symmetry directions in the unit cell (i.e., Brillouin zone in the reciprocal or wave vector space). A useful way of obtaining these curves is by means of ab initio calculations within the DFT using standard electronic

structure packages, such as Vienna ab initio Simulation Package (VASP), and by invoking either local-density approximation (LDA) for the electron density or the more accurate generalized gradient approximation (GGA). The latter is implemented with a specific choice of basis, for example, projector-augmented wave basis. Lattice dynamics is carried out for the different (austenite and martensite) crystal structures by first constructing the dynamical matrix. Its diagonalization allows one to find the phonon dispersion relations.

The DFT is a ground-state theory that essentially depends on the total number of electrons in the unit cell. In addition to calculating electronic structure, lattice parameters, bonding properties (e.g., bulk modulus), elastic constants, crystal structures, atomic positions, and equation of state, the DFT allows one to study the energy pathway associated with the strain (distortion)-shuffle transformation and the corresponding energy landscape as well as the (asymmetric) energy density barrier between the austenite and martensite. Recent calculations for $B2 \rightarrow B19$ and $B19'$ transformations have provided fundamental parameters for TiNi and related alloys (Ref 31).

The Ginzburg-Landau free energy functional can be fit to the first-principles energy data to obtain the coefficients of the Landau free energy (Ref 32).

Shuffle Transitions. Previously, the proper ferroelastic transitions were considered. This method can be readily extended (including the Ginzburg-Landau free energy and elastic compatibility) to the study of improper ferroelastics (e.g., shuffle-driven transitions such as $B2 \rightarrow B19'$ in TiNi) (Ref 33). However, now the elastic energy is considered only up to the harmonic order, whereas the primary order parameter has anharmonic contributions. For example, for a two-dimensional shuffle-driven transition on a square lattice, the Ginzburg-Landau free energy is given by (Ref 31, 34):

$$
\begin{aligned}
F(\vec{Q}) = \; & \alpha_1 \left(Q_x^2 + Q_y^2 \right) + \alpha_{11} \left(Q_x^4 + Q_y^4 \right) \\
& + \alpha_{12} Q_x^2 Q_y^2 + \alpha_{111} \left(Q_x^6 + Q_y^6 \right) \\
& + \alpha_{112} \left(Q_x^2 Q_y^4 + Q_x^4 Q_y^2 \right) \\
& + \frac{g_1}{2} \left(Q_{x,x}^2 + Q_{y,y}^2 \right) + \frac{g_2}{2} \left(Q_{x,y}^2 + Q_{y,x}^2 \right) \\
& + g_3 Q_{x,x} Q_{y,y} + \frac{1}{2} A_1 e_1^2 + \frac{1}{2} A_2 e_2^2 + \frac{1}{2} A_3 e_3^2 \\
& + \beta_1 e_1 \left(Q_x^2 + Q_y^2 \right) + \beta_2 e_2 \left(Q_x^2 - Q_y^2 \right) \\
& + \beta_3 e_3 Q_x Q_y
\end{aligned}
$$

(Eq 11)

where Q_x and Q_y are the shuffle components.

Martensitic Nucleation

The application of the Ginzburg-Landau approach to rigorous solutions for the structure of a martensitic nucleus has clarified

longstanding issues in the theory of martensitic nucleation. Olson and Roitburd (Ref 35) reviewed the full range of possible theoretical mechanisms of martensitic nucleation set forth in the pioneering work of Morris Cohen. A central issue has been whether nucleation operates in the "classical" limit, where a critical nucleus has the full strain amplitude corresponding to fully formed martensite, or exhibits "nonclassical" behavior with significantly reduced strain amplitude, resembling a local lattice instability. As first described by Olson and Cohen (Ref 36), the Ginzburg-Landau theoretical framework allows a rigorous description of the competition between these modes of nucleation, when adapted to the two- and three-dimensional variational solutions describing critical nuclei. Using the case of a simple coherent shear transformation, Reid et al. (Ref 27, 36–38) applied such Ginzburg-Landau descriptions to obtain rigorous two-dimensional variational solutions for heterogeneous nucleation at linear defects, using an element-free Galerkin numerical method. For the case of a group of dislocations of sufficient potency to account for nucleation at the martensite start (M_s) temperature, Fig. 5 (a) shows the computed strain field (right) when the material is described as linear elastic (left). The defect provides a significant volume of highly distorted material. Representing conditions far above the T_o equilibrium temperature, Fig. 5(b) shows the computed field (right) under conditions corresponding to the first appearance of a second energy well (left), representing the martensite structure. The variational solution shows that such a defect field becomes unstable with respect to the formation of a locally stabilized classical (fully formed) martensite embryo well above the transformation temperature. This solution constitutes rigorous confirmation of the "pre-existing embryo" hypothesis of Kaufman and Cohen in 1958 (Ref 39). Operational nucleation at the M_s temperature then corresponds to the growth startup of such classical embryos with kinetic control by processes of interfacial motion.

A central concept of quantitative martensite kinetic theory (Ref 40) is a potency distribution of nucleation sites, as originally proposed by Machlin and Cohen (Ref 41). For low-potency heterogeneous nucleation, Fig. 6 (Ref 38) shows the computed fields (right) at temperatures below T_o (left) for the case of a single lattice dislocation, which would contribute to nucleation at temperatures well below M_s. For the smaller distorted volume in this class of defect, gradient energy suppresses the formation of a local classical embryo, so that nucleation abruptly transforms the crystal in Fig. 6 (c) without a pre-existing classical embryo. The defect field in this case resembles the "strain embryo" concept of the nonclassical reaction path model first proposed by Cohen et al. in 1949 (Ref 42). The variational solutions of Fig. 6 confirm that the strain amplitude of the critical nucleus in this weak defect limit shows some departure from the classical limit.
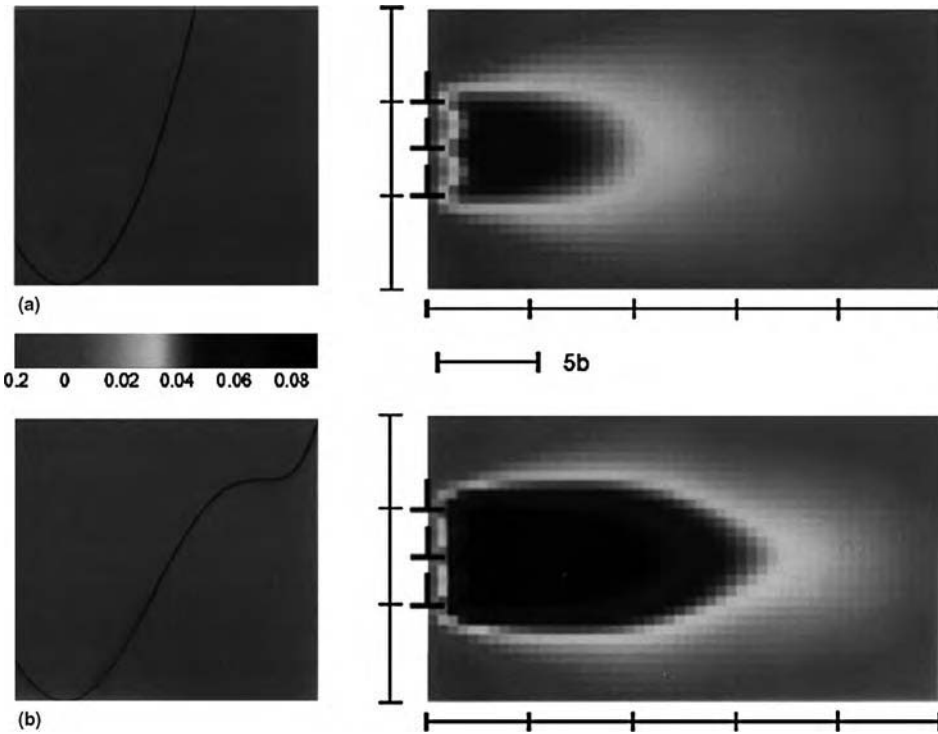
0.2　0　0.02　0.04　0.06　0.08

5b

Fig. 5 Heterogeneous nucleation at strong defects. Two-dimensional variational solutions to Ginzburg-Landau model for (a) linear elastic material and (b) nonlinear material at onset of martensite mechanical stability. Length scale in units of lattice dislocation Burgers vector, b. Source: Ref 27

Stronger departures from classical behavior can be expected in the limit of homogeneous martensitic nucleation, which requires high driving forces approaching the critical driving force for lattice instability. The data points in Fig. 7 represent the measured critical driving force for homogeneous face-centered cubic (fcc) → body-centered cubic (bcc) martensitic nucleation in defect-free 10 nm iron-cobalt particles coherently precipitated in a copper matrix, as studied by Lin et al. (Ref 43). The driving force is normalized to the critical driving force for fcc lattice instability determined from DFT calculations. The solid curve shows the theoretical prediction of a three-dimensional model of a single-domain, fully coherent bcc nucleus based on a rigorous one-dimensional Ginzburg-Landau description, using the gradient energy coefficient as a fitting parameter (Ref 44). Significant departure of the critical nucleus strain amplitude from the classical limit is predicted. With the model parameters fixed, the upper dashed curve shows the predicted critical driving force for nucleation at full strain amplitude. The model predicts that in this limit of homogeneous coherent fcc → bcc nucleation, nonclassical effects reduce the critical driving force for nucleation by 15 to 20%.

These examples of the incorporation of rigorous nonlinear physics in two- and three-dimensional variational solutions of critical nucleus structure provide important guidelines for appropriate approximations in the engineering application of martensite kinetic theory. For the case of finely dispersed metastable particles, transformation controlled by weak defect or homogeneous nucleation at high driving forces can require significant nonclassical corrections. For bulk polycrystals, where transformation is dominated by strong defects at relatively low driving forces, classical heterogeneous nucleation theory (Ref 45) can be applied with high confidence.

Nucleation in the low-driving-force classical limit is predicted to involve relatively sharp interfaces, as supported by direct high-resolution electron microscopy observations of interfacial core structure (Ref 46) in this regime. Sharp interface models are therefore appropriate. This includes discrete dislocation models, whose application in martensitic transformation modeling has been comprehensively reviewed by Olson and Cohen (Ref 47). The earliest quantitative solution to the heterogeneous martensitic nucleation problem is based on this approach, as depicted in Fig. 8 (Ref 48). Equivalent to the case described in Fig. 5, Fig. 8(a) represents an interfacial defect as a group of "extrinsic" interfacial dislocations. A martensitic embryo formed at such a defect is represented by an array of discrete partial transformation or coherency dislocations (Fig. 8b), which account for the transformation strain while maintaining interfacial coherency, and an array of lattice dislocations (Fig. 8c), accounting for the lattice-invariant deformation of a semicoherent embryo. The energetics of formation of such an embryo can equivalently be described as the dissociation of the defect of

Fig. 8(a) to form the coherency dislocations of Fig. 8(b) or the elastic interaction of a semicoherent embryo of Fig. 8(b and c) with the defect of Fig. 8(a). Both descriptions predict barrier-less nucleation, where an embryo with a thickness of n close-packed planes of interplanar spacing d is derived from the defect interaction. The value of n scales with the resolved Burgers vector content of the defect and thus defines the defect potency. Consideration of an interfacial work, w_f, for the glide of a martensitic interface in a solid solution gives a critical free energy change (driving force) per volume for barrier-less nucleation at a defect of potency, n, as (Ref 48):

$$\Delta g_{\mathrm{crit}}(n) = -\left[\frac{2\gamma}{nd} + g_o + w_f\right] \qquad \text{(Eq 12)}$$

where γ is the semicoherent interfacial energy per area. The term g_o was originally interpreted as a shape-independent elastic energy associated with distortion in the interface at small embryo size. Detailed comparison of nucleation in thermoelastic and nonthermoelastic transformation in the iron-platinum system (Ref 49) has led to a reinterpretation of this quantity as a dislocation forest hardening term for autocatalytic nucleation in the plastic zone associated with the nonthermoelastic growth mode described in the section by that name that follows.

Based on direct measurements of interfacial mobility in thermoelastic single-interface transformations, the theory of thermally activated dislocation mobility has been adapted to martensitic interfaces (Ref 50, 51) to predict the form of the composition and temperature dependence of the w_f interfacial frictional work term. Finding a characteristic potency, n, for the defects controlling the M_s of typical polycrystals, a kinetic database has been established for the solution-hardening coefficients of alloying elements in steels (Ref 52), predicting both athermal and thermal contributions to w_f, the latter contribution describing observed isothermal nucleation behavior. Compared to traditional empirical linear regression models for the composition dependence of M_s, the combination of the nonlinear thermodynamics and kinetic w_f behavior of Eq 12 has provided a significant improvement in the accuracy of M_s prediction in the design of complex high-alloy steels.

Martensitic Growth

Thermoelastic Growth. Addressing the second unit process of Fig. 1, Wang and Khachaturyan (Ref 53) employed the time-dependent Ginzburg-Landau equation, incorporating a linear-viscous dynamic response, in the three-dimensional dynamic simulation of a cubic-tetragonal martensitic transformation with internal shuffles (improper ferroelastic), giving the sequence of structures depicted in Fig. 9. Starting from a supercritical embryo of arbitrary shape, the detailed simulation predicts the dynamic growth form of an internally twinned plate,
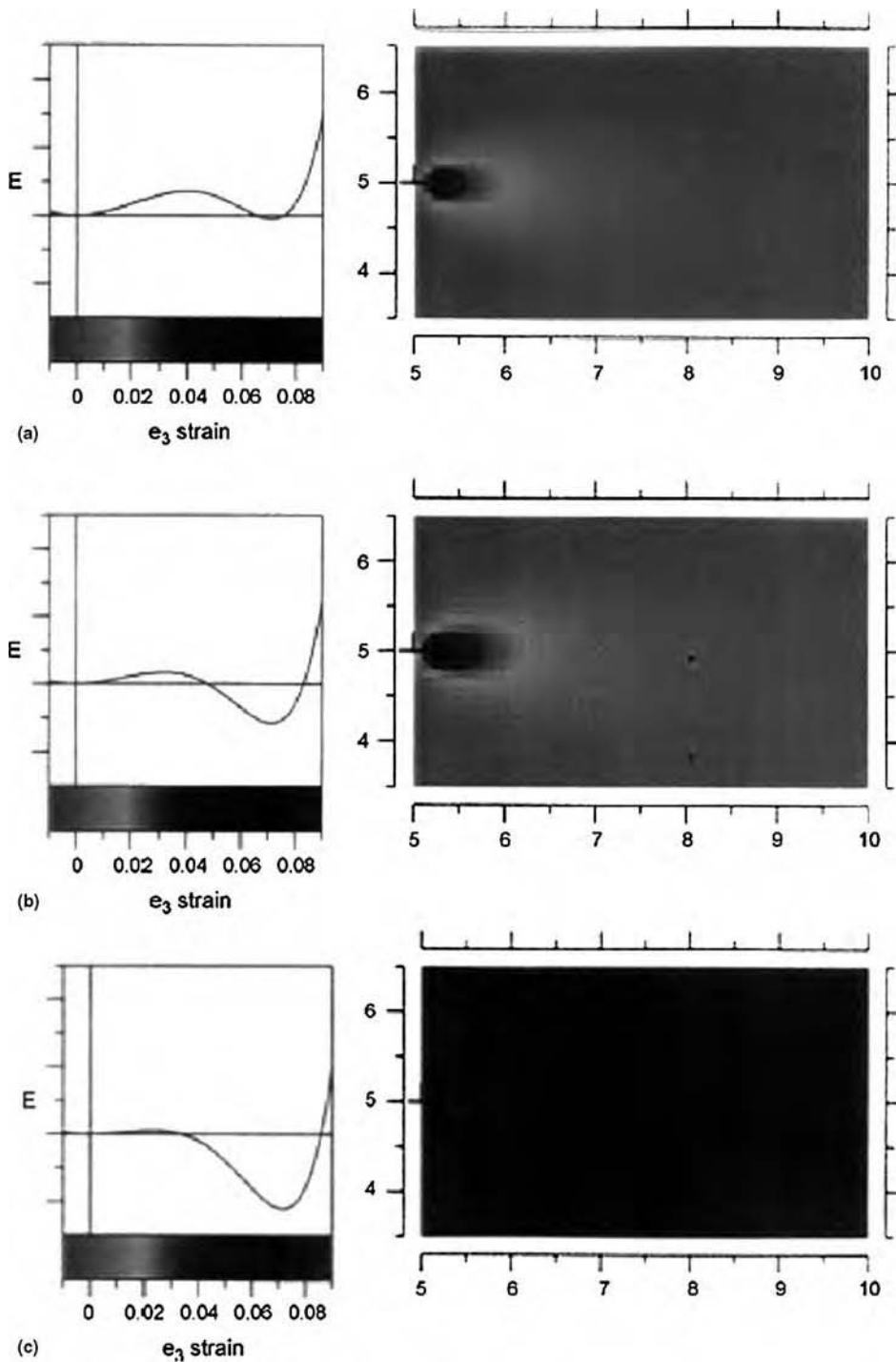
(a)

(b)

(c)

Fig. 6 Heterogeneous nucleation at weak defects. Two-dimensional variational solutions to Ginzburg-Landau model for single lattice dislocation with increasing driving force (a and b), leading to transformed crystal at critical driving force for nucleation in (c). Length scale in units of lattice dislocation Burgers vector, b. Source: Ref 38

LIVE GRAPH
Click here to view

Fig. 7 Normalized critical driving force for homogeneous coherent face-centered cubic → body-centered cubic nucleation in defect-free iron-cobalt particles, comparing three-dimensional Ginzburg-Landau variational solution for nucleus (solid curve) with experimental data points. Horizontal dashed line denotes elastic energy threshold for coherent nucleation. Upper dashed curve represents classical nucleus of full strain amplitude. Source: Ref 44

which is strikingly similar to the oblate spheroidal, semicoherent embryo assumed by Kaufman and Cohen (Ref 39) and Knapp and Dehlinger (Ref 54) and predicted by the growth dynamic model of Raghavan and Cohen (Ref 55).

Nonthermoelastic Growth. The model assumptions of the unit growth simulation of Fig. 9 correspond to a thermoelastic system.

A model of unit growth dynamics in nonthermoelastic systems demonstrated the important role of (size- and rate-dependent) plastic accommodation of the transformation shape strain in governing the technologically important lath/plate morphological transition in steels (Ref 47, 56). The complex interactions that occur during nonthermoelastic growth have

been modeled by Haezebrouck (Ref 56) using simulations of dynamic growth events in FeNi alloys. The interactions considered are summarized in Fig. 10. As in the thermoelastic case, the driving force for interfacial motion arises from the chemical free energy change (a), with restraining forces arising from interfacial energy (b) and elastic strain energy (c). Differentiating the total energy with respect to particle radius, r, and semithickness, c, these terms all influence the net driving force per unit area acting on the interface (d) for radial growth and thickening. Through interfacial mobility relations (e), these net forces determine the interfacial velocities in the r and c directions. For a first-order phase transformation with a latent heat, the interfacial velocity influences the heat transfer across the interface and thus affects the temperature at the interface (f). This in turn affects the interfacial mobility (e), modifies the chemical free energy change (a) at the interface, and weakly affects the interfacial energy (b). The overall interaction thus far discussed would also apply to the initial radial growth events in thermoelastic transformations.

In a nonthermoelastic transformation, the interfacial velocity also determines the plastic strain rate (g) imposed on the parent phase. This influences the extent of plastic accommodation (h), which is also dependent on particle size through the conditions for slip nucleation in the particle stress field. Plastic accommodation modifies the elastic strain energy (c) and introduces two important dissipations: the work
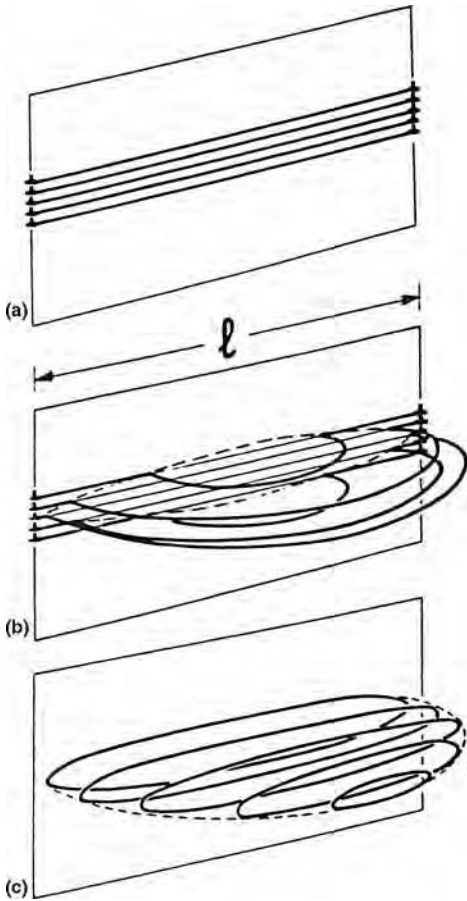
Fig. 8 Martensitic nucleation by dislocation dissociation. (a) Nucleating defect. (b) Dissociation of defect to produce $a/18$ [112] partial dislocations. (c) Simultaneous generation of lattice dislocations. Source: Ref 48
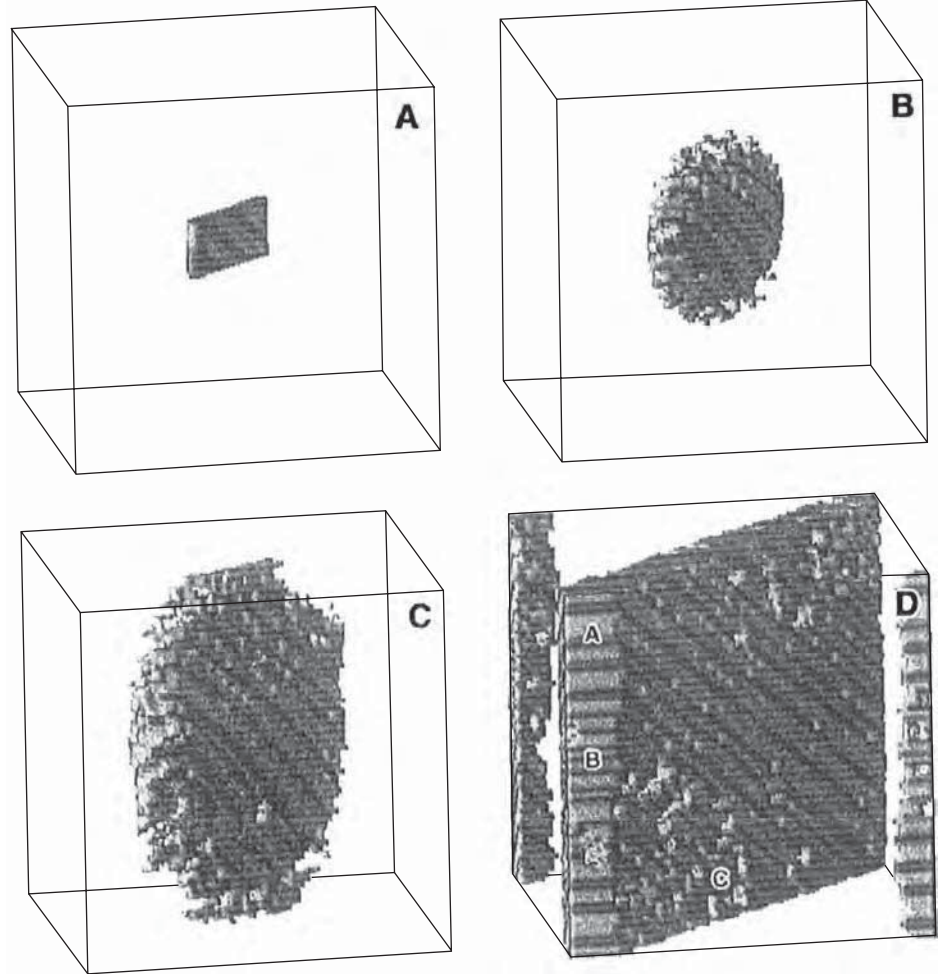

Fig. 9 Simulation of growth of supercritical semicoherent (internally twinned) martensitic embryo in nonlinear, nonlocal continuum model. Source: Ref 53

of creating the plastic zone (i) and the additional interface frictional work (j) for moving the interface through strain-hardened material. These dissipations modify the net interfacial driving forces (d).

As discussed in the previous section, interfacial mobility parameters for thermally activated behavior in ferrous alloys can be estimated from nucleation kinetic data; phonon drag parameters are estimated from studies of slip dislocation mobility. Steady-state heat transfer during radial growth of a martensitic particle has been analyzed by Lyubov and Roitburd (Ref 57), and the transient behavior during thickening has been treated by Nishiyama et al. (Ref 58). From the form of the behavior obtained for the latter case, the interface temperature at time t can be taken to be a function of $\overline{v^2}$, where $\overline{v^2}$ is the time-averaged mean-square velocity. Using available alloy thermodynamic parameters and elastic constant data, all the necessary parameters can then be defined to describe the uppermost loop of Fig. 10.

The mechanics of plastic accommodation can be treated using available slip kinetic data for the alloys of interest and adapting concepts from fracture mechanics developed for crack-tip plasticity. Using the plastic zone solution for a mode III shear crack (Ref 59), the radial plastic zone size, R_p, is estimated as:

$$R_p = r[1 + 2.5(\overline{\tau}_f/\tau_i)]^{-2} \qquad (Eq\ 13)$$

where $\overline{\tau}_f$ is the matrix effective shear strength, and τ_i is the internal shear stress in the particle, as defined by the Eshelby elastic particle calculation (Ref 60). Using the same model for the form of the plastic distribution, the average plastic strain rate imposed in the plastic zone by a radial interfacial velocity, v_r, is expressed by:

$$\dot{\gamma} = 2v_r\overline{\tau}_f/\mu R_p \qquad (Eq\ 14)$$

From the form of calculated elastic stress fields in the c-direction (Ref 61), the plastic-zone size normal to the particle habit is estimated as:

$$C_p = \frac{1}{2}r(1 - \overline{\tau}_f/\tau_i) \qquad (Eq\ 15)$$

Both the dependence of the local plastic strain rate on v_c and the total plastic displacement across the zone can be estimated by assuming a constant plastic strain gradient in the zone and fitting to experimental displacement fields based on detailed analysis of surface-relief effects from isolated martensite plates in ferrous alloys. Modification of the invariant plane-strain elastic strain energy (and the associated particle internal stress parameter, τ_i) is taken into account by a radial plastic-zone size correction to the effective particle radius (Ref 62) and by an effective transformation shear strain that incorporates accumulated plastic displacements averaged over the particle volume. Integration of the plastic strain fields defines the work of plastic-zone formation. The primary influence of strain hardening is assumed to lie in the athermal resistance to interfacial motion, and this is taken to vary linearly with the average plastic strain in each zone (r and c directions treated separately). The magnitude of the strain-hardening effect, within the range of theoretical possibilities, is estimated from the thermodynamic conditions for which radial growth is observed to cease in thermal-gradient and composition-gradient experiments (Ref 63).
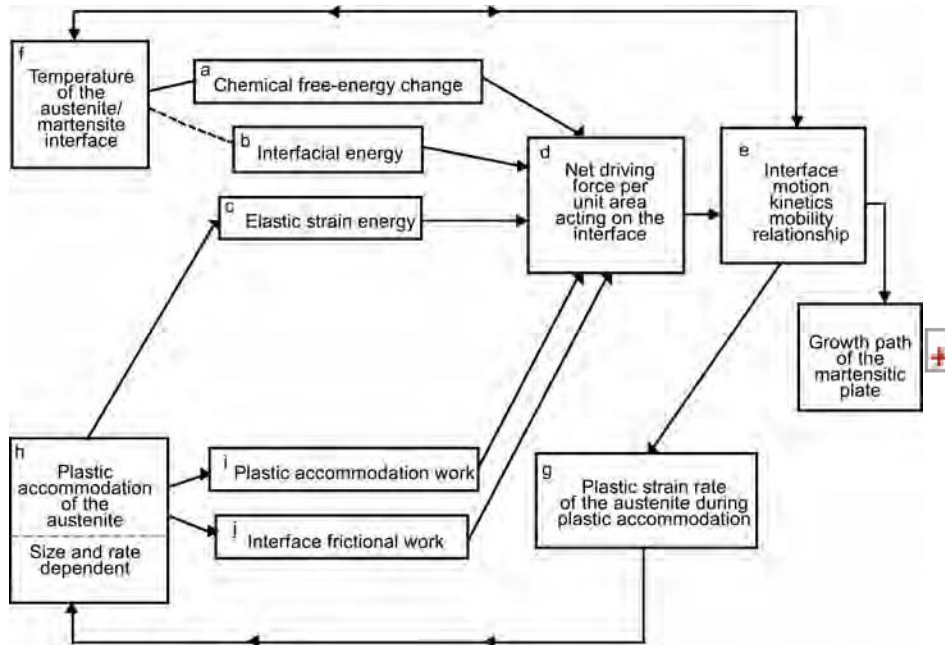
Fig. 10 Dynamic interactions during nonthermoelastic growth of a martensitic particle. Source: Ref 56
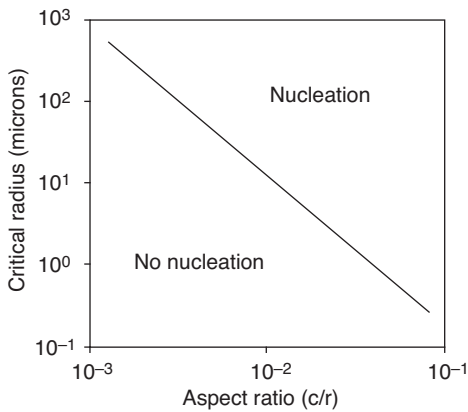

Fig. 11 Critical condition for barrier-less nucleation of slip dislocations in stress field of plate periphery. Source: Ref 56

LIVE GRAPH
Click here to view

Nucleation of slip dislocations in the particle stress field is regarded as a necessary condition for plastic-zone formation. A linear elastic calculation of the conditions for barrier-less nucleation of slip dislocations in the stress field at the plate periphery gives a critical plate radius as a function of plate aspect ratio, plotted in Fig. 11. For an aspect ratio of a few percent, a plate radius of a few micrometers is required to nucleate slip. When this nucleation condition is met, a radial plastic-zone size can be calculated by expressing the matrix strength, $\bar{\tau}_f$, as a function of $\dot{\gamma}$. The latter function is denoted by the schematic solid curve in Fig. 12. Here, region I ($\tau_\mu < \bar{\tau}_f < \hat{\tau}$) corresponds to thermally activated behavior, linear region II represents phonon drag control, and region III is a
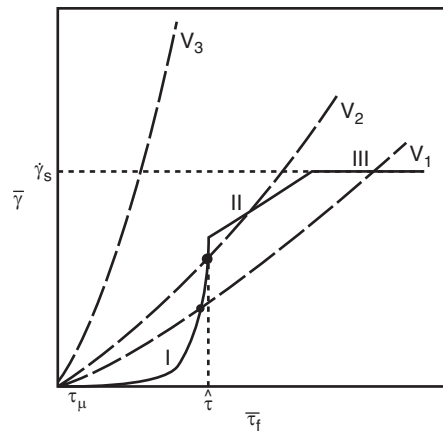
Fig. 12 Conditions for radial plastic-zone formation. At a given radial velocity, dashed curve representing imposed plastic strain rate defined by Eq 13 and 14 must intersect solid flow-stress curve. Source: Ref 57

relativistic limit where $\dot{\gamma}$ is not allowed to exceed $\dot{\gamma}_s$, corresponding to a dislocation velocity equal to the elastic shear wave velocity. The dashed curves indicate the imposed strain rate in the plastic zone as a function of $\bar{\tau}_f$, depicting three different radial interface velocities. For the velocities v_1 and v_2, physically meaningful plastic-zone solutions correspond to the curve intersections shown by the heavy points. At the higher velocity, v_3, no solution exists, and a radial plastic zone cannot form. Similar behavior is obtained for determination of the plastic-zone size in the c-direction.

Results of simulations of growth events at the M_s temperature for two FeNi alloys are

LIVE GRAPH
Click here to view
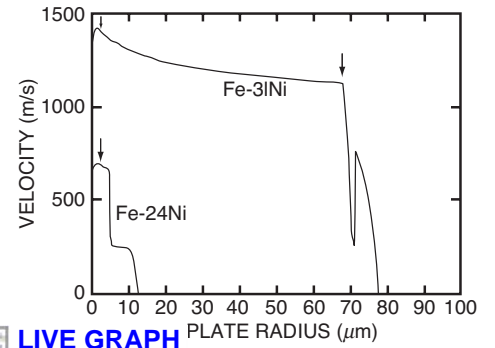
Fig. 13 Summary of simulation results for growth events at martensite start temperatures of Fe-31Ni and Fe-24Ni alloys. Source: Ref 57

summarized by plots of radial velocity versus plate radius in Fig. 13. For the Fe-31Ni case, radial growth rapidly accelerates to a velocity near one-third the shear wave velocity. The condition for slip nucleation is met at the point indicated by the first arrow, but the interface velocity is sufficiently high that a plastic zone cannot form. As the plate grows, its increasing tip radius makes heat transfer less effective, and the interface temperature increases, causing a gradual deceleration. At the point marked by the second arrow, a critical velocity is reached where plastic-zone formation is possible. Abrupt formation of a plastic zone causes a strong deceleration and erratic growth behavior, leading to complete cessation of radial growth at a plate radius of ≈80 μm. This is then followed by an additional increment of plate thickening.

For the case of the Fe-24Ni alloy, which forms lath martensite, a lower peak velocity is achieved during growth, due primarily to increased phonon drag at its higher M_s temperature. Soon after the slip nucleation condition is achieved, as indicated by the arrow, extensive plastic accommodation brings growth to a halt. In contrast to the Fe-31Ni growth involving a macroscopic elastic event, the major part of the growth occurs in a plastic state and stops at a size that is smaller by an order of magnitude.

The behavior described here identifies three basic modes of martensitic growth:

- *Elastic:* When significant plastic accommodation is absent, the behavior described in Fig. 9 occurs, corresponding to thermoelastic plate martensite.
- *Elastic/plastic:* When plastic accommodation is delayed by a high initial growth rate, a macroscopic elastically accommodated "midrib" forms, followed by thickening and growth cessation in an irreversible plastic mode. This corresponds to nonthermoelastic plate martensite (as in Fe-31Ni, Fig. 13).
- *Fully plastic:* When plastic accommodation is not significantly delayed beyond the slip nucleation condition, the entire growth event can be considered to take place in a plastic condition. This corresponds to the lath martensite of Fe-24Ni.

Although the approximations employed in the simulations of Fig. 13 are based on small-scale yielding and cannot accurately describe the highly plastic behavior of the Fe-24Ni alloy, the model does predict a substantial transition in growth mode that correlates with the lath/plate morphological transition and provides a likely physical basis for it. While radial symmetry was imposed in these calculations, anisotropic plastic-zone interactions could provide some tendency toward a lath shape during fully plastic growth. The macroscopic elastic growth events common to the thermoelastic and nonthermoelastic plate martensite would be consistent with the success of habit plane analyses for these cases, based on elastic strain energy minimization alone. The more complex deformations possible during fully plastic growth may favor interfacial energy minimization, promoting habits near close-packed planes, as observed for lath martensites.

Working within a mesoscopic continuum nonequilibrium thermodynamic framework (Ref 64), Levitas (Ref 65) has undertaken more rigorous finite-element simulations of unit growth in a plastic matrix. Contours of computed accumulated plastic strain for directional growth of a martensite lath are represented in Fig. 14. Calculating the incremental work of elemental growth steps offers insights into the evolution of morphology in a plastic environment. The behavior is found to be strongly path dependent, and non-local plasticity models could be employed to more realistically treat size-dependent plastic flow at the 0.1 μm length scales of interest.

Overall Kinetics

While the basic characteristics of individual nucleation and growth events determine many

essential features of martensitic transformations, the overall kinetics of the reaction also depend on the nature of the interactions between martensitic particles. As represented in Fig. 3 and 4, the strain-energy reducing rearrangements allowed by particle translation will play an important role in the course of thermoelastic transformations. For nonthermoelastic behavior, where particles are frozen into the configuration in which they originally form, the nature of autocatalytic nucleation plays a more important role in microstructural development.

In addition to particle-shape-related differences in the character of the stress fields of martensitic laths and plates, the macroscopic elastic growth event predicted for the plate martensite of Fig. 13 suggests different types of autocatalysis. With plastic-zone formation and autocatalytic nucleation as competitive modes of stress-field accommodation, the intense elastic event can provide an autocatalytic nucleation site of high potency, in line with the "bursting" form of autocatalysis encountered in plate martensites. Similar to Hall-Petch strengthening, austenitic grain refinement can reduce the intensity of these elastic stress fields and inhibit the grain-to-grain spread of autocatalytic nucleation. The effect is manifested in a grain size dependence of the martensite bursting temperature (Ref 66) and an influence of grain size on the relative roles of the grain-to-grain spreading and intragranular filling-in modes of autocatalysis in controlling microstructure development (Ref 67, 68). The latter effects are important to the overall transformation kinetics through the dependence of average plate volume, $\bar{V}$, on the extent of transformation (Ref 69). As represented by the example in Fig. 15 (Ref 40), the behavior measured during the filling-in mode can be modeled as a

simple power-law function of the phase fraction, f:

$$\bar{V} = \bar{V}_1 f^{-m} \qquad \text{(Eq 16)}$$

where m is an exponent near 1. As denoted by the horizontal dashed lines in Fig. 15, the behavior during the initial spreading mode of transformation is represented by truncation of Eq 16 for f below a characteristic level. While this behavior has been well established by stereological measurements in nonthermoelastic ferrous alloys, surface-relief observations have more recently demonstrated that it also applies to thermoelastic TiNi alloys (Ref 70).

When the dependence of $\bar{V}$ on the transformed volume fraction, f, was established, the overall kinetics of a nucleation-controlled martensitic transformation was modeled by Pati and Cohen (Ref 71) by a rate equation of the form:

$$\dot{f} = (N_i + pf - N_v)(1-f)V\,v\exp\left(-\frac{Q}{RT}\right). \qquad \text{(Eq 17)}$$

Here, N_i is the density of the most-potent initial nucleation sites; p defines the density of autocatalytically generated sites; N_v is the number of plates per unit volume, accounting for sites that have operated; the $(1-f)$ factor accounts for potential sites swept up by the transformation; V is the instantaneous mean plate volume obtained by differentiation of Eq 16; while v and Q are an attempt frequency and activation energy defined by the thermal component of the interfacial frictional work, w_f, in Eq 12. This model has been successfully fit to experimental transformation curves of f versus time, t. Because N_v is equal to $f/\bar{V}$, the shape of the $f(t)$ function is sensitive to the product $p\bar{V}$, curving upward when $p\bar{V} > 1$. For $p\bar{V} < 1$, $f(t)$ curves downward, leading to a finite saturation level of transformation at a given reaction temperature.

The model of Eq 17 employs the simplifying approximation of a singly-activated process. The derivation of the activation energy, Q, from Eq 12, employing the thermal contributions to w_f, predicts that Q will be a function of defect potency, n. Small-particle experiments in metals and ceramics show that the pre-existing defects corresponding to N_i in Eq 17 are in fact distributed in potency and described by an exponential cumulative potency distribution (Ref 72):

$$N_i(\geq n) = N_i^o \exp(-\alpha n) \qquad \text{(Eq 18)}$$

with $\alpha \approx 0.84$. Sensitive acoustic emission measurements detecting earliest nucleation events have further prescribed the dependence of the total amplitude, N_i^o, on grain size, D (Ref 73). It is expected that the autocatalytic nucleation described by the factor p in Eq 17 should also be distributed in potency, n. In this case, its mechanistic origin in growth events suggests a function distributed about a characteristic mode.
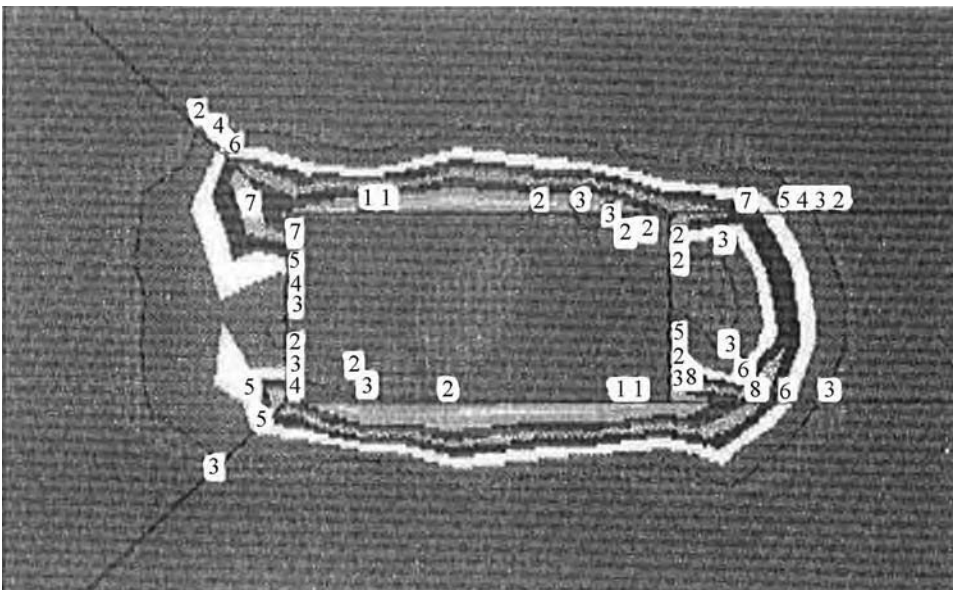

Fig. 14 Computed plastic strain contours for growth of lath martensite in a ferrous alloy. Source: Ref 65
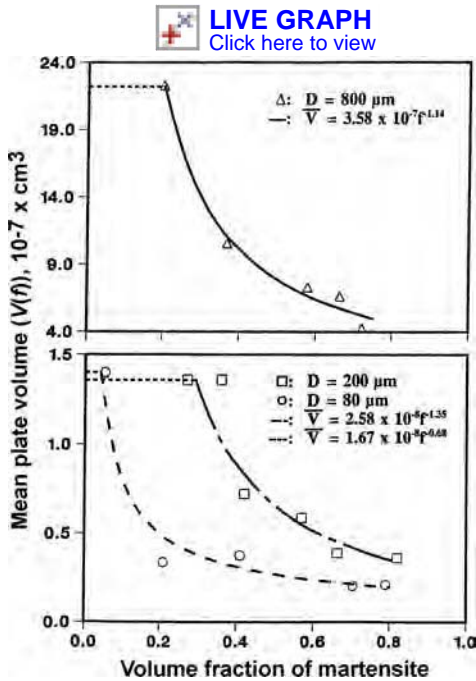
Fig. 15 Measured mean volume of martensitic plates versus volume fraction of martensite in an Fe-32.3Ni alloy with austenitic grain size D = 80, 200, and 800 μm
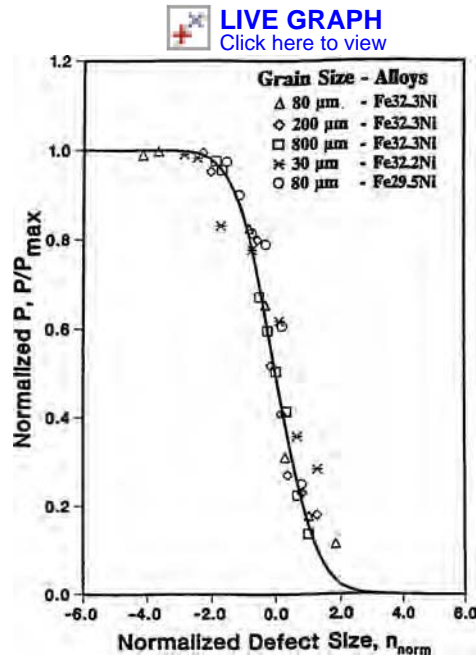
Fig. 16 Normalized cumulative form of the autocatalytic-defect distribution (P/P_{max}) versus normalized defect size $((n - \bar{n})/\sigma)$

Fig. 17 Calculated transformation temperatures and initial nucleation probabilities as a function of grain size for Fe-32.3Ni alloy. The solid circles represent the very first plate-formation temperatures from high-sensitivity acoustic-emission measurements of Ref 70, and the triangular points are experimental data for 20% martensite.

Based on these concepts, Lin et al. (Ref 40) developed a comprehensive kinetic model for martensitic transformations, allowing the Q-values associated with N_i and p to be distributed in a manner predicted by potency distributions with respect to n. Equation 17 is then generalized by an integration of the form:

$$\frac{dN_v}{dt} = \int_o^Q \left[\left(\frac{dN_i}{dQ} + f \frac{dP}{dQ} - \frac{dN_v}{dQ} \right) (1-f)v \exp\left(-\frac{Q}{kT}\right) \right] dQ \qquad \text{(Eq 19)}$$

where dN_i/dQ is the activation-energy distribution of pre-existing defects, dP/dQ is the activation-energy distribution of autocatalytic defects, and dN_v/dQ is the plate-number-density distribution as transformed from defects having an activation energy, Q. With the measured parameters of Eq 18 as input, together with the measured $\bar{V}(f)$ behavior of Fig. 15, the evolution of f with temperature and time defines a cumulative autocatalytic $p(n)$ function well described by a Gaussian distribution, as shown in Fig. 16.

With the athermal and thermal components of w_f in Eq 12, this model describes the full range of athermal and isothermal behaviors exhibited by ferrous martensites defining the evolution of f, N_v, and $\bar{V}$ with both temperature and driving force. As summarized by Fig. 17, an opposite grain-size dependence of initial and autocatalytic potency distributions also accounts for a transition from smooth to burst character with increasing grain size.

More recent application to ferrous lath martensites has demonstrated applicability of the same functional forms of $\bar{V}(f)$ and $p(n)$ (Ref 74). Numerical implementation as the CryoMART simulator (Ref 75) has applied the model in optimization of cryogenic treatments for minimization of retained austenite in high-performance steels. Surface-relief measurements have shown that the same Gaussian $p(n)$ form allows the model to accurately describe the evolution of transformation in thermoelastic TiNi alloys (Ref 37, 70).

Conclusions

Rigorous application of continuum Ginzburg-Landau models informed by atomistic-energy calculations has answered key questions in the mechanism and kinetics of martensitic transformations. This in turn has informed appropriate approximations for the treatment of behaviors of higher complexity, such as morphological transitions associated with nonthermoelastic growth, and the quantitative modeling of overall transformation kinetics in both thermoelastic and nonthermoelastic systems. The opportunity now exists to employ large-scale Ginzburg-Landau simulations to predict the $\bar{V}(f)$ evolution and $p(n)$ autocatalytic potency distributions to compare with the corresponding functions derived from experiment.

Current models and simulations have already provided significant support for computational materials design and process optimization. Improved model fidelity will enhance these capabilities and support higher levels of optimization in integrated materials and product design.

The models presented here provide the framework for the prediction of behaviors under stress, including nonthermoelastic transformation plasticity, thermoelast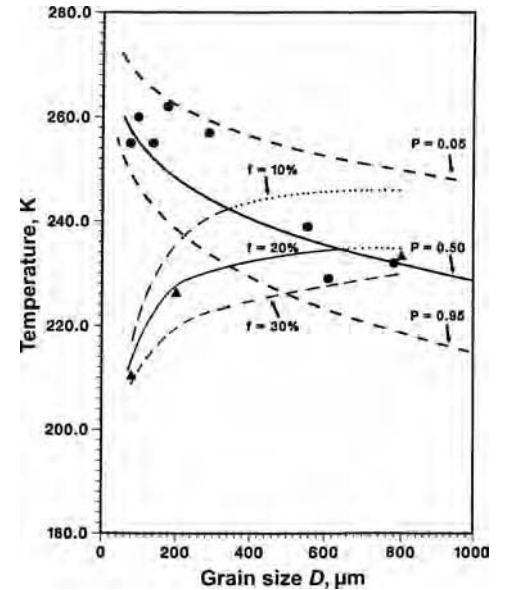ic superelasticity, and shape memory behavior. The nonthermoelastic applications are reviewed in detail elsewhere (Ref 76–78). Rigorous application to thermoelastic systems is a current opportunity.

REFERENCES

1. C.S. Smith, *A Search for Structure,* MIT Press, Cambridge, MA, 1981
2. M. Cohen, *Mater. Sci. Eng.,* Vol 25, 1976, p 3
3. G.B. Olson, *Science,* Vol 277 (5330), 1997, p 1237
4. J.C. Tolédano and P. Tolédano, *The Landau Theory of Phase Transitions,* World Scientific, Singapore, 1987
5. G.R. Barsch and J.A. Krumhansl, *Phys. Rev. Lett.,* Vol 53, 1984, p 1069

6. G.R. Barsch and J.A. Krumhansl, *Metall. Trans. A,* Vol 18, 1988, p 761

7. S.R. Shenoy, T. Lookman, A. Saxena, and A.R. Bishop, *Phys. Rev. B,* Vol 60, 1999, p R12537

8. S.H. Curnoe and A.E. Jacobs, *Phys. Rev. B,* Vol 63, 2001, p 094110

9. T. Lookman, S.R. Shenoy, K.O. Rasmussen, A. Saxena, and A.R. Bishop, *Phys. Rev. B,* Vol 67, 2003, p 024114

10. K.O. Rasmussen, T. Lookman, A. Saxena, A.R. Bishop, R.C. Albers, and S.R. Shenoy, *Phys. Rev. Lett.,* Vol 87, 2001, p 055704

11. A.E. Jacobs, S.H. Curnoe, and R.C. Desai, *Phys. Rev. B,* Vol 68, 2003, p 224104

12. D.M. Hatch, T. Lookman, A. Saxena, and S.R. Shenoy, *Phys. Rev. B,* Vol 68, 2003, p 104105

13. H.T. Stokes and D.M. Hatch, *Isotropy Subgroups of the 230 Crystallographic Space Groups,* World Scientific, Singapore, 1988

14. A.G. Khachaturyan, *Theory of Structural Transformations in Solids,* Wiley, New York, 1983

15. J.M. Ball and R.D. James, *Arch. Rational Mech. Anal.,* Vol 100, 1987, p 13

16. R.D. James and K.F. Hane, *Acta Mater.,* Vol 48, 2000, p 197

17. S.F. Borg, *Fundamentals of Engineering Elasticity,* World Scientific, Singapore, 1990

18. M. Baus and R. Lovett, Phys. Rev. Lett., Vol 65, 1990, p 1781

19. M. Baus and R. Lovett, Phys. Rev. A, Vol 44, 1991, p 1211

20. J. Sapriel, *Phys. Rev. B,* Vol 12, 1975, p 5128

21. C. Boulesteix, B. Yangui, M. Ben Salem, C. Manolikas, and S. Amelinckx, *J. Phys.,* Vol 47, 1986, p 461

22. R. Ahluwalia, T. Lookman, and A. Saxena, *Phys. Rev. Lett.,* Vol 91, 2003, p 055501

23. R. Ahluwalia, T. Lookman, A. Saxena, and R.C. Albers, *Acta Mater.,* Vol 52, 2004, p 209

24. R. Ahluwalia, T. Lookman, A. Saxena, and S.R. Shenoy, *Phase Transitions,* Vol 77, 2004, p 457

25. C. Manolikas and S. Amelinckx, *Phys. Status Solidi (a),* Vol 60, 1980, p 607

26. C. Manolikas and S. Amelinckx, *Phys. Status Solidi (a),* Vol 61, 1980, p 179

27. A.C.E. Reid, G.B. Olson, and B. Moran, *Phase Transitions,* Vol 69, 1999, p 309

28. T. Lookman, S.R. Shenoy, and A. Saxena, to be published

29. Y.H. Wen, Y.Z. Wang, and L.Q. Chen, *Philos. Mag. A,* Vol 80, 2000, p 1967

30. H.L. Hu and L.Q. Chen, *Mater. Sci. Eng. A,* Vol 238, 1997, p 182

31. S. Kibey, H. Sehitoglu, and D.D. Johnson, *Acta Mater.,* Vol 57, 2009, p 1624

32. M. Sanati, A. Saxena, T. Lookman, and R.C. Albers, *Phys. Rev. B,* Vol 63, 2001, p 4114

33. G.R. Barsch, *Mater. Sci. Forum,* Vol 327–328, 2000, p 367

34. R. Ahluwalia and W. Cao, *Phys. Rev. B,* Vol 63, 2001, p 012103

35. G.B. Olson and A.L. Roitburd, in *Martensite,* G.B. Olson and W.S. Owen, Ed., ASM International, 1992, p 149

36. G.B. Olson and M. Cohen, *J. Phys.,* Vol 43, 1982, p C4–C75

37. G.B. Olson, *Mater. Sci. Eng. A,* Vol 273–275, 1999, p 11–20

38. A.C.E. Reid and G.B. Olson, *Mater. Sci. Eng. A,* Vol 309–310, 2001, p 370

39. L. Kaufman and M. Cohen, *Prog. Met. Phys.,* Vol 1, 1958, p 165

40. M. Lin, G.B. Olson, and M. Cohen, *Metall. Trans. A,* Vol 23, 1992, p 2987

41. E.S. Machlin and M. Cohen, *Trans. AIME,* Vol 194, 1952, p 489

42. M. Cohen, E.S. Machlin, and V.G. Paranjpe, *Thermodynamics in Physical Metallurgy,* American Society for Metals, 1949, p 242

43. M. Lin, G.B. Olson, and M. Cohen, *Acta. Metall. Mater.,* Vol 41, 1993, p 253

44. Y.A. Chu, B. Moran, and G.B. Olson, *Metall. Mater. Trans. A,* Vol 31, 2000, p 1321

45. G.B. Olson and M. Cohen, *Metall. Trans. A,* Vol 7, 1976, p 1897

46. F.-R. Chen and G.B. Olson, *Proc. XIIth Int. Conf. Electron Microscopy,* 1990

47. G.B. Olson and M. Cohen, Chap. 37, *Dislocations in Solids,* Vol 7, F.R.N. Nabarro, Ed., North-Holland, Amsterdam, 1986, p 295–407

48. G.B. Olson and M. Cohen, *Metall. Trans. A,* Vol 7, 1976, p 1905–1914

49. G.B. Olson and H. Ohtsuka, *Proc. Int. Symp. on Shape Memory Materials,* Int. Acad. Publishers, Beijing, 1994, p 17–23

50. M. Grujicic, G.B. Olson, and W.S. Owen, *Metall. Trans. A,* Vol 16, 1985, p 1713

51. M. Grujicic and G.B. Olson, *Interface Sci.,* Vol 6 (No. 155), 1998, p 155–164

52. G. Ghosh and G.B. Olson, *Acta Metall. Mater.,* Vol 42, 1994, p 3361–3370

53. Y. Wang and A.G. Khachaturyan, *Acta Mater.,* Vol 45, 1997, p 759

54. H. Knapp and U. Dehlinger, *Acta Metall.,* Vol 4, 1956, p 289

55. V. Raghavan and M. Cohen, *Acta Metall.,* Vol 20, l972, p 779

56. D.M. Haezebrouck, "Nucleation and Growth of a Martensitic Particle," Sc.D. thesis, MIT, 1987

57. B.Y. Lyubov and A.L. Roitburd, *Dokl. Akad. Nauk SSSR,* Vol 13l, 1960, p 809

58. Z. Nishiyama, A. Tsubaki, H. Suzuki, and Y. Yamada. *J. Phys. Soc. Jpn.,* Vol 13, 1958, p 1084

59. J.A.H. Hult and F.A. McClintock, in *Proc. Ninth Int. Congress Applied Mechanics,* Vol 8 (Brussels, Belgium), 1956, p 51

60. J.D. Eshelby, *Prog. Solid Mech.,* Vol 2, 1961, p 89

61. H.C. Ling and W.S. Owen, *Acta Metall.,* Vol 29, 1981, p 1721

62. F.A. McClintock and G.R. Irwin, in *Symp. Fracture Toughness Testing and Applications, ASTM Spec. Tech. Pub. 381,* 1965, p 84

63. G.A. Knorovsky, "Autocatalysis of Martensitic Transformations," Sc.D. thesis, MIT, 1977

64. V.I. Levitas, *J. Mech. Phys. Solids,* Vol 45, 1997, p 923–1203

65. V.I. Levitas, A.V. Idesman, G.B. Olson, and E. Stein, *Philos. Mag. A,* Vol 82 (No. 3), 2002, p 429–462

66. M. Umemoto and W.S. Owen, *Metall. Trans.,* Vol 5, 1974, p 2041

67. J.RC. Guimaraes, *Acta Metall.,* Vol 26, 1978, p 1591

68. J.R.C. Guimaraes and J.C. Gomes, *Metall. Trans. A,* Vol 10, 1979, p 109

69. P.-H. Chang, P.G. Winchell, and G.L. Liedl, *Metall. Trans. A,* Vol 14, 1983, p 163

70. A. McGeorge, M.S. thesis, Northwestern University, 1996

71. S.R. Pati and M. Cohen, *Acta Metall.,* Vol 17, 1969, p 189, 1327

72. M. Cohen and G.B. Olson, *New Aspects of Martensitic Transformations*, *Proc. First JIM International Symposium (JIMIS-1)* (Kobe, Japan), *Suppl. Trans. JIM,* Vol 17, 1976, p 93

73. G.B. Olson, K. Tsuzaki, and M. Cohen, *Turnbull Symposium, Phase Transitions in Condensed Systems*, G.S. Cargill, F. Spaepen, and K.N. Tu, Ed., MRS, 1987, p 129

74. G. Ghosh and G.B. Olson, *J. Phys. (France) IV,* Vol 112, 2003, p 139–142

75. G.B. Olson, *Mater. Sci. Eng. A,* Vol 438, 2006, p 48–54

76. G.B. Olson, *J. Phys. (France) V*, Colloque C1, 1996, p C1–407

77. G.B. Olson, Chap. 9, *Deformation, Processing and Structure*, 1982 ASM Materials Science Seminar, G. Krauss, Ed., American Society for Metals, 1984, p 391–424

78. G.B. Olson and M. Cohen, *Proc. Earl R. Parker Symposium on Structure/Property Relationships,* S.D. Antalovich, W.W. Gerberich, and R.O. Ritchie, Ed., TMS-AIME, Warrendale, PA, 1986, p 367

Modeling of Nucleation Processes

Emmanuel Clouet, CEA, DEN, Service de Recherches de Métallurgie Physique, France

NUCLEATION is the onset of a first-order phase transition by which a metastable phase transforms into a more stable one. Such a phase transition occurs when a system initially in equilibrium is destabilized by the change of an external parameter such as temperature or pressure. If the perturbation is small enough, the system does not become unstable but rather stays metastable. In diffusive transformations, the system then evolves through nucleation, the growth and coarsening of a second phase. Such a phase transformation is found in many situations in materials science, such as condensation of liquid droplets from a supersaturated vapor, solidification, precipitation from a supersaturated solid solution, and so on. The initial stage of all these different processes can be well described within the same framework, currently known as the classical nucleation theory.

Since its initial formulation in 1927 by Volmer, Weber, and Farkas (Ref 1, 2) and its modification in 1935 by Becker and Döring (Ref 3), the classical nucleation theory has been a suitable tool to model the nucleation stage in phase transformations. The success of this theory relies on its simplicity and on the few parameters required to predict the nucleation rate, that is, the number of clusters of the new phase appearing per unit of time and volume. It allows rationalizing experimental measurements, predicting the consequences of a change of the control parameters such as temperature or supersaturation, and describing the nucleation stage in mesoscopic modeling of phase transformations.

This article first describes the results obtained by Volmer, Weber, Farkas, Becker, and Döring (Ref 1–3), which constitute the classical nucleation theory. These results are the predictions of the precipitate size distribution, steady-state nucleation rate, and incubation time. This theory describes the nucleating system as a homogeneous phase where heterophase fluctuations occur. Some of these fluctuations reach a large enough size that they can continue to grow and lead to the formation of precipitates. The nucleating system is thus envisioned mainly from a thermodynamic viewpoint. The key controlling parameters are the nucleation driving force and the interface free energy. A kinetic approach, cluster dynamics,

can also be used to describe nucleation. This constitutes the second part of this article. Here, a master equation describes the time evolution of the system, which is modeled as a cluster gas. The key parameters are the cluster condensation and evaporation rates. Both approaches are different in their description of the nucleating system and their needed input parameters. They are nevertheless closely related. Predictions of the classical nucleation theory have actually been derived from the same master equation used by cluster dynamics (Ref 3), and extensions of classical nucleation theory always start from this master equation. In this article, the links as well as the difference between both descriptions are emphasized. Since its initial formulation, the classical nucleation theory has been enriched, mainly by Binder and Stauffer (Ref 4–6), to take into account the fact that clusters other than monomers can migrate and react. It has also been extended to multicomponent systems (Ref 7–12). These generalizations of the initial formalism are presented at the end of the second part.

Thermodynamic Approach

Conditions for Nucleation

Nucleation occurs when a homogeneous phase initially in stable thermal equilibrium is put in a state where it becomes metastable by the variation of a controlling parameter. In the following case, the controlling parameter is the temperature, and the initial system is quenched through a first-order phase transition in a two-phase region. The system then tends to evolve toward a more stable state and to reach its equilibrium. Because the parent phase is not unstable, this transformation cannot proceed through the continuous development of growing infinitesimal perturbations delocalized in the whole phase, that is, by spinodal decomposition (Ref 13, 14). Such perturbations in a metastable state increase the free energy. As a consequence, they can appear because of

thermal fluctuations, but they naturally decay. To reach its equilibrium, the system must overcome an energy barrier to directly form clusters of the new equilibrium phase, a process known as nucleation.

This difference between a metastable and an unstable state, as well as between nucleation and spinodal decomposition, is better understood through the following example. Consider a system corresponding to a binary mixture of two elements, A and B, with a fixed atomic fraction x of B elements. Such a system can be a solid or a liquid solution, for instance. Assume that the free energy per atom, $G(x)$, of this system is known for every composition x and is given by the function plotted in Fig. 1. A two-phase region given by the common tangent construction exists at the considered temperature; the equilibrium state of binary mixtures with an intermediate composition x^0 between x^e and y^e corresponds to a mixture of two phases having the compositions x^e and y^e. A homogeneous system with a composition x^0 will then separate into these two equilibrium phases. The variation of the free energy can be examined if this transformation happens through the development of infinitesimal fluctuations. In that purpose, consider a small perturbation corresponding to a separation of the initially homogeneous system into two phases having the compositions $x^0 + dx_1$ and $x^0 + dx_2$. For the perturbation to be small, assume $|dx_1| \ll 1$ and $|dx_2| \ll 1$. If f_1 is the fraction of phase 1, matter conservation imposes the following relation between both compositions:

$$f_1 dx_1 + (1 - f_1)dx_2 = 0 \qquad \text{(Eq 1)}$$

The free energy variation associated with this unmixing is given by:

$$\begin{aligned} \Delta G &= f_1 G(x^0 + dx_1) + (1 - f_1)G(x^0 + dx_2) \\ &\quad - G(x^0) \\ &= \tfrac{1}{2}\left[f_1 dx_1^2 + (1 - f_1)dx_2^2\right]G''(x^0) + o(dx_1^2) \end{aligned}$$
$$\text{(Eq 2)}$$

The first derivative $G'(x^0)$ of the free energy does not appear in Eq 2 because of the relation in Eq 1. The sign of the free-energy variation is

thus governed by the second derivative $G''(x^0)$ of the free energy. If this second derivative is negative, the initial infinitesimal perturbation decreases the free energy (Fig. 2a). It can therefore develop until the system reaches its two-phase equilibrium state. This is the regime of spinodal decomposition. In Fig. 1, the free-energy second derivative changes its signs in x^s and y^s; all homogeneous systems with a composition between these limits are unstable and evolve spontaneously to equilibrium. On the other hand, if the composition x^0 is higher than the equilibrium composition x^e but smaller than the spinodal limit x^s, the homogeneous binary mixture is metastable. Because the second derivative of the free energy is positive, any infinitesimal perturbation increases the free energy (Fig. 2b) and will therefore decay. To reach its equilibrium state, the system must overcome an energy barrier, and phase separation occurs by nucleation of the new equilibrium phase with the composition y^e.

The Capillary Approximation

In the nucleation regime, the system evolves through the formation of well-defined and localized fluctuations corresponding to clusters of the new equilibrium phase. The formation free energies of these clusters are well described by the capillary approximation. This assumes that two contributions enter this free energy (Fig. 3):

- *Volume contribution*: By forming a cluster of the new phase, the system decreases its free energy. The gain is directly proportional to the volume of the cluster or, equivalently, to the number, n, of atoms forming the cluster. This is the nucleation driving force.
- *Surface contribution*: One needs to create an interface between the parent phase and the cluster of the new phase. This interface has a cost that is proportional to the surface area of the cluster or, equivalently, to $n^{(d-1)/d}$, where d is the dimension of the system.

The following is restricted to the three-dimensional case. The formation free energy of a cluster containing n atoms is then given by:

$$\Delta G_n = n\Delta G^{\mathrm{nuc}} + n^{2/3}A\sigma \qquad \text{(Eq 3)}$$

where ΔG^{nuc} is the nucleation free energy, σ is the interface free energy, and A is a geometric factor. If the interface free energy is isotropic, the equilibrium shape of the cluster is a sphere. The corresponding geometric factor is then $A = (36\pi\Omega_1^2)^{1/3}$, where Ω_1 is the volume of a monomer. For anisotropic interface free energy, one can use the Wulff construction (Ref 15, 16) to determine the equilibrium shape, that is, the shape with minimum free energy for a given volume, and deduce an average interface free energy corresponding to a hypothetical spherical cluster having the same volume and the same interface energy as the real one, which may be facetted. An example is given in Ref 17 for precipitates with {100}, {110}, and {111} interfaces.

The nucleation free energy is obtained by considering the difference of chemical potentials in the parent and in the equilibrium phases for all atoms composing the cluster:
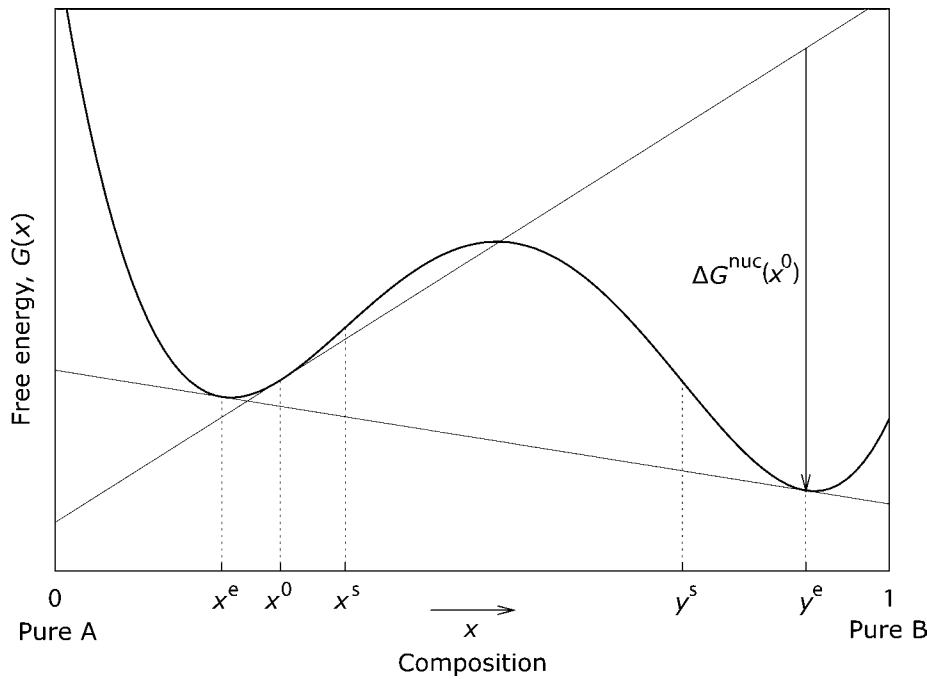

Fig. 1 Sketch of the free energy of a binary mixture quenched in a two-phase region. The bold line is the free energy per atom $G(x)$ of the homogeneous system. The compositions x^e and y^e of the equilibrium phases are given by the common tangent construction. The spinodal limits x^s and y^s define the unstable region. $\Delta G^{\mathrm{nuc}}(x^0)$ is the nucleation free energy of the metastable homogeneous system of composition x^0.
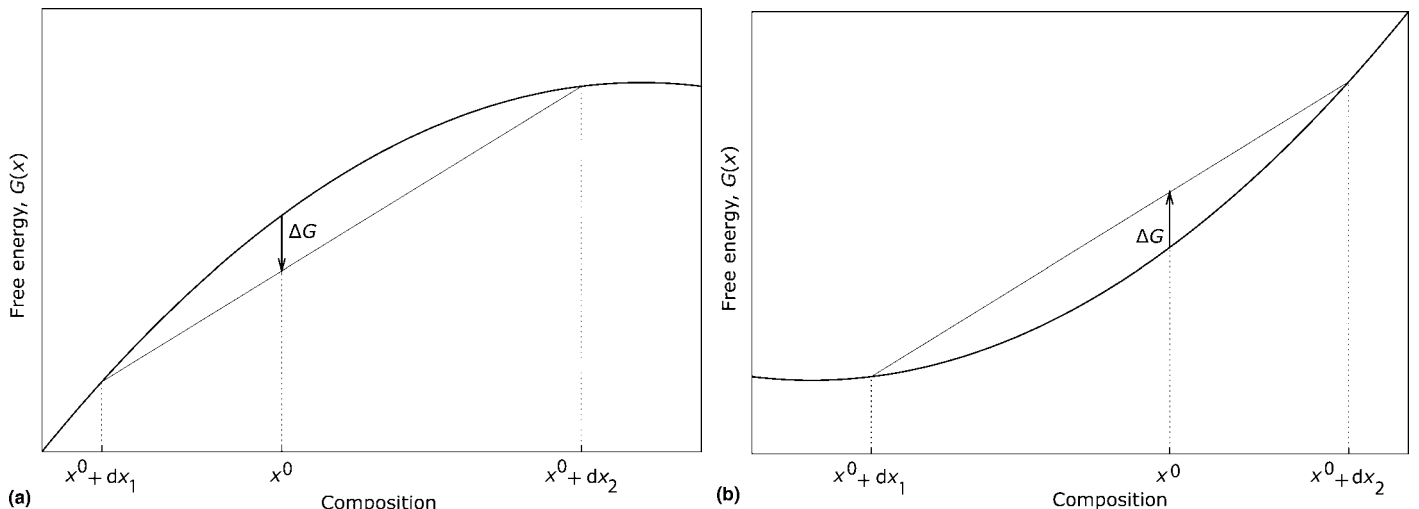

Fig. 2 Variation ΔG of the free energy corresponding to the spontaneous unmixing of a homogeneous system of composition x^0 in two phases of respective compositions $x^0 + dx_1$ and $x^0 + dx_2$. (a) Spinodal regime ($G''(x^0) < 0$). (b) Nucleation regime ($G''(x^0) > 0$)
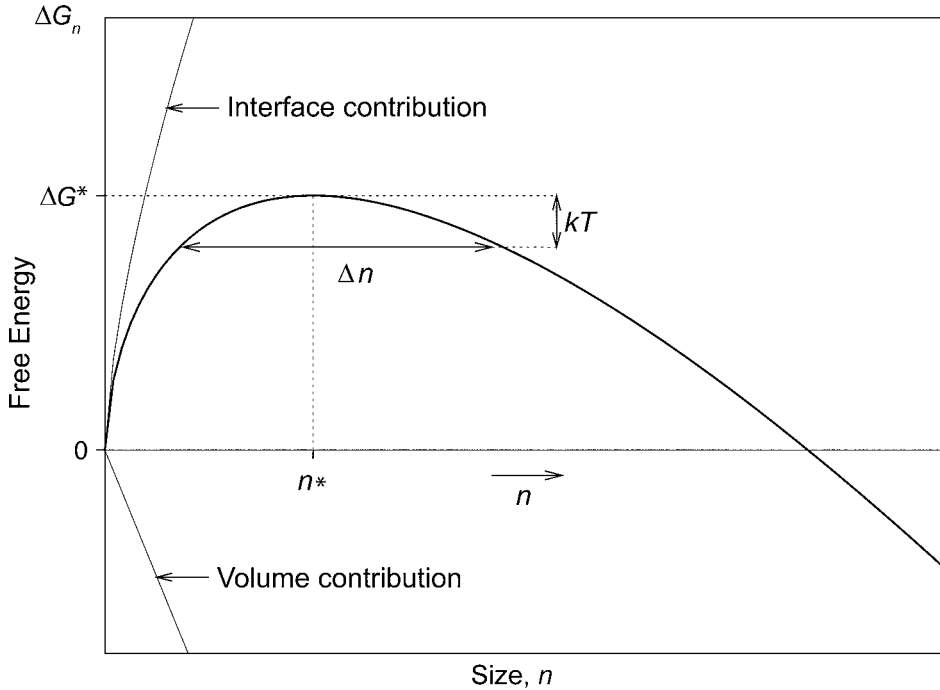
Fig. 3 Variation of the cluster formation free energy ΔG_n with the number, n, of atoms they contain as described by Eq 3. n^* is the critical size and ΔG^* the corresponding free energy. The size interval Δn characterizes the energy profile around the critical size and is directly linked to the Zeldovich factor (Eq 10).
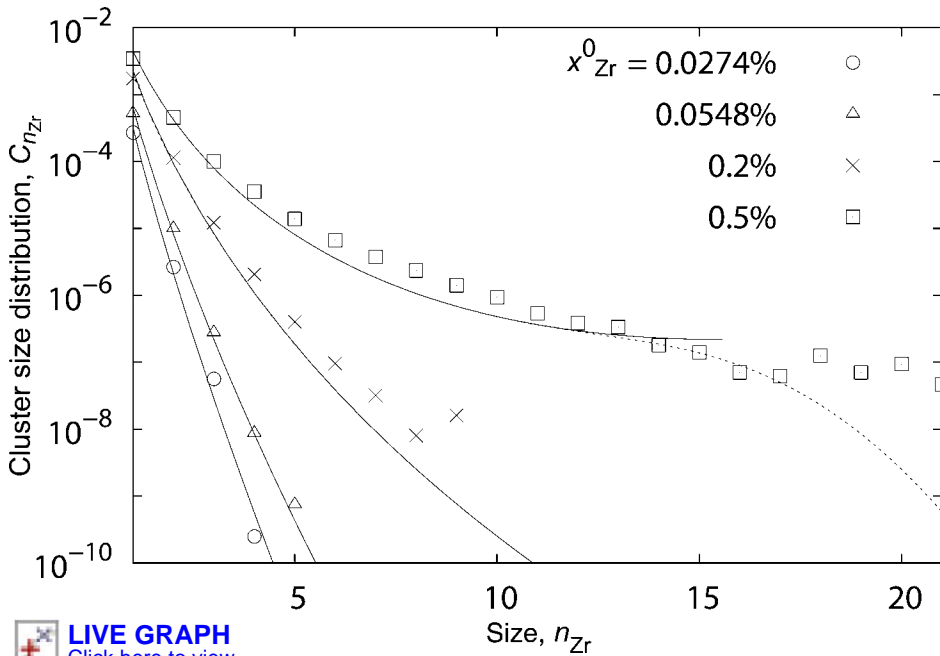

LIVE GRAPH
Click here to view

Fig. 4 Dependence on the nominal concentration x^0_{Zr} of the cluster size distribution of an aluminum solid solution at 500 °C. At this temperature, the solubility limit is $x^e_{Zr} = 0.0548$ %. Symbols correspond to atomic simulations (kinetic Monte Carlo) (Ref 18) and lines to predictions of the classical nucleation theory. Full lines correspond to the equilibrium cluster size distribution (Eq 7) for $n \leq n^*$ and dotted lines to the steady-state distributions (Eqs 8 and 85).

$$\Delta G^{\mathrm{nuc}} = \sum_i y_i^e (\mu_i^e - \mu_i^0) \qquad \text{(Eq 4)}$$

where y_i^e is the atomic fraction of the type i atom in the nucleating equilibrium phase, and μ_i^e and μ_i^0 are the corresponding chemical potentials in the nucleating equilibrium phase

and the parent phase, respectively. When the parent phase is metastable, chemical potentials in this phase are higher than the ones at equilibrium. The nucleation free energy given by Eq 4 is therefore negative. Classic expressions of the nucleation free energy are given at the end of

this section in some simple cases. For negative nucleation driving force, because of the competition between the volume and the interface contributions, the cluster formation free energy (Eq 3) shows a maximum for a given critical size, n^*, as illustrated in Fig. 3. n^* corresponds to the size at which the first derivative of ΔG_n is equal to zero, thus leading to:

$$n^* = \left(-\tfrac{2}{3} \frac{A\sigma}{\Delta G^{\mathrm{nuc}}}\right)^3 \qquad \text{(Eq 5)}$$

and the corresponding formation free energy:

$$\Delta G^* = \Delta G_{n^*} = \tfrac{4}{27} \frac{(A\sigma)^3}{(\Delta G^{\mathrm{nuc}})^2} \qquad \text{(Eq 6)}$$

Below this critical size, the energy of growing clusters increases because of the interface predominance at small sizes. Clusters in this size range are therefore unstable; if a cluster is formed, it will tend to redissolve. Nevertheless, unstable clusters can be found in the parent phase because of thermal fluctuations. The size distribution of these clusters is given by:

$$C_n^{\mathrm{eq}} = C_0 \exp\left(-\frac{\Delta G_n}{kT}\right) \qquad \text{(Eq 7)}$$

where C_0 is the atomic fraction of sites accessible to the clusters. For precipitation in the solid state, for instance, all lattice sites can receive a cluster, and therefore $C_0 = 1$. The validity of the size distribution (Eq 7) can be demonstrated for an undersaturated system ($\Delta G^{\mathrm{nuc}} \geq 0$) using a lattice gas model (compare with "Cluster Gas Thermodynamics" in the "Kinetic Approach" section of this article). For a supersaturated system, one assumes that the system reaches a steady state where clusters smaller than the critical size still obey the distribution (Eq 7).

Comparisons with atomic simulations have shown that Eq 7 correctly describes the size distribution of subcritical clusters. An example of such a comparison is given in Fig. 4 for aluminum-zirconium alloys, leading to the coherent precipitation of L1$_2$ Al$_3$Zr compounds (Ref 17); size distributions are given for undersaturated, saturated, and supersaturated solid solutions. A similar comparison leading to the same conclusion can be found in Ref 18 for an unmixing alloy on a body-centered cubic lattice, or in Ref 19 and 20 for the magnetization reversal of an Ising model in two and three dimensions, respectively.

The kinetic approach developed further in this article shows that the steady-state distribution in a nucleating system slightly deviates from the equilibrium distribution (Eq 7) around the critical size. An exact expression of the steady-state distribution has been obtained by Kashiev (Ref 21, 22). In the critical size interval, Δn, which is precisely defined as follows, it can be approximated by:

$$C_n^{\mathrm{st}} = [\tfrac{1}{2} - Z(n - n^*)]C_n^{\mathrm{eq}} \qquad \text{(Eq 8)}$$

The Zeldovich factor, Z, appearing in this equation is a function of the second derivative

of the cluster formation free energy at the critical size:

$$Z = \sqrt{-\frac{1}{2\pi kT}\frac{\partial^2 \Delta G_n}{\partial n^2}\bigg|_{n=n^*}} = \frac{3(\Delta G^{nuc})^2}{4\sqrt{\pi kT}(A\sigma)^{3/2}}$$

(Eq 9)

The physical meaning of the Zeldovich factor can be seen in Fig. 3, which sketches the variation of the cluster formation free energy with their size. If the formation free energy was harmonic, the size interval where the difference between the cluster free energy and the nucleation barrier, ΔG^*, is smaller than the thermal energy, kT, would be given by:

$$\Delta n = \frac{2}{\sqrt{\pi}}\frac{1}{Z}$$

(Eq 10)

The Zeldovich factor therefore characterizes the flatness of the energy profile around the critical size. Equation 8 shows that steady-state cluster concentrations in the critical region are reduced compared to the equilibrium distribution. For the critical size, a factor ½ appears in front of the equilibrium concentration.

Steady-State Nucleation Rate

When the nucleation barrier, ΔG^*, is high enough compared to the thermal energy, kT, the metastable state of the system contains thermal fluctuations well described by the distribution (Eq 7). Sometimes, one of these fluctuations will reach and overcome the critical size. It can then continue to grow and become more and more stable. Classical nucleation theory assumes that the system reaches a steady state, and it then shows that stable nuclei appear at a rate given by (Ref 3):

$$J^{st} = \beta^* Z C_0 \exp\left(-\frac{\Delta G^*}{kT}\right)$$

(Eq 11)

where β^* is the rate at which a critical cluster grows, and Z is the Zeldovich factor (Eq 9). This factor has been introduced by Becker and Döring (Ref 3) to describe cluster fluctuations around the critical size and, in particular, the probability for a stable nucleus to redissolve. $ZC_0 \exp(-\Delta G^*/kT)$ is therefore the number of critical clusters that reach a size large enough that they can continuously grow. The initial expression of the nucleation rate derived by Volmer and Weber (Ref 1) and by Farkas (Ref 2) did not consider this Zeldovich factor and led to an overestimation of the nucleation rate. A small Zeldovich factor corresponds to a flat energy profile around the critical size. Critical clusters experience size variations that are mainly random and not really driven by their decrease in energy. Some of them will redissolve and not fall in the stable region. This explains why the nucleation rate is reduced by the Zeldovich factor. A more rigorous derivation of the nucleation rate where the Zeldovich factor naturally appears is given in the section

"The Link with Classical Nucleation Theory" in this article.

An expression for the growing rate β^* of the critical cluster is needed. If the growth-limiting process is the reaction at the interface to attach the atoms on the critical cluster (ballistic regime), β^* is then proportional to the cluster area. Assuming that this reaction is controlled by one type of atom, the following expression is obtained (Ref 23):

$$\beta^* = 4\pi r^{*2} \frac{\lambda_i \Gamma_i}{\Omega} \frac{x_i^0}{y_i^e}$$

(Eq 12)

where r^* is the radius of the critical cluster, λ_i is the distance corresponding to the atom last jump to become attached to the critical cluster, Γ_i is the corresponding reaction frequency, and Ω is the volume corresponding to one atomic site. x_i^0 and y_i^e are the respective atomic fraction of the jumping atoms in the metastable parent phase and the stable nucleating phase.

For solid - solid phase transformations, the critical cluster growth is usually controlled by the long-range diffusion of solute atoms. The critical condensation rate is then obtained by solving the classical diffusion problem associated with a growing spherical particle. If diffusion of only one type of atom limits the growth, and all other atomic species diffuse sufficiently fast enough so that the cluster composition instantaneously adjusts itself, one obtains (Ref 23):

$$\beta^* = 4\pi r^* \frac{D_i}{\Omega} \frac{x_i^0}{y_i^e}$$

(Eq 13)

where D_i is the diffusion coefficient of type i atoms. In a multicomponent alloy, when

diffusion coefficients of different atomic species have close values and when the composition of the critical cluster can vary, one must use the linked flux analysis presented in the section "Cluster Dynamics" in this article. In all cases, the growth rate is proportional to the cluster radius in this diffusive regime.

Both events, that is, the long-range diffusion and the reaction at the interface, can be simultaneously taken into account. The corresponding expression of the condensation rate has been derived by Waite (Ref 24).

Transient Nucleation

A transient regime exists before the nucleation rate reaches its stationary value (Eq 11). One conventionally defines an incubation time, or a time lag, to characterize this transient regime. This is defined as the intercept with the time axis of the tangent to the curve representing the variations of the nuclei density (Fig. 5). Exact expressions of the incubation time have been obtained as a series of the initial and steady-state cluster size distributions (Ref 26, 27). Different approximations have then been made to evaluate this series and obtain closed forms of the incubation time. They all lead to an incubation time:

$$t_{inc}(n^*) = \theta_0 \frac{1}{\pi Z^2 \beta^*} = \theta_0 \frac{16kT(A\sigma)^3}{9(\Delta G^{nuc})^4 \beta^*}$$

(Eq 14)

where the factor θ_0 depends on the chosen approximation and is close to 1 (Ref 28). Some authors obtained a factor θ_0 that depends slightly on the temperature and the shape of the
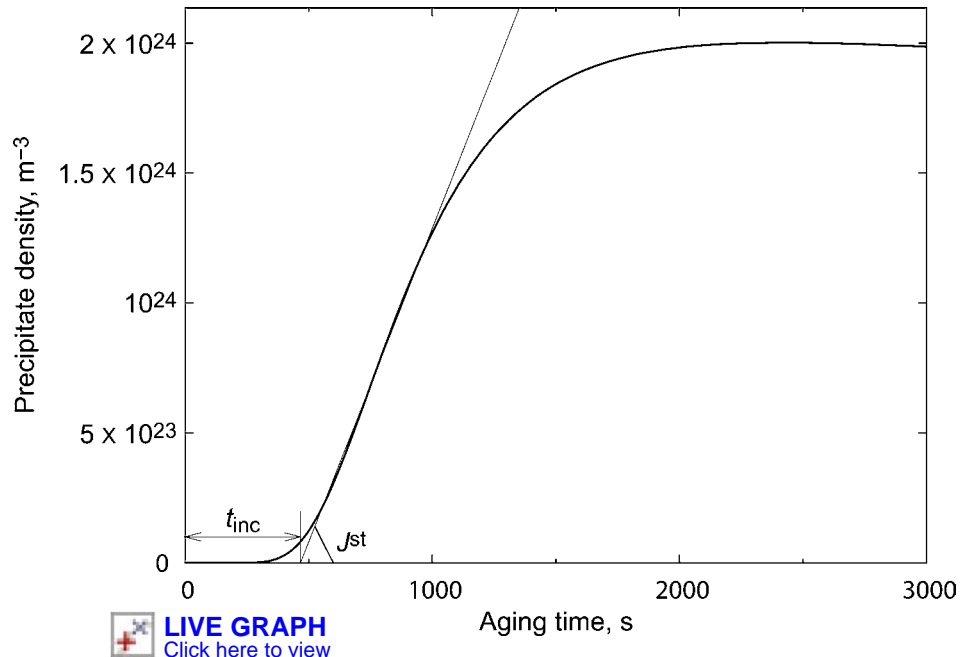

Fig. 5 Precipitate density as a function of aging time for an aluminum solid solution containing 0.18 at.% Sc aged at 300 °C. The time evolution obtained from cluster dynamics simulations (Ref 25) allows the definition of a steady-state nucleation rate, J^{st}, and an incubation time, t_{inc}.

cluster formation free energy around the critical size (Ref 26, 29). One can stress, however, that a precise value of this factor is seldom, if ever, needed. As is shown later, the incubation time depends on too many parameters to be known precisely experimentally. Equation 14 allows describing its main variation when the temperature or the nucleation driving force are changed; this is usually enough to model incubation and nucleation.

This expression (Eq 14) of the incubation time can be obtained from simple physical considerations (Ref 23, 30). The steady state will be reached once the clusters have grown sufficiently far away from the critical size. Supercritical clusters have a negligible probability to decay when their size becomes greater than $n^* + \frac{1}{2}\Delta n$. Δn characterizes the width of the critical region (Fig. 3) and is related to the Zeldovich factor via Eq 10. Because the energy profile is flat in this neighborhood, clusters make a random walk in the size space, with a constant jump frequency β^*. Accordingly, the corresponding time needed to diffuse from n^* to $n^* + \frac{1}{2}\Delta n$ is:

$$t_{inc}(n^*) \sim \frac{\Delta n^2}{4\beta^*} \qquad (Eq\ 15)$$

This leads to the expression (Eq 14) with a factor $\theta_0 = 1$.

Different approximations of the nucleation rate in this transient regime can also be found in the literature. Kelton et al. (Ref 28) have compared these approximations with exact results obtained, thanks to a numerical integration of the kinetic equations describing nucleation. They concluded that the best-suited approximation to describe the transient nucleation rate is the one obtained by Kashchiev (Ref 21, 22):

$$J(t) = J^{st}\left[1 + 2C\sum_{m=1}^{\infty}(-1)^m \exp\left(-\frac{m^2 t}{\tau}\right)\right] \qquad (Eq\ 16)$$

where $C = 1$ for a system initially prepared in a state far from its nucleating metastable state. The time constant is given by:

$$\tau = \frac{4}{\pi^3}\frac{1}{Z^2\beta^*} \qquad (Eq\ 17)$$

When $t > \tau$, one can retain only the first term in the sum appearing in Eq 16. Usually, it is even enough to assume that the nucleation rate behaves like the Heaviside step function; that is, that the nucleation rate reaches its stationary value after an incubation time where no nucleation occurs. The incubation time corresponding to Eq 16 is $\pi^2\tau/6$. Therefore, in the Kashchiev treatment, the factor in Eq 14 is $\theta_0 = \frac{2}{3}$.

It is worth saying that the incubation time and the associated transient regime depend on the conditions in which the system has been prepared. Equations 14 and 16 implicitly assume that the quench was done from infinite temperature; no cluster around the critical one existed

at the initial time. This may not be true. For instance, the system could have been prepared in an equilibrium state corresponding to a slightly higher temperature where it was stable and then quenched in a metastable state. A cluster distribution corresponding to this higher temperature already exists before the beginning of the phase transformation. If the temperature difference of the quench is small, these preexisting clusters will reduce the incubation period. The dependence of the incubation time on the initial conditions has been observed, for instance, in atomic simulations for an unmixing binary alloy (Ref 18). Starting from a random solid solution corresponding to an infinite temperature preparation, an incubation time is observed before nucleation reaches its steady state. If the alloy is annealed above its solubility limit before a quench, the incubation stage disappears if the temperature difference of the quench is not too high. Kashchiev considered the effect of this initial cluster distribution on nucleation in the case of a change in pressure (Ref 22, 31). His results can be easily generalized (Ref 28). To do so, the supersaturation variation is introduced:

$$\Delta s = \frac{\Delta G^{nuc}(t=0^-)}{kT(t=0^-)} - \frac{\Delta G^{nuc}(t=0^+)}{kT(t=0^+)} \qquad (Eq\ 18)$$

where $t = 0^-$ means that thermodynamic quantities are calculated for the initial state in which the system has been prepared, and $t = 0^+$ for the state where nucleation occurs. In his derivation, Kashchiev assumed that the interface free energy of the clusters is the same in both stable and metastable states. The constant entering in the expression (Eq 16) of the transient nucleation rate is then:

$$C = 1 - \frac{\Delta s}{Z}\exp(-n^*\Delta s) \qquad (Eq\ 19)$$

and the corresponding incubation time is multiplied by this constant C. The supersaturation variation, Δs, is positive; otherwise, nucleation would have happened in the initial state in which the system has been prepared. The existence of an initial cluster size distribution therefore always reduces the incubation time. Nevertheless, C rapidly tends to 1 when the thermodynamic states $t = 0^-$ and $t = 0^+$ become too different.

By definition, the nucleation rate does not depend on the cluster size in the stationary regime. This property is used to advantage in Eq 11 to calculate the steady-state nucleation rate, J^{st}, at the critical size. However, the time needed for the stationary regime to develop will, of course, vary with the cluster size. This means that the incubation time depends on the cluster size at which it is measured. The previously defined incubation time corresponds to the critical size. However, the smallest cluster size that one can detect experimentally may be significantly larger than the critical size. Therefore, it is necessary to describe the variation with the

cluster size of the incubation time. This problem has been solved by Wu (Ref 26) and Shneidman and Weinberg (Ref 29), who showed that the incubation time measured at size n is:

$$t_{inc}(n) = t_{inc}(n^*) + \frac{1}{2\pi Z^2\beta^*}\left\{\theta_1 + \ln\left[\sqrt{\pi}Z(n-n^*)\right]\right\} \qquad (Eq\ 20)$$

for $n > n^* + \Delta n$, that is, a cluster size outside the critical region. The constant θ_1 is 1 in the expression obtained by Wu and $\theta_1 = \gamma/4 + \ln(2)/2$ for Shneidman and Weinberg, where $\gamma \sim 0.5772$ is Euler's constant.

All the aforementioned expressions are obtained in the parabolic approximation, that is, assuming that the cluster formation free energy is well described by its harmonic expansion around the critical size. According to Shneidman and Weinberg (Ref 29), this approximation is highly accurate when calculating the steady-state nucleation rate, but its validity is limited for the incubation time. When considering the exact shape of the cluster formation free energy (Eq 3), the expression of the incubation time then depends on the model used for the absorption rate. In all cases, the incubation time at the critical size can be written:

$$t_{inc}(n^*) = \frac{1}{2\pi Z^2\beta^*}\left[\frac{\gamma}{2} + \ln\left(\sqrt{\pi}Zn^*\right) - \theta_3\right] \qquad (Eq\ 21)$$

where the constant θ_3 differs from 0 when the parabolic approximation is not used. In the ballistic regime, when the condensation rate is proportional to the cluster surface, such as in Eq 12, the authors obtained $\theta_3 = 1 - \ln(3)$. The incubation time measured at size n is then:

$$t_{inc}(n) = \frac{1}{2\pi Z^2\beta^*}\left\{\left(\frac{n}{n^*}\right)^{1/3} + \ln\left[\left(\frac{n}{n^*}\right)^{1/3} - 1\right]\right.$$
$$\left. + \gamma - 2 + \ln\left[\frac{6\Delta G^*}{kT}\right]\right\}, \quad \forall n > n^* + \frac{1}{2}\Delta n \qquad (Eq\ 22)$$

In the diffusive regime, when the condensation rate is proportional to the cluster radius, such as in Eq 13, $\theta_3 = 3/2 - \ln(3)$ and:

$$t_{inc}(n) = \frac{1}{2\pi Z^2\beta^*}\left\{\frac{1}{2}\left[\left(\frac{n}{n^*}\right)^{1/3} + 2\right]^2 + \ln\left[\left(\frac{n}{n^*}\right)^{1/3} - 1\right]\right.$$
$$\left. + \gamma - \frac{7}{2} + \ln\left[\frac{6\Delta G^*}{kT}\right]\right\}, \quad \forall n > n^* + \frac{1}{2}\Delta n. \qquad (Eq\ 23)$$

It should be stressed that all these expressions for incubation time have been obtained in the continuous limit valid for large clusters. The expressions should be used only when the nucleation barrier ΔG^* is high enough for the critical size not being too small.

Heterogeneous Nucleation

Until now, only homogeneous nucleation has been considered; it was assumed that nuclei can form anywhere in the system. However, it may

require less energy for the nuclei to form heterogeneously on preferred nucleation sites. These sites can be at the interface with existing impurities or some lattice defects such as grain boundaries or dislocations. The classical theory also allows modeling heterogeneous nucleation after some slight modifications. The first modification is that the parameter C_0 appearing in the cluster size distribution (Eq 7) is now the number of sites where heterogeneous precipitation can take place. One also needs to take into account the decrease of the nuclei free energy when they are located at a preferred nucleation site. Such a decrease usually arises from a gain in the interface free energy; it is more favorable for the nuclei to form on an already existing interface, because the cost to create the interface between the old and the new phases is reduced.

First consider the case where the cluster wets the substrate and has a cap shape (Fig. 6a). Electrodeposition is one example where this happens (Ref 32). Three different interface free energies must be considered:

- σ between the parent and the nucleating phase
- σ_s between the parent phase and the substrate
- σ_i between the nucleating phase and the substrate

The wetting angle is then defined by the Young equilibrium equation (Ref 16):

$$\cos \theta_w = \frac{\sigma_s - \sigma_i}{\sigma} \tag{Eq 24}$$

The wetting leads to a cap shape only if the interface free energies obey the inequalities $-\sigma \leq \sigma_s - \sigma_i \leq \sigma$. If the difference $\sigma_s - \sigma_i$ is smaller than $-\sigma$, then $\theta_w = \pi$, and the wetting is not possible because an unwet cluster costs less energy. On the other hand, if the difference is greater than σ, the wetting is complete, and one can no longer define a cap shape because the nucleating phase will uniformly cover the interface.

The cluster free energy takes the same expression as the one given by the capillary approximation in the homogeneous case (Eq 3). To calculate the geometric factor A appearing in this expression, the radius, R, of the cap must be defined. The cluster volume is then given by (Ref 22):
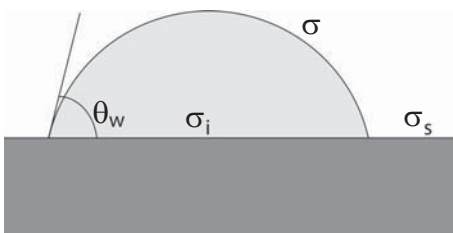
$$n\Omega_1 = \tfrac{1}{3}\pi R^3(2 + \cos \theta_w)(1 - \cos \theta_w)^2$$
$$= \pi R^3 \frac{(2\sigma + \sigma_s - \sigma_i)(\sigma - \sigma_s + \sigma_i)^2}{3\sigma^3} \tag{Eq 25}$$

and the free energy associated with the whole cluster interface by:

$$n^{2/3}A\sigma = \pi R^2 \left[\sigma 2(1 - \cos \theta_w) + (\sigma_i - \sigma_s)\sin^2 \theta_w \right]$$
$$= \pi R^2(\sigma - \sigma_s + \sigma_i)$$
$$\times \frac{2\sigma^2 + (\sigma_i - \sigma_s)(\sigma + \sigma_s - \sigma_i)}{\sigma^2} \tag{Eq 26}$$

Eliminating the variable R between Eq 25 and 26, one obtains the expression of the geometric factor appearing in the capillary approximation:

$$A = \left(9\pi\Omega_1^2\right)^{1/3} \frac{2\sigma^2 - (\sigma_s - \sigma_i)(\sigma + \sigma_s - \sigma_i)}{\sigma(2\sigma + \sigma_s - \sigma_i)^{2/3}(\sigma - \sigma_s + \sigma_i)^{1/3}} \tag{Eq 27}$$

When $\sigma_s - \sigma_i = \sigma$, the unwetting is complete; one recovers the geometric factor $A = \left(36\pi\Omega_1^2\right)^{1/3}$ corresponding to a spherical cluster. With this expression of the geometric factor and the correct value of the parameter C_0, all expressions obtained for homogeneous nucleation can also be used for heterogeneous nucleation.

Nuclei can also have a lens shape (Fig. 6b). In such a case, the two wetting angles are defined by (Ref 22):

$$\cos \theta_w = (\sigma_s^2 + \sigma^2 - \sigma_i^2)/2\sigma_s\sigma$$
$$\cos \theta_s = (\sigma_s^2 + \sigma^2 - \sigma_i^2)/2\sigma_s\sigma_i \tag{Eq 28}$$

The geometric factor corresponding to this lens shape is obtained using the same method as previously; one expresses the volume and the interface energy of the two caps composing the cluster and then eliminates the cap radii between these two equations.

Examples

It is worth having a closer look at some examples—solidification and precipitation in the solid state—and giving an approximated expression of the nucleation free energy in these simple cases.

Example 1: Solidification. A single component liquid that was initially at equilibrium is quenched at a temperature, T, below its melting temperature, T_m. Because the liquid and the solid have the same composition, the nucleation free energy is simply the free-energy difference between the liquid and the solid states at the temperature T. If the undercooling is small, one can ignore the difference in the specific heats of the liquid and the solid. The nucleation free energy is then proportional to the latent heat of fusion per atom, L (Ref 16):

$$\Delta G^{nuc} = L\frac{T - T_m}{T_m} \tag{Eq 29}$$

When the undercooling is large, Eq 29 may be not precise enough. One can then consider the next term in the Taylor expansion (Ref 22), leading to:

$$\Delta G^{nuc} = L\frac{T - T_m}{T_m} - \Delta C_p \frac{(T - T_m)^2}{2T_m} \tag{Eq 30}$$

where $\Delta C_p = C_p^{liq}(T_m) - C_p^{sol}(T_m)$ is the difference in the heat capacities of the liquid and solid phases.

Equation 29 shows how the steady-state nucleation rate varies with the quenching temperature. Assuming that the interface free energy is constant and that the condensation rate, β^*, simply obeys an Arrhenius law, one obtains the nucleation rate given by:

$$J^{st} = \lambda \frac{(T - T_m)^2}{\sqrt{T}} \exp\left[-\left(\frac{A}{T} + \frac{B}{T(T - T_m)^2} \right) \right] \tag{Eq 31}$$

where λ, A, and B are positive constants. The nucleation rate corresponding to this equation is sketched in Fig. 7. The main variations are given by the exponential. As a consequence, there is a temperature window in which the nucleation rate is substantial. For high temperatures close to the melting temperature, T_m, the nucleation free energy is small and leads to a negligible nucleation rate. At low temperatures, the nucleation rate is also negligible because of the Arrhenian behavior of the kinetic factor and the critical cluster concentration (Eq 7). The nucleation rate can be measured only at intermediate
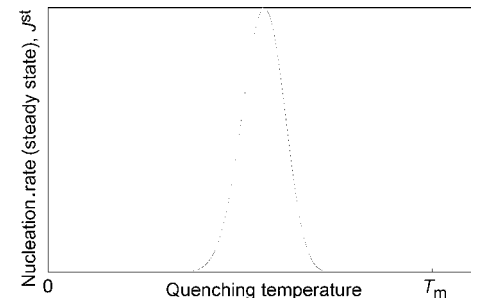

Fig. 6 Possible shapes of a nucleus in heterogeneous nucleation. (a) Cap shape. (b) Lens shape

Fig. 7 Variation with quenching temperature of steady-state nucleation rate in the case of solidification (Eq 31). T_m, melting temperature

temperatures. Such conclusions on the nucleation rate are not specific to solidification but are encountered in any nucleation experiment.

Example 2: Precipitation in the Solid State. In this example, it is necessary to take into account elastic effects. The free energy is thus divided between a chemical and an elastic contribution.

Chemical Contribution. For the binary mixture, whose free energy per atom $G(x)$ is sketched in Fig. 1, the homogeneous metastable phase of composition x^0 has a nucleation free energy given by:

$$\Delta G^{\mathrm{nuc}}\left(x^0\right) = (1 - y^{\mathrm{e}})\left[\mu_A\left(y^{\mathrm{e}}\right) - \mu_A\left(x^0\right)\right] \\ + y^{\mathrm{e}}\left[\mu_B\left(y^{\mathrm{e}}\right) - \mu_B\left(x^0\right)\right] \quad \text{(Eq 32)}$$

A and B atom chemical potentials are respectively defined as the first derivatives of the total free energy with respect to the number N_A and N_B of A and B atoms. This leads to the following expressions:

$$\mu_A(x) = G(x) - xG'(x) \\ \mu_B(x) = G(x) + (1-x)G'(x) \quad \text{(Eq 33)}$$

which check the property $(N_A + N_B)G(x) = N_A\mu_A + N_B\mu_B$.

Incorporating these expressions in Eq 32:

$$\Delta G^{\mathrm{nuc}}\left(x^0\right) = G(y^{\mathrm{e}}) - G\left(x^0\right) - \left(y^{\mathrm{e}} - x^0\right)G'\left(x^0\right) \quad \text{(Eq 34)}$$

This shows that the nucleation free energy corresponds to the difference, calculated in the point of abscissa y^{e}, between the free energy and the tangent in x^0, as illustrated in Fig. 1.

It should be stressed, however, that this construction does not correspond to the maximal nucleation driving force. If the stoichiometry of the precipitates is allowed to vary, the maximal nucleation driving force is obtained for a cluster composition y^0 corresponding to the point where the tangent to the free energy is parallel to the tangent in point of abscissa x^0 (Fig. 8). Such a deviation of the nucleating phase from its equilibrium may be important to consider. An example is the precipitation of carbonitride precipitates in steels (Ref 33). The following considers that the free energy well defining the nucleating phase is deep enough that the compositions y^{e} and y^0 can be assumed identical. This question of the precipitate composition is revisited in "Nonstoichiometric Clusters" of the "Kinetic Approach" section of this article, where a general framework to treat variations of the precipitate composition is presented.

To go further, one must consider a precise function for the free energy. The regular solid solution is a convenient energetic model that is representative of a binary alloy. In this model, the free energy per atom is:

$$G(x) = kT[x\ln(x) + (1-x)\ln(1-x)] + x(1-x)\omega \quad \text{(Eq 35)}$$

where ω is the interaction parameter. When this parameter is positive, the alloy tends to unmix at low temperature, and the corresponding phase diagram possesses a two-phase region. For temperatures lower than $\omega/2k$, the free energy indeed has two minima, and its variation with the composition is similar to the one sketched in Fig. 1. The nucleation free energy of a solid solution quenched in a metastable state, as given by this thermodynamic model, is:

$$\Delta G^{\mathrm{nuc}}\left(x^0\right) = (1 - y^{\mathrm{e}})kT\ln\left(\frac{1 - x^{\mathrm{e}}}{1 - x^0}\right) \\ + y^{\mathrm{e}}kT\ln\left(\frac{x^{\mathrm{e}}}{x^0}\right) + \omega\left(x^0 - x^{\mathrm{e}}\right) \quad \text{(Eq 36)}$$

A useful approximation of this expression is the dilute limit corresponding to a small solubility limit, $x^{\mathrm{e}} \ll 1$, and a small nominal concentration, $x^0 \ll 1$. In that case, one can keep only the major contribution in the nucleation free energy, leading to:

$$\Delta G^{\mathrm{nuc}}\left(x^0\right) = y^{\mathrm{e}}kT\ln\left(\frac{x^{\mathrm{e}}}{x^0}\right) \quad \text{(Eq 37)}$$

This generally gives a good approximation of the nucleation free energy at low temperature for not-too-high supersaturations. It then allows predicting the main consequences of a variation of the solid-solution nominal composition on the nucleation. This approximation of the nucleation free energy in the dilute limit can be easily generalized to a multicomponent alloy.

Other thermodynamic approaches can be used to obtain expressions of the nucleation free energy. It is possible, for instance, to describe interactions between atoms with an Ising model. Chemical potentials entering in Eq 4 can then be calculated with the help of current thermodynamic approximations, such as mean-field approximations and low- or high-temperature expansions (Ref 34). Using the simple Bragg-Williams mean-field approximation, one indeed recovers the expression (Eq 36) corresponding to the regular solid-solution model. An example of this approach, starting from an atomic model, is given in Ref 17 and 35 for a face-centered cubic solid solution leading to the nucleation of a stoichiometric compound with the $L1_2$ structure, such as aluminum-zirconium or aluminum-scandium alloys. On the other hand, it is possible to use an experimental thermodynamic database, such as the ones based on the Calphad approach (Ref 36, 37), to calculate the nucleation free energy.

Elastic Contribution. Usually, the precipitating phase has a different structure or molar volume from the parent phase. If the interface between both phases remains coherent, an elastic contribution must be taken into account in the formation free energy of the clusters (Eq 3). Similar to the "chemical" nucleation free energy, this elastic contribution varies linearly with the volume, V, of the cluster. Its sign is always positive because there is an extra energy cost to maintain coherency at the interface. One can illustrate this elastic contribution by considering the case of a precipitating phase having a slightly different equilibrium volume from the parent phase, as
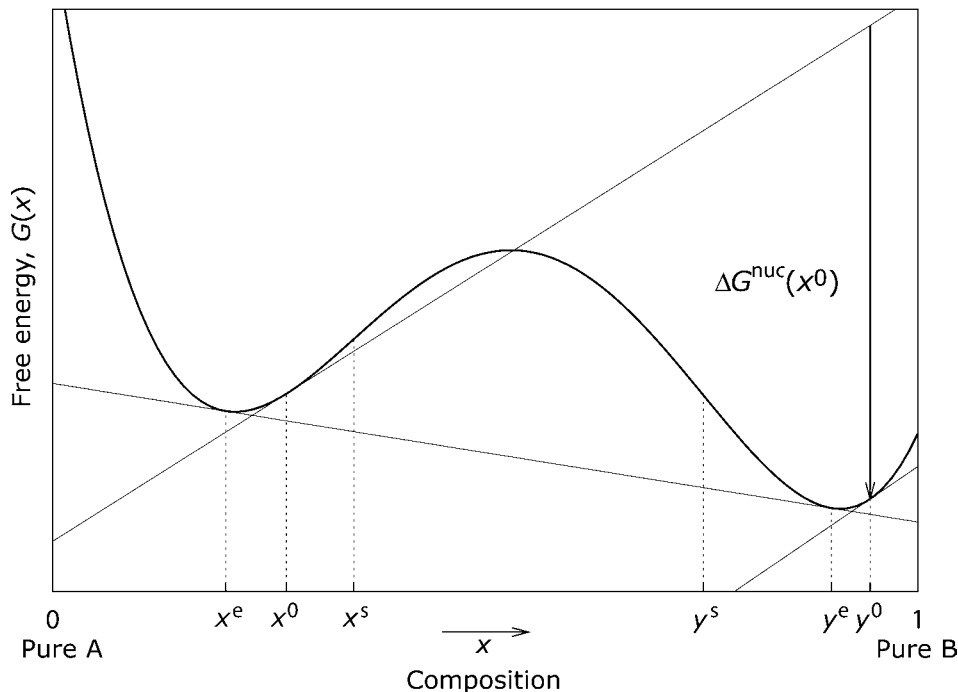

Fig. 8 Parallel tangent construction leading to the maximal nucleation driving force for precipitates having the composition y^0

well as different elastic constants. For the sake of simplicity, the assumption is that both phases have an isotropic elastic behavior characterized by their Lamé coefficients, λ and μ for the parent phase, and λ' and μ' for the precipitating phase. If a and $a(1 + \delta)$ are the respective lattice parameters of the two phases, the elastic energy necessary to embed a spherical cluster of volume V in an infinite elastic medium corresponding to the parent phase is:

$$\Delta G^{\mathrm{el}} = V \frac{6\mu(3\lambda' + 2\mu')}{3\lambda' + 2\mu' + 4\mu} \delta^2 \qquad \text{(Eq 38)}$$

The model of the elastic inclusion and inhomogeneity developed by Eshelby (Ref 38–40) allows calculating the elastic energy in more complicated situations, when the inhomogeneity elastic behavior is anisotropic or when the inclusion stress-free strain is different from a simple pure dilatation. One can also deduce from this model the cluster shape minimizing its elastic self-energy. Nevertheless, this model is tractable only when the inclusion is an ellipsoid. When the elastic contribution becomes important compared to the interface one, the shape of the critical cluster strongly deviates from an ellipsoid. One can then use a diffuse interface phase-field model to determine the critical nucleus morphology and determine the associated nucleation activation energy (Ref 41, 42). However, in all cases, the extra energy cost arising from elasticity is positive and proportional to the inclusion volume. It thus reduces the absolute value of the nucleation driving force.

This inclusion model allows deriving the cluster self-elastic energy. However, the interaction of the cluster with the surrounding microstructure is ignored. In particular, one does not consider the elastic interaction between different clusters. Such an interaction is long range and cannot always be neglected. It may lead to self-organized morphological patterns due to preferred nucleation sites around already existing clusters. In the case where the strain induced by the microstructure varies slowly compared to the size of the nucleating cluster, it has been shown that the interaction elastic energy depends linearly on the cluster volume and is independent of its shape (Ref 43). This interaction energy, whose sign is not fixed, depends on the position of the cluster. It can be considered in the cluster formation free energy (Eq 3) to model strain-enhanced nucleation. Such a model is able to predict, for instance, variation of the nucleation driving force near an existing precipitate between the elastically soft and hard directions. A natural way to develop such a model is to use a phase-field approach (see the Appendix to this article).

Kinetic Approach

Predictions of the classical nucleation theory, that is, the steady-state nucleation rate and the incubation time, are approximated solutions of

kinetic equations describing the time evolution of the system. Instead of using results of the classical nucleation theory, one can integrate these kinetic equations numerically. This kinetic approach is known as cluster dynamics. It rests on the description of the system undergoing phase separation as a gas of clusters that grows and decays by absorbing and emitting other clusters. In this section, the cluster gas thermodynamic formalism used by cluster dynamics is described first. Kinetic equations simulating the phase transformation are presented. Finally, the link with classical nucleation theory is shown. It is generally assumed that the stoichiometry of the nucleating phase cannot vary. This is thus equivalent to considering the nucleation of clusters with a fixed composition that is known a priori. The end of this section shows how this strong assumption can be removed when one is interested in the nucleation of a multicomponent phase with a varying composition.

Cluster Gas Thermodynamics

The system is described as a gas of noninteracting clusters having a fixed stoichiometry corresponding to that of the precipitating phase at equilibrium with the parent phase. Clusters are groups of atoms that are linked by a neighborhood relation. If one wants to model precipitation in an unmixing alloy, for instance, one can consider that all solute atoms that are closer than a cutoff distance belong to the same cluster. No distinction is made between clusters belonging to the old or to the new phase. In this modeling approach, clusters are defined by a single parameter: their size or the number, n, of atoms they contain. The term G_n is the free energy of a cluster containing n atoms embedded in the solvent. G_n is a free energy and not simply an energy because of the configurational entropy; for a given cluster size, there can be different configurations having different energies. Thus, the associated partition function must be considered. If D_n^i is the number of configurations having the energy H_n^i for a cluster of size n, the cluster free energy is then defined as:

$$G_n = -kT \ln \left[\sum_i D_n^i \exp\left(-H_n^i / kT\right) \right] \qquad \text{(Eq 39)}$$

It is formally possible to divide this free energy into a volume and an interface contribution such as in the capillary approximation, except that the interface free energy, σ_n, may now depend on the cluster size. This free energy corresponds to an interface between the stoichiometric cluster and the pure solvent. Thus, in three dimensions:

$$G_n = n g^{\mathrm{e}} + n^{2/3} (36\pi\Omega^2)^{1/3} \sigma_n \qquad \text{(Eq 40)}$$

where g^{e} is the free energy per atom of the bulk equilibrium precipitating phase, that is, without any interface. This is, by definition, the sum of

the chemical potentials, μ_i^{e}, for each constituent of the cluster modulated by its atomic fraction, y_i^{e}:

$$g^{\mathrm{e}} = \sum_i y_i^{\mathrm{e}} \mu_i^{\mathrm{e}} \qquad \text{(Eq 41)}$$

The interface free energy, σ_n, entering in Eq 40 is an average isotropic parameter, and clusters, on average, are therefore assumed to be spherical. One important difference with the capillary approximation is that this interface free energy now depends on the size n of the cluster. It is possible to compute the cluster free energy, G_n, starting from an energetic model describing interactions between atoms. For small clusters, one can directly enumerate the different configurations, i, accessible to a cluster of size n, and then directly build the free energy (Eq 39) (Ref 17, 32, 44). Because the degeneracy D_n^i grows very rapidly with the size of the cluster, this approach is limited to small clusters. For larger clusters, one can sample thermodynamic averages with Monte Carlo simulations to compute the free-energy difference between a cluster of size n and one of size $n + 1$ at a given temperature (Ref 44, 45). These simulations have shown that, in three dimensions, the size dependence of the interface free energy is well described for large enough clusters by a generalized capillary approximation:

$$\sigma_n = \sigma \left(1 + c n^{-1/3} + d n^{-2/3} + e n^{-2/3} \ln(n) \right) \qquad \text{(Eq 42)}$$

where the temperature-dependent coefficients c, d and e correspond to the line, the point, and the undulation contributions to the interface free energy (Ref 45). They take into account the interface curvature. The asymptotic limit of Eq 42 corresponds to the constant interface free energy of the classic capillary approximation, which also depends on temperature.

Some other expressions have been proposed in the literature for the size dependence of the interface free energy. Gibbs (Ref 46) indeed obtained a differential equation of this size dependence. Integrating this expression, Tolman (Ref 47) obtained the following expression:

$$\sigma_n = \sigma \left[1 + \left(\frac{n_0}{n} \right)^{1/3} \right]^{-2} \qquad \text{(Eq 43)}$$

where n_0 is a parameter. One can see, however, that Eq 42 and 43 are equivalent up to the order $o(n^{-1/3})$ when n tends to infinity.

Consider an assembly composed of noninteracting clusters and model thermodynamics in the cluster gas approximation of Frenkel (Ref 48). If N_n is the number of clusters containing n atoms, the free energy of the system is given by:

$$G = G_0 + \sum_{n=1}^{\infty} N_n G_n - kT \ln(W) \qquad \text{(Eq 44)}$$

where G_0 is the free energy in the absence of clusters, and W is the number of different configurations accessible to the cluster assembly.

Assuming that each cluster, whatever its size, lies only on one site, and neglecting around each cluster all excluded sites that cannot be occupied by any other cluster, this number is simply given by:

$$W = \frac{N_0!}{\left(N_0 - \sum_{n=1}^{\infty} N_n\right)! \prod_{n=1}^{\infty} N_n!} \qquad \text{(Eq 45)}$$

where N_0 is the number of sites accessible to the cluster. Application of the Stirling formula leads to the following estimation for the free energy:

$$G = G_0 + \sum_{n=1}^{\infty} N_n G_n + kT \sum_{n=1}^{\infty} N_n \ln(N_n)$$
$$+ kT\left(N_0 - \sum_{n=1}^{\infty} N_n\right) \ln\left(N_0 - \sum_{n=1}^{\infty} N_n\right)$$
$$- N_0 \ln(N_0) \qquad \text{(Eq 46)}$$

The equilibrium cluster size distribution can be deduced from this free energy. This distribution is obtained by minimizing Eq 46 under the constraint that the total number of atoms included in the clusters is fixed. Therefore, a Lagrange multiplier, μ, is introduced, and the grand canonical free energy is defined:

$$G - \mu \sum_{n=1}^{\infty} nN_n \qquad \text{(Eq 47)}$$

The minimization of this grand canonical free energy with respect to the variables N_n leads to the equilibrium cluster size distribution that should check the equation:

$$\frac{N_n^{\text{eq}}}{N_0 - \sum_{n=1}^{\infty} N_n^{\text{eq}}} = \exp\left(-\frac{G_n - n\mu}{kT}\right) \qquad \text{(Eq 48)}$$

The assumption of noninteracting clusters used to derive this equation is only valid in the dilute limit. It is therefore reasonable to neglect in Eq 48 the sum appearing in the rightside denominator compared to the number of accessible sites N_0. At equilibrium, the atomic fraction of clusters containing n atoms is then:

$$C_n^{\text{eq}} = \frac{N_n^{\text{eq}}}{N_s} = C_0 \exp\left(-\frac{G_n - n\mu}{kT}\right) \qquad \text{(Eq 49)}$$

where $C_0 = N_0/N_s$, and N_s is the total number of sites. For homogeneous nucleation, all sites can act as nucleation centers: $N_0 = N_s$ and $C_0 = 1$. Sometimes, Eq 49 is written in its equivalent form:

$$C_n^{\text{eq}} = C_0\left(\frac{C_1^{\text{eq}}}{C_0}\right)^n \exp\left(-\frac{G_n - nG_1}{kT}\right) \qquad \text{(Eq 50)}$$

The quantities $G_n - n\mu$ and $G_n - nG_1$ should nevertheless not be confused; the first one is the cluster formation free energy in a cluster gas characterized by the parameter μ, whereas the last one is the energy difference between the cluster and the equivalent number of monomers. The following uses Eq 49 because it

allows a direct link with the capillary approximation used by the classical nucleation theory.

It is interesting to understand the physical meaning of the Lagrange multiplier μ appearing in Eq 49. At equilibrium, the grand canonical free energy (Eq 47) is at a minimum. Then, for all sizes n:

$$\mu = \frac{1}{n}\frac{\partial G}{\partial N_n} \qquad \text{(Eq 51)}$$

To calculate this derivative, the total number of atoms of type i is introduced:

$$M_i = y_i^{\text{e}} \sum_{n=1}^{\infty} nN_n \qquad \text{(Eq 52)}$$

Equation 51 is equivalent to:

$$\mu = \frac{1}{n}\sum_i \frac{\partial G}{\partial M_i}\frac{\partial M_i}{\partial N_n}$$
$$= \sum_i \mu_i^0 y_i^{\text{e}} \qquad \text{(Eq 53)}$$

which uses the definition of the chemical potential—first derivative of the total free energy with respect to the number of atoms. Therefore, the Lagrange multiplier is nothing else than the chemical potentials of the different atomic species modulated by their atomic fraction. The fact that only one Lagrange multiplier is needed, and not one for each constituent, is a consequence of the initial assumption that the clusters have a fixed composition corresponding to the equilibrium one, y_i^{e}. Using the expression (Eq 40) of the cluster free energy and the definition (Eq 41) of the volume contribution, the equilibrium cluster size distribution given by the capillary approximation is recovered:

$$C_n^{\text{eq}} = C_0 \exp\left(-\frac{\Delta G_n}{kT}\right) \qquad \text{(Eq 54)}$$

with:

$$\Delta G_n = n\Delta G^{\text{nuc}} + n^{2/3}\left(36\pi\Omega_1^2\right)^{1/3}\sigma_n \qquad \text{(Eq 55)}$$

The nucleation free energy has the same expression as the one used in classical nucleation theory (Eq 4), but now the interface free energy depends on the cluster size.

It should be stressed that the cluster gas approximation is a thermodynamic model by itself; thermodynamic quantities such as chemical potentials are results and not input parameters of the model (Ref 49). This has important consequences for the kinetic approach of nucleation developed in the next section; in contrast with classical nucleation theory, one does not need to calculate the nucleation driving force to input it in the modeling.

One can use this cluster gas thermodynamic model to calculate the composition of the parent phase at the coexistence point between the parent and the nucleating phase, that is, the solubility limit. This coexistence point is defined by the

equality of the chemical potentials μ_i^0 and μ_i^{e}. The nucleation free energy is thus null, and only the interface contributes to the cluster formation free energy (Eq 55). At the coexistence point, the composition of the parent phase is then:

$$x_i^{\text{e}} = y_i^{\text{e}} \sum_{n=1}^{\infty} n \exp\left(-\frac{n^{2/3}(36\pi\Omega_1)^2\sigma_n}{kT}\right) \qquad \text{(Eq 56)}$$

The interface free energy fixes the solubility limit in the parent phase. This interface free energy is actually the key parameter of the nucleation kinetic approach. Even if its dependence on the cluster size is small, it is generally important to take it into account, because all thermodynamic quantities derive from it, and it enters in exponential terms such as in Eq 56.

Cluster Dynamics

For the sake of simplicity, in the following subsections homogeneous nucleation is considered. All monomers can be assumed equivalent; one does not need to distinguish between monomers lying on nucleation sites and free monomers.

Master Equation. Kinetics is described thanks to a master equation that gives the time evolution of the cluster size distribution. In many cases, one can assume that only monomers migrate. Therefore, this assumption is considered first and later the case where all clusters are mobile. When only monomers can migrate, the probability of observing a cluster containing n atoms obeys the differential equations:

$$\frac{\partial C_n}{\partial t} = J_{n-1\to n} - J_{n\to n+1} \quad \forall n \geq 2$$
$$\frac{\partial C_1}{\partial t} = -2J_{1\to 2} - \sum_{n\geq 2} J_{n\to n+1} \qquad \text{(Eq 57)}$$

where $J_{n\to n+1}$ is the cluster flux from the class of size n to the class $n+1$. This flux can be written:

$$J_{n\to n+1} = \beta_n C_n - \alpha_{n+1} C_{n+1} \qquad \text{(Eq 58)}$$

where β_n is the probability per unit time for one monomer to impinge on a cluster of size n, and α_n is the probability for one monomer to leave a cluster of size n.

Condensation Rate. Expression of the condensation rate β_n can be obtained from physical considerations. This condensation rate must be proportional to the monomer concentration and can generally be written:

$$\beta_n = b_n C_1 \qquad \text{(Eq 59)}$$

where b_n is an intrinsic property of the cluster of size n. In the ballistic regime, this factor is proportional to the surface of the cluster and to the jump frequency, Γ_1, of the monomer to impinge on the cluster. In the diffusion regime, this factor is proportional to the cluster radius and to the monomer diffusion coefficient, D_1. A general expression of the condensation rate, covering the ballistic and the diffusion regime,

has been proposed by Waite (Ref 24), who obtained:

$$\beta_n = 4\pi \frac{R_n^2}{R_n + \kappa} \frac{D_1}{\Omega_1} C_1 \qquad \text{(Eq 60)}$$

where Ω_1 is the monomer volume, and R_n is the cluster capture radius. It can be assumed that this radius is close to the one corresponding to the more compact cluster shape, that is, a sphere, leading to:

$$R_n = \left(\frac{3n\Omega_1}{4\pi}\right)^{1/3} \qquad \text{(Eq 61)}$$

The distance, κ, is given by the relation:

$$\kappa = \frac{D_1}{\lambda_1 \Gamma_1} \qquad \text{(Eq 62)}$$

where λ_1 is the distance corresponding to the monomer last jump to become attached to the cluster. If $R_n \ll \kappa$, one recovers the expression of the condensation rate in the ballistic regime, and in the diffusive regime if $R_n \gg \kappa$. Equation 60 therefore shows that condensation on small clusters is generally controlled by ballistic reactions, and condensation on big clusters by diffusion.

The expressions used by classical nucleation theory for the condensation rate (Eq 12 and 13) are similar to the ballistic and diffusion limits of Eq 62. Nevertheless, a difference appears because the condensation rate of the classical nucleation theory is proportional to the solute concentration and not to the monomer concentration, as in Eq 62. It thus makes use of the total solute diffusion coefficient or jump frequency and not of the monomer diffusion coefficient or jump frequency. For a dilute system, one can consider that all the solute is contained in monomers. The condensation rates used by both approaches are then equivalent. However, the difference may be important for more concentrated systems. This point has been thoroughly discussed by Martin (Ref 49), who showed the equivalence in the dilute limit.

Evaporation Rate. By contrast with the condensation rate, the evaporation rate, α_n cannot generally be obtained directly. It has to be deduced from β_n using the equilibrium cluster size distribution (Eq 49). The evaporation rate is obtained assuming that it is an intrinsic property of the cluster and does not depend on the embedding system. Therefore, it is assumed that the cluster has enough time to explore all its configurations between the arrival and the departure of a monomer. This assumption is coherent with the fact that the clusters are only described through their sizes. Thus, α_n should not depend on the saturation of the embedding system. It could be obtained, in particular, by considering any undersaturated system. Such a system is stable, and there should be no energy dissipation. This involves all fluxes $J_{n \to n+1}$

equaling zero. Using Eq 58, the following is obtained:

$$\alpha_{n+1} = \bar{\alpha}_{n+1}(\mu) = \bar{\beta}_n(\mu) \frac{\bar{C}_n(\mu)}{\bar{C}_{n+1}(\mu)} \qquad \text{(Eq 63)}$$

where overlined quantities are evaluated in the system at equilibrium characterized by its effective chemical potential, μ. In particular, the cluster size distribution is the equilibrium relation given by Eq 49. Using the expression (Eq 59) for the condensation rate, this finally leads to the following expression for the evaporation rate:

$$\alpha_{n+1} = b_n C_0 \exp[(G_{n+1} - G_n - G_1)/kT] \qquad \text{(Eq 64)}$$

Because the condensation rate varies linearly with the monomer concentration, the contribution of the effective chemical potential cancels out in the expression (Eq 63) of α_n. The starting assumption is recovered; the evaporation rate depends only on the cluster free energy and not on the overall state of the cluster gas characterized by the effective chemical potential, μ. Using the generalized capillary approximation (Eq 55), one can show that the evaporation rate actually depends only on the cluster interface free energy:

$$\alpha_{n+1} = b_n C_0 \exp\Big\{\left(36\pi\Omega_1^2\right)^{1/3} \\ \times [(n+1)^{2/3}\sigma_{n+1} - n^{2/3}\sigma_n - \sigma_1]/kT\Big\} \qquad \text{(Eq 65)}$$

The evaporation rate is then independent of the nucleation free energy, ΔG^{nuc}, which does not appear in any parameter. The nucleation free energy is implicit in cluster dynamics; there is no need to know it, but, if needed, one can calculate it from the cluster gas thermodynamic. This is in contrast with classical nucleation theory, where the nucleation free energy is an input parameter. On the other hand, cluster dynamics is very sensitive to the interface free energy as it appears in an exponential in the expression (Eq 65) of the evaporation rate. It is very important to have a correct evaluation of this interface free energy, especially of its variations with the cluster size, at least for small sizes.

In this approach, the evaporation rate is derived assuming that it is an intrinsic property of the cluster. Sometimes, one derives this parameter assuming instead that a hypothetical constrained equilibrium exists for the clusters in the supersaturated system; the equilibrium cluster size distribution (Eq 49) is taken to hold, although the system is supersaturated and cannot be at equilibrium. The evaporation rate is then obtained by imposing a detailed balance for Eq 58 with respect to this constrained equilibrium. Comparison with atomic simulation of the magnetization reversal of an Ising model (Ref 20) has shown that this constraint equilibrium assumption is good. The same conclusion

was reached for subcritical clusters in the case of precipitation in the solid state (Ref 50). Katz and Wiedersich (Ref 51) pointed out that this constrained equilibrium assumption generally leads to the same expression of the evaporation rate as the intrinsic property assumption. In particular, this is true when the condensation rate varies linearly with the monomer concentration, as is the case here (Eq 59).

When the growth and decay of clusters is controlled by a reaction at the interface (ballistic regime), it is also possible to directly compute the condensation and evaporation rates (Ref 44). An atomistic model is used to describe the physical process at the atomic scale, and the corresponding rates are obtained by thermal averaging through Monte Carlo sampling. Detailed balance is now imposed at the atomic scale. This ensures that the detailed balance at the cluster scale, as given by Eq 63, is also checked. A huge computational effort is required, but this could be optimized by calculating the cluster interface free energies and their condensation and evaporation rates at the same time.

Numerical Scheme. The evolution of the cluster size distribution is obtained by integrating the set of equations (Eq 57). A direct approach can become cumbersome because the number of differential equations varies linearly with the size of the largest cluster. The maximum size of the cluster that can be considered is therefore limited by the number of differential equations that can be integrated. This problem can be circumvented by noticing that a detailed description is important only for small cluster sizes where quantities vary rapidly. For large sizes, variations are smoother, and an approximated description can be used. The easiest approach to do so is to consider that the size n is now a continuous variable. One can then develop Eq 57 and 58 to the second order about n, and the system evolution is described by the Fokker-Planck equation (Ref 52):

$$\frac{\partial C_n}{\partial t} = -\frac{\partial}{\partial n}[(\beta_n - \alpha_n)C_n] + \frac{1}{2}\frac{\partial^2}{\partial n^2}[(\beta_n + \alpha_n)C_n] \qquad \text{(Eq 66)}$$

This continuous equation can be solved numerically by discretizing the continuous variable n. The best way to handle large cluster sizes is to use a varying increment greater than 1 and increasing with the cluster size. A convenient solution is an increment growing at a constant rate λ. The variable n is then discretized according to:

$$n_j = j, \quad \forall j \leq n_d$$
$$n_j = n_d + \frac{1 - \lambda^{j-n_d}}{1 - \lambda}, \quad \forall j \geq n_d \qquad \text{(Eq 67)}$$

where n_d is the number of classes for which the discrete equation (Eq 57) is used. Above this size, one integrates instead the discretized version of Eq 66, which is:

$$\frac{\partial C_{n_j}}{\partial t} = \frac{1}{n_{j+1} - n_{j-1}} \left[\left(\beta_{n_j} - \alpha_{n_j} \right) + \frac{\beta_{n_j} + \alpha_{n_j}}{n_j - n_{j-1}} \right.$$
$$\left. - \frac{\partial}{\partial n} \left(\beta_{n_j} + \alpha_{n_j} \right) \right] C_{n_{j-1}}$$
$$+ \left[-\frac{\beta_{n_j} + \alpha_{n_j}}{\left(n_{j+1} - n_j \right) \left(n_j - n_{j-1} \right)} - \frac{\partial}{\partial n} \left(\beta_{n_j} - \alpha_{n_j} \right) \right.$$
$$\left. - \frac{1}{2} \frac{\partial^2}{\partial n^2} \left(\beta_{n_j} + \alpha_{n_j} \right) \right] C_{n_j}$$
$$+ \frac{1}{n_{j+1} - n_{j-1}} \left[-\left(\beta_{n_j} - \alpha_{n_j} \right) + \frac{\beta_{n_j} + \alpha_{n_j}}{n_{j+1} - n_j} \right.$$
$$\left. + \frac{\partial}{\partial n} \left(\beta_{n_j} + \alpha_{n_j} \right) \right] C_{n_{j+1}} \tag{Eq 68}$$

The evolution of the monomer concentration is approximated by:

$$\frac{\partial C_1}{\partial t} = -2\beta_1 C_1 + \alpha_2 C_2 + \sum_{j \geq 2}^{n_d} \left(\alpha_{n_j} - \beta_{n_j} \right) C_{n_j}$$
$$+ \sum_{j \geq n_d} \left(\alpha_{n_j} - \beta_{n_j} \right) \frac{n_{j+1} - n_{j-1}}{2} C_{n_j} \tag{Eq 69}$$

This numerical scheme is simple and allows large cluster sizes to be reached with a reasonable number of differential equations. Typically, it is possible to simulate clusters containing more than 4 million atoms by using 100 discrete classes and 400 continuous classes with a growing increment rate $\lambda = 1.03$. It should nevertheless be mentioned that this numerical scheme does not strictly conserve the matter. By using reasonable values for the discretization parameters λ and n_d, the losses are generally insignificant, but, in any case, they must be checked afterward to see if they are acceptable. One must also verify that the concentration of the largest size has not evolved at the end of the simulation.

Another numerical approach has been proposed by Kiritani (Ref 53) to solve the set of differential equations (Eq 57) while allowing large cluster sizes to be reached. His grouping method consists of replacing a group of master equations by only one equation representing the class. It assumes that the number of clusters of each size in a group is the same and that the condensation and evaporation rates for clusters in a group do not vary. Unfortunately, it has been shown that the result can be very bad if the grouping is not carried out properly (Ref 54). Furthermore, as in the previous scheme, it does not strictly conserve the matter, even with an optimized grouping. Golubov et al. (Ref 55) proposed a new grouping method that can conserve the matter. For this purpose, the first and second moments of each group are considered, and two equations for each class are obtained. The first moment equation controls the time evolution of the cluster size distribution, and the second moment equation ensures the matter conservation. Such a numerical scheme therefore requires twice as many equations as the one proposed previously.

The Link with Classical Nucleation Theory

The main results of classical nucleation theory have actually been derived from cluster dynamics, that is, from the master equation (Eq 57) describing the time evolution of the cluster population. This derivation is interesting because it allows a better understanding of the assumptions behind the classical nucleation theory. Moreover, it provides insights into how this theory can be further developed to broaden the range where it applies. In the following subsection, the definition of the critical size in cluster dynamics is compared with the classical ones, and then it is shown how the steady-state nucleation rate and the corresponding cluster size distribution can be derived from the master equation. The derivation of the incubation time is not given here but can be found in Ref 26 and 27, for instance.

Critical Size. Subcritical clusters are unstable; they have a higher probability to decay than to grow. On the contrary, supercritical clusters are stable and have a higher probability to grow than to decay. The critical size n^* is then defined as the size for which the condensation rate equals the evaporation rate:

$$\beta_{n^*} = \alpha_{n^*} \tag{Eq 70}$$

This definition is actually different from the one used by the classical nucleation theory, where the critical size is the size at which the cluster formation free energy is maximum. One can show that these two definitions are consistent and lead to the same expression in the limit of large cluster sizes. To do so, rewrite Eq 70 using the expressions of the condensation rate (Eq 59) and of the evaporation rate (Eq 64):

$$b_{n^*} C_1 = b_{n^*-1} C_0 \exp\left(\frac{G_{n^*} - G_{n^*-1} - G_1}{kT} \right) \tag{Eq 71}$$

Then, assume that monomers are at local equilibrium; their concentration C_1 obeys the equilibrium cluster size distribution (Eq 49). One can thus eliminate in Eq 71 the monomer free energy G_1:

$$b_{n^*} \exp\left(\frac{\mu}{kT} \right) = b_{n^*-1} \exp\left(\frac{G_{n^*} - G_{n^*-1}}{kT} \right) \tag{Eq 72}$$

Using the definition $\Delta G_n = G_n - n\mu$ of the cluster formation free energy, Eq 72 can be rewritten:

$$\frac{b_{n^*-1}}{b_{n^*}} \exp\left(\frac{\Delta G_{n^*} - \Delta G_{n^*-1}}{kT} \right) = 1 \tag{Eq 73}$$

Using Eq 55 to express the cluster formation free energy, one finally obtains that the critical size verifies:

$$\frac{b_{n^*-1}}{b_{n^*}} \exp\left\{ \frac{\begin{array}{c} \Delta G^{\mathrm{nuc}} + \left(36\pi\Omega_1^2 \right)^{1/3} \\ \times \left[n^{*2/3} \sigma_{n^*} - (n^*-1)^{2/3} \sigma_{n^*-1} \right] \end{array}}{kT} \right\} = 1 \tag{Eq 74}$$

To go further, one needs to take the limit corresponding to large cluster sizes. One can then neglect the size dependence of the condensation rate prefactor, $b_{n^*-1} \sim b_{n^*}$, and of the cluster interface free energy, $\sigma_{n^*-1} \sim \sigma_{n^*} \sim \sigma$. At the critical size, one should therefore check:

$$\Delta G^{\mathrm{nuc}} + \left(36\pi\Omega_1^2 \right)^{1/3} \left[n^{*2/3} - (n^*-1)^{2/3} \right] \sigma = 0 \tag{Eq 75}$$

A limited expansion of Eq 75 for large sizes leads to the result:

$$n^* = \left[-\frac{2}{3} \frac{\left(36\pi\Omega_1^2 \right)^{1/3} \sigma}{\Delta G^{\mathrm{nuc}}} \right]^3 \tag{Eq 76}$$

One therefore recovers Eq 5 of the critical size with a geometric factor A corresponding to spherical clusters. The critical size considered by classical nucleation theory corresponds to the one of cluster dynamics in the limit of large cluster sizes. However, when the critical size is small, both definitions may differ. This coherence of both definitions at large size and this deviation at small sizes has been observed in atomic simulations (Ref 20).

The Steady-State Nucleation Rate. One can calculate the steady-state nucleation rate, J^{st}, corresponding to the master equation (Eq 57). To do so, one must make two assumptions:

- There is a small size below which clusters have their equilibrium concentration, given by Eq 49. Clusters smaller than the critical size appear and disappear spontaneously through thermal fluctuations, and their concentrations stay roughly at equilibrium. The smaller the cluster, the better this assumption. The most convenient choice is therefore to impose thermal equilibrium for monomers:

$$C_1(t) = C_1^{\mathrm{eq}} = C_0 \exp\left(-\frac{G_1 - \mu(t)}{kT} \right) \tag{Eq 77}$$

- There is a maximum cluster size, N, above which the cluster concentration remains null: $C_N(t) = 0$. This assumption cannot be checked for a true steady state without invoking a demon that removes clusters that appear at the size N and dissolves them into monomers. Nevertheless, one can always define at a given time a size large enough so that the cluster distribution did not propagate to this size.

By definition, the steady-state nucleation rate can be calculated at any cluster size n. At the

steady-state, all cluster concentrations remain constant. As a consequence:

$$\frac{\partial J_{n \to n+1}}{\partial n} = 0 \qquad \text{(Eq 78)}$$

and the steady-state nucleation rate can be calculated at any given cluster size. Using the expression of the cluster flux (Eq 58) with Eqs 59 and 64 for the condensation and evaporation rates, one obtains:

$$J^{\text{st}} = b_n \left\{ C_1 C_n - C_{n+1} C_0 \exp\left[\frac{G_{n+1} - G_n - G_1}{kT} \right] \right\},$$

$$n = b_n C_1 \exp\left[-\frac{G_n - n\mu}{kT} \right] \left\{ C_n \exp\left[\frac{G_n - n\mu}{kT} \right] \right.$$

$$\left. - C_{n+1} \exp\left[\frac{G_{n+1} - (n+1)\mu}{kT} \right] \right\} \qquad \text{(Eq 79)}$$

This equation uses the fact that monomers are at equilibrium (Eq 77) to go from the first to the second line. After rearranging the terms between the left and right sides, a sum between a minimal and a maximal size is derived:

$$\sum_{n=n_1}^{n_2} \frac{J^{\text{st}}}{b_n C_1 C_0 \exp\left[-\frac{G_n - n\mu}{kT} \right]}$$

$$= \frac{C_{n_1}}{C_0 \exp\left[-\frac{G_{n_1} - n_1\mu}{kT} \right]} - \frac{C_{n_2+1}}{C_0 \exp\left[-\frac{G_{n_2+1} - (n_2+1)\mu}{kT} \right]} \qquad \text{(Eq 80)}$$

$n_1 = 1$ is chosen so that the first term on the right side is equal to 1. With $n_2 = N - 1$, the second term is null; it is assumed $C_N = 0$, and the exponential is tending to ∞ for high enough N. This results in:

$$J^{\text{st}} = C_1 C_0 \frac{1}{\sum_{n=1}^{N-1} \frac{1}{b_n} \exp\left[\frac{G_n - n\mu}{kT} \right]} \qquad \text{(Eq 81)}$$

This gives an exact expression of the steady-state nucleation rate under both of the previous assumptions.

The sum appearing in Eq 81 can be easily evaluated. To do so, a continuous approximation is made to transform the sum into an integral. The cluster formation free energy, $\Delta G_n = G_n - n\mu$, presents a maximum at the critical size n^*. As a consequence, the main contribution to the integral arises from sizes around the critical size and can be evaluated by a Taylor expansion around n^*. Finally, neglecting the variations of b_n in front of the exponential leads to:

$$J^{\text{st}} = C_1 C_0 b_{n^*}$$

$$\times \frac{1}{\int_1^{N-1} \exp\left[\left(\Delta G_{n^*} + \frac{1}{2} \frac{\partial^2 \Delta G_n}{\partial n^2} \Big|_{n=n^*} (n - n^*)^2 \right) \Big/ kT \right] dn} \qquad \text{(Eq 82)}$$

Changing the integration limits in $-\infty$ and $+\infty$, the result of classical nucleation theory is recovered:

$$J^{\text{st}} = \beta^* Z C_0 \exp\left(-\frac{\Delta G^*}{kT} \right) \qquad \text{(Eq 83)}$$

where $\beta^* = C_1 b_{n^*}$, $\Delta G^* = \Delta G_{n^*}$ and the Zeldovich factor is given by Eq 9.

The Steady-State Cluster Size Distribution. Once the steady-state nucleation rate is known, one can easily obtain the corresponding cluster size distribution. Equation 80 is again used with the limits $n_2 = N - 1$, so that the last term on the right side is still null, and with $n_1 = n$, the size for calculating the cluster concentration. This leads to the result:

$$C_n^{\text{st}} = C_0 \exp\left[-\frac{G_n - n\mu}{kT} \right] \sum_{j=n}^{N} \frac{J^{\text{st}}}{b_n C_1 C_0 \exp\left[-\frac{G_j - j\mu}{kT} \right]} \qquad \text{(Eq 84)}$$

Similar to the steady-state nucleation rate, the sum can be evaluated by making a continuous approximation, developing the cluster formation free energy around the critical size, and considering the limit $N \to \infty$. One obtains:

$$C_n^{\text{st}} = C_0 \exp\left[-\frac{G_n - n\mu}{kT} \right] \frac{J^{\text{st}}}{C_1 C_0 b_{n^*}}$$

$$\int_n^\infty \exp\left[\left(\Delta G_{n^*} + \frac{1}{2} \frac{\partial^2 \Delta G_j}{\partial j^2} (j - n^*)^2 \right) \Big/ kT \right] dj$$

$$= \frac{1}{2} \text{erfc}[\sqrt{\pi} Z (n - n^*)] C_0 \exp\left[-\frac{G_n - n\mu}{kT} \right] \qquad \text{(Eq 85)}$$

The stationary distribution therefore corresponds to the equilibrium one, reduced by a factor varying from 0 for large sizes to 1 for small sizes. Well below the critical size, that is, for $n \leq n^* - \Delta n/2$ with Δn given by Eq 10, this factor differs only slightly from 1, and the stationary distribution corresponds to the equilibrium one. At the critical size n^*, this factor is exactly one-half, and in the vicinity of n^*, the stationary distribution can be approximated with Eq 8.

Discussion. This derivation of quantities predicted by classical nucleation theory from cluster dynamics formalism enlightens the approximations made by this theory. It assumes that the supersaturation is not too high, so that the critical size is large enough. This allows one to consider the size as a continuous instead of a discrete variable and to make a finite expansion of key parameters around the critical size. Classical nucleation theory may therefore appear as more restricted than the kinetic approach based on the master equation (Eq 57), but the situation is not so simple.

One severe restriction of cluster dynamics is the thermodynamic model on which it relies. It is based on the cluster gas model of Frenkel (Ref 48), which is valid for a dilute system. Strictly speaking, cluster dynamics should only be used in the dilute case. If one wants to study more concentrated systems, the cluster gas model must be extended. Such an extension has been performed by Lépinoux (Ref 56) and is presented in the section "Configurational Frustrations between Clusters" in this article. On the

other hand, classical nucleation theory does not rely on the cluster gas thermodynamic model. Instead, it makes use of the nucleation driving force that may be calculated with any thermodynamic model, in particular, one better suited to concentrated systems. Therefore, it is not a problem to use the classical nucleation theory to study concentrated systems as long as one correctly calculates the nucleation driving force.

Both formalisms also differ in the way they describe the parent and the nucleating phases. In the classical theory, one differentiates both phases, and nucleation is described through hetero-phase fluctuations corresponding to precipitates embedded in the parent phase. Such a differentiation does not appear in the kinetic approach, where one deals with only one system that is described as a gas of clusters having a fixed stoichiometry and embedded in a pure solvent. This description difference may become relevant when modeling concentrated systems, because the values of the input parameters may then differ according to the chosen, thermodynamic or kinetic, approach. This is the case, for instance, of the interface free energy. In cluster dynamics, this corresponds to the energy cost of an interface between a cluster with a fixed stoichiometry and the pure solvent, whereas in classical nucleation theory, one should consider that the precipitate and the parent phase are not pure and that solubility exists in both phases. The concentration appearing in the expression of the condensation rate may also differ between both approaches, as already quoted in the previous section, "Condensation Rate," in this article. This is either the monomer concentration (cluster dynamics) or the total solute concentration (classical nucleation theory).

All these subtle differences between cluster dynamics and classical nucleation theory have been discussed by Martin (Ref 49) in the case of precipitation in the solid state. He showed that both approaches were consistent and led to the same expressions in the dilute limit.

Extensions of Cluster Dynamics

The master equation (Eq 57) can be modified to describe nucleation under less restricted conditions than the ones of the previous subsections and then to build extensions of the cluster dynamics. In particular, the assumptions that only monomers can react and that clusters have a fixed stoichiometry corresponding to the equilibrium nucleating phase can be removed.

Mobile Clusters. Until now, it was assumed that only monomers are mobile. This assumption is not always valid. There is no reason to think, for instance, that all clusters except monomers are immobile in solidification. Diffusion of small clusters can also happen in solid-solid phase transformations. An interesting example is copper precipitation in iron, where atomic simulations have revealed that clusters containing up to several tens of copper atoms can be much more mobile than individual copper atoms (Ref 57). The master equation should

therefore be modified to account for reactions involving clusters larger than monomers. Such a generalization of the cluster dynamics formalism has been performed by Binder and Stauffer (Ref 4–6).

The probability of observing a cluster containing n atoms now obeys the generalized master equation:

$$\frac{\partial C_n}{\partial t} = \frac{1}{2}\sum_{n'=1}^{n-1} J(n-n', n' \to n) - \sum_{n'=1}^{\infty} J(n, n' \to n + n') - J(n, n \to 2n)$$

(Eq 86)

where $J(n,n' \to n + n')$ is the cluster flux corresponding to the reaction between the classes n and n' to the class $n + n'$. The factor ½ appearing in Eq 86 accounts for overcounting the pairs $\{n-n',n'\}$ in the summation. In this equation, one should not forget that the reactions $n + n \rightleftharpoons 2n$ involve two clusters of size n. When reactions are limited to reactions involving monomers, n' can only take the value 1, and $n-1$ in the first sum and 1 in the second sum; the classical master Eq 57 of cluster dynamics is recovered.

The cluster flux is the difference between the condensation of two clusters of sizes n and n' and the splitting of a cluster of size $n + n'$ into two clusters of sizes n and n':

$$J(n, n' \to n + n') = b(n, n' \to n + n')C_n C_{n'} - \alpha(n + n' \to n, n')C_{n+n'}$$

(Eq 87)

One then obtains an expression for the absorption coefficient $b(n, n')$. If the reaction is limited by the cluster diffusion, this coefficient is given by (Ref 24):

$$b(n, n' \to n + n') = 4\pi R_{n,n'}\frac{D_n + D_{n'}}{\Omega_1}, \quad \forall n \neq n'$$

$$b(n, n \to 2n) = 4\pi R_{n,n}\frac{D_n}{\Omega_1}$$

(Eq 88)

where $R_{n,n'}$ is a capture radius and can be approximated by the sum of the two reacting cluster radii. In the expression (Eq 13) used in the classical nucleation theory, this capture radius was identified with the radius of the critical cluster, and the monomer radius was neglected.

If the diffusion coefficients of the n-mers are not known, one can use an approximation proposed by Binder et al. (Ref 58, 59). They simply consider that cluster diffusion is due to jumps of atoms located at the interface. When an atom jumps over a distance r_s with a frequency Γ_s, the center of gravity of the cluster jumps over r_s/n. Since the number of possible jumps at the interface increases with its area as $n^{2/3}$, D_n depends on n as:

$$D_n = \Gamma_s\left(\frac{r_s}{n}\right)^2 n^{2/3} = D_1 n^{-4/3}$$

(Eq 89)

In the case of precipitation in the solid state, one should not forget that substitutional atoms diffuse through exchange with vacancies and that a vacancy enrichment at the cluster interface is possible. In such a case, Eq 89 must be corrected with a prefactor to consider the vacancy concentration at the interface (Ref 57). This vacancy segregation is the reason why clusters containing several copper atoms are more mobile than monomers in iron (Ref 57).

The evaporation rate is still obtained by assuming that it is an intrinsic property of the cluster (or imposing a constrained equilibrium), thus leading to:

$$\alpha(n + n' \to n, n') = b(n, n' \to n + n')C_0 \exp\left(\frac{G_{n+n'} - G_n - G_{n'}}{kT}\right)$$

(Eq 90)

All parameters are thus determined, and the master equation (Eq 86) can be numerically integrated.

Binder and Stauffer (Ref 4, 6) also extended classical nucleation theory to obtain expressions of the steady-state nucleation rate and the incubation time, taking into account the mobility of all clusters. They started from the master equation (Eq 86) and imposed the detailed balance corresponding to Eq 90. They obtained expressions similar to the classical ones—Eq 11 for the steady-state nucleation rate and Eq 14 for the incubation time—except that now the growing rate, β*, of the critical cluster incorporates contributions of all clusters. This growing rate is given by:

$$\beta^* = \sum_{n=1}^{n_c} b(n^*, n \to n + n^*)n^2 C_0 \exp\left(-\frac{G_n - n\mu}{kT}\right)$$

(Eq 91)

where n_c is a cut-off size corresponding to the correlation length of thermal fluctuations. It seems reasonable to identify this cut-off size with the critical size n^*. When only reactions involving monomers can occur, the sum in Eq 91 is limited to the term $n = 1$, and one recovers the classical growing rate $\beta^* = b_{n^*}C_1^{\text{eq}}$. When reactions involving other clusters are possible, this growing rate increases. The mobility of small clusters therefore leads to an increase of the nucleation rate and a decrease of the incubation time by the same factor.

Nonstoichiometric Clusters. Until now, it has been assumed that clusters have a fixed stoichiometry corresponding to the equilibrium of the nucleating phase. In some systems, the composition of the nucleating phase can vary. One therefore must extend cluster dynamics to allow the cluster stoichiometry to vary (Ref 7).

To illustrate such an extension of the formalism, consider the example of a system where the nucleating phase is composed of two elements, A and B, and assume that the composition can vary. A cluster is then a group of A and B atoms that are linked by a neighborhood relation. If the clusters are homogeneous (no segregation of one element at the interface, for instance), they can simply be described by two variables: the number i and j of elements A and B they contain. Therefore, $G_{i,j}$ is the free energy of such a cluster. If the system is undersaturated, one can show that the concentration of $\{i,j\}$ clusters is given by the distribution:

$$C_{i,j}^{\text{eq}} = C_0 \exp-\left(-\frac{G_{i,j} - i\mu_A - j\mu_B}{kT}\right)$$

(Eq 92)

where μ_A and μ_B are Lagrange multipliers ensuring matter conservation for A and B and are related to their chemical potentials.

It is assumed that only monomers are mobile. The time evolution of clusters containing i A elements and j B elements is then governed by the master equation:

$$\frac{\partial C_{i,j}}{\partial t} = J_{i-1,j\to i,j} - J_{i,j\to i+1,j} + J_{i,j-1\to i,j} - J_{i,j\to i,j+1},$$
$$\forall\{i,j\} \neq \{1,0\} \text{ and } \{i,j\} \neq \{0,1\}$$

$$\frac{\partial C_{1,0}}{\partial t} = -\sum_{i\geq 0}\sum_{j\geq 0} J_{i,j\to i+1,j} - J_{1,0\to 2,0}$$

$$\frac{\partial C_{0,1}}{\partial t} = -\sum_{i\geq 0}\sum_{j\geq 0} J_{i,j\to i,j+1} - J_{0,1\to 0,2}$$

(Eq 93)

Fluxes are written as a difference between the evaporation and the condensation of a monomer:

$$J_{i,j\to i+1,j} = b_{i,j\to i+1,j}C_{1,0}C_{i,j} - \alpha_{i+1,j\to i,j}C_{i+1,j}$$
$$J_{i,j\to i,j+1} = b_{i,j\to i,j+1}C_{0,1}C_{i,j} - \alpha_{i,j+1\to i,j}C_{i,j+1}$$

(Eq 94)

The condensation and evaporation rates are still linked by a detailed balance condition, leading to the relations:

$$\alpha_{i+1,j\to i,j} = b_{i,j\to i+1,j}\exp\left(\frac{G_{i+1,j} - G_{i,j} - G_{1,0}}{kT}\right)$$
$$\alpha_{i,j+1\to i,j} = b_{i,j\to i,j+1}\exp\left(\frac{G_{i,j+1} - G_{i,j} - G_{0,1}}{kT}\right)$$

(Eq 95)

One therefore needs a physical modeling of the condensation process to express the coefficients $b_{i,j\to i+1,j}$ and $b_{i,j\to i,j+1}$. The evaporation rates are then obtained by Eq 95, and the kinetics are obtained by integration of Eq 93.

Starting from the master equation (Eq 93), the classical nucleation theory has been extended to treat a multicomponent system. This was first performed by Reiss (Ref 7) for a binary system such as the one considered here and then extended by Hirschfelder (Ref 8) to a general multicomponent system. Both authors assumed that the growth of the critical nucleus was entirely driven by the free energy. It was realized later by Stauffer (Ref 9) that the growth direction in the $\{i,j\}$ plane may also be affected by the condensation coefficients, especially when coefficients corresponding to A and B condensation have very different values. He proposed an expression of the steady-state nucleation rate for the binary system that was then extended to a multicomponent system by

Trinkaus (Ref 10). All these approaches calculated the steady-state nucleation rate in the vicinity of the critical nucleus. Wu (Ref 11) instead defined a global nucleation rate that should correspond more closely to what can be measured experimentally. The following gives the expression of the steady-state nucleation rate for a binary system obeying the master equation (Eq 93) in the local approach as derived by Vehkamäki (Ref 12). Expressions in the more general case—multicomponent systems and mobile clusters other than monomers—can be found in the cited references.

The critical cluster corresponds to the saddle point of the cluster formation free energy $\Delta G_{i,j} = G_{i,j} - i\mu_A - j\mu_B$ appearing in the equilibrium distribution (Eq 92). It is thus defined by the equations:

$$\frac{\partial \Delta G_{i,j}}{\partial i} = 0 \quad \text{and} \quad \frac{\partial \Delta G_{i,j}}{\partial j} = 0 \qquad \text{(Eq 96)}$$

ΔG^* is the corresponding formation free energy, and H^* is the Hessian matrix calculated for the critical cluster:

$$H^* = \begin{pmatrix} \left.\frac{\partial^2 \Delta G_{i,j}}{\partial i^2}\right|_{\{i,j\}^*} & \left.\frac{\partial^2 \Delta G_{i,j}}{\partial i \partial j}\right|_{\{i,j\}^*} \\ \left.\frac{\partial^2 \Delta G_{i,j}}{\partial i \partial j}\right|_{\{i,j\}^*} & \left.\frac{\partial^2 \Delta G_{i,j}}{\partial j^2}\right|_{\{i,j\}^*} \end{pmatrix} \qquad \text{(Eq 97)}$$

This Hessian matrix has two eigenvalues. One of them is negative and gives the direction in the $\{i,j\}$ space corresponding to the maximal decrease of the critical cluster free energy. In the approach of Reiss and Hirschfelder, this direction corresponds to the nucleation flow. Nevertheless, one should generally take into account that the condensation rates for A and B elements may be different, because this will impact the direction of the nucleation flow. Therefore, a new matrix characterizing the condensation process for the critical cluster is defined:

$$B^* = \left.\begin{pmatrix} C_{1,0} b_{i,j \to i+1,j} & 0 \\ 0 & C_{0,1} b_{i,j \to i,j+1} \end{pmatrix}\right|_{\{i,j\}^*} \qquad \text{(Eq 98)}$$

The fact that this matrix is diagonal reflects the assumption that only reactions involving monomers are possible. The angle θ of the nucleation flow in the $\{i,j\}$ space is then defined by:

$$\tan\theta = \frac{-H_{11}^* B_{11}^* + H_{22}^* B_{22}^* - \sqrt{4 H_{12}^{*2} B_{11}^* B_{22}^* + \left(H_{11}^* B_{11}^* - H_{22}^* B_{22}^*\right)^2}}{2 H_{12}^* B_{11}^*} \qquad \text{(Eq 99)}$$

The equivalent of the Zeldovitch factor is given by:

$$Z = -\frac{H_{11}^* + 2 H_{12}^* \tan\theta + H_{22}^* \tan^2\theta}{(1 + \tan^2\theta)\sqrt{|\det(H^*)|}} \qquad \text{(Eq 100)}$$

and the average growth rate of the critical cluster by:

$$\beta^* = \frac{\det(B^*)}{B_{11}^* \sin^2\theta + B_{22}^* \cos^2\theta} \qquad \text{(Eq 101)}$$

With these definitions, the steady-state nucleation rate keeps its usual expression:

$$J^{st} = \beta^* Z C_0 \exp\left(-\frac{\Delta G^*}{kT}\right) \qquad \text{(Eq 102)}$$

Configurational Frustrations between Clusters. Cluster dynamics simulations rely on the cluster gas approximation derived in the section "Cluster Gas Thermodynamics" in this article. This thermodynamic approximation, initially introduced by Frenkel (Ref 48), is strictly valid only in the dilute limit. It indeed assumes that the space occupied by the clusters can be neglected when computing the configurational partition function (Eq 45) of the cluster gas; each cluster occupies only one site, no matter its size. Lépinoux (Ref 56) has shown that this approximation can be improved to properly take into account frustrations between clusters, that is, the space forbidden to a given cluster by other clusters. This allows the modeling of systems that are not as dilute as required by Frenkel's treatment.

Note that $V_{j,n}$ is the number of sites that a cluster of size j forbids to a cluster of size n. According to Lépinoux (Ref 56), the equilibrium cluster size distribution is given by:

$$C_n^{eq} = C_0 \exp\left(-\frac{G_n - n\mu}{kT}\right) \exp\left(-\sum_j C_j^{eq} V_{j,n}\right) \qquad \text{(Eq 103)}$$

or equivalently:

$$C_n^{eq} = C_0 \left(\frac{C_1^{eq}}{C_0}\right)^n \exp\left(-\frac{G_n - nG_1}{kT}\right) \exp\left[-\sum_j C_j^{eq}\left(V_{j,n} - nV_{1,n}\right)\right] \qquad \text{(Eq 104)}$$

It is clear that Frenkel's approximation corresponds to neglecting all exclusion volumes ($V_{j,n} = 0$). When exclusion volumes are considered, only an implicit expression of the size distribution is obtained; equilibrium cluster size concentrations, C_j^{eq}, are required to evaluate the right side of Eq 103 or 104. A self-consistent loop can be used to evaluate the equilibrium distribution, starting from the distribution given by Frenkel's approximation (Eq 49 or 50).

The exclusion volumes can be approximated by identifying a cluster of size n with a sphere of radius R_n. This leads to:

$$V_{j,n} = \frac{4\pi}{3}\left(R_j + R_n\right)^3 \qquad \text{(Eq 105)}$$

The radii R_n depend on the temperature because a cluster becomes less compact with higher temperatures due to its configurational entropy. Nevertheless, it can be reasonably assumed that these radii are close to the ones corresponding to the more compact cluster shape (Ref 56), and Eq 61 can be used.

The second step is to obtain the kinetic coefficients α_n and β_n. As previously mentioned, the condensation rate β_n is obtained by the proper physical modeling of the condensation process, leading to an expression of the form in Eq 59. However, it is no longer possible to assume that the evaporation rate is an intrinsic property of the cluster; the obtained expression would violate the assumption because of the frustration contribution in the cluster size distribution. The constrained equilibrium is not satisfactory either, because it leads to a diverging frustration correction and hence diverging evaporation rates in supersaturated systems. There is actually no framework that allows rigorously deriving the evaporation rate from the condensation rate, taking into account cluster frustrations. It seems that the most reasonable scheme is to consider that the classical expression (Eq 64) of the evaporation rate must be corrected from frustrations caused by the instantaneous cluster size distribution and not by a hypothetical equilibrium one:

$$\alpha_{n+1}(t) = b_n C_0 \exp\left(\frac{G_{n+1} - G_n - G_1}{kT}\right) \exp\left[\sum_j C_j(t)\left(V_{j,n+1} - V_{j,n} - V_{j,1}\right)\right] \qquad \text{(Eq 106)}$$

This set of condensation and evaporation rates ensures that the cluster distribution evolves toward the equilibrium distribution given by Eq 103 for subcritical clusters. Equation 106 clearly shows that the evaporation rate is no longer an intrinsic property of the cluster, because it now depends on the whole cluster distribution. Moreover, because this parameter depends on the instantaneous concentrations $C_j(t)$, it must be calculated at each time step. When the system is dilute, the frustration correction in Eq 106 becomes negligible, and the classical expression of the condensation rate is recovered. Comparisons with atomic simulations (Ref 50, 56) have shown that this treatment of cluster frustrations greatly improves the ability of cluster dynamics to describe nucleation kinetics for high supersaturations.

Limitations of the Cluster Description

The previous extensions of cluster dynamics have allowed the removal of two limitations of classical nucleation theories due to initial simplifying assumptions:

- Only monomers are mobile, and therefore, only reactions involving monomers are possible.
- The cluster stoichiometry is fixed and known a priori. It is assumed to correspond to the composition of the nucleating phase at equilibrium with the mother phase.

The extension to mobile clusters is quite straightforward, and that to nonstoichiometric clusters shows that it was possible to take into account a nonfixed cluster composition.

The composition of the nucleating cluster was found to be the one minimizing the work, ΔG^*, necessary to form them.

Nevertheless, some limitations still remain for this nucleation modeling approach. One of these limitations arises from the needed assumption that clusters are homogeneous. This assumption is induced by the fact that clusters are only described by the number of elements they contain. This is not always valid because segregation may occur in some systems; it can be more favorable for one element to lie at the interface between the cluster and the matrix instead of in the core of the cluster. In such a case, it is necessary to introduce at least one more parameter to describe the cluster structure. Binder and Stauffer (Ref 4) have extended cluster dynamics formalism to incorporate additional parameters describing cluster internal degrees of freedom, but the application of the formalism appears quite intricate.

Cahn and Hilliard (Ref 60) proposed a modeling approach different from the classical one presented here, which is based on a cluster description. Their approach agrees with the classical one at low supersaturations and underlines some limitations of the classical approach with increasing supersaturations. They showed that the work, ΔG^*, required to form a critical cluster becomes progressively less than that given by the classical theory and continuously approaches zero at the spinodal limit, thus for a finite supersaturation. By contrast, the classical theory predicts that this work becomes zero only for an infinite supersaturation (Eq 6). Moreover, the classical theory assumes that clusters are homogeneous and that their composition is the one minimizing the work, ΔG^*. Cahn and Hilliard showed that the composition at the center of the nucleus approaches that of the exterior mother phase when the supersaturation tends to the spinodal limit, and that the interface becomes more diffuse until eventually no part of the nucleus is even approximately homogeneous. The last disagreement found with the classical theory is the variation of the critical cluster size. They showed that this size first decreases, passes through a minimum, and then increases to become infinite when the supersaturation increases and approaches the spinodal limit. Nevertheless, some recent experiment observations (Ref 61, 62) have contradicted this last point, showing no divergence of the cluster critical size when approaching the spinodal limit.

Conclusions

Two different approaches based on an equivalent cluster description can therefore be used to model nucleation in a phase-separating system. In the classical nucleation theory, one obtains expressions of the nucleation rate and the incubation time. These expressions depend on a limited number of input parameters: the nucleation driving force, the interface free energy, and the condensation rate. On the other hand, the kinetic description of nucleation relies on a master equation. Cluster dynamics simulations, that is, the integration of this master equation, allow the time evolution of the cluster size distribution to be obtained. The input parameters needed by such simulations are the cluster condensation rates and the cluster free energies. At variance with classical nucleation theory, no external thermodynamic model is needed to calculate the nucleation driving force; cluster dynamics simulations possess their own thermodynamic model, the cluster gas. As shown previously, both approaches are intrinsically linked, but it is worth saying that they differ in the way they can be used to model the kinetics of phase transformations. Classical nucleation theory is able to model only the nucleation stage. To model the whole kinetics, one must couple this theory with classical descriptions of the growth and coarsening stage. Such a coupling can be done following the Wagner and Kampmann approach (Ref 63, 64). On the other hand, the cluster dynamics modeling approach is not restricted to the nucleation stage. It also predicts growth and coarsening kinetics. To conclude, this cluster approach is well suited when one knows what the nucleating new phase looks like. Such information is not always available a priori. One then must use other modeling techniques. These can be atomic simulations, such as molecular dynamics (Ref 65, 66), for condensation of a gas into a liquid or crystallisation of a liquid, or kinetic Monte Carlo (Ref 67–69) for solid-solid phase transformations, or phase-field simulations (see the Appendix at the end of this article). These simulations are computationally much more time-consuming and, as a consequence, are limited to the study of high enough supersaturations. Nevertheless, they can be very useful for understanding what happens in the nucleation stage and then building a classical model based on a cluster description and extending the range of supersaturations that can be simulated. Moreover, these atomic or phase-field simulations can be a convenient way to calculate the input parameters needed by classical theories.

Appendix—Phase-Field Simulations

The phase-field approach describes the different phases through continuous fields such as the atomic concentration or long-range-order parameters. The spatial and temporal evolution of the microstructure is then driven by differential equations obeyed by these fields. Because this technique is the object of the article "Phase-Field Modeling of Microstructure Evolution" in this Volume, this Appendix addresses how nucleation can be handled in such simulations.

The main advantage of phase-field simulations is that all spatial information on the microstructure is obtained. This is in contrast with classical approaches where a limited number of information is known, such as the flux of nucleating particles (classical nucleation theory) or the cluster size distribution (cluster dynamics). This may make the phase-field approach an attractive technique for modeling nucleation in specific situations. Indeed, such simulations perfectly take into account phase inhomogeneities. These inhomogeneities can be, for instance, a solute segregation in the vicinity of a defect such as a dislocation. Phase-field simulations therefore allow the description of heterogeneous nucleation associated with a variation of the driving force. Moreover, in the case of solid-solid phase transformations, the elastic energy is fully contained in the calculation of the system free energy (Ref 41). One therefore does not need a specific expression for the elastic self-energy of a nucleating particle nor for its elastic interaction with the existing microstructure. The correlated and collective nucleation due to elastic interaction between precipitates is naturally described. Two different roads have been proposed to include nucleation in phase-field simulations.

One can use the phase-field approach to calculate spatial variations of the concentrations and the order parameters describing the different phases as well as the inhomogeneous strain created by the microstructure. One then calculates the nucleation free energy as a function of the local phase fields and the local strain. Finally, the expression of the nucleation rate given by the classical theory is used to seed the phase-field simulations with new nuclei (Ref 70–72). In this way, one obtains a spatial variation of the nucleation rate caused by the microstructure inhomogeneities.

The phase-field approach offers another way to model nucleation without relying on the classical theory. One can add to the equations describing the phase-field evolution a stochastic term through a Langevin force to describe thermal fluctuations. This allows nucleation to proceed. Phase-field simulations can then naturally describe the spatial and temporal evolution of the microstructure, from the nucleation to the coarsening stage (Ref 73–75). Nevertheless, this description is usually only qualitative; to obtain a fully quantitative modeling, the amplitude of the Langevin force must be carefully set. In particular, it must depend on the coarse-graining size similar to the other ingredients of the simulation (chemical potentials, mobilities, stiffness coefficients) (Ref 76). Such phase-field simulations that naturally handle nucleation through thermal fluctuations suffer from the small time-step needed to catch the rare event of a nucleating particle. On the other hand, simulations using an explicit description of the nucleation do not have this drawback.

REFERENCES

1. M. Volmer and A. Weber, Keimbildung in Übersättigten Gebilden, *Z. Phys. Chem. (Leipzig)*, 119, 1926, p 277
2. L. Farkas, Keimbildungsgeschwindigkeit in Übersättigten Dämpfen, *Z. Phys. Chem. (Leipzig)*, 125, 1927, p 239
3. R. Becker and W. Döring, Kinetische Behandlung der Keimbildung in Übersättigten Dämpfen, *Ann. Phys. (Leipzig)*, 24, 1935, p 719
4. K. Binder and D. Stauffer, Statistical Theory of Nucleation, Condensation and Coagulation, *Adv. Phys.*, 25, 1976, p 343
5. K. Binder, Theory for the Dynamics of "Clusters." II. Critical Diffusion in Binary Systems and the Kinetics of Phase Separation, *Phys. Rev. B*, 15, 1977, p 4425
6. K. Binder, Theory of First-Order Phase Transitions, *Rep. Prog. Phys.*, 50, 1987, p 783
7. H. Reiss, The Kinetics of Phase Transitions in Binary Systems, *J. Chem. Phys.*, 18, 1950, p 840
8. J.O. Hirschfelder, Kinetics of Homogeneous Nucleation on Many-Component Systems, *J. Chem. Phys.*, 61, 1974, p 2690
9. D. Stauffer, Kinetic Theory of Two-Component ("Hetero-Molecular") Nucleation and Condensation, *J. Aerosol. Sci.*, 7, 1976, p 319
10. H. Trinkaus, Theory of the Nucleation of Multicomponent Precipitates, *Phys. Rev. B*, 27, 1983, p 7372
11. D.T. Wu, General Approach to Barrier Crossing in Multicomponent Nucleation, *J. Chem. Phys.*, 99, 1993, p 1990
12. H. Vehkamäki, *Classical Nucleation Theory in Multicomponent Systems*, Springer, Berlin, 2006
13. J.W. Cahn, Phase Separation by Spinodal Decomposition in Isotropic Systems, *J. Chem. Phys.*, **42**, 1965, p 93
14. K. Binder, Spinodal Decomposition, *Materials Science and Technology, A Comprehensive Treatment*, R.W. Cahn, P. Haasen, and E.J. Kramer, Ed., VCH, Weinheim, Vol. 5, 1991, pp. 405–471
15. G. Wulff, Zur Frage der Geschwindigkeit des Wachstums und der Auflösung der Kristallflächen, *Z. Kristallogr.*, 34, 1901, p 449
16. D.A. Porter and K.E. Easterling, *Phase Transformations in Metals and Alloys*, Chapman & Hall, London, 1992
17. E. Clouet, M. Nastar, and C. Sigli, Nucleation of Al₃Zr and Al₃Sc in Aluminum Alloys: From Kinetic Monte Carlo Simulations to Classical Theory, *Phys. Rev. B*, 69, 2004, p 064109
18. F. Soisson and G. Martin, Monte-Carlo Simulations of the Decomposition of Metastable Solid Solutions: Transient and Steady-State Nucleation Kinetics, *Phys. Rev. B*, 62, 2000, p 203
19. V.A. Shneidman, K.A. Jackson, and K.M. Beatty, Nucleation and Growth of a Stable Phase in an Ising-Type System, *Phys. Rev. B*, 59, 1999, p 3579
20. L. Maibaum, Phase Transformation near the Classical Limit of Stability, *Phys. Rev. Lett.*, 101, 2008, p 256102
21. D. Kashchiev, Solution of the Non-Steady State Problem in Nucleation Kinetics, *Surf. Sci.*, 14, 1969, p 209
22. D. Kashchiev, *Nucleation: Basic Theory with Applications*, Butterworth Heinemann, Oxford, 2000
23. G. Martin, The Theories of Unmixing Kinetics of Solid Solutions, *Solid State Phase Transformation in Metals and Alloys*, Les Éditions de Physique, Orsay, France, 1980, pp. 337–406
24. T.R. Waite, General Theory of Bimolecular Reaction Rates in Solids and Liquids, *J. Chem. Phys.*, 28, 1958, p 103
25. E. Clouet, A. Barbu, L. Laé, and G. Martin, Precipitation Kinetics of Al₃Zr and Al₃Sc in Aluminum Alloys Modeled with Cluster Dynamics, *Acta Mater.*, 53, 2005, p 2313
26. D.T. Wu, The Time Lag in Nucleation Theory, *J. Chem. Phys.*, 97, 1992, p 2644
27. V.A. Shneidman and M.C. Weinberg, Transient Nucleation Induction Time from the Birth-Death Equations, *J. Chem. Phys.*, 97, 1992, p 3629
28. K.F. Kelton, A.L. Greer, and C.V. Thompson, Transient Nucleation in Condensed Systems, *J. Chem. Phys.*, 79, 1983, p 6261
29. V.A. Shneidman and M.C. Weinberg, Induction Time in Transient Nucleation Theory, *J. Chem. Phys.*, 97, 1992, p 3621
30. J. Feder, K.C. Russell, J. Lothe, and G.M. Pound, Homogeneous Nucleation and Growth of Droplets in Vapours, *Adv. Phys.*, 15, 1966, p 111
31. D. Kashchiev, Nucleation at Existing Cluster Size Distributions, *Surf. Sci.*, 18, 1969, p 389
32. F. Berthier, B. Legrand, J. Creuze, and R. Tétot, Atomistic Investigation of the Kolmogorov-Johnson-Mehl-Avrami Law in Electrodeposition Process, *J. Electroanal. Chem.*, 561, 2004, p 37; Ag/Cu (001) Electrodeposition: Beyond the Classical Nucleation Theory, *J. Electroanal. Chem.*, 562, 2004, p 127
33. P. Maugis and M. Gouné, Kinetics of Vanadium Carbonitride Precipitation in Steel: A Computer Model, *Acta Mater.*, 53, 2005, p 3359
34. F. Ducastelle, *Order and Phase Stability in Alloys*, North-Holland, Amsterdam, 1991
35. E. Clouet and M. Nastar, Classical Nucleation Theory in Ordering Alloys Precipitating with L1₂ Structure, *Phys. Rev. B*, 75, 2007, p 132102
36. P. Spencer, A Brief History of CALPHAD, *Calphad*, 32, 2008, p 1
37. N. Saunders and A.P. Miodownik, *CALPHAD — Calculation of Phase Diagrams — A Comprehensive Guide*, Pergamon, Oxford, 1998
38. J.D. Eshelby, The Determination of the Elastic Field of an Ellipsoidal Inclusion, and Related Problems, *Proc. Roy. Soc. Lond. A*, 241, 1957, p 376
39. J.D. Eshelby, The Elastic Field Outside an Ellipsoidal Inclusion, *Proc. Roy. Soc. Lond. A*, 252, 1959, p 561
40. J.D. Eshelby, Elastic Inclusion and Inhomogeneities, *Progress in Solid Mechanics*, I.N. Sneddon and R. Hill, Ed., North Holland, Vol. 2, 1961, pp. 87–140
41. A.G. Khachaturyan, *Theory of Structural Transformations in Solids,* Wiley, New York, 1983
42. L. Zhang, L.-Q. Chen, and Q. Du, Morphology of Critical Nuclei in Solid-State Phase Transformations, *Phys. Rev. Lett.*, 98, 2007, p 265703
43. C. Shen, J.P. Simmons, and Y. Wang, Effect of Elastic Interaction on Nucleation: I. Calculation of the Strain Energy of Nucleus Formation in an Elastically Anisotropic Crystal of Arbitrary Microstructure, *Acta Mater.*, 54, 2006, p 5617
44. J. Lépinoux, Interfacial Reaction Rates and Free Energy of Cubic Clusters, *Philos. Mag.*, 85, 2005, p 3585
45. A. Perini, G. Jacucci, and G. Martin, Cluster Free Energy in the Simple-Cubic Ising Model, *Phys. Rev. B*, 29, 1984, p 2689; Interfacial Contribution to Cluster Free Energy, *Surf. Sci.*, 144, 1984, p 53
46. J.W. Gibbs, *Collected Works*, Vol. 1, *Thermodynamics*, Longmans-Green, New York, 1928
47. R.C. Tolman, The Effect of Droplet Size on Surface Tension, *J. Chem. Phys.*, 17, 1949, p 333
48. J. Frenkel, *Kinetic Theory of Liquids*, Dover Publications, New York, 1955
49. G. Martin, Reconciling the Classical Nucleation Theory and Atomic Scale Observations and Modeling, *Adv. Eng. Mater.*, 8, 2006, p 1231
50. J. Lépinoux, Modelling Precipitation in Binary Alloys by Cluster Dynamics, *Acta Mater.*, 57, 2009, p 1086
51. J.L. Katz and H. Wiedersich, Nucleation Theory without Maxwell Demons, *J. Colloid and Interface Science*, 61, 1977, p 351
52. N.M. Ghoniem and S. Sharafat, A Numerical Solution to the Fokker-Planck Equation Describing the Evolution of the Interstitial Loop Microstructure during Irradiation, *J. Nucl. Mater.*, 92, 1980, p 121
53. M. Kiritani, Analysis of the Clustering Process of Supersaturated Lattice Vacancies, *J. Phys. Soc. Jpn.*, 35, 1973, p 95
54. M. Koiwa, On the Validity of the Grouping Method —Comments on "Analysis of the Clustering Process of Supersaturated Lattice Vacancies," *J. Phys. Soc. Jpn.*, 37, 1974, p 1532
55. S.I. Golubov, A.M. Ovcharenko, A.V. Barashev, and B.N. Singh, Grouping Method for the Approximate Solution of a Kinetic Equation Describing the Evolution of Point-Defect Clusters, *Philos. Mag. A*, 81, 2001, p 643

56. J. Lépinoux, Contribution of Matrix Frustration to the Free Energy of Cluster Distributions in Binary Alloys, *Philos. Mag.*, 86, 2006, p 5053

57. F. Soisson and C.-C. Fu, Cu-Precipitation Kinetics in α-Fe from Atomistic Simulations: Vacancy-Trapping Effects and Cu-Cluster Mobility, *Phys. Rev. B*, 76, 2007, p 214102

58. K. Binder and D. Stauffer, Theory for the Slowing Down of the Relaxation and Spinodal Decomposition of Binary Mixtures, *Phys. Rev. Lett.*, 33, 1974, p 1006

59. K. Binder and M.H. Kalos, "Critical Clusters" in a Supersaturated Vapor: Theory and Monte Carlo Simulation, *J. Stat. Phys.*, 22, 1980, p 363

60. J.W. Cahn and J.E. Hilliard, Free Energy of a Nonuniform System. III. Nucleation in a Two-Component Incompressible Fluid, *J. Chem. Phys.*, 31, 1959, p 688

61. A. Pan, T. Rappl, D. Chandler, and N. Balsara, Neutron Scattering and Monte Carlo Determination of the Variation of the Critical Nucleus Size with Quench Depth, *J. Phys. Chem. B*, 110, 2006, p 3692

62. P.G. Debenedetti, Thermodynamics: When a Phase is Born, *Nature*, 441, 2006, p 168

63. R. Wagner and R. Kampmann, Homogeneous Second Phase Precipitation, *Materials Science and Technology, A Comprehensive Treatment*, R.W. Cahn, P. Haasen, and E.J. Kramer, Ed., Vol. 5, pp. 213–303, VCH, Weinheim, 1991

64. M. Perez, M. Dumont, and D. Acevedo-Reyes, Implementation of Classical Nucleation and Growth Theories for Precipitation, *Acta Mater.*, 56, 2008, p 2119

65. M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987

66. D. Frenkel and B. Smit, *Understanding Molecular Simulation — From Algorithms to Applications*, Academic Press, San Diego, 2001

67. G. Martin and F. Soisson, Kinetic Monte Carlo Method to Model Diffusion Controlled Phase Transformations in the Solid State, *Handbook of Materials Modeling*, S. Yip, Ed., pp. 2223–2248, Springer, The Netherlands, 2005

68. P. Bellon, Kinetic Monte Carlo Simulations in Crystalline Alloys: Principles and Selected Applications, *Thermodynamics, Microstructure, and Plasticity*, A. Finel, D. Mazière, and M. Veron, Ed., pp. 395–409, Kluwer Academic, Dordrecht, 2002

69. F. Soisson, Applications of Monte Carlo Simulations to the Kinetics of Phase Transformations, *Thermodynamics, Microstructure, and Plasticity*, A. Finel, D. Mazière, and M. Veron, Ed., pp. 427–436, Kluwer Academic, Dordrecht, 2002

70. J.P. Simmons, C. Shen, and Y. Wang, Phase Field Modeling of Simultaneous Nucleation and Growth by Explicitly Incorporating Nucleation Events, *Scripta Mater.*, 43, 2000, p 935

71. Y.H. Wen, J.P. Simmons, C. Shen, C. Woodward, and Y. Wang, Phase-Field Modeling of Bimodal Particle Size Distributions during Continuous Cooling, *Acta Mater.*, 51, 2003, p 1123

72. C. Shen, J.P. Simmons, and Y. Wang, Effect of Elastic Interaction on Nucleation: II. Implementation of Strain Energy of Nucleus Formation in the Phase Field Method, *Acta Mater.*, 55, 2007, p 1457

73. Y. Wang, H.-Y. Wang, L.-Q. Chen, and A.G. Khachaturyan, Microstructural Development of Coherent Tetragonal Precipitates in Magnesium-Partially-Stabilized Zirconia: A Computer Simulation, *J. Am. Ceram. Soc.*, 78, 1995, p 657

74. Y. Le Bouar, A. Loiseau, and A.G. Khachaturyan, Origin of Chessboard-Like Structures in Decomposing Alloys. Theoretical Model and Computer Simulation, *Acta Mater.*, 46, 1998, p 2777

75. Y.H. Wen, Y. Wang, and L.Q. Chen, Phase-Field Simulation of Domain Structure Evolution during a Coherent Hexagonal-to-Orthorhombic Transformation, *Philos. Mag. A*, 80, 2000, p 1967

76. Q. Bronchart, Y. Le Bouar, and A. Finel, New Coarse-Grained Derivation of a Phase Field Model for Precipitation, *Phys. Rev. Lett.*, 100, 2008, p 015702

SELECTED REFERENCES

Focusing on nucleation

• F.F. Abraham, Homogeneous Nucleation Theory—The Pretransition Theory of Vapor Condensation, *Advances in Theoretical Chemistry,* H. Eyring and D. Henderson, Ed., Academic Press, New York, 1974

• K.F. Kelton, Crystal Nucleation in Liquids and Glasses, *Solid State Physics*, H. Ehrenreich and D. Turnbull, Ed., Academic Press, New York, Vol. 45, 1991, pp. 75–177

• K.C. Russell, Nucleation in Solids: The Induction and Steady State Effects, *Adv. Colloid Interface Sci.*, 13, 1980, p 205

• D.T. Wu, Nucleation Theory, *Solid State Physics*, H. Ehrenreich and F. Spaepen, Ed., Academic Press, New York, Vol. 50, 1997, pp. 37–187

Phase transformations

• J.W. Christian, *The Theory of Transformations in Metals and Alloys—Part I: Equilibrium and General Kinetic Theory*, 2nd ed., Pergamon Press, Oxford, 1975

• A.G. Khachaturyan, *Theory of Structural Transformations in Solids*, Wiley, New York, 1983

Models of Recrystallization

Frank Montheillet, CNRS, Ecole Nationale Supérioure des Mines de Saint-Etienne, France
John J. Jonas, Birks Professor of Metallurgy Emeritus, McGill University, Canada

DURING HOT PROCESSING, metallic materials can undergo various types of recrystallization, which dramatically modify their microstructures and crystallographic textures. Therefore, recrystallization is to a large extent responsible for their final mechanical properties. Various types of recrystallization must be considered.

Static recrystallization (SRX) occurs (most generally) after room-temperature deformation during further heat treatment. In the initial fully annealed state (Fig. 1a), the dislocation density of the material is very low ($\rho \approx 0.01$ $\mu m/\mu m^3$). Strain hardening makes it grow to 10^4 $\mu m/\mu m^3$ (Fig. 1b). Since it was formerly believed that metals lost their crystalline character after large strain deformation, the generation of new grains upon subsequent heating was termed recrystallization. It is now well known, however, that deformed initial grains are still present (although more or less strongly deformed) in the cold-worked state (Fig. 1c), and they are merely replaced by new grains during static recrystallization (Fig. 1e). Note that here *static* means that no strain is imposed to the material during recrystallization. Static recrystallization is similar to a phase transition.

Dynamic recrystallization (DRX) takes place instead during hot working, which means that the generation of new grains is combined in an intricate way with other phenomena associated with high-temperature straining, such as work hardening and dynamic recovery (Fig. 1d). Two types of DRX can, in fact, be distinguished.

Continuous Dynamic Recrystallization. When dynamic recovery is very efficient (i.e., in metals where high stacking-fault energy favors dislocation mobility), the initial grains can undergo quite large strains without losing their individualities. In the absence of a nucleation mechanism, the new grain structure results from the progressive fragmentation of the latter, by the generation of subgrain boundaries (polygonization) and their evolution into high-angle boundaries. This mechanism is referred to as continuous dynamic recrystallization (CDRX). In addition, initial grain boundaries become elongated and serrated at large

strains and also contribute to the grain fragmentation, a process known as geometric dynamic recrystallization. CDRX has been observed mainly in hot worked aluminum alloys (Ref 1) and ferritic steels (Ref 2). However, quite similar microstructures are also generated during cold or warm severe plastic deformation (SPD), produced by equal channel angular extrusion (ECAE) (Ref 3), high pressure torsion (HPT) (Ref 4), or merely large strain torsion

testing (Ref 5). Ultrafine grain strucutres can be obtained in this way.

Discontinuous Dynamic Recrystallization. When dynamic recovery is weak, local concentrations of dislocations lead to various types of nucleation events. The large gradients of dislocation density induce considerable grain-boundary migration velocities such that CDRX cannot occur. The new structure then results from repeated cycles of nucleation, growth,
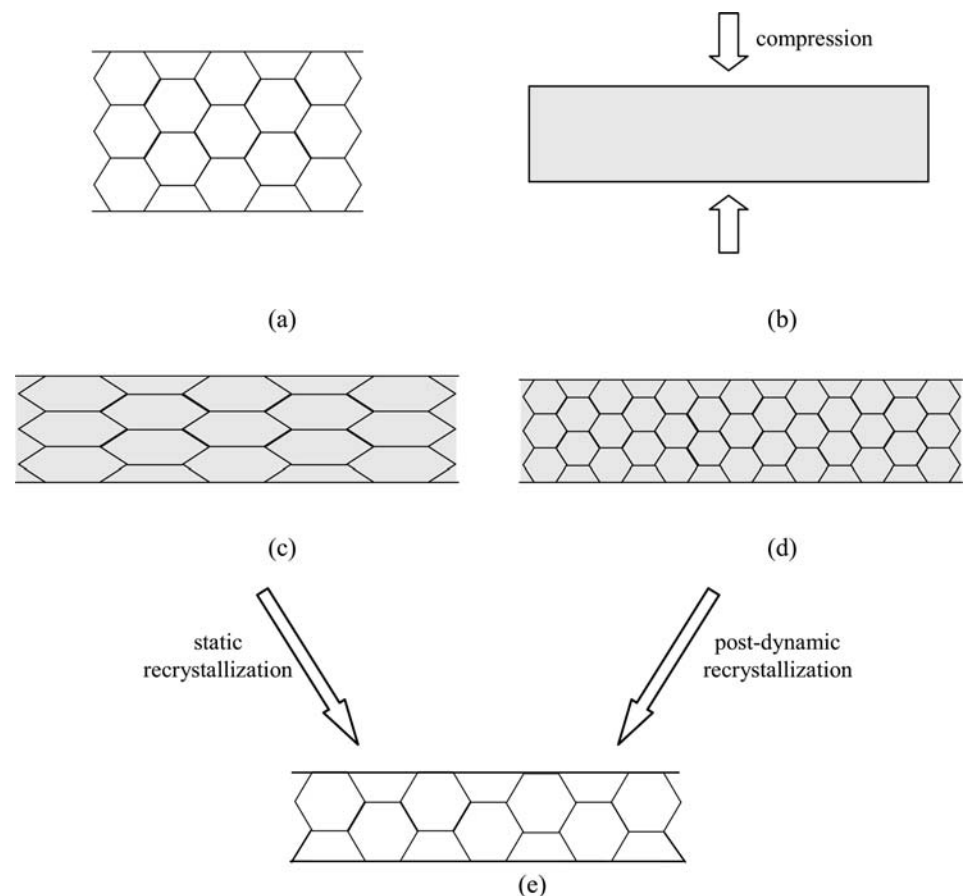
Fig. 1 Schematic representations of (a) the fully annealed state of a polycrystalline metal, (b) the former interpretation of a cold-worked specimen (amorphous state), (c) the modern interpretation of the latter, (d) a hot-worked dynamically recrystallized microstructure, and (e) a postdynamically recrystallized state

and shrinkage of grains. This mechanism is called discontinuous dynamic recrystallization (DDRX). It should be noted that for a long time it was the only recognized DRX mechanism. Moreover, in some cases, hot deformation is likely to take place according to mixed or intermediate mechanisms.

In both cases, a steady state of flow stress, microstructure, and crystallographic texture is observed at large (DDRX) or very large (CDRX) strains (in the latter case, the steady state cannot generally be achieved in industrial processes). In such a state, the material has reached a dynamic equilibrium with the external prescribed conditions (strain rate and temperature) and behaves as a dissipative structure that converts the mechanical energy input into heat.

The basic difference between static and dynamic recrystallization is also illustrated in Fig. 2, which shows the level of stored (elastic) energy associated with dislocations. During steady-state DRX, the latter is larger than in the annealed state but less than after cold working.

Recrystallization and the Avrami Kinetics

Historical Background

Jewelers and armor makers have known since biblical times that a heavily worked piece of metal can be softened by heating. This takes place largely by means of recrystallization, a mechanism that is clearly influenced by a number of experimental parameters in a manner that is outlined subsequently. (Recovery also plays a role but is not discussed here.) Without the occurrence of recrystallization, it would not be possible to manufacture the fairly thin sheets of steel from which auto bodies and many household appliances are made. Similar remarks apply to many other industrial processes. The progress of recrystallization can be described by what are known as Avrami kinetics (Ref 6, 7), which are sometimes referred to

instead as Johnson-Mehl-Avrami-Kolmogorov (JMAK) kinetics (Ref 8, 9).

The Avrami theory was developed to take into account the following general observations regarding the progress of recrystallization (Ref 10):

1. A minimum amount of deformation is necessary to initiate recrystallization.
2. The greater the amount of deformation, the lower the temperature required for initiation.
3. Increasing the annealing time decreases the initiation temperature.
4. The final grain size decreases with the amount of deformation and increases with the deformation temperature.
5. More deformation is required, for a given temperature and time, when the original grain size is coarser.
6. More deformation is required to produce a specific amount of hardening as the temperature of working is increased.

Such behavior can be interpreted in broad terms by considering the individual characteristics of nucleation and growth (Ref 11). For example, the existence of a critical strain (item 1) is based on the need to establish a minimum dislocation density (or local lattice curvature) in order to create a nucleus. Since nucleation is thermally activated, longer annealing times and higher temperatures increase the probability of producing a nucleus (items 2 and 3). The need for thermal activation, in turn, accounts for the presence of an incubation time before recrystallization can be observed.

The final grain size will depend on the balance between the rates of nucleation and growth. Because the nucleation rate increases with the amount of work hardening, while the growth rate increases with temperature, it is clear that the grain size will decrease with the imposed strain and increase with the temperature. Given that most nucleation takes place at or near grain boundaries, a finer initial grain size will lead to a higher nucleus density (item 5). Finally, item 6 is based on the dependence of dislocation density on the deformation temperature, decreasing

as it does with increased temperature. Similar remarks apply to strain rate: An increase in the latter increases the net dislocation density at a given strain and temperature and therefore the nucleation rate. In this case, higher strain rates are associated with finer final grain sizes.

The Avrami analysis (Ref 6, 7) was based on the following assumptions and simplifications:

1. The potential nuclei are already present in the deformed material and are randomly distributed.
2. All the nuclei are either active from the beginning of heating (one extreme), or else they form at a constant rate (the other extreme).
3. The nuclei may grow isotropically in three dimensions, two dimensions, or one dimension (i.e., there can be polyhedral, platelike, or linear growth).
4. A correction must be made for the impingement of growing grains upon one another; that is, transformation can no longer occur within the overlapped regions.

The constant formation rate (item 2) taken together with the three-dimensional growth rate (item 3) leads to a transformed (recrystallized) volume that increases with the fourth power of the time. This can account for the observation of what appears to be an incubation time.

The JMAK Model for Static Recrystallization

Defining X as the recrystallized volume fraction, the Avrami analysis leads to the following general relationship:

$$X = 1 - \exp(-k\,t^n) = 1 - \exp[-k'(t/t_R)^n] \quad \text{(Eq 1)}$$

Here, $k = k'/t_R{}^n$ is the Avrami constant; k' depends on the definition of t_R (see the following); t is the annealing time; t_R is a characteristic time, defined in more detail later; and n is the Avrami time exponent. The general equation for n on the basis of physical models is $n = q\,d + B$, where q is 1 for interface-controlled growth, q is $\frac{1}{2}$ for diffusion-controlled growth, d is the dimensionality of growth, and B is 0 and 1 under site-saturation nucleation conditions and continuous nucleation at a constant rate, respectively (Ref 12). Nevertheless, most experimental values fall in the range 1 to 2. In the original paper by Avrami (Ref 7), n was = 3, 2, and 1 under site-saturation conditions for three-, two-, and one-dimensional growth, respectively. Similarly, under continuous nucleation conditions, n was = 4, 3, and 2, respectively, for the same three classes of dimensionality. In Avrami's work, no account was taken of diffusion-controlled growth.

For algebraic purposes, it is useful to define t_R as the time required for a given fraction of recrystallization. In the general case, $k' = -\ln(1-f)$, where f is the fraction selected. Thus, if the times associated with 5, 50, and 95% recrystallization are of interest, for
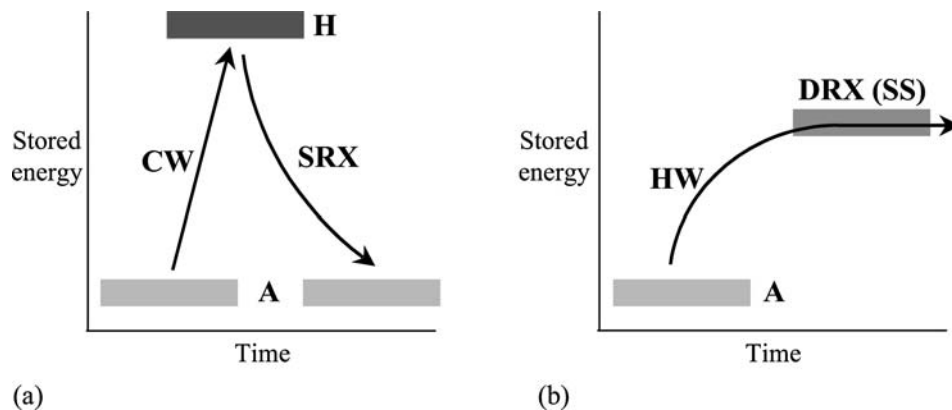
Fig. 2 (a) Static recrystallization (SRX) that follows cold working (CW) involves the transition between a high-energy work-hardened state (H) to the low-energy annealed state (A). (b) Dynamic recrystallization (DRX) taking place during hot working (HW) leads at large strains to a steady-state (SS) microstructure with intermediate energy level.

example, $k' = 0.05$, 0.69, and 2.99, respectively. For most calculations, the time for 50% recrystallization, $t_{50\%}$ or t_{50}, is employed, leading to:

$$X = 1 - \exp[-0.69\,(t/t_{50})^n] \qquad \text{(Eq 2)}$$

where $k = 0.69/t_{50}^n$.

The dependence of t_{50} on the experimental conditions, that is, on temperature, strain, strain rate, composition, and so on, is considered in some detail later. If the value of n is known, the aforementioned relation can be used to evaluate the ratio t_{95}/t_5, for example, for different materials.

For plotting purposes, it is useful to take the double log of $1/(1-X)$ so that the progress of recrystallization $1 - X$ can be displayed against time t. The previous equation can then be rewritten as:

$$\log\{\ln[1/(1-X)]\} = [\log\,(0.69) - n\,\log\,(t_{50})] \\ + n\,\log\,(t) \\ = \log\,(k) + n\,\log\,(t)$$

$$\text{(Eq 3)}$$
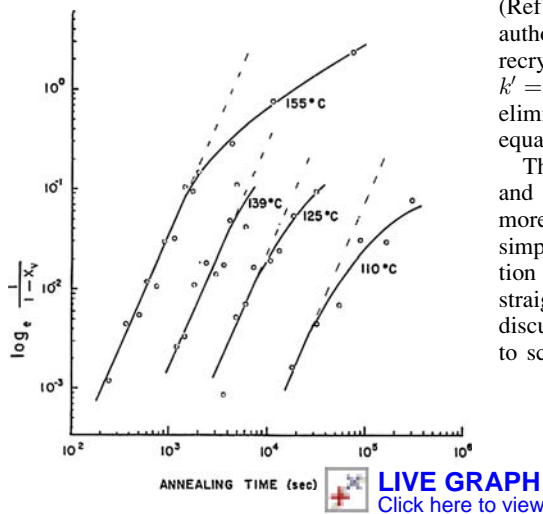
Fig. 3 Log [ln 1/(1 – X)] plotted as a function of the logarithm of the annealing time at various temperatures in zone-refined aluminum containing 0.004% Cu cold rolled to 40% reduction at 0 °C. Source: Ref 13

Table 1   Approximate recrystallization temperatures for various metals

Metal	Quality	Temperature °C	K
Lead	99.999%	<0	<273
Aluminum	99.999%	75	348
Aluminum	Commercial purity	275	548
Aluminum	+1% Mn	400	673
Copper	99.999%	100	373
Copper	Commercial purity	200–250	473–523
Copper	+2% Be	250	523
Iron	Pure	450	723
Molybdenum	Sintered	1000	1273
Molybdenum	0.5% Ti	1400	1673
Molybdenum	10% Nb	1750	2023
Tungsten	Sintered	1200	1473

Adapted from Ref 14

Such plots are frequently linear, signifying that both k and n are constant. An example of such a plot is presented in Fig. 3. Numerous other examples can be found in the literature.

For practical purposes, the time required to soften a material (produce recrystallization) in 1 h is of considerable interest. Some typical values are displayed in Table 1 (Ref 14). Here, because the values are taken from the work of different researchers, the amount of cold work is not constant. The temperatures, in K, are generally about 0.4 T_m or above, where T_m is the melting temperature in absolute degrees. Nevertheless, as can be seen, the addition of alloying elements increases the recrystallization temperature by appreciable amounts. This is due to the effects of both solute drag and boundary pinning by particles.

Another useful way of describing and comparing the recrystallization kinetics of different materials is in terms of their characteristic times. An example is given in Fig. 4, where materials of different purities are examined, together with the dependence of the characteristic time, t_R, on the absolute temperature of deformation (Ref 15). Note that in this case, however, the authors have chosen to use the time for 63% recrystallization, because under these conditions $k' = -\ln(1-f) = -\ln(0.37) = 1.0$, thus eliminating the former from the Avrami equation.

The effects of strain, strain rate, composition, and temperature on t_{50} will be considered in more detail as follows. For the moment, it is simply of interest to compare the recrystallization temperatures of different materials in a straightforward manner. As expected from the discussion of Fig. 3, the latter can be expected to scale with the absolute melting temperature

of the metal. This is illustrated in Fig. 5, where the excellence of this correlation can be seen.

Effect of the Rates of Nucleation and Growth on Recrystallization Kinetics. Here, the simple case is considered where recrystallization is being continuously nucleated throughout the material at a constant rate N. It is also assumed that the nuclei are randomly distributed. (The alternative case of site saturation can be treated in a similar manner.) The volume of a spherical new grain, before impingement, will therefore be given by (Ref 16):

$$V = \tfrac{4}{3}\pi r^3 = \tfrac{4}{3}\pi(vt)^3 \qquad \text{(Eq 4)}$$

where r is the radius of the sphere, t is the time, and v is the growth velocity, which is taken here to be constant with time. Because of continuing nucleation, prior to impingement, the volume fraction recrystallized, X, at any time is given by:

$$X = \frac{4}{3}\pi N v^3 \int_0^t u^3 du = \frac{\pi}{3} N v^3 t^4 \qquad \text{(Eq 5)}$$

This equation is only valid for small values of X, that is, prior to impingement. Once impingement begins, a correction must be made for the effect of overlapping. This leads to the Avrami equation in the form:

$$X = 1 - \exp(-\frac{\pi}{3} N v^3 t^4) \qquad \text{(Eq 6)}$$

where $k = \pi/3N v^3$ and $n = 4$ for constant rates of nucleation and of three-dimensional growth. The other examples of dimensionality and conditions of nucleation can be readily described in a similar manner.

Given that the $\exp\left[-\pi/3N v^3(t_{50})^4\right] = 0.5$ when recrystallization is half complete, it
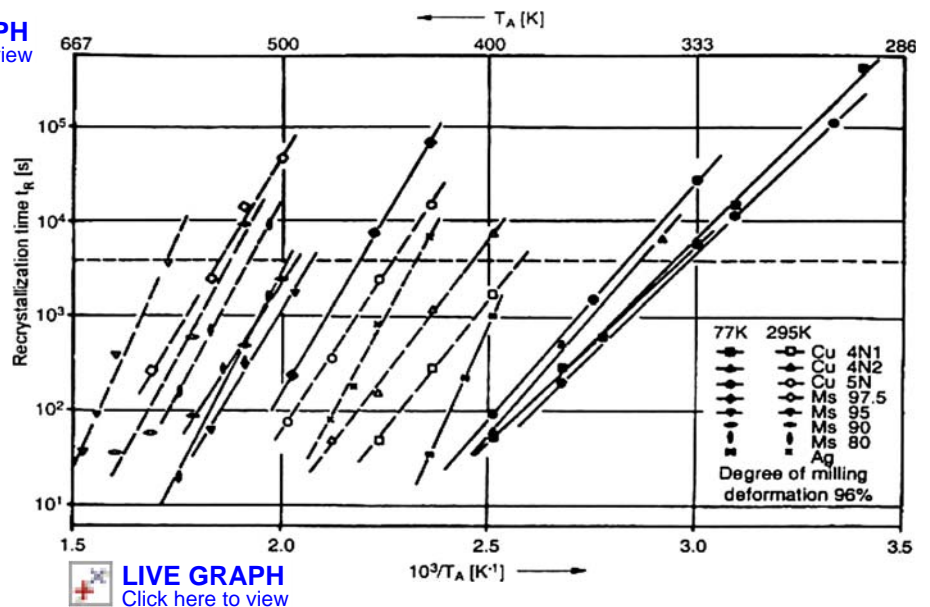
Fig. 4 Dependence of the characteristic time (t_R) on annealing temperature (T_A) for different materials and rolling temperatures. Source: Ref 15

can be seen that t_{50} is given by $(0.69/k)^{1/n}$, which is equivalent to:

$$t_{50} = 0.9/(N^{1/4} v^{3/4}) \qquad \text{(Eq 7)}$$

under the present conditions. This relation makes it clear that rapid recrystallization is associated with rapid rates of nucleation and growth.

Effect of the Rates of Nucleation and Growth on the Recrystallized Grain Size. The rates of nucleation and growth also determine the new grain size after recrystallization. For example, the new grain size d is approximately given by $d \approx 2\,v\,t_{50}$, which, for the case discussed previously, works out to $d \approx 1.8\,(v/N)^{1/4}$. This agrees with the intuitive expectation that the grain size increases somewhat slowly with v and decreases in a similar manner with N.

Both the grain-boundary mobility and nucleation rate are thermally activated processes, as indicated previously, and can therefore be associated with activation energies Q_v and Q_N, respectively. Thus, their rates can be expressed as:

$$v = v_0 \exp(-Q_v/kT) \qquad \text{(Eq 8a)}$$

$$N = N_0 \exp(-Q_N/kT) \qquad \text{(Eq 8b)}$$

Here, the pre-exponentials v_0 and N_0 are considered to be temperature independent (Ref 15).

In this way, t_{50} can be rewritten in the following form:

$$t_{50} = \left[0.9/(N_0 v_0^3)^{1/4} \right] \exp\left[\frac{Q_N + 3Q_v}{4kT} \right] \qquad \text{(Eq 9)}$$

It follows that the slope of an Arrhenius plot of t_{50} or t_R versus $1/T$, which corresponds to the apparent activation energy for recrystallization, Q_{SRX}, should be approximately equal to $(Q_N + 3Q_v)/4$.

The effects of N, v, Q_N, and Q_v on the recrystallized grain size can be derived in a similar manner. Thus, the aforementioned expression for d can be expanded as follows:

$$d \approx 1.8\,(v_0/N_0)^{1/4} \exp\left[\frac{Q_N - Q_v}{4kT} \right] \qquad \text{(Eq 10)}$$

From this equation, it is apparent that the final grain size is determined by a competition between the nucleation rate and the growth rate, where higher nucleation rates lead to finer grain sizes and higher growth rates to coarser sizes. It also follows that the recrystallized grain size will be approximately temperature independent if Q_N and Q_v are approximately equal. Another important outcome of this analysis involves the effects of prior strain on N and v. Both of these rates increase with strain, N the more rapidly. Thus, both t_{50} and d decrease with the prior strain.

It should be noted that the expressions discussed in these two sections are valid for the ideal case of continuous, homogeneous nucleation and isotropic, three-dimensional growth. Equivalent relations can be readily derived for the other dimensional growth conditions and for the case of site saturation. These are not reviewed here.

Because continuous dynamic recrystallization (CDRX), discussed later in this article, does not involve nucleation and growth, the aforementioned treatment only applies to static (SRX) and discontinuous dynamic recrystallization (DDRX).

Effects of Strain, Strain Rate, and Temperature on t_{50}. It is important to be able to predict and quantify the rate of static recrystallization under industrial conditions, because of its effect on the flow stress and therefore on the modeling of rolling load. For this reason, it was proposed

by Sellars (Ref 17) that the effects of the preceding parameters on t_{50} can be described as follows:

$$t_{50} = A\,\varepsilon^{-p}\,\dot{\varepsilon}^{-q}\,d_0^r\,\exp(Q_{SRX}/RT) \qquad \text{(Eq 11)}$$

Some typical values for p, q, and Q_{SRX} are presented in Table 2 for seven steels (Ref 18). Substituting these typical values in Eq 11, it can be seen that the strain ε has quite a significant effect—doubling ε reduces t_{50} by a factor of 4. Conversely, t_{50} is much less sensitive to the strain rate $\dot{\varepsilon}$ of the prior pass. Although the dependence on prior grain size d_0 is not shown in the table, the value of r is generally about 2, so that increasing the prior grain size increases t_{50}, as has already been mentioned.

During the hot rolling of austenite, and that of most face-centered cubic metals, dynamic recrystallization is induced after a certain critical strain. Once dynamic recrystallization is well under way, the rate of recrystallization after unloading becomes much more rapid, and the parameters listed previously are no longer valid. This phenomenon is referred to as metadynamic recrystallization (MDRX) or postdynamic recrystallization. After large enough prestrains, the rate of postdeformation softening actually becomes strain independent. As in this case, nucleation has taken place during deformation; the behavior is largely controlled by growth of the grains nucleated during the prior deformation (and partly by the nucleation of new grains in the strain-hardened regions, i.e. by conventional SRX). For this reason, the activation energy for MDRX ($\approx 3\,Q_v/4$ only, see Eq 8 to 10) is always lower than for SRX $[(Q_N + 3\,Q_v)/4]$. In a similar manner, because the initial microstructure is now replaced by the dynamic microstructure, which is strain-rate dependent, the rate dependence of MDRX is approximately double that of SRX. This relatively complex topic is not treated here but is discussed in greater detail in Ref 18 and 20.

Kinetics of Dynamic Recrystallization

Although not immediately apparent, it is also possible to describe the kinetics of dynamic recrystallization with the aid of the Avrami relations. This is because the time variable applicable to static conditions can be replaced by the strain using the relation $\varepsilon = \dot{\varepsilon}\,t$ (as long
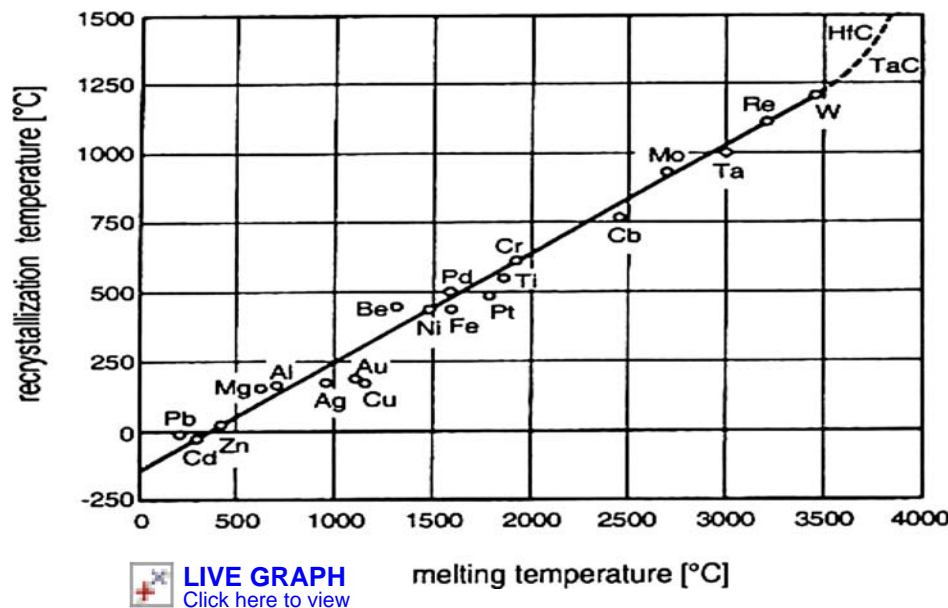
Fig. 5 Recrystallization temperatures of various metals as a function of their melting temperatures. Source: Ref 10

Table 2 t_{50} parameters for static recrystallization in seven steels

Grade	p	q	Q_{SRX}, kJ·mol⁻¹
SS 304	1.48	0.42	207
C-Mn	2	0.34	215
HSLA	2	0.37	330
IF	2	0.37	250
IF	1.9	0.4	192
IF (B)	1.9	0.4	293
X65	2.5	0.3	390

Source: Ref 18

as the strain rate is constant). A further modification involves the introduction of the critical strain, ε_c, for the initiation of dynamic recrystallization (DRX), so that the appropriate strain variable becomes $(\varepsilon - \varepsilon_c)$. (Here, ε_c plays the role of the incubation time in conventional recrystallization.) In turn, the critical strain can be determined by the method of Poliak and Jonas (Ref 21); however, according to classical metallographic observations, it is generally taken equal to $\frac{5}{6}\varepsilon_p$.

A suitable expression must now be introduced to represent the flow curve expected in the absence of dynamic recrystallization, that is, due to dynamic recovery alone. It is shown schematically in Fig. 6(a), where it is identified as σ_{recov}. This curve is described by the relation:

$$\sigma = \left[\sigma_\infty^2 - \left(\sigma_\infty^2 - \sigma_e^2\right)\exp\left(-r\varepsilon\right)\right]^{1/2} \qquad \text{(Eq 12)}$$

Here, σ_e is the yield stress; the recovery saturation stress, σ_∞, is determined from the intercept of the $\sigma \cdot d\sigma/d\varepsilon$ versus σ^2 curve, and the value of the recovery parameter, r, from its slope (Ref 18, 19). Note that the appropriate σ/ε data are only taken from the portion of the work-hardening curve prior to σ_c, ε_c.

With the aid of these definitions, the softening attributable to DRX, defined with respect to the dynamic recovery flow curve, can be specified as follows:

$$X = 1 - \exp[-k_2(\varepsilon - \varepsilon_c)^n] \qquad \text{(Eq 13a)}$$

Here, k_2 is the Avrami constant applicable to softening by DRX. Note that X refers here to the fractional softening and not the volume fraction of recrystallization employed previously.

A schematic diagram depicting the Avrami kinetics pertaining to Fig. 6(a) is presented in Fig. 6(b), with $n \approx 2.5$. The strong similarity to Fig. 3 for the case of the volume fraction recrystallized during static recrystallization is

evident. A somewhat similar analysis was employed by Stewart et al. (Ref 22) to estimate the volume fraction of new grains (not the fractional softening) associated with DRX.

An approximate description of the kinetics as well as the magnitude of the softening is the following, which does not require knowledge of σ_{recov}, σ_∞, or ε_c. For this purpose, the peak strain, ε_p, is used in place of ε_c. Then, the amount of softening is defined with respect to σ_p and not σ_∞. Under these conditions, X is given instead by:

$$X = 1 - \exp\left[-k_2'(\varepsilon - \varepsilon_p)^{n'}\right] \qquad \text{(Eq 13b)}$$

A plot of the softening depicted in Fig. 6(a) in terms of this alternative relation is shown in Fig. 7, from which it can be seen that the second DRX relation can also be used to describe the softening taking place under DRX conditions.

Mesoscale Approach for DDRX

By contrast to the Avrami phenomenological approach, the mesoscale (or grain scale) models take the basic physical mechanisms into account, that is, strain hardening and dynamic recovery, nucleation, grain-boundary migration, and the associated softening effects. Their main advantage, however, is to predict both flow stress and microstructure changes (dislocation densities, crystallite or grain sizes) during hot deformation. In this section, the basic framework of the mesoscale approach for DDRX is first presented, including the three basic equations for:

- Grain size changes
- Strain hardening and dynamic recovery
- Nucleation

The model is then illustrated by some numerical predictions. Finally, a simplified analysis is given for the steady state, in the case of the

power-law strain-hardening equation. Closed-form equations for the flow stress and the average recrystallized grain size are derived, which allow the Derby relationship and the Jonas and Sakai criterion to be discussed.

Geometrical Framework

The material is made of a set of N (number per unit volume) interacting crystals where, more specifically, each grain is considered as an inclusion embedded in a matrix, the properties of which will be obtained by averaging that of all the grains in the aggregate. Such an approach is close to the self-consistent models commonly used in the continuum mechanics of heterogeneous materials. It precludes topological effects, such as necklacing, which are expected, however, to be second-order effects and do not even take place during the steady state. As shown as follows, this average-field approach leads to a simple formulation of the model, leading, in most cases, to quasi-analytical predictions. Three-dimensional aggregates of equiaxed grains are often represented by stacks of tetrakaidecahedra (or truncated octahedra), which can fill space without voids or overlapping. However, calculations are much easier when ellipsoids or spheroids are used. Even in that case, when normal grain-boundary migration takes place (i.e., the normal component of the local grain-boundary velocity remains constant over the whole area), the ellipsoidal shape is not conserved, which leads us to consider merely spherical grains (Ref 23). This approximation is likely to hold for DDRX, since grains generally disappear before their shapes become significantly flattened or elongated. Then, each grain will be specified at any time by only two parameters, for example, its diameter, D_i, and average dislocation density, ρ_i, while $\bar{\rho}$ will denote the average dislocation density of the material.

(a)

LIVE GRAPH
Click here to view

(b)
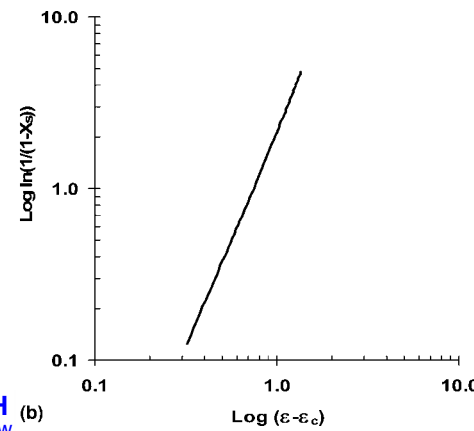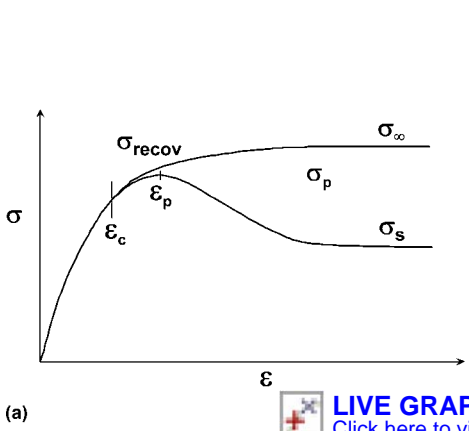
Fig. 6 (a) Stress-strain diagram of dynamic recovery and dynamic recrystallization (DRX) flow curves. The critical strain, ε_c, and peak strain, ε_p, are identified. The amount of softening attributable to DRX is defined as $\sigma_{recov} - \sigma$. The fractional softening X is then given by the following expression: $X = (\sigma_{recov} - \sigma)/(\sigma_{recov} - \sigma)_{max} = (\sigma_{recov} - \sigma)/(\sigma_\infty - \sigma_s)$. (b) Avrami plot of the schematic softening data of (a) prepared using ε_c and $\sigma_{recov} - \sigma$

LIVE GRAPH
Click here to view
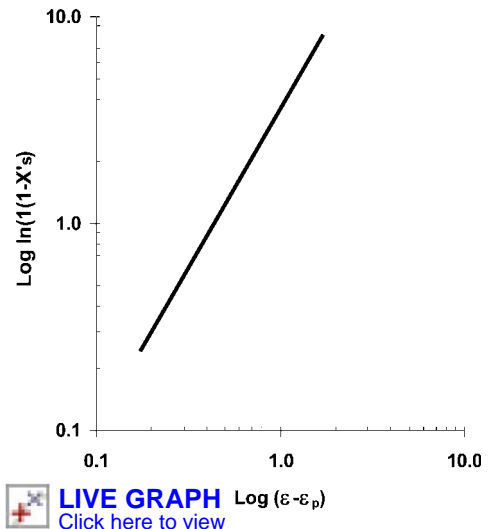
Fig. 7 Avrami plot of the softening data of Fig. 6(a) prepared using ε_p and $\sigma_p - \sigma$

Instantaneous homogenization of the dislocation densities is assumed to take place within the grains.

Basic Equations

The evolution of the aforementioned dynamical system of interacting grains is ruled by three types of equation, pertaining to grain-boundary migration, dislocation density changes, and grain nucleation, respectively (Ref 24, 25 and 34).

Grain Growth and Shrinkage. The dislocation density difference between a grain and the surrounding matrix is assumed to be the predominant driving force for dynamic migration of grain boundaries (Fig. 8). Hence:

$$\dot{D}_i = \frac{dD_i}{dt} = 2M\tau\,(\bar{\rho} - \rho_i) \qquad \text{(Eq 14a)}$$

or

$$\frac{dD_i}{d\varepsilon} = \frac{2M\tau}{\dot{\varepsilon}}\,(\bar{\rho} - \rho_i) \qquad \text{(Eq 14b)}$$

for each grain, where the strain rate, $\dot{\varepsilon}$, is considered as identical for all the grains (Taylor assumption). M denotes the grain-boundary mobility [m^4/(J·s)], and $\tau = \mu b^2$ (J/m) is the line energy of the dislocations, μ is the shear modulus, and b is the Burgers vector. The driving energy for the grain boundary is $\Delta G = \tau(\bar{\rho} - \rho_i)$, and the factor 2 means that two opposite parts of a spherical grain boundary move into opposite directions. Note that the product $M\tau$ has the simple dimension of a material flux (m^3/s). In the following, it will be convenient to use the micrometer as unit length, and therefore μm/μm^3 for dislocation densities and μm^3/s for $M\tau$. Equations 14(a) and (b) mean that a grain either grows or shrinks according to whether its dislocation density is less or greater than the average dislocation density $\bar{\rho}$ of the material. Since ρ_i mostly increases monotonically with time, the grain reaches its maximum size when $\rho_i = \bar{\rho}$.

Volume Conservation. The time derivative of the overall volume $V = \sum D_i^3$ is proportional to $\sum D_i^2\dot{D}_i$, whence, according to Eq 14(a) and (b):

$$\dot{V} \propto 2M\tau\Big(\bar{\rho}\sum D_i^2 - \sum \rho_i D_i^2\Big) \qquad \text{(Eq 15)}$$

Volume conservation therefore requires $\bar{\rho}$ to be defined as:

$$\bar{\rho} = \frac{\sum \rho_i D_i^2}{\sum D_i^2} \qquad \text{(Eq 16)}$$

When the aforementioned summation is restricted to the first neighbors of the grain under consideration, Eq 16 can be given a physical interpretation: It merely means that grain interactions with their neighbors are proportional to their surfaces rather than their volumes.

Strain Hardening and Dynamic Recovery. The dislocation density change within a grain can be described by any equation accounting for the effects of strain hardening and dynamic recovery. Two of them are used as follows, namely the power-law equation:

$$\frac{d\rho}{d\varepsilon} = \frac{H^{v+1}}{\rho^v} \qquad \text{(Eq 17)}$$

where $v \geq 0$, and H has the dimensions of dislocation density (μm^{-2}), since it leads to closed-form expressions for the steady state. In the absence of DRX (low strains), it gives by integration:

$$\rho = (v+1)^{1/(v+1)}H\left[\varepsilon + \frac{\rho_0^{v+1}}{(v+1)H^{v+1}}\right]^{1/(v+1)} \qquad \text{(Eq 18)}$$

where ρ_0 is the initial value of the dislocation density. Using the classical relationship $\sigma = \alpha\mu b\sqrt{\rho}$ ($\alpha \approx 1$) relating the dislocation density to the flow stress, this yields:

$$\sigma = K\,(\varepsilon + \varepsilon_1)^n \qquad \text{(Eq 19)}$$

where K and ε_1 are constants, and $n = 1/(2v+2)$, which is the well-known power-law strain-hardening (Swift) equation. Note that the dislocation density and the flow stress grow indefinitely at large strains.

The Yoshie-Laasraoui-Jonas (YLJ) equation is (Ref 26):

$$\frac{d\rho}{d\varepsilon} = h - r\rho \qquad \text{(Eq 20)}$$

where the two terms on the right side represent strain hardening (h) and dynamic recovery (rρ). Note that h has the dimension of a dislocation density, whereas r is dimensionless. Integration of Eq 20 gives:

$$\rho = \rho_\infty - (\rho_\infty - \rho_0)\exp(-r\varepsilon) \qquad \text{(Eq 21)}$$

where ρ_0 is again the initial dislocation density. At large strains, ρ tends to the asymptotic value $\rho_\infty = h/r$. The associated flow stress can be written in the form:

$$\sigma_0 = \left[\sigma_\infty^2 - (\sigma_\infty^2 - \sigma_e^2)\exp(-r\varepsilon)\right]^{1/2} \qquad \text{(Eq 22)}$$

where σ_e and σ_∞ are the yield stress and steady-state stress, respectively. Equations similar to Eq 20, for example, proposed by Estrin and Mecking (Ref 27) or Stüwe and Hertel (Ref 28), can be employed as well. In contrast to the power law, they lead to a steady-state flow stress at large strains. However, they are unable to model the flow softening usually associated with DRX.

Effect of Grain-Boundary Migration on the Dislocation Density. A first cause of softening during hot working can be expected to arise from the annihilation of dislocations (belonging to subgrain boundaries or not) by moving grain boundaries. Such an effect will be referred to as boundary-migration-induced softening (BMIS). In the mesoscale DDRX model, during grain growth the volume increment during time dt is almost free of dislocations (density ρ_0), as shown in Fig. 9. Under the assumption of instantaneous homogenization, a simple geometric derivation leads to the following BMIS term:

$$\left(\frac{d\rho_i}{d\varepsilon}\right)_{BMIS} = -3\frac{\rho_i - \rho_0}{D_i}\frac{dD_i}{d\varepsilon} \qquad \text{(Eq 23)}$$

for a grain of current size D_i and dislocation density ρ_i, with $dD_i/dt > 0$, that is, $\rho_i < \bar{\rho}$ according to Eq 14 When a grain shrinks, however, its

Fig. 8 Schematic representation of a grain in the aggregate, which grows when (a) $\rho_i \leq \bar{\rho}$ and shrinks when (b) $\rho_i > \bar{\rho}$
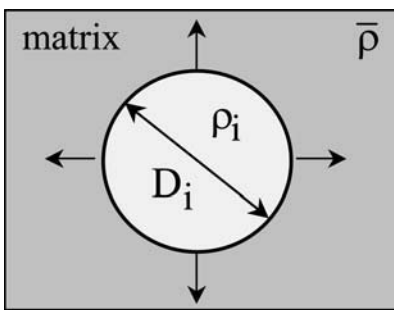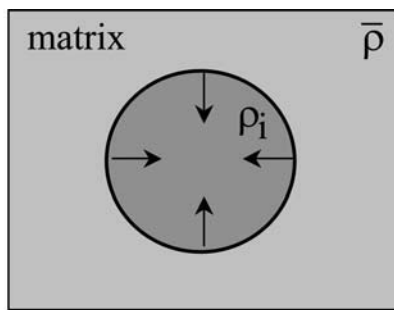
(a) $\rho_i < \bar{\rho}$: growth

(b) $\rho_i > \bar{\rho}$: shrinkage
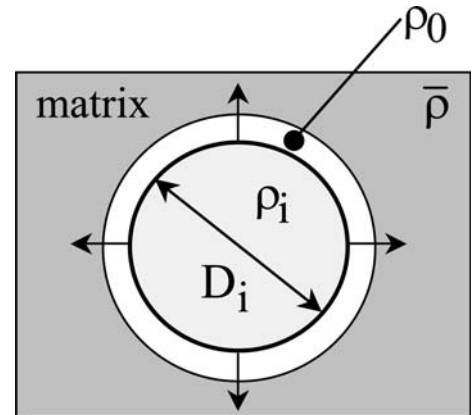
Fig. 9 Schematic representation of the boundary-migration-induced softening mechanism associated with grain growth during discontinuous dynamic recrystallization

dislocation density is not affected by boundary migration, since the volume swept by the boundary is simultaneously removed from the grain (BMIS is then implicitly accounted for in $\bar{\rho}$).

Finally, the dislocation density changes in a grain of the aggregate are given by:

$$\frac{d\rho_i}{d\varepsilon} = f_i(\rho_i) - 3\frac{\rho_i - \rho_0}{D_i}\frac{dD_i}{d\varepsilon} \text{ if } \rho_i < \bar{\rho} \qquad \text{(Eq 24a)}$$

$$\frac{d\rho_i}{d\varepsilon} = f_i(\rho_i) \text{ if } \rho_i > \bar{\rho} \qquad \text{(Eq 24b)}$$

where $f_i(\rho_i)$ stands for Eq 17 or 20 or a similar relationship. Parameters involved in f_i (i.e., H_i, ν_i, h_i, r_i) may depend on the grain under consideration. In particular, in the YLJ formulation, variations of the strain-hardening parameter h are expected due to the different crystallographic orientations of the grain in the aggregate. On the other hand, possible variations of r can be discarded to a first approximation.

Nucleation. Physical mechanisms for the nucleation of new grains during hot deformation (e.g., grain-boundary bulging, subgrain rotation, thermal twinning) are not yet well established and are likely to depend on the material and accumulated strain. In the mesoscale DDRX model, a general equation for the number of new grains generated per unit time and unit volume is introduced:

$$\frac{dN^+}{dt} = k_N \bar{\rho}^p \sum D_i^2 \qquad \text{(Eq 25a)}$$

or

$$\frac{dN^+}{d\varepsilon} = \frac{k_N \bar{\rho}^p}{\dot{\varepsilon}} \sum D_i^2 \qquad \text{(Eq 25b)}$$

where k_N (μm^{2p-2}/s) is a nucleation parameter. Here, dN^+/dt is assumed to be proportional to the total grain-boundary area in the aggregate, since DRX new grains are generally observed to nucleate at grain boundaries, and $\dot{n} = k_N \bar{\rho}^p$ represents the classical nucleation rate per unit area (the exponent p is estimated subsequently). Volume nucleation of the new grains could be alternatively accounted for by a cubic dependence with respect to the D_i. The N^+ increases monotonically during straining; in the numerical version of the model, whenever it reaches the critical value 1, a nucleus of size $D = 0$ and dislocation density ρ_0 is added to the set of grains, and N^+ is reset to zero.

Strain Rate and Temperature Dependence of the Parameters. The grain-boundary mobility M and nucleation parameter k_N are assumed to be rate-insensitive increasing functions of temperature, while the various strain-hardening and dynamic recovery parameters (H, ν, h, r) are likely to be both temperature and strain-rate dependent. Power-law and Arrhenius relationships have sometimes been used (Ref 24, 29):

$$M = M_0 \exp\left(-\frac{Q_M}{RT}\right) \qquad \text{(Eq 26a)}$$

$$k_N = k_{N0} \exp\left(-\frac{Q_N}{RT}\right) \qquad \text{(Eq 26b)}$$

$$h = h_0 \left(\frac{\dot{\varepsilon}}{\dot{\varepsilon}_0}\right)^{m_h} \exp\left(\frac{m_h Q_h}{RT}\right) \qquad \text{(Eq 27a)}$$

$$r = r_0 \left(\frac{\dot{\varepsilon}}{\dot{\varepsilon}_0}\right)^{-m_r} \exp\left(-\frac{m_r Q_r}{RT}\right) \qquad \text{(Eq 27b)}$$

where M_0, k_{N0}, h_0, and r_0 are constants, usually $\dot{\varepsilon}_0 = 1 s^{-1}$, R is the gas constant, and m_h (>0) and m_r (>0) are the strain-rate sensitivities of h and r, respectively; similarly, Q_M, Q_N, Q_h, and Q_r denote the apparent activation energies of M, k_N, h, and r. Equations 26(a) and (b) indicate that M and k_N both increase with temperature. On the other hand, Eq 27(a) and (b) show that h increases with strain rate and decays with increasing temperature, whereas the opposite holds for r. A set of values of h, r, and their strain-rate and temperature dependences are given in Table 3 for various materials. Combining Eq 27 with Eq 21 gives:

$$\sigma_\infty = \alpha \mu b \sqrt{\frac{h}{r}}$$
$$= \alpha \mu b \sqrt{\frac{h_0}{r_0}} \left(\frac{\dot{\varepsilon}}{\dot{\varepsilon}_0}\right)^{(m_h + m_r)/2} \exp\left(\frac{m_h Q_h + m_r Q_r}{2RT}\right) \qquad \text{(Eq 28)}$$

This means that the steady-state flow stress predicted by the YLJ equation (i.e., without DRX) exhibits, in turn, power-law and Arrhenius-type dependences, which justifies the aforementioned choices for h and r. Similar assumptions could be proposed for the strain-rate and temperature dependence of H and ν in the power-law constitutive Eq 17. Note that the model parameters can also be changed by the presence of solutes and second-phase precipitates in a given alloy.

The strain-hardening and dynamic recovery parameters (H, ν, h, r) can be readily determined from the experimental stress-strain relationships for each deformation temperature and strain rate. However, only the portion of the work-hardening curve prior to the onset of DRX should be used (i.e., $\varepsilon < \varepsilon_c$). Direct fitting procedures with Eq 19 or 22 are very efficient. Alternatively, special plots can be used to reduce the problem to a simple linear regression.

Some Numerical Predictions

To obtain numerical results, the equations of the previous section can be applied step by step as follows:

1. Choose a suitable stress-strain equation and determine the relevant parameters, for example, H, ν, or h, r.
2. Estimate M and k_N. (For M, literature data are available; however, a trial-and-error method shall be used at least for k_N.)
3. Define a set of grains D_i (i = N) of average size D_0.
4. Then, for each time (or strain) increment:
   a. Compute new D_i's (Eq 14a or b). If D_i is less than some critical value close to zero, remove the grain from the set.
   b. Compute new ρ_i's (Eq 24a and b).
   c. Compute new N^+. If $N^+ > 1$, add a new grain to the set.
   d. Compute new $\bar{\rho}$ (Eq 16), σ, and $\bar{D}$.
   e. Compute the recrystallized fraction, defined as the ratio of volume of new grains/total volume.

Figure 10(a) shows stress-strain curves predicted by the DDRX mesoscale model using material parameters for a high-purity 304L stainless steel deformed at 1050 °C and 10^{-2} s^{-1}. The strain-hardening parameters h_i (Eq 20) and initial grain sizes D_i^0 were uniformly distributed within intervals $h_0 \pm \Delta h$ and $D_0 \pm \Delta D$, with $\Delta h = 0.05 h_0$ and $\Delta D = 0.3 D_0$, respectively. The average grain size evolutions are displayed in Fig.10(b), which shows that grain size increase or moderate decrease is associated with oscillating stress-strain curves, whereas single-peak flow curves are predicted when grain size is strongly reduced. This is in good agreement

Table 3 Strain-hardening and dynamic recovery parameters of various materials

Material	Range of validity		Parameters						Reference
	Temperature °C	Strain rate, s^{-1}	h_0, μm^{-2}	m_h	Q_h, kJ·mol^{-1}	r_0	m_r	Q_r, kJ·mol^{-1}	
Aluminum alloy 6060	350–500	10^{-2}–3	35.89	0.15	90	2497	0.15	210	30
Stainless steel 304(a)	850–1100	10^{-3}–10^{-1}	1.03	0.27	200	...	0	r increases from 2 (850 °C) to 7 (1090 °C)	31
Stainless steel 304L(b)	850–1150	10^{-4}–10^{-1}	0.48	0.4	280(c)	45	0.066	280(c)	32
Stainless steel 304H(d)	850–1150	10^{-4}–10^{-1}	4.65	0.26	280(c)	110	0.096	280(c)	32
Pure Ni	800–1000	0.03–0.1	2.42	0.14	335	3–10	≈0	r increases with temperature ≈0	33
Ni-1Nb	800–1000	0.03–0.1	10.85	0.14	335	≈6	≈0	r increases with temperature ≈0	33

(a) High purity, $< 5 \times 10^{-6}$ C. (b) Low carbon, <0.03C. (c) Prescribed value. (d) High carbon, <0.10C. Note: When $m_r = 0$ it is not possible to estimate Q_r from Eq 27b so behavior of r is given.

with experimental data and is rationalized by the Jonas and Sakai criterion as follows. Finally, Fig.10(c) illustrates the strain dependence of the fraction recrystallized, which can be defined as the volume fraction of grains that were not present at the onset of straining. It is worthwhile to note, however, that it is not equivalent to its counterpart in static recrystallization, since here, recrystallized grains can contain a wide range of dislocation densities, according to their specific strain. Figure 10(c) shows clearly

(a)

(b)

(c)
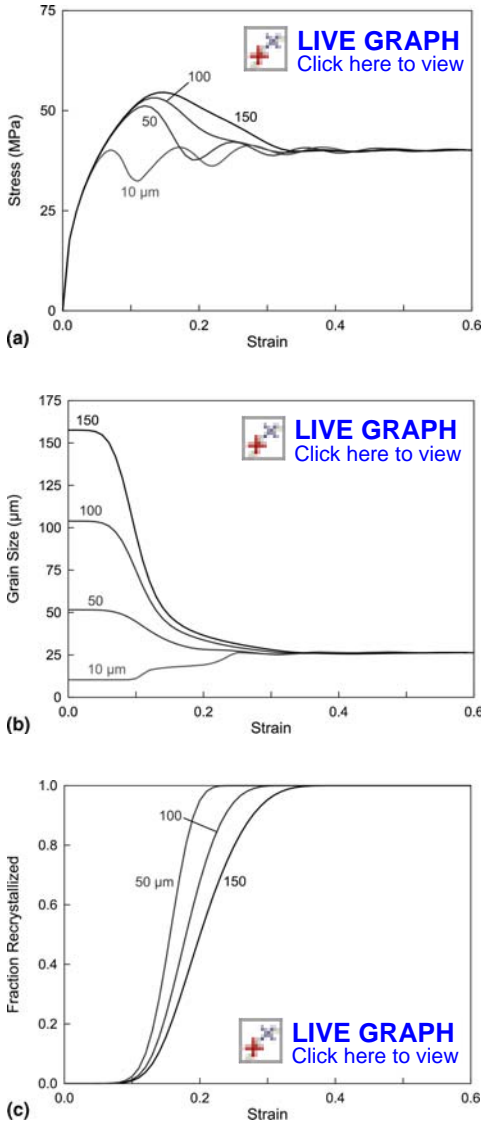
Fig. 10 (a) Stress-strain curves showing the numerical response of the system. The initial grain size varies between 10 and 150 μm, and the curves from the multiple peak behavior to the single-peak one. The stress tends to a unique steady-state value. Material parameters used here are $M\tau = 0.278$ μm³ · s⁻¹, $k_N = 5.8 \times 10^{-7}$ μm⁴ · s⁻¹, p = 3, $h_0 = 80.9$ μm⁻², and r = 5 (refer to the Yoshie-Laasraoui-Jonas equation). (b) Strain dependence of the average grain size. The initial grain size varies between 10 and 150 μm. The average grain size tends to a unique steady-state value. Same parameters as in (a). (c) Strain dependence of the fraction recrystallized. The initial grain size varies between 50 and 150 μm. Same parameters as in (a)

that increasing the initial grain size decreases the rate of DDRX, which is meaningful since nucleation has been assumed to take place at grain boundaries.

Simplified Analysis for the Steady State

The previous numerical results show that the system of equations used to model DDRX naturally leads to steady-state behavior, where the average flow stress, σ, and average grain size, $\bar{D}$, no longer depend on time (or strain). Furthermore, assuming that all the grains of the aggregate have the same behavior, that is, their constitutive parameters h and r are the same, their histories described by the variables $D(\varepsilon)$ and $\rho(\varepsilon)$ are identical (here, ε is the specific strain of the grain under consideration). The ergodic assumption can then be applied to the steady-state system. This means that the investigation of the whole set of grains at any given time is tantamount to the observation of any single grain along its lifetime, from its nucleation (ε = 0) to its disappearance (ε = ε_{end}). Accordingly, summations over the grains can be replaced by integrals with respect to time or strain. For example, the average grain size can be written:

$$\bar{D} = \frac{\sum D_i}{N} = \frac{1}{\varepsilon_{\text{end}}} \int_0^{\varepsilon_{\text{end}}} D(\varepsilon)d\varepsilon \qquad \text{(Eq 29)}$$

This approach is illustrated using the power law (Eq 17) without the BMIS term, since, in this case, closed-form analytical predictions can be obtained.

Flow stress and average grain size with $\rho_0 = 0$ (Eq 18) can be written:

$$\rho = (\nu + 1)^{1/(\nu+1)} H \varepsilon^{1/(\nu+1)} \qquad \text{(Eq 30)}$$

The combination of Eq 14(b) and 30 gives:

$$D = \frac{2M\tau}{\dot{\varepsilon}} H \varepsilon \left[\frac{\bar{\rho}}{H} - \frac{(\nu+1)^{(\nu+2)/(\nu+1)}}{\nu+2} \varepsilon^{1/(\nu+1)} \right] \qquad \text{(Eq 31)}$$

assuming that the initial grain (nucleus) size at ε = 0 is zero. The aforementioned relationship shows that D also vanishes for:

$$\varepsilon_{\text{end}} = \frac{(\nu+2)^{\nu+1}}{(\nu+1)^{\nu+2}} \left(\frac{\bar{\rho}}{H} \right)^{\nu+1} \qquad \text{(Eq 32)}$$

(end of grain life) and reaches its maximum value:

$$D_{\max} = \frac{2M\tau}{\dot{\varepsilon}} H \frac{1}{(\nu+1)(\nu+2)} \left(\frac{\bar{\rho}}{H} \right)^{\nu+2} \qquad \text{(Eq 33)}$$

for

$$\varepsilon_{\max} = \frac{1}{\nu+1} \left(\frac{\bar{\rho}}{H} \right)^{\nu+1} \qquad \text{(Eq 34)}$$

The strain dependences of ρ and D are illustrated in Fig. 11 in the case of 304L steel deformed at 1050 °C and two different strain rates. Numerical values of the parameters are given in Table 4.

In its integral form, Eq 25(b) becomes:

$$\frac{k_N \bar{\rho}^p}{\dot{\varepsilon}} \int_0^{\varepsilon_{\text{end}}} D^2 d\varepsilon = 1 \qquad \text{(Eq 35)}$$

which means that exactly one new grain is nucleated on average over the life of any grain
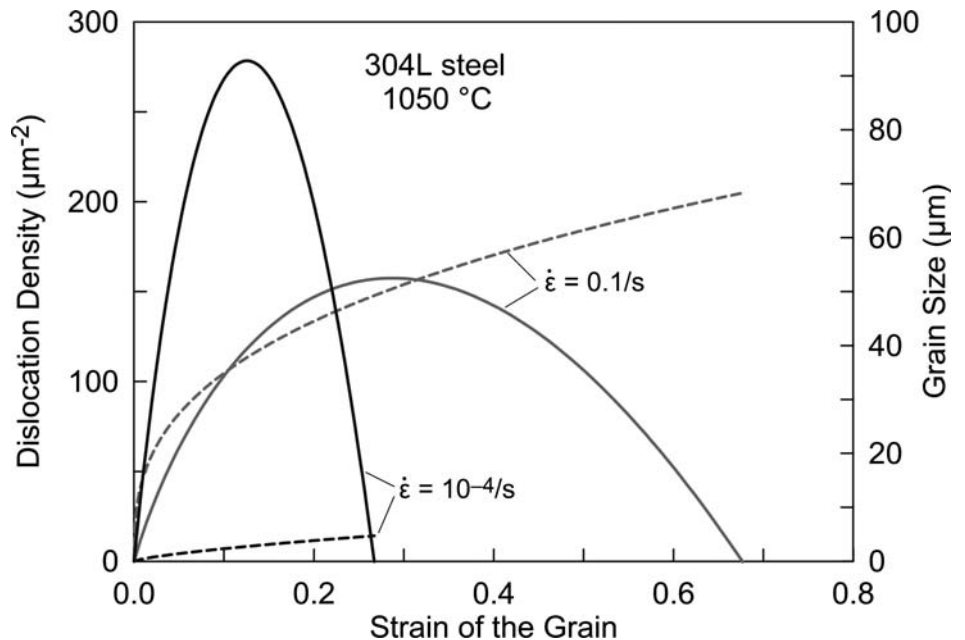
Fig. 11 Strain dependence of the grain size D (solid lines) and the dislocation density ρ (broken lines) over the lifetime of a grain during steady-state discontinuous dynamic recrystallization. Data pertaining to 304 steel were used at 1050 °C and two strain rates (Table 4).

Table 4   Parameters used in Fig. 11

Strain-hardening and recovery parameters	Strain rate ($\dot{\varepsilon}$), s^{-1}	
	10^{-4}	10^{-1}
H, μm^{-2}	28.7	162.8
ν, unitless	0.40	1.85
$M\tau$, μm^3/s	0.011	0.232
$\bar{\rho}$, μm^{-2}	8.29	151.81

Source: Ref 34

during steady state. Substituting Eq 31 into Eq 35 and solving for $\bar{\rho}$ gives:

$$\bar{\rho} = \left[C_1 \frac{(H^{\nu+1}\dot{\varepsilon})^3}{M^2\tau^2 k_N} \right]^{\frac{1}{p+3\nu+5}} \quad \text{(Eq 36a)}$$

where

$$C_1 = \tfrac{3}{8}(3\nu+4)(3\nu+5)\frac{(\nu+1)^{3(\nu+2)}}{(\nu+2)^{3(\nu+1)}} \quad \text{(Eq 36b)}$$

from which the steady-state flow stress $\sigma_s = \alpha\mu b\sqrt{\bar{\rho}}$ is readily derived as a function of strain rate and the material parameters H, ν, M, k_N, and p (which can themselves be both strain-rate and temperature dependent):

$$\sigma_s = \alpha\mu b \left[C_1 \frac{(H^{\nu+1}\dot{\varepsilon})^3}{M^2\tau^2 k_N} \right]^{\frac{1}{2(p+3\nu+5)}} \quad \text{(Eq 37)}$$

The average grain size can now be computed from Eq 29:

$$\bar{D} = C_2 \left[\frac{(M\tau)^{p+\nu+1}}{(H^{\nu+1}\dot{\varepsilon})^{p-1} k_N^{\nu+2}} \right]^{\frac{1}{p+3\nu+5}} \quad \text{(Eq 38a)}$$

where

$$C_2 = \frac{1}{2\nu+3}\frac{(\nu+2)^{\nu+1}}{(\nu+1)^{\nu+2}} C_1^{(\nu+2)/(p+3\nu+5)} \quad \text{(Eq 38b)}$$

From the previous expressions for σ_s and $\bar{D}$, which are both easily measurable quantities, closed-form formulae can be derived for the parameters $M\tau$ (where M is the grain-boundary mobility) and k_N (nucleation rate), which, in contrast, are quite difficult to determine directly. Indeed, solving Eq 37 and 38(a) for $M\tau$ and k_N gives:

$$M\tau = (2\nu+3)\frac{(\nu+1)^{\nu+2}}{(\nu+2)^{\nu+1}} H^{\nu+1}\dot{\varepsilon}\frac{\bar{D}}{(\sigma_s/\alpha\mu b)^{2(\nu+2)}} \quad \text{(Eq 39a)}$$

$$k_N = \tfrac{3}{8}\frac{(3\nu+4)(3\nu+5)}{(2\nu+3)^2}\frac{(\nu+1)^{\nu+2}}{(\nu+2)^{\nu+1}}\frac{H^{\nu+1}\dot{\varepsilon}}{(\sigma_s/\alpha\mu b)^{2(p+\nu+1)}\bar{D}^2} \quad \text{(Eq 39b)}$$

where the parameters H and ν accounting for strain hardening and dynamic recovery can also be measured from the transient part of the stress-strain curves. In the special case where ν = 0 (parabolic hardening) and assuming p = 3 (see previous), quite simple expressions are found:

$$M\tau = \tfrac{3}{2} H\dot{\varepsilon}\frac{\bar{D}}{(\sigma_s/\alpha\mu b)^4} \quad \text{(Eq 40a)}$$

$$k_N = \tfrac{5}{12}\frac{H\dot{\varepsilon}}{(\sigma_s/\alpha\mu b)^8\bar{D}^2} \quad \text{(Eq 40b)}$$

The Derby Relationship. It is quite remarkable that Eq 37 and 38(a) can be solved for $H^{\nu+1}\dot{\varepsilon}$, thus providing a simple relationship between $\bar{D}$ and σ_s:

$$\sigma_s = \alpha\mu b \left(\frac{M\tau}{k_N}\frac{C_\nu}{\bar{D}^3} \right)^{1/2(p-1)} \quad \text{(Eq 41a)}$$

where

$$C_\nu = \tfrac{3}{8}\frac{(3\nu+4)(3\nu+5)}{(2\nu+3)^3} \quad \text{(Eq 41b)}$$

which can be written in the form:

$$\sigma_s = \frac{k_D}{\bar{D}^a} \quad \text{(Eq 41c)}$$

where the exponent a = $3/[2(p-1)]$.

The aforementioned inverse power-law correlation has been reported a long time ago from a number of experimental data pertaining to various materials (undergoing both DDRX and CDRX) by Derby (Ref 35). The empirical exponent a of $\bar{D}$ is generally observed to range between 0.5 and 1, with the majority of measurements around 0.75. This indicates that the nucleation exponent p should be chosen close to 3, which gives a = $\tfrac{3}{4}$. Furthermore, the empirical Derby relationship seems to be almost independent of strain rate and temperature, provided the flow stress is normalized by μ. As a consequence, k_D in Eq 41(c) is likely to be independent of (or weakly dependent on) $\dot{\varepsilon}$ and T. In particular, the apparent activation energies of M and k_N are expected to be similar.

Strain Rate and Temperature Dependence of the Steady-State Flow Stress. The strain-rate sensitivity of the steady-state flow stress, defined as $m = \partial\ln\sigma_s/\partial\ln\dot{\varepsilon}$, can be derived from Eq 41(a), assuming to a first approximation that ν and therefore C_ν are independent of $\dot{\varepsilon}$ (as well as $M\tau$ and k_N; compare with Eq 26 and 27):

$$m = \frac{3[(\nu+1)m_H+1]}{2(p+3\nu+5)} \quad \text{(Eq 42)}$$

where $m_H = \partial\ln H/\partial\ln\dot{\varepsilon}$ is the strain-rate sensitivity of the hardening parameter H. For ν = 0 (parabolic hardening, $n = 0.5$), ν = 4 ($n = 0.1$), and p = 3, this leads to $m = (\tfrac{3}{16})(1 + m_H)$ and $m = (\tfrac{3}{40})(1 + 5m_H)$, respectively, which indicates that the weight of m_H in the overall strain-rate sensitivity increases with ν, that is, when strain hardening weakens.

In the same way, the apparent activation energy of the flow stress, defined as $Q = (R/m)[\partial\ln\sigma_s/\partial(1/T)]$, is derived in turn from Eq 41(a):

$$Q = \frac{3(\nu+1)m_H Q_H + 2Q_M + Q_N}{3[(\nu+1)m_H+1]} \quad \text{(Eq 43)}$$

where $Q_H = (R/m_H)[\partial\ln H/\partial(1/T)]$, and R is the gas constant. According to this equation, the overall activation energy is a combination of the activation energies pertaining to strain hardening (and dynamic recovery), grain-boundary migration, and grain nucleation. Whenever $Q_H \approx Q_M \approx Q_N \approx Q_a$ (the activation energy of diffusion), Eq 43 reduces to $Q \approx Q_a$.

The Sakai and Jonas Criterion

In the transient stage of DDRX, the flow curves exhibit a transition from a single-peak behavior at large values of the Zener-Hollomon parameter $Z = \dot{\varepsilon}\exp(Q/RT)$ to an oscillatory behavior when Z is low. Sakai and Jonas (Ref 36) proposed a criterion involving the initial grain size D_0 and the steady-state grain size $\bar{D}$. If DDRX leads to grain refinement (more precisely, if $D_0 \geq 2\bar{D}$), the initial microstructure contains few nucleation sites (on the grain boundaries). The recrystallization process is associated with necklace formation until new nuclei appear, which prevents synchronization. The transient flow curve then exhibits one single peak. By contrast, if the average grain size increases ($D_0 \leq 2\bar{D}$), the presence of a large number of nucleation sites allows synchronized recrystallization. The transient flow curve is then characterized by a series of oscillations. The mesoscale DDRX model leads to a physical background to the Sakai and Jonas relationship.

The average grain-boundary velocity can first be estimated from the steady-state model:

$$\bar{v}_{GB} = \frac{D_{max}}{t_{end}} = \frac{D_{max}}{\varepsilon_{end}}\dot{\varepsilon} \quad \text{(Eq 44)}$$

Using again the simple power-law model for strain hardening, the combination of Eq 32 and 33 gives:

$$\bar{v}_{GB} = 2M\tau\frac{(\nu+1)^{\nu+1}}{(\nu+2)^{\nu+2}}\bar{\rho} \quad \text{(Eq 45)}$$

On the other hand, the nucleation rate per unit time and area is $\dot{n} = k_N\bar{\rho}^3$, assuming a nucleation exponent p = 3 (see previous text). The ratio $v_{GB}/\dot{n}$ (which has dimensions of a volume) can be used to estimate the influence of grain-boundary velocity with respect to nucleation rate. Deriving $\bar{\rho}$ as a function of $\bar{D}$ from Eq 41(a) leads to:

$$\frac{\bar{v}_{GB}}{\dot{n}} = C_\nu'\bar{D}^3 \quad \text{(Eq 46a)}$$

where

$$C_\nu' = \frac{16}{3}\frac{(\nu+1)^{\nu+1}}{(\nu+2)^{\nu+2}}\frac{(2\nu+3)^3}{(3\nu+4)(3\nu+5)} \quad \text{(Eq 46b)}$$

The factor C_ν' decreases from 1.800 for ν = 0 to 1.748 for ν = 4 and will therefore be taken

equal to 2 for the sake of simplicity. $D_0 \geq 2\bar{D}$ is then equivalent to:

$$\frac{\bar{v}_{GB}}{\dot{n}} \leq \frac{D_0^3}{4} \qquad \text{(Eq 47a)}$$

for grain refinement to occur (single-peak flow curve) and conversely:

$$\frac{\bar{v}_{GB}}{\dot{n}} > \frac{D_0^3}{4} \qquad \text{(Eq 47b)}$$

for grain coarsening (multiple-peak flow curve).

Mesoscale Approach for CDRX

Similar to DDRX, the following mesoscale (or grain scale) model for CDRX allows to take into account the basic physical mechanisms and to predict microstructural evolutions occurring during hot deformation.

The basic equations pertaining to DDRX must be modified to deal with CDRX, since it is relevant in that case to focus on subgrains and grain boundaries instead of grains (Ref 29). The main features of the CDRX mesoscale model are first described, after which it is illustrated by some examples.

Description of the Model. Figure 12 shows a schematic representation of a CDRX microstructure, made of an aggregate of crystallites. Each of them is delimited partly by (high-angle) grain boundaries (HAB) and partly by (low-angle) subgrain boundaries (LAB), as commonly observed at large strains in high-stacking-fault-energy metals (e.g., aluminum alloys, ferritic steels). The model involves three internal variables: the average dislocation density ρ_i inside the crystallites, their average diameter D (scalar parameters), and the distribution function $\varphi(\theta)$ of the subgrain-boundary misorientations. The latter is defined as follows: $\varphi(\theta)d\theta$ is the area fraction of LAB, with misorientations ranging between θ and $\theta + d\theta$. Here, θ denotes the minimum angle of misorientation between two adjacent crystals. It ranges between θ_0, the angle at which a LAB can be considered to be generated, and $\theta_c \approx 15°$, the angle beyond which the
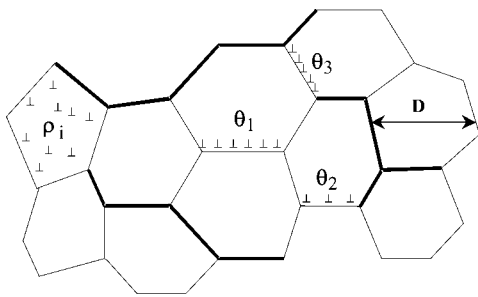
boundary can no longer be considered as an array of dislocations and therefore transforms into a HAB. The area fraction of LAB is then given by:

$$f_{LAB} = \int_{\theta_0}^{\theta_c} \varphi(\theta)\,d\theta \qquad \text{(Eq 48)}$$

whence, obviously, $f_{HAB} = 1 - f_{LAB}$.

Furthermore, the dislocation length per unit area in a LAB of misorientation angle θ is:

$$L = n\,\theta/b \qquad \text{(Eq 49)}$$

where n = 1, 2, or 3, according to whether the boundary is made of an array of 1, 2, or 3 sets of parallel dislocations, and b is the Burgers vector length. Thus, the density of dislocations in the LABs is:

$$\rho_{LAB} = \int_{\theta_0}^{\theta_c} L\,S\,\varphi(\theta)\,d\theta \qquad \text{(Eq 50)}$$

where S is the total area of boundaries per unit volume.

Assuming the average crystallite size D equals the mean intercept length (accounting for all boundaries), the stereological relationship $D = 2/S$ is used in the following (Ref 37).

The flow chart in Fig. 13 summarizes the behavior of dislocations during CDRX. The main part $d\rho_i^{-(1)}$ of the dislocation density $d\rho_i^+$ generated during a strain increment $d\varepsilon$ is subject to dynamic recovery. A fraction α of $d\rho_i^{-(1)}$ is consumed by the creation of new subgrain boundaries (with misorientation angle θ_0), while the remaining fraction $1 - \alpha$ is absorbed by pre-existent LAB or HAB, in direct ratio to their respective area fractions f_{LAB} and f_{HAB}. In turn, the absorption of dislocations by subgrain boundaries induces, on average, a progressive increase in their misorientation angles, as indicated by the gray arrows in the diagram. Whenever θ reaches the critical value θ_c, the LAB continuously transforms into HAB, which then merely acts as a sink for dislocations. Finally, grain-boundary migration leads to the elimination of an additional part $d\rho_i^{-(2)}$ of the dislocations and a boundary area dS^- (black arrows).

Basic Equations. Similar to DDRX (Eq 20 and 24a), the evolution of the dislocation density ρ_i inside the crystallites can be assumed to take the form of a modified Laasraoui-Jonas equation:

$$\frac{d\rho_i}{d\varepsilon} = h - r\,\rho_i - \rho_i\,\frac{dV}{d\varepsilon} \qquad \text{(Eq 51)}$$

where, as usual, h and r denote the strain-hardening and dynamic recovery parameters, which

Fig. 12 Schematic representation of a continuous dynamic recrystallization microstructure made up of an aggregate of crystallites. High- and low-angle grain boundaries are represented with thick and fine lines, respectively.
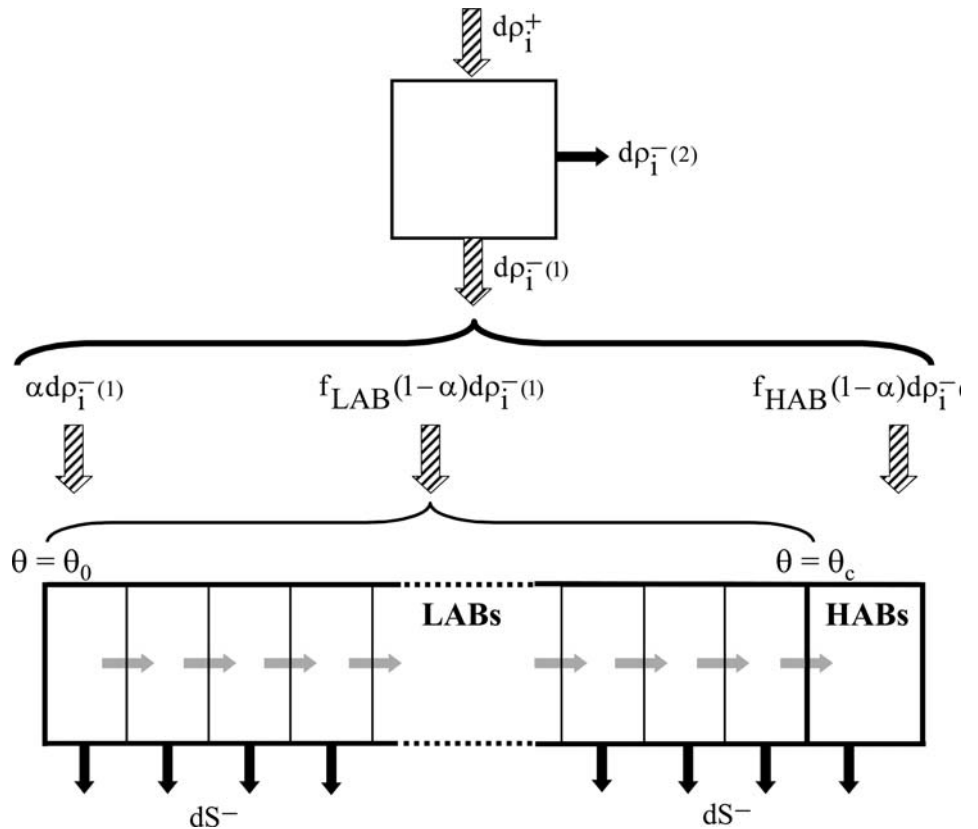
Fig. 13 Flow chart schematically illustrating the behavior of dislocations produced by strain hardening during continuous dynamic recrystallization (hatched arrows), the continuous increase of low-angle boundary (LAB) misorientations (gray arrows), and the absorption of dislocations and boundaries by high-angle grain boundaries (HABs, black arrows)

are generally dependent on strain rate and temperature (Eq 27a and b), and dV is the volume swept by mobile boundaries during the strain increment $d\varepsilon$. The last term on the right side of Eq 51 is associated with BMIS, already introduced for DDRX. Assuming for the sake of simplicity that boundary migration is restricted to HAB, it is easy to find:

$$\frac{dV}{d\varepsilon} = \frac{2\, f_{HAB}\, v_m}{D\dot{\varepsilon}} \qquad \text{(Eq 52)}$$

where v_m is the rate of grain-boundary migration, which is likely to depend on temperature. Simple geometric considerations lead to a second basic equation:

$$\frac{d\theta}{d\varepsilon} = \frac{b}{2n}\,(1-\alpha)\, r\rho_i\, D \qquad \text{(Eq 53)}$$

The change in the total area of boundaries per unit volume is $dS = dS^+ - dS^-$, where dS^+ is the area of LAB generated during the strain increment $d\varepsilon$, and $dS^- = S\, dV$ is the area annihilated by the movement of HAB. Therefore:

$$\frac{dS^+}{d\varepsilon} = \frac{b}{n\theta_0}\,\alpha r\rho_i \qquad \text{(Eq 54a)}$$

and

$$\frac{dS^-}{d\varepsilon} = \frac{f_{HAB}\, S^2\, v_m}{\dot{\varepsilon}} \qquad \text{(Eq 54b)}$$

Finally, the evolution of the distribution function $\varphi(\theta)$ is obtained using a first-order series expansion in $d\varepsilon$ in the form (Ref 29):

$$\varphi(\theta+d\theta, \varepsilon+d\varepsilon) = \left[1 + \frac{1}{D}\frac{dD}{d\varepsilon} - dV\right]\varphi(\theta,\varepsilon) \qquad \text{(Eq 55)}$$

The material flow stress is then derived from the aforementioned microstructural parameters.

It depends primarily on the dislocation density ρ_i inside the crystallites. However, an additional term involving the dislocation density ρ_{LAB} within the subgrain boundaries is usually introduced, leading to:

$$\sigma = \alpha\mu b\big(A_1\sqrt{\rho_i} + A_2\sqrt{\rho_{LAB}}\big) \qquad \text{(Eq 56)}$$

where $A_1 \approx 1$ and $A_2 << A_1$.

Some Predictions of the CDRX Mesoscale Model. To obtain numerical results, the equations of the previous section can be applied step by step as follows:

1. Choose a suitable stress-strain equation and determine the relevant parameters, for example, h, r for the Laasraoui-Jonas relationship.
2. Estimate the average grain-boundary migration rate v_m using literature data and/or a trial-and-error method.
3. Choose initial D, ρ_i, and $\varphi(\theta)$ (e.g., $\varphi(\theta) = 0$ for all θ if no LABs are present in the initial state). Choose $\alpha < 1$ by trial and error. (Tests have shown that $\alpha < 0.1$.)
4. Then, for each time (or strain) increment:
   a. Compute new S (Eq 54a, b), $D = 2/S$, f_{LAB} (Eq 48), and f_{HAB}.
   b. Compute new ρ_i (Eq 51, 52).
   c. Compute new $\varphi(\theta)$ (Eq 55) and ρ_{LAB} (Eq 50).
   d. Finally, compute σ (Eq 56).

Figure 14(a) shows stress-strain curves obtained from the aforementioned set of equations, using physical parameters pertaining to AA1200-grade aluminum. Two variants of Eq 56 were used here, namely $A_1 = 1$, $A_2 = 0$ (solid lines) and $A_1 = 0.9$, $A_2 = 0.1$ (broken lines). The strain dependence of the crystallite size is illustrated in Fig. 14(b) at 460 °C and 0.01 s^{-1} for various initial grain sizes. Comparison of these two figures with Fig. 10(a) and (b)

shows that CDRX takes place much more slowly than DDRX. In particular, a significantly larger strain is required to reach the steady state. Furthermore, no oscillations of the flow stress are visible, and this holds whatever the straining conditions, in agreement with experimental data. By contrast, commonly observed flow softening is not correctly reproduced, which may be due to the fact that texture changes are not accounted for in the model.

With the conditions that the total area of boundaries per unit volume S and the dislocation density inside the crystallites ρ_i remain constant, the mesoscale CDRX model leads to analytical expressions for the steady-state flow stress and microstructural parameters (Ref 29). In particular, it predicts a decreasing exponential form for the distribution function of the subgrain-boundary misorientations:

$$\varphi(\theta) = \varphi(\theta_0)\exp[-\varphi(\theta_0)(\theta-\theta_0)] \qquad \text{(Eq 57a)}$$

where

$$\varphi(\theta_0) = \frac{\alpha}{(1-\alpha)\theta_0} \qquad \text{(Eq 57b)}$$

REFERENCES

1. S. Gourdet and F. Montheillet, *Mater. Sci. Eng. A*, Vol 283, 2000, p 274
2. T.R. De Oliveira and F. Montheillet, *Steel Research Int.* Vol 79, 2008, p 497
3. K. Máthis and E.F. Rauch, *Mater. Sci. Eng. A*, Vol 462, 2007, p 248
4. A. Vorhauer and R. Pippan, *Metall. Mater. Trans. A*, Vol 39, 2008, p 417
5. S.M. Lim, M. El Wahabi, C. Desrayaud, and F. Montheillet, *Mater. Sci. Eng. A*, Vol 460–461, 2007, p 532
6. M. Avrami, *J. Chem. Phys.*, Vol 7, 1939, p 1103
7. M. Avrami, *J. Chem. Phys.,* Vol 8, 1940, p 212
8. W.A. Johnson and R.F. Mehl, *Trans. AIME*, Vol 135, 1939, p 416
9. A.N. Kolmogorov, *Izv. Acad. Nauk SSSR, Ser. Mat.,* Vol 3, 1937, p 355
10. J.E. Burke and D. Turnbull, *Progr. In Met. Phys.,* Vol 3, 1952, p 220
11. R.W. Cahn, *Physical Metallurgy*, North-Holland Publishing Co., Amsterdam, 1965, p 925
12. J.J. Li, J.C. Wang, Q. Xu, and G.C. Yang, *Acta Mater.,* Vol 55, 2007, p 825
13. R.A. Vandermeer and P. Gordon, *Recovery and Recrystallization of Metals*, Interscience Publ., New York, 1963, p 211
14. J.H. Brophy, R.M. Rose, and J. Wulff, *Thermodynamics of Structure*, John Wiley & Sons, N.Y., 1964, p 120
15. G. Gottstein, *Physical Foundations of Materials Science*, Springer Verlag, Berlin, 2004, p 334–335
16. D.A. Porter and K.E. Easterling, Van Nostrand Reinhold Co. Ltd., Wokingham, Berkshire, U.K., 1988, p 289
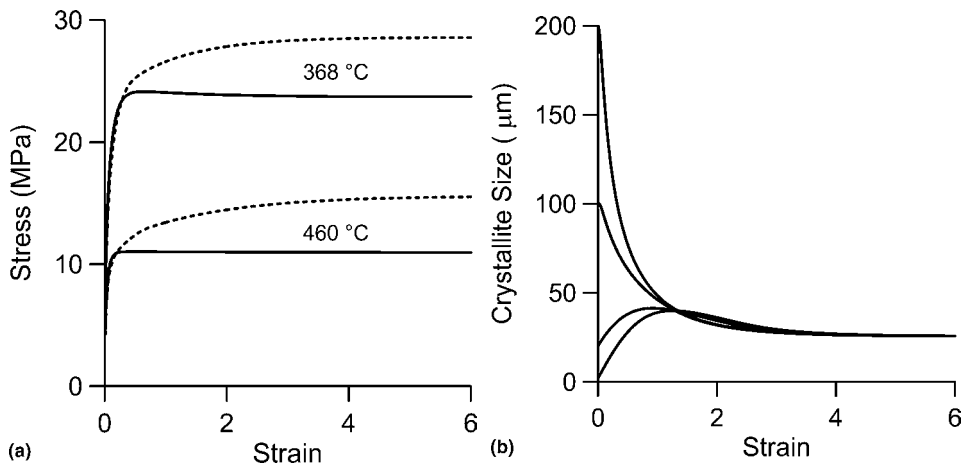
Fig. 14 (a) Stress-strain curves predicted by the continuous dynamic recrystallization (CDRX) model for AA1200-grade aluminum at 368 and 460 °C and a strain rate of 0.1 s^{-1} for two variants of Eq 56: $A_1 = 1$, $A_2 = 0$ (solid lines) and $A_1 = 0.9$, $A_2 = 0.1$ (broken lines). The material parameters used here are h = 158.3 μm^{-2}, r = 9.48, and v_m = 0.95 μm · s^{-1}. (b) Strain dependence of the crystallite size predicted by the CDRX model for AA1200-grade aluminum at 460 °C and 0.01 s^{-1} strain rate for various initial grain sizes. Same material parameters as in (a)

17. C.M. Sellars, *Hot Working and Forming Processes*, G. Davies and C.M Sellars, Ed., The Metals Society, London, 1980, p 3
18. P.D. Hodgson, J.J. Jonas, and C.H.J. Davies, *Handbook of Thermal Process Modeling of Steels*, C.H. Gur, Ed., in press, 2008
19. J.J. Jonas, X. Quelennec, L. Jiang, and E. Martin, *Acta Mater.*, Vol 57, 2009, p 2748
20. A. Dehghan-Manshadi, J.J. Jonas, P.D. Hodgson, and M.R. Barnett, *ISIJ Int.,* Vol 48, 2008, p 208
21. E.I. Poliak and J.J. Jonas, *Acta Metall. et Mater.,* Vol 44, 1996, p 127
22. G.R. Stewart, A.M. Elwazri, S. Yue, and J.J. Jonas, *Mat. Sci. & Technol.,* Vol 22, 2006, p 519
23. J.P. Thomas, F. Montheillet, and S.L. Semiatin, *Metall. Mater. Trans. A,* Vol 38, 2007, p 2095
24. F. Montheillet, Moving Grain Boundaries during Hot Deformation of Metals: Dynamic recrystallization. *Moving Interfaces in Crystalline Solids*, CISM Courses and Lectures—No. 453, SpringerWien, New York, 2004, p 203
25. F. Montheillet, O. Lurdos, and G. Damamme, *Acta Mater.*, Vol 57, p 1602
26. A. Laasraoui and J.J. Jonas, *Metall. Trans. A,* Vol 22, 1991, p 1545
27. Y. Estrin and H. Mecking, *Acta Metall.,* Vol 32, 1984, p 57
28. J. Hertel, *Ein Modell der dynamischen Erholung von Stählen während der Warmumformung*. Doktor-Ingenieurs Dissertation, Fakultät für Bergbau und Hüttenwesen der Rheinisch-Westfälischen Technischen Hochschule Aachen, 1979
29. S. Gourdet and F. Montheillet, *Acta Mater.,* Vol 51, 2003, p 2685
30. C. Chovet-Sauvage, "Evolution des microstructures et des textures en grande déformation à chaud d'un alliage Al-Mg-Si", PhD. Thesis, Ecole des Mines, Saint-Etienne, France, 2000
31. L. Gavard, "Recristallisation dynamique d'aciers inoxydables austénitiques de haute pureté", PhD. Thesis, Ecole des Mines, Saint-Etienne, France, 2001
32. M. El Wahabi, J.M.Cabrera, and J.M. Prado, *Mater. Sci. Eng. A,* Vol 343, 2003, p 116
33. F. Montheillet and J. Le Coze, EOARD Contract No. FA8655-06-M-4001, Final Report, 2007
34. O. Lurdos, PhD. Thesis, Ecole des Mines, Saint-Etienne, France, 2008
35. B. Derby, *Scripta Metall. Mater.,* Vol 27, 1992, p 1581
36. T. Sakai and J.J. Jonas, *Acta Metall.,* Vol 32, 1984, p 189
37. R.T. De Hoff and F.W. Rhines, *Quantitative Microscopy*, McGraw-Hill, New York, 1968

Crystal-Plasticity Fundamentals

Henry R. Piehler, Carnegie Mellon University

THE PURPOSE of this article is to enable the reader to understand through examples the fundamentals of crystal plasticity. The rich historical development of crystal plasticity is traced in some detail to enable the reader to appreciate and critically analyze more sophisticated recent approaches to crystal-plasticity modeling. While most of the examples involve cubic metals deforming by rate-insensitive plastic flow, the concepts outlined here can be generalized to other crystal structures and loading conditions as well.

Schmid's Law

Crystal plasticity has as its origin Schmid's law, which states that crystallographic slip is initiated when a critical resolved shear stress on a slip plane in a slip direction is reached. As shown in Fig. 1, crystallographic slip initiates in uniaxial tension when the resolved shear stress
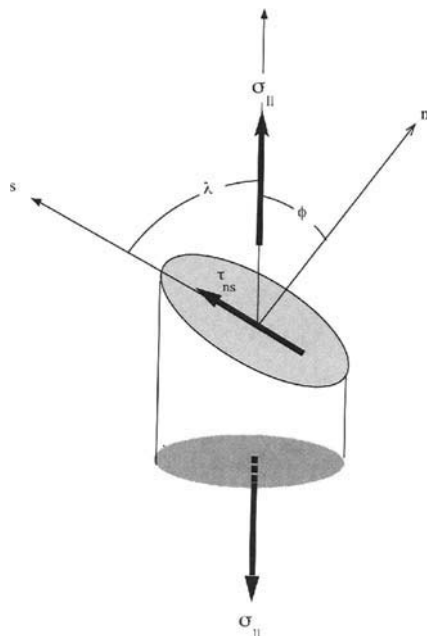
on the slip plane in the slip direction reaches a critical value k (Ref 1). This critical resolved shear stress criterion can be expressed as:

$$\tau_{ns} = \sigma \cos \lambda \cos \phi = k \qquad \text{(Eq 1)}$$

where τ_{ns} is the shear stress on the slip plane, with normal n in the slip direction s; σ is the applied uniaxial stress; $\cos \lambda$ is the cosine of the angle between the tensile axis and the slip direction; $\cos \phi$ is the cosine of the angle between the tensile axis and the slip plane normal; and k is the critical resolved shear stress. If $\hat{\mathbf{l}}$, $\hat{\mathbf{s}}$, and $\hat{\mathbf{n}}$ are unit vectors along the tensile axis l, the slip direction s, and the slip plane normal n, $\cos \lambda$ and $\cos \phi$ can be found from the dot products $\hat{\mathbf{l}} \cdot \hat{\mathbf{s}}$ and $\hat{\mathbf{l}} \cdot \hat{\mathbf{n}}$.

For cubic metals, it is convenient to express the unit vectors $\hat{\mathbf{l}}$, $\hat{\mathbf{s}}$, and $\hat{\mathbf{n}}$ in terms of the unit vectors $\hat{\mathbf{i}}_1$, $\hat{\mathbf{i}}_2$, and $\hat{\mathbf{i}}_3$ along the [100], [010], and [001] crystallographic axes, respectively. For example, if a uniaxial stress is applied in the [001] direction of a face-centered cubic (fcc) single crystal, slip will occur on the (111) plane in the $[0\bar{1}1]$ direction when the normal stress σ reaches k/cos λ cos ϕ (Eq 1). For slip to occur on this $(111)[0\bar{1}1]$ slip system in response to a uniaxial normal stress applied along the [001] direction:

$$\hat{\mathbf{l}} = \hat{\mathbf{i}}_3, \hat{\mathbf{n}} = \frac{1}{\sqrt{3}}[\hat{\mathbf{i}}_1 + \hat{\mathbf{i}}_2 + \hat{\mathbf{i}}_3], \text{ and } \hat{\mathbf{s}} = \frac{1}{\sqrt{2}}[-\hat{\mathbf{i}}_2 + \hat{\mathbf{i}}_3]$$

Performing the dot products $\hat{\mathbf{l}} \cdot \hat{\mathbf{s}}$ and $\hat{\mathbf{l}} \cdot \hat{\mathbf{n}}$ yields:

$$\sigma_{[001]} = \sqrt{6}k = 2.45\,k$$

It should be noted that the yield stress must be at least 2 times the critical resolved shear stress, the minimum value achievable when both the slip plane normal and the slip direction are oriented at 45° to the uniaxial stress axis. As seen from Eq 1 and indicated in Table 1, Schmid's law is identical for fcc metals deforming on {111}<110> systems and body-centered cubic metals deforming on {110}<111> systems; only the identity of ϕ and λ are interchanged.

One of the earliest crystal-plasticity calculations was the determination by Sachs (Ref 2) of the uniaxial yield stress of an isotropic fcc

metal in terms of the critical resolved shear stress k. Sachs' calculation was an average over the stereographic triangle of the uniaxial stresses necessary to initiate slip on the most highly stressed {111}<110> slip system(s). This isostress model predicted that the uniaxial yield stress of an isotropic fcc metal should equal 2.22k. It is also noteworthy that the single-crystal yield stress in the [111] direction is equal to:

$$1.5\sqrt{6}k$$

50% higher than the isostress yield stress for the [001] orientation calculated earlier and 65% above the Sachs average. It is this dependence of the single-crystal yield stress on crystallographic orientation that provides the basis for texture hardening (Ref 3).

Generalized Schmid's Law

Schmid's law for the response of crystals to a uniaxial normal stress can be generalized to three dimensions to include both normal and shear stresses. A critical resolved shear stress criterion can again be used, this time in response to all six components of stress, chosen initially for convenience along the cubic axes of an fcc crystal. These stresses along the cubic axes include the three normal stresses, σ_{11}, σ_{22}, σ_{33}, and the three independent shear stresses, $\sigma_{12} = \sigma_{21}$, $\sigma_{23} = \sigma_{32}$, and $\sigma_{31} = \sigma_{31}$. These stresses can be referred to as σ_{ij}, where i and $j = 1, 2, 3$.

Expressions for the resolved shear stresses on the twelve crystallographic slip systems (and their negatives) are written in terms of the six independent stresses referred to the cubic axes of the fcc crystal. The notation used by Bishop and Hill (Ref 4, 5) and Bishop (Ref 6) for the twelve slip systems is shown in Fig. 2 and Table 1, where the individual slip planes and their corresponding slip directions are identified, along with the shear strains associated with these systems. The generalized Schmid's law is obtained using a tensor transformation to sum the contributions to each of the twelve shear stresses from the stresses σ_{ij} along the cubic axes of the fcc crystal. This generalized Schmid's law for the critical resolved shear stress τ_{ns} on each of the twelve fcc slip systems is:

Fig. 1 Representation of Schmid's critical resolved shear stress criterion: $\tau_{ns} = \sigma \cos \lambda \cos \phi = k$

Table 1 Slip system notation used by Bishop and Hill for cubic crystals

Slip plane (or direction)	(111)	$(\bar{1}\bar{1}1)$	$(\bar{1}11)$	$(1\bar{1}1)$
Slip direction (or plane)	$[01\bar{1}][\bar{1}01][1\bar{1}0]$	$[0\bar{1}\bar{1}][101][\bar{1}10]$	$[01\bar{1}][101][\bar{1}10]$	$[0\bar{1}\bar{1}][\bar{1}01][110]$
Shear strain	a_1, a_2, a_3	b_1, b_2, b_3	c_1, c_2, c_3	d_1, d_2, d_3
Slip system designation	$(a_1), (a_2), (a_3)$	$(b_1), (b_2), (b_3)$	$(c_1), (c_2), (c_3)$	$(d_1), (d_2), (d_3)$
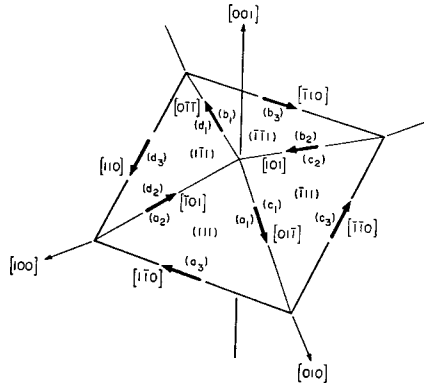
Fig. 2 Schematic representation of Bishop-Hill slip system notation

$$\tau_{ns} = l_{n1}l_{s1}\sigma_{11} + l_{n2}l_{s2}\sigma_{22} + l_{n3}l_{s3}\sigma_{33} + l_{n1}l_{s2}\sigma_{12} + l_{n2}l_{s3}\sigma_{23} + l_{n3}l_{s1}\sigma_{31} + l_{n2}l_{s1}\sigma_{21} + l_{n3}l_{s2}\sigma_{32} + l_{n1}l_{s3}\sigma_{13}$$

(Eq 2a)

or

$$\tau_{ns} = \sum_{j=1}^{3}\sum_{i=1}^{3} l_{ni}l_{sj}\sigma_{ij}$$

(Eq 2b)

where l_{ni} and l_{sj} are the direction cosines of the angle between the particular slip plane normal n and the *i*th cubic axis and the direction cosine of the angle between the particular slip direction s and the *j*th cubic axis. Equation 2(a) can also be written using the summation convention that the repeated indices *i* and *j* are summed from 1 to 3, or:

$$\tau_{ns} = l_{ni}l_{sj}\sigma_{ij}$$

(Eq 2c)

These direction cosines can again be found from the dot products $l_{ni} = \hat{\mathbf{n}} \cdot \hat{\mathbf{i}}_i$ and $l_{sj} = \hat{\mathbf{s}} \cdot \hat{\mathbf{i}}_j$. Calculating the generalized Schmid's law for the $(111)[0\bar{1}1]$ system considered previously gives:

$$l_{n1} = l_{n2} = l_{n3} = \frac{1}{\sqrt{3}} \quad \text{and}$$

$$l_{s1} = 0, l_{s2} = -\frac{1}{\sqrt{2}}, l_{s3} = \frac{1}{\sqrt{2}}$$

and the generalized Schmid's law for this system becomes:

$$\tau_{ns} = \frac{1}{\sqrt{6}}(-\sigma_{22} + \sigma_{33} - \sigma_{12} + \sigma_{23} - \sigma_{32} + \sigma_{13})$$

Since $\sigma_{23} = \sigma_{32}$, this reduces to:

$$\tau_{ns} = \frac{1}{\sqrt{6}}(\sigma_{22} + \sigma_{33} - \sigma_{12} + \sigma_{13})$$

This $(111)[0\bar{1}1]$ system is designated as $-a_1$ by Bishop and Hill (Ref 4, Table 1), whose procedure will subsequently make use of all 12 of these generalized Schmid's law expressions. Notice that the critical resolved shear stress reduces to the same result calculated previously for uniaxial tension along the [001] direction if only σ_{33} is applied.

Taylor Model

G.I. Taylor (Ref 7) noted that the isostress model used by Sachs to calculate the uniaxial yield stress of an isotropic fcc aggregate in terms of the critical resolved shear stress k failed to satisfy the strain compatibility requirements among grains for plastic deformation of an isotropic material deformed in uniaxial tension. These isostrain uniaxial tension requirements are that, for a given plastic strain $d\varepsilon$ along the length l of a single crystal, the normal strains along two axes perpendicular to l are equal to $-d\varepsilon/2$ (from constancy of volume requirements), and all shear strains are equal to zero. Taylor then proceeded to calculate the uniaxial yield stress of an isotropic fcc aggregate in terms of the critical resolved shear stress using his isostrain model, which is described next.

Taylor's approach sought to determine the crystallographic shear strains resulting from the operation of a particular set of five independent slip systems that will accommodate the uniaxial tension isostrain requirements imposed along the primed axes associated with a particular fcc crystal orientation. The first step is to find the plastic strain increments referred to the cubic axes of an fcc single crystal resulting from the imposition of the plastic strains imposed along primed specimen axes. If the imposed strain increments along the primed specimen axes are designated as $d\varepsilon_{k'l'}$ using the summation convention, the strain increments along cubic axes can be written in terms of these imposed strains as:

$$d\varepsilon_{ij} = l_{ik''}l_{jl'}d\varepsilon_{k'l'}$$

(Eq 3)

The results of this isostrain requirement for uniaxial tension along various orientations can then be used to relate the plastic strains along the cubic axes and the contributions of each of the twelve possible slip systems (four <111> planes each containing three {110} slip directions), that is:

$$d\varepsilon_{ij} = \sum_{n=1}^{4}\sum_{s=1}^{3} l_{in}l_{js}d\varepsilon_{ns}$$

(Eq 4a)

where $d\varepsilon_{ns}$ is 1 of 12 true plastic strains associated with the operation of particular slip system with slip plane normal $\hat{\mathbf{n}}$ and a slip direction $\hat{\mathbf{s}}$. The true shear strain $d\varepsilon_{ns}$ is equal to ½ the simple shear strain γ_{ns}, which contains a rotation equal to (½) $d\gamma_{ns}$. This relationship between crystallographic or simple shear, pure or tensorial shear, and rotation is shown in Fig. 3. Using the nomenclature of Bishop and Hill contained in Table 1, the relationships between the five independent components of strain referred to the cubic axes and the 12 crystallographic simple shear strains a_i (Eq 4a) become:

$$d\varepsilon_{11} = \frac{1}{\sqrt{6}}(-2a_2 + 2a_3 - 2a_5 + 2a_6 - 2a_8 + 2a_9 - 2a_{11} + 2a_{12})$$

$$d\varepsilon_{22} = \frac{1}{\sqrt{6}}(-2a_1 - 2a_3 + 2a_4 - 2a_6 + 2a_7 - 2a_9 + 2a_{10} - 2a_{12})$$

$$d\varepsilon_{33} = \frac{1}{\sqrt{6}}(+2a_1 + 2a_2 - 2a_4 - 2a_5 - 2a_7 - 2a_8 - 2a_{10} + 2a_{11})$$

$$d\varepsilon_{12} = \frac{1}{\sqrt{6}}(+a_1 - a_2 + a_4 - a_5 - a_7 + a_8 - a_{10} + a_{11})$$

$$d\varepsilon_{23} = \frac{1}{\sqrt{6}}(+a_2 - a_3 - a_5 + a_6 + a_8 - a_9 - a_{11} + a_{12})$$

$$d\varepsilon_{31} = \frac{1}{\sqrt{6}}(+a_1 + a_3 + a_4 - a_6 + a_7 - a_9 - a_{10} + a_{12})$$

(Eq 4b)

Since there are five independent components of strain ε_{ij} along the cubic axes (reduced from six by the constancy of volume requirement $d\varepsilon_{11} + d\varepsilon_{22} + d\varepsilon_{33} = 0$), at least five slip systems (a's) must operate for Eq 4(b) to be inverted to find the amounts of shear on particular slip systems that must operate to accommodate the strains imposed along the cubic axes. This requirement for the simultaneous operation of five independent slip systems to accommodate an arbitrary strain state in an fcc crystal was first pointed out by von Mises (Ref 8). Since one can find more than one set of five independent shear strains to accommodate a given imposed strain state, the solution for ε_{ij} in terms of a_i (Eq 4b) is not unique.

Taylor selected the particular set of five independent crystallographic simple shear strains a_i that accommodated the imposed macroscopic strain state $\varepsilon_{k''l'}$ with a minimum of the total shear strain (or shear strain energy):

$$\sum_{i=1}^{5} a_i$$

Therefore, to calculate the value of the uniaxial yield stress for an isotropic fcc crystalline aggregate using Taylor's isostrain minimum
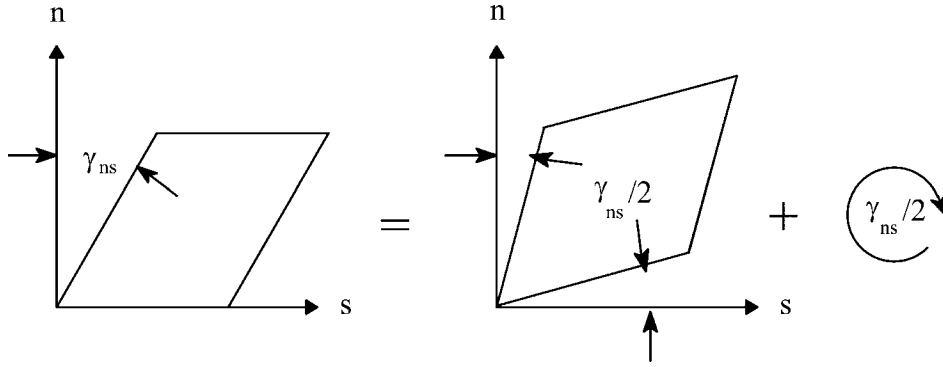
Fig. 3 Decomposition of a simple crystallographic shear into a pure or tensorial shear plus a rotation

Table 2 Bishop-Hill shear stress expressions

Slip system	Yield expression
$\pm(a_1)$	$A - G + H = \pm\sqrt{6}\,k$
$\pm(a_2)$	$B - H + F = \pm\sqrt{6}\,k$
$\pm(a_3)$	$C - F + G = \pm\sqrt{6}\,k$
$\pm(b_1)$	$A + G + H = \pm\sqrt{6}\,k$
$\pm(b_2)$	$B - H - F = \pm\sqrt{6}\,k$
$\pm(b_3)$	$C + F - G = \pm\sqrt{6}\,k$
$\pm(c_1)$	$A + G - H = \pm\sqrt{6}\,k$
$\pm(c_2)$	$B + H + F = \pm\sqrt{6}\,k$
$\pm(c_3)$	$C - F - G = \pm\sqrt{6}\,k$
$\pm(d_1)$	$A - G - H = \pm\sqrt{6}\,k$
$\pm(d_2)$	$B + H - F = \pm\sqrt{6}\,k$
$\pm(d_3)$	$C + F + G = \pm\sqrt{6}\,k$

shear criterion, one starts by imposing the plastic strain increment $d\varepsilon_{11} = d\varepsilon$ along the sample tensile axis and $-d\varepsilon/2$ along two axes perpendicular to this axis. This strain state referred to the sample axes is then transformed to the cubic axes of the fcc crystal using Eq 3. The particular set of five independent crystallographic shear strains a_i that accommodate the imposed strains is found next by looking for a set of five a_i's that satisfy Eq 4(b) and, at the same time, minimize the total crystallographic shear:

$$\sum_{i=1}^{5} a_i$$

Knowing the minimum total crystallographic shear for a particular crystal, one can find the value of the uniaxial yield stress in terms of the critical resolved shear stress k by equating the strain energy from slip to the external work done:

$$k\sum_{i=1}^{5} a_i = \sigma d\varepsilon$$

or

$$\sigma = k\sum_{i=1}^{5} a_i/d\varepsilon = Mk$$

where M, the ratio of the sum of the crystallographic shears to the uniaxial axisymmetric extension resulting from these shears, is the Taylor factor. More generally, M is defined as the sum of the crystallographic shears per unit effective strain.

Averaging the results for M over the stereographic triangle using this isostrain procedure, Taylor predicted that the uniaxial yield stress of an isotropic fcc aggregate is equal to 3.06k.

Taylor's minimum shear criterion was initially based on intuition but was later shown by Bishop and Hill (Ref 4, 5) to be rigorously correct and equivalent to their procedure described subsequently. Taylor also did not address the question as to whether or not one could find stress states that will simultaneously reach the critical resolved shear stress k on at least five independent slip systems while remaining below

k on the remaining systems, a question that again was addressed subsequently by Bishop and Hill (Ref 4, 5). Even if such stress states could be found, however, the issue of stress compatibility across grain boundaries remained. Nevertheless, Taylor's isostrain model remains a significant advance in crystal plasticity and is also currently used as the basis for several numerical codes capable of calculating Taylor factors for arbitrary imposed strain states rather than just axisymmetric deformation.

Bishop-Hill Procedure

Like that of Taylor (Ref 7), the procedure of Bishop and Hill (Ref 4, 5) and Bishop (Ref 6) is an isostrain model. However, their stress-based approach sought to directly find stress states that could simultaneously operate at least five independent slip systems. Bishop and Hill began by examining the generalized Schmid's law requirements for operation of the 12 {111} <110> fcc slip systems. They chose to describe this yielding in terms of the new stresses:

$$A = \sigma_{22} - \sigma_{33}, B = \sigma_{33} - \sigma_{11}, C = \sigma_{11} - \sigma_{22},$$
$$F = \sigma_{23}, G = \sigma_{31}, \text{ and } H = \sigma_{12}$$

(Eq 5)

which leads to the 12 (24 with negatives) critical resolved shear stress expressions given in Table 2.

These 12 yield expressions (24 with negatives) can be plotted in three separate three-dimensional stress spaces having coordinates A, G, and H; B, H, and F; and C, F, and G. The yield condition for a particular slip system is represented by a plane in one of the three three-dimensional stress spaces. These planes make equal angles to the three coordinate axes; their intersection forms three separate octahedral, as shown in Fig. 4. The intersection of these planes along lines and at corners represents polyslip stress states. One notes that there are two types of polyslip stress states: one that is situated at the vertex of an octahedron (points x_1 and x_2 in Fig. 4) and the other that lies on a line oriented at 45° to two coordinate axes (points y_1, y_2, y_3, and y_4). The former will

simultaneously operate four slip systems, the latter two. Hence, any polyslip stress state that simultaneously operates at least five slip systems will actually operate six or eight. Slip will be activated on eight slip systems if the stress state is located at two vertices (e.g., points x_1 and x_2) or at one vertex and two 45° lines (e.g., points x_1, y_1, y_2). Six systems will be activated if the stress state is located along three 45° lines (e.g., points y_1, y_3, and y_4). (One vertex and one 45° line would violate the condition that A + B + C = 0.) A summary of all permissible polyslip stress states is given in Table 2. The three cases considered previously correspond to stress states −2, −12, and −16. As can be shown with reference to the octahedral in Fig. 4, stress state −2 will operate the eight slip systems $-(a_2)$, (a_3), $-(b_2)$, (b_3), $-(c_2)$, (c_3), $-(d_2)$, and (d_3); stress state −12 will operate the eight systems $-(a_2)$, (a_3), $-(b_2)$, (b_3), $-(c_2)$, (c_3), $-(d_1)$, and (d_3); and stress state −16 will operate the six systems $-(b_2)$, (b_3), $-(c_1)$, (c_2), $-(d_1)$, and (d_3). A similar examination of the remaining polyslip stress states in Table 3 will show that the stress states 1 to 12 simultaneously operate eight slip systems, while stress states 13 to 28 operate six.

The particular state of stress that acts to accommodate an imposed state of strain can be found by selecting from the 28 (56 with negatives) permissible fcc polyslip stress states shown in Table 3, which shows the particular stress state that maximizes the external work done. This principle of maximum work was first derived rigorously by Bishop and Hill (Ref 4, 5), who also showed it to be equivalent to the minimum shear approach taken earlier by Taylor (Ref 7). The increment of work done, with reference to the cubic axes, is:

$$dW = \sigma_{11}d\varepsilon_{11} + \sigma_{22}d\varepsilon_{22} + \sigma_{33}d\varepsilon_{33} + 2\sigma_{12}d\varepsilon_{12} + 2\sigma_{23}d\varepsilon_{23} + 2\sigma_{31}d\varepsilon_{31}$$

(Eq 6a)

or, in Bishop-Hill notation:

$$dW = -Bd\varepsilon_{11} + Ad\varepsilon_{22} + 2Fd\varepsilon_{23} + 2Gd\varepsilon_{31} + 2Hd\varepsilon_{12}$$

(Eq 6b)

In the actual calculation procedure, the strain increments $d\varepsilon_{ij}$ along the cubic axes are first
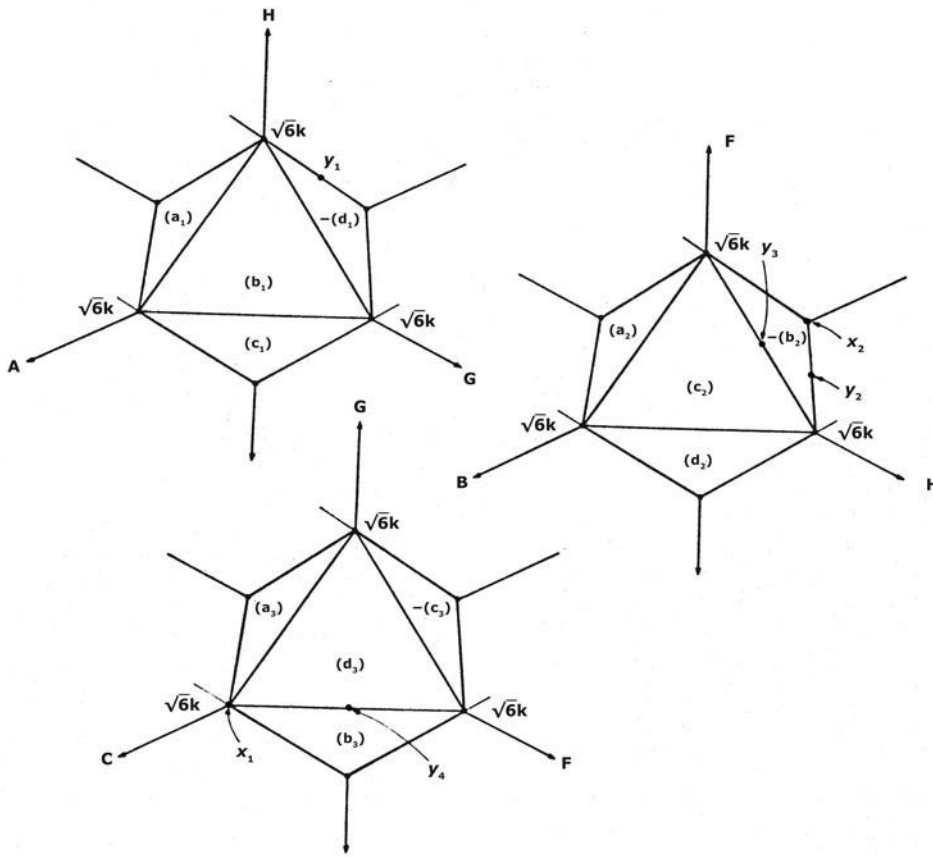
Fig. 4 Generalized Schmid's law plotted as octahedra using three separate coordinate systems

Table 3 Bishop-Hill stress states that simultaneously operate six or eight slip systems

Stress state number	Stress terms					
	$A/\sqrt{6}\,k$	$B/\sqrt{6}\,k$	$C/\sqrt{6}\,k$	$F/\sqrt{6}\,k$	$G/\sqrt{6}\,k$	$H/\sqrt{6}\,k$
1	1	−1	0	0	0	0
2	0	1	−1	0	0	0
3	−1	0	1	0	0	0
4	0	0	0	1	0	0
5	0	0	0	0	1	0
6	0	0	0	0	0	1
7	½	−1	½	0	½	0
8	½	−1	½	0	−½	0
9	−1	½	½	½	0	0
10	−1	½	½	−½	0	0
11	½	½	−1	0	0	½
12	½	½	−1	0	0	−½
13	½	0	−½	½	0	½
14	½	0	−½	−½	0	½
15	½	0	−½	½	0	−½
16	½	0	−½	−½	0	−½
17	0	−½	½	0	½	½
18	0	−½	½	0	−½	½
19	0	−½	½	0	½	−½
20	0	−½	½	0	−½	−½
21	−½	½	0	½	½	0
22	−½	½	0	−½	½	0
23	−½	½	0	½	−½	0
24	−½	½	0	−½	−½	0
25	0	0	0	½	½	−½
26	0	0	0	½	−½	½
27	0	0	0	−½	½	½
28	0	0	0	½	½	½

written in terms of the imposed macroscopic strain state $d\varepsilon_{k'l'}$ and substituted into Eq 6(b). The values of the 28 (56) permissible stress states contained in Table 2 are then substituted sequentially into Eq 6(b). That particular permissible stress state that maximizes dW for the given imposed strain state will operate the six or eight fcc slip systems that can accommodate this imposed strain.

Like Taylor before them, Bishop and Hill used an isostrain assumption to calculate the yield stress of an isotropic fcc crystalline aggregate in terms of the critical resolved shear stress k. Since, as shown by Bishop and Hill, their maximum work and Taylor's minimum shear criteria are equivalent, not surprisingly their value for the uniaxial yield stress of an isotropic fcc crystalline aggregate was also 3.06k.

Bounds for Yield Loci from Two-Dimensional Sachs and Bishop-Hill Averages

The isostress and isostrain approaches to calculating the uniaxial yield stress of an isotropic fcc aggregate in terms of the critical resolved shear stress k can be generalized to two dimensions to calculate yield loci for single crystals or ideal textural components (Ref 9). Individual single-crystal yield loci can then be averaged to calculate yield loci for crystalline aggregates. These rather simple two-dimensional bounds are often sufficient to answer engineering questions regarding texture hardening, for example. The basic approach is again illustrated for fcc metals deforming by rate-insensitive {111}<110> slip, but this general approach is applicable to other crystal structures with other slip behaviors as well.

Isostress or lower-bound yield loci for sheets having single-crystal orientation are found by:

- Applying a biaxial principal stress state along the rolling and transverse directions of a single-crystal or ideal texture component
- Calculating the resolved shear stresses on the 12 {111}<110> slip systems resulting from this biaxial principal stress state applied along the rolling and transverse directions
- Setting the 12 resolved shear stresses equal to k, the critical resolved shear stress
- Plotting the 12 lines representing yielding in the biaxial principal stress space
- Forming the lower-bound isostress yield locus from the intersection of these 12 lines

In general, if the rolling, transverse, and sheet normal direction are designated as r, t, and p, the six independent macroscopic stress components are σ_{rr}, σ_{tt}, σ_{pp}, σ_{rt}, σ_{tp}, and σ_{pr}. The stresses τ_{ns} acting on the {111}<110> slip systems can then be written in terms of these macroscopic stresses using the transformation:

$$\tau_{ns} = l_{nr}l_{sr}\sigma_{rr} + l_{nt}l_{st}\sigma_{tt} + l_{np}l_{sp}\sigma_{pp} + 2l_{nr}l_{st}\sigma_{rt}$$
$$+ 2l_{nt}l_{sp}\sigma_{tp} + 2l_{np}l_{sr}\sigma_{pr}$$

$$(\text{Eq } 7)$$

For example, for the ideal cube-on-corner texture, t is taken along $[1\bar{1}0]$, r is along $[11\bar{2}]$, and p is along $[111]$; the corresponding unit vectors are:

$$\hat{\mathbf{t}} = \frac{1}{\sqrt{2}}(\hat{\mathbf{i}}_1 - \hat{\mathbf{i}}_2), \hat{\mathbf{r}} = \frac{1}{\sqrt{6}}(\hat{\mathbf{i}}_1 + \hat{\mathbf{i}}_2 - 2\hat{\mathbf{i}}_3), \quad \text{and}$$

$$\hat{\mathbf{p}} = \frac{1}{\sqrt{3}}(\hat{\mathbf{i}}_1 + \hat{\mathbf{i}}_2 + \hat{\mathbf{i}}_3)$$

Calculating the direction cosines using the appropriate dot products and substitution into Eq 7 give the expressions in Table 4.

A lower bound to the yield locus for this ideal cube-on-corner crystal can be found directly from the equations in Table 4 by setting $\sigma_{pp} = \sigma_{rt} = \sigma_{tp} = \sigma_{tr} = 0$. The operation of each slip system will be governed by a line in $\sigma_{rr}-\sigma_{tt}$ space; the intersections of these lines form the lower bound to the yield locus for the ideal cube-on-corner crystal (Fig. 5).

A similar procedure is used to relate the strain increments along the rolling, transverse, and sheet normal directions of the ideal cube-on-corner orientation and the summed contributions from the 12 crystallographic shears, yielding:

$$d\varepsilon_{rr} = \frac{1}{3\sqrt{6}}(-2b_1 + 2b_2 - 3c_1 + c_2 + 2c_3 - d_1 + 3d_2 - 2d_3)$$

$$d\varepsilon_{tt} = \frac{1}{\sqrt{6}}(c_1 - c_2 + d_1 - d_2)$$

$$d\varepsilon_{pp} = \frac{2}{3\sqrt{6}}(b_1 - b_2 + c_2 - c_3 - d_1 + d_3)$$

$$d\varepsilon_{rt} = \frac{1}{3\sqrt{2}}(-b_1 - b_2 + 2b_3 - c_1 + c_3 - d_2 + d_3)$$

$$d\varepsilon_{tp} = 1/12(-3a_1 - 3a_2 + 6a_3 - b_1 - b_2 + 2b_3 - c_1 - 3c_2 + 4c_3 - 3d_1 - d_2 + 4d_3)$$

$$d\varepsilon_{pr} = \frac{1}{12\sqrt{3}}(9a_1 - 9a_2 + 7b_1 - 7b_2 + 3c_1 - 5c_2 + 2c_3 + 5d_1 - 3d_2 - 2d_3)$$

$$(\text{Eq } 8)$$

These strain-slip equations are used subsequently.

An upper bound to the yield locus for this same crystal is be found by:

- Imposing different ratios of the principal strains ε_{rr} and ε_{tt}
- Transforming these principal strains to the cubic axes
- Using the maximum work principle to select the Bishop-Hill stress state that operates the slip systems necessary to accommodate this strain imposed along the cubic axes
- Transforming these Bishop-Hill stresses to the rolling (r), transverse (t), and sheet normal

Table 4 Stress transformation in Bishop-Hill notation

Slip system	Shear stress expression
$\pm(a_1)$	$-\frac{1}{2}\sigma_{tp} + \frac{1}{2}\sqrt{3}\sigma_{pr} = \pm\sqrt{6}\,k$
$\pm(a_2)$	$-\frac{1}{2}\sigma_{tp} - \frac{1}{2}\sqrt{3}\sigma_{pr} = \pm\sqrt{6}\,k$
$\pm(a_3)$	$\sigma_{tp} = \pm\sqrt{6}\,k$
$\pm(b_1)$	$-\frac{2}{3\sqrt{6}}\sigma_{rr} + \frac{2}{3\sqrt{6}}\sigma_{pp} - \frac{2}{3\sqrt{2}}\sigma_{rt} - \frac{1}{6}\sigma_{tp} + \frac{7}{6\sqrt{3}}\sigma_{pr} = \pm\sqrt{6}\,k$
$\pm(b_2)$	$\frac{2}{3\sqrt{6}}\sigma_{rr} - \frac{2}{3\sqrt{6}}\sigma_{pp} - \frac{2}{3\sqrt{2}}\sigma_{rt} - \frac{1}{6}\sigma_{tp} - \frac{7}{6\sqrt{3}}\sigma_{pr} = \pm\sqrt{6}\,k$
$\pm(b_3)$	$\frac{4}{3\sqrt{2}}\sigma_{rt} + \frac{1}{3}\sigma_{tp} = \pm\sqrt{6}\,k$
$\pm(c_1)$	$-\frac{1}{\sqrt{6}}\sigma_{rr} + \frac{1}{\sqrt{6}}\sigma_{tt} - \frac{2}{3\sqrt{2}}\sigma_{rt} - \frac{1}{6}\sigma_{tp} + \frac{1}{2\sqrt{3}}\sigma_{pr} = \pm\sqrt{6}\,k$
$\pm(c_2)$	$\frac{1}{3\sqrt{6}}\sigma_{rr} - \frac{1}{\sqrt{6}}\sigma_{tt} + \frac{2}{3\sqrt{6}}\sigma_{pp} - \frac{1}{2}\sigma_{tp} - \frac{5}{6\sqrt{3}}\sigma_{pr} = \pm\sqrt{6}\,k$
$\pm(c_3)$	$\frac{2}{3\sqrt{6}}\sigma_{rr} - \frac{2}{3\sqrt{6}}\sigma_{tt} + \frac{2}{3\sqrt{2}}\sigma_{rt} + \frac{2}{3}\sigma_{tp} + \frac{1}{3\sqrt{3}}\sigma_{pr} = \pm\sqrt{6}\,k$
$\pm(d_1)$	$-\frac{1}{3\sqrt{6}}\sigma_{rr} + \frac{1}{\sqrt{6}}\sigma_{tt} - \frac{2}{3\sqrt{6}}\sigma_{pp} - \frac{1}{2}\sigma_{tp} + \frac{5}{6\sqrt{3}}\sigma_{pr} = \pm\sqrt{6}\,k$
$\pm(d_2)$	$\frac{1}{\sqrt{6}}\sigma_{rr} - \frac{1}{\sqrt{6}}\sigma_{tt} - \frac{2}{3\sqrt{2}}\sigma_{rt} - \frac{1}{6}\sigma_{tp} - \frac{1}{2\sqrt{3}}\sigma_{pr} = \pm\sqrt{6}\,k$
$\pm(d_3)$	$-\frac{2}{3\sqrt{6}}\sigma_{rr} + \frac{2}{3\sqrt{6}}\sigma_{pp} + \frac{2}{3\sqrt{2}}\sigma_{rt} + \frac{2}{3}\sigma_{tp} - \frac{1}{3\sqrt{3}}\sigma_{pr} = \pm\sqrt{6}\,k$

(p) coordinate system and supplementing these stresses with the condition that $\sigma_{pp} = 0$
- Plotting the results in $\sigma_{rr}-s_{tt}$ space

Using the maximum work principle, the resulting Bishop-Hill stress states for this crystal to accommodate biaxial principal strains in the ideal cube-on-corner crystal are No. 6, 24, 25, and 28, the specific stress state depending on the imposed strain ratio. After performing the transformation outlined previously, there results the upper-bound yield locus shown in Fig. 6.

Figures 5 and 6, together with the normality condition (Ref 4, 5), are used to determine bounds for the biaxial stresses needed to enforce a specific strain state (in this case plane strain) as well as to illustrate the physical differences between these two bounds for the plane-strain yield stress.

Applying the normality condition as shown in Fig. 5, plane strain in the transverse direction ($d\varepsilon$ horizontal), one sees that the biaxial stress state needed to initiate this deformation is:

$$\sigma_{rr} = \sigma_{tt} = \frac{3}{2}\sqrt{6}k$$

This stress state simultaneously operates two branches of the yield locus; the first (horizontal) branch of the yield locus initiates yielding on systems $-(b_1)$, $+(b_2)$, $+(c_3)$, and $-(d_3)$; the second on systems $-(c_2)$ and $+(d_1)$.

Before proceeding further, it should first be noted that the strains $d\varepsilon_{rr}$ and $d\varepsilon_{tt}$ that result from the operation of each particular system are normal to their respective branch of the yield locus. Focusing first on the horizontal branch and letting slip occur by a unit amount in any one of the systems $-(b_1)$, $+(b_2)$, $+(c_3)$, and $-(d_3)$, the strain state from Eq 8 for each of these four systems is:

$$d\varepsilon_{rr} = -d\varepsilon_{pp} = \frac{2}{3\sqrt{6}}d\varepsilon \text{ and } d\varepsilon_{tt} = 0$$

which is normal to the horizontal branch. Similarly, if slip occurs by a unit amount on either system $-(c_2)$ or (d_1), the resulting strain is:

$$d\varepsilon_{rr} = -\frac{\sqrt{6}}{3}, d\varepsilon_{tt} = \sqrt{6}, \quad \text{and} \quad d\varepsilon_{pp} = -\frac{2\sqrt{6}}{3}$$

which again is perpendicular to its branch of the yield locus. This two-dimensional normality holds true even if the biaxial principal stresses activate systems that result in shear strains as well as normal strains; that is, the principal directions of stress and strain do not coincide.

Plane strain in the transverse direction of the cube-on-corner crystal is achieved by combining the strain vectors normal to the two branches of the yield locus to achieve the strain state $d\varepsilon_{rr} = 0$, $d\varepsilon_{tt} = 1$, and $d\varepsilon_{pp} = -1$. This is achieved by adjusting the amounts of shear on the operative systems to be:

$$b_1 = d_3 = -\frac{\sqrt{6}}{8}, b_2 = c_3 = \frac{\sqrt{6}}{8}, c_2 = -\frac{\sqrt{6}}{2}, \text{ and}$$

$$d_1 = \frac{\sqrt{6}}{2}$$

Substituting these values into Eq 8 results in plane strain in the transverse direction, since $d\varepsilon_{rr} = 0$, $d\varepsilon_{tt} = 1$, and $d\varepsilon_{pp} = -1$. However, only two of the shear strains, $d\varepsilon_{rt}$ and $d\varepsilon_{tp}$, vanish; the remaining shear strain ($d\varepsilon_{pr}$) is equal to:

$$\frac{15\sqrt{2}}{8}$$

In order to impose a state of transverse plane strain in the absence of shear strain on this cube-on-corner crystal, it is necessary both to alter the levels of σ_{rr} and σ_{tt} and to apply an additional shear stress σ_{pr}. After transforming Bishop-Hill stress state No. 6:

$$\sigma_{12} = \sqrt{6}k$$

to the r, t, and p coordinate system and imposing the condition that $\sigma_{pp} = 0$, one finds that:

$$\sigma_{rr} = \frac{\sqrt{6}}{3}k, \sigma_{tt} = \frac{5\sqrt{6}}{3}k, \sigma_{pr} = -\frac{2\sqrt{3}}{3}k, \text{ and}$$

$$\sigma_{pp} = \sigma_{rt} = \sigma_{tp} = 0$$

By substituting these stresses into Eq 8, one can verify that the critical resolved shear stress is
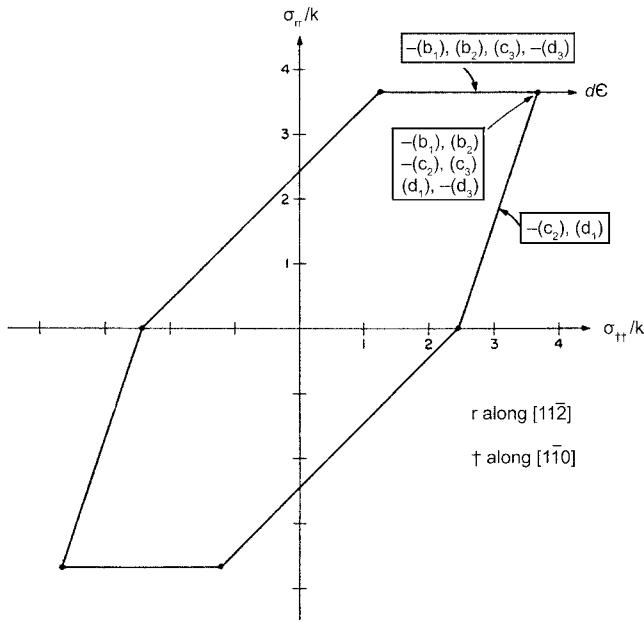
Fig. 5 Lower-bound yield locus for ideal cube-on-corner texture. Sheet normal is [111].
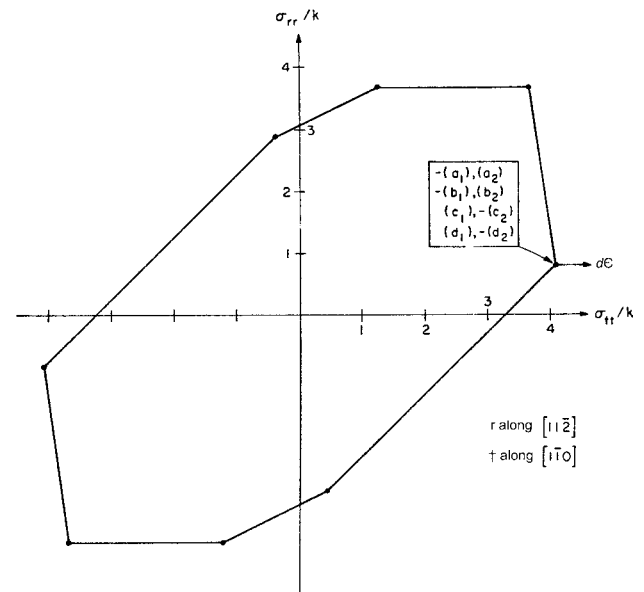
Fig. 6 Upper-bound yield locus for ideal cube-on-corner texture. Sheet normal is [111].

reached simultaneously on systems $-(a_1)$, (a_2), $-(b_1)$, (b_2), (c_1), $-(c_2)$, (d_1), and $-(d_2)$. Plane strain in the absence of shear strain can be achieved if the amounts of shear on these operative systems are:

$$a_1 = \frac{\sqrt{6}}{6}, a_2 = -\frac{\sqrt{6}}{6}, b_1 = -\frac{\sqrt{6}}{2}, b_2 = \frac{\sqrt{6}}{2}, c_1 = d_1$$
$$= \frac{\sqrt{6}}{4}, \text{ and } c_2 = d_2 = -\frac{\sqrt{6}}{4}$$

Thus, for the ideal cube-on-corner orientation, transverse plane strain in the absence of shear stress is accompanied by the additional shear strain $d\varepsilon_{pr}$; transverse plane strain in the absence of shear strain, on the other hand, requires the additional shear stress σ_{pr}. Hence, in neither case do the principal directions of stress and strain coincide. This lack of coincidence of the principal directions is not restricted to the cube-on-corner orientation. It should also be pointed out that the aforementioned solution for the amounts of shear needed to accommodate transverse plane strain in this ideal cube-on-corner orientation is not unique. Hence, the accompanying rotations leading to textural changes are not unique as well.

There are various ways in which the slip behaviors of multiple-crystal orientations can be combined to find yield loci for crystalline aggregates. One such average is the original yield locus for an isotropic fcc polycrystalline aggregate calculated by Bishop and Hill (Ref 5). Their averaging process rounded the edges that resulted from the intersections of the lines governing the operation of specific slip systems and resulted in a yield locus that is midway between the Tresca and the von Mises locus.

Since the Bishop-Hill procedure is based on an isostrain model (violating stress continuity among grains), this yield locus is an upper bound to the actual locus.

Recent Developments

The lack of a unique prediction of operative slip systems resulting from the rate-insensitive Bishop-Hill procedure was remedied by using a rate-sensitive model, first introduced by Hutchinson (Ref 10). An example of a computer code that incorporates rate sensitivity is the widely used Los Alamos polycrystal plasticity (LApp) code (Ref 11, 12). The LApp is a modified Taylor isostrain model that incorporates not only rate sensitivity but also grain shape and relaxed constraints (Ref 12).

Several slip behaviors other than $<111>$ $\{110\}$ slip in fcc metals have been studied in some detail. It should again be pointed out that $<111>\{110\}$ slip in fcc metals is equivalent to $<110>\{111\}$ slip in body-centered cubic (bcc) metals (often referred to as restricted glide), since this only involves switching the identities of the angles λ and ϕ and hence leaves Schmid's law unchanged. Another approach to modeling bcc slip behavior is pencil glide, where slip is assumed to occur with equal ease on the most highly stressed planes containing the $<111>$ slip directions (Ref 13). This slip plane relaxation has resulted in a change of the Taylor factor for an isotropic crystalline aggregate from 3.06 for bcc restricted glide to 2.73 for bcc pencil glide (Ref 14). The plastic deformation of hexagonal close-packed (hcp) crystals has also been studied (Ref 15, 16).

Modeling the plastic behavior of hcp metals is inherently more complicated, because deformation typically occurs by twinning in addition to slip (Ref 12).

While most current computer modeling efforts are isostrain based (Taylor or Bishop and Hill), many problems, especially those associated with surface behavior, are better described using simpler crystallographic approaches. A case in point is the characterization of surface roughening during plane-strain plastic deformation of aluminum alloy sheet (Ref 17). Orientation imaging and scanning electron microscopy of individual surface grains were used to characterize surface roughening resulting from the presence of both slip-banded valley-forming grains and nonslip-banded hill-forming grains. What was found was that slip-banded valley-forming grains correlated with Schmid factor-based unconstrained measures rather than constrained measures based on the Taylor factor. This is but one example of the caution needed to select the plasticity model appropriate to the problem being addressed.

REFERENCES

1. E. Schmid, *Proc. Intern. Congr. Appl. Mech., Delft, 1924*, p 342–352, J. Waitman, Jr., Delft, 1925
2. G. Sachs, Zur Ableitung einer Fliessbedingung, *Z. Ver. deut.Eng.*, Vol 72, p 734–736, 1928
3. W.A. Backofen, W.F. Hosford, Jr., and J.J. Burke, Texture Hardening, *Trans. ASM*, Vol 55, p 264, 1962

4. J.F.W. Bishop and R. Hill, A Theory of the Plastic Distortion of a Polycrystalline Aggregate under Combined Stresses, *Phil. Mag.*, Vol 42, p 414, 1951

5. J.F.W. Bishop and R. Hill, A Theoretical Derivation of the Plastic Properties of a Polycrystalline Face-Centered Metal, *Phil. Mag.*, Vol 42, p 1298, 1951

6. J.F.W. Bishop, A Theoretical Examination of the Plastic Deformation of Crystals by Glide, *Phil. Mag.*, Vol 44, p 51, 1953

7. R. von Mises, Mechanik der plastischen Formanderung der Kristallen, *Z. Angew. Math. Mech.*, Vol 8, p 161, 1928

8. G.I. Taylor, Plastic Strain in Metals, *J. Inst. Metals*, Vol 62, p 307, 1938

9. H.R. Piehler and W.A. Backofen, The Prediction of Anisotropic Yield Loci for Textured Sheets, *Textures in Research and Practice*, G. Wassermann and G. Grewen, Ed., Springer Verlag, Berlin, 1969, p 436–443

10. J.W. Hutchinson, Bounds and Self-Consistent Estimates for Creep of Polycrystalline Materials, *Proc. Roy. Soc. London A,* Vol 348, No. 1652, (Feb. 10, 1976), p 101–127

11. U.F. Kocks, G.R.Canova, C.N. Tome, A.D. Rollett, and S.I. Wright, *Computer Code LA-CC-88-6* (Los Alamos, NM: Los Alamos National Laboratory)

12. U.F. Kocks, C.N. Tome, and H.-R. Wenk, *Texture and Anisotropy*, Cambridge University Press, 1998

13. H.R. Piehler and W.A. Backofen, A Theoretical Examination of the Plastic Properties of Body-Centered Cubic Crystals Deforming by <111> Pencil Glide, *Met. Trans.,* Vol 2, p 249–255, 1971

14. J.M. Rosenberg and H.R. Piehler, Calculation of the Taylor Factor and Lattice Rotations for Bcc Metals Deforming by Pencil Glide, *Met. Trans.*, Vol 2, p 257–259, 1971

15. D.R. Thornburg and H.R. Piehler, An Analysis of Constrained Deformation by Slip and Twinning in Hexagonal Close-Packed Metals and Alloys, *Met. Trans. A*, Vol 6A, p 1511–1523, 1975

16. C. Tome and U.F. Kocks, The Yield Surface of HCP Single Crystals, *Acta Metall*, Vol 33, p 603–621, 1985

17. Y.S. Choi, H.R. Piehler, and A.D. Rollett, Formation of Mesoscale Roughening in 6022-T4 Al Sheets Deformed in Plane-Strain Tension*, Met. and Mat. Trans. A*, Vol 35A, p 513–524, 2004

Self-Consistent Modeling of Texture Evolution

David Dye, Imperial College, London, United Kingdom

Introduction

When considering the macroscopic properties of a polycrystalline metal or alloy, the texture (Ref 1), that is, the crystal orientation distribution, is the primary contributor to elastic and plastic anisotropy arising from the anisotropy of single crystals. Of course, the grain morphology and microstructural arrangement of phases will also often be important. Therefore, it is often desirable to be able to measure, represent, and model textures and their evolution due to materials processing operations such as metal forming, heat treating, and casting.

Self-consistent models are a particular class of models in continuum micromechanics, that is, the field concerned with making predictions of the properties and evolution of aggregates whose single-crystal deformation behavior is known. The simplest estimates of elastic properties are those that assume equal strain in all the grains in the aggregate (Voigt, 1887) or equal stress (Reuss, 1929). Sachs (1928) and Taylor (1938) extended the approach from elasticity to plasticity, with Sachs applying the first slip system to reach its critical resolved shear stress and Taylor requiring all grains to undergo the same, macroscopic, strain.

Eshelby's solution to the elastic problem around an ellipsoidal inclusion via a Green's function analysis (Ref 2) then allowed a revolution, with the invention of the self-consistent method elaborated by Kröner (Ref 3) for elasticity and then for plasticity by Hill (Ref 4) and Hutchinson (Ref 5).

The foundation-stone of the approach is the solution of the strain in an ellipsoidal single crystal c embedded in a homogenous effective medium (Fig. 1). The strain rate in the ellipsoid $\dot{\varepsilon}_c$ is constant and given by:

$$\dot{\varepsilon}_c = A_c \dot{\bar{\varepsilon}} = (L^* + L_c)^{-1}(L^* + \overline{L})\dot{\bar{\varepsilon}} \qquad \text{(Eq 1)}$$

where the medium strain rate is $\dot{\bar{\varepsilon}}$, A_c is a concentration tensor, and $\overline{L}, L_c$, and L^* are the stiffness of the medium, crystal, and the overall constraint, respectively. L^* depends solely on

the medium and grain shape and can be obtained by integration (Ref 4–6):

$$L^* = \Lambda^{-1} - \overline{L} \qquad \text{(Eq 2)}$$

$$\Lambda_{ijnm} = \frac{1}{4\pi} \int_{\theta=0}^{\pi} \int_{\phi=0}^{1\pi} \left(\hat{U}_{im} k_n k_j + \hat{U}_{jm} k_n k_i \right.$$
$$\left. + \hat{U}_{in} k_m k_j + \hat{U}_{jn} k_m k_i \right) \sin\theta \, d\theta \, d\phi \qquad \text{(Eq 3)}$$

where $\hat{U}$ is given by $\overline{L}_{ijkl}\hat{U}_{km}k_jk_l = \delta_{im}$, $k_1 = \sin\theta \cos\phi/a_1$, $k_2 = \sin\theta \sin\phi/a_2$, and $k_3 = \cos\Phi/a_3$. θ and Φ are spherical co-ordinates for the ellipsoid, while the a_i are the semi-lengths of the ellipsoid axes.

A self-consistent model is then constructed by finding the strain in every grain in the aggregate and then assigning the average to the medium to obtain an estimate for the medium stiffness $\overline{L}$:

$$\overline{L} = <L_c A_c> \qquad \text{(Eq 4)}$$

The solution is then iterated until a stable estimate of $\overline{L}$ is found. Therefore, both stress and strain compatibility are satisfied in the average sense.

Plasticity within each grain is modeled using a viscoplastic flow rule to obtain the instantaneous (tangent) stiffness of the grain, yielding a viscoplastic self-consistent (VPSC) model. For example (Ref 7), a nonlinear viscous response can be assumed:

$$\dot{\gamma}^s = \dot{\gamma}_0 \left(\frac{m_k^s \sigma_k'}{\tau_c^s} \right)^n \qquad \text{(Eq 5)}$$

where $\dot{\gamma}^s$ is the shear rate on slip system s, $m_k^s \sigma_k' = \tau^s$ is the resolved shear stress on the slip system from the applied deviatoric stress σ' and Schmid tensor m, τ_c is a threshold stress, $\dot{\gamma}_0$ is a reference shear strain rate, and n is the rate law exponent. The Schmid tensor is calculated from the (unit) slip plane normal vector $\mathbf{n}$ and the (unit) vector in the slip direction or Burgers vector $\mathbf{b}$ from $m_{ij}^s = \frac{1}{2}(\mathbf{n}_i\mathbf{b}_j + \mathbf{n}_j\mathbf{b}_i)$ and is by definition symmetric and traceless.

The overall strain rate in the grain is then given by:

$$\dot{\varepsilon}_i = \left\{ \dot{\gamma}_0 \sum_{s=1}^{S} \frac{m_i^s m_j^s}{\tau_c^s} \left(\frac{m_k^s \sigma_k'}{\tau_c^s} \right)^{n-1} \right\} \sigma_j' \qquad \text{(Eq 6)}$$

$$= L_{ij}^{c(\text{sec})} \sigma_j' \qquad \text{(Eq 7)}$$

where the sum is over all the slip systems s, and $L^{c(\text{sec})}$ is the secant viscoplastic compliance modulus of the grain.

The instantaneous or tangent modulus in Eq 1 and 2 is then found to be (Ref 7):

$$L_c = L_{ij}^{\text{tg}} = n L_{ij}^{\text{sec}} \qquad \text{(Eq 8)}$$

by performing a Taylor expansion in the vicinity of the current point or the secant modulus. The grains will rotate during the deformation process, giving rise to texture change. This rotation $\dot{\omega}_{ij}$ has three components: that due to the macroscopic distortion $\dot{\Omega}_{ij}$, which is the antisymmetric component of the applied distortion tensor; the rotation of the Eshelby ellipsoid from the Eshelby rotation tensor Π; and the rotation of the crystal lattice due to slip:

$$\dot{\omega}_{ij} = \dot{\Omega}_{ij} + \Pi_{ijkl} S_{klmn}^{-1} \bar{\dot{\varepsilon}}_{mn} - \sum_s \frac{1}{2}(b_i n_j - b_j n_i)^s \dot{\gamma}^s \qquad \text{(Eq 9)}$$

where S is the Eshelby tensor (Ref 2). The attraction of this type of modeling is that
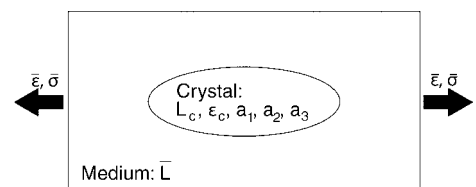
Fig. 1 Schematic of the Eshelby inclusion, highlighting the symbols used. Many grains are inserted as the inclusion, and the average is used to estimate the stiffness of the surrounding medium.

polycrystal flow can be modeled using only three fitting parameters, $\dot{\gamma}^s$, τ^s and n, per slip mode. Of course, other single-crystal plasticity models can be treated within the approach. In addition, these parameters can be constrained to conform to relationships for the temperature and strain-rate dependence, where these are known.

One example is the elastic-plastic self-consistent modeling approach, where grain rotation is neglected, but conventional elastoplasticity is treated as the flow rule (Ref 8). It is common practice to use such a model when studying the operative deformation mechanisms during the elastic-plastic transition and the intergranular strains that arise, measured by neutron or synchrotron x-ray diffraction (Ref 9–11).

Measuring and Representing Textures

Textures, that is, crystal orientation distributions, are traditionally measured using laboratory x-ray diffraction. Here, only a summary overview is given. The variation in intensity of an elastic scattering peak {hkl} or {$hkil$} is measured as a function of sample orientation, (ω, ϕ). Because a grain orientation is defined by three angles (Φ, ψ_1, ψ_2), three or more peaks must be measured in order to collect sufficient information about the orientation distribution function (ODF) (Ref 12, 13). The ODF is then reconstructed from the measured data by a function-fitting method, for example, fitting a spherical harmonic description or via the WIMV approach (Ref 14, 15).

When lab x-ray data are used, the data must be corrected for variation in the scattering geometry with incident angle in the reflection mode, which ultimately leads to adsorbtion at high incident angles; therefore, measurements are not usually made at incident angles greater than 80°. This reduces the amount of data associated with each reflection, leading to the usual practice of using the four most intense, nonredundant peaks from each phase. When using the spherical harmorics approach, the production of "ghosts" associated with underdetermination of the problem at high angles is often problematic, and therefore, the WIMV approach is usually preferred. Often data will be symetrized during the ODF fitting process in line with the known sample and crystal symmetry and partial, mutually consistent pole figures then plotted, for example, quadrants for fourfold sample symmetry.

Neutron diffraction may also be used for texture measurement. Here, sampling volumes are typically in the range of 1 to 20 mm^3. Lab x-ray measurements, in comparison, only measure those grains in the surface few tens of micrometers illuminated by the diffraction spot, which may be ~5 mm in diameter. Therefore, particularly for large-grained materials, neutron diffraction measurements are preferred because many more grains will typically be measured.

On a conventional constant-wavelength diffractometer, the principle is still the same, with the sample being rotated through every symmetric angle. The ODF is then reconstructed as for lab x-ray measurements but without the problem of ghosts.

A more direct method is electron backscatter diffraction (EBSD). Here, the Kikuchi lines obtained from backscattered electrons in the scanning electron microscope (SEM) are used to analyze the orientation of a grain under the electron beam. If scans are performed over a large enough area relative to the grain size to obtain a reasonable statistical sampling— maybe 10^3 grains—then the macroscopic texture may be measured directly. Of course, this is only a side effect of measurement of the local misorientations, and a good field emission SEM and sample preparation are required, but with modern EBSD equipment and computer analysis being capable of approximately 10^2 EBSD pattern measurements per second, obtaining complete textures in a few hours in this fashion is now possible, compared to ~4 h for lab x-ray measurements or ~12 h for neutron measurements at a reactor source.

One newer method is the use of a time-of-flight neutron diffractometer with large detector coverage (Ref 16). Here, diffraction data are obtained over a broad d-spacing range across a number of orientations simultaneously, thereby obtaining many incomplete pole figures simultaneously. Rotation of the sample is then used to partially fill these pole figures. The large number of diffraction peaks collected mean that complete ODFs can be collected with only six sample rotations, in approximately 30 to 60 min. The ODFs are obtained by simultaneous fitting of the diffraction data from each detector for each sample setting in a Reitveld analysis program such as MAUD (Ref 17) or GSAS (Ref 18), usually using the spherical harmonic approach. This method is only in widespread use at the HIPPO and GEM diffractometers at Los Alamos and ISIS, United Kingdom, but could be more widely applied, for example, at the SNS.

Finally, at synchrotron x-ray facilities, the use of area detectors for high-energy diffraction from engineering samples has become increasingly popular, where complete diffraction rings can be collected in transmission. If enough diffraction rings are collected, it again becomes possible to reconstruct the entire orientation distribution by Reitveld refinement (Ref 19), here without rotating the sample. Because satisfactory patterns can be collected in times as short as 100 ms, this allows the real-time evolution of textures to be characterized, for example, during deformation processing and heat treatment.

When presenting pole figures, customarily equal-area projections are used such that, visually, an average multiple of random deviation of 1 is obtained. Similarly for ODFs, spherical equal-area representations are generally preferred (Ref 14). For analyzing textures, the

general aim is to use a software that can treat data from all five of the measurements described previously for random sampling grids, and that will output orientation distributions both graphically and numerically. Three sets of software are in widespread use: popLA (Ref 15), which takes the form of a suite of routines originally written at Los Alamos; BEARTEX (Ref 12); and the simultaneous Reitveld and texture code MAUD (Ref 17). The GSAS Reitveld refinement code is also often used for simultaneous fitting of transmission diffraction spectra by spherical harmonics (Ref 18), although popLA or some other software must then be used for visualization.

An ODF can be inputted into a self-consistent texture evolution problem as a starting parameter in two ways: either a random or gridded set of orientations can be tracked and each assigned a weight corresponding to the frequency in the ODF, or alternatively, the starting set of orientations can be sampled from the measured ODF Generally, the prior approach is taken because it ensures that all possible orientations are treated in the simulation, even if they do not form a significant component of the starting texture. The final set of rotated grain orientations and weights must then be manipulated back into an orientation distribution function, for example, using popLA.

Predictions of Texture Evolution

Single-Phase Materials Deforming by Slip. Compared to Taylor-type deformation models, VPSC models are far more successful in predicting texture evolution (Ref 7). In particular, for realistic assumptions about the operative deformation mechanisms, even in single-phase face-centered cubic (fcc) materials, where the deformation anisotropy is relatively low, Taylor models tend to overpredict the sharpness of the textures that result, whereas self-consistent models provide softer interactions and hence more realistic textures (Ref 20). However, it should be noted that the textures predicted remain slightly sharper than those observed experimentally.

For simulations of recrystallization, the stored energy must be known and will vary according to the amount of deformation in grains in different orientations; hence, polycrystal models of deformation can be very useful. In bcc iron, VPSC modeling has been used to determine the amounts of strain, again yielding qualitatively different results to a Taylor full-constraints approach; a VPSC model predicts that the γ-fiber ({111} rolling plane) is softest, whereas a Taylor model predicts that the α-fiber (< 110 > rolling direction) is softest at the end of rolling (Ref 21).

Deformation textures observed due to torsion deformation at both room temperature and at 264 °C in fcc nickel and nickel-cobalt alloys of varying stacking fault energy (SFE) have also been modeled using the VPSC approach

(Ref 22) (Fig. 2). Here again, VPSC models performed much more satisfactorily than Taylor models, particularly with updating of the grain shapes during deformation. However, it was found that, regardless of whether twinning was incorporated in the model, the subdivision of a grain into cell blocks, controlled by slip planarity, resulted in the activation of additional slip systems, and therefore, some of the details of the textures of low-SFE materials could not be reproduced.

The VPSC models have also been used to predict the yield surface (R-ratio, etc.) in rolled Al5019 sheet, in both the H48 and O temper conditions (Ref 23), again showing much better agreement than a full-constraints (Taylor) model. Here, it was found that grain shape was not particularly critical to the results obtained, but that when the material was annealed, the prediction was less accurate, presumably due to some details around the recrystallization process.

The approach has even been used to determine the deformation mechanisms operative in iron at pressures representative of the Earth's core from the textures produced in high-pressure experiments using diamond anvil cells at an x-ray synchrotron (Ref 24).

Generally, it appears that there is a consensus that VPSC models predict deformation textures well where there is an absence of change in the micromechanisms that operate during deformation. This point has been made rather elegantly by Engler (Ref 25), who examined copper-manganese alloys, which do not change in SFE with manganese addition and therefore where partial dislocations and twinning are not involved, and still found that the textures change in character, due to, for example, short-range ordering leading to planar slip and shear band formation as yield strength increases. However, some aspects of this change, such as an alteration in work-hardening rate, can be included in a VPSC model. Where these effects are absent, for example, in AA1050 with a high SFE and low solute content, then the textures produced by both shear and tension of sheet material can be reproduced satisfactorily (Ref 26).

The VPSC approach has also been applied to the earing problem in cup-drawing of AA3104, successfully predicting the ear profiles as a function of initial texture after cold rolling subsequent to initial hot rolling (Ref 27) (Fig. 3). Finite-element (FE) models using simplified anisotropic behavior were used to derive the stress and strain boundary conditions.

There have been some attempts to include the effect of grain-to-grain coupling within the VPSC approach. Here, a second ellipsoid is placed in the effective medium (Ref 28), allowing interaction between crystals placed on each site (Ref 29). This, then, allows grain interactions to be modeled in a so-called grain co-rotation scheme in Al7554 (Ref 30), which particularly affect the evolution of low-angle misorientations and, overall, act to soften the textures produced by compression testing.
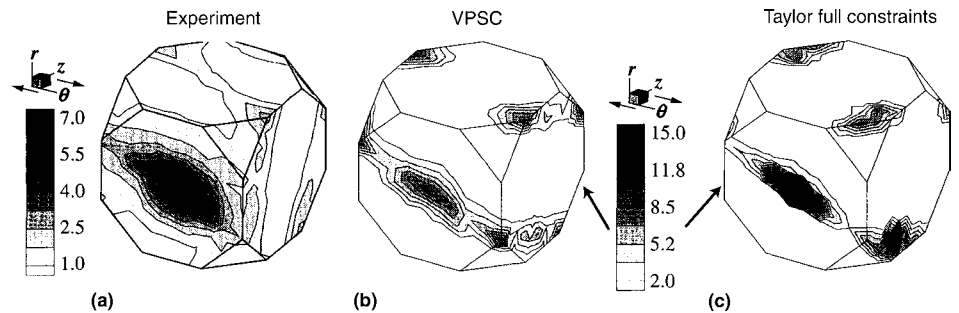
Fig. 2 (a) Texture (plotted in Rodrigues' space) produced in initially untextured high-purity nickel following room-temperature torsion to a strain of 4.0, compared to (b) Viscoplastic self-consistent (VPSC) and (c) Taylor full-constraints models. Source: Ref 22
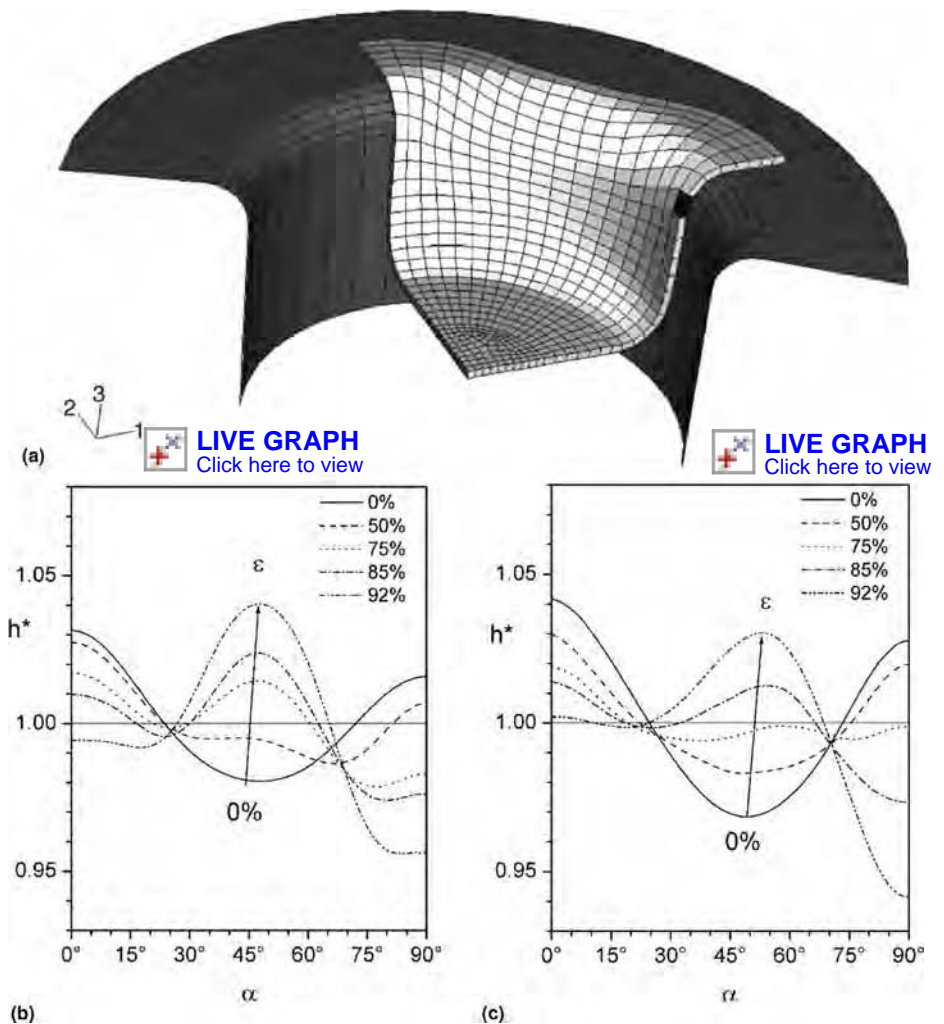
Fig. 3 Simulation of earing using a viscoplastic self-consistent model. (a) Finite-element model used to provide the boundary conditions. (b) Experimental and (c) modeled normalized earing profiles (h^*) as a function of prerolling amount. Source: Ref 27

Grain subdivision during high shear processing (equal-channel angular extrusion, or ECAE) has also been incorporated into the VPSC approach (Ref 31), rationalizing the effect of different routes on the effectiveness of grain refinement and the changes in grain morphology in fcc aluminum. This work was

continued in copper (Ref 32), and again FE simulations were found to be helpful in providing the input deformation history to the texture simulation; here, the grain co-rotation scheme (Ref 30) was found to be helpful in enforcing the soft coupling between adjacent grains. The evolution of bcc textures during

ECAE has also been modeled using the VPSC approach (Ref 33), as has the evolution of texture in pure aluminum (Ref 34). Finally, in the ECAE of initially untextured powder metallergy beryllium, deforming primarily by $<c + a>$ slip at elevated temperatures, the textures could be reproduced in form, but the overall texture strengths were found to be too strong; due to the nature of the strain path, basal $<a>$ slip was largely suppressed (Ref 35).

Grain-size dependence in the flow law has also been examined using a one-site VPSC model in copper, both with a single grain size and with a bimodal mixture of coarse (10 μm) and fine (0.7 μm) grains (Ref 36). A Hall-Petch dependence could be incorporated in the flow rule and the model then used to design bimodal microstructures of tailored strength and ductility. An obstacle hardening law has also been implemented to simulate irradiation hardening in ferritic steels (Ref 37), with a measured initial texture.

Single-Phase Materials with Twinning. Twinning results in significant reorientations of individual grains and therefore significant texture changes despite the small amount of deformation that twinning can accommodate (Ref 38). Therefore, twinning has been the subject of significant effort within the self-consistent modeling community, particularly in the technologically important zirconium and titanium systems, as well as in brasses, Hadfield steels, zinc, and magnesium. More recently, this has become of importance in examining twinning-induced plasticity steels. However, twinning is difficult to model because:

- The factors controlling the stress state that induces twinning are not well understood.
- The potential for multiple twin orientations in a single grain gives rise to a potential bifurcation in the number of grains in the problem.
- The effect of twin boundaries on the subsequent micromechanics of a twinned gain are nontrivial. The effect of the twinning event itself on the strain in the parent grain are clearly important and yet, until recently, undetermined.

The original VPSC model (Ref 7) included a description of twinning based on the volume fraction transfer concept. There, twinning is assumed to be associated with a critical stress and gives rise to an associated twin strain. When twinning occurs, volume fraction (weight) is reassigned between the twinning grain and the corresponding twinned orientation, avoiding the bifurcation problem. Qualitatively, this approach was successful in reproducing the observed textures in near-alpha Zircaloy-2, much more so than a Taylor full-constraints model. In addition, the model was used to examine the texture evolution during the final rolling of Zircaloy-4 tube (Ref 39), with the result that the main features of the texture were reproduced, but the magnitudes were overpredicted.

Another approach is the preferred twin reorientation (PTR) scheme, where a grain is transferred to a new orientation when its overall twin fraction reaches 50%, using the most active twin orientation. This, then, allows the hardening due to twinning to be modeled by changing the flow law to account for the additional boundaries, in a fashion that accounts for the ease with which some dislocations will pass through the twin boundary compared to others (Ref 40). Here, the simulation correctly reproduced the hardening response of a Hadfield steel and the increase in strength of the twinned regions.

The PTR model has also been applied to the texture evolution of magnesium-lithium and magnesium-yttrium (Ref 41), where the increase in $<c + a>$ slip activity with alloying, similar to that observed with aluminum additions in α-titanium, was successfully detected and employed to rationalize the changes in texture evolution observed during plane-strain compression. This work has been extended to AZ31B (Mg-3Al-1Zn wt%), finding that the high tensile ductility observed

in many magnesium alloys could only be explained by the presence of $<c + a>$ slip as well as tensile twinning, as observed in transmission electron microscopy and inferred using a polycrystal model calibrated against the observed texture evolution (Ref 42) (Fig. 4). In scenarios where magnesium is found to twin very significantly, such as the tensile testing of AM20 (Mg-2Al-0.5Mn-0.2Zn wt%), the textures can also be reproduced remarkably accurately (Ref 43).

In beryllium, which is highly anisotropic and which twins very easily at room temperature and high strain rates, VPSC models have also been used to rationalize the observed texture changes during in-plane and compression deformation, both with and without an initial rolling texture (Ref 44). The factor 2 strength variations observed were correctly predicted, although again, the texture changes were slightly overpredicted.

More recently, following Kalidindi (Ref 45) and Salem (Ref 46), a composite grain (CG) flow model has been implemented for twinning in the VPSC framework (Ref 47), (Fig. 5).
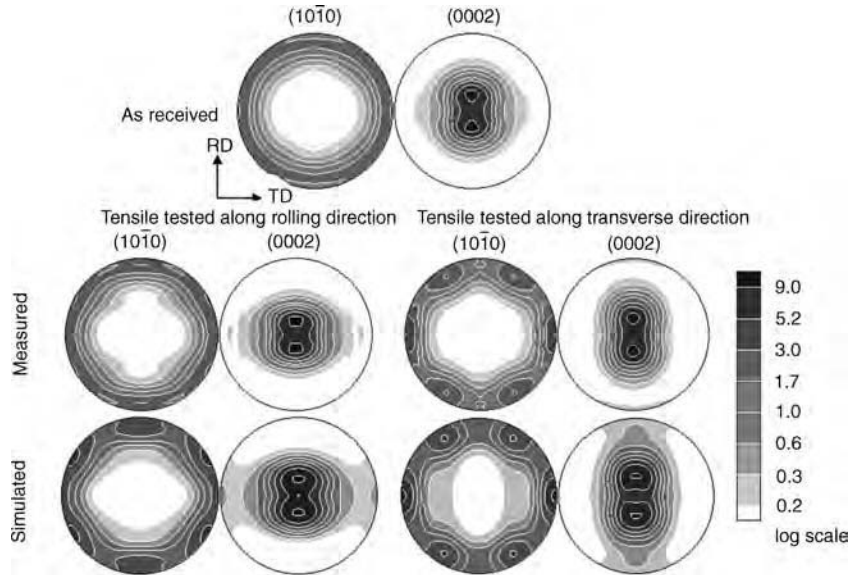
Fig. 4 Texture evolution due to tensile testing of magnesium alloy AZ31B rolled sheet to a strain of ~0.11; comparison between model and experiment. Source: Ref 41
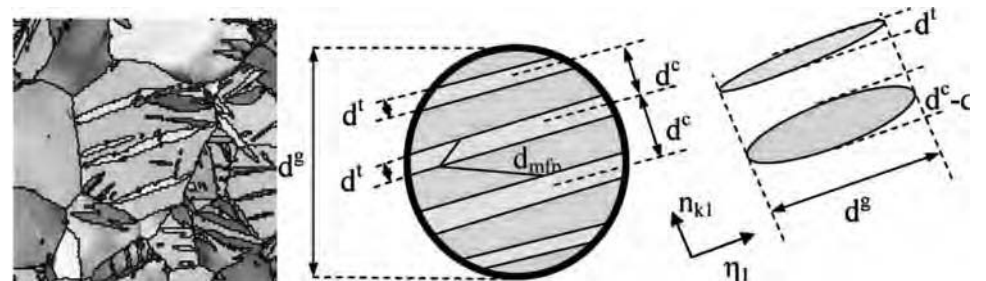
Fig. 5 Electron backscatter diffraction image of twinned zirconium. In the composite grain model, the twins are modeled explicitly within the grain, defining various microstructural parameters such as the twin thickness, orientation relationship, and so on. Source: Ref 47

There, an approach is taken that is more similar to that taken for martensitic transformations; the twin lamellae are treated explicitly within the micromechanical model for the individual grain, such that effects such as the mean free path between twins, the requirement for strain compatibility across the composite grain, the rise in Hall-Petch hardening, and the requirement for geometrically necessary dislocations can then all be included in the model. The result is that the detail of the hardening and texture response of rolled zirconium to in-plane (IPC) and through-thickness compression (TTC) experiments, as well as mixed IPC-TTC load paths, is much improved at temperatures (76 K) where twinning is the dominant concern.

Two-Phase Aggregates. Many materials are composed of dual-phase composites, such as duplex steels and α + β titanium alloys. Therefore, it is desirable to be able to model such situations within codes that aim to predict texture evolution. The simplest way to perform this would be to incorporate both phases as individual grains within the homogenous medium, but such an approach would ignore the grain interactions that are central to the texture evolution.

The two-site VPSC model, where two inclusions are modeled within the medium (Ref 29), was developed for this purpose. It was first applied to Ti-6Al-4V, where there exists a Burgers relationship $(0002)_\alpha \parallel (110)_\beta$, $[1\bar{2}10]_\alpha \parallel [111]_\beta$ between the two phases (Ref 48). The effect of the grain interaction was to promote prismatic slip over basal slip in the α-phase, minimizing the discontinuity in distortion across the habit plane. However, the concentration of flow into the low-volume-fraction β-phase resulted in an overly strong texture in the bcc phase.

The model has also been applied to powder metallurgy silver-nickel and copper-iron blends, with no orienation relationship between the two phases (Ref 49), finding that, if the compatibility requirement between the two sites was treated as a fitting parameter, reasonable agreement between the textures could be obtained. Iron-copper blends were examined in more detail by Commentz (Ref 50), finding that the two-site VPSC model predicted the observed textures best for copper (soft) inclusions in an iron (hard) matrix, rather than the reverse, but in all cases, the texture strengths tended to be overestimated.

Semiatin (Ref 51) has also developed a self-consistent model for the hot working of equiaxed α/β Ti-6Al-4V, focusing on the use of power law viscoplasticity where the strengths of each phase are composition dependent, based on the temperature dependence of the volume fractions and compositions from thermodynamic models and experimental observations, with physically-based activation energies and stress exponents. The model successfully reproduced the flow behavior observed during hot working.

The CG approach has also been taken to model the deformation of copper-niobium multilayers, where perfect strain compatibility between the grains must be satisfied (Ref 52), reproducing the form and magnitude of the textures resulting from rolling. Here, the interfaces are incoherent, and therefore, there is no dislocation coupling between the phases.

Deformation twinning observed in the α-phase of Ti-6Al-4V during high-temperature (800 °C) ECAE has also been modeled, although only using a single-phase VPSC model (Ref 53). The observed α-textures could be reproduced in form and magnitude, although the exact texture pattern was not reproduced, presumably due to the absence of the β-phase.

The single-site VPSC formulation has also been adapted to model pearlitic and bainitic steels (Ref 54), incorporating the micromechanical effect of the pearlite lamellae into the hardening formulation and performing an Eshelby analysis to account for the effect of carbides in upper bainite. The mechanical properties and effect of torsion on the texture anisotropy were successfully reproduced.

Coupling to Finite-Element Models. In many situations, such as rolling, the strain paths within the material are known, but more generally, as in forging, the strain paths must be computed by means of a FE model and will vary spatially throughout the part being produced. In principle the textures that evolve will influence the deformation behavior, and therefore, the texture evolution model should be incorporated within the FE model rather than be treated as a postprocessing operation. Given that a single strain-step in a VPSC model can be calculated quite quickly on modern computers (on the order of 20 ms), this would still allow for a typical 10,000-element FE simulation to be run in a tractable period of time, particularly given parallel computing of the VPSC step. Indeed, this relatively rapid computation time is a major attraction of self-consistent models over competing approaches, such as crystal-plasticity finite-element modeling (CPFEM), which are more usually applied to studying the detail of grain interactions.

Some initial steps in this direction have already been made, with the interfacing of a VPSC model with an explicit FE code (Ref 55, 56). The spatial variation component of the problem studied, the response of pure zirconium bent beams, presented no additional challenges beyond the interfacing of the code compared to the authors' other work on texture simulations of similar strain paths in zirconium.

Variational Formulations. As pointed out by Castañeda and others (Ref 57–59), other bounds can be deduced than the Voigt and Reuss approaches, especially the Hashin-Strikman (H-S) bounds, which make use of a variational approach to determine the bounds on the behavior of nonlinear composites. A key finding is that, for very anisotropic crystals, the self-consistent approach may violate the H-S bounds. This has led to the development of self-consistent models that take account of the second-order field fluctuations within the grains, variously termed the affine formulations (Ref 60), the variational self-consistent approach of Gilormini (Ref 61), the second-order approach (Ref 62), or based on fast Fourier transforms (Ref 63). Generally, it is found that previous estimates of texture evolution and overall strength from the conventional, tangent self-consistent approach are accurate for low-anisostropy cubic materials, are in reasonable agreement for moderately anisotropic hexagonal close-packed metals such as titanium and zirconium, but are less robust for strongly anisotropic materials such as ice. It should be emphasized that these issues are only of significant concern for very high strength anisotropy. However, it is expected that over time, these variational approaches will be used more widely for texture simulations.

Recrystallization. Because the plastic strain energy can be estimated from the deformation system activity during a polycrystal simulation (Ref 21), and because this is one of the major driving forces for recrystallization, the output from a VPSC model can be used in a subsequent recrystallization model to simulate recrystallization textures subsequent to plastic deformation. This has been performed, allowing significant insight to be gained into the textures observed in minerals (Ref 64).

Other Texture Simulation Approaches. It should be mentioned that there are other approaches to texture simulation than the self-consistent family of models. One research thread in particular has focused on developing Taylor-type models. Here, the full-constraints (FC) Taylor theory (same strain in all grains) has led to the relaxed-constraints approach (relaxing a few of the strain components) and subsequently to multigrain relaxed-constraints models (ALAMEL or GIA) models. These are significantly more successful at predicting texture evolution in both interstitial-free steels and AA5182 than Taylor FC models (Ref 65).

An alternative approach is to model the grains explicitly in a FE code, as in the CPFEM approach (Ref 66–69). These approaches are generally very successful at modeling deformation textures, especially if each grain is modeled using multiple elements in order to allow for strain variations within each grain. However, they are of order 1000× more computationally expensive (Ref 66). They are reviewed in the following article, "Crystal-Scale Simulations Using Finite-Element Formulations" in this Volume.

Concluding Remarks

Self-consistent models have been developed over the last 15 to 20 years to model deformation textures, evolving from the starting point of Taylor FC models, based on the Eshelby inclusion approach. They are computationally quick and can be implemented into macroscale FE simulations of components. Many successful examples of self-consistent models

compared to real textures can be found in the literature, although micromechanical processes that tend to reduce the level of texture development should be taken into account. They can be implemented rather simply by fitting to the macroscopic flow curve with few fitting parameters, or additional physics can be included by reference to the actual material structure, thermal and strain-rate dependence of the deformation mechanisms, and so on. More recently, enhancements have been proposed to the approach that improve the behavior for strongly anisotropic materials.

REFERENCES

1. I.L. Dillamore and W.T. Roberts, Preferred Orientation in Wrought and Annealed Metals, *Metall. Rev.,* Vol 10 (No. 39), 1965, p 271–380
2. J.D. Eshelby, The Determination of the Elastic Field of an Ellipsoidal Inclusion, and Related Problem, *Proc. R. Soc. (London) A,* Vol 241, 1957, p 376–396
3. E. Kröner, Berechnung der Elastischen Konstanten des Vielkristalls aus den Konstanten des Einkristalls, *Z. Phys.,* Vol 151, 1958, p 504–518
4. R. Hill, Continuum Micro-Mechanics of Elastoplastic Polycrystals, *J. Mech. Phys. Solids,* Vol 13, 1965, p 89–101
5. J.W. Hutchinson, Elastic-Plastic Behavior of Polycrystalline Metals and Composites, *Proc. R. Soc. (London) A,* Vol 319, 1970, p 247–272
6. D. Dye, H.J. Stone, and R.C. Reed, A Two Phase Elastic-Plastic Self-Consistent Model for the Accumulation of Microstrains in Waspaloy, *Acta Mater.,* Vol 49, 2001, p 1271–1283
7. R.A. Lebensohn and C.N. Tomé, A Self-Consistent Anisotropic Approach for the Simulation of Plastic Deformation and Texture Development of Polycrystals: Application to Zirconium Alloys, *Acta Metall. Mater.,* Vol 41 (No. 9), 1993, p 2611–2624
8. P.A. Turner and C.N. Tomé, A Study of Residual Stresses in Zircaloy-2 with Rod Texture, *Acta Metall. Mater.,* Vol 42 (No. 12), 1994, p 4143–4153
9. T.M. Holden, R.A. Holt, and A.P. Clarke, Intergranular Strains in Inconel-600 and the Impact on Interpreting Stress Fields in Bent Steam-Generator Tubing, *Mater. Sci. Eng. A,* Vol 246, 1998, p 180–198
10. D. Dye, H.J. Stone, and R.C. Reed, Intergranular and Interphase Microstresses, *Curr. Opin. Solid State Mater. Sci.,* Vol 5, 2001, p 31–37
11. S.L. Raghunathan, A.M. Stapleton, R.J. Dashwood, M. Jackson, and D. Dye, Micromechanics of Ti-10V-2Fe-3Al: In Situ Synchrotron Characterisation and Modelling, *Acta Mater.,* Vol 55, 2007, p 6861–6872
12. U.F. Kocks, C.N. Tomé, and H.-R. Wenk, *Texture and Anisotropy,* Cambridge University Press, Cambridge, U.K., 1998
13. H.J. Bunge and C. Esling, *Quantitative Texture Analysis,* DGM, Germany, 1982
14. H.R. Wenk, S. Matthies, and G.W. Vinel, Some Basic Concepts of Texture Analysis and Comparison of 3 Methods to Calculate Orientation Distributions from Pole Figures, *J. Appl. Crystallogr.,* Vol 21 (No. 4), 1988, p 285–304
15. J.S. Kallend, U.F. Kocks, A.D. Rollett, and H.R. Wenk, Operational Texture Analysis, *Mater. Sci. Eng. A,* Vol 132, 1991, p 1–11
16. H.R. Wenk, L. Lutterotti, and S. Vogel, Texture Analysis with the New HIPPO TOF Diffractometer, *Nucl. Instrum. Methods Phys. Res. A,* Vol 515, 2003, p 575–588
17. H.R. Wenk, S. Matthies, and L. Lutterotti, Texture Analysis from Diffraction Spectra, *Mater. Sci. Forum,* Vol 157–162, 1994, p 473–480
18. A.C. Larson and R.B. Von Dreele, "General Structure Analysis System (GSAS)," Technical Report LAUR 86-748, Los Alamos National Laboratory, 1994
19. H.R. Wenk and S. Grigull, Synchrotron Texture Analysis with Area Detectors, *J. Appl. Crystallogr.,* Vol 36, 2003, p 1040–1049
20. C. Maurice and J.H. Driver, Hot Rolling Textures of FCC Metals, Part 2: Numerical Simulations, *Acta Mater.,* Vol 45, 1997, p 4639–4649
21. B. Bacroix, A. Miroux, and O. Castelnau, Simulation of the Orientation Dependence of Stored Energy during Rolling Deformation of Low Carbon Steels, *Model. Simul. Mater. Sci. Eng.,* Vol 7 (No. 5), 1999, p 851–864
22. D.A. Hughes, R.A. Lebensohn, H.R. Wenk, and A. Kumar, Stacking Fault Energy and Microstructure Effects on Torsion Texture Evolution, *Proc. R. Soc. (London) A,* Vol 456 (No. 1996), 2000, p 921–953
23. S.H. Choi, J.C. Brem, F. Barlat, and K.H. Oh, Macroscopic Anisotropy in AA5019A Sheets, *Acta Mater.,* Vol 48 (No. 8), 2000, p 1853–1863
24. H.R. Wenk, S. Matthies, R.J. Hemley, H.K. Mao, and J. Shu, The Plastic Deformation of Iron at Pressures of the Earth's Inner Core, *Nature,* Vol 405 (No. 6790), 2000, p 1044–1047
25. O. Engler, Deformation and Texture of Copper-Manganese Alloys, *Acta Mater.,* Vol 48 (No. 20), 2000, p 4827–4840
26. A.B. Lopes, F. Barlat, J.J. Gracio, J.F.F. Duarte, and E.F. Rauch, Effect of Texture and Microstructure on Strain Hardening Anisotropy for Aluminum Deformed in Uniaxial Tension and Simple Shear, *Int. J. Plast.,* Vol 19 (No. 1), 2003, p 1–22
27. O. Engler and S. Katz, Simulation of Earing Profiles from Texture Data by Means of a Visco-Plastic Self-Consistent Polycrystal Plasticity Approach, *Mater. Sci. Eng. A,* Vol 373 (No. 1–2), 2004, p 350–362
28. M. Berveiller, O. Fassi-Fehri, and A. Hihi, The Problem of Two Plastic and Heterogeneous Inclusions in an Anisotropic Medium, *Int. J. Eng. Sci.,* Vol 25 (No. 6), 1987, p 691–709
29. R.A. Lebensohn and G.R. Canova, A Self-Consistent Approach for Modelling Texture Development of Two-Phase Polycrystals: Application to Titanium Alloys, *Acta Mater.,* Vol 45 (No. 9), 1997, p 3687– 3694
30. C.N. Tomé, R.A. Lebensohn, and C.T. Necker, Mechanical Anisotropy and Grain Interaction in Recrystallized Aluminum, *Metall. Mater. Trans. A,* Vol 33, 2002, p 2635–2648
31. I.J. Beyerlein, R.A. Lebensohn, and C.N. Tomé, Modeling Texture and Microstructural Evolution in the Equal Channel Angular Extrusion Process, *Mater. Sci. Eng. A,* Vol 345 (No. 1–2), 2003, p 122–138
32. S.Y. Li, I.J. Beyerlein, D.J. Alexander, and S.C. Vogel, Texture Evolution during Multi-Pass Equal Channel Angular Extrusion of Copper: Neutron Diffraction Characterization and Polycrystal Modeling, *Acta Mater.,* Vol 53, 2005, p 2111–2125
33. S. Li and I.J. Beyerlein, Modelling Texture Evolution in Equal Channel Angular Extrusion of BCC Materials: Effects of Processing Route and Initial Texture, *Model. Simul. Mater. Sci. Eng.,* Vol 13, 2005, p 509–530
34. A.P. Zhilyaev, D.L. Swisher, K. Oh-ishi, T.G. Langdon, and T.R. McNelley, Microtexture and Microstructure Evolution during Processing of Pure Aluminum by Repetitive ECAP, *Mater. Sci. Eng. A,* Vol 429, 2006, p 137–148
35. R.D. Field, K.T. Hartwig, C.T. Necker, J.F. Bingert, and S.R. Agnew, Equal-Channel Angular Extrusion of Beryllium, *Metall. Mater. Trans. A,* Vol 33, 2002, p 965–972
36. B. Raeisinia, C.W. Sinclair, W.J. Poole, and C.N. Tomé, On the Impact of Grain Size Distribution on the Plastic Behaviour of Polycrystalline Metals, *Model. Simul. Mater. Sci. Eng.,* Vol 16, 2008
37. C. Deo, C. Tomé, R. Lebensohn, and S. Maloy, Modeling and Simulation of Irradiation Hardening in Structural Ferritic Steels for Advanced Nuclear Reactors, *J. Nucl. Mater.,* Vol 377, 2008, p 136–140
38. J.W. Cristian and S. Mahajan, Deformation Twinning, *Prog. Mater. Sci.,* Vol 39, 1995, p 1–157
39. R.A. Lebensohn, M.I. Gonzalez, C.N. Tomé, and A.A. Pochettino, Measurement and Prediction of Texture Development during a Rolling Sequence of Zircaloy-4 Tubes, *J. Nucl. Mater.,* Vol 229, 1996, p 57–64
40. I. Karaman, H. Sehitoglu, A.J. Beaudoin, Y.I. Chumlyakov, H.J. Maier, and C.N. Tomé, Modeling the Deformation Behavior

of Hadfield Steel Single and Polycrystals due to Twinning and Slip, *Acta Mater.*, Vol 48, 2000, p 2031–2047

41. S.R. Agnew, M.H. Yoo, and C.N. Tomé, Application of Texture Simulation to Understanding Mechanical Behavior of Mg and Solid Solution Alloys Containing Li or Y, *Acta Mater.*, Vol 49, 2001, p 4277– 4289

42. S.R. Agnew and O. Duygulu, Plastic Anisotropy and the Role of Non-Basal Slip in Magnesium Alloy AZ31B, *Int. J. Plast.*, Vol 21, 2005, p 1161–1193

43. S.B. Yi, H.G. Brokmeier, R.E. Bolmaro, K.U. Kainer, and T. Lippmann, In Situ Measurements of Texture Variations during a Tensile Loading of Mg-Alloy AM20 Using Synchrotron X-Ray Radiation, *Scr. Mater.*, Vol 51, 2004, p 455–460

44. D.W. Brown, S.P. Abeln, W.R. Blumenthal, M.A.M. Bourke, M.C. Mataya, and C.-N. Tomé, Development of Crystallographic Texture during High Rate Deformation of Rolled and Hot-Pressed Beryllium, *Metall. Mater. Trans. A,* Vol 36, 2005, p 929–939

45. S.R. Kalidindi, Modeling Anisotropic Strain Hardening and Deformation Textures in Low Stacking Fault Energy FCC Metals, *Int. J. Plast.*, Vol 17, 2001, p 837–860

46. A.A. Salem, S.R. Kalidindi, and S.L. Semiatin, Strain Hardening Due to Deformation Twinning in Alpha-Titanium: Constitutive Relations and Crystal-Plasticity Modeling, *Acta Mater.*, Vol 53, 2005, p 3495–3502

47. G. Proust, C.N. Tomé, and G.C. Kaschner, Modeling Texture, Twinning and Hardening Evolution during Deformation of Hexagonal Materials, *Acta Mater.*, Vol 55, 2007, p 2137–2148

48. W.G. Burgers, *Physica,* Vol 1, 1934, p 561–586

49. R.E. Bolmaro, R.A. Lebensohn, and H.G. Brokmeier, Crystal Spin in Two-Sites Self Consistent Models: From Kinematics to Kinetics, *Comput. Mater. Sci.*, Vol 9, 1997, p 237–250

50. B. Commentz, C. Hartig, and H. Mecking, Micromechanical Interaction in Two-Phase Iron-Copper Polycrystals, *Comput. Mater. Sci.*, Vol 16, 1999, p 237–247

51. S.L. Semiatin, F. Montheillet, G. Shen, and J.J. Jonas, Self-Consistent Modeling of the Flow Behavior of Wrought Alpha/Beta Titanium Alloys under Isothermal and Nonisothermal Hot-Working Conditions, *Metall. Mater. Trans. A,* Vol 33, 2002, p 2719–2727

52. K. Al-Fadhalah, C.N. Tomé, A.J. Beaudoin, I.M. Robertson, J.P. Hirth, and A. Misra, Modeling Texture Evolution during Rolling of a Cu-Nb Multilayered System, *Philos. Mag.*, Vol 85, 2005, p 1419–1440

53. G.G. Yapici, I. Karaman, and Z.-P. Luo, Mechanical Twinning and Texture Evolution in Severely Deformed Ti-6Al-4V at High Temperatures, *Acta Mater.*, Vol 54, 2006, p 3755–3771

54. D. Canadinc, H. Sehitoglu, H.J. Maier, and P. Kurath, On the Incorporation of Length Scales Associated with Pearlitic and Bainitic Microstructures into a Visco-Plastic Self-Consistent Model, *Mater. Sci. Eng. A,* Vol 485, 2008, p 258–271

55. C.N. Tomé, P.J. Maudlin, R.A. Lebensohn, and G.C. Kaschner, Mechanical Response of Zirconium, Part I: Derivation of a Polycrystal Constitutive Law and Finite Element Analysis, *Acta Mater.*, Vol 49, 2001, p 3085–3096

56. G.C. Kaschner, J.F. Bingert, C. Liu, M.L. Lovato, P.J. Maudlin, M.G. Stout, and C.N. Tomé, Mechanical Response of Zirconium, Part II: Experimental and Finite Element Analysis of Bent Beams, *Acta Mater.*, Vol 49, 2001, p 3097–3108

57. P. Ponte Castañeda and P. Suquet, Nonlinear Composites, *Adv. App. Mech.*, Vol 34, 1998, p 171–302

58. A. Zaoui, Continuum Micromechanics: Survey, *J. Eng. Mech.*, Vol 128 (No. 8), 2002, p 808–816

59. P. Gilormini, A Shortcoming of the Classical Nonlinear Extension of the Self-Consistent Model, *C.R. Acad. Sci. Ser. II,* Vol 320, 1995, p 115–122

60. R. Masson, M. Bornert, P. Suquet, and A. Zaoui, An Affine Formulation for the Prediction of the Effective Properties of Nonlinear Composites and Polycrystals, *J. Mech. Phys. Solids,* Vol 48, 2000, p 1203–1227

61. P. Liu, P. Gilormini, and P.P. Castañeda, Variational Self-Consistent Estimates for Texture Evolution in Viscoplastic Polycrystals, *Acta Mater.*, Vol 51, 2003, p 5425–5437

62. R.A. Lebensohn, C.N. Tomé, and P. Ponte Castañeda, Self-Consistent Modelling of the Mechanical Behaviour of Viscoplastic Polycrystals Incorporating Intragranular Field Fluctuations, *Philos. Mag.*, Vol 87, 2007, p 4287–4322

63. R.A. Lebensohn, Y. Liu, and P.P. Castañeda, On the Accuracy of the Self-Consistent Approximation for Polycrystals: Comparison with Full-Field Numerical Simulations, *Acta Mater.*, Vol 52, 2004, p 5347– 5361

64. H.R. Wenk, G. Canova, Y. Brechet, and L. Flandin, A Deformation-Based Model for Recrystallization of Anisotropic Materials, *Acta Mater.*, Vol 45, 1997, p 3283–3296

65. M. Seefeldt, P. Van Houtte, S.Y. Li, and L. Delannay, Deformation Texture Prediction: From the Taylor Model to the Advanced Lamel Model, *Int. J. Plast.*, Vol 21 (No. 3), 2005, p 589–624

66. P. Bate, Modelling Deformation Microstructure with the Crystal Plasticity Finite-Element Method, *Philos. Trans.R. Soc. (London) A,* Vol 357, 1999, p 1589–1601

67. P.S. Bate and J.Q. da Fonseca, Texture Development in the Cold Rolling of IF Steel, *Mater. Sci. Eng. A,* Vol 380, 2004, p 365–377

68. A. Kumar and P.R. Dawson, Computational Modeling of FCC Deformation Textures over Rodrigues' Space, *Acta Mater.*, Vol 48, 2000, p 2719–2736

69. P.R. Dawson, S.R. MacEwen, and P.D. Wu, Advances in Sheet Metal Forming Analyses: Dealing with Mechanical Anisotropy from Crystallographic Texture, *Int. Mater. Rev.*, Vol 48, 2003, p 86–122

Crystal-Scale Simulations Using Finite-Element Formulations

P.R. Dawson and D.E. Boyce, Cornell University

SOLID METALLIC MATERIALS typically are crystalline, existing with well-defined lattice structures. The lattices possess certain symmetries dependent on the packing arrangements of atoms. Accompanying the symmetries are anisotropies of the mechanical properties, both elastic and plastic, of the individual crystals. Anisotropy is evident even at the continuum scale in aggregates of crystals having preferred orientations of the crystal lattices, known as crystallographic texture. The behaviors observed during processing of metallic solids are influenced by the presence of the anisotropic properties. Further, processing affects the anisotropy by altering the crystallographic texture and other features of the microstructural state.

To model the processing of metallic materials effectively, especially those aspects of a process that are strongly influenced by anisotropy, the properties inherited from the crystal structure must appear explicitly in the model formulation. Such aspects include the evolving state of the material during processing as well as its condition subsequent to processing. Constitutive models posed at the crystal scale are intended to quantify the behaviors that derive from the crystalline structure of the material. The coefficients of the elastic stiffness reflect the crystal symmetry directly. The equations for plastic flow are constructed from the premise of slip occurring on close-packed planes in close-packed directions. Both relations exhibit the anisotropy present at the crystal level and together provide a physically-motivated description of a crystalline material behavior. The anisotropic behavior at the crystal level implies that stress and deformation is not uniform over the volume of a loaded polycrystalline aggregate. The ability of crystal to carry load depends on the spatial orientation of its lattice, so in an aggregate of crystals displaying a range of orientations, there exists a range of properties in relation to the load. Finite-element formulations offer a powerful methodology for dealing with this complexity. The properties can vary spatially according to the domains of the crystals and can incorporate the observed elastic and plastic behaviors. Over the past two decades, it has become the dominant method for simulating processing of metallic materials.

The goal of this article is to introduce the reader to finite-element simulations in which single-crystal properties provide the basis for the constitutive model, to lay out the basic equations that need to be solved, and to provide examples illustrating the types of simulations amenable to this approach. Two physical size scales are considered. In the smaller, crystal scale, the crystals themselves are discretized with finite elements. A group (or aggregate) of crystals is represented by a mesh in which each element is part of a (or an entire) single crystal. Depending on how finely resolved the crystals are, the direct interactions of hundreds to tens of thousands of crystals can be modeled in this way. However, modeling thousands of crystals still accounts for only a small piece of material, and ultimately, the practical application of polycrystal modeling is at the larger, continuum scale. This is the scale of actual components or parts with elements of the finite-element mesh being much larger than individual crystals of the material. At this level, the material description can no longer include detailed interactions of every crystal but instead must approximate the net effect of the crystal-scale behavior. At each computational point in the body (e.g., each quadrature point in a finite element), an entire aggregate of crystals is modeled but in a way that is only weakly coupled to the continuum scale.

The two scales are shown in Fig. 1. In Fig. 1(a), the polycrystal is being modeled. Each constituent crystal is modeled geometrically by one or more elements, and properties are computed directly from the single-crystal slip model. The continuum scale is illustrated in Fig. 1(b). There, the body is subdivided into finite elements, with each element deriving its properties from a polycrystal model based on relatively simple rules for the crystal interactions. For both the crystal and continuum scales, basic frameworks of the formulations are presented and then illustrated with representative examples. The examples illustrate typical issues that can be addressed at each scale by modeling. At the crystal scale, issues often center around the influence of the polycrystalline microstructure on a material behavior. Simulations at this scale can combine morphological (size and shape) and topological (grain or phase connectivity) features with the anisotropy of both the elastic and plastic properties. Using this capability, one can examine grain interaction effects on load sharing, strain hardening, and damage initiation at the level of individual grains embedded in a polycrystal. At the continuum scale, it is possible to investigate how forming effects the development of microstructural attributes such as the crystallographic texture, as well as how attributes such as the
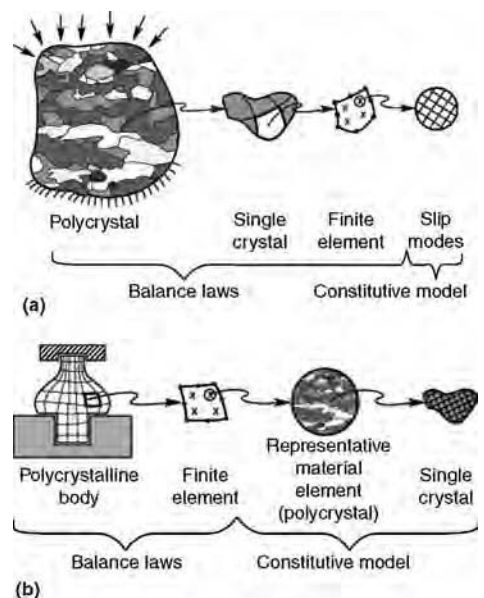
Fig. 1 Schematic diagrams indicating the characteristic sizes of crystals relative to the entire body being modeled. (a) Crystal scale. (b) Continuum scale

texture influence the outcome of a forming process. Examples of each include texture evolution in rolling or extrusion and earing or localized thinning in the forming of textured sheet metal. In addition, one can directly introduce anisotropy accompanying texture into simulations designed to assess the impact of factors such as residual-stress distributions or orientation dependence of stiffness and strength on the service performance of a component.

Crystal Elastoplasticity—Theory, Methods, and Applications

Metals are capable of deforming elastically and plastically by a number of different physical mechanisms. Plastic flow occurs by different combinations of slip, twinning, and diffusion, depending on the regime of temperature and stress (or strain rate). A processing window can be defined in terms of the active ranges of strain rate and temperature, which in turn determines the dominant deformation modes. That processing window here is limited to one in which slip is the dominant mode of plastic deformation and is characterized by moderate strain rate and moderate homologous temperature (Ref 1).

The following notation convention is employed: plain fonts are used for scalars, math bold fonts are used for vectors, and math sans serif fonts are used for higher-order tensors or matrices. Examples of a scalar, a vector, and a tensor are, respectively: a, $\boldsymbol{a}$, and a. In addition, lower-case variables associated with either the stress or the motion apply to quantities at the crystal scale, while upper-case letters apply to quantities at the continuum scale. A superscript $'$ refers to the deviatoric part of a tensor, and a superposed dot indicates material time differentiation.

Elastic and Plastic Behaviors of Single Crystals. An elastoplastic deformation within a single crystal consists of elastic and plastic parts, as well as a rotation. Within the specified processing window, the plastic deformation occurs through crystallographic slip between atomic planes of the crystal lattice. The elastic deformations are associated with lengthening or shortening of the interatomic distances. The equations for the elastic and plastic behaviors are presented first separately. This is followed by a discussion of the kinematic framework needed to splice the elastic and plastic responses into a compatible motion. The elastic response follows a linear relation (Hooke's law):

$$\boldsymbol{\tau} = \mathsf{c}\mathsf{e} \tag{Eq 1}$$

where c the tensor containing elastic moduli for the appropriate crystal symmetry (Ref 2), and e is the elastic strain. (More precise definitions of the strain and stress relative to deformed and undeformed configurations follow.) The Kirchhoff stress, $\boldsymbol{\tau}$, is related to the Cauchy

stress, $\boldsymbol{\sigma}$, through $\boldsymbol{\tau} = \beta\boldsymbol{\sigma}$, where $\beta = \det(\mathsf{I} + \mathsf{e})$, and I is the second-order identity tensor.

The plastic shearing rate on the α-slip system, $\dot{\gamma}^\alpha$, is related to the crystal stress by the power law relation (Ref 3):

$$\dot{\gamma}^\alpha = \dot{\gamma}0\left(\frac{|\tau^\alpha|}{g^\alpha}\right)^{\frac{1}{m}}\mathrm{sgn}(\tau^\alpha) \tag{Eq 2}$$

where g^α is the slip system strength, $\dot{\gamma}_0$ is a reference shear rate, and m is the rate sensitivity of slip. The resolved shear stress, τ^α, is the plastic work rate conjugate to $\dot{\gamma}^\alpha$ and is the projection of the deviatoric part of the Kirchhoff stress, $\boldsymbol{\tau}'$, on the α-slip system as:

$$\tau^\alpha = \mathrm{tr}(\mathsf{p}^\alpha\boldsymbol{\tau}') \quad \text{where}$$
$$\mathsf{p}^\alpha = \mathrm{sym}(\mathsf{T}^\alpha) = \mathrm{sym}(\boldsymbol{b}^\alpha \otimes \mathsf{m}^\alpha) \tag{Eq 3}$$

where T^α is the Schmid tenor, formed from the product of the vector, $\boldsymbol{b}^\alpha$, the slip direction, and the vector, $\boldsymbol{m}^\alpha$, the normal to the slip plane, associated with the α-slip system (Ref 4).

The slip system strength, g^α, evolves according to the following hardening rule:

$$\dot{g}^\alpha = h_0\left(\frac{g_s(\dot{\gamma}) - g^\alpha}{g_s(\dot{\gamma}) - g_0}\right)^n \dot{\gamma} \tag{Eq 4}$$

where:

$$g_s(\dot{\gamma}) = g_1\left(\frac{\dot{\gamma}}{\dot{\gamma}_1}\right)^{m'} \quad \text{and} \quad \dot{\gamma} = \sum_\alpha |\dot{\gamma}^\alpha| \tag{Eq 5}$$

Here, $\dot{\gamma}^\alpha$ is the net shear strain rate in the crystal, $g_s(\dot{\gamma})$ is the saturation hardness, and h_0, g_0, n, g_1, $\dot{\gamma}_1$, and m' are slip system hardening parameters. In the basic form of the crystal-plasticity model, the slip system strengths are taken to be the same.

Central to the behavior of crystalline materials is the anisotropy associated with the elastic and plastic behaviors. The elastic moduli and the slip systems derive their values in part from geometric attributes of the crystal lattice. Thus, it is essential to specify the orientation of the crystallographic lattices in space with respect to a reference. A number of options are available to parameterize the lattice orientation. The one chosen here is the angle-axis parameterization of Rodrigues, as described by Frank (Ref 5, 6):

$$\boldsymbol{r} = \boldsymbol{n}\tan\frac{\phi}{2} \tag{Eq 6}$$

where $\boldsymbol{n}$ is the rotation axis, and ϕ is the angle of rotation about this axis. A Rodrigues vector specifies the rotation needed to take base vectors aligned with the reference axes to coincidence with axes embedded with the lattice of a crystal in its spatial orientation. The fundamental region for cubic crystal structure is depicted in Fig. 2(a). Fundamental regions are subregions of orientation space, constructed to contain only one of every set of orientations that are equivalent under crystal symmetries.
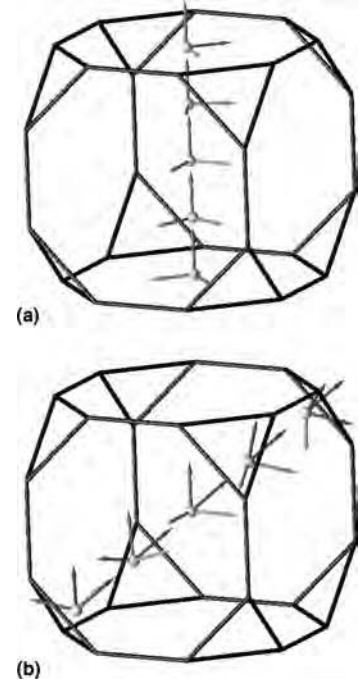
Fig. 2 Fundamental region for cubic symmetry crystals showing several orientations differing only by rotations about (a) a 001 axis of rotation and (b) a $\bar{1}11$ axis of rotation. The crystal axes are shown in gray scale. The $\bar{1}11$ axis of rotation is colored red.

Consequently, these are free of the ambiguity associated with assigning crystals to orientations taken from all of orientation space. The location of the surfaces of the fundamental region are defined by the crystal symmetries. The orientations of crystal axes are shown for several orientations lying along one line through the fundamental region in Fig. 2(b). One can see that these orientations are different only by rotation about a single axis and that the end points are symmetrically equivalent orientations. Lines through the fundamental region constitute crystallographic fibers, and for the case of the Rodrigues parameterization, fibers are always straight lines.

When the lattice orientations of a population of crystals are not uniformly distributed over the fundamental region, that population is said to have preferred orientation or texture. Mathematically, this is represented by an orientation distribution function (ODF), $A(\boldsymbol{r})$ (Ref 7). The ODF is defined over an appropriately reduced fundamental region of orientation space. More precisely, the ODF describes the local crystal density over the fundamental region, so that the crystal volume fraction enclosed within region Ω_{fr} of the fundamental region is given by:

$$v_f = \int_{\Omega_{fr}} A(\boldsymbol{r})d\Omega \tag{Eq 7}$$

where $d\Omega = \sqrt{\det\mathsf{g}}\,dr_1 dr_2 dr_3$ is a volume element of orientation space, and g is the metric tensor of the space. $A(\boldsymbol{r})$ is normalized to satisfy

the condition that its integral over the fundamental region is unity.

Kinematic Framework for Deforming Crystals and Polycrystals. The two modes of straining, elastic and plastic, jointly must produce a single compatible deformation. As depicted in Fig. 3, this is accomplished by a multiplicative decomposition of the deformation gradient **f**. A multiplicative decomposition is, in essence, a sequence of deformations. In this case, the plastic slip occurs first, followed by a rotation, and ending with the elastic stretching. The time interval for the process is shrunk toward a limit of zero, rendering the rate form of decomposition. The individual contributions to the deformation thus are occurring simultaneously over the course of a complete deformation history. However, the intermediate configurations are generated and retained. One intermediate configuration distinguishes the current configuration of the body from one that would be obtained by removing the stress. The change between this unloaded configuration and the current configuration defines the elastic strains. The rate form also includes terms that accommodate purely rotational motion, or spins, that must appear for the kinematics to be consistent in the presence of geometric nonlinearities from large strains and rotations. The three-part decomposition is written as:

$$\mathbf{f} = \mathbf{f}^{\flat}\mathbf{f}^{\star}\mathbf{f}^{\sharp} = \mathbf{v}^{\flat}\mathbf{r}^{\star}\mathbf{f}^{\sharp} \qquad (Eq\ 8)$$

where $\mathbf{f}^{\sharp}$ is the purely plastic part of **f** arising from slip; $\mathbf{f}^{\star}$ is the lattice rotation, which may be written as $\mathbf{r}^{\star}$; and $\mathbf{f}^{\flat}$ is the elastic part of **f** (Ref 8). (Superscript symbols $\flat$, $\sharp$, and $\star$ refer to mapping over a change of configuration.) The deformation gradient $\mathbf{f}^{\sharp}$ can be used to define an intermediate configuration, $\hat{\mathcal{B}}$, which is a relaxed configuration obtained by unloading without rotation from the current configuration $\mathcal{B}$. Using this interpretation of $\hat{\mathcal{B}}$, the symmetric left elastic stretch tensor, $\mathbf{v}^{\flat}$, is introduced. For the case of small elastic strains, $\mathbf{v}^{\flat} = \mathbf{I} + \mathbf{e}^{\flat}$, where $\|\mathbf{e}^{\flat}\| \ll 1$. From this decomposition, the kinematics then are expressed in rate form as:

$$\ell = \dot{\mathbf{f}}\mathbf{f}^{-1} = \mathbf{d} + \mathbf{w} \qquad (Eq\ 9)$$

where **d** is the deformation rate tensor, and **w** is the spin tensor, expressed in the current
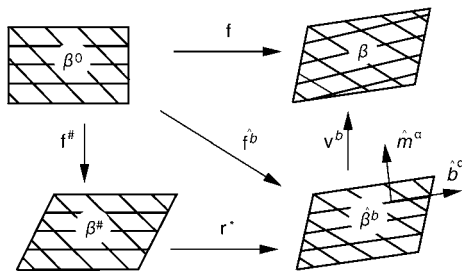
Fig. 3 Elementary kinematic decomposition for motion at the crystal level with plastic flow occurring via slip

configuration $\mathcal{B}$. These terms may be split into spherical and deviatoric parts, respectively, as:

$$\text{tr}(\mathbf{d}) = \text{tr}(\mathbf{e}^{\flat}) \qquad (Eq\ 10)$$

and

$$\mathbf{d}' = \mathbf{e}^{\flat\prime} + \hat{\mathbf{d}}^{\sharp\prime} + \mathbf{e}^{\flat\prime}\hat{\mathbf{w}}^{\sharp} - \hat{\mathbf{w}}^{\sharp}\mathbf{e}^{\flat\prime} \qquad (Eq\ 11)$$

$$\mathbf{w} = \hat{\mathbf{w}}^{\sharp} + \mathbf{e}^{\flat\prime}\hat{\mathbf{d}}^{\sharp\prime} - \hat{\mathbf{d}}^{\sharp\prime}\,\mathbf{e}^{\flat\prime} \qquad (Eq\ 12)$$

Here, $\mathbf{d}'$ and $\mathbf{e}^{\flat\prime}$ are the deviatoric components of **d** and $\mathbf{e}^{\flat}$, respectively. The ":" superscript indicates mapping forward by $\mathbf{r}^{\star}$ according to:

$$\hat{\mathbf{d}}^{\sharp\prime} = \mathbf{r}^{\star}\mathbf{d}^{\sharp\prime}\mathbf{r}^{\star T} \quad \text{and} \quad \hat{\mathbf{w}}^{\sharp} = \mathbf{r}^{\star}\,\mathbf{w}^{\sharp}\mathbf{r}^{\star T} \qquad (Eq\ 13)$$

to define the plastic deformation rate tensor, $\hat{\mathbf{d}}^{\sharp\prime}$, and the plastic spin tensor, $\hat{\mathbf{w}}^{\sharp}$, in the relaxed configuration $\hat{\mathcal{B}}$.

The viscoplastic flow rule is derived from the crystallographic slip and is defined as:

$$\hat{\mathbf{l}}^{\sharp} = \hat{\mathbf{d}}^{\sharp\prime} + \hat{\mathbf{w}}^{\sharp} = \dot{\mathbf{r}}^{\star}\mathbf{r}^{\star T} + \sum_{\alpha}\dot{\gamma}^{\alpha}(\mathbf{T}^{\alpha}) \qquad (Eq\ 14)$$

The assumed slip systems for the face-centered cubic crystals, for example, are the 12 systems with (110) directions and ⟨111⟩ normals. The symmetric and skew symmetric parts of the plastic velocity gradient, $\hat{\mathbf{d}}^{\sharp\prime}$ and $\hat{\mathbf{w}}^{\sharp}$, respectively, are defined as:

$$\hat{\mathbf{d}}^{\sharp\prime} = \sum_{\alpha}\dot{\gamma}^{\alpha}\hat{\mathbf{p}}^{\alpha} \quad \text{and} \quad \hat{\mathbf{w}}^{\sharp} = \dot{\mathbf{r}}^{\star}\mathbf{r}^{\star T} + \sum_{\alpha}\dot{\gamma}^{\alpha}\hat{\mathbf{q}}^{\alpha} \qquad (Eq\ 15)$$

where $\hat{\mathbf{q}}^{\alpha}$ is the skew part of $\hat{\boldsymbol{b}}^{\alpha} \otimes \hat{\boldsymbol{m}}^{\alpha}$.

The lattice orientation evolves as a consequence of the spin and is given by:

$$\dot{\mathbf{r}} = \mathbf{v} = \frac{1}{2}\boldsymbol{\omega} + (\boldsymbol{\omega}\cdot\mathbf{r})\mathbf{r} + \boldsymbol{\omega}\times\mathbf{r} \qquad (Eq\ 16)$$

where

$$\boldsymbol{\omega} = \text{vect}\left(\hat{\mathbf{w}}^{\sharp} - \sum_{\alpha}\dot{\gamma}^{\alpha}\hat{\mathbf{q}}^{\alpha}\right) \qquad (Eq\ 17)$$

Refinements to the Single-Crystal Constitutive Equations. The crystal constitutive equations presented here, especially those describing the plastic deformation by slip, are quite basic. They capture the fundamental observations that slip occurs on close-packed planes in close-packed direction and that the strength is elevated by strain hardening. Nevertheless, they can be improved upon in a variety of ways.

- *Competing deformation mechanisms*: Twinning and phase transformations can be incorporated by extending the kinematics to include additional deformation mechanisms that are driven by the crystal stress (Ref 9, 10).
- *Kinetics of plastic flow*: The simple power law relating the slip system shear rate to

the resolved shear stress can be generalized to include dependence on the resolved normal stress and to allow other functional forms, such as an exponential behavior, that can capture real aspects of slip beyond that possible using a simple power law (Ref 11, 12).

- *Hardening rules*: The modified form of the Voce law can be extended to include the influences of the various stages of hardening and the details of the distribution of slip. Other factors that influence the strength, such as grain or slip domain size, can be introduced (Ref 13, 14).

In contrast to extending the basic formulation, it is possible to simplify it by neglecting some aspects of the behavior. For example, by neglecting the elastic response, the kinematic decomposition can be reduced to one having slip followed by the lattice rotation. The computational requirements for integrating the constitutive equations (consistently updating the stress over an elastic-plastic deformation) are lessened substantially because the only mode of deformation is the plastic one. Such a simplification is useful for large strain deformations under monotonic loading, such as occur in many forming operations. It is not useful for cyclic loading applications.

Crystal-Scale Finite-Element Formulation. Various avenues are available to analysts for incorporating the constitutive equations for the elastic and plastic behaviors of single crystals in finite-element formulations. The avenues range from using commercial packages that accommodate custom routines for constitutive models to building special-purpose software with capabilities designed particularly for modeling polycrystals. Employing commercial codes is a popular choice because the analyst can focus attention on the material and its behavior rather than the methods needed to solve a boundary-value problem. Subroutines are publicly available that interface with well-known commercial codes (Ref 15). Developing custom software can be more expensive in many respects but offers the flexibility and control not afforded through interfaces with commercial packages. Such software often is found in research environments where new or different approaches are explored prior to their widespread acceptance (Ref 16). Another possibility is the use of niche codes that are more mature than research codes but are not intended as general purpose nor may not be publicly distributed (Ref 17). Crystal elastic-plastic models have been successfully implemented and applied within all these avenues.

Here, the salient points of one finite-element formulation that forms the basis of a custom-purpose research code are summarized. Results presented in sections that follow were obtained with this formulation. For the crystals scale, a weighted residual is formed on the equilibrium equation as:

$$R_u = - \int_{\mathcal{B}} \text{tr} \ (\sigma'^T \ \text{grad}\psi)d\mathcal{B} + \int_{\mathcal{B}} \pi \ \text{div}\psi d\mathcal{B}$$
$$+ \int_{\partial\mathcal{B}} t \cdot \psi d\Gamma + \int_{\mathcal{B}} \iota \cdot \psi d\mathcal{B}$$

where ψ are vector weighting functions, t is the traction vector, ι is a body force per unit volume, $\mathcal{B}$ is the volume of the body, and $\partial\mathcal{B}$ is its surface. The deviatoric Cauchy stress, σ', and mean stress (negative of the pressure, π) sum to the total Cauchy stress:

$$\sigma = \sigma' - \pi I \qquad \text{(Eq 18)}$$

Traction or velocity is specified over the boundary, $\partial\mathcal{B}$.

The stress is replaced ultimately with the velocity field through introduction of the constitutive equations and the kinematic relation defining the velocity gradient. The first step in this process is to introduce a difference expression for the elastic strain rate as:

$$\left\{ e^{\flat} \right\} = \frac{1}{\Delta t}\left(\left\{ e^{\flat} \right\} - \left\{ e^{\flat}_0 \right\}\right) \qquad \text{(Eq 19)}$$

This is separated into the volumetric part:

$$-\pi = \frac{\kappa \Delta t}{\beta} \text{tr}\{d\} + \frac{\kappa}{\beta}\text{tr}\left\{ e^{\flat}_0 \right\} \qquad \text{(Eq 20)}$$

and the deviatoric part:

$$\{d'\} = \frac{1}{\Delta t}\left\{ e^{\flat} \right\} + \{\hat{d}^{\sharp}\} + [\hat{w}^{\sharp}]\left\{ e^{\flat'} \right\} - \frac{1}{\Delta t}\left\{ e^{\flat'}_0 \right\}$$
$$\text{(Eq 21)}$$

After removing the volumetric portion of the elastic response in Eq 1, the resulting relation for the deviatioric response may be inverted to obtain:

$$\left\{ e^{\flat'} \right\} = [c']^{-1}\{\tau'\} \qquad \text{(Eq 22)}$$

and combining it with a relation obtained from the merger of Eq 2, 3, and 15:

$$\{\hat{d}^{\sharp}\} = [m]\{\tau'\}$$
$$\text{where } [m] = \sum_{\alpha} \left(\frac{f(\tau^{\alpha}, g)}{\tau^{\alpha}}\right)\{p^{\alpha}\}\{p^{\alpha}\}^T \qquad \text{(Eq 23)}$$

results in a matrix equation for the stress in terms of the total deformation rate:

$$\{\sigma'\} = [s](\{d'\} - \{h\}) \qquad \text{(Eq 24)}$$

where:

$$[s]^{-1} = \frac{\beta}{\Delta t}[c']^{-1} + \beta[m]$$

$$\text{and } \{h\} = [\hat{w}^{\sharp}]\left\{ e^{\flat'} \right\} - \frac{1}{\Delta t}\left\{ e^{\flat'}_0 \right\} \qquad \text{(Eq 25)}$$

Equations 20 and 24 are substituted into Eq 18 to eliminate the explicit appearance of the stress.

The trial functions are introduced for the velocity, leading to a matrix equation for the nodal point velocities at the end of a time step.

This is a nonlinear system, involving the elastic strain at the end of the time step as well as the velocity, that is solved iteratively. The lattice orientations and slip system strengths are updated over the time step by integrating Eq 16 and 4 numerically.

2.5 Virtual Sample Instantiation. All finite-element simulations of polycrystalline samples are carried out on a mesh in which the individual elements are associated with crystals. A mesh having a specified assignment of crystal attributes to particular finite elements is referred to as a virtual sample. It represents one of many possible ways that crystals may be arranged within a volume of the material. The methods to create virtual samples range from quite simple to very complex, depending on which of the microstructural attributes, and to what degree, are being replicated. Perhaps the most simplistic is to define a mesh of regularly shaped elements and to each of these assign a crystal lattice orientation randomly from a uniform distribution. From there, one can build virtual polycrystals that include aspects of the grain size and shape distributions, lattice orientations that are drawn from nonuniform distributions (textures), spatial correlations of lattice orientation, and phase topologies, including anisotropic distributions of the continuity. Commonly employed methods use tesselations, both regular and irregular, to define the grain geometry and direct mapping of three-dimensional images obtained from experimental characterization (for example, by serial sectioning).

2.6 Postprocessing for Comparison to Experiments. Comparison to experiment is critical to gaining confidence in the fidelity of simulation. This is true both when new methods are under development and when new applications emerge. Experiments involving in situ loading of specimens in facilities capable of x-ray and neutron diffraction measurements provide data at the scale of crystals. Elastic (lattice) strains can be inferred from shifts in diffraction peaks under changes in the local stress state. Simulation results can be analyzed to extract data equivalent to that obtained from diffraction peaks. Besides comparisons of lattice strains, one also can examine texture evolution from changes in the peak intensity and strain hardening from changes in the peak width.

The basic steps of this postprocessing operation are laid out in Fig. 4. The first step in the postprocessing operation is to determine which subset of the orientations should be considered active when simulating a particular diffraction peak. Any element whose orientation (accounting for crystal symmetries) is such that the designated crystal plane normal lies within a solid angle, Ω, of the scattering vector is considered to be active in contributing to the simulated diffraction peak. Experimentally, only those crystals that align with the scattering vector to a high degree of resolution will contribute to the peak. Even with tens of thousands of elements, the simulations cannot produce the degree of resolution comparable to the experiment, so typically, a larger resolution angle, Ω,

Finite element virtual specimen with dodecahedra crystals in the interior

Average values of the normal component of the lattice strain of all crystals lying on the indicated fiber with increasing levels of applied stress

Crystals within the virtual specimen satisfying a specified Bragg condition

Location of crystal lattice orientations lying on a fiber within the fundamental region

Fig. 4 Postprocessing finite-element simulation data for comparison to diffraction data

is specified in the simulations than is set by the slit dimensions of the diffractometer. The resolution angle should be as small as possible to maximize the fidelity to the experiment but at the same time large enough to provide meaningful statistics. The lattice strains, given by the elastic strain tensor, are averaged over that set of crystals. The component of the average strain tensor in the direction of the specified plane normal (scattering vector) is then available for comparison to experiment. Further, changes in the numbers of elements found to align with the scattering vector associate with changes of the peak intensity; similarly, changes in the average slip system strength for those elements may be compared with the evolving peak width.

Example 1: Elastic Strains within Grains. The simulation results for the response of a stainless alloy are compared to experimental data taken by neutron diffraction. Figure 5 shows the tensile specimens, undeformed and deformed, used in the in situ loading experiments. Only the central portion of the gage section of the specimen was discretized, using approximately 56,000 elements: Each finite element of the mesh is an eight-noded brick and is treated as an individual crystal and so is assigned initial values for its own orientation, slip system strength, and elastic strain tensor. All of these state descriptors are free to evolve during the course of the simulation. The crystal lattice orientations were initialized by sampling the ODF for the specimen material. The initial value of slip system strength was taken to be constant across all elements, and the initial value of the elastic stretch tensor was taken as the identity (corresponding to zero initial lattice strain). With this level of resolution, the simulations capture stress variations between crystals. However, because the order of the elemental interpolation is low (trilinear velocity field

within crystals), stress fluctuations around dislocations are not modeled explicitly. A full description of the experiments and simulations is given in Ref 18 together with values of the constitutive model parameters.

These samples were loaded in tension to the point of strain localization, with several unloading/reloading episodes being performed at fairly regular intervals. Although the macroscopic loading condition is simple tension, the responses of the individual crystals vary from the nominal stress due to the anisotropy in the elastic and plastic behaviors. The goal of the simulations is to demonstrate the ability of the crystal-based formulation to capture the variations in stress at the scale of individual crystals.

Also shown in Fig. 5 is the finite-element mesh of the gage section of the specimen together with several subsets of the crystals in the sample. Here, each element is a distinct crystal and is colored according to the axial component of its elastic strain. As is evident, the more compliant the crystal direction, the greater the strain, on average. These subsets of crystals correspond to different Bragg conditions for which there exists lattice strain measurements. The comparison is shown in Fig. 6 for the most (200) and least compliant (222) sets of crystals while the specimen remains under load and after it has been unloaded. Strains under load increase with the overall strain of the sample as a consequence of strain hardening. Recall that strain hardening is incorporated via evolution of the slip system strength. The ordinate scale of the two plots differs by a factor of 2, showing the pronounced difference in elastic strains among crystals with compliant or stiff orientations relative to the loading axis. After unloading, the crystals do not return to an unstrained state but rather exhibit significant residual strains. On average, these are positive (extension) for the

(200) set of crystals and negative (compression) for the (222) set of crystals. Nonzero elastic strains indicate the presence of residual stresses

(a)

(b)

(221) (311) (200) (220)
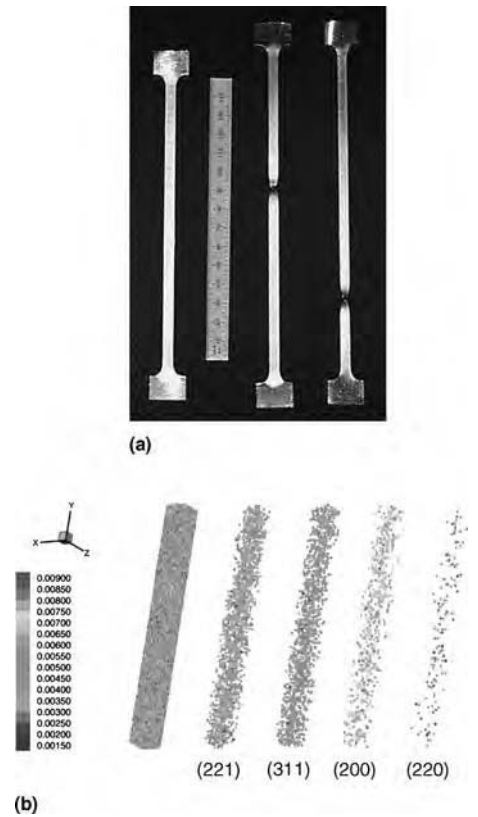
Fig. 5 Tensile specimen loaded to the point of failure. (a) Photographs of the samples in the undeformed and deformed conditions. (b) Finite-element mesh of the gage section of the mesh showing axial strains. One insert shows the full mesh, while the others show only those crystals satisfying the indicated Bragg condition for an axial scattering vector.

(a) LIVE GRAPH Click here to view

(b) LIVE GRAPH Click here to view
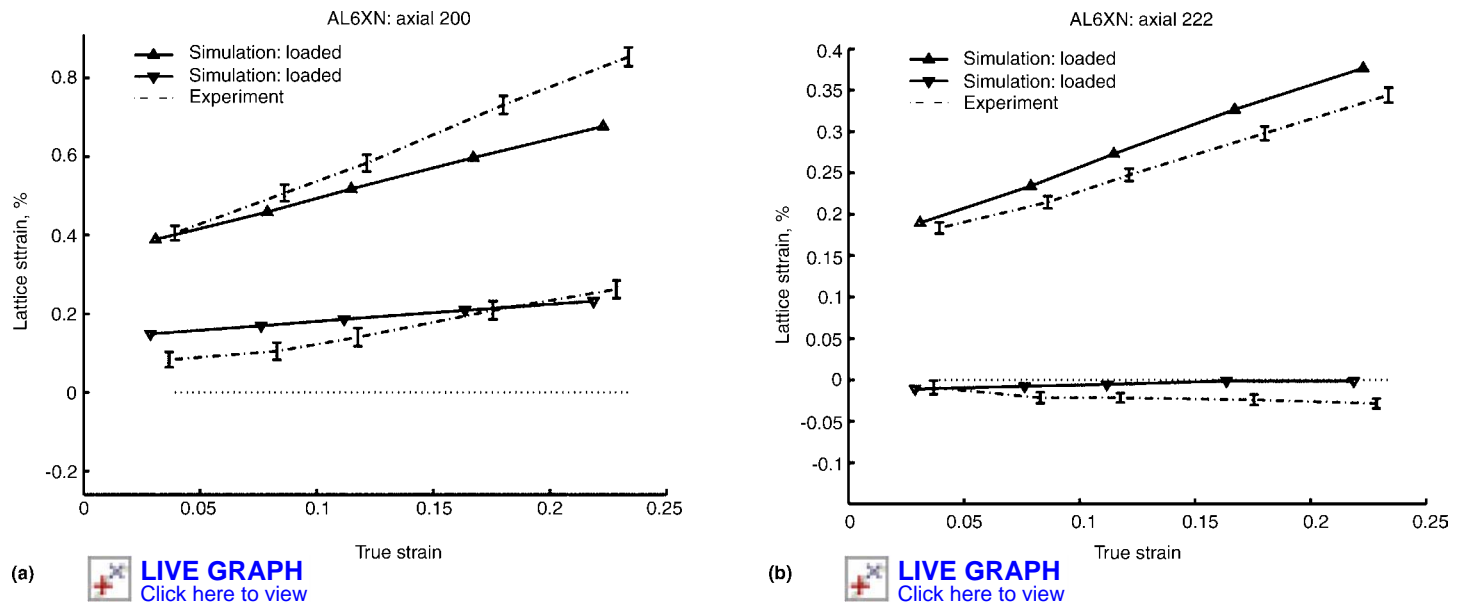
Fig. 6 Comparison of the computed and measured axial lattice strains. (a) 200 crystals. (b) 222 crystals. Experimental points are shown with estimated error bars.

at the crystal scale. Thus, even for the case of simple tension, the heterogeneity of deformation at the crystal scale induces residual stresses following plasticity. The simulations capture these trends well.

In addition to the positions of Bragg peaks changing under load, the peak profiles evolve with deformation. A connection between the broadening of the Bragg peaks and the evolution of the slip system strength is anticipated because both the peak width and the strength can be related to the dislocation density. Specifically, both are proportional to the square root of the dislocation density, so that a direct dependence is expected. For the stainless steel experiments discussed here, a correlation between the experimental peak width and the simulated strain hardening has been drawn (Ref 19). In computing the correlation, the contribution to peak broadening from variations in crystal lattice strain is subtracted from the total increase in peak width (lattice strains within a grain give rise to a peak shift; when there is a spread in lattice strains over the grains within a diffraction volume, the net effect of the strain variations is an apparent broadening of the peak). The difference is attributed to the dislocation density increase and correlated with it, using:

$$f(\theta)\beta_d = b_1 + b_2 g^* \qquad (Eq\ 26)$$

where β_d is the dislocation-based increase in the peak width, and g^* is the normalized slip system strength (computed as the ratio of the average slip system strength in the diffracting crystals to the initial slip system strength). The plot of the dislocation-based broadening, β_d, after scaling by the cosine of the Bragg angle to account for diffraction geometry, versus the normalized slip system strength is shown in Fig. 7. The correlation is good beyond the initial stages of hardening once the normalized slip system strength, g^*, exceeds approximately 1.2.

Example 2: Bulk Texture Evolution. The evolution of texture during processing has received considerable attention from modelers in the past. One application of particular interest is the texture evolution that accompanies the rolling of face-centered cubic metals. The rolling textures may be studied by comparing the relative strengths of ideal texture components, which, in terms of the ODF, is the value of $A(\mathbf{r})$ at specified values of $\mathbf{r}$. The copper-type texture is comprised principally of the copper, brass, and S components. The brass-type texture is dominated by the brass and Goss components. The positions of these ideal components in the Rodrigues fundamental region (Ref 20) are shown in Fig. 8. Finite-element formulations at the crystal scale facilitate the study of the influence of the deformation heterogeneity on the predicted textures. These predictions can be compared and contrasted to those from alternative approaches, such as applying mean field assumptions of isostrain or isostress conditions.

Here, the predicted textures obtained from two different slip system hardening assumptions, isotropic and latent hardening, are compared; a more complete description of the results is given in Ref 21. Isotropic hardening is given by the relatively simple hardening model shown in Eq 4. For anisotropic hardening, an interplanar latent hardening law elevates the slip system strength as a consequence of plastic straining. This hardening model causes slip on one plane to strengthen systems on other planes more than it strengthens systems sharing that plane. The model implemented conforms to a simple Voce (Ref 22) hardening format:

$$\dot{g}^\alpha = h_o \frac{(g_s - g^\alpha)}{(g_s - g_o)} M_{\alpha\beta}\dot{\gamma}^\beta \qquad (Eq\ 27)$$

where g^α, g_s, g_o, h_o, and $M_{\alpha\beta}$ are hardening parameters. For the complete set of slip systems, $M_{\alpha\beta}$ constitutes a matrix of coefficients.

$M_{\alpha\beta} = a$ when the α slip system shares the same slip plane with the β slip system. Likewise, $M_{\alpha\beta} = b$ when the α slip system does not share a slip plane with the β slip system. The latent hardening ratio (LHR) is then defined as b/a. If the LHR is greater than unity, slip on a given slip plane hardens slip systems on other planes at a higher rate than it hardens itself; if the LHR is unity, the isotropic hardening assumption is recovered. Latent hardening rules influence the morphology of the single-crystal yield surface. The changing yield surface in turn affects slip system activation and lattice reorientation velocities in the polycrystal, thus influencing texture evolution. Experimental data have shown that LHRs can be as high as 2.0 but more often are approximately 1.4 (Ref 4).

Simulations were performed at three LHRs: 1.0, 1.4, and 1.8. To obtain the single-crystal hardening parameters appearing in Eq 27, the
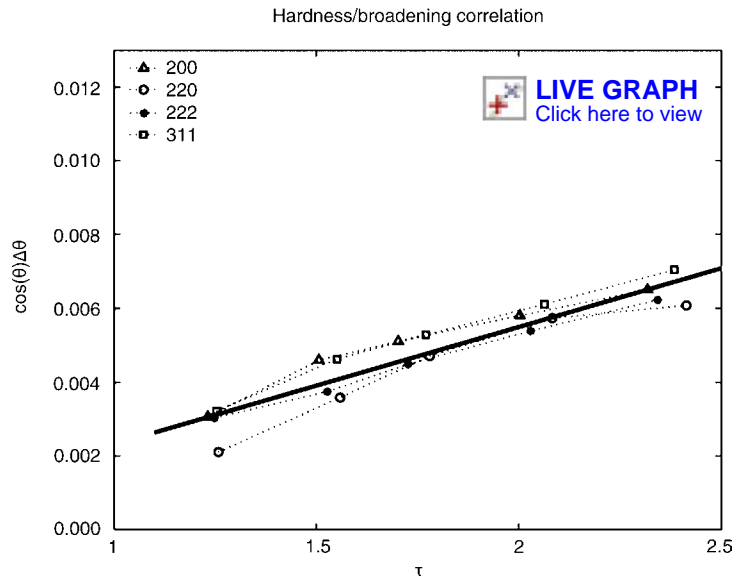
Fig. 7 Correlation of the normalized slip system strength, τ, with evolution of peak width for crystals having various crystallographic directions aligned with the specimen transverse direction

Component	r1	r2	r3
Brass	$(\sqrt{2}-1)(\sqrt{3}-\sqrt{2})$	$\sqrt{2}-1$	$\sqrt{3}-\sqrt{2}$
Copper	$\sqrt{3}-\sqrt{2}$	$(\sqrt{2}-1)(\sqrt{3}-\sqrt{2})$	$\sqrt{2}-1$
Goss	0	$\sqrt{2}-1$	0
S	0.2346	–0.2535	–0.2906
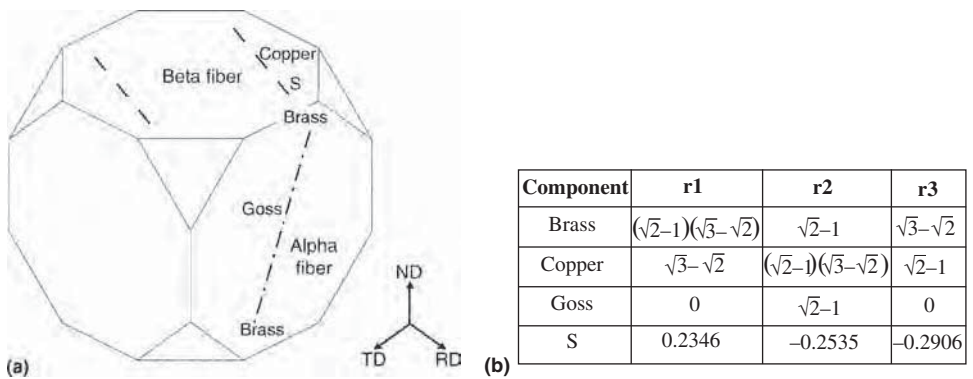
Fig. 8 (a) Location of face-centered cubic rolling components in the Rodrigues fundamental region. (b) Coordinates of rolling components in Rodrigues space. The existence of orthorhombic sample symmetry results in multiple ideal component locations in orientation space. Note: <r1,r2,r3> directions correspond to the <TD,RD,ND> (transverse, rolling, normal) directions of a rolling process.

macroscopic response of each LHR was matched to an experimental stress-strain curve for brass subjected to tensile loading. Matching the macroscopic response was important for two reasons. First, it was important to be sure that the samples exhibited realistic mechanical responses. Second, for the effects of latent hardening to be studied, the specimens needed to be under similar stress levels. Because LHRs affect the net strengthing of the crystal, each LHR must have different single-crystal hardening parameters to match the given stress-strain curve. Otherwise, the texture differences may have been attributed to the differences in macroscopic stress state. Greater detail regarding the choice of hardening parameters is provided in Ref 21. The elastic moduli are those of pure copper.

Simulations were performed using a virtual polycrystal constructed of an arrangement of space-filling dodecahedra, with each representing a single grain, as shown in Fig. 9. Each dodecahedral grain is made up of 48 higher-order, tetrahedral elements. Discretization at the subgrain level enables spatial variation of the deformation within the grain. The virtual specimens shown in Fig. 9 contain 81,000 tetrahedral elements and 1098 full dodecahedral grains. It should be noted that there are a number of partial grains on the surface that are required to create the full cube shown in Fig. 9(a). Lattice orientations were assigned to all the elements in the mesh from a random sample drawn from a uniform (untextured) orientation distribution. All the elements within a crystal were assigned the same initial orientation, although those orientations could evolve independently thereafter.

The virtual specimens were placed under plane-strain compression, which can be considered an idealized rolling operation. In an actual rolling operation, frictional constraints applied by the rolls as well as the geometric constraint of the workpiece itself prevent deformation in the transverse direction (TD, parallel to the rolls). The deformation is then limited to the rolling direction (RD, the extension direction) and the normal direction (ND, the compression direction). This results in plane strain at the macroscopic level, although deformation can occur in TD within the polycrystal. Initially,

each virtual sample was deformed to a target effective strain (typically, 30%) at a strain rate of $0.001\,\mathrm{s}^{-1}$. Effective strain is defined as $\bar{\varepsilon} = \int \bar{d}\,dt$, where $\bar{d}$ is proportional to the second invariant of the symmetric part of the velocity gradient:

$$\bar{d} = \sqrt{\frac{2}{3}\mathbf{d} : \mathbf{d}}$$

Orientation distribution functions were determined and plotted in Rodrigues orientation space for three different effective strains: 10, 20, and 30% (Fig. 10a,b, and c, respectively). Qualitatively similar textures result for all three LHRs. At all of the strain levels, the development of a standard rolling texture is evident, with regions of high orientation density developing along the α and β fibers. By an effective strain of 10%, the crystals are beginning to align themselves along these fibers. At a strain of 20%, a rolling texture pattern has emerged; at a strain of 30%, the rolling texture is well established.

The effects of latent hardening on texture development at low strains are too subtle to be seen from a simple visual comparison of the resultant ODFs. Difference orientation distribution functions (DODEs) allow the ODFs to be compared in greater detail. A DODF is simply a subtraction of one ODF from another:

$$\delta A(\mathbf{r})_{12} = A(\mathbf{r})_1 - A(\mathbf{r})_2 \qquad \text{(Eq 28)}$$

A peak in a DODF indicates that the likelihood of there being crystal lattices corresponding to the peak orientation is higher in the material state represented by $A(\mathbf{r})_1$ than in the material state represented by $A(\mathbf{r})_2$ (Eq 28). Figure 11 shows the DODF between a LHR of 1.8 and a LHR of 1.0 at a strain level of 30% for the finite-element simulations (Fig. 11a) and for simulations based on an isostrain assumption (Fig. 11b). In the finite-element simulations, the higher LHR mainly leads to an increase in strength along the α-fiber but does not correspond to an equivalent increase along the β-fiber. This indicates that latent hardening is affecting the direction of reorientation. The crystals are migrating to different regions of orientation space rather than simply arriving in

the same location at a faster rate. The increase in the strength of the α-fiber, without a corresponding increase in the β-fiber, indicates that a brass texture is forming. In contrast, the isostrain assumption gives opposite trends to those produced when heterogeneous deformations are allowed within the finite-element framework. The finite-element virtual polycrystal captures the combined effects of latent

(a)

(b)
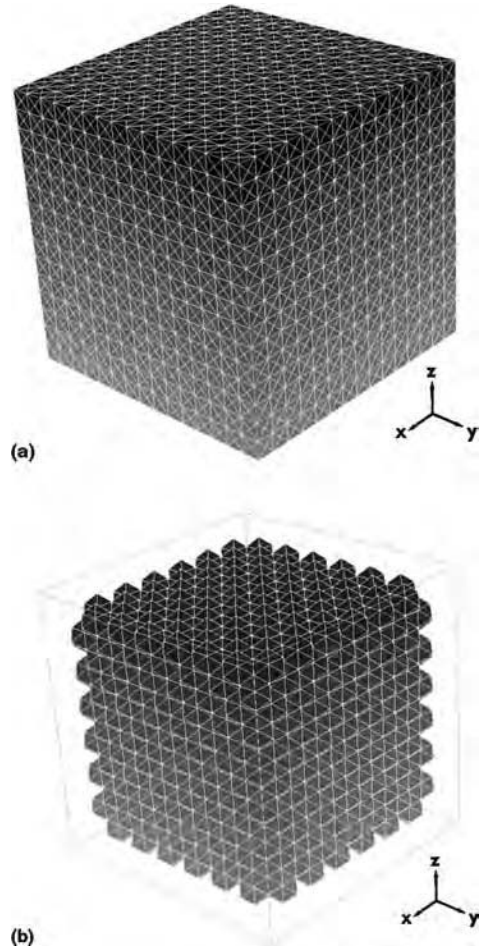
Fig. 9 Virtual specimen. (a) Full mesh discretization with 81,000 elements. (b) Mesh showing 1098 full dodecahedral grains. Crystals are colored according to position to aid in visualization.
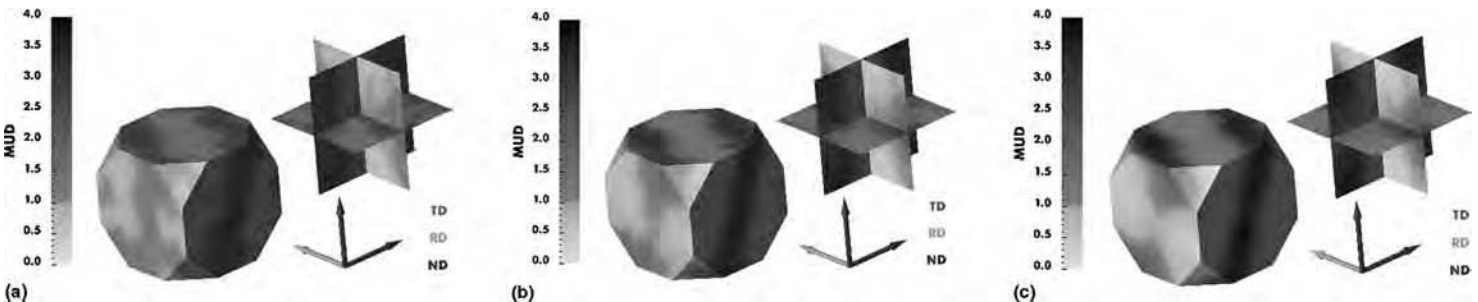
Fig. 10 Resultant orientation distribution functions from sample 1 with a latent hardening rate of 1.4 at effective strains of (a) 10% (b) 20% and (c) 30% after 5° Gaussian smoothing. MUD, multiples of uniform distribution
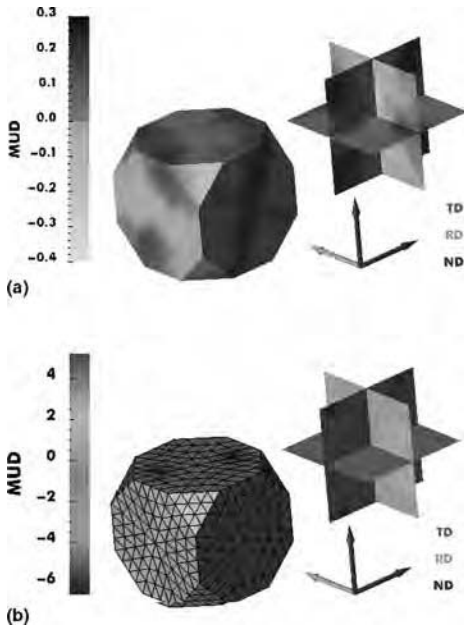
Fig. 11 Difference orientation distribution function for a latent hardening rate (LHR) of 1.8 minus LHR 1.0 for (a) finite-element virtual polycrystal with 5° Gaussian smoothing and (b) Taylor assumption. MUD, multiples of uniform distribution

hardening and heterogeneous deformation over the polycrystal on the formation of brass textures.

Application to the Continuum Scale

Engineering components typically are much larger in volume than can be modeled using methods that resolve every grain. Yet, the plastic deformation mechanisms are ones of slip and twinning that are modeled well with constitutive equations rooted at the crystal scale. It is possible to devise methodologies that retain benefits of a crystal-scale theory in a formulation appropriate for continuum-scale simulations by making assumptions that link the motion and stress between the two scales. In this section, one such methodology is presented in which the scales are linked by assuming that every crystal in a volume associated with a continuum point experiences the same deformation gradient as the continuum point. Other assumptions are possible, however, although space limits their review.

Polycrystal Behavior. For the continuum-scale model, the material properties derive from orientational averages over a representative volume of single crystals associated with continuum points throughout the domain. Continuum-scale quantities of stress ($\boldsymbol{\Sigma}$) and velocity gradient ($\mathbf{L}$) are defined for a volume, $\mathcal{B}$, containing a population of crystals by integrating tractions, t, and the velocities, $\boldsymbol{u}$, over the volume surface, $\partial\mathcal{B}$. For the stress:

$$\boldsymbol{\Sigma} = \frac{1}{\mathcal{B}} \int_{\partial\mathcal{B}} \mathbf{x} \otimes t \, d\Gamma \qquad \text{(Eq 29)}$$

and for the velocity gradient:

$$\mathbf{L} = \frac{1}{\mathcal{B}} \int_{\partial\mathcal{B}} \boldsymbol{n} \otimes \boldsymbol{u} \, d\Gamma \qquad \text{(Eq 30)}$$

where $\boldsymbol{x}$ is the position, and $\boldsymbol{n}$ is the surface normal. For sufficiently smooth distributions of stress and velocity gradient, these surface integrals can be converted to integrals over the volume. Again, for the stress:

$$\boldsymbol{\Sigma} = \langle \boldsymbol{\sigma} \rangle = \frac{1}{\mathcal{B}} \int_{\mathcal{B}} \boldsymbol{\sigma} \, d\mathcal{B} \qquad \text{(Eq 31)}$$

and for the velocity gradient:

$$\mathbf{L} = \langle \mathbf{l} \rangle = \frac{1}{\mathcal{B}} \int_{\mathcal{B}} \ell \, d\mathcal{B} \qquad \text{(Eq 32)}$$

Using the ODF, a change of variables can be introduced to map these integrals to the fundamental region. For any of the fields, including the stress and velocity gradient, an average is given by:

$$\Phi = \langle \phi \rangle = \int_{\Omega_{\text{fr}}} A(\boldsymbol{r}) \phi(\boldsymbol{r}) \, d\Omega \qquad \text{(Eq 33)}$$

In this way, the crystal-scale quantities link to their continuum-scale counterparts, as shown schematically in Fig. 1b. Applying such averaging to constitutive equations from the crystal scale provides formulae for the mechanical properties.

Finite-Element Formulation for Continuum Scales. As pointed out in the section "Crystal-Scale Finite-Element Formulation" in this article, several avenues are possible for implementing particular constitutive equations in finite-element formulations. The same comments apply to continuum-scale constitutive models based on averaged behaviors of polycrystals. Again, the formulation used in a custom-purpose research code is presented because it illustrates the main features needed in any of the approaches. Note that the formulations for the crystal-scale and continuum-scale responses are similar, except that the continuum formulation uses averaged quantities associated with the motion, stress, and properties. Similarly, for the continuum scale, the weak form of equilibrium is written as:

$$R_u = -\int_{\mathcal{B}} \text{tr} \left(\boldsymbol{\Sigma}'^{\mathrm{T}} \text{ grad} \boldsymbol{\Psi} \right) d\mathcal{B} + \int_{\mathcal{B}} \Pi \text{ div} \boldsymbol{\Psi} \, d\mathcal{B}$$
$$+ \int_{\partial\mathcal{B}} \boldsymbol{T} \cdot \boldsymbol{\Psi} d\Gamma + \int_{\mathcal{B}} \mathcal{J} \cdot \boldsymbol{\Psi} \, d\mathcal{B}$$

where $\boldsymbol{\Psi}$ are vector weighting functions, $\boldsymbol{T}$ is the traction vector, $\mathcal{J}$ is a body force per unit volume, $\mathcal{B}$ is the volume of the body, and $\partial\mathcal{B}$ is its surface. The continuum Cauchy stress has been decomposed into a pressure, Π, and a deviatoric part, $\boldsymbol{\Sigma}'$:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}' - \Pi\mathbf{I} \qquad \text{(Eq 34)}$$

A matrix equation for the deviatoric stress in terms of the deformation rate, now for the continuum quantities, results in the same fashion as for the crystal scale. This gives:

$$\{\Sigma'\} = [S](\{D'\} - \{H\}) \qquad \text{(Eq 35)}$$

To evaluate the stiffness, [S], from the corresponding crystal quantities, the continuum velocity gradient is partitioned among the crystals in the aggregate. Such an assumption is known as a linking or mean field hypothesis. An isostrain assumption is one possibility that equates the deformation field of the individual crystals with that of the polycrystal (Ref 23). Enforcing this condition ensures that intergranular compatibility and global equilibrium is satisfied but not local equilibrium (Ref 24–26). The specific condition used is $\mathbf{l} = \mathbf{L}$, giving:

$$[S] = \langle [s] \rangle \{H\} = \langle [s]\{h\} \rangle \qquad \text{(Eq 36)}$$

From this point, the crystal-scale and continuum-scale formulations are quite similar. The numerical solution gives the nodal velocities as a function of time, together with the history of the elastic strain, lattice orientations, and slip system strengths. For the continuum formulation, however, there is an aggregate of crystals at each continuum point where properties are evaluated with an associated distribution of lattice orientations and the single-crystal properties.

Alternatives within the Continuum-Scale Formulation. The continuum-scale formulation outlined in this article possesses the necessary basic elements to simulate the response of a material that displays anisotropy in its strength due to the influence of crystallographic texture. That anisotropy may evolve in concert with evolution of texture with deformation. Alternatives to the particular choices made for many aspects of the formulation do exist, of course. A few of those are listed as follows:

- In lieu of performing averages over an aggregate of crystals to obtain the plastic stiffness in Eq 35, one could employ a yield surface: a surface in stress space that defines the stress necessary to induce plastic flow. Functions that define yield surfaces range from extensions of the isotropic model proposed by von Mises to piecewise polynomial representations that employ parametric mappings of elementary patches to form a complete surface (Ref 27).

- The representation of the state and its evolution can be tackled in a variety of ways. The state descriptors (texture and slip system strength, in this case) can be defined locally at quadrature points within the finite elements or as entire fields over the body or workpiece. If the definition is strictly local, then the evolution equations for the state descriptors at each quadrature can be integrated independently of those at other quadrature points. In contrast, if the representation of state is done using fields defined

over the full body, then global methods may be needed. Galerkin-weighted residual approaches work well in this regard.

- As with the crystal-scale formulation, it is possible to simplify the formulation by neglecting the elastic response, thereby gaining substantial computational advantage. Such a simplification may be valid under loadings in which the straining is dominated throughout the application by the plastic response. Issues of texture evolution and shape changes can be addressed effectively in many large-deformation forming operations under this simplication.

Example 3: Residual Stresses from Bending. The continuum-scale formulation is illustrated with the calculation of stress distributions in a stainless steel component (the same material, AL6XN, discussed in example 1.) The component has a "U" shape with tabs for the application of opposing loads, applied horizontally in the picture (Fig. 12). Under this loading, the component acts as a curved beam, developing macroscopic gradients in the stress and strain distributions when its ends are pulled in opposite directions. The state of loading resembles that of pure bending but also includes an
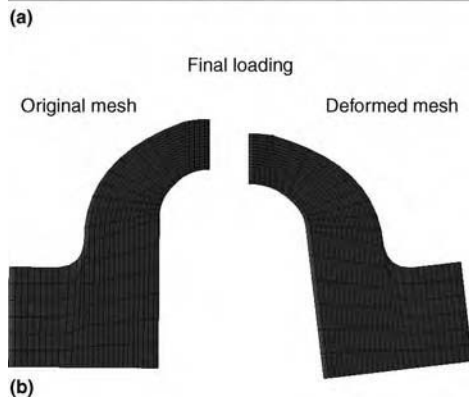
axial component that arises from loading the specimen with a set of opposing forces. The finite-element mesh for a symmetric half of the specimen at the macroscopic scale is shown in Fig. 12. This mesh contains a total of 1548 elements, with each element having assigned a total of 128 crystal orientations to define the texture within it. The initial values for the orientations were defined from a random sampling of a measured ODF. Again, there exist measurements of crystal lattice strains obtained by neutron diffraction to serve as a basis of comparison.

The measured loading sequence, shown in Fig. 13, was applied to the macroscopic mesh. The complete sequence included a total of 7 unloading/reloading episodes, as well as a final unloading sequence. This history was sufficient to produce plastic deformations along the centerplane where the moment is highest, imparting a permanent shape change to the specimen, as shown in Fig. 12. The macroscopic stress

distributions in the loaded and unloaded conditions of the final unloading episode are shown in Fig. 14.

Elements in the macroscale mesh are comparable to, but somewhat smaller than, the experimental diffraction volumes. From the macroscopic scale results, loading histories for elements along the centerplane were extracted, as shown in Fig. 15. Specifically, the velocity history of the surface of a macroscopic element was imposed on the corresponding surface of the crystal-scale mesh, thereby forcing the boundary of the polycrystal to move in the same way as the boundary of the macroscopic element. Figure 16 shows one of the crystal-scale meshes. It contains 1000 elements with each element corresponding to an individual crystal; the complete mesh, or polycrystal, coincides with one of the elements along the centerplane of the macroscopic mesh. Figure 15 shows the history of 15 elements adjacent to the centerplane. These span the same cross section as
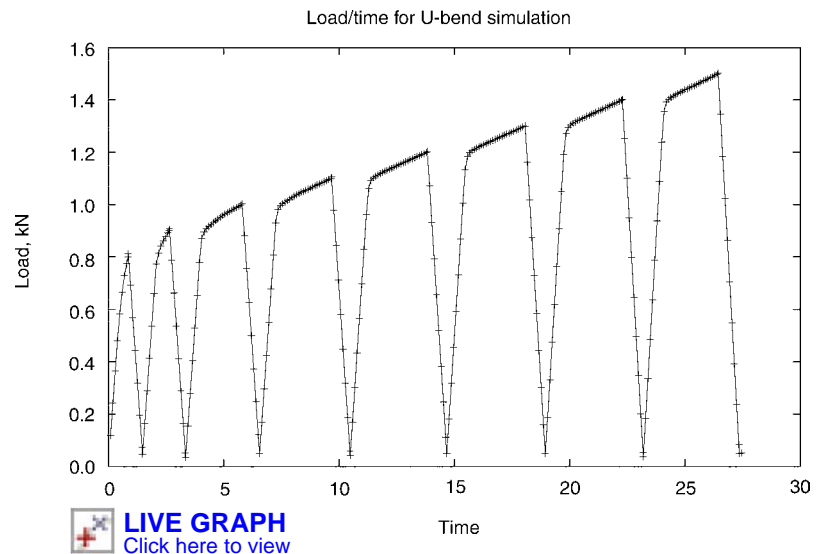
Fig. 13 Load history on bending specimen. Scale is in seconds; hold times for diffraction measurements have been omitted.

Fig. 12 (a) Photograph of undeformed and deformed component. (b) Undeformed and deformed meshes for the component

Fig. 14 Normal stress components at the initiation (left) and conclusion (right) of the final unloading episode
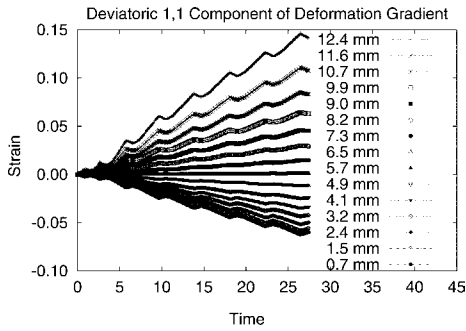
Fig. 15 Deformation gradient histories extracted from the continuum-scale simulation and applied to the crystal-scale meshes corresponding to diffraction volumes along the specimen centerplane. Scale is in seconds; hold times for diffraction measurements have been omitted.
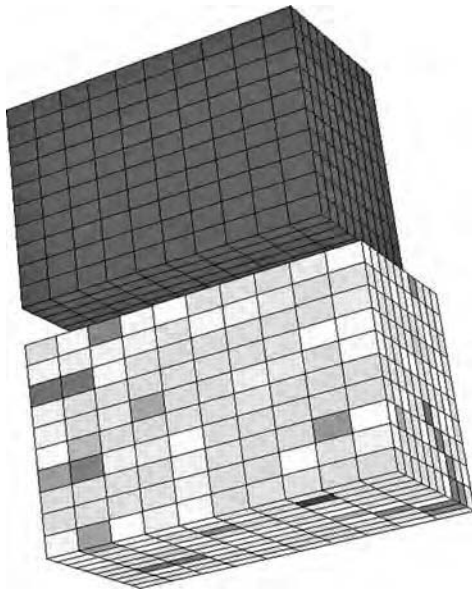
Fig. 16 Diffraction volume (microscopic) finite-element mesh for the bending specimen simulations. Upper image shows the undeformed mesh; the lower image shows the final mesh, where color indicates relative magnitude of the *xx* component of stress.

Fig. 17 Comparisons of measured and computed lattice strains for crystals having a 004 direction aligned with the specimen *x*-direction. Loaded and unloaded states are shown.
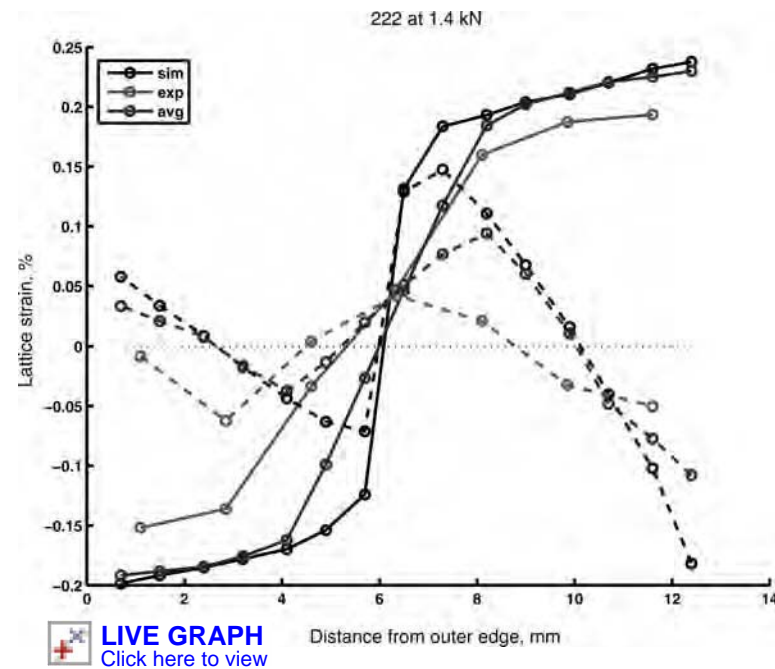
Fig. 18 Comparisons of measured and computed lattice strains for crystals having a 222 direction aligned with the specimen *x*-direction. Loaded and unloaded states are shown.

examined experimentally, although the centroids of the simulation and experimental diffraction volumes do not exactly coincide.

Comparisons between measured and computed lattice strains are shown for the final complete unloading episode. For both the loaded and unloaded states, lattice strains for 004 crystals are shown in Fig. 17 and strains for 222 crystals are shown in Fig. 18. One can readily see the bending profile expected for a plastic hinge of a strain-hardening material in the loaded state and the subsequent residual profile for the unloaded state. The computed strains are somewhat higher, but the difference between the loaded and unloaded states compares well between experiment and simulation. The 004 strains are considerable larger (factor

of 2) than the 222 strains, a trend that is captured by the simulation. Probably the most notable difference between simulation and experiment is the sharpness of the computed profile near the neutral axis in the unloaded state. This trend is much more muted in the experiment. The size of the diffraction volume may influence this difference. By averaging lattice strains in adjacent elements, the results correspond to enlarging the diffraction volumes in

the simulations. As expected, these compare more closely to experiments.

Example 4: Deformation of a Welded Joint. Another example at the continuum-scale formulation is the simulation of the deformation of a tensile specimen having a weld in the center of its gage section. The simulation mimics one case in a set of physical experiments that were conducted on specimens cut from a plate having a double-pass friction stir weld. This example,

while relatively simple in geometry, illustrates several aspects of modeling at this scale. One aspect is the initialization of the mechanical state of the material. Another is the impact of anisotropic properties on the deformation. A third aspect is how the deformation induces evolution of the state, which, for polycrystalline materials, includes the evolution of the texture. Changing texture implies evolving mechanical properties, especially in the strength anisotropy. This example illustrates how simulations can loosely couple the continuum and crystal scales.

In the experiments, several samples were cut from the welded plate with their tensile axes aligned with the plate TD. The welding process leaves a strong structural heterogeneity in the material, as can be seen in Fig. 19. Each sample exhibited a unique cross section, as can be seen in Fig. 20. Discussion of the motivation behind the experimental design, as well as the simulation of all of the specimens, can be found in Ref 28. Results are presented only for one of the samples, which had an initial cross section of 3.94 by 3.34 mm.

Electronic backscatter diffraction maps were completed in both the weld zone and the base metal for the purpose of initializing the texture. These are shown in Fig. 21 for both the weld zone and the base metal in terms of a Rodrigues parameterization of orientation. The base metal exhibits no strong texture components and exhibits a peak value only slightly higher than the value for a uniform distribution. In the weld zone, the material exhibits a diffuse cube texture. Grain sizes generally are much smaller in the weld zone than in the base metal. Microhardness measurements were taken over the weld zone and regions of the base metal adjoining it. These indicated that the mixing associated with the welding process substantially hardened the material in the weld zone.

Initializing the state of the material involves specifying the initial texture and slip system strength for the continuum-scale formulation discussed here. The welding process leaves a weld zone that is distinct from the base metal but relatively uniform within itself. The transition from the weld zone to the base metal is quite sharp. Thus, for the simulation, two zones are considered, base metal and weld zone, each with uniform properties. The initial textures in the weld zone and base metal were specified by the measured ODFs. The initial slip system strength distribution was qualitatively estimated from the microhardness maps.

The multiscale methodology used in the simulations is illustrated in Fig. 22. The figure shows the simulated specimen before deformation on the left and after deformation on the right. Certain elements were selected from the base metal and the weld zone, and the volume contained in each of these elements was tracked over the entire continuum-scale simulation. Then, each of these elements was discretized with 8000 elements, with each element representing an individual crystal to define the crystal-scale model. Additional simulations were performed on each
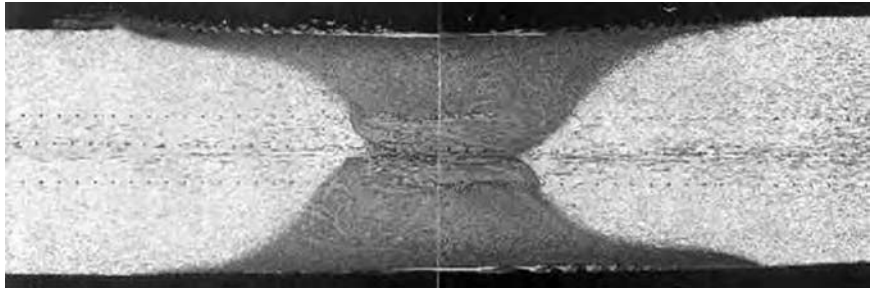
Fig. 19 Micrograph showing base metal and stirred regions within the friction-stir-welded microstructure
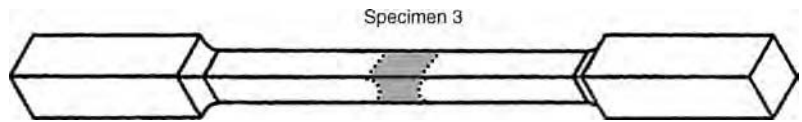
Fig. 20 Initial geometry for the tensile specimen cut from the welded plate. Shaded region shows approximate location of stirred zone depicted in Fig. 19.
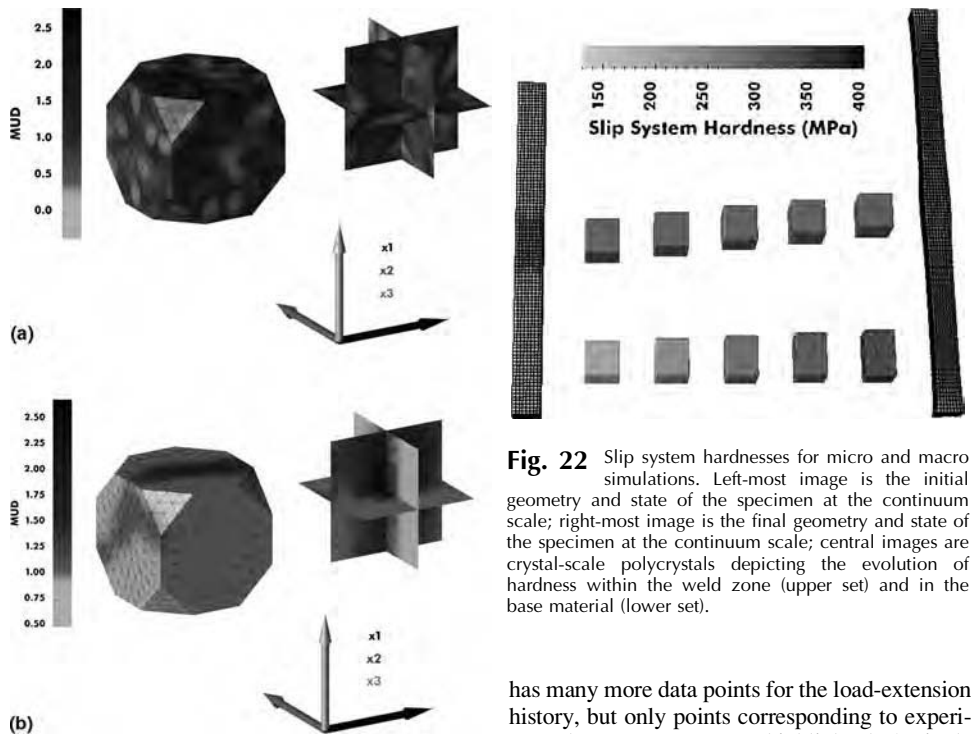
Fig. 21 Orientation distribution functions shown on the surface and coordinate planes in Rodrigues space. Scale is in multiples of the uniform distribution (MUD). (a) Base metal. (b) Weld zone

Fig. 22 Slip system hardnesses for micro and macro simulations. Left-most image is the initial geometry and state of the specimen at the continuum scale; right-most image is the final geometry and state of the specimen at the continuum scale; central images are crystal-scale polycrystals depicting the evolution of hardness within the weld zone (upper set) and in the base material (lower set).

crystal-scale model. The motion of the continuum-scale element provided boundary conditions for the crystal-scale model. The figure shows the element shape and slip system hardnesses from the initial condition through the four peak load levels for the crystal-scale specimens.

The load-extension record is presented in Fig. 23. The experimental record consists of the data points connected by dashed lines indicating the sequence of measurements. The simulation

has many more data points for the load-extension history, but only points corresponding to experimental measurements are highlighted. A single set of material parameters was chosen to match the experimental extension records at the specified loads, effectively using this macroscopic experimental data to calibrate the work-hardening parameters of the model.

Comparisons of the final shapes of each specimen are presented in Fig. 24, showing contour plots of specimen thickness and line plots of in-plane edge profile. Due to physical limitations of the coordinate measuring system used, the measured thickness profiles do not extend to the edge and are shown as the top rectangles in the contour plots. The simulated data are presented directly below and show the entire specimen. The line plots contain both experimental
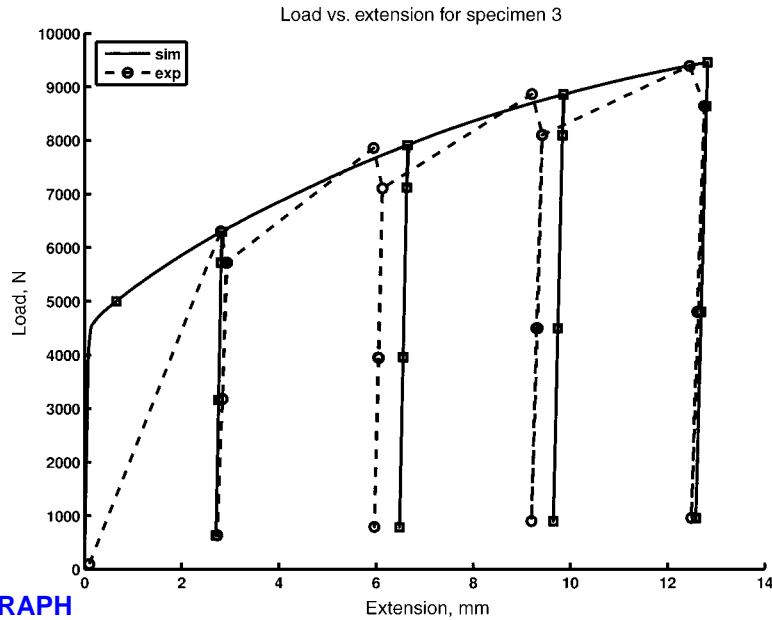
Fig. 23 Load-extension record for sample 3 showing the measured and computed loads. The experimental loads do not follow the dotted lines; the lines are provided only to aid the reader in following the sequence of load measurements.
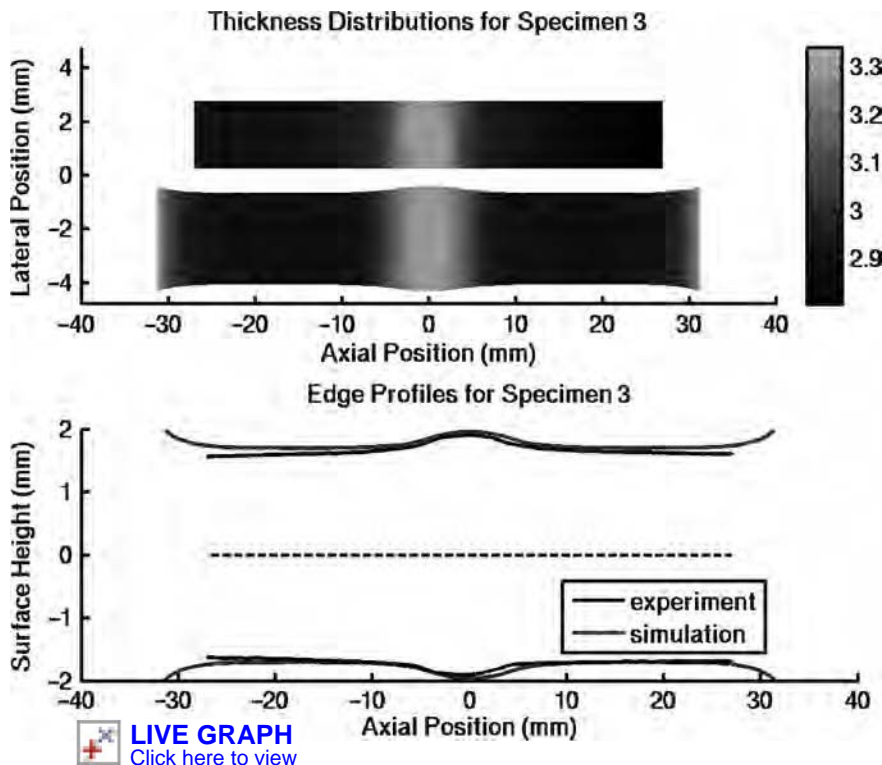
Fig. 24 Comparisons of simulated and experimental deformed profile. The upper images show contours of thickness for experiment and simulation over the full specimen (the grid of measured values did not extend to the edge of the specimen, however). The lower image shows computed and measured values of the profile along the midplane of the specimen.

Comparisons to lattice strains are available in Ref 28. The simulated ODF for the base metal at the end of the loading sequence is presented in Fig. 25; note that the peak value is nearly twice that of Fig. 21. The welded material experiences very little deformation early in the experiment, and its texture just begins to change in the last load levels, so it is not presented. In the experiment, four reflections were measured. The experimental results show the relative change in intensity for each *hkl* and are plotted as the the ratio of the integrated intensity to the original integrated intensity of the undeformed specimen. The simulations use the ratios of the number of crystals in the diffraction condition to the number before deformation. A cutoff angle of 5° is used to ensure that enough crystals were in the diffraction condition. For the (311) direction, with high multiplicity, 337 crystals out of 8000 were lit up initially. The lowest initial count was for the (004) direction with only 117 crystals in the simulated diffraction condition. The (222) direction began with 205 lit-up crystals, and the (220) direction had 127.

Figure 26 compares the simulated relative peak intensities in the base metal with the experimental ones. Overall, the simulations captured the changes well. The strongest changes are in the (222) reflection, which increases the most, and in the (220) reflection, which decreases significantly. Both behaviors are captured well by the simulations. The other two reflections, (311) and (004), show less change both in experiment and simulation. The simulation matches the (004) results well for the first several measurements but does not catch the downturn in the experimental data near the end. For the (311) reflection, both data sets remain relatively flat, but the simulation results are consistently lower.

Summary and Conclusions

This article is devoted to explaining how crystal plasticity is implemented within finite-element formulations. Two physical length scales are considered. The first is that of crystals in which the computational domain is a polycrystal consisting of thousands of crystals. Finite elements constitute individual crystals or parts of a crystal. The second scale is that of an engineering component with characteristic dimensions far larger than that of any single crystal within it. In this case, the volume of an element from the finite-element mesh is considerably greater than that of a polycrystal used to evaluate effective properties. The overall theoretical formulation involving anisotropic elasticity and restricted slip plasticity (also anisotropic) captures many aspects of the observed behaviors of materials under mechanical loading. These include:

- Elastic anisotropy from individual crystals
- Stress distributions within and among crystals

and simulated results. The relative amount of deformation between the welded zone and the base metal is a key feature of the deformed shape. For the simulations to capture this feature, the qualitative estimates of the initial slip

system hardness distribution derived from microhardness measurements were refined.

The simulations are compared to the diffraction experiments, looking in particular at the evolution of the texture in the base metal.

Fig. 25 Final texture orientation distribution function for base metal for specimen 3. MUD, multiples of uniform distribution
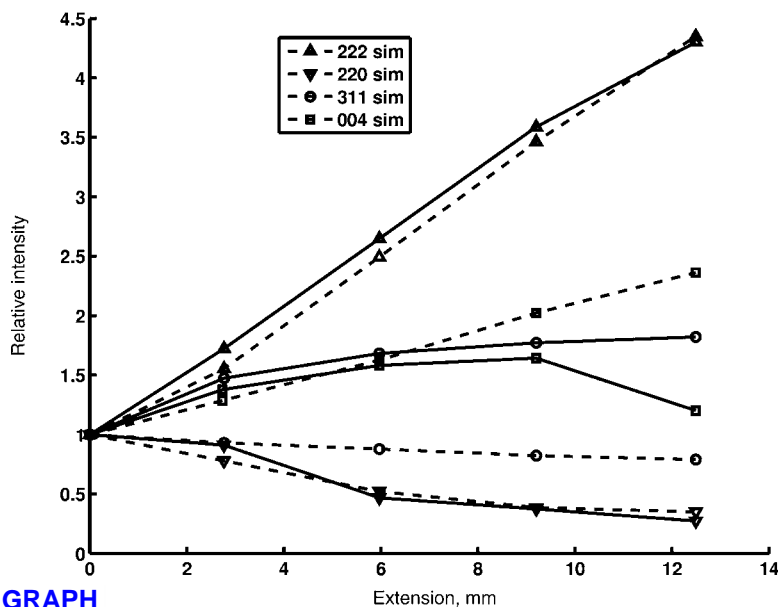
Fig. 26 Measured and computed peak intensity changes during deformation for several diffraction peaks

- Dependence of the plastic flow on crystallographic orientations
- Influence of strain hardening
- Evolution of texture over large plastic strains

Through four examples, the capabilities of formulations are illustrated at these two length scales, with the intention of demonstrating successful application of each formulation to gaining a better understanding of several important issues. At the crystal scale, the simulations of in situ diffraction experiments give insight into how polycrystals collectively respond to loading, showing for crystals within a polycrystal the dependence on interactions with neighboring grains as well as on crystal lattice orientation itself. The simulations at this scale also were used to investigate whether or not latent hardening together with grain-to-grain heterogeneity of deformation could be responsible for the emergence of a brass-type texture. These simulations demonstrated how hardening assumptions manifest themselves in texture evolution differently, depending on how grain interactions are modeled. At the continuum scale, the simulations of bending were used to illustrate the inclusion of plastic anisotropy due to the presence of crystallographic texture. Macroscopic residual-stress distributions were computed, and through weak coupling to the crystal scale, compared directly to diffraction measurements. Finally, the modeling of a specimen cut from a friction-stir-welded plate demonstrated how the state of a material with very strong pre-existing microstructure variations can be characterized and used to initialize simulations of deformations that further alter the microstructure and promote substantial spatial gradients of the strain field.

ACKNOWLEDGMENTS

REFERENCES

1. M. Ashby, Mechanisms of Deformation and Fracture, *Adv. Appl. Mech.,* Vol 25, 1983, 117–177
2. J.F. Nye, *Physical Properties of Crystal; Their Representation by Tensors and Matrices,* Oxford Science Publications, 1957
3. J. Hutchinson, Bounds and self-consistent estimates for creep of polycrystalline materials, *Proc. R. Soc.* (*London*) *A,* Vol 238 1976, p 101–127
4. U.F. Kocks, C.N. Tome, and H. Wenk, Ed., *Texture and Anisotropy,* Cambridge University Press, 1998
5. F. Frank, Orientation Mapping, *Eighth International Conference on Textures of Materials,* J.S. Kallend and G. Gottstein, Ed. The Metallurgical Society, Warrendale, PA, 1988, p 3–13
6. S. Becker and S. Panchanadeeswaran, Crystal Rotations Represented as Rodrigues' Vectors, *Textures Microstruct.,* Vol 10, 1989, p 167–194
7. H. Bunge, *Texture Analysis in Materials Science,* Butterworth, London, 1982
8. P.R. Dawson and E.B. Marin, Computational Mechanics for Metal Deformation Processes Using Polycrystal Plasticity, *Advances in Applied Mechanics,* Vol 34, E. van der Giessen and T.Y. Wu, Ed., Academic Press, 1998, p 78–169
9. C. Tome, R.A. Lebenson, and U.F. Kocks, A Model for Texture Development Dominated by Deformation Twinning: Application to Zirconium Alloys, *Acta Metall. Mater.,* Vol 39, 1991, p 2667–2680
10. N.R. Barton, D.J. Benson, and R. Becker, Crystal Level Continuum Modeling of

Phase Transformations: The Transformation in Iron, *Model. Simul. Mater. Sci. Eng.,* Vol 13, 2005, p 707–731

11. S. Kok, A.J. Beaudoin, and D.A. Tortorelli, A Polycrystal Plasticity Model Based on the Mechanical Threshold, *Int. J. Plast.,* Vol 18, 2005, p 715–741

12. J.M. Gerken and P.R. Dawson, A Crystal Plasticity Model That Incorporates Stresses and Strains due to Slip Gradients, *J. Mech. Phys. Solids,* Vol 56, 2008, p 1651–1672

13. S. Kok, A.J. Beaudoin, and D.A. Tortorelli, On the Development of Stage IV Hardening Using a Model Based on the Mechanical threshold, *Acta Mater.* Vol 50, 2002, p 1653–1667

14. H. Turkmen, M. Miller, P. Dawson, and J. Moosebrugger, A Slip-Based Model for Strength Evolution during Cyclic Loading, *J. Eng. Mater. Technol.,* Vol 126, 2004 p 329–338

15. S.R. Kalidindi, C.A. Bronkhorst, and L. Anand, Crystallographic Texture Evolution in Bulk Deformation Processing of fcc Metals, *J. Mech. Phys. Solids,* Vol 40, 1992, p 537–579

16. E. Marin and P. Dawson, Elastoplastic Finite Element Analyses of Metal Deformations Using Polycrystal Constitutive Models, *Comput. Meth. Appl. Mech. Eng.,* Vol 165, 1998, p 24–41

17. N.R. Barton, J. Knap, A. Arsenlis, R. Becker, R.D. Hornung, and D.R. Jefferson, Embedded Polycrystal Plasticity and Adaptive Sampling, *Int. J. Plast.,* Vol 24, 2008, p 242–266

18. P.R. Dawson, D.E. Boyce, and R.B. Rogge, Issues in Modeling Heterogeneous Deformations in Polycrystalline Metals Using Multiscale Approaches, *Comput. Model. Eng. Sci.,* Vol 10 (No. 2), 2005, p 123–141

19. P.R. Dawson, D.E. Boyce, and R. Rogge, Correlation of Diffraction Peak Broadening and Crystal Strengthening in Finite Element Simulations, *Mater. Sci. Eng. A,* Vol 399 (No. 1–2), 2005, p 13–25

20. A. Kumar and P. Dawson, Computational Modeling of f.c.c. Deformation Textures over Rodrigues Space, *Acta Mater.,* Vol 48, 2000, p 2719–2736

21. M. Miraglia, P. Dawson, and T. Leffers, On the Influence of Mechanical Environment on the Emergence of Brass Textures in fcc Metals, *Acta Mater.,* Vol 55, 2007, p 799–812

22. E. Voce, A Practical Strain-Hardening Function, *Acta Metall.,* Vol 51, 1948, p 210–226

23. G. Taylor, Plastic Strain in Metals, *J. Inst. Met.,* Vol 62, 1938, p 307–324

24. R. Asaro and A. Needleman, Texture Developement and Strain Hardening in Rate Dependent Polycrystals, *Acta Metall.,* Vol 33 (No. 6), 1985, p 923–953

25. E. Marin and P. Dawson, Elastoplastic Finite Element Analyses of Metal Deformations Using Polycrystal Constitutive Models, *Comput. Methods Appl. Mech. Eng.,* Vol 165, 1998, p 23–41

26. E. Marin and P. Dawson, On Modeling the Elasto-Viscoplastic Response of Metals Using Polycrystal Plasticity, *Comput. Methods Appl. Mech. Eng.,* Vol 165, 1998, p 1–21

27. P. Dawson, S. MacEwen, and P. Wu, Advances in Sheet Metal Forming Analyses: Dealing with Mechanical Anisotropy from Crystallographic Texture, *Int. Mater. Rev.,* Vol 48 (No. 2), 2003, p 86–122

28. D.E. Boyce, P.R. Dawson, B. Sidle, and T. Gnaupel-Herold, A Multiscale Methodology for Deformation Modeling Applied to Friction Stir Welded Steel, *Comput. Mater. Sci.,* Vol 38, 2006, p 158–175

Cellular Automaton Models of Recrystallization

C.H.J. Davies, Monash University, Australia

CELLULAR AUTOMATA are a subset of so-called finite-state machines and have application in disciplines as diverse as computer-generated imagery of crowd, flock, and swarm behavior through tools to optimize parameters in complex multiparameter models, to the simulation of microstructural evolution. The study of cellular automata (CA) is a discipline in its own right, and the reader is referred to other sources (Ref 1) for a survey of the mathematics of CA. The power of the approach lies in its fundamental simplicity. A cellular automaton is essentially a collection of rules governing system behavior replicated over an array that may contain many instances of the rules in one, two, or three dimensions. Each instance interacts with its neighbors in a predefined manner, and thus, the global behavior emerges from interactions at a local scale.

Thus, for example, flock behavior is not explicitly programmed into a simulation but rather results from a rule which determines that each member of the flock must avoid its nearest neighbors while maintaining a given speed. From this simple rule evolves the flocking behavior that one marvels at when witnessing a large flock of birds or school of fish. In microstructural simulations, rules that govern the interactions of microstructural subunits (which are given a length scale that is some appropriate factor smaller than the smallest relevant microstructural feature) may be based on the exchange of dislocations in a simulation of solid-state microstructural evolution, from which the Avrami equation for recrystallization kinetics is "rediscovered."

An important point is that the philosophy underlying the CA approach mimics our apprehension of the physical world. Bird "A" does not know what trajectory bird "B" is taking several meters away in the same flock, but their movements are somehow connected. Similarly, grain "X" does not know that grain "Y" is recrystallizing some tens or hundreds of micrometers away in a block of metal, but their behavior can be connected through the kinetics of recrystallization. It is this localization of the rules that distinguish CA simulations (and other representational simulations such as Monte Carlo simulations) from mean field analytical models and models employing a homogeneous effective medium.

This article examines how CA can be applied to the simulation of static and dynamic recrystallization. It describes the approach and how it has evolved since the early 1990s. The framework elements necessary for an operational cellular automaton are described, with the aim of providing the reader with an understanding of the necessary terminology, but will not provide algorithms for constructing a cellular automaton. Other references (e.g., Ref 2) contain more detailed information on CA algorithms and provide the reader with the necessary detail to write their own cellular automaton code. The importance of calibration with experimental data and integrated multiscale modeling (combining a cellular automaton with finite-element models) is highlighted with respect to recrystallization simulations and, in particular, multiphenomena models. The evolution of the field from regular to irregular CA and future outlooks are also addressed.

The Cellular Automaton Method

Cellular automata operate at mesoscopic length scales on the order of subgrains and grains for the simulation of recrystallization but can be coupled with simulations at other length scales to produce powerful multiscale models. Cellular automata can be used in two ways:

- As a conventional simulation engine, which requires input data and can interpolate within the domain of the data and, to some extent, extrapolate beyond that domain
- As an inverse engine, wherein the CA is used to perform "what if?" virtual experiments, the results of which are only compared to the final results of an experiment

The first approach is used, for example, where a simulation is required to assist with exploring options for processing a material, whereas the second can be used to test hypotheses about the physical phenomena underlying recrystallization—perhaps the sites for nucleation or the effect of local texture distribution on the recrystallization. It is often the case that the two extreme cases are combined: Assumptions are made about nucleation in a simulation that is used to explore a domain of known data.

The algorithms for simulation of recrystallization using CA are fairly well established, and any of several can be used to determine the kinetics of a recrystallizing microstructure. The Johnson-Mehl-Avrami-Kolmororov (JMAK) relationship is "rediscovered" from two- and three-dimensional simulations that are not explicitly programmed with a JMAK exponent; in fact, this rediscovery should be seen as one of the tests of validity of any new CA algorithm. In this regard, CA simulations reveal little more about the process of recrystallization than was known 70 years ago. The power of the approach is that the local spatial evolution can be coupled with the temporal evolution, but to include the spatial component, the CA needs experimental data. There are four or five steps in the CA simulation of recrystallization, and each is examined in this article:

1. Defining the CA framework
2. Generating the initial microstructure
3. Distributing nuclei of recrystallized grains
4. Growing the recrystallized grains
5. Updating the dislocation density, which is implemented for simulations of dynamic recrystallization or in situations where significant static recovery occurs concurrently with static recrystallization

The Cellular Automaton Framework

The CA framework consists of the size of the CA, the order (dimensionality), the interaction domain (neighborhood), and the boundary conditions. In this description the term *size of the CA* should be thought of as interchangeable

with the number of elements or cells, whereas the term *scale of the CA* is used for the calibration of CA dimensions to physical dimensions. For microstructural simulations, we are usually only interested in two- or three-dimensional order are of interest because of their correspondence to the methods of imaging or representing bulk microstructures. The upper limit on the size of a cellular automaton is fixed by computer storage capacity, but rarely is this limit reached. It is more likely to be governed by the pragmatic measure of the time required to run the simulation, and this is dictated by computer clock speed and programming language (whether the language is interpreted or compiled).The lower limit to the number of elements is determined by the phenomenon to be simulated and the boundary conditions. The size and scale of the CA must be sufficient to capture the microstructural variation in the physical system. Too few cells in the array will introduce errors because the microstructure and/or the texture is inadequately represented. This is especially a problem with the nucleation of recrystallized grains: Too few nuclei lead to errors in the kinetics and the texture resulting from the simulation. Once the size and order of the array are prescribed, a cellular automaton framework requires a neighborhood definition, which specifies the cells in an array that are permitted to interact with one another, and the boundary conditions prevailing at the edges (faces in three dimensions) of the array.

Typical regular neighborhoods on a square lattice are the von Neumann, Moore, and modified-Moore types (Ref 3) (Fig. 1), all of which can be described in two dimensions, as in the figure, or extended without much effort to three dimensions. Hexagonal (triangular) neighborhoods may be employed, as well as irregular neighborhoods (Ref 2), which are discussed later in this article. Each neighborhood yields a characteristic grain shape prior to impingement that is believable to a lesser (von Neumann) or greater (modified Moore) extent (Fig. 2).

In many cases, a CA simulation of recrystallization uses periodic boundary conditions at each edge (two dimensional) or face (three dimensional) so that the CA wraps around on itself, which, for a two-dimensional CA, results in the plane of the CA lying over a torus. Periodic boundary conditions are appropriate for near-isothermal (or constant heat/cooling rate), near-isostrain situations. For example, a two-dimensional CA representing microstructural evolution on an RD-TD plane of a rolled plate or sheet may not suit all situations (the RD-ND plane of a rolled plate)(Ref 4). (Following convention, the principal axes of a rolled plate are the rolling direction (RD), transverse direction (TD), and normal direction, or (ND).)) In such cases, one or both periodic boundaries must be relaxed to mirror boundary conditions (the cells at the edge reflected), and this introduces errors that depend on the size of the simulation (Ref 5, 6).

Early CA simulations of recrystallization use a two- (Ref 3, 7, 8) or three-dimensional (Ref 9, 10) CA to simulate recrystallization in a homogeneous medium and are essentially aimed at verifying the technique for use in solid-state simulation and exploring the limits of the framework. The simulations are run under the assumptions of the JMAK theory:

- Randomly distributed and homogeneous nucleation
- Either site saturated or constant nucleation rate
- Constant and isotropic growth rate

Compared to the well-known JMAK results, the kinetics of recrystallization are adequately reproduced, apart from minor deviations at the start and end of recrystallization, which reflect the oft-overlooked fact that the JMAK equation is a limit for an infinitely large number of grains, rather than an equality (Ref 6). The characterization of the deformed structure as a homogeneous medium implies a mean field effect of the surroundings on the growth of recrystallized grains, something that is known to be a gross approximation but is consistent with JMAK theory. This work demonstrates that under the assumptions of the theory, the JMAK exponent does not depend on the type of neighborhood used (Ref 3, 8).
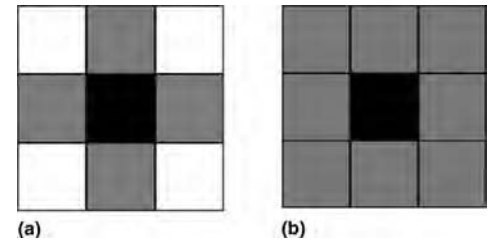
Fig. 1 Neighborhood definitions. (a) von Neumann neighborhood. If (i,j) is the index of the central cell, neighbors are $(i-1,j)$, $(i+1,j)$, $(i,j-1)$, $(i,j+1)$. (b) Moore neighborhood. Neighbors are $(i-1,j)$, $(i+1,j)$, $(i,j-1)$, $(i,j+1)$, $(i-1,j-1)$, $(i+1,j-1)$, $(i+1,j-1)$, $(i+1,j+1)$. The modified-Moore neighborhood operates over two time-step intervals, alternating between the diagonal from top left and the diagonal from top right.
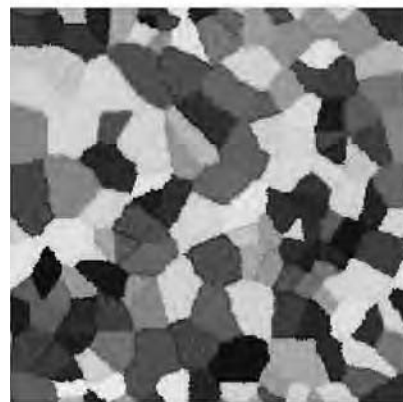
Fig. 2 Microstructures resulting from differing neighborhood definitions. (a) von Neumann neighborhood. Note the strong dependence of the orientation of grain boundaries on the neighborhood definition. (b) Modified-Moore neighborhood. Simulations times are arbitrary time steps.
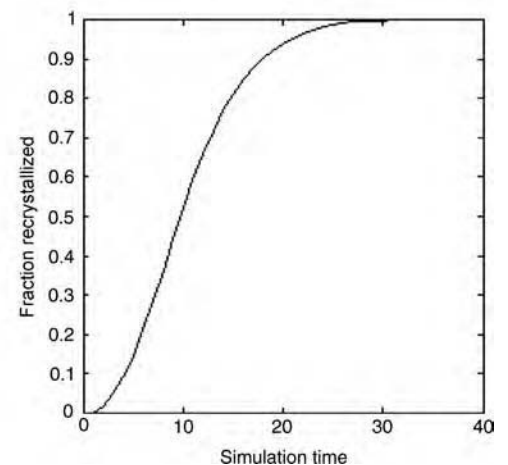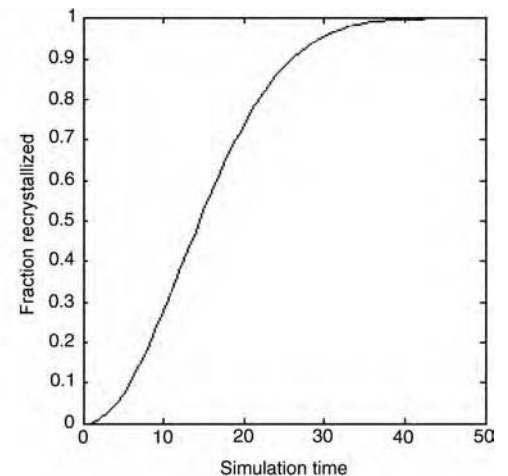
Once the framework of the CA is established, the CA must be scaled to physical lengths and times. This is achieved both through the generation of the initial microstructure (length) and through the rules for nucleation and growth (length and time).

Generating the Initial Microstructure

How the initial microstructure is incorporated into the simulation is a philosophical question as well as a practical one. The microstructure can itself be the result of a simulation or model, or it can be captured in an explicit manner from experimental data. Simulations of the microstructure can be more compact, whereas explicit data capture can be more realistic, and the modeler must decide what level of information is necessary to capture the essential features to be modeled. For example, the kinetics of dynamic recrystallization of copper (Ref 11) and austenite (Ref 12) can be largely reproduced without the inclusion of any texture information. The advantage of the CA approach is that phenomena such as necklace-type microstructures and oscillations in the flow stress curve emerge as a result of the simulation, rather than being explicitly programmed, and are more easily modeled than by analytical means. Such geometrical CA are used (Ref 11, 13) to confirm—it is emphasized, without explicit programming of the event—the observations of Sakai et al. (Ref 14) that if the initial grain size is greater than twice the steady-state grain size, then the flow stress curve will exhibit multiple peaks as a result of dynamic recrystallization. Careful consideration of the length scale is important to the success of a simulation. The scale, usually defined by the distance between cell centers, must be such that the area (volume) is representative of the microstructure and microstructure distribution (grain size and texture) yet not so great as to mask phenomena of interest.

Seeding Algorithms. The simplest technique for simulating the starting microstructure is to generate a geometry using a seeding algorithm based on a cellular automaton (or a Monte Carlo) simulation, which requires input of the number density of grains (Ref 11, 15). Seeding the initial microstructure has the advantage that it can be grown in two or three dimensions but may lack other microstructural state information, such as texture and dislocation density. The dislocation density can be included for static recrystallization by distributing a known average dislocation density over the simulation field. Genetic algorithms are employed to distribute the dislocation density over the synthetic microstructure (Ref 16), but this approach is in its infancy; presumably, the technique could be extended to include texture distributions. Notwithstanding their simplicity, such starting microstructures are useful for investigating the geometric features of recrystallization (Ref 11, 15).

Because crystallographic texture plays such a large role in recrystallization, it is often desirable to start with a more complete description of the microstructure. Texture can be incorporated by statistical sampling of data and subsequent mapping on the CA (Ref 17), by direct import of data from electron backscattered diffraction (EBSD) in the scanning electron microscope (Ref 18), or by coupling the CA simulation with a spatially discrete deformation simulation such as a finite-element or crystal-plasticity simulation (Ref 19–22). Each of these techniques is discussed in the following paragraphs.

Sampling and Mapping Techniques. Statistical sampling methods can reproduce faithfully the volume fraction of dominant texture components, although they can significantly sharpen the discretized texture compared to the experimental measurement (Ref 17, 23), and this can adversely affect the kinetics of the simulation and the resultant texture. This problem is not exclusive to CA simulations. Orientations can then be assigned to grains according to the volume fraction of each texture component. However, this does not reproduce the correlated grain-boundary character distribution (GBCD), which may be important for microstructures in which texture banding is prevalent and/or important in the recrystallization process (for example, in warm-rolled steels). Preserving the GBCD in conjunction with the volume fraction of texture components mapped from statistical sampling is much more challenging and requires the manipulation of the texture until the GBCD of the simulation matches the experimental GBCD to a given predefined tolerance (Ref 24). Adding a correlation with grain-size distribution further increases the complexity of the task, and, as far as the author knows, this three-way correlation has yet to be accomplished.

Importing data directly from EBSD experiments does away with the need to correlate texture, misorientation, and grain size, because this information is a result of the experiment (Ref 18). Furthermore, the scale of the CA can be coupled to the scale of the EBSD image (Ref 25). However, EBSD data are often imperfect and must therefore be processed before being used as input to a simulation. Decisions must be made about the completion of grain boundaries, minimum viable grain sizes (based on numbers of pixels), and the errors associated with limitations of the spatial and angular resolution inherent to the experiment (Ref 26). Crucial for a recrystallization simulation is the need to make some assumption about the stored-energy distribution based on the EBSD pattern quality (Ref 27). The EBSD maps are maps on a surface, albeit one that can be sectioned from deep inside a sample, and implicit in the use of EBSD data as the starting microstructure for a simulation is acknowledgment of the two-dimensional nature of the data. However, three-dimensional EBSD data obtained by serial sectioning using focused ion beam field emission

gun scanning electron microscopy (FIB-FEG-SEM) are becoming more common (Ref 26), but the issues of translating the often imperfect data to the simulation persist. As a consequence, EBSD maps themselves (rather than the secondary data that can be derived from EBSD experiments) are perhaps best used for validation of a simulation rather than as the input microstructure.

Cellular Automata Coupled with Deformation Simulations. Coupled simulations (Ref 12, 20, 25, 28) circumvent the issues of recasting or correcting data derived from imaging techniques. The broad principle is the same in each case: The phenomena that govern recrystallization, such as the stored energy/dislocation density, are calculated from a finite-element or texture model and are passed to the CA. Assumptions made in the deformation model necessarily influence the microstructural model.

Finite-element (Ref 10, 28) and crystal-plasticity finite-element (Ref 22, 25) methods have been used to generate input data upon which the CA subsequently operates to simulate static recrystallization (Fig. 3). Translation of the finite-element information to the CA is accomplished by mapping the finite-element mesh onto the CA grid. (Ref 22, 28) However, it is important to note that the mapping of the finite-element mesh onto the CA grid is not one-to-one, and thus, there is an averaging of data in one or both directions of data transfer. A practical requirement in these coupled models is that the CA cell spacing is smaller than the finite-element node spacing, and this means that clusters of CA cells adopt the same macroscopic state variables as one another, based on the distribution of the finite-element nodes (Ref 22). The consequences of this are remarked upon by Raabe and Becker (Ref 22), but there does not seem to have been a systematic study of the sensitivity of the microstructural model to differences between the spacing of finite-element nodes compared to CA cells. For example, early finite-element/CA-coupled models successfully embedded the CA wholly within a single element. Gottstein's group (Ref 21) use the grain interaction model as the starting point for simulations of static recrystallization, and simple dislocation models can be embedded in the CA to simulate both dynamic dislocation evolution and static recovery (Ref 10–13). Updating of the dislocation density is explored later.

Nucleation and Growth of Recrystallized Grains

Nucleation and growth are captured in transition rules that determine how cells in the array change their state and include the initialization of recrystallization (nucleation rules), and how cells respond to like and unlike cells in their neighborhood (growth rules). Temporal calibration enters the simulation through these transition rules.
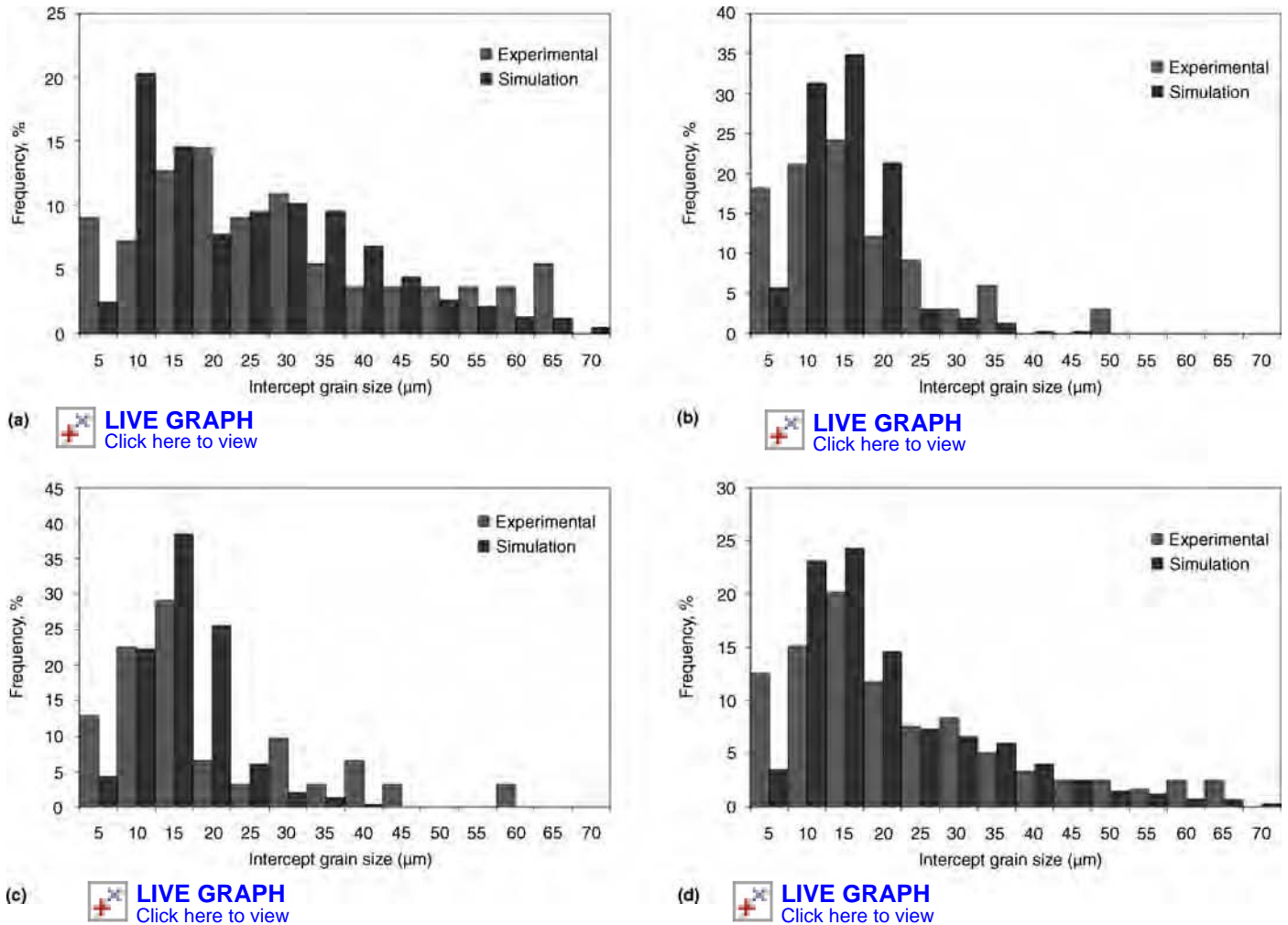
Fig. 3 Comparison of simulated and experimental grain-size distribution for each texture class arising from recrystallization in AA1050. (a) Cube class. (b) Random class. (c) Rolling class. (d) All classes combined. Source: Ref 29 with permission from Elsevier

Probability of Nucleation. The transition rule that dictates the probability of nucleation, P_N, is governed by the equation:

$$P_N = \frac{\dot{N}^Y dt}{N_{CA}} \quad \text{(Eq 1)}$$

where $\dot{N}^Y$ is the nucleation rate, and N_{CA} is the number of cells in the cellular automaton, with dt the time step. The nucleation rate may be a global rate or specific to a texture component (Ref 29) and/or location such as a grain boundary or grain-boundary triple point. Of course, this texture- and site-specific information requires additional experimental detail and conditional statements in the simulation code.

Nucleation of recrystallization in a CA simulation is generally accomplished by an imposed rule based on a cell achieving a critical dislocation density (Ref 12, 20, 25) set by the modeler. This critical dislocation density can be calculated (Ref 12, 20) based on the analytical models of Sandstrom and Lagneborg (Ref 30) or set at some fraction of the maximum dislocation density. Alternatively, the nucleation rate can

be characterized as a function of processing parameters, such as the Zener-Hollomon parameter, and strain and used as direct input to the CA simulation (Ref 31). In each case, the nucleation criterion is not derived from a model of nucleation per se but rather is a proxy for events as yet poorly characterized that lead to the observation at the mesoscale of a new strain-free grain. Mukhopadhyay et al. (Ref 21) have used a CA simulation in passing in an inverse manner to test three nucleation scenarios and found a noticeable difference between the three scenarios, without comment on the applicability of each. Analytical models that are quantitatively site specific (Ref 32) are eminently suitable for CA simulations (Ref 33, 34).

The probability of growth P_G, is, in its simplest form:

$$P_G = \frac{v dt}{s_{CA}} \quad \text{(Eq 2)}$$

where v is the interface migration rate, and s_{CA} is the cell center-to-center distance. The probability of growth can be rendered unity (certain

transformation of an unrecrystallized cell to a recrystallized state) by appropriate selection of the time step, dt. Similar to the nucleation rules, growth rules can depend on a specific texture component (e.g., measurements made on texture "classes," (Ref 29) or based on high-mobility $40° <111>$ boundaries in face-centered cubic metals, (Ref 10) and the stored energy of the region into which the newly recrystallized grain is growing. A key advantage of CA over other representational simulation approaches is that a cellular automaton can be readily calibrated directly to experimental timescale through the interface velocity entering the numerator of the equation governing the probability of growth, and to experimental lengths by ascribing a length to the cell dimensions that enter the denominator of the same equation.

Interface migration is incorporated into the transition rule representing the probability of growth in a neighborhood (Eq 2), and in this respect, one is not tracking a moving boundary, as such, but rather the expansion of a cluster of cells with the same orientation. The boundary is defined by the orientation difference (as it is

in Monte Carlo simulations, Ref 35) rather than as a surface (as is the case in vertex simulations). Interface migration rates can be successfully determined directly from experimental data using the Cahn-Hagel approach (Ref 29), with good correlation between experimental and simulated microstructures (Fig. 4). However, it is preferable to use a more indirect approach, because the Cahn-Hagel approach requires that a migration rate be determined by texture component class. Interface migration remains largely based on the relationship between the migration rate, v, and the mobility, M, and driving pressure, P (Ref 36):

$$v = MP \tag{Eq 3}$$

where the mobility is an Arrhenius rate expression varying with temperature, T:

$$M = M_0 \exp\left(-\frac{Q_b}{RT}\right) \tag{Eq 4}$$

where M_0 is a pre-exponential constant, Q_b is the activation energy, R is the gas constant, and T is absolute temperature. The driving pressure is either expressed as a direct function of dislocation density (Ref 21) or of the energy of low-angle boundaries (Ref 36). In either instance, solute or Zener drag can be included as an additive term in the driving pressure relationship (Ref 21, 25).

Interface migration models are heavily reliant on assumptions regarding the mobility of boundaries (in particular), but this is one area in which CA models can be used to test those assumptions by a comparison of simulated and experimental results.

Updating the Dislocation Density. In many cases, it is necessary to also include a model for the evolution of the parent microstructure during recrystallization, either because of significant recovery during static recrystallization (e.g., Ref 10, 17) or because of dislocation evolution during dynamic recrystallization (Ref 11, 12). Concurrent recovery during static recrystallization can also be incorporated in time-dependent form for a dislocation density, ρ, at time, t (Ref 17):

$$\rho(t) = \rho_0 \cdot \exp\left(-\frac{t}{t_0}\right) \tag{Eq 5}$$

where ρ_0 and t_0 are constants. This reduces the driving pressure as the simulation progresses and effectively captures the recrystallization behavior in, for example, aluminum alloys (Ref 10).

A single-variable internal-state evolution equation can be used in the case of dynamic recrystallization to calculate the local dislocation density, ρ, as it varies with strain, ε (Ref 11, 12). In the absence of recrystallization, the dislocation density evolves with strain in the form:

$$\frac{d\rho}{d\varepsilon} = k_1 - k_2\rho \tag{Eq 6}$$

or

$$\frac{d\rho}{d\varepsilon} = \frac{1}{bd} + k_1\sqrt{\rho} - k_2\rho \tag{Eq 7}$$

where k_1 and k_2 are constants. The values of these constants are calculated by analyzing the work-hardening curves from experimental studies on the relevant material (Ref 11, 12) based on common procedures (Ref 37, 38). Once a recrystallized nucleus forms, its dislocation density is reduced catastrophically to zero before continuing to evolve in the following time step. Using this approach, Ding and Guo (Ref 11) have been able to achieve close agreement between simulation kinetics and grain sizes and experimental data in copper. Yazdipour et al. (Ref 12) have achieved similar agreement in steels.

In the case of dynamic recrystallization, CA coupled with conventional finite-element models rely on the two-way coupling of the CA and finite-element model at each finite-element Gauss point (Ref 20), in contrast to the one-way coupling used in simulations of static recrystallization. State variables determined by the finite-element model are distributed across the CA, and the consequent mechanical response of the CA as a result of microstructural evolution is handed back to the finite-element model (Ref 20). In this way, the finite-element and CA models can be updated as deformation progresses, and dynamic recrystallization can be simulated (Fig. 5). Reminiscent of the mapping of CA and finite-element grids in static recrystallization, where the finite-element grid captures multiple CA cells with an element (Fig. 6), the situation in coupled dynamic recrystallization simulations is more difficult, and artifacts of visualization
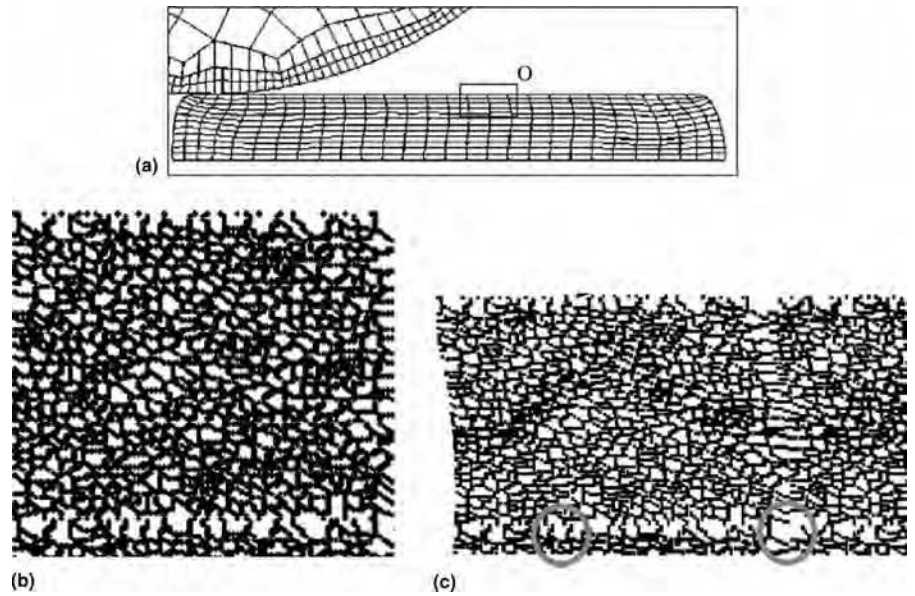
Fig. 4 The cellular automaton finite-element (CAFÉ) model of hot rolling of steel. (a) Slab exiting the rolling gap after it has been rolled at 30% reduction in thickness. (b) Initial cellular automaton microstructure with equiaxed grains. (c) Microstructure near the slab surface within box "O" after rolling. Note grains elongated in the rolling direction. Open regions, such as those shown by the encircled examples, are because grain-boundary continuity is not enforced across element boundaries. Source: Ref 20 with permission from The Royal Society and the author
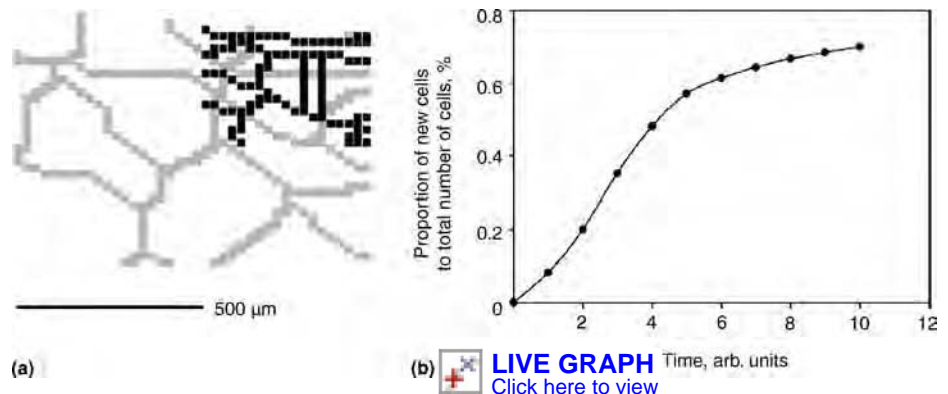
Fig. 5 Recrystallization modeling. (a) Original undeformed cellular automaton (CA) grain structure in gray and the recrystallized grains in black. (b) Ratio of the recrystallization nuclei cells to the total number of cells within a CA attached to one Gauss point. Source: Ref 20 with permission from The Royal Society and the author
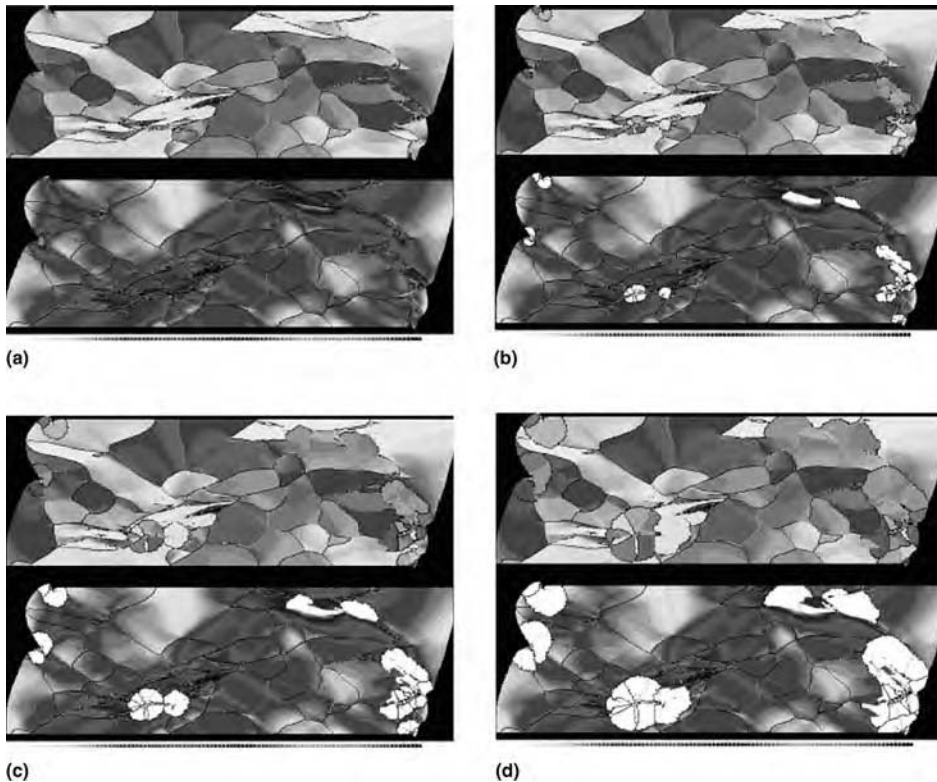
Fig. 6 Two-dimensional simulations of primary static recrystallization in a deformed aluminum polycrystal on the basis of crystal-plasticity finite-element data. The figure shows the change in both microtexture (upper images) and dislocation density (lower images), which was derived from the value of the accumulated crystallographic shear, as a function of the annealing time during isothermal recrystallization. The white areas in the lower images indicate a stored dislocation density of zero; i.e., they are recrystallized. The black lines in both figures indicate misorientations above 15°, and the thin gray lines indicate misorientations between 5° and 15°, regardless of the rotation axis. The orientation image given in the upper figures represents different crystal orientations by different gray levels. The simulation parameters are: annealing temperature, 800 K; site-saturated nucleation conditions; kinetic instability criterion, misorientation above 15°; thermodynamic instability criterion, dislocation density larger than 70% of the maximum occurring value; maximum occurring driving force, 1 MPa; activation energy of the grain-boundary mobility, 1.46 eV; pre-exponential factor of the grain-boundary mobility, $M_0 = 8.3 \times 10^{-3}$ m^{3}N^{-1}s^{-1}; and mesh size of the cellular automaton grid (scaling length), $\lambda m = 61.9$ µm. The images show a crystal in the following states of recrystallization: (a) 0, (b) 3, (c) 13, and (d) 22%. (Figures (e) to (h) from the original are omitted.) Source: Figure and caption text reproduced from Ref 22 with permission of IOP Publishing Ltd. and the author

scale—those that describe nucleation, interface migration, recovery, and precipitation.

Nucleation and growth of recrystallizing grains are governed in conventional CA by the probabilities of occurrence of the events (Eq 1, 2), and these require the CA to be coupled to analytical submodels. Significant experimental and computational effort is necessary to develop nucleation models and interface migration models. One can speculate that submodels could develop along a number of paths. As was remarked upon earlier, the advantage of CA simulations is that they allow the global rules to be "rediscovered" from local interactions in much the same way as mean-field rate equations are conventionally derived from real experimental data. The inconsistency then is that the submodels that govern dislocation generation, recovery, precipitation, interface migration, and the nucleation of recrystallized grains are mean-field models. Scalable CA (Ref 21) would appear to allow the modeler to introduce at least some measure of local interaction concurrently to phenomena that operate and can be simulated at different length scales, such as CA simulations (as opposed to analytical models already incorporated into through-process models) of precipitation and precipitate coarsening (Ref 21, 42), concurrent with recrystallization simulations. Such a multiscale multiphenomena model would be philosophically attractive, but this must be tempered with the question accompanying any modeling activity: Is any further insight gained by investing the significant resources required to accomplish this task? Thus, in tandem with these philosophically attractive models, development of appropriate analytical models will continue, and these will not be exclusive to CA. Increased routine use of FIB-FEGSEM will lead to advances in the models governing nucleation and interface migration. Increasing computer storage and memory capacity will facilitate both approaches.

Developments in Cellular Automaton Simulations

Developments in CA simulations have largely focused on algorithm developments that allow the CA to store data more efficiently or to execute more efficiently. The cellular operator or adaptive CA model extends the basic CA by introducing adaptive scaling to the model, and this allows a significantly higher misorientation resolution to be simulated. Although relatively new at the time of writing, the scalable CA would seem to lend itself to more accurate modeling of nucleation but for the dearth of physically realistic nucleation models.

Irregular Cellular Automata. Another exciting development in CA algorithms is the irregular CA, introduced by Janssens (Ref 39) and adopted by others (Ref 12). Instead of a fixed neighborhood definition, the irregular CA defines a neighborhood as a capture radius over a random grid. In this way, the CA algorithm is also altered in a subtle but significant way so that cells are captured by the central cell of the neighborhood, rather than the central cell being captured by cells in its neighborhood, the so-called inverse updating rule (Ref 40). The inverse updating rule enables the CA to subsequently model curvature-driven boundary migration (grain growth) and thereby to incorporate several grain topology models in the one simulation, unlike hybrid models (Ref 41) that combine Monte Carlo and CA approaches.

Thus, CA algorithms are relatively well developed, and with the introduction of scalable CA (Ref 21) and irregular random CA (Ref 39, 40), all the tools exist for multiscale, multiphenomena models. Mesoscopic length-scale CA simulations are effectively coupled to macroscopic simulations and models, and the key developments in CA are therefore likely to be in the submodels at the microscopic length

Summary

Cellular automaton simulations are a class of simulation tools that allow global behavior to emerge from local interaction rules in much the same way that recrystallization is perceived to occur in real systems. Unlike other simulation methods at the mesoscopic length scale, they are readily coupled to spatio-temporal rules and data determined by experimental observation. As a consequence, CA are able to be coupled to simulations and models across multiple length scales and offer the opportunity to be used in an inverse manner to test theories and assumptions underlying the microstructural phenomena contributing to recrystallization. As CA have become more sophisticated, one can expect them to be used to test the assumptions about the physical phenomena contributing to recrystallization, by a combination of rule-testing and validation by comparison with experimental data.

(but not state variables) can be introduced (Fig. 5).

REFERENCES

1. S. Wolfram, *A New Kind of Science,* Wolfram Media, 2002

2. K.F.G. Janseens, Chapter 4, Cellular Automata, *Computational Materials Engineering, An Introduction to Microstructure Evolution,* K. Janssens, D. Raabe, B. Nestler, E. Kozeschnik, and M. Miodownik, Elsevier, 2007, p 109–150

3. H.W. Hesselbarth and I.R. Göbel, Simulation of Recrystallization by Cellular Automata, *Acta Metall. Mater.,* Vol 39, 1991, p 2135–2143

4. C.H.J. Davies, The Cellular Automaton Simulation of Metal Rolling, *Multi-Scale Phenomena in Materials—Experiments and Modelling,* MRS Symposium Proceedings, Vol 578, I.M. Robertson et al., Ed., 1999, p 457–468

5. S. Fletcher, M. Thomson, and T. Tran, Growth of Circular Crystals in a Circular Region, *J. Chem. Soc., Faraday Trans. I,* Vol 80, 1984, p 1867–1874

6. S. Fletcher, M. Thomson, and T. Tran, Numerical Analysis of 2D Nucleation, Growth, Collision Processes, *J. Electroanal. Chem.,* Vol 199, 1986, p 241–247

7. C.F. Pezzee and D.C. Dunand, *Acta Metall. Mater.,* Vol 42, 1994, p 1509–1524

8. C.H.J. Davies, The Effect of Neighbourhood on the Kinetics of a Cellular Automaton Recrystallisation Model, *Scr. Metall. Mater.,* Vol 33, 1995, p 1139–1143

9. C.H.J. Davies, Growth of Nuclei in a Cellular Automaton Simulation of Recrystallisation, *Scr. Mater.,* Vol 36, 1997, p 35–40

10. V. Marx, F.R. Reher, and G. Gottstein, Simulation of Primary Recrystallization Using a Modified Three Dimensional Cellular Automaton, *Acta Mater.,* Vol 47, 1999, p 1219–1230

11. R. Ding and Z.X. Guo, Coupled Quantitative Simulation of Microstructural Evolution and Plastic Flow during Dynamic Recrystallization, *Acta Mater.,* Vol 49, 2001, p 3163–3175

12. N. Yazdipour, C.H.J. Davies, and P.D. Hodgson, Microstructural Modeling of Dynamic Recrystallization Using Irregular Cellular Automata, *Comput. Mater. Sci.,* Vol 44, 2008, p 566–576

13. R.L. Goetz and V. Seetharaman, Modeling Dynamic Recrystallization Using Cellular Automata, *Scr. Metall.,* Vol 38, 1998, p 405–413

14. T. Sakai, M.G. Akben, and J.J. Jonas, *Acta Metall.,* Vol 31, 1983, p 631

15. R.L. Goetz, Particle Stimulated Nucleation during Dynamic Recrystallization Using a Cellular Automaton Model, *Scr. Mater.,* Vol 52, 2005, p 851–856

16. S. Ghosh, P. Gabane, A. Bose, and N. Chakraborti, Modeling of Recrystallization in Cold Rolled Copper Using Inverse Cellular Automata and Genetic Algorithms, *Comput. Mater. Sci.,* Vol 45, 2009, p 96–103

17. V. Marx, D. Raabe, O. Engler, and G. Gottstein, Simulation of the Texture Evolution during Annealing of Cold Rolled bcc and fcc Metals Using a Cellular Automaton Approach, *Textures Microstruct.,* Vol 28, 1997, p 211–218

18. D. Raabe and L. Hantcherli, 2D Cellular Automaton Simulation of the Recrystallization Texture of an IF Steel Sheet Under Consideration of Zener Pinning, *Comput. Mater. Sci.,* Vol 34, 2005, p 299–313

19. H. Aretz, R. Luce, M. Wolsker, R. Kopp, M. Goerdeler, V. Marx, G. Pomana, and G. Gottstein, *Model. Simul. Mater. Sci. Eng.,* Vol 8, 2000, p 881–891

20. S. Das, A. Shterenlikht, I.C. Howard, and E.J. Palmiere, A General Method for Coupling Microstructural Response with Structural Performance, *Proc. R. Soc. A,* Vol 462, 2006, p 2085–2096

21. P. Mukhhopadhyay, M. Loeck, and G. Gottstein, A Cellular Operator Model for the Simulation of Static Recrystallization, *Acta Mater.,* Vol 55, 2007, p 551–564

22. D. Raabe and R.C. Becker, Coupling of a Crystal Plasticity Finite Element Model with a Probabilistic Cellular Automaton for Simulating Primary Static Recrystallization in Aluminium, *Model. Simul. Mater. Sci. Eng.,* Vol 8, 2000, p 445–462

23. P. Eisenlohr and F. Roters, Selecting a Set of Discrete Orientations for Accurate Texture Reconstruction, *Comput. Mater. Sci.,* Vol 42, 2008, p 670–678

24. M. Miodownik, A.W. Godfrey, E.A. Holm, and D.A. Hughes, On Boundary Misorientation Distribution Functions and How to Incorporate Them into Three-Dimensional Models of Microstructural Evolution, *Acta Mater.,* Vol 47, 1999, p 2661–2668

25. D. Raabe, Mesoscale Simulations of Recrystallization Texture and Microstructures, *Adv. Eng. Mater.,* Vol 3, 2001, p 745–752

26. S. Zaefferer, Application of Orientation Microscopy in SEM and TEM for the Study of Texture Formation during Recrystallisation Processes, *Mater. Sci. Forum,* Vol 495–497, 2005, p 3–12

27. D. Raabe, Multiscale Recrystallization Models for the Prediction of Crystallographic Textures with Respect to Process Simulation, *J. Strain Anal.,* Vol 42, 2007, p 253–268

28. R. Luce, M. Wolske, R. Kopp, V. Marx, F. Roters, and G. Gottstein, Integrated Microstructural Simulation for Multistep Forming Processes, *Proc. Fourth Conference on Recrystallization and Related Phenomena,* T. Sakai and H.G. Suzuki, Ed., The Japan Institute of Metals, 1999, p 739–744

29. C.H.J. Davies and L. Hong, The Cellular Automaton Simulation of Static Recrystallization in Cold Rolled AA1050, *Scr. Mater.,* Vol 40, 1999, p 1145–1150

30. R. Sandstrom and R. Lagneborg, Model for Hot Working Occurring by Recrystallization, *Acta Metall.,* Vol 23, 1975, p 387–398

31. C.H.J. Davies, Intelligent Materials Design and Processing: Simulating Microstructural Evolution during Metal Processing, *The Brimacombe Memorial Symposium,* G. Irons and A.W. Cramb, Ed., CIM, Toronto, 2000, p 685–694

32. F. Lefevre-Schlick, Y. Brechet, H.S. Zurob, G. Purdy, and D. Embury, On the Activation of Recrystallization Nucleation Sites in Cu and Fe, *Mater. Sci. Eng. A,* Vol 502, 2009, p 70–78

33. R.L. Goetz and V. Seetharaman, Static Recrystallization with Homogeneous and Heterogeneous Nucleation Using a Cellular Automata Model, *Metall. Mater. Trans. A,* Vol 29, p 2307–2321

34. P.R. Rios, L.O. Pereira, F.F. Oliveira, W.L.S. Assis, and J.A. Castro, Impingement Function for Nucleation on Non-Random Sites, *Acta Mater.,* Vol 55, 2007, p 4339–4348

35. M.A. Miodownik, A Review of Microstructural Computer Models Used to Simulate Grain Growth and Recrystallisation in Aluminium Alloys, *J. Light Met.,* Vol 2, 2002, p 125–135

36. F.J. Humphreys and M.J. Hatherly, *Recrystallization and Related Annealing Phenomena,* Pergamon Press, 1996

37. H. Mecking and U.F. Kocks, Kinetics of Flow and Strain-Hardening, *Acta Metall.,* Vol 29, 1981, p 1865–1875

38. Y. Estrin, Dislocation-Density-Related Constitutive Modeling, *Unified Constitutive Laws of Plastic Deformation,* A.S. Krausz and K. Krausz, Ed., Academic Press, London, 1996, p 69–106

39. K.G.F. Janssens, Random Grid, Three-Dimensional, Space-Time Coupled Cellular Automata for the Simulation of Recrystallization and Grain Growth, *Model. Simul. Mater. Sci. Eng.,* Vol 11, 2003, p 157–171

40. K.G.F. Janssens, An Introductory Review of Cellular Automata Modeling of Moving Grain Boundaries in Polycrystalline Materials, *Math. Comput. Simul.,* 2009, in press

41. A.D. Rollett and D. Raabe, A Hybrid Model for Mesoscopic Simulation of Recrystallization, *Comput. Mater. Sci.,* Vol 21, 2001, p 69–78

42. M.J. Young and C.H.J. Davies, Cellular Automaton Modelling of Precipitate Coarsening, *Scr. Mater.,* Vol 41, 1999, p 697–701

Monte Carlo Models for Grain Growth and Recrystallization

Mark Miodownik, King's College London, United Kingdom

COMPLEX MODELS of microstructural evolution are required to predict texture, misorientation distribution, and grain-size distributions of a final rolled or extruded product. However, there is a major difficulty: Microstructural evolution is a many-variable, multiscale problem, and even such apparently simple phenomena such as grain growth and recrystallization are still not well understood. For instance, boundary energy and boundary mobility are not smooth functions of boundary misorientation, or boundary plane, or composition, or temperature (Ref 1). It is not just that the boundaries are complicated entities, the grain structures themselves are three-dimensional (3-D) interconnected boundary networks whose evolution cannot be deduced solely from the individual behavior of an average boundary. Explicitly including the topology of the boundary network into models is important because the misorientation of a boundary of a growing grain is defined not only by its crystallography but also by the crystallography of the grain into which it is growing. This article focuses on one method for doing this, the Monte Carlo Potts model, which is typically used to model grain growth, Zener-Smith pinning, abnormal grain growth, and recrystallization, as shown, for example, in Fig. 1.

This article is set out in the following way: First, the basics of the model are introduced, giving the reader enough information to perform their own simulations and a flavor of what the Potts model is capable of modeling; the importance of gaining practical experience of simulations as a way of understanding the emergent behavior of such Monte Carlo systems cannot be emphasized enough. A section then deals with how to incorporate experimental parameters and how to validate the model by comparing the observed behavior quantitatively with theory. Industrial applications of the model are dealt with next, covering anisotropic grain growth, abnormal grain growth, grain growth stagnation, and recrystallization. Finally, an algorithms section is included to give the reader a wide selection of efficient Potts algorithms, such as boundary-site models,
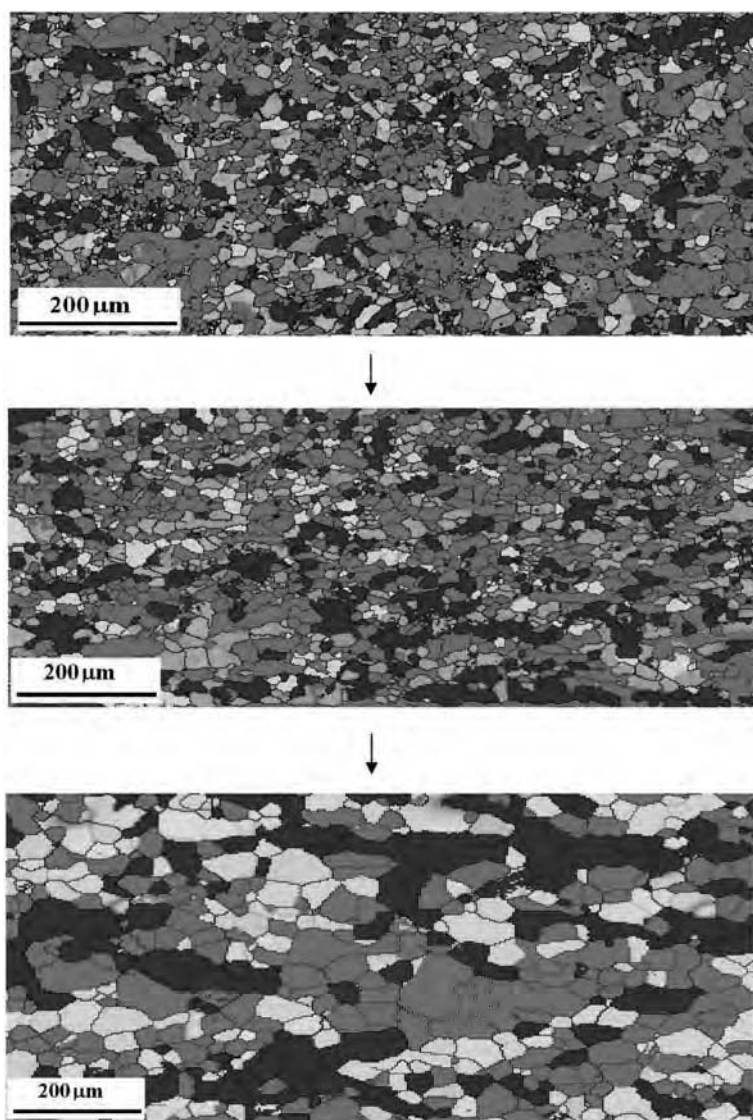
Fig. 1 Potts model simulation of the microstructural evolution of a silicon steel. Grains that are part of a <110> fiber parallel to the sheet normal, within 15° of the <110> axis, are shown in light gray; <111> fiber grains are shown in white; and <100> fiber grains are shown in dark gray. Source: Ref 2

n-fold way models, and parallel models, which are needed to simulate large-scale industrial applications.

The Method

The Potts model simulates the effects of interface energy on the topology of boundary networks. Figure 2 shows the evolution of such system simulated using the Potts model. In this simulation, the grain boundaries are associated with an isotropic excess energy, which has a profound influence on the network topology because it implies that in two-dimension 2-D the number of boundaries impinging on a vertex is always equal to three. In other words, only triple points with 120° vertex angles are stable (Ref 3). Fourfold and higher vertices, if formed, will always dissociate into the relevant number of triple points. This topological law comes out of the model, rather than being imposed on it, which is one of the remarkable features of the model that make it so attractive. Other implications of modeling the interface energy of a network of boundaries are discussed later in the article. First, the basics of the model are discussed.

The Monte Carlo Potts model is a generalization of the Ising model (Ref 4–6). The state of the system is discretized into a regular array of lattice sites and is described in terms of the set of Q crystallographic orientations, which are associated with each lattice site, $s_i \in \{0, Q\}$, where i labels the lattice site. The system defines a boundary between unlike orientations and no interface between like orientations.

In the most basic Potts model, the Q orientations can be represented as integers. In the isotropic case, the energy associated with the boundaries between orientations is described by an energy function, γ, of the form:

$$\gamma(s_i, s_j) = \begin{cases} 0 & \text{for } s_i = s_j \\ \frac{J}{2} & \text{for } s_i \neq s_j \end{cases} \qquad \text{(Eq 1)}$$

where i represents a site, j is its neighbor, and $J(>0)$ is an interfacial energy constant of the system. Thus, the energy of the system can be written as a sum over the spatial distribution of the orientations, called a Hamiltonian, as follows:

$$E = \sum_{i=1}^{N} \sum_{j=1}^{z} \gamma(s_i, s_j) \qquad \text{(Eq 2)}$$

where N is the total number of lattice sites in the system, and z is the maximum number of neighbors. The definition of neighbor orientation depends on the dimension and co-ordination of the lattice, which is covered later. It is the reduction of interface energy that drives the microstructural evolution shown in Fig. 2.

The energy function may also include a volume energy term associated with a volume

driving force, $H(>0)$, which lowers the energy of one type of orientation relative to the other:

$$E = \sum_{i=1}^{N} \sum_{j=1}^{z} \gamma(s_i, s_j) - H \sum_{i=1}^{N} s_i \qquad \text{(Eq 3)}$$

where the H term represents a sum over orientations $s_i = 1$. J scales with the interfacial energy of the system in which the single crystal has zero energy; all polycrystalline states have positive energy that scales with the total boundary area. H may represent stored energies that arise in the case of deformed structures and so provide a driving force for recrystallization.

Dynamics. Microstructural evolution is simulated by using a Monte Carlo method to sample different states of the system. The method is extremely simple in principle: Choose a site at random, propose a change in orientation, calculate the change in energy, ΔE, associated with that orientation swap, and accept or reject the change based on ΔE. There are two main methods for performing such dynamics. The first is Glauber dynamics, where the orientations are unconserved. A lattice site is chosen at random, a new orientation is proposed for the site, a ΔE is computed, and the change is accepted or rejected depending on a probability transition function $P(\Delta E)$. The second method deals with situations where the volume fraction of each orientation type is conserved; it is called Kawasaki dynamics. Here, a lattice site is chosen, a neighboring site is chosen, and then a swap of the orientations is proposed. Again, the ΔE is computed, and the change is accepted or rejected depending on $P(\Delta E)$.

Both Glauber and Kawasaki dynamics require the definition of the probability transition function. There are two common choices: the metropolis function:

$$P(\Delta E) = \begin{cases} 1 & \text{if } \Delta E \leq 0 \\ \exp\dfrac{-\Delta E}{kT_s} & \text{if } \Delta E > 0 \end{cases} \qquad \text{(Eq 4)}$$

and the symmetric function:

$$P(\Delta E) = \frac{1}{2}\left\{1 - \tanh\frac{\Delta E}{2kT_s}\right\} \qquad \text{(Eq 5)}$$

where kT_s defines a thermal energy of the simulation; it is analogous to the thermal energy of experimental systems but is not directly related. The choice of the probability function has no effect on the thermodynamics of the system, although the choice of the functional form of $P(\Delta E)$ does slightly affect the dynamics of boundary motion.

The time required to attempt a single orientation swap, whether successful or unsuccessful, is defined arbitrarily as τ. On average, it takes N_τ to visit each site on the simulation lattice once; this is defined as one Monte Carlo time step, or 1 MCS (Ref 4).

Simulation Variables. In 2-D simulations, hexagonal or square lattices are used to discretize space, as shown in Fig. 3. In a simple square
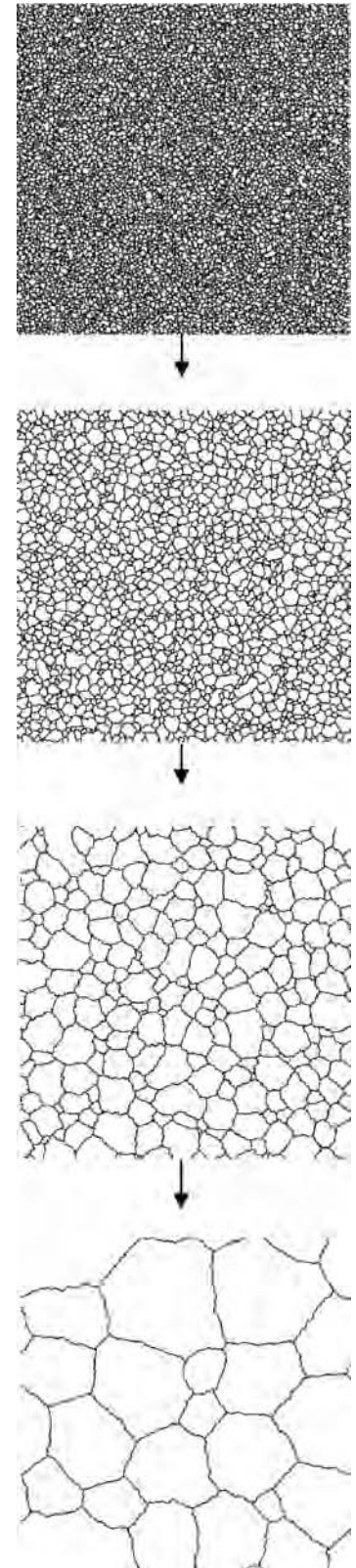
Fig. 2 Microstructural evolution of an initially random distribution of spins on a two-dimensional square lattice using the Potts model, periodic boundary conditions, metropolis spin dynamics, and $kT_s = 0.5$. The initial configuration of spins was set by allocating each lattice a random spin of $s_i \in \{0, Q\}$.
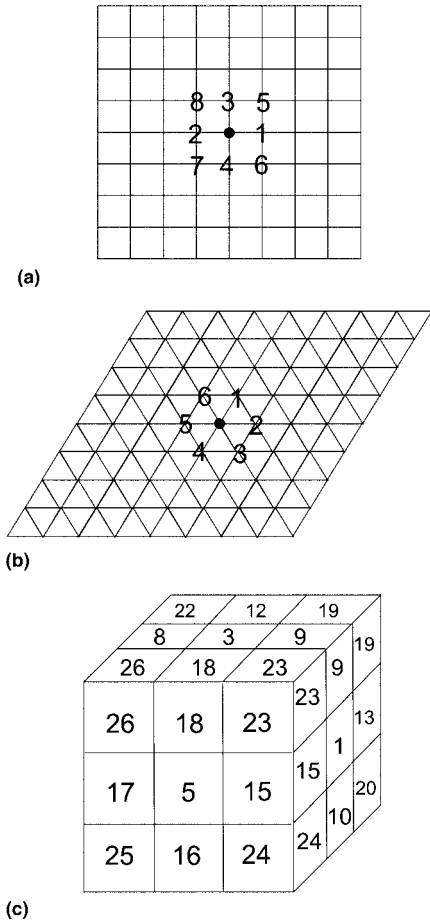
Fig. 3 Different types of lattice and the neighbor co-ordination used in the Potts model. (a) Two-dimensional (2-D) square lattice. (b) 2-D triangular lattice. (c) Three-dimensional simple cubic lattice

Fig. 4 Different types of boundary conditions used in the Potts model. (a) Surface boundary condition. (b) Mirror boundary condition. (c) Periodic boundary condition. (d) Skew-periodic boundary condition

lattice, a site may be defined to have only the four first-nearest neighbors, labeled 1 to 4 in Fig. 3(a), but more often, the eight first- and second-nearest neighbors, shown in Fig. 3(a), are used. In the triangular lattice, the six first-nearest neighbors, shown in Fig. 3(b), are sufficient. In 3-D, the simple cubic lattice is commonly used with the 26 first-, second-and third-nearest neighbors, as shown in Fig. 3(c). The choice of neighbor co-ordination strongly affects the type of boundary shapes that are favored in the system. In 2-D square lattices, boundaries with planes 0°, 45°, and 90° are strongly favored and can cause facets to form. These low-energy facets obviously have a corollary with the faceting of atomic planes, and although they can be studied by similar techniques, they are problematic when using the model to simulate isotropic boundary shapes, because they impose an anisotropy into the simulations. There are a number of ways of mitigating against these lattice effects. First, some lattices and neighbor co-ordinations have less intrinsic anisotropy. For instance, the triangular lattice in 2-D, using the six first neighbors, has

the lowest anisotropy of any 2-D regular lattice (Ref 7), and the simple cubic lattice, using 26 nearest neighbors, is the most effective in 3-D. Another method to mitigate against the unwanted influence of the lattice is to not have one at all by using a random lattice (Ref 8).

There exists a set of lattice sites that lie at the boundaries of the simulation area/volume. These sites must be treated differently than the internal sites of the lattice. In the case where the boundaries represent the edge of the simulated system, these sites will have fewer nearest neighbors. They may also have a solid/liquid or a solid/gas surface energy associated with them, or they may have a certain concentration of solute imposed on them as a boundary condition (Fig. 4a). Another type of commonly used boundary condition is that of mirror boundary conditions. In this case, the sites on a boundary are mirrored, so that the neighbor shell of the boundary site is comprised of those defined by a reflection transformation (Fig. 4b).
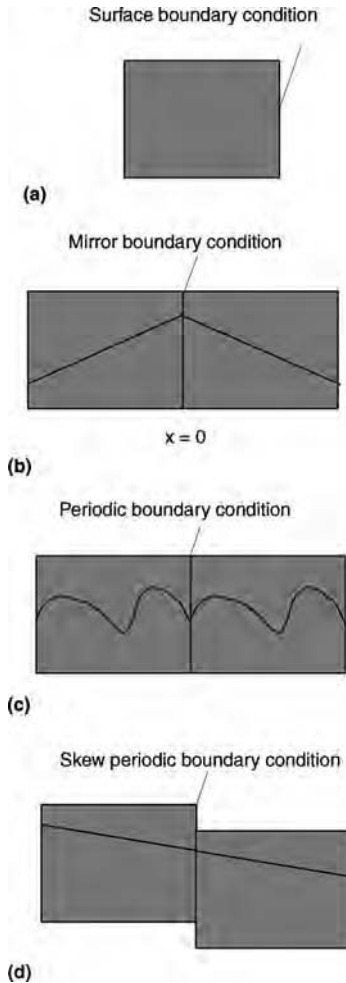
For example, in a 2-D square lattice, a mirror boundary condition applied at the $x = 0$ boundary simulates the effect that the orientations for all sites $x < 0$ exactly mirror those for $x > 0$. Perhaps the most popular type of boundary condition is a periodic boundary condition. In this case, the edges of the simulation effectively wrap around and contact the opposite edges, creating a toroidal simulation area in the 2-D case. These boundary conditions are relatively simple to impose; it simply involves assigning the nearest neighbors of each edge site to the opposite boundary edge (Fig. 4c). The boundary conditions for the simulation volume are usually implicitly encoded in the function that returns the neighbor sites of each site. A variant of the periodic boundary condition is the skew-periodic boundary condition. In this case, the edges of the simulation wrap around, but at the boundary, a vector displacement parallel to the boundary is imposed (Fig. 4d). This type of boundary condition is used when simulating flat boundaries that have a nonperpendicular intersection angle with a simulation area/volume boundary.

Boundary Energy. So far, boundaries have been treated as if they are isotropic interfaces, and Fig. 2 shows the general result of simulations carried out with such underlying assumptions. To perform simulations relevant to experimental systems, γ needs to be implemented as a function of the crystallographic misorientation across the boundary. This is relatively straightforward because it only requires the modification of Hamiltonian. The simplest way to begin to understand how these systems behave is to introduce a new identifier that, along with the orientation, is associated with each lattice site. Thus, each lattice site has both an orientation identifier, s_i, and a component identifier, η_i. The component identifier carries information necessary to calculate the anisotropic nature of the boundary, while the orientation identifier continues to be used to calculate whether a particular site is on a boundary. The reason for not considering the full crystallographic complexity of real grain structures at this stage is that even this small increase in complexity has an enormous impact on microstructural evolution of the systems, and so, it is worth observing such effects. In such a system, the Hamiltonian becomes:

$$E = \sum_{i=1}^{N} \sum_{j=1}^{z} \gamma(s_i, s_j, \eta_i, \eta_j) \qquad \text{(Eq 6)}$$

This Hamiltonian represents only the boundary energy of the system and so implicitly assumes that the bulk free energy of each component is the same, but their boundary energy is different.

This changes the equilibrium condition at the nodes (where three or more boundaries meet). If these are to remain in a state of local equilibrium and maintain the equilibrium angles defined by the boundary energies, then
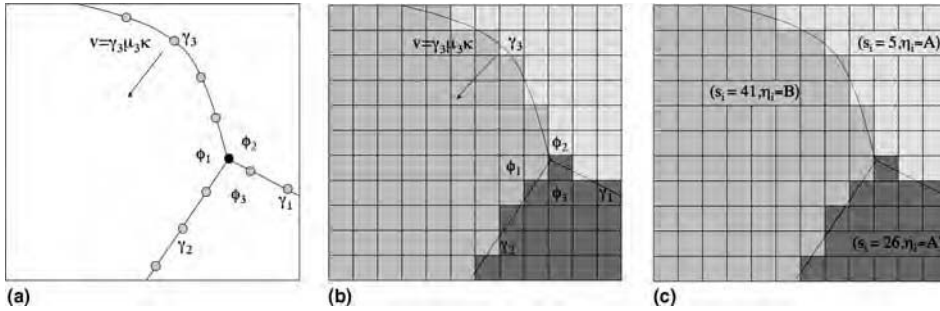
Fig. 5 Relationship between boundary energy and node angle. (a) Continuum system. (b) Monte Carlo Potts model. Each grain orientation is represented by a different gray scale; the boundaries are sharp, being implicitly defined between sites of different orientations. (c) Implementation of components and spins into the model

neglecting torque forces, the angles at the nodes in 2-D are given by the relation:

$$\frac{\gamma_1}{\sin\phi_1} = \frac{\gamma_2}{\sin\phi_2} = \frac{\gamma_3}{\sin\phi_3} \qquad \text{(Eq 7)}$$

where γ_i are boundary energies, and ϕ_i are the angles at the triple point, as illustrated in Fig. 5(a). What this means in practice is that triple points are no longer thermodynamically constrained to be 120°. Not just that, but triple points in 2-D and quadijunction points in 3-D are no longer the only stable node configurations. This makes possible a vast array of different boundary network morphologies. Figure 5(b) shows how a node is represented in the Potts model on a square lattice. Note that the triple-point angles are discrete quantities that depend not just on boundary energies but also on the type of lattice.

The simplest such system that can be considered is a polycrystal with only two types of component, A and B; three types of boundary, A-A, B-B, and A-B; and three boundary energies, J_{AA}, J_{BB}, and J_{AB}, which gives:

$$\gamma(s_i, s_j, \eta_i, \eta_j) = \begin{cases} 0 & \text{in the grain interiors} \\ & (s_i = s_j, \eta_i = \eta_j) \\ \frac{J_{AA}}{2} & \text{for A-A boundaries} \\ & (s_i \neq s_j, \eta_i = \eta_j = A) \\ \frac{J_{BB}}{2} & \text{for B-B boundaries} \\ & (s_i \neq s_j, \eta_i = \eta_j = B) \\ \frac{J_{AB}}{2} & \text{for A-B boundaries} \\ & (s_i \neq s_j, \eta_i \neq \eta_j) \end{cases} \qquad \text{(Eq 8)}$$

The behavior of the system can be easiest understood by considering the the dimensionless parameters $R_A = J_{AA}/J_{AB}$ and $R_B = J_{BB}/J_{AB}$. When $R_A = R_B = 1$, the system is isotropic. When $R_A = R_B > 1$, the relative energy of the A-B boundaries decreases in relation to the A-A and B-B boundaries; thus, during evolution, the system will try to minimize the area or length of A-A and B-B boundaries in favor of A-B boundaries and so minimize the energy of the system. Figure 6 shows how the system self-organizes into a mosaic structure that minimizes the length of A-A and B-B boundaries. The mosaic structure is itself then able to coarsen in a self-similar manner. Figure 7

shows the equilibrium structures formed under a variety of other conditions. If $R_A = R_B > 1$, the A-A and B-B boundaries are favored over A-B boundaries, and the system self-orders to segregate the A and B components and thus minimize boundary energy. Figure 7(b) shows such a structure, which orders the A component and the B component into separate enclaves and can be contrasted with Fig. 7(a), which shows the random distribution of A and B components that comes in the isotropic case when $R_A = R_B = 1$. Figure 7(c) shows what happens when $R_A > R_B$; the system gets rid of the high-energy A-A boundaries altogether. Figure 7(d) shows another example of mosaic-type structures that are formed when $R_A = R_B < 1.0$. Figure 7(e) shows another example of $R_A > R_B$, but this time where $\gamma_{BB} = \gamma_{AB}$. Here, the A-component grains are not removed, because it is only the A-A boundaries that are high energy; however, they do become an isolated component. Figure 7(f) shows the effect of using $kT_s = 0$ with anisotropic energies. Because of the high lattice pinning present, the structure shows a very high degree of planarity in the low-boundary planes that are at 45° to the simulation lattice.

Note that in many of these 2-D anisotropic energy systems, four grain junctions (quadrijunctions) are possible. The angles at which boundaries meet in a quadrijunction are not uniquely determined by an energy balance. Instead, the angle of a grain corner in a stable quadrijunction must be greater than or equal to the angle of the same corner in the triple junction formed when the quadrijunction fluctuates into two triple junctions. This angular flexibility has an important effect on the kinetics. Systems in which quadrijunctions are unstable undergo normal grain growth. When quadrijunctions are stable (due to the ratios of R_A and R_B), grain growth can stop due to the flexibility of these junctions to change their angles and thus eliminate boundary curvature (Ref 9). Both greater number of texture components and more realistic energy functions, such as the Read-Shockley function, can be incorporated into the model, as is shown later.

Boundary Mobility. To simulate the case where the mobility is also a function of the boundary character, $\mu(s_i, s_j, \eta_i, \eta_j)$, the probability
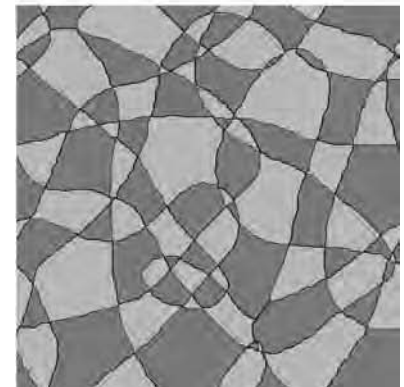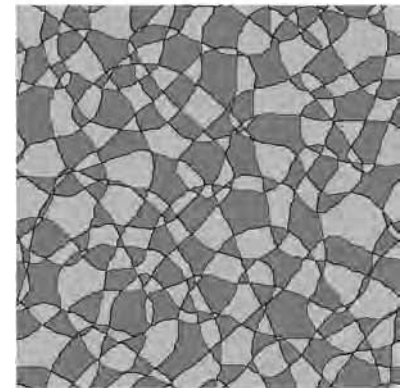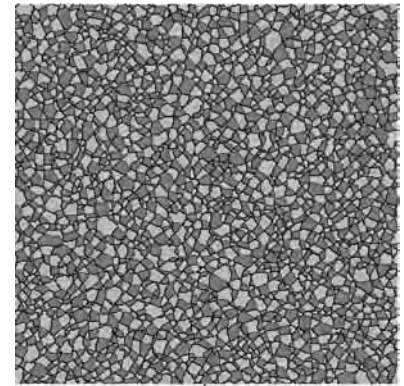
Fig. 6 Evolution of microstructure during a Potts model simulation of a two-component system in which the initial distribution of components is equal and $R_A = R_B = 0.5$. The A and B components are differentiated by the gray scale. The simulation was performed using a square (1,2) lattice, Glauber dynamics, metropolis transition probability function, and $kT_s = 0$.

transition function is modified so that the probability of a orientation swap is proportional to the mobility of that boundary. The metropolis probability transition function then becomes:

$$P(\Delta E) = \begin{cases} p_0 & \text{if} \quad \Delta E \leq 0 \\ p_0 \exp\frac{-\Delta E}{kT} & \text{if} \quad \Delta E > 0 \end{cases} \qquad \text{(Eq 9)}$$

where $p_0 = \frac{\mu(s_i, s_j, \eta_i, \eta_j)}{\mu_m}$, and μ_m is the maximum mobility in the system.

For the simplest model system with two components, A and B, p_0 is reduced to a simple binary function:

$$\mu(s_i, s_j, \eta_i, \eta_j) = \begin{cases} 0 & \text{in the grain interiors} \\ & (s_i = s_j, \eta_i = \eta_j) \\ M_{\text{AA}} & \text{for A-A boundaries} \\ & (s_i \neq s_j, \eta_i = \eta_j = \text{A}) \\ M_{\text{BB}} & \text{for B-B boundaries} \\ & (s_i \neq s_j, \eta_i = \eta_j = \text{B}) \\ M_{\text{AB}} & \text{for A-B boundaries} \\ & (s_i \neq s_j, \eta_i \neq \eta_j) \end{cases}$$

(Eq 10)

When $M_{\text{AA}} = M_{\text{BB}} = M_{\text{AB}}$, the system becomes isotropic, and normal grain growth is observed. In the case where $M_{\text{AA}} = M_{\text{BB}} < M_{\text{AB}}$, the A-B boundaries have a higher mobility than the A-A and B-B boundaries, which means that isolated A or B grains grow rapidly if they have a size advantage over their neighbors. Equally, if they do not have a size advantage and are not favored to grow, the mobility advantage acts to shrink these grains. In systems where one component is initially in the minority, this leads to a phenomenon called abnormal grain growth, where the minority component grows to become the majority component (Fig. 8). More sophisticated mobility functions can be incorporated into the model to simulate the effect of temperature gradients, because mobility of interfaces is often a function of temperature.

Note that increasing temperature does not increase the mobility of the boundaries. This is obvious from the form of probability transition functions, where the intrinsic mobility of the boundaries is set at $kT_s = 0$. Thus, temperature just serves to roughen the boundaries and so prevent lattice effects. How then is the effect of "real" temperature simulated? It is done by making the intrinsic mobility $\mu(s_i, s_j)$ a function of real temperature, T_r.

Pinning Systems. Figure 9 shows the evolution of the microstructure for a Potts model simulation using Glauber dynamics for component A and where component B was inert. Note how, in the initial stages, the microstructure of component A evolves, but then boundary motion becomes increasingly slow until the microstructure becomes pinned. The incorporation of an inert component does not require any modification of the Hamiltonian of the system; it requires only the modification of the Potts algorithm so that sites of the inert component never undergo orientation or component swaps. Note that there is no explicit implementation of any pinning mechanism or pinning force. The pinning is a complex emergent phenomenon that arises out of the interaction of many parts of the system (Ref 10). By using Kawasaki dynamics, coarsening of second-phase arrays can be simulated. Once again, the phase volume fractions are conversed, but unlike the pinning simulations, the A-A, A-B, and B-B boundaries have different energies.

The theory of pinning and the ability of Potts model simulations to capture the physics of real
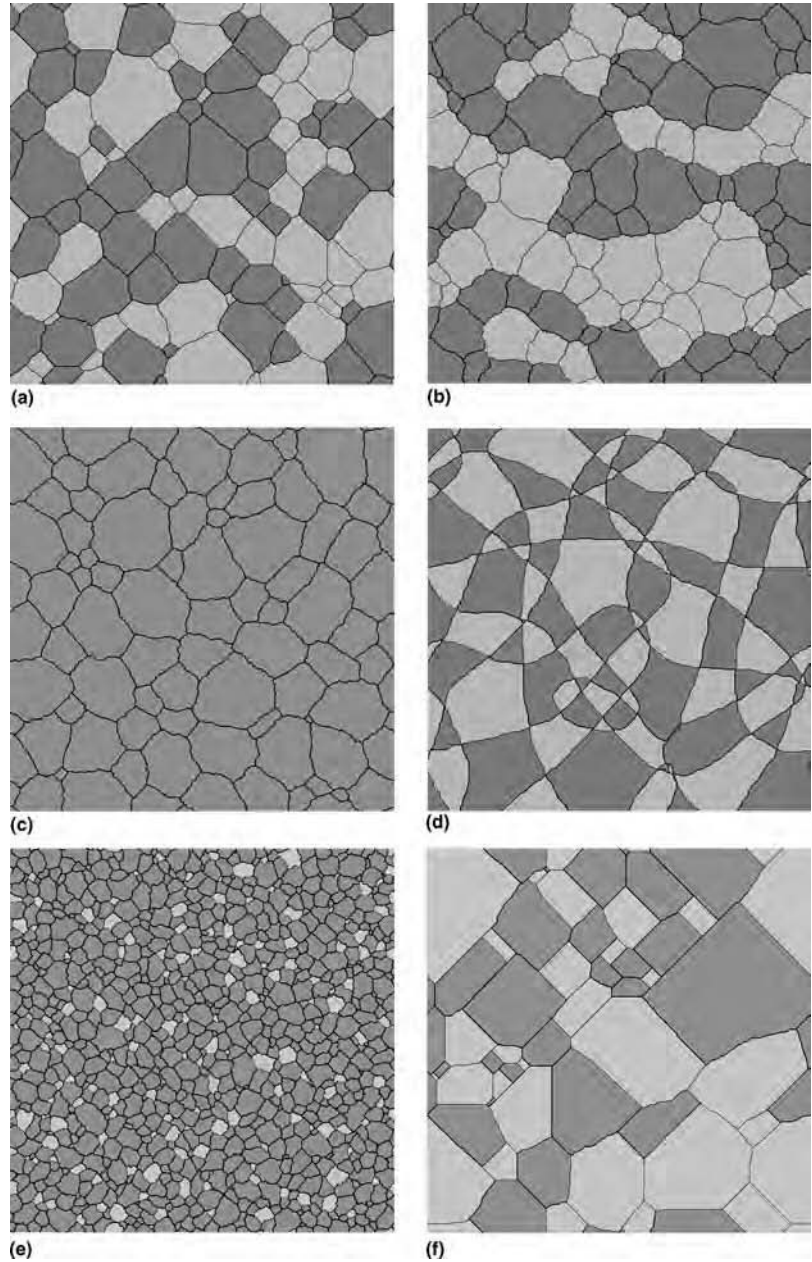
Fig. 7 Effect of anisotropic boundary energy on microstructure during grain growth of a two-component system in which the initial distribution of the A and B components was equal. (a) $R_A = R_B = 1$, $kT_s = 0.75$; the isotropic case. (b) $R_A = R_B = 1.5$, $kT_s = 0.75$. (c) $R_A > R_B$, $R_A = 1$, $R_B = 0.67$, $kT_s = 0.75$. (d) $R_A = R_B = 0.5$, $kT_s = 0$. (e) $R_A > R_B$, $R_A = 1.5$, $R_B = 1$, $kT_s = 0.75$. (f) $R_A = R_B = 1.3$, $kT_s = 0$. The simulations were performed using a square (1,2) lattice, Glauber dynamics, and metropolis transition probability function.

pinned systems are dealt with more fully later. For the moment, it should be noted that incorporating pinning phases into the model is relatively straightforward, and many different pinning-phase morphologies can be modeled. Figure 10 shows the wide range of pinning-phase morphologies and pinned structures possible in Potts model simulations. However, performing simulations at nonzero T_s allows boundaries to depin from particles thermally, especially when they are small. Also, it should be noted that the physics of pinning are

dimensionally dependent, which means that 2-D and 3-D pinning are very different phenomena (Ref 10, 11). This is important when using these simulation techniques to understand real microstructural phenomena such as Zener-Smith pinning.

Stored Energy. The Potts model can be used to simulate the effect of a volume stored energy. In this case, the state of each cell is described by a crystallographic orientation and scalar variable that describes the stored energy. Thus, the Hamiltonian for the system becomes:
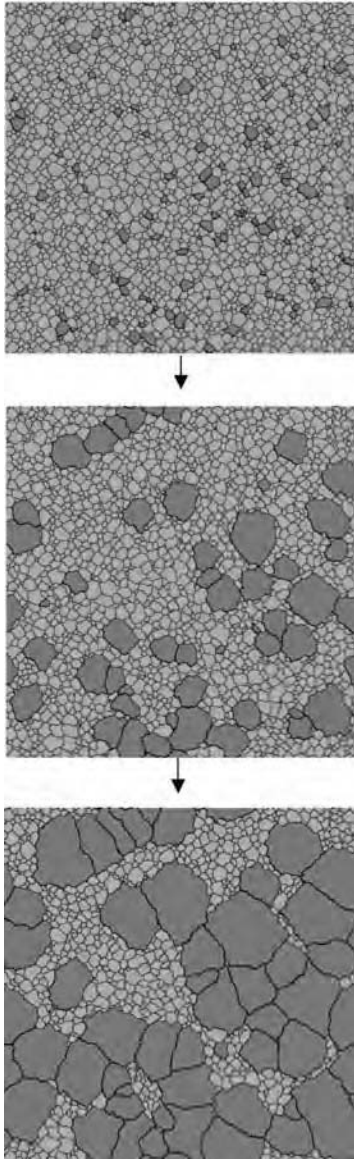
Fig. 8 Evolution of microstructure during a Potts model simulation of a two-component system in which the initial distribution of components is unequal and the A-B boundaries have a mobility advantage: $f_B = 0.05$, $M_A = M_B = 1$, $M_{AB} = 100$. The A and B components are differentiated by the gray scale. The simulation was performed using a square (1,2) lattice, Glauber dynamics, metropolis transition probability function, and $kT_s = 0.75$.

$$E = \sum_{i=1}^{N} \sum_{j=1}^{z} \gamma(s_i, s_j) + h_i \qquad \text{(Eq 11)}$$

where h_i represents the stored energy term for each site and:

$$\gamma(s_i, s_j) = \begin{cases} 0 & \text{in the grain interiors} \\ \frac{J}{2} & \text{at the grain boundaries} \end{cases} \qquad \text{(Eq 12)}$$

In the simplest system, every site is allocated the same stored energy, $h_i = H$, and so, the ratio of H/J determines the behavior of the system. For low values of H/J, normal grain growth is
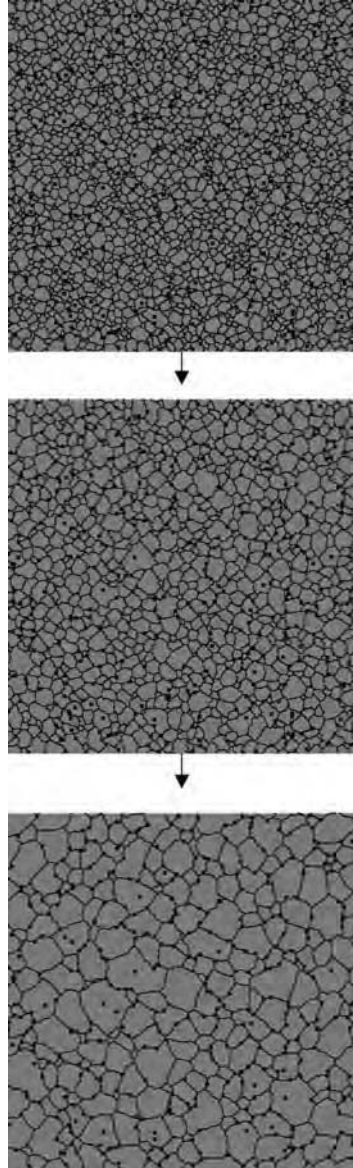
Fig. 9 Potts model simulation carried out on a square lattice, using Glauber dynamics and $kT_s = 0.75$. The second phase has an unchangeable index and so pins the primary phase. The simulations were performed using a square (1,2) lattice, Glauber dynamics, metropolis transition probability function, and $kT_s = 0.75$.

observed, but as H/J increases, the stored energy term dominates. Figure 11 shows the microstructural evolution of such a system in which each grain in an initial equiaxed grain structure is assigned a uniform level of stored energy. As the boundaries move, driven both by the stored energy driving force and the curvature driving force, they sweep out areas of zero stored energy. This enhanced boundary motion gives some grains an advantage, and they grow to consume the others. Once all the stored energy is consumed, the system reverts to normal grain-growth behavior. The ratio of the stored energy and the interfacial energy is a crucial variable in the system. Differential

boundary energies and mobilities can also be employed. If the stored energy of the system is very large compared to the interfacial energy, then boundary growth becomes chaotic.

Incorporating Experimental Parameters into the Potts Model

Textures and Misorientation Distributions. It is essential to capture real textures and misorientation distributions in the Potts model if it is to simulate the behavior of real experimental systems. For 2-D simulations, the most straightforward way of doing this is to incorporate the microstructural information and the crystallographic information directly from the electron backscatter diffraction (EBSD) data set. Typically, this means that each lattice site in the simulation is allocated a unique orientation number, and a table is created that correlates the orientation number with the Euler angles corresponding to the crystallographic orientation of the grain. A normal Potts model simulation can then be performed, with the crystallographic information of each lattice site being used to plot the evolution of microstructure and development of textures. Now, the sites no longer are allocated integers but instead a discrete crystallographic orientation, O_i. In such as system, the Hamiltonian becomes:

$$E = \sum_{i=1}^{N} \sum_{j=1}^{z} \gamma(O_i, O_j) \qquad \text{(Eq 13)}$$

Although this modification seems straightforward enough, there are some important issues. First, in such EBSD maps there is a good deal of noise that corresponds to some pixels being identified as single-site grains when, in fact, they are more likely to be a site whose orientation was incorrectly measured. This kind of noise may not be due just to incorrect measurement; in the case of deformed microstructures, the dislocation and other defects may be associated with low-angle subboundaries. Also, because the map is a 2-D section of a 3-D microstructure, some topological features may appear to be noise when, in fact, they are the tip of a grain protruding into the 2-D section. For these and many other reasons, importing a microstructure into a Potts model often requires a clean-up filter to be applied so that these effects can be mitigated and features that are not going to be included in the model can be removed. However, it is obvious that using these filters can also distort the data set, in some cases changing the fundamentals of the microstructure to be studied.

These 2-D microstructure maps are extracted from a small volume of the material. This volume is represented in the model only through the boundary conditions. Thus, choice of the boundary conditions is important when performing simulations and also when interpreting the results. Choosing periodic boundary
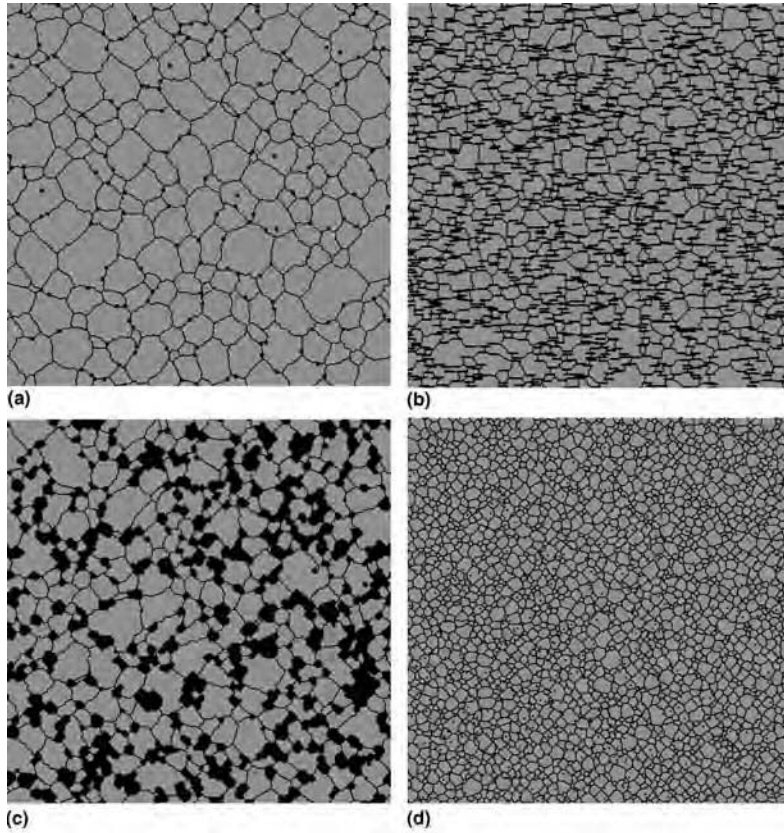
Fig. 10 Illustrating the different types of pinning-phase morphology that can be simulated in Potts model simulations. (a) Square particles (2 × 2 sites). (b) Aligned rods (10 × 1 sites). (c) Immobile grains (average size 10 sites). (d) Dispersoids (1 site). The simulations were performed using a square (1,2) lattice, Glauber dynamics, metropolis transition probability function, and $kT_s = 0.75$.

conditions is not an option, because there will not be continuity across the simulation boundaries. The choice of mirror or free-surface boundaries is available; both have implications. Furthermore, the fact that a 2-D simulation is being performed of a 3-D phenomenon must also be taken into account. The upshot of these factors is that extreme care should be taken when carrying out and interpreting such simulations. The best practice is to carry out a large number of simulations using a large number of different input microstructures and to measure the evolution of average characteristics, for example, the average texture, misorientation distribution function (MDF), grain size, and so on. It is when these averaged quantities are compared with experimental results that meaningful conclusions and predictions may be drawn.

There are no routine methods for extracting the necessary 3-D information from experiment. It is possible to combine EBSD with serial sectioning, but this is a very labor-intensive task and still leaves the problem of how to extrapolate between the sections. The 3-D x-ray tomography methods have recently become possible using high-energy focused synchrotron x-ray sources (Ref 12), but, at the moment, the resolution is low, and again, the method is not widely available.

Another approach to this problem is to use computation methods to reconstruct an equivalent 3-D microstructure with the grain size, grain size distribution, texture, and MDF, because obtaining these characteristics of the 3-D microstructure from experiment is straightforward (Ref 13). The first step is to obtain a 3-D microstructure with the right grain size and grain-size distribution. This is done by using a 3-D Potts model and anisotropic mobility to grow an appropriate microstructure using trial and error. This is easy for equiaxed microstructures and less easy for more complicated microstructures. Next, the experimental texture is discretized into Q orientations and allocated randomly to the orientations of the grains of the 3-D microstructure. This produces a 3-D microstructure with the correct texture but random MDF. This MDF is calculated and quantized into n_b bins, such that S_k is the number of boundaries with misorientations between $k\Delta\theta$ and $(k + 1) \Delta\theta$, $k = 0, 1, \ldots, n_b$. A system Hamiltonian is defined as the sum of the squared differences between S_k^m and S_k^{exp}:
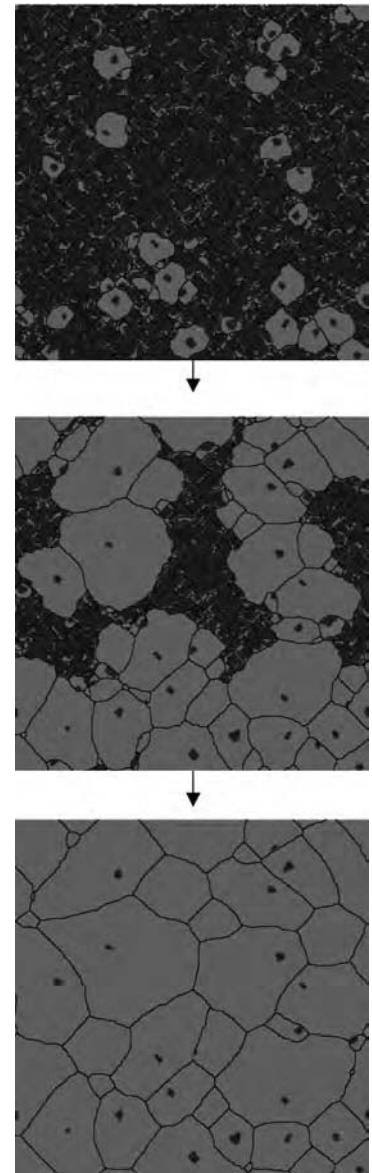
Fig. 11 Effect of stored energy on an evolving microstructure. An initial equiaxed grain structure is assigned uniform levels of stored energy (H/J = 0.8), indicated by a gray scale. Strain-assisted boundary motion occurs, which mimics abnormal grain-growth behavior. The simulation was performed using a square (1,2) lattice, Glauber dynamics, metropolis transition probability function, and $kT_s = 0$.

$$H_{mdf} = \sum_{k=0}^{k=n_b} \left(S_k^m - S_k^{\mathrm{exp}}\right)^2 \qquad \text{(Eq 14)}$$

where S_k^m defines the MDF of the model, and S_k^{exp} defines the experimental MDF. H_{MDF} is a state variable providing a measure of the difference between the model MDF and the experimental MDF. It is equal to zero when the model MDF and the experimental MDF are identical. A Monte Carlo algorithm can be used to minimize H_{MDF} and, in doing so, construct the desired MDF. The method is as follows.

Two grains are chosen at random, and the H_{MDF} due to swapping their orientations is calculated. The probability $p(H_{MDF})$ that the swap is accepted is a metropolis function (Ref 13).

This swap method is effective and produces a starting 3-D microstructure with a texture and MDF that are identical to the experiment (Fig. 12). It is not elegant. More ambitious ways of reconstructing 3-D microstructures from 2-D metrics that integrate the microstructure generation, texture generation, and MDF optimization steps into one step have been proposed. Unfortunately, none yet have been shown to work. Progress on 3-D x-ray methods may make the swap method redundant in the future (Ref 12).

Incorporating Realistic Energies and Mobilities. Read and Shockley (Ref 14) derived an analytical expression for the energy (per unit area) of a low-angle grain boundary comprised of a regular array of dislocations. The boundary energy can be expressed as a function of the misorientation:

$$\gamma = \gamma_0 \theta (A - \ln \theta) \qquad \text{(Eq 15)}$$

The parameters γ_0 and A are related to elastic constants and properties of the dislocation cores: γ_0 sets the overall energy scale, and A adjusts the angle of the maximum grain-boundary energy. For large-angle grain boundaries, this model would not be expected to be valid, because the dislocation cores would overlap substantially, and their interaction could not be neglected. Nevertheless, this formula has been successfully fit to experimental grain-boundary energies for wide misorientation angles. Thus, a normalized version of Eq 15

can be used to model the functional form of a general grain boundary in the Potts model:

$$J_{RS} = J_0 \left(\frac{\theta}{\theta_m} \right) \left\{ 1 - \ln \left(\frac{\theta}{\theta_m} \right) \right\} \qquad \text{(Eq 16)}$$

where θ_m is the misorientation angle that results in the maximum boundary energy of the system. Experimentally, it is observed to lie between 10° and 30°, depending on the system (Ref 15). The grain-boundary energy function for the Potts model then becomes:

$$\gamma(O_i, O_j) = \begin{cases} 0 & \text{in the grain interiors } (O_i = O_j) \\ J_{RS}/2 & \text{for boundaries } (O_i \neq O_j) \end{cases}$$
$$\text{(Eq 17)}$$

Clearly, in most real systems, mobility is also a function of the boundary character: $\mu(O_i, O_j)$. Thus, the probability transition function must be modified so that the probability of an orientation swap is proportional to the mobility of that boundary. The metropolis probability transition function then becomes:

$$P(\Delta E) = \begin{cases} p_0 & \text{if } \Delta E \leq 0 \\ p_0 \exp(-\Delta E/kT_s) & \text{if } \Delta E > 0 \end{cases} \quad \text{(Eq 18)}$$

where $p_0 = \mu(O_i, O_j)/\mu_m$, and μ_m is the maximum mobility in the system. Note that these are normalized mobilities measured from experiment and have a wide range of functional forms. Obtaining these data from experiment is often nontrivial, and, like boundary energies, these mobilities may also be a function of boundary plane (especially in the case of twins) as well as composition.

Although the implementation of the Read-Shockley energy function seems to be a straightforward extension of the model to change the boundary energy, it has another implicit effect, which is to change the node angles of the boundaries. This changes the boundary curvature acting on a boundary and thus the driving force on that boundary. If systems with a continuous range of boundary energies, and thus a continuous range of node angles, are to be simulated, the discrete nature of the simulation lattice must not affect these angles.

One way to check this is to consider a model geometry, such as that shown in Fig. 13 of a system with a constant driving force for motion and in which the triple points have invariant geometry. A similar approach is taken by experimentalists studying boundary and triple-point mobility (Ref 15). The grain structure is columnar, with two grains, B and C, capped by a third grain, A. Boundary conditions are periodic in the x-direction and fixed in the y-direction. There are two boundary misorientations in the system: θ_1 is the misorientation angle of the A/B and A/C boundaries, and θ_2 is the misorientation angle of the B/C boundaries. There are two triple junctions in the system, and the geometry is arranged such that these two are identical and symmetric: from Eq 7, the equilibrium junction angle where θ_1 is the energy of the A/B and A/C boundaries, and θ_2 is the energy of the B/C boundaries.

The driving force acting on the boundary is γ_2/D. Assuming that the driving force is proportional to the velocity of the boundary, the boundary velocity in the y-direction is:
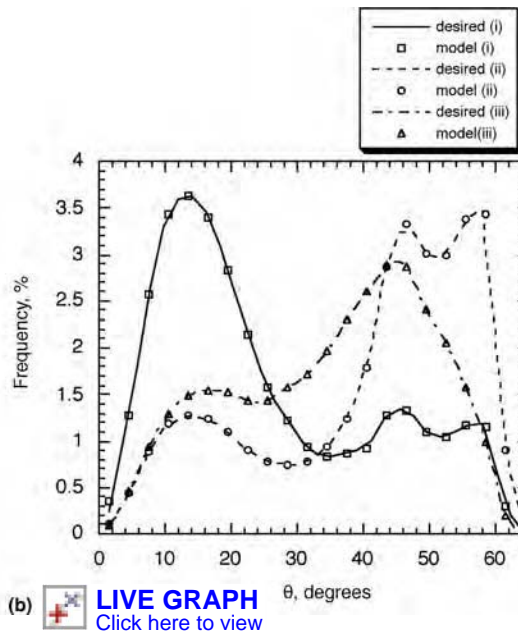
Fig. 12 (a) Three-dimensional equiaxed microstructure grown using the Potts model. (b) Graph showing the desired and the achieved misorientation distribution functions (MDFs) generated by discretizing a texture, allocating orientations to the grains, and then using the swap method to achieve the desired MDF. Source: Ref 10
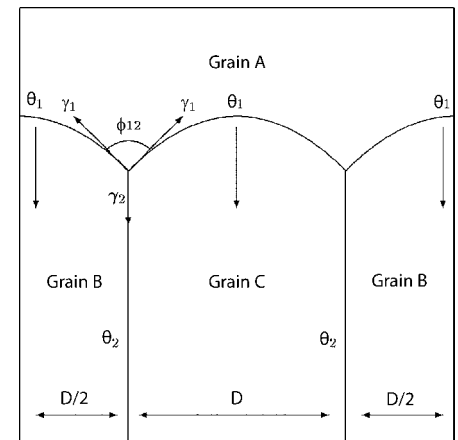
Fig. 13 Boundary geometry used to validate the Q-state Potts model for anisotropic grain growth. Boundary conditions are continuous in the x-direction and fixed in the y-direction. The boundary between grain A and grains B and C is the only boundary that moves. θ_1 is the misorientation between grains A and B and also between grains A and C. θ_2 is the misorientation between grains B and C. The equilibrium angle of each triple point, ϕ_{12}, is defined by the ratio of the boundary energies that intersect at the triple point, $\gamma(\theta_1)$ and $\gamma(\theta_2)$.

$$\frac{dy}{dt} = \frac{\mu_1 \gamma_2}{D} \qquad \text{(Eq 19)}$$

where μ_1 is the intrinsic mobility of the A/B and A/C boundaries.

To examine the validity of the Q-state Potts method, a nominal γ_2 is set, and dy/dt is measured with time. By finding the regime in which dy/dt is constant and using Eq 19, the effective γ_2 can be extracted. Figure 14(a) compares the measured γ_2 to the nominal γ_2. It can be seen that for large γ_2 (i.e., high misorientations), there is good agreement between the simulation and the theory. However, as γ_2 decreases, a deviation from theory is observed; the effective γ_2 becomes constant as the nominal γ_2 continues to decrease. This deviation occurs around $\gamma_2 = 0.4\gamma_1$, corresponding to a misorientation angle $\theta_2 = 2°$ when $\theta_1 = 15°$.

This behavior has its origin in the discrete nature of the lattice. As θ_2 becomes smaller relative to θ_1, the equilibrium triple-junction angle, ϕ_{12}, becomes larger until it approaches 180° and the A/B and A/C boundaries become nearly flat. Because junction angles must be changed by the addition or removal of an entire step in a discrete lattice, small differences in the junction angle cannot be resolved. That is, at some point, the last step is removed, the boundary becomes flat, and the triple-junction angle cannot change with further decreases in γ_2. Because the triple-junction angle defines boundary curvature, it also defines the driving force. Thus, if this angle becomes invariant at some γ_2, so does the driving force acting on the boundary. This effect is unavoidable in these discrete lattice simulations and hence is a limit to the range of anisotropies that the model can simulate. For simulations on the square lattice, the limit is reached at approximately $\gamma_2 = 0.4\gamma_1$, when $\phi_{12} = 157°$; larger triple-junction angles cannot be resolved.

Note that this effect limits only the maximum triple-junction angle and thus the range of boundary energies (anisotropy) that may be resolved. It does not limit the absolute value of the boundary energy. For example, a system of $\theta = 1°$ boundaries, each with energy $\gamma = 0.25$, has 120° triple junctions and can be successfully simulated by the Q-state Potts model. The triple-junction limitation must be considered only if a higher angle boundary (in this case, $\theta \geq 4°$) must be included in the system.

The limitation on energetic anisotropy does not affect the model ability to simulate nonuniform boundary mobility. Because mobility is independent of curvature, it is unaffected by triple-junction angles. Figure 14(b) shows the linear relationship between mobility and velocity in the Q-state Potts model over 4 orders of magnitude (Ref 16).

Applications

Grain Growth. In 2-D networks formed through the action of the minimization of isotropic surface energy, the average number of boundaries per grain is six. Therefore, the only stable network is a hexagonal array of grains, where each grain has six neighbors, and the 120° vertex angles at the triple points can be satisfied by straight boundaries. These boundaries having no curvature have no net force acting on them and so remain static. Any networks that deviate from this regular array inevitably contain some grains with less than six sides and some with more than six sides. If the triple points maintain their 120° angles, then the array must contain curved boundaries. Curvature-driven migration then causes the system to evolve. The boundaries of grains with less than six sides are concave (curved toward the center of a grain), and so, boundary migration makes these grains shrink. Grains with more than six sides have convex boundaries, and so, these grains grow. In other words, the competing requirements of space filling and surface tension cause large grains to grow and small grains to shrink. This forms the basis of a remarkable law proposed by von Neumann (Ref 3), which states that the growth rate of a 2-D cell with area, A, and N_s sides is given by:

$$\frac{dA}{dt} = c(N_s - 6) \qquad \text{(Eq 20)}$$

where c is a constant. This result has been shown to be correct for both 2-D soap froths and 2-D grain structures (Ref 17) and has recently been generalized to 3-D (Ref 18).

Although it is obvious that the self-organizing behavior of the Q-state Potts model resembles the phenomenon of grain growth, as shown in Fig. 2, the question arises: How closely do the simulations compare to the experimental measurements of grain growth? Experimentally, it is observed that the grain-size distribution, when normalized by the average grain size, remains constant during grain growth. This means that even though some grains grow, while others shrink, the grain ensemble remains self-similar. This type of phenomenon is called normal grain growth. The grain-size distribution and the topological distribution derived from 3-D Potts model simulations of isotropic grain growth are also observed to be time-invariant and in agreement with experimental data (Ref 5).

The rate at which the average size increases is another parameter by which experimentalists measure normal grain growth. The kinetics of grain growth are characterized by the parabolic equation:

$$R_{av}^n - R_0^n = A_{gg}t \qquad \text{(Eq 21)}$$

where R_0 is the initial grain size, and A_{gg} is a constant. The grain-growth exponent, n, has been the focus of much of the debate in the grain-growth community. Hillert's theoretical derivation (Ref 19) gives $n = 2$, but most experiments show grain-growth exponents much greater than this; typically, the values lie between $n = 2.5$ and $n = 4$. It has been argued that impurity effects may be responsible for the deviation from the ideal value. However, even data from a wide range of ultrapure metals show considerable deviation from $n = 2$. The 3-D Potts model simulations of isotropic grain growth show grain-growth exponents in the range of $2 < n < 2.5$. The measured exponent depends on many variables of the system but importantly on the size of the system, kT_s, Q, and the initial distribution of the grain size. Issues about why the grain-growth exponent is so sensitive to these variables have yet to be definitely resolved.

Anisotropic Grain Growth. Figure 15 shows the evolution of such a system using realistic textures, misorientation distributions using Read-Shockley energies, and anisotropic mobilities in which the initial microstructure has a strong texture $<100>$ cube texture. The system undergoes normal grain growth, which causes a tightening of the texture. The boundaries are colored to show their misorientation,

(a) LIVE GRAPH
Click here to view

(b) LIVE GRAPH
Click here to view
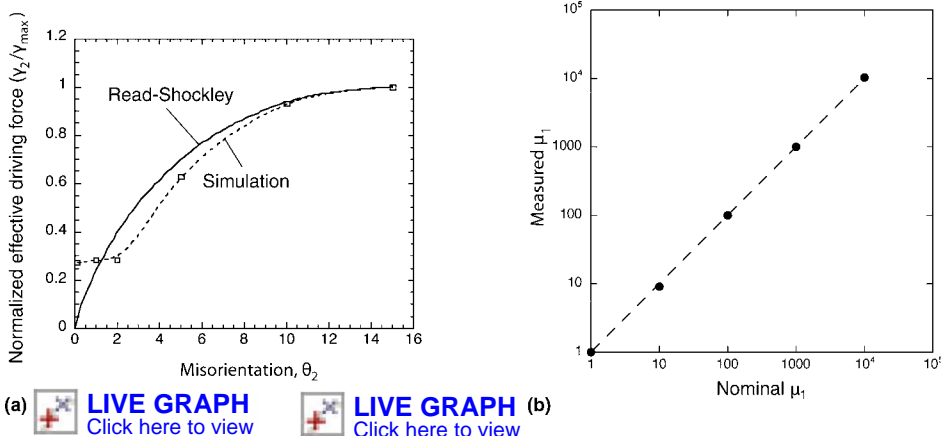
Fig. 14 (a) Measured γ_2 versus nominal γ_2 for Potts model simulations of boundary motion in the system illustrated in Fig. 13; $kT_s = 0.5$. (b) Measured μ_1 versus nominal μ_1 for Potts model simulations of boundary motion in the system illustrated in Fig. 13 with $\mu_2 = 1$
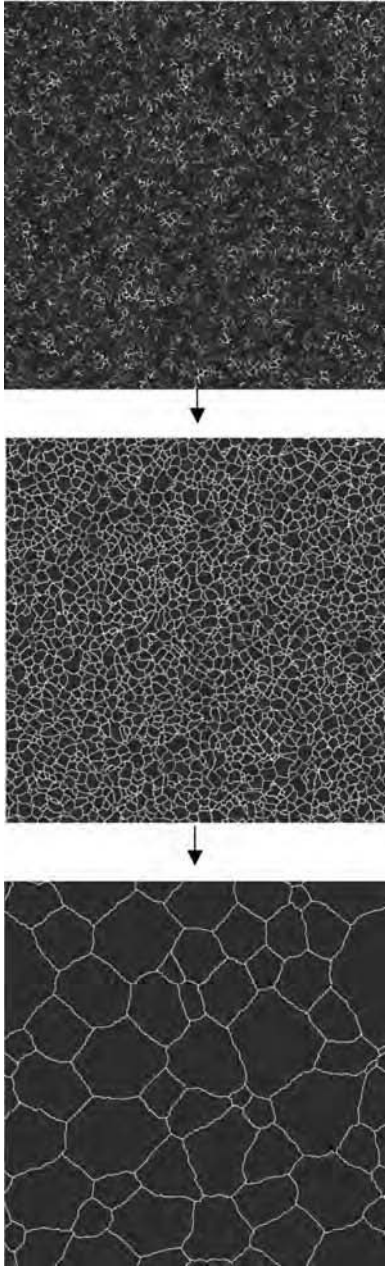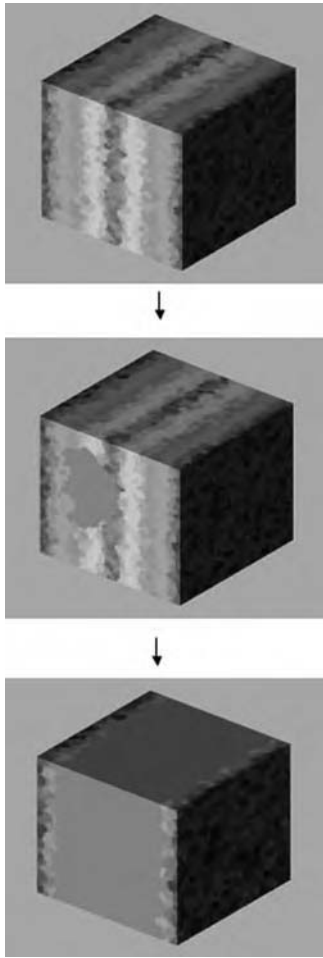
Fig. 15 Evolution of microstructure during a Potts model simulation of anisotropic grain growth of a single-texture component, using Read-Shockley energies and uniform mobilities. The simulation was performed using a square (1,2) lattice, Glauber dynamics, metropolis transition probability function, and $kT_s = 0.5$.

Fig. 16 Potts model simulation of anisotropic grain growth. (a) Relationship between misorientation distribution function (MDF) of the evolved system and the energy function. (b) Two-dimensional microstructure growth showing the multijunctions that form with highly anisotropic energy functions

LIVE GRAPH
Click here to view

with black being high misorientation and white being low misorientation. Note how all the high-misorientation boundaries (dark colored) are removed from the system during grain growth, with all the boundaries becoming white. This causes a reduction in the average misorientation and a narrowing misorientation distribution. This effect is observed experimentally and is due to the high-energy boundaries being replaced by low-misorientation boundaries.

Turning off anisotropic mobility and performing the same set of simulations with Read-Shockley energies and uniform mobilities reveals that mobility has little effect on the texture changes, and the energy function is the dominant factor (Ref 20).

The coincidence site lattice (CSL) description of grain boundaries is a geometric model based on the fact that, for certain misorientation rotations, a fraction $1/\Sigma$ of atomic lattice sites will be coincident. The resulting boundary is termed a CSL boundary and is characterized by Σ. Exact CSL boundaries are seldom observed in general materials, because the CSL relationship requires three independent boundary degrees of freedom to assume particular values. Brandon (Ref 21) introduced the concept of an acceptance criterion, which admits a wider range of CSL boundaries. Misorientations with angular deviations of less than $\Delta\theta\Sigma = 15°/\Sigma^{1/2}$ from the true CSL misorientation are assumed to be within the Σ CSL.

The Read-Shockley derivation requires that the array of boundary dislocations be spaced uniformly by some multiple of the Burgers vector, b. A CSL boundary can be viewed as a secondary array of dislocations with spacing b/Σ imposed on this primary array. As such, the contribution to the grain-boundary energy from the CSL can be modeled:

$$\delta J_\Sigma = \begin{cases} -\frac{J_0}{\Sigma}\left\{1 - \frac{\delta\theta}{\delta\theta_\Sigma}\left[1 - \ln\left(\frac{\delta\theta}{\delta\theta_\Sigma}\right)\right]\right\} & \text{for } \delta\theta < \delta\theta_\Sigma \\ 0 & \text{for } \delta\theta \geq \delta\theta_\Sigma \end{cases}$$
(Eq 22)

where $\delta\theta$ is the misorientation angle between the true CSL rotation and the actual grain-boundary rotation, $\delta\theta_\Sigma$ parameterizes the width of the energy well and is given by the Brandon criterion, and J_0 determines its depth, which maintains consistency with the non-CSL boundary model in Eq 16. The total energy of a boundary in this system is thus given by the sum of Eq 16 and 22. Note that for non-CSL boundaries, the contribution from Eq 22 is zero; Fig. 16(a) shows the form of such an energy function.

This type of simulation shows some interesting differences between modeling the full 3-D crystallographic orientations of a crystal and the 2-D crystallographic orientation (Ref 22). In the latter case, each grain requires only a scalar index to denote its orientation, and the misorientation θ is then easily calculated as a sum. In such a system, energy cusps of CSL, as shown in Fig. 16(a), have a profound effect on grain growth, with the produced MDF mirroring the energy function and a large fraction of the boundaries forming multijunctions, as shown in Fig. 16(b). However, if the same simulations are carried out in which each grain requires three Euler angles to denote its orientation, then the evolution is very different. The extra degrees of freedom that exist in Euler space mean that the chances of forming a boundary within the CSL limit become much smaller. The vast majority of boundaries that are classified as CSLs do not have significantly reduced energy; in fact, 95% of nominal $\Sigma 5$ boundaries have energy within 5% of the random, high-angle boundary energy. Even if the misorientation angle of the generic CSL is close to that of the exact CSL, the axis need not be close to the true axis. Therefore, most nominal CSL boundaries have energy near that of non-CSL boundaries and should not be morphologically enhanced during grain growth (Ref 22).

Abnormal Grain Growth. Under certain conditions, grain growth breaks down, and a small percentage of grains grow abnormally to consume the microstructure. Where these abnormal grains come from, or what causes them to form, is still an open research question, but what is clear is that, by its very nature, abnormal grain growth is a statistical process. One cannot point to a particular grain and determine whether it will grow abnormally; rather, one can predict a probability that it will do so. The reason for this unpredictability is that the growth of a grain depends also on the growth of its neighbors, which, in turn, depends on their neighbors and so on. This means that simulations of abnormal grain growth must be
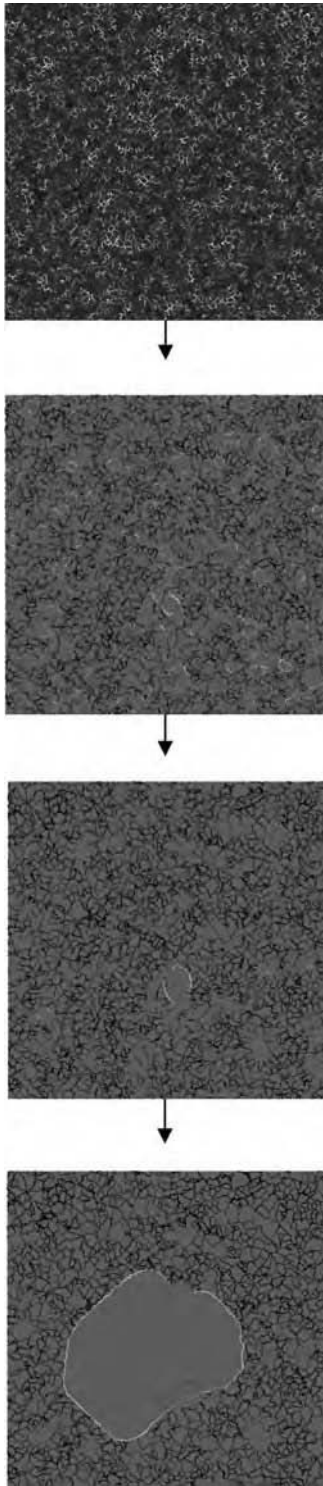
Fig. 17 Evolution of microstructure during a Potts model simulation of anisotropic grain growth of a single-texture component, using Read-Shockley energies and anisotropic mobilities to show the emergence of an abnormal grain

Fig. 18 Time rate of change in the size of an abnormally growing subgrain normalized by the mean size of the normal subgrains. The solid line is for the abnormal grain shown in Fig. 17, and the dotted line gives the results of the analytical theory of Rollett and Mullins (Ref 23) for the same subgrain. At intermediate times, the agreement between theory and simulation is excellent, indicating that the abnormal growth is driven by an energy and/or mobility advantage, as predicted by the theory. MCS, Monte Carlo time step

LIVE GRAPH
Click here to view

large enough to capture these rare events. The rarer the event, the larger the simulation must be. The alternative to performing very large simulations is to carry out many smaller simulations; either way, it is important to statistically sample the system.

Figure 17 shows the time evolution of an anisotropic system with Read-Shockley energies and anisotropic mobilities. The gray scale of the boundaries is proportional to their misorientation; thus, the darker boundaries have higher misorientation. Initially, the grains undergo normal growth, in which they remain relatively uniform in size, but the average boundary misorientation decreases. However, in this particular simulation, one subgrain begins to grow discontinuously. This subgrain has an orientation far from the mean, so that the mean misorientation of its boundaries is larger than average, as indicated by its light boundaries in Fig. 17. As it grows, its high misorientation boundaries come to dominate the system, and the average boundary misorientation increases. In texture space, an initial tightening of texture is observed during normal growth, followed by a shift to a new texture as the abnormal grain envelops the system.

The growth rate of the abnormal subgrain, as calculated from the Mullins and Rollett (Ref 23) theory, is plotted against its actual growth rate, as measured in the simulation in Fig. 18. At early times, stochastic effects cause the data to be quite noisy, and at late times, when the abnormal subgrain has consumed much of the system, finite-size effects become important. However, the agreement is good at intermediate times, indicating that the abnormal subgrain is growing as predicted by theory. This comparison with theory shows that when abnormal subgrains appear, they behave as they should, but theory does not explain how many such events should be expected. Is the observed occurrence of abnormal events reasonable?

One of the virtues of computer simulation is the ability to decouple physical parameters to help determine underlying mechanisms. In the simulations of abnormal subgrain growth discussed previously, both boundary energy and mobility varied with misorientation, and no stress or surface effects were included. Thus, boundary energy, mobility, or both are implicated in abnormal growth. To determine which, simulations were performed with uniform boundary mobility and Read-Shockley energies, in which case abnormal growth is not observed. If simulations are run with a gently varying mobility function of Read-Shockley energies, again, no abnormal growth occurs. However, when a highly anisotropic mobility function is used with boundary energy constant and uniform, abnormal growth occurs as before. Further simulations have shown that to observe abnormal growth in these highly textured systems, it is necessary and sufficient to have some very high mobility boundaries available to the system. With this knowledge, a model for the abnormal growth process that predicted nucleation rates was developed and compared with experiment (Ref 16). The effect of texture gradient on grain growth using the same MDF and energy and mobility functions was also investigated. Figure 19 demonstrates an example of such a simulation, which shows an abnormal grain growing rapidly through a texture gradient, until it meets grains of a similar orientation, at which point it slows down, and normal grain growth occurs.

Recrystallization. Figure 20 shows an example of a 3-D Potts model simulation of recrystallization in an extruded aluminum rod. There are many important issues that must be dealt with to model recrystallization correctly. Clearly, the microstructure, texture, and misorientation distribution are important, as is the overall macroscopic shape of the sample, as shown in Fig. 20. However, perhaps the most important variable is the stored energy and its correlation with the microstructure. This is not only because the stored energy is the main driving force of recrystallization but also because the nature and density of the stored energy determine the nucleation rate.

Stored energy is the main driving force in recrystallization, but in general, it is not easy to measure. Recently, the pattern quality of EBSD Kikuchi images has been used as an index of the work-hardening state and hence the dislocation density on a local scale. At best, this is an accurate qualitative indication of the distribution of stored energy in the system. It is also, of necessity, a 2-D measure. When importing these structures, the pattern quality is turned into the scalar variable, h_i, describing the local stored energy. An alternative way of modeling the distribution of stored energy is to perform simulations of plastic deformation and take the output of these simulations as the input for the Potts model simulations. These polycrystal-plasticity simulations are notoriously difficult to perform accurately, but they

Fig. 19 Evolution of microstructure during a Potts model simulation of anisotropic grain growth in a texture gradient, using Read-Shockley energies and anisotropic mobilities. The simulation was performed using a square (1,2) lattice, Glauber dynamics, metropolis transition probability function, and $kT_s = 0.5$.

Fig. 20 Series of snapshots from a Potts model simulation of an extruded aluminum rod recrystallizing with site-saturated surface nucleation. The light gray indicates a recrystallized grain; the dark-gray grains are unrecrystallized. The system is a $50 \times 50 \times 200$ cylinder, with periodic boundary conditions in the axial (z) direction. Source: Ref 6

Fig. 21 Boundary/particle interaction. (a) TiN particle interacting with a carbon steel grain boundary. (b) The simulation was performed using a square (1,2) lattice, Glauber dynamics, metropolis transition probability function, and $kT_s = 0.5$.

do have the advantage that the variables which describe the deformed state are known. Once again, these can be converted into the scalar quantities, h_i, needed by the Potts model.

Once the structure has been imported into the model, the next problem concerns the nucleation criteria to produce strain-free nuclei in the structure. The simulations shown in Fig. 20 were performed with site-saturated surface nucleation, which means that all the nuclei were allocated to sites on the surface at the beginning of the simulation, and recrystallization occurred as a result of growth and competition between these nuclei. The precise nucleation model is extremely important in these simulations, because it dictates the final grain size and the texture of the simulations. Thus, great care must be taken when implementing and interpreting the nucleation model.

The growth part of the model, which really is the only part of such simulations that concerns the Potts model, affects only the kinetics of grain growth. It has been mentioned before that there are no absolute length and time scales in Potts model simulations. Determining critical length and time scales is not usually a problem, although this cannot be done a priori but must be performed by some calibration procedure, which involves measuring the starting grain size in sites and equating this to the imported grain size measured in micrometers. The annealing time can be calibrated in a similar manner. However, users of the Potts model should be aware that there is a difficulty with the kinetics of the simulations, as mentioned in the section "Stored Energy" in this article. The driving force for growth is not proportional to the stored energy, and the boundaries are all driven forward with equal force. Thus, the kinetics are unlikely to be accurate even when the calibrations have been carried out, unless boundary velocity is independent of driving force. Rollett and Raabe have proposed a hybrid model of recrystallization that involves
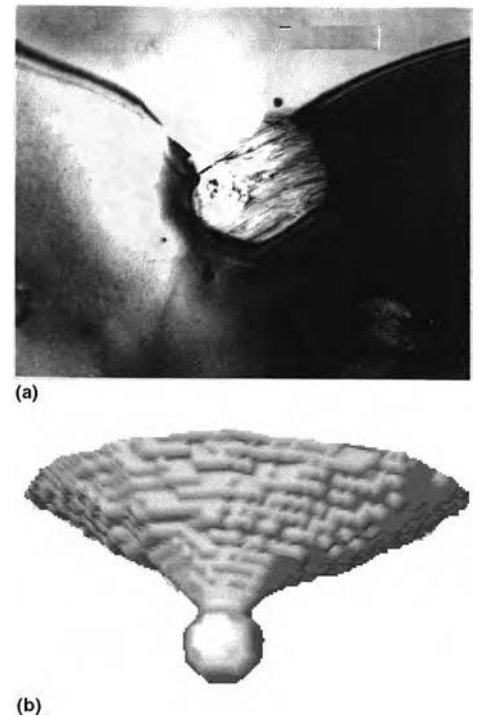
the coupling of the Potts model with a cellular automaton (CA) to get around this problem (Ref 24). This model is ideal if the aim is to model recrystallization and the subsequent grain growth in a single simulation, but in general, CAs are more appropriate for modeling recrystallization than Monte Carlo models (Ref 24).

The Zener-Smith pinning theory describes how second-phase particles hinder the motion of grain boundaries during microstructural evolution. The theory has been used to explain a wide range of microstructural phenomena, such as grain-growth stagnation and abnormal grain growth. Incorporating particles into metal alloys and ceramics to control grain size during thermomechanical processing is row standard industrial practice. Figure 21 shows the shape of the boundary being pulled away from a particle in two different systems: a TiN particle in steel and the Potts model. The origin of the pinning force is the extra energy required to create the dimpled boundary shape, and, as can be seen from Fig. 21, the Potts model appears to work well qualitatively. However, it is not enough for the model to appear to give the right shape; the correspondence must be quantified.

If the particle is incoherent with the matrix, and the interface tension, γ, at the boundary-particle interface is isotropic, this results in an angle of $\pi/2$ at the boundary-particle interface. Under these circumstances, the boundary

exhibits a characteristic dimple shape, which is a minimal surface described as a catenoid of revolution. Hellman and Hillert (Ref 25) derived the catenoid boundary shape as a function of z_o, the distance between the particle center and the spherical boundary cap (Fig. 22a):

$$z_o = a \left[b + \cosh^{-1} \left(\frac{\rho}{a} \right)^{1/2} - \frac{\rho}{a} \left(1 - \left(1 - \frac{a}{\rho} \right)^{1/2} \right) \right]$$

(Eq 23)

where θ is the bypass angle, r is the particle radius, ρ is the radius of the boundary curvature, $a = r \sin \theta \cos \theta$, and $b = \sec \theta \cosh^{-1} (\csc \theta)$.

Figure 22(b) shows the comparison between the Potts model and the theoretical predictions. The pinning force can be calculated by measuring the derivative of the energy with distance traveled by the boundary. The only complication with this is knowing the distance of the boundary, because locally it changes shape dramatically. Thus, these calculations require large simulations to be carried out in which the local dimple is a small perturbation on the overall boundary shape. Figure 22(c) shows the normalized force (F/F_p) versus normalized displacement (z_o/r) for the model and theory. Note that boundary detachment occurs at a bypass angle of 70°; this means that although the maximum pinning force is independent of ρ/r, the point of boundary detachment, the "stickiness" of a particle/boundary interaction, is not.

What these simulations show is that the Potts model gives good agreement with the theoretical curve, although there is disagreement with the details of boundary detachment. This provides confidence to scale up the model and look at the effect of arrays of particles on grain growth. Figure 23(a) shows a snapshot of a simulation carried out to investigate the effect of volume fraction of particles on grain growth. By carrying out a systematic study of the effect of volume fraction on pinned grain size, it is possible to plot the relationship and compare with the experiment. Figure 23(b) and (c) show that the Potts model does a good job of describing the phenomenon (Ref 11). As a result of the effectiveness of the method, there is now quite a bit of interest in using the method to study abnormal grain growth in the presence of particles.

Algorithms

The simplest algorithm for implementing the Potts model is called the vanilla algorithm. This involves setting up the initial conditions of the lattice, then choosing a random change to an orientation (to one of the Q other orientations) at a random lattice site and accepting or rejecting the change based on the change to the total energy of the system, ΔE, computed via the

relevant Hamiltonian. Glauber or Kawasaki dynamics can be employed using the metropolis or symmetric probability functions. The time required to attempt a single orientation swap, whether successful or unsuccessful, is defined arbitrarily as τ, and so, 1 MCS is defined as N-attempted flips. This vanilla algorithm requires several decisions to choose the site and the new orientation, which are fundamentally random. It should be emphasized that for large lattices and large run times, this requires the generation of a large number of random numbers. If the source of these random numbers has a low repeat signature or does not distribute the random number uniform on {0, 1}, then artificial patterns of behavior that are not due to curvature driving forces will be observed.

This vanilla algorithm is sufficiently fast for small 2-D systems. However, to use the model to simulate industrially relevant applications, as outlined in the previous section, large 3-D lattices and long simulation times are required. Such work usually requires the use of speed-up algorithms. This section presents three such algorithms, each of which uses a different speed-up strategy. All of them have their advantages and disadvantages, and so, it is ultimately for the user to decide which is best to employ.

The boundary-site algorithm is the simplest of the speed-up algorithms. It relies on the fact that when simulating interface migration, it is only the boundary sites that are important. The swaps that occur within the domains are, in fact, problematic, because they represent a nucleation event that, for most applications, is unphysical. Excluding these events by using $kT_s = 0$ has the problem of causing lattice effects. The boundary-site algorithm is similar to the vanilla algorithm, but it excludes these events a priori, even at finite temperatures. The algorithm is much more efficient for simulating boundary migration than vanilla Potts algorithms, because, in most systems of interest, the boundary sites are not only the most important sites but also in the minority. Thus, by excluding nonboundary sites from potential orientation swaps, more central processing unit time is spent on simulating boundary migration.

There are many ways of ensuring that only boundary sites are sampled for potential orientation swaps, each of which has an associated computational overhead and memory allocation implications. The easiest way is to modify the vanilla Potts model so that after a site is selected at random, a neighbor check is performed, and if the site is found not to be a boundary site, time is incremented and a new site selected. Further efficiency gains can be obtained by maintaining a list of the boundary sites and only picking sites from that list. The maintenance of this list has a computational overhead associated with it, and so, the overall efficiency of the algorithm depends on the boundary-site/interior-site ratio of the system. In this case, incrementing the time is also more complicated, because the algorithm must accurately calculate the effective time, τ_b, associated which each attempted boundary-site swap. In

(a)

(b)

LIVE GRAPH
Click here to view

(c)

LIVE GRAPH
Click here to view
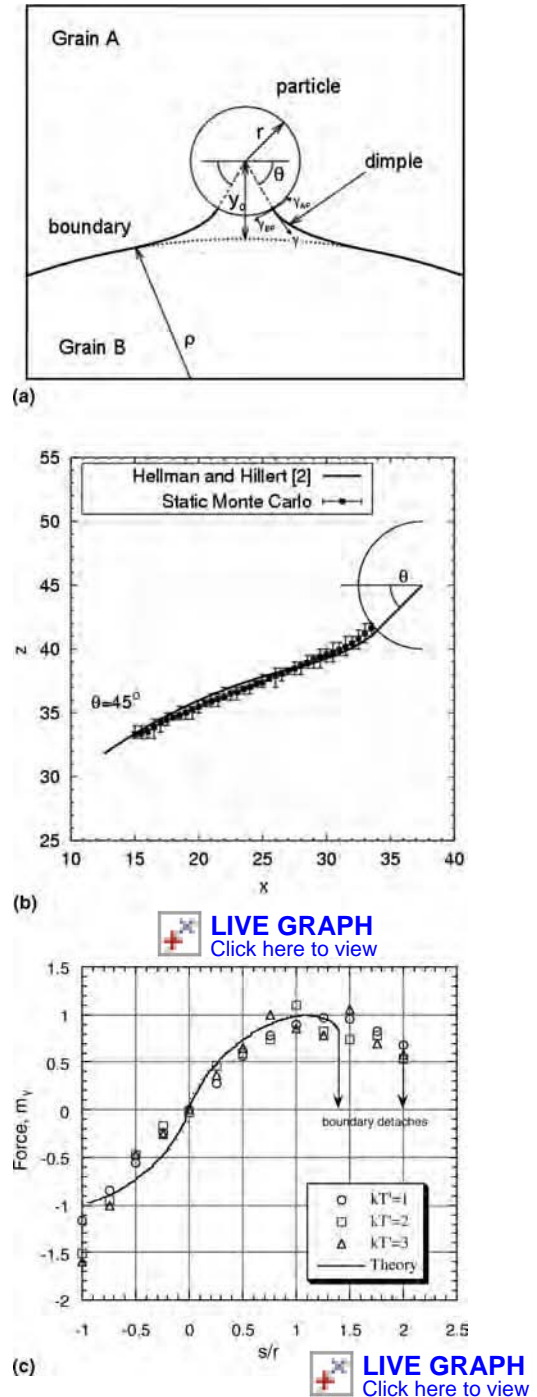
Fig. 22 (a) Schematic of the formation of a dimple during grain-boundary bypass of a particle. r, radius of the particle; ρ, radius of the boundary curvature; θ, boundary bypass angle; γ boundary surface tension; γ_{AP} and γ_{BP}, the two particle/boundary surface tensions; and y_o, distance of the boundary from the particle center. (b) Comparison of the dimple shape produced by a Potts model and theory. (c) Comparison of the pinning force produced by a Potts model and theory

such a simple algorithm, τ_b is equal to the number of boundary sites in the list divided by N.

Further speedup can be obtained by only allowing boundary sites to swap to an
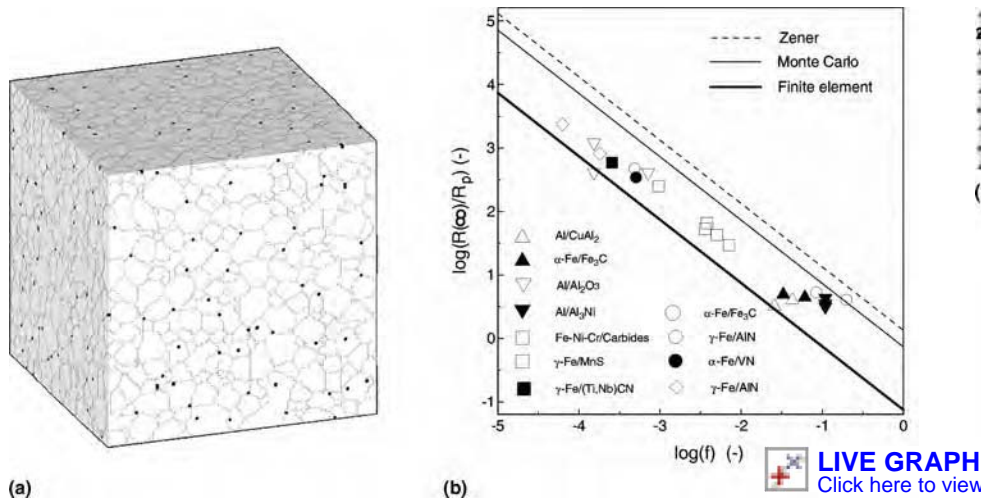
Fig. 23 (a) Snapshot of a pinned microstructure in a Potts model simulation of Zener pinning on a $400 \times 400 \times 400$ lattice, using particles with sizes $3 \times 3 \times 3$. (b) Comparison of pinned grain size with experimental data. Source: Ref 26

Fig. 24 Checkerboarding of lattices. (a) Two-dimensional (2-D) square lattice using only first-nearest neighbors. (b) 2-D square lattice using first- and second-nearest neighbors. (c) Sharing of an interface strip allows two processors to simultaneously update their sites using checkerboarding.

orientation of one of their neighbors. Once again, this is more physically realistic in that it rules out the boundary nucleation events that can occur at finite temperatures. Choosing one orientation randomly from the set of available neighbor orientations again has implications for τ_b. The effective speedup is different for each site and depends on its neighborhood. Because this algorithm is actually doing something slightly different from the vanilla Potts, an exact quantitative match is not relevant. However, a check should always be performed with every new code to make sure that grain growth obeys von Neumann laws, with an invariant normalized grain-size distribution and an average grain size that obeys a parabolic growth law.

A speed-up algorithm that exactly corresponds to the vanilla Potts model is known as the *n*-fold way algorithm and was first developed by Bortz et al. (Ref 27). The algorithm eliminates all unsuccessful orientation swap attempts a priori, so that all orientation swaps result in domain evolution. It requires a list of all possible successful orientation swaps to be compiled, along with their transition energies, ΔE. An activity $\pi_i = P\,(\Delta E_i)$ of the N possible orientation swaps is defined (Ref 27) with the total system activity, A, as given by:

$$A_n = \sum_{i=1}^{N} \pi_i \tag{Eq 24}$$

Each site is visited with a probability weighted by its activity, an effective orientation swap is performed, the site and its neighbors are re-evaluated, and π_i is updated.

In the conventional Potts model, the simulation is incremented after each attempted orientation swap. In the *n*-fold way, each orientation swap is successful, so the time increment must be scaled by the average time

between successful flips in the vanilla algorithm. This time increment is:

$$\Delta t = -(\tau/A_n)\ln R \tag{Eq 25}$$

where R is a random number uniformly distributed on the interval (0,1). This time increment decreases as the total system activity decreases, reflecting an increased efficiency of the algorithm. Hassold and Holm (Ref 4) have shown that using this type of *n*-fold way algorithm results in identical simulation dynamics in substantially less computation time. The algorithm is efficient in the case where either $kT_s = 0$ or when $kT_s \neq 0$, when all orientation swaps are confined to the boundary sites. For other cases, the increase in computational effort involved in compiling and updating lists of the active sites usually outweighs the advantages of the algorithm.

The Potts algorithm does not readily parallelize in the obvious way of assigning each of P processors a subset of the lattice sites. This is because two or more processors may pick adjacent sites. If this occurs, then, when the two processors attempt to calculate ΔE for an orientation swap, they will each do so using incorrect information about neighboring orientations (each other). This violates the detailed balance that demands that two or more sites may not be flipped simultaneously if their interaction affects the Hamiltonian. A valid parallelization method is to use a checkerboarding approach in which each processor simultaneously attempts to swap orientations in a subdomain in which all sites share no neighbors with other subdomains.

Checkerboarding involves assigning each lattice site one of a finite set of flavors. For example, Fig. 24(a) shows a 2-D square lattice checkerboarded with two flavors: those denoted with circular symbols and those with triangular symbols. This can be used to perform Potts
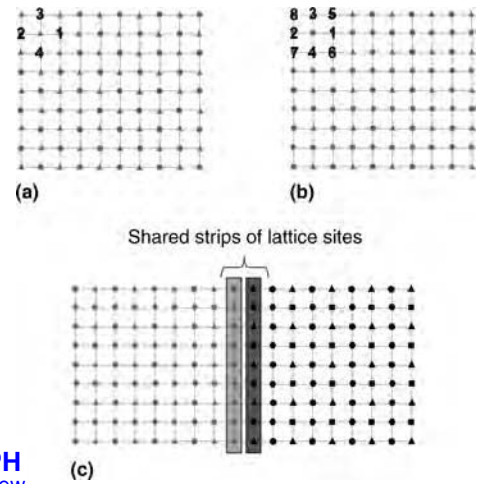
model simulations using first-nearest neighbors, because each of the circle sites has no nearest neighbors that are circles. Similarly, each of the triangular sites has no nearest neighbors that are triangles. Thus, instead of picking sites randomly, all the circle sites can be updated simultaneously without the result of any flip attempt affecting the other circle sites. This is also true for the triangular sites. Algorithmically, a loop over the flavors replaces the need to randomly pick sites and thus removes the possibility that two processors may try to flip to adjacent sites.

The checkerboarding method can be applied to 2-D square lattices using first- and second-nearest neighbors, but in this case, four flavors of lattice sites are needed to ensure that simultaneous updating does not invalidate detailed balance. In 2-D triangular lattices using first-nearest neighbors, three unique flavors are needed. In 3-D using the 26 first-, second-, and third-nearest neighbors, eight flavors are needed. By using checkerboarding, P processors can be most efficiently employed if the lattice is subdivided into P contiguous domains. Each processor also must store a copy of the narrow strips (2-D) or planes (3-D) of the lattice sites that immediately adjoin its subdomain and are actually owned by neighboring processors. The important point is that the processors are all synchronized to update the same flavor of lattice site. This algorithm is highly parallel, with the only communication cost being the local exchanges of boundary orientations between neighboring processors. These exchanges also serve as synchronization points in the loop over flavors to ensure that all processors work on the same flavor at the same time. In practice, as long as the processor subdomains are of reasonable size, the communication costs are only a few percent of the total run time, and thus, the algorithm can simulate large lattices with parallel efficiencies of over 90% (Ref 28).

Final Remarks

The aim of this article has been to give an introduction to the Potts model and a practical guide to programming and performing simulations. The beauty of the Potts model is that is a simple way to model complex systems by modeling local physics, in particular, the effect of surface tension phenomena on the development and evolution of microstructure. At all times, it has been important to compare the model with theory and experimental data; this instinct is essential to any modeler. It is easy to make a model yield pretty pictures that appear to have a correspondence with a "real" phenomenon, but quantification of the simulation is the only way to use the model as a method to gain physical insights and understanding. Finally, it is important to note that a model is just that, a model, and the benefit of it is as much to guide the experiment and to hone the intuition about physical phenomena as it is to make predictions. The major role of computer models is to reduce the number of experiments that must be carried out and to highlight what variables are key to understanding the results. The Potts model should be seen in this light; it is a guide to the intuition. Above all, it is a medium of communication between experimentalists and theoreticians.

REFERENCES

1. G. Gottstein, D.A. Molodov, and L.S. Shvindlerman, Grain Boundary Migration in Metals: Recent Developments, *Interface Sci.*, Vol 6, 1998, p 7–22
2. H. Yan, M.A. Miodownik, and V. Randle, Experimental and Computer Model Investigations of the Texture Evolution of a Non-Oriented Silicon Steel, *J. Mater. Sci. Technol.*, Vol 24, 2008, p 705–710
3. J. von Neumann (in discussion with Smith), *Metal Interfaces*, American Society for Metals, Cleveland, OH, 1952, p 108
4. G.N. Hassold and E.A. Holm, A Fast Serial Algorithm for the Finite Temperature Quenched Potts Model, *Comput. Phys.*, Vol 7, 1993, p 97–107
5. M.P. Anderson, G.S. Grest, and D.J. Srolovitz, *Scr. Metall.*, Vol 19, 1985, p 225–230
6. E.A. Holm and C.C. Battaile, *JOM*, Vol 53, 2001, p 203
7. A.E. Lobkovsky, A. Karma, M.I. Mendelev, M. Haataja, and D.J. Srolovitz, Grain Shape, Grain Boundary Mobility and the Herring Relation, *Acta Mater.*, Vol 52, 2004, p 285–292
8. K.G.F. Janssens, Random Grid, Three Dimensional, Space-Time Coupled Cellular Automata for the Simulation of Recrystallization and Grain Growth, *Model. Simul. Mater. Sci. Eng.*, Vol 11 (No. 2), 2003, p 157–171
9. E.A. Holm, D.J. Srolovitz, and J.W. Cahn, Microstructural Evolution in Two-Dimensional Two-Phase Polycrystals, *Acta Metall. Mater.*, Vol 41, 1993, p 1119–1136
10. M.A. Miodownik, J.W. Martin, and A. Cerezo, *Philos. Mag. A*, Vol 79 (No.1), 1999, p 20322
11. A. Harun, E.A. Holm, M.P. Clode, and M.A. Miodownik, On Computer Simulation Methods to Model Zener Pinning, *Acta Mater.*, Vol 54, 2006, p 3261–3273
12. H.F. Poulsen, 3DXRD: Grain Maps, Grain Dynamics and Grain Refinement, *Crystallogr. Rev.*, Vol 10 (No. 1), 2004, p 29–43
13. M.A. Miodownik, A. Godfrey, E.A. Holm, and D.A. Hughes, On Boundary Misorientation Distribution Functions and How to Incorporate Them into Three-Dimensional Models of Microstructural Evolution, *Acta Mater.*, Vol 47, 1999, p 2661–2668
14. W.T. Read and W. Shockley, *Phys. Rev. B*, Vol 78, 1950, p 275
15. A.P. Sutton and R.W. Balluffi, *Interfaces in Crystalline Materials*, Oxford Science Publications, Oxford, U.K., 1995
16. E.A. Holm, M.A. Miodownik, and A. Rollett, On Abnormal Subgrain Growth and the Origin of Recrystallization Nuclei, *Acta Mater.*, Vol 51, 2003, p 2701–2716
17. W.W. Mullins, Two-Dimensional Motion of Idealized Grain Boundaries, *J. Appl. Phys.*, Vol 27, p 900
18. R.D. MacPherson and D.J. Srolovitz, The von Neumann Relation Generalized to Coarsening of Three-Dimensional Microstructures, *Nature*, Vol 446, 2007, p 1053–1055
19. M. Hillert, On the Theory of Normal and Abnormal Grain Growth, *Acta Metall.*, Vol 13, 1965, p 227
20. G.N. Hassold, E.A. Holm, and M.A. Miodownik, The Influence of Anisotropic Boundary Properties on the Evolution of Misorientation Distribution during Grain Growth, *Acta Mater.*, Vol 49, 2001, p 2981–2991
21. C. Maurice, 2- and 3-D Curvature Driven Vertex Simulations of Grain Growth, *Proceedings of the First Joint International Conference on Recrystallization and Grain Growth*, Vol 1, Springer-Verlag, 2001, p 123–134
22. G.N. Hassold, E.A. Holm, and M.A. Miodownik, On the Accumulation of CSL Boundaries during Grain Growth, *Mater. Sci. Technol.*, Vol 19, 2003, p 683–687
23. A.D. Rollett and W.W. Mullins, *Scr. Mater.*, Vol 36, 1997, p 975
24. A.D. Rollett and D. Raabe, A Hybrid Model for Mesoscopic Simulation of Recrystallization, *Comput. Mater. Sci.*, Vol 21, 2001, p 69–78
25. P. Hellman and M. Hillert, On the Effect of Second-Phase Particles on Grain Growth, *Scand. J. Metall.*, Vol 4, 1975, p 211–219
26. G. Couturier, R. Doherty, C. Maurice, and R. Fortunier, *Acta Mater.*, Vol 53 (No. 4), 2005, p 977–989
27. A.B. Bortz, M.H. Kalos, and J. Liebowitz, *J. Comp. Phys.*, Vol 17, 1975, p 10–18
28. S.A. Wright, S.J. Plimpton, T.P. Swiler, R.M. Fye, M.F. Young, and E.A. Holm, "Potts-Model Grain Growth Simulations: Parallel Algorithms and Applications," Sandia Report, SAND-97, Sandia National Laboratories, 1997

SELECTED REFERENCES

- J.E. Burke and D. Turnbull, Recrystallisation and Grain Growth, *Prog. Met. Phys.*, Vol 3, 1952, p 220–292
- C. Herring, Surface Tension as a Motivation for Sintering, *The Physics of Powder Metallurgy*, R. Gomer et al., Ed., McGraw-Hill, New York, NY, 1949
- F.K. Hwang, D.S. Richards, and P. Winter, The Steiner Tree Problem, *Annals of Discrete Mathematics*, Elsevier, Amsterdam, Netherlands, 1992
- S. Ling and M.P. Anderson, Monte Carlo Simulation of Grain Growth and Recrystallization in Polycrystalline Materials, *JOM*, Vol 44, Sept 1992, p 30–36
- M.I. Mendelev, D.J. Srolovitz, G. Gottstien, and L.S. Shvindlerman, *J. Mater. Res.*, Vol 17, 2002, p 234
- A.D. Rollett, D.J. Srolovitz, and M.P. Anderson, Simulation and Theory of Abnormal Grain Growth—Anisotropic Grain Boundary Energies and Mobilities, *Acta Metall. Mater.*, Vol 37, 1988, p 1227–1240
- A.D. Rollett, D.J. Srolovitz, M.P. Anderson, et al., Computer-Simulation of Recrystallization. 3. Influence of a Dispersion of Fine Particles, *Acta Metall. Mater.*, Vol 40, 1992, p 3475–3495
- A.D. Rollett, D.J. Srolovitz, and R.D. Doherty, Computer-Simulation of Recrystallization in Non-Uniformly Deformed Metals, *Acta Metall. Mater.*, Vol 37, 1989, p 627–639
- C.S. Smith, Grain Shapes and Other Metallurgical Applications of Topology, *Metal Interfaces*, American Society for Metals, Cleveland, OH, p 65–108
- D. Solas, P. Geber, T. Baudin, and R. Penelle, Monte Carlo Method for Simulating Grain Growth in 3D: Influence of Lattice Site Arrangements, *Mater. Sci. Forum*, Vol 467–470, 2004, p 1117–1122
- M. Winning, G. Gottstien, and L.S. Shvindlerman, *Acta Mater.*, Vol 50, 2002, p 353

Network and Vertex Models for Grain Growth

L.A. Barrales Mora, V. Mohles, G. Gottstein, and L.S. Shvindlerman,* Institut für Metallkunde und Metallphysik, Aachen, Germany

NETWORK MODELS and their applications for the simulation of various physical phenomena related to grain-boundary migration are reviewed. The implementation of typical two- and three-dimensional network models is introduced and comprehensively detailed. The implementation of such models is subdivided into three steps: acquisition and discretization of the microstructure, formulation of the equation of motion, and implementation of the topological transformations. To demonstrate the predictive power of network models and their flexibility, prominent examples of the simulation of physical phenomena are presented, that is, the simulation of two- and three-dimensional normal grain growth, grain growth in the presence of an extrinsic driving force, and the effect of a finite mobility of boundary junctions on grain growth.

Various models have been used for the simulation of grain-boundary migration and related phenomena, in particular, grain growth and recrystallization, notably Monte Carlo (Ref 1–4), phase field (Ref 5, 6), and network models (Ref 7–24). However, models based on cellular automata (Ref 25, 26), finite elements (Ref 27), and molecular dynamics (Ref 28) can also be found in the literature. Among them, the network models stand out for the clarity of the physics used for the description of grain-boundary migration. The network models received their name from the procedure in which the microstructure of a polycrystal is abstracted, because the only required information in these models is the interfacial topology of the microstructure. These models have been referred to in different terms, for example, vertex model (Ref 11, 12, 20), partial differential equation model (Ref 21), surface evolver model (Ref 29), and so on, but all of them belong to the category of network models.

History of Development

The first reported attempt of a network model is due to Fullman in 1952 (Ref 7), who simulated grain growth in a two-dimensional (2-D) polycrystal. He represented a 2-D microstructure (cellular pattern) only by its triple junctions. Then, the evolution of the microstructure results only from the movement of the triple junctions whose velocities were calculated manually from the line tension and length of the three adjacent grain boundaries. The model was picked up in 1983 by Weaire and Kermode (Ref 8, 9), who developed a rather complex algorithm based on the continuous equilibration of the topological network through the displacement of the triple junctions. Interestingly, in this model the displacements of the triple junctions were not directly calculated but only adjusted to obtain the values of the dihedral angles and grain areas at each simulation step. More akin to Fullman's approach, Soares et al. (Ref 10) developed a model for the simulation of grain growth, with the difference that they neglected the curvature of the grain boundaries and calculated the displacement of the triple junctions only from the line tensions of the adjacent grain boundaries. Frost et al. (Ref 30) also considered the curvature of the grain boundaries by defining them as an array of nodes. In their model, the velocities of these nodes and thus their displacements were calculated from the local curvature. However, the displacements of the triple junctions were not calculated but only adjusted to attain the equilibrium dihedral angle. This simple modification generated the possibility of directly simulating grain-boundary migration.

A different approach was considered by Kawasaki et al. (Ref 11), who represented the microstructure by only their triple junctions, called vertices, in a manner similar to the approaches of Fullman (Ref 7) and Soares et al. (Ref 10). However, in the approach of Kawasaki et al., the velocities of the triple junctions resulted from the solution of a system of Lagrangian equations for the dissipation of the potential energy (stemming from the interfacial energy) of a microstructure through the movement of the triple junctions. As in the previous models, the equilibrium at triple junctions had to be enforced because of the constraints imposed by straight grain boundaries. To improve the accuracy of the approach and to address curvature-driven boundary motion, Weygand et al. (Ref 18) introduced nodes (virtual vertices) along the grain boundary that permit the representation of curved elements and thus do not have to enforce equilibrium at triple junctions. Due to its simplicity, the vertex approach has remained popular, and some other implementations have been proposed to overcome the limitations of straight grain boundaries. In particular, Maurice and Humphreys (Ref 19), based on an idea by Svoboda (Ref 31), formulated what they called an explicit curvature-driven vertex model. In their approach, the curvature of the grain boundaries was approximated by circular arcs and subsequently used to calculate the velocity and displacement of the triple junctions. An advantage of this formulation is that it can be easily extended to three dimension (3-D), because it only relies on the positions of the triple junctions in 2-D or quadruple junctions in 3-D and the grain-boundary energy.

Owing to the difficult implementation of the topological transformations that take place in the course of grain growth, only a few 3-D network models have been developed so far. Prominent examples are Brakke's surface evolver (Ref 29) and the 3-D vertex model by Fuchizaki et al. (Ref 16).

*Russian Academy of Sciences

All of these previously mentioned models or modifications thereof have been used in a variety of applications pertinent to grain-boundary migration. The most prominent applications are detailed in a subsequent section, but to begin with, the implementation and the problems of typical 2-D and 3-D network models are comprehensively addressed in the following section. This implementation can be subdivided into three parts: initialization and discretization of the microstructure, equation of motion, and topological transformations.

Initialization and Discretization of the Microstructure Model

Initialization of the Microstructure. The generation of a microstructure in a numerical simulation code is crucial for every simulation model. In 2-D, the microstructure can be obtained even from experimental sampling by light optical or electron microscopy, notably by orientation microscopy via electron bockscatter diffraction in a scanning electron microscope and digitization of the information (Fig. 1). In 3-D, the problem is more complicated, because there are very few real 3-D observation techniques available so far. Generally, the experimental methods only allow the analysis of small sample volumes and thus are often not representative for the entire microstructure of a specimen (Ref 32–35).

An alternative to experimental 3-D sampling is the reconstruction of a 3-D microstructure from numerous successive 2-D slices of the sample (serial sectioning) (Ref 36–38). The necessary effort for this task is, nevertheless, enormous if the acquisition is conducted manually. Fortunately, automated 3-D orientation microscopy has recently become available, facilitating this task (Ref 39–42). However, the information provided by these methods normally does not include the grain boundaries (i.e., does not provide the microstructure as a topological network) but discretizes only the grains as small pieces of volume with a uniform orientation. For several simulation models, such as cellular automata, phase-field models, Monte Carlo models, and so on, this information is sufficient because they require the discretization of the volume, precisely through small volume elements. Conversely, for the network models, this information is neither adequate nor sufficient, because they demand a determination and discretization of the grain boundaries, that is, curves in 2-D and surfaces in 3-D. Additionally, for the simulations of the topological network information about the grains and their orientation, they require number and position of the triple and quadruple junctions and the misorientation relationships across the grain boundaries. Special algorithms that allow the extraction of these data and the identification of the topological elements can be used. However, they are rather complicated and require substantial computational efforts (Ref 43).

The numerical alternative is the use of tessellations. A tessellation (also called mosaic) is a regular arrangement of polygons (2-D) (regular in the sense that there are no gaps or overlaps between the elements of the tessellation), polyhedra (3-D), or polytopes (*n*-dimensional) that comprises a determined space. A tessellation can be bounded or unbounded, depending on the spatial-boundary conditions. There are several tessellations in nature, for example, a polycrystal or the honeycomb of a beehive (Ref 44).

For a simulation of grain growth, so-called Voronoi tessellations can be used and are extensively described in the literature (Ref 4, 11, 12, 18). This kind of tessellation is simple and provides the necessary information for the simulation of grain growth with network models, that is, the topological network. However, existing algorithms for the construction of the tessellation are either complicated or slow; a naïve implementation can result in both. An efficient implementation is not a trivial task and demands some programming skills. For a detailed description of the different methods for the construction of Voronoi tessellations, refer to Ref 45 and 46.

The concept of Voronoi tessellations was introduced more than a century ago (Ref 47). The formal definition is rather simple: Given a defined (not necessarily finite) space, S, let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of points within S, which are called partition sites (Ref 45). A partition of the space is obtained by assigning every point to its nearest site. All points assigned to p_i form a Voronoi region $V(p_i)$, which is also referred to as a Voronoi polygon (in 2-D and in a bounded space) (Ref 45). The $V(p_i)$ consists of all points that are at least as close to p_i as to any other site, explicitly

$$V(p_i) = \left\{ x : |p_i - x| \leq |p_j - x| \forall j \neq i \right\} \qquad \text{(Eq 1)}$$

Evidently, the sites compete for the largest region possible. Those points that belong to more than one site form the borders for each Voronoi region; in 2-D this border is a line, whereas in 3-D it is a surface. The set of all these borders forms the Voronoi diagram. In Fig. 2, a 2-D Voronoi tessellation is shown. This tessellation was generated by ten sites in a bounded space S (Fig. 2a). Each site has influence over a limited region of the space S, and each region is bounded by straight lines that form a polygon (Fig. 2b). An example of a bounded 3-D Voronoi tessellation can be seen in Fig. 3.

In polycrystalline materials, a Voronoi region can be associated with a grain or crystal, and the Voronoi diagram represents the grain boundaries. This method renders the topology of the tessellation and therefore lends itself as an input for network models. The main problem
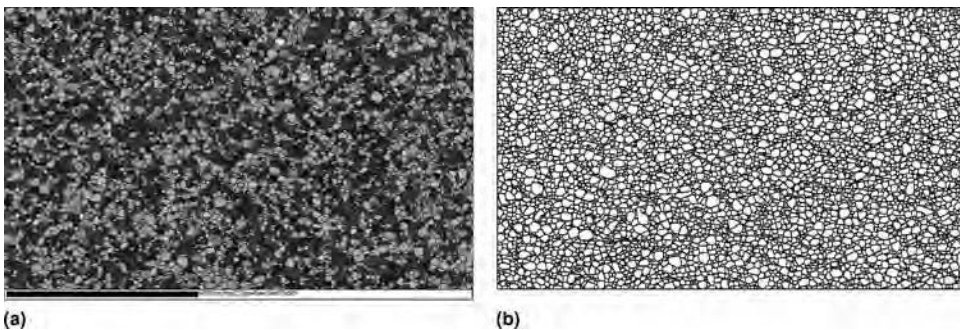
Fig. 1 Titanium microstructure electron backscatter diffraction (EBSD). (a) EBSD mapping of pure titanium. (b) Discretization in a topological network for the simulations
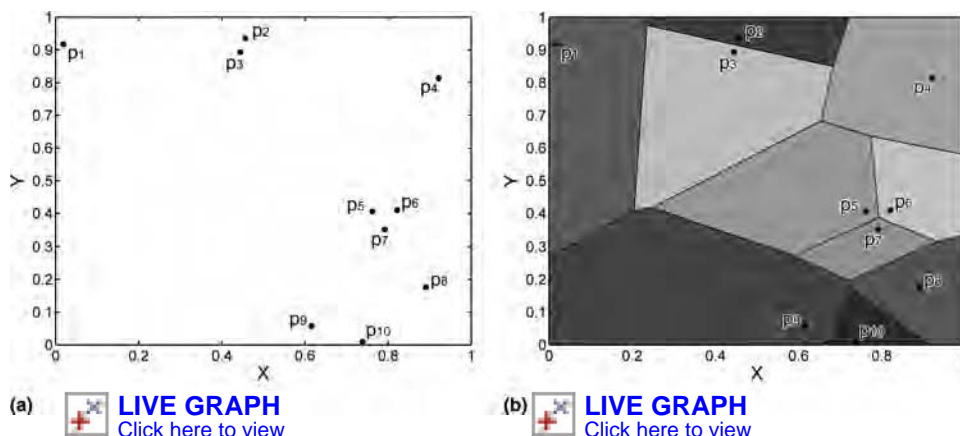
Fig. 2 Concept of tessellation in a space, S. (a) Randomly generated partition sites ($p_1 \ldots p_{10}$) for the Voronoi tessellation. (b) Corresponding Voronoi diagram. Model in color

with Voronoi tessellations is that they do not yield realistic grain-size distributions. To overcome this difficulty, some modifications to the original Voronoi tessellation have been proposed (Ref 48). For instance, the radical plane method for the packing of spheres allows the partitioning of space by considering the size associated with a site, and thus, it is capable of mimicking an arbitrary size distribution. This mosaic differs from a Voronoi tessellation in the definition of the planes delimiting the regions of the tessellation. It can be noticed from Eq 1 that for a Voronoi tessellation, this plane intersects in the midpoint between two sites (bisecting plane), which results in tessellations with a narrow size distribution because all Voronoi regions tend to have similar grain sizes despite randomly distributed sites. Conversely, in the radical plane method, the boundaries are defined by the set of points with equal power k^2 with respect to two spheres or circles (Fig. 4a). The power of point Q with respect to circle C_1 in Fig. 4a is defined as $k_1^2 = \overline{P_1Q}^2 - r_1^2$. For instance, the points fulfilling the condition $k_1^2 = k_3^2$, that is, $\overline{P_1Q}^2 - r_1^2 = \overline{P_3Q}^2 - r_3^2$, form the radical line for the circles C_1 and C_3 (Fig. 4a). This line is perpendicular to the center line at Q. It can also be noticed that for touching spheres, the radical plane is tangential to the spheres (Ref 49). This causes the cells of the tessellation to have a size similar to that of the spheres, and thus, the size distribution of the spheres essentially determines the size distribution of the tessellation (Fig. 4b). The topological properties of the radical plane construction are the same as for the Voronoi tessellation; that is, a vertex belongs to four cells (quadruple junction), an edge to three (triple line), and a face to two cells (grain boundary). Hence, tessellations are excellent tools for the generation of microstructures to be used in simulations.

Discretization of 2-D Grain Boundaries. For the discretization of 2-D and 3-D grain boundaries, it is necessary to first define the resolution, that is, the necessary minimal length. This length not only serves the purpose of defining the density of the grain boundaries (Fig. 5) but also triggers the topological transformations, as addressed later. Because the resolution remains constant during a simulation, operations that reduce or increase the density of nodes composing the grain boundary must
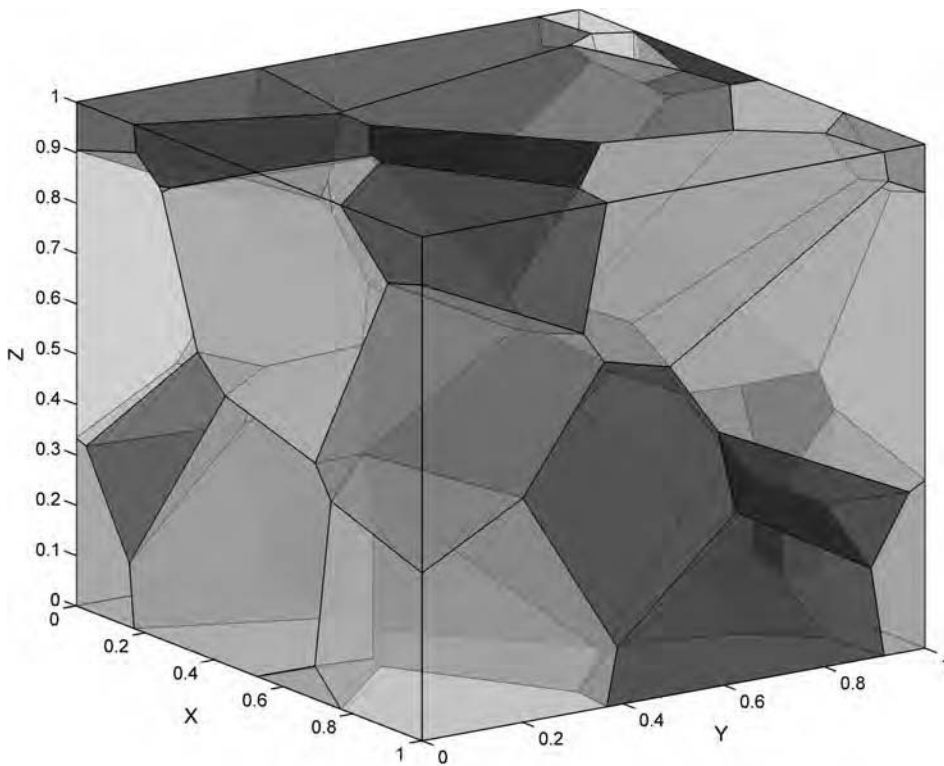
Fig. 3 Three-dimensional Voronoi tessellation. Model in color
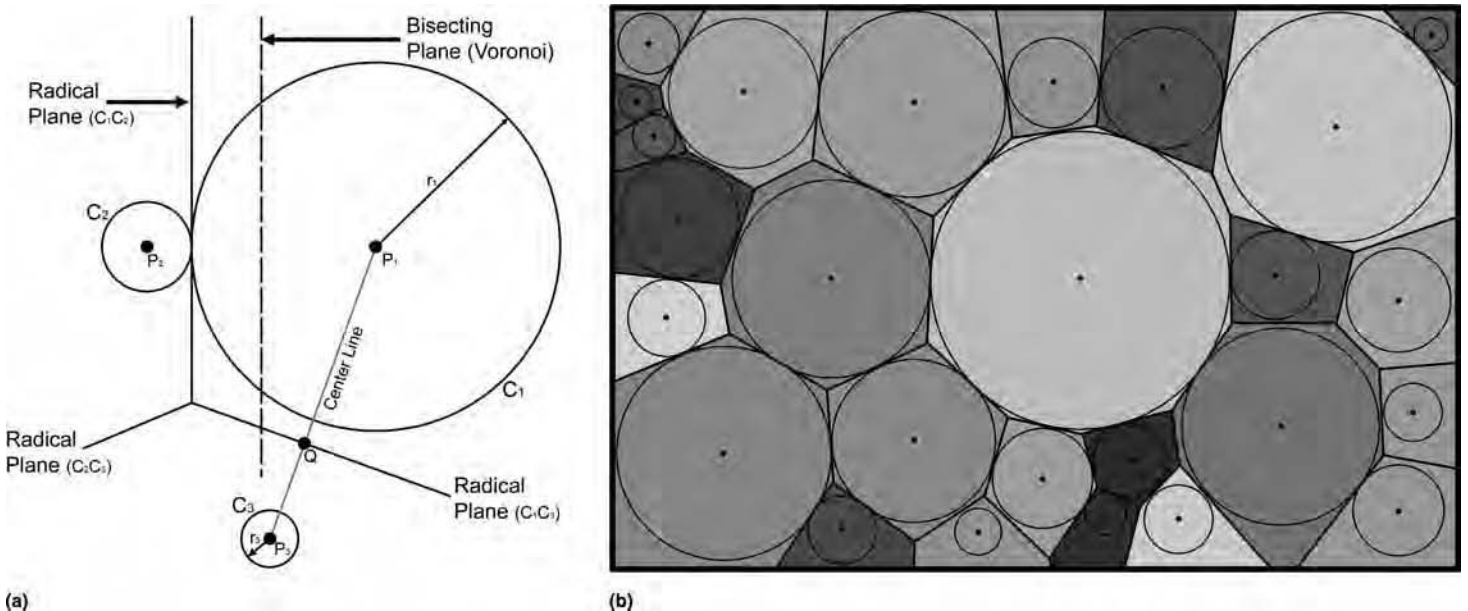
Fig. 4 (a) Definition of the plane delimiting a cell. The solid lines represent the radical planes; the hashed line denotes the bisecting plane. (b) Tessellation obtained with the radical plane method. The size of the cells matches approximately the size of the circles. Model in color

be implemented. For instance, assume a given resolution Δr. Once the spacing between next-neighbor nodes of a grain boundary falls below the critical size of $0.5\Delta r$, the corresponding segment loses any physical meaning for the simulation and can be deleted. This procedure is illustrated in Fig. 5(b), where the segment $\overline{lm}$ has reached a length less than the critical size and therefore is deleted. This requires the deletion of point l and the relocation of point m to the midpoint, m', of the original segment $\overline{lm}$.

The inverse operation is shown in Fig. 5(c), where the segment $\overline{no}$ has a length larger than $2\Delta r$. In order to keep an essentially uniform distance between nodes, a new node, p, at the midpoint of the segment is introduced. As seen in Fig. 5(c), the segment $\overline{no}$ is now replaced by the two segments $\overline{np}$ and $\overline{po}$, both of which have approximately a length of Δr. It is stressed that the first operation (shortening) is absolutely necessary. This is because the numerical error increases when the nodes approach each other, and, depending on the solver of the equation of motion, this can lead to either large deviations from the real solution or critical slowing down of the simulation. It is also noted that the shortening operation triggers the necessary topological transformations, since a grain boundary with a length equal to Δr has zero node density, and thus, its shortening represents the formation of a quadruple junction. By

contrast, the second operation is less important to be implemented, because the density of nodes along the grain boundary does not drastically affect the results of the simulations (Ref 50).

Discretization of 3-D Grain Boundaries. In the 3-D case, the grain boundaries are internal surfaces and can be discretized into triangular facets. In the 2-D case, grain boundaries and triple junctions constitute the microstructure, while in 3-D an additional element must be added, namely, the quadruple junction. A quadruple junction is the geometrical point where four grains meet. In addition, grain boundaries and triple junctions now require a 3-D description; that is, a grain boundary is represented by a surface, and the triple junction now has the character of a line. All these elements are shown in Fig. 6.

The procedure for the discretization depends on how the initial input was provided. Such input must deliver the spatial positions of the grains (center of mass), the grain boundaries enclosing it, and the coordinates of the vertices conforming the grain boundaries. If any of this information is missing, a construction of the grain arrangement becomes impossible. If this information is available, the first step is to discretize the grain boundaries in triangular facets. The choice of the triangle as the minimal surface element is because this geometrical element is the only one with all corners on the same plane (i.e., all corner points of the triangle

are coplanar). The selection of other surface elements is possible, but the calculation of velocity of such elements and the topology are more complicated, because for more complex geometries, some of the corners of the discretizing element may be noncoplanar.

The grain boundaries are first discretized by a segmentation of the faces of the polyhedra that are generated by the Voronoi tessellation. For this, it is necessary to first calculate the geometric centroid O of the grain boundaries in order to construct the triangular facets. In Fig. 7, a polyhedron and the discretization procedure of its faces are shown. The first step is depicted in Fig. 7(b). Fuchizaki et al. (Ref 16) and Weygand et al. (Ref 18) discretized the grains only as shown in Fig. 7(b). However, the subdivision of the grain boundaries in only these coarse triangular facets cannot adequately represent the curvature of the grain boundaries.

To overcome this problem, a recursive and continuous segmentation in further triangular facets is introduced. As in the 2-D case, a minimal length Δr of the edges of the triangular facets is first defined. This length is not only necessary for the discretization of the grain boundaries but is also required for triggering the topological transformations. The procedure for the segmentation of the facets is as follows. When the distance between two vertices is larger than $2\Delta r$, a new vertex in the middle of the segment is introduced, and the facets attached to the edge are then split into two. The procedure is repeated until all edges have a length of approximately Δr. In Fig. 8, an arbitrary grain boundary, whose shape corresponds to a regular hexagon, is discretized in triangular segments that are generated from the edges and the geometric centroid of the hexagon (Fig. 8a). Because the length of all edges is much larger than $2\Delta r$, a new vertex (denoted by "1" in Fig. 8b) is introduced on one of the edges of the grain boundary. Correspondingly, the facet is sectioned into two new triangles (dotted line connecting 1 to O); the same operation is repeated for all edges, for example, 2 and 3. Thereafter, the diagonals from O to the vertices of the edges are split (4 and 5), and finally, the first edge introduced at the beginning of the segmentation procedure is sectioned (6). The repetition of the process leads to the discretization of the grain boundary, as shown in Fig. 8(c). The length of all edges is now very close to Δr. It is stressed that the
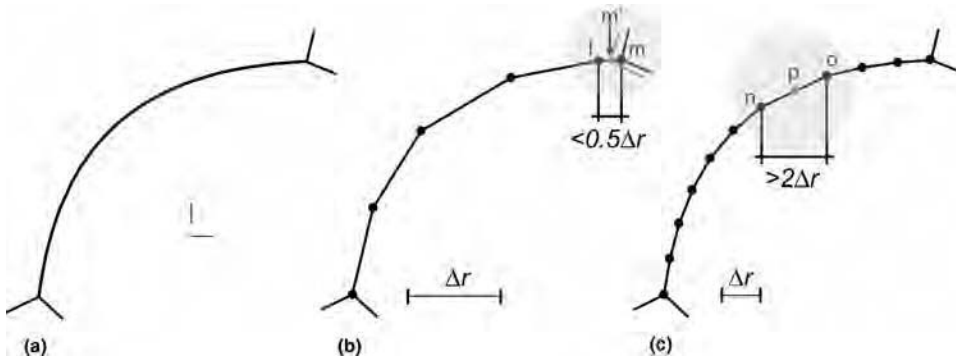
Fig. 5 Resolution of grain boundary. (a) Continuous grain boundary. (b) Discretization of the grain boundary with a low resolution and (c) with a high resolution. The operations for the elimination of grain-boundary segments and introduction of new nodes, respectively, are shown.
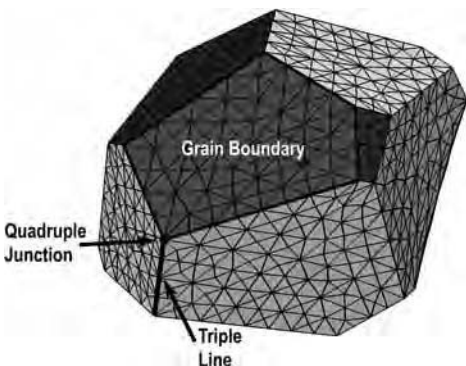
Fig. 6 Discretized boundaries of a grain and their junctions with adjacent grains. Model in color
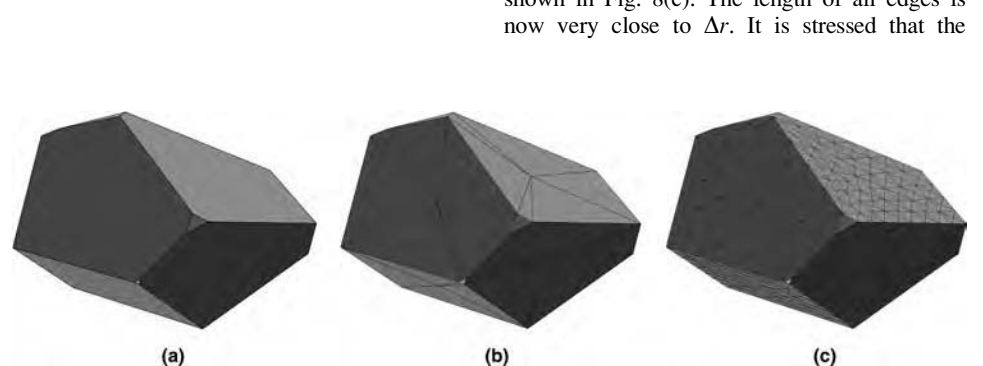
Fig. 7 Sequence of the discretization procedure. (a) Original polyhedron. (b) Grain after the introduction of the first triangular facets. (c) After sequential segmentation. Model in color
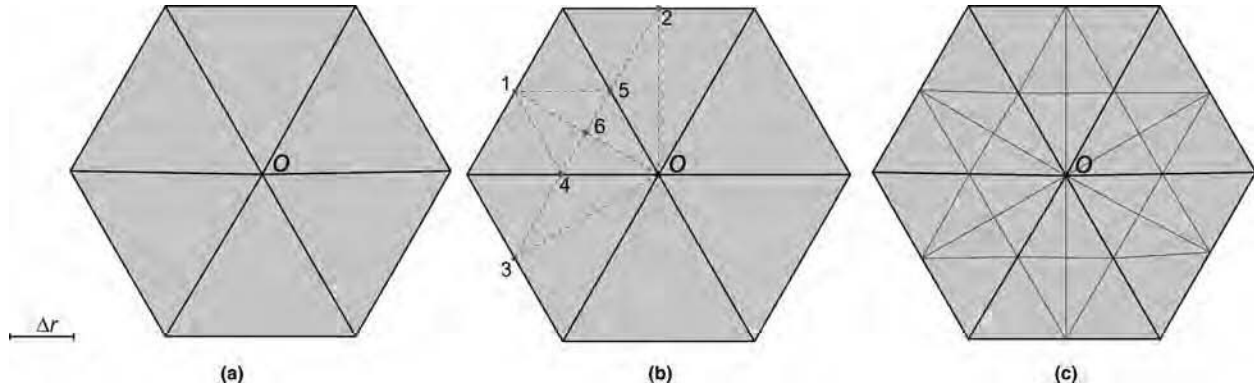
Fig. 8 Adding Triangular facets. (a) Arbitrary grain boundary after initial discretization. (b) The distance between two points cannot be larger than twice the predefined critical distance Δr. (c) Result after the segmentation procedure
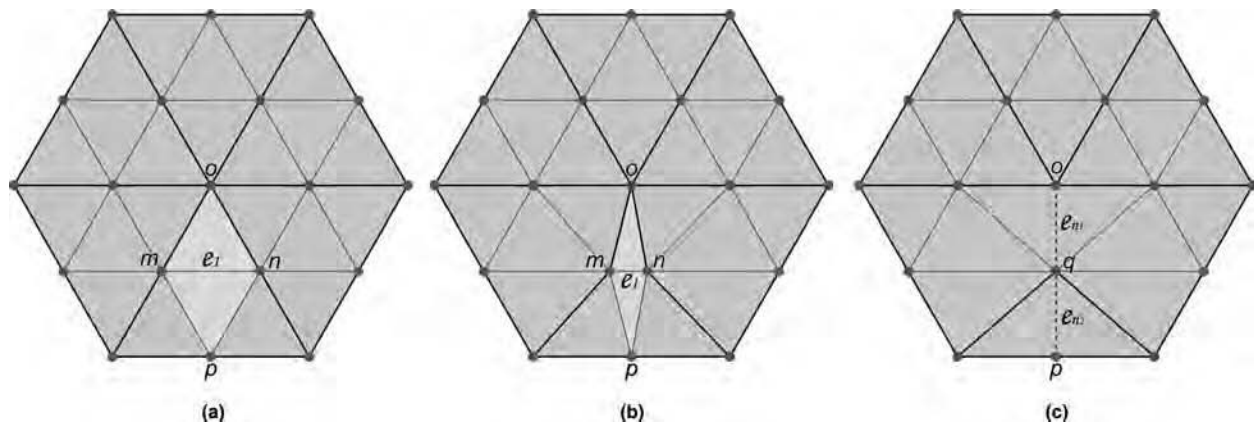
Fig. 9 Facet elimination process that takes place once a facet edge is smaller than the critical size. (a) Initial grain boundary. (b) Distance between two nodes is less than 0.5Δr. (c) Grain boundary after the elimination of the short edge and thus of the adjoining facets

procedure described here is only one of many possible segmentation procedures (Ref 29).

This segmentation procedure can be used to increase the density of triangular facets (and nodes) when a surface expands, similar to the 2-D case when the length of a grain-boundary segment increases and is then split (Fig. 5c). Conversely, a procedure to eliminate triangular facets once they have shrunk below a critical size is necessary as well. This process of facet elimination is the fundamental mechanism for the reduction of the grain-boundary area and for triggering the topological transformations when certain conditions are met.

To explain facet elimination, refer to Fig. 9 (a), which shows a hexagonal surface that has been discretized into regular triangular facets. Consider the facets attached to the edge e_1, that is, the triangles Δmno and Δmnp. Due to attraction forces, the points m and n approach each other (Fig. 9b). The distance between them becomes eventually smaller than $0.5\Delta r$, which is the condition set to trigger facet elimination. As the length of the edge e_1 tends to zero, so does the area of Δmno and Δmnp. For this reason, it is now possible to remove these facets. The point q located at the midpoint of e_1 replaces the vertices m and n (i.e., the length of

the edge e_1 is set to zero). The attached facets are removed and replaced by the edges e_{n1} ($\overline{oq}$) and e_{n2} ($\overline{pq}$), as shown in Fig. 9(c) by the dotted lines. A topological transformation can be triggered when the facet elimination procedure meets determined conditions that alter the topology of the system. In the current case, only two facets are involved (highlighted in Fig. 9a and b), because the edge e_1 lies on the grain boundary. However, if the considered edge belongs to a triple line, then three facets (lying on the three adjacent grain boundaries) are involved.

Equation of Motion

Equation of Motion for the 2-D Model. In the 2-D implementation of the model, the equation of motion for the grain boundary and for the triple junctions is considered separately. A grain boundary moves with a velocity (V):

$$\nu = m_{gb}\, p \tag{Eq 2}$$

where m_{gb} is the grain boundary mobility, and p is the driving force. For the general case when more than one driving force is applied to the grain boundary, the velocity reads:

$$\nu = m_{gb} \sum_{i}^{n} p_i \tag{Eq 3}$$

Because the capillary driving force $p = T \cdot \kappa$, owing to grain-boundary curvature, is intrinsic for polycrystals, separating it out is recommended:

$$\nu = m_{gb}T\kappa + m_{gb} \sum_{i}^{n-1} p_i$$
$$= m_{gb}\kappa\left(\gamma + \frac{\partial^2 \gamma}{\partial \varphi^2}\right) + m_{gb} \sum_{i}^{n-1} p_i \tag{Eq 4}$$

where κ is the curvature of the grain boundary, and T is the line tension, which is the sum of the grain-boundary energy, γ, and its second derivative with respect to the inclination angle, φ, which defines the spatial orientation of the grain boundary. It is noted that the boundary energy and mobility depend on grain-boundary misorientation.

Consider the grain boundary in Fig. 10. The variables used in Eq 4 are also indicated. To explicitly define Eq 4, only the curvature of the grain boundary is needed. The curvature can be calculated from the local geometry at each point b_i along the grain boundary with the adjacent points b_{i-1} and b_{i+1} (Ref 23, 51).

The path vector or tangential vector describes the position of a discretizing point of the grain boundary with respect to the global coordinate system. In Fig. 10, the components of this vector are t_x and t_y, and hence, the vector normal to the boundary at point b_i is $(-t_y, t_x)$. The normal velocity is then given by:

$$\vec{\nu}_n = \begin{pmatrix} -t_x \\ t_y \end{pmatrix} m_{gb} \kappa \left(\gamma + \frac{\partial^2 \gamma}{\partial \varphi^2} \right) \qquad \text{(Eq 5)}$$

In some cases, it is inconvenient to explicitly calculate the curvature (e.g., irregular grain boundaries), so instead, Eq 5 is reduced to a more fundamental expression reflecting only the fact that the capillary driving force stems from the line tension of the grain boundary. This also permits the possibility of taking into account the stiffness of the grain boundary. In Fig. 11, the forces exerted on a node of the grain boundary are shown. Apparently, these forces stem from two sources: the line tension of the boundary and the torque, M, experienced by a boundary segment due to the anisotropy of the energy with the inclination angle (Ref 52, 53). The velocity of a node can be determined from the sum of the acting forces:

$$\nu = m_{gb} \left[\frac{\left(\vec{F}_1 + \vec{F}_2 \right)}{\Delta s_{12}} - \sum_i^n M_i(\varphi)/\Delta s_i \right]$$
$$\approx m_{gb} \left[\gamma(\vec{u}_1 + \vec{u}_2)/(\Delta s_1 + \Delta s_2) - \sum_i^n M_i(\varphi)/\Delta s_i \right]$$
$$\text{(Eq 6)}$$

where the vector $\vec{u}$ is the unit vector of $\vec{s}_i$:

$$\vec{u}_i = \frac{\vec{s}_i}{\|s_i\|} \qquad \text{(Eq 7)}$$

and the sum of the unit vectors $(\vec{u}_1 + \vec{u}_2)$ defines the direction of motion. The torque $M(\varphi)$ is the first derivative of the energy with respect of the inclination angle:

$$M(\varphi) = -\frac{\partial \gamma}{\partial \varphi} \qquad \text{(Eq 8)}$$

As seen in Ref 52 and 53, Eq 6 is equivalent to Eq 5 but with the advantage that, by avoiding the calculation of the curvature, it is possible to consider the dependency of grain-boundary energy on misorientation and inclination.

The equation of motion of the triple junction can be calculated in a very similar way. The velocity of the triple junction is given by the product of the triple-junction mobility and the sum of forces acting on the triple junction:

$$\vec{\nu}_{tj} = m_{tj} \sum_i^n \left(\vec{F}_{gb}^i + \vec{p}_i \right) \qquad \text{(Eq 9)}$$

where the line tensions, $\vec{F}_{gb}^i$, of the boundary and any external force, $\vec{p}_i$, exerted on the triple junction have been separated; m_{tj} is the triple-junction mobility. As for the equation of motion for the grain boundaries, it is possible to expand Eq 9 to include the forces $\vec{F}_{M_i}$ acting on the triple junction due to the torque:

$$\vec{F}_{gb} = \sum_i^n \frac{\gamma_i \cdot \vec{t}_i}{\Delta s_i} + \vec{F}_{M_i} = \sum_i^n \frac{\gamma_i \cdot \vec{t}_i - M(\varphi)_i \cdot \vec{n}_i}{\Delta s_i}$$
$$\text{(Eq 10a)}$$

where $\vec{t}$ and $\vec{n}$ are the tangential and normal unit vectors, respectively, at the triple junction of a grain-boundary segment. In terms of the
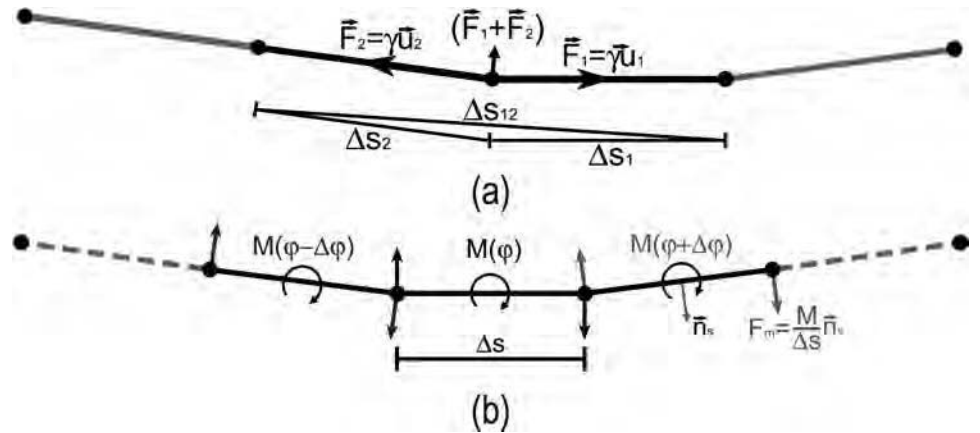
Fig. 11 Schematic representation of the forces at a node of a grain boundary. (a) Tensile forces. (b) Each rigid grain-boundary segment experiences a torque due to the anisotropy of the grain-boundary energy with the inclination angle, which can be substituted by a pair of forces applied at the ends of the boundary segments (nodes).
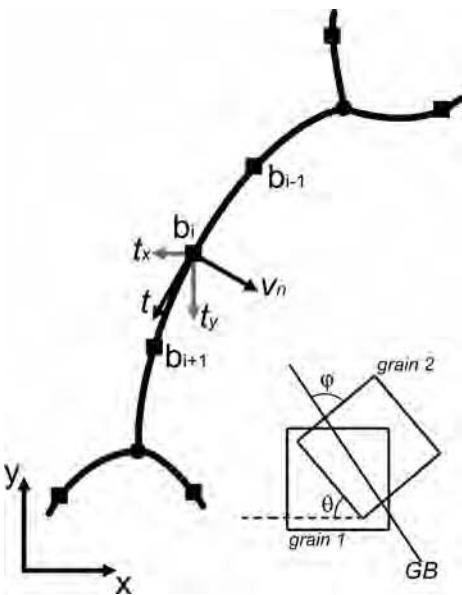
Fig. 10 Discretized grain boundary. t_0 represents the tangential vector to the boundary at point b_i; v_n is the normal velocity at b_i; and t_x and t_y are the trajectory/path vectors along the boundary. As seen in the sketch in the right side of the figure, the angle θ represents the misorientation of the grain boundary (orientation relationship of the neighboring grains) whereas φ is the spatial position of the grain boundary (inclination).
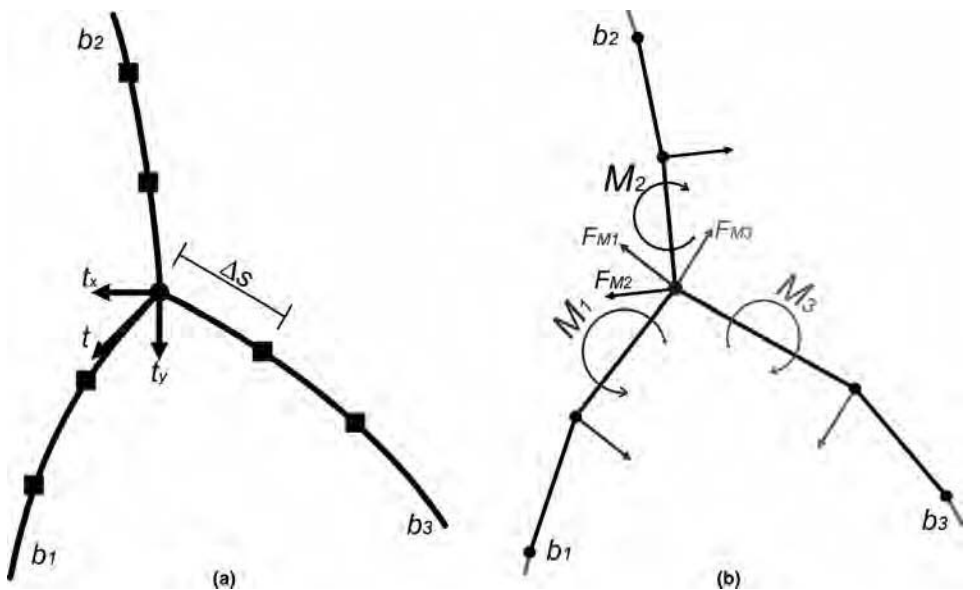
Fig. 12 Triple junctions, (a) Variables used for calculation of triple-junction velocity. (b) The torque caused by the dependency of the grain-boundary energy on the inclination angle induces another force to be considered at the triple junction

variables depicted in Fig. 11 and 12, Eq 10(a) can be rewritten as:

$$\vec{F}_{gb} = \sum_i^n \frac{\gamma_i \cdot \begin{pmatrix} t_{xi} \\ t_{yi} \end{pmatrix} - M(\varphi)_i \cdot \begin{pmatrix} n_{xi} \\ n_{yi} \end{pmatrix}}{\Delta s_i} \quad \text{(Eq 10b)}$$

It is noted that the sign of the torque, M, determines the direction of the normal vector $\vec{n}$. In turn, Δs is the length of a segment of boundary i adjacent to the triple junction (Fig. 12). The application of Eq 9 will lead to an equilibrium of forces at the triple junctions. For instance, under isotropic conditions, the dihedral angles between grain boundaries will be 120°, and thus, no motion of the triple junction will occur. However, the motion of the curved grain boundary introduces a perturbation of the equilibrium of forces that consequently leads to motion of the triple junction. It is stressed that in this model, the motion of the triple junctions does not result from artificially enforcing an equilibrium of the dihedral angles, but instead, the equilibrium angle will be attained by the displacement of the triple junction (Ref 54).

Equation of Motion for the 3-D Model. The elemental discrete unit for 3-D grain boundaries is a triangle. The surface of such a triangle exerts a force on its vertices to decrease its area. These forces at vertices (nodes) arise from the surface tensions of the triangular facets that adjoin the respective vertex and can be calculated from the local geometry (Ref 23). Figure 13 shows a vertex and the facets attached to it. The force F_{f1} exerted by the shaded facet f_1 on the vertex P_0 is given by:

$$\vec{F}_{f_1} = \frac{\gamma}{2} \cdot \frac{\vec{s}_0 \times (\vec{s}_1 \times \vec{s}_0)}{\|\vec{s}_1 \times \vec{s}_0\|} \quad \text{(Eq 11)}$$

where, as in the 2-D case, γ is the grain-boundary energy, and $\vec{s}_0$, $\vec{s}_1$, $\vec{s}_2$ are the edge vectors of the triangular facet, which are fully determined by the position of its vertices (P_0, P_1, and P_2), as indicated in Fig. 13.

The sum of the forces over all facets attached to a vertex leads to the net force:

$$\vec{F}_{sum} = \frac{\gamma}{2} \cdot \sum_{i=1}^n \frac{\vec{s}_{0i} \times (\vec{s}_{1i} \times \vec{s}_{0i})}{\|\vec{s}_{1i} \times \vec{s}_{0i}\|} \quad \text{(Eq 12)}$$

with n being the total number of facets meeting at P_0. Then, the velocity of the vertex reads:

$$\vec{\nu} = m_{eff} \cdot \vec{F}_{sum} = \frac{1}{D_f} \cdot \vec{F}_{sum} \quad \text{(Eq 13)}$$

where m_{eff} is the effective mobility of the vertex, and D_f is the drag factor, which is defined as:

$$D_f = \frac{A_n}{m_{gb}} + \frac{\Delta s}{m_{tl}} + \frac{1}{m_{qp}} \quad \text{(Eq 14)}$$

where m_{gb}, m_{tl}, and m_{qp} are, respectively, the grain-boundary, triple-line, and quadruple-junction mobilities; Δs is the mean spacing of vertices on a triple line; and A_n is the projected area of the surrounding facets perpendicular to the direction of motion. Depending on the position of a node, it can be affected by more than one mobility. For instance, a vertex acting as a quadruple junction has its own mobility, but its motion also causes a displacement of small portions of the attached triple lines and grain boundaries. Therefore, their respective mobilities must be considered as well as indicated in Eq 14. The motion of a node lying on the surface of a grain boundary causes exclusively a grain-boundary displacement. In such case, the triple-line and quadruple-junction mobilities are regarded as infinite.

The discretization of internal surfaces by triangular facets also permits the calculation of the metrics of a grain with simple vector operations. These geometrical properties of a grain are relevant to an analysis of 3-D grain growth, as addressed in subsequent sections.

Topological Transformations

The topological properties of the crystals change during the course of grain growth because they lose faces (grain boundaries), edges (triple lines), and vertices (quadruple junctions). One fundamental operation during grain growth that triggers a change of the topology of a polycrystal is the loss of a triangular grain boundary. When a triangular grain boundary collapses, the edges shared with other grains are deleted, and the Euler formula is no longer valid, because the collapsing grain boundary produces an edge with more than three grain boundaries. (The Euler formula gives a general formulation of 3-D topological networks with a specific number of grains, g, grain boundaries, f, triple lines, e, and quadruple junctions, v, and reads $e + g = f + v$. Ref 55). Consequently, the topological order must be restored. For this, new triangular grain boundaries are created, and edges with multiple adjoining grain boundaries are reconnected until the topology is restored; that is, the Euler formula is valid again. This structuring is called topological transformation.
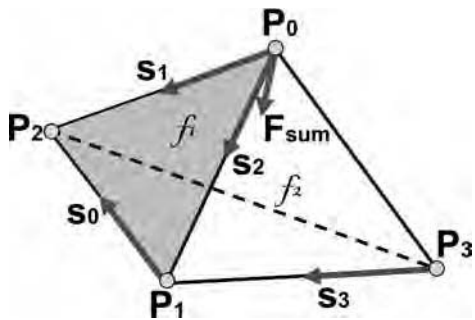
In nature, this is managed simply by atomic rearrangements to generate the configuration with the lowest energy. This problem is automatically taken care of in Monte Carlo and phase-field models because these models follow the steepest gradient of the free energy.

Because the microstructure in network models is only represented by connected points, the collapse of a boundary means that the distance between two triple junctions approaches zero. When this occurs, the network must be reconnected in order to minimize the free energy. The most difficult part of the implementation of network models is the programming of the topological transformations. The basic transformations are, however, well documented (Ref 56), and their implementation has succeeded to a great extent (Ref 15, 16, 20, 24). There are a finite number of transformations only if the grain-boundary energy is constant (isotropic case). If the energy of the grain boundaries is not uniform (anisotropic case), the occurrence of a transformation will depend on the energetic equilibrium of the grain boundaries attached to an edge. A transformation will only take place if it represents a decrease of the free energy of the system. In an anisotropic system, the Euler formula is not valid anymore, because multiple junctions can be stable, even if only for a very short time.

Basic Topological Transformations in 2-D. For an implementation of 2-D vertex models, normally only two topological operations must be considered (Ref 11): neighbor switching (Fig. 14a) (usually denoted as T1) and triangle elimination (T2) (Fig. 14b). These operations are triggered when the distance Δs between triple junctions falls below a certain predefined critical value, ε. Operation T2 is not necessary because it can be achieved by successive applications of T1 to the boundaries of the triangular grains. However, this operation is normally implemented because it reduces the complexity

Fig. 13 Three-dimensional grain with grain boundaries discretized in triangular facets. The forces at triple junctions result from the surface tensions of the attached facets.
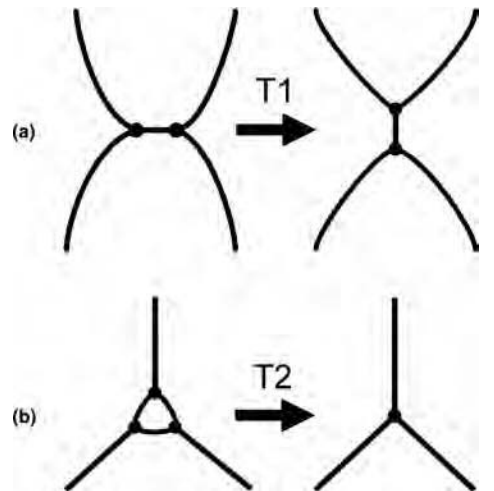
Fig. 14 Two-dimensional topological transformations. (a) Neighbor switching, T1. (b) Triangle elimination, T2
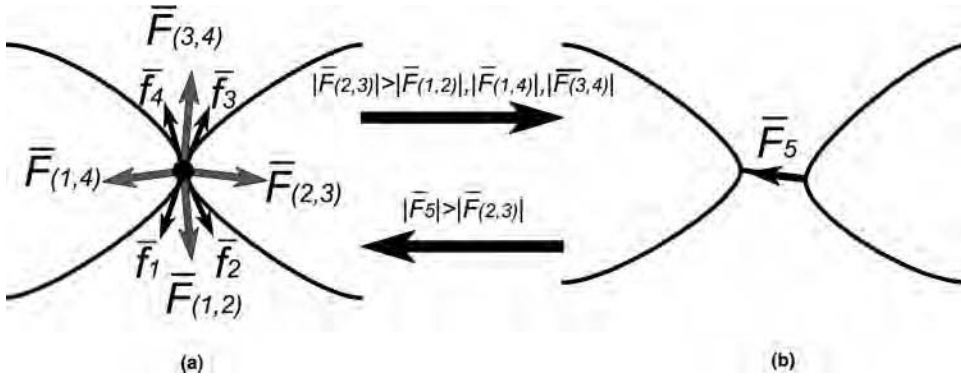
Fig. 15 Quadruple junction. (a) A quadruple junction is allowed to form; the forces $\vec{f}_1, \vec{f}_2, \vec{f}_3, \vec{f}_4$ and $\vec{f}_{ij} = \vec{f}_i + \vec{f}_j$ are calculated. (b) If the split of the quadruple junction leads to an increase of the free energy of the system, the quadruple junction is a stable configuration.
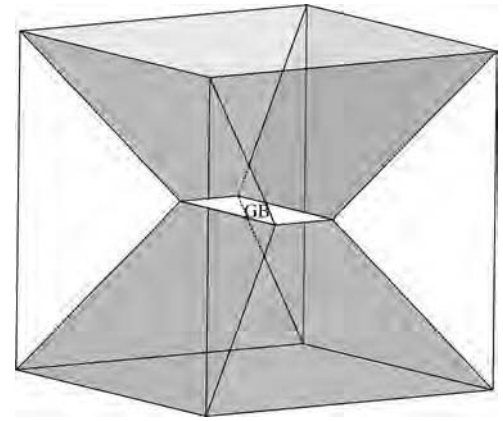
Fig. 16 To demonstrate topological transformations, a small system is used. Six grains fill the shown cube; in the figure, only two of them are shown, but the remaining grains completely fill the space. GB, grain boundary. Image in color
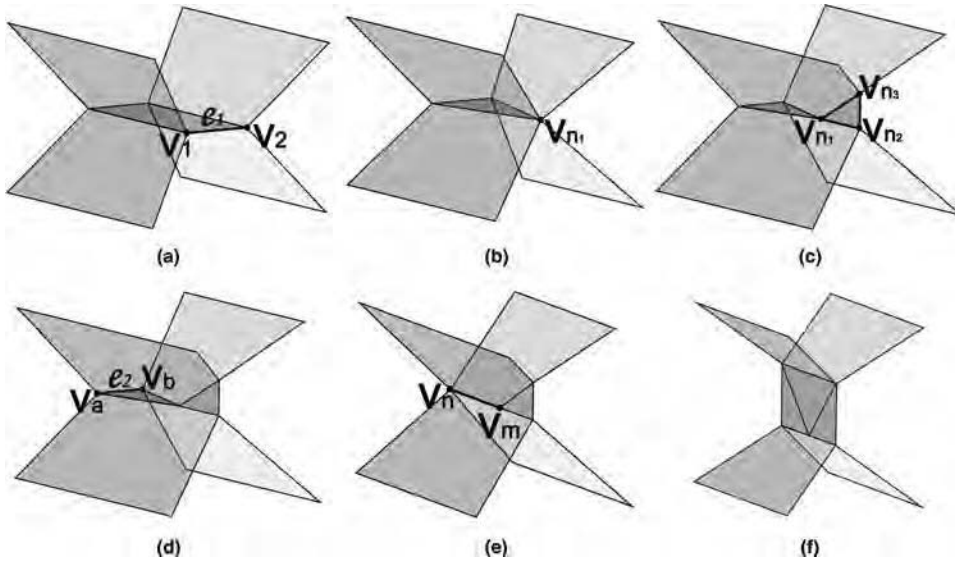
Fig. 17 Fundamental topological transformations. (a)–(c) Triple-line collapse: (a) original configuration, (b) collapse of the triple line e_1, and (c) restoration of the topology by the introduction of a new triangular grain boundary. (d) and (e) Grain-boundary collapse: (d) configuration prior to the collapse, and (e) collapse and formation of a quadruple line (v_n-v_m). (f) Restoration of the correct topology by introduction of the necessary triangular facets

of the model, whereas the topology of the network remains intact.

It is emphasized that T1 introduces a nonphysical procedure by avoiding the formation of a quadruple junction. Usually, such junctions are unstable and automatically split into two triple junctions and a new boundary. However, under anisotropic conditions, where the energies of the grain boundaries can vary, this may not be the case.

To overcome this problem, it is possible to allow the formation of the quadruple junctions and to formulate a rule that will determine whether the split is energetically favorable. The procedure is as follows. Once a quadruple junction is formed, the forces, $\vec{f}_i$, exerted on the junction by each attached grain boundary are calculated. The largest sum of the forces of two adjacent grain boundaries of a grain, denoted as $\vec{F}_{ij}$, allows determining the direction of the split. However, if the formation of the

new grain boundary leads to an increase of free energy of the system, the quadruple junction is considered stable. This procedure is schematically shown in Fig. 15.

Basic Topological Transformations in 3-D. For an illustration of the basic topological transformations in 3-D, the grain system is shown in Fig. 16. This system is composed of six grains, but only two of them are shown. The rectangular grain boundary between them is used to exemplify the basic topological transformations.

There are three basic topological transformations in the 3-D case: triple-line collapse, grainboundary collapse, and grain collapse. A tripleline (e_1) collapse occurs when the distance between two quadruple junctions (v_1 and v_2 in Fig. 17a) approaches zero, provided that the grain boundaries attached to the triple line e_1 have more than three sides. Once the triple line collapses, the point v_{n1} is formed (Fig. 17b),

which connects nine grain boundaries and five grains and thus violates the Euler formula. The operation necessary for the restoration of the topology is shown in Fig. 17(c). A new triangular grain boundary (formed by v_{n1}, v_{n2}, and v_{n3}) is introduced, and the topological net is reconnected to this new element.

The second operation takes place when a triple line (formed by v_a and v_b in Fig. 17d) of a triangular grain boundary collapses. This event leads to the formation of a quadruple line (v_n-v_m) that, as in the 2-D case, must split into two energetically favorable triple lines. This operation is achieved by introduction of the necessary triangular facets that split the grain boundaries, as shown in Fig. 17(f).

The collapse of a tetrahedral grain does not represent an alteration of the topology, because this event leaves the topology of the network intact. However, grains with more than four faces can also instantly collapse if their size is comparable to the minimal size allowed in the simulation. The collapse of such grains does not leave an intact topology, and thus, more restoration operations must be applied (Ref 57).

Applications

2-D Normal Grain Growth Simulation (Fig. 18, 19, 20). Grain growth is a phenomenon that frequently occurs in nature. In any event, the properties of the topological network and their change during the evolution of grain growth are of interest. In fact, the first network models (Ref 30) were developed for the study of soap froths (Ref 58). Many characteristics of the networks can be studied or measured by means of computer simulations, in particular, the grain-size distribution, the kinetics of grain growth, and the topological properties. Moreover, simulation of normal grain growth offers an easy method for the validation of theoretical models, which usually assume isotropy of grain-boundary energy and mobility. Under

such assumption, the phenomenon is sometimes denoted as ideal normal grain growth. For instance, for ideal normal grain growth, parabolic kinetics, a scaling regime, and the fulfillment of the von Neumann-Mullins equation are expected. In fact, in a comprehensive investigation, Maurice (Ref 59) studied the behavior of several network models and confirmed that all of them essentially meet these requirements.

For a rough impression of such simulations, Fig. 18 shows two snapshots of the simulated evolution of a polycrystal for different times during grain growth. A distinct coarsening of the grain arrangement can be observed in Fig. 18(b), and it can be anticipated that the simulation reaches a scaling regime because the microstructural evolution shown in Fig. 18 seems to reveal self-similarity also noticeable in the invariance of the grain size distribution with time (Fig. 20). Self-similar growth is characterized by the temporal evolution of the average grain size, $\bar{D}$, as a power law (Ref 12),

namely, $\bar{D} \sim t^{1/2}$. This follows from the assumption that the growth rate $\left(\frac{1}{2}\frac{d\bar{D}}{dt}\right)$ of a grain is proportional to $1/R$, where $1/R$ is the grain-boundary curvature, and $R \sim \bar{D}$. Thus, integration yields a parabolic growth law (Ref 60). Consequently, the evolution of the mean grain area, $\bar{A} \sim \bar{D}^2$, with time follows a linear function, as seen in Fig. 19. While all known network models predict linear kinetics for the mean grain area, they predict different slopes, $\bar{A}$, ranging from $0.5m_{gb}\gamma$ (Ref 7) up to $20m_{gb}\gamma$ (Ref 10), with $m_{gb}\gamma \leq \bar{A} \geq 2m_{gb}\gamma$ for most of the models (Ref 8, 11, 12, 18, 61). This is because the slope depends strongly on the features of the simulation model. However, the largest deviations are to be expected in models that do not consider the curvature of the boundaries at all (Ref 10). The scaling regime is defined by the invariance of the grain-size distribution with time. All network models have shown such behavior and reveal only slight differences in the shape of the grain-size distribution. An example for a simulated grain-size distribution for a microstructure with initially 10,000 grains is given in Fig. 20. The distribution remains the same, even if only 5000 and 2500 grains are left.

Because different implementations of the network models can lead to different results for the kinetics or grain-size distribution, the compliance with the von Neumann-Mullins relation (Ref 62) is a more definite test for a model. This relationship associates the growth rate (dA/dt) of a 2-D grain to only the number of sides of that grain, which is also referred to as the topological class, n (Ref 62):

$$\frac{dA}{dt} = m_{gb}\gamma\frac{\pi}{3}(n - 6) \qquad \text{(Eq 15)}$$

The equation simply means that the growth rate of the grains does not depend on the shape or the size of a grain but only on its topological class. Because this relation is an exact analytical solution for the growth rate, it must be obeyed by any simulation model. In Fig. 21, the simulated growth rate is plotted as a function of the topological class, n. The result is a straight line with an intersection at $n = 6$, as expected from Eq 15. The linear fit of the simulations yields a slope of $1.0452m_{gb}\gamma$. The calculated slope from the von Neumann-Mullins relation is $1.047m_{gb}\gamma$; hence, the deviation is below 1%. This accuracy denotes that the physics of grain growth have been accurately implemented in the network model. The largest deviations are found for grains with only three sides ($n = 3$) and grains with high topological class ($n > 9$). This is caused by numerical artifacts, because when a grain becomes very small, the curvature of its grain boundaries can no longer be properly represented. Because large grains (high n) are surrounded by many small grains, they are more prone to suffer deviations. For the same reason, large grains are also short-lived during a simulation, because the surroundings change rapidly and alter the topology of these grains. It is stressed, however, that the von Neumann-Mullins relation must be observed at any time and by any grain. The symbols in Fig. 21 correspond to
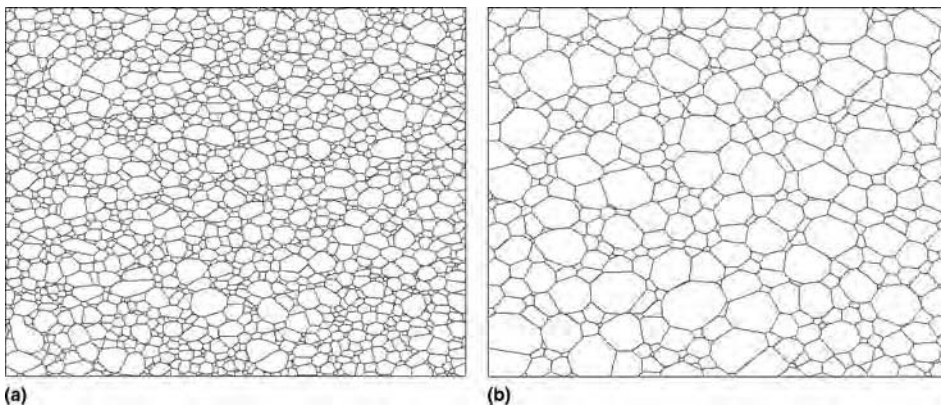
Fig. 18 Evolution of a simulated topological network during grain growth for two different times, t_a and t_b, where $t_a << t_b$
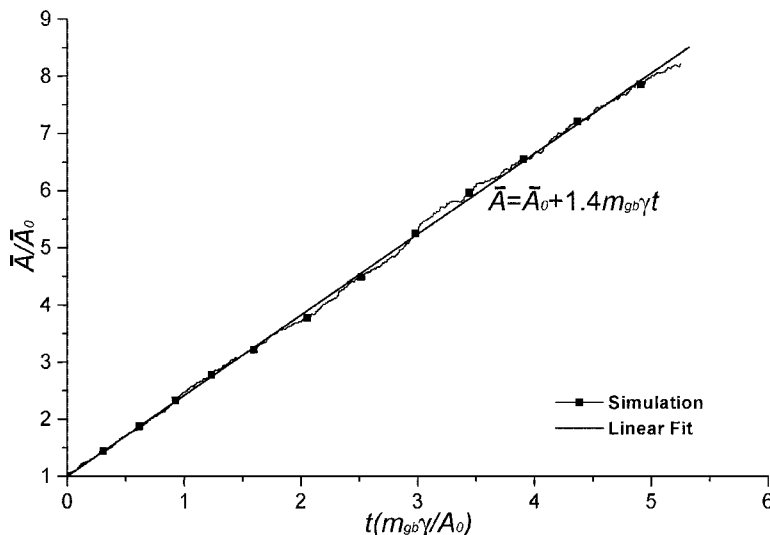
Fig. 19 Evolution of the mean grain area, $\bar{A}$, with time. The results show a linear dependence of $\bar{A}$ with time. Correspondingly, the mean grain size, $\bar{D}$, follows a parabolic behavior, $\bar{D} \sim t^{1/2}$. Source: Ref 61
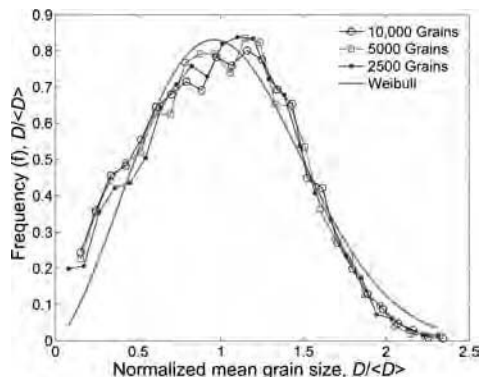
Fig. 20 Invariance of the grain-size distribution with time. Source: Ref 23

more than one million measurements taken for all grains of a microstructure during the whole grain-growth simulation.

Simulation of 3-D Normal Grain Growth. The simulation of 3-D grain growth with network models presents a considerable challenge, and, in fact, only a few attempts can be found in the literature. With such models, one can study the same properties for 3-D topological networks as addressed for 2-D approaches.

The mean grain size $\bar{D} = 2\bar{R}$ is still expected to grow proportionally to $t^{1/2}$, and thus, R^2 should yield a linear relationship with time. This has been corroborated by simulations of the few existing 3-D network models (Ref 16, 20, 57). Unfortunately, a comparison of these simulations is not possible, because details of

the simulations are usually not given in the references, but different slopes, $\bar{A}$, are also expected for the same reasons cited in the previous section. In Fig. 22, the grain-growth kinetics taken from (Ref 20) are shown. The symbols in this figure denote the simulation data, whereas the straight line is a linear fit with a slope of approximately 1.53 $m_{gb}\gamma$.

The 3-D grain-size distribution also reaches a steady state (Ref 16, 20, 57). The results (Fig. 23) not only attest to a self-similar growth of the grains but also prove that the shape of the steady-state grain-size distribution is essentially independent of the 3-D network model used. This means that the shape of the distribution is characteristic for ideal 3-D grain growth. However, it is stressed that the agreement of

these distributions from simulations with experimental results is rather poor. This is normally attributed to pinning factors and anisotropy of the grain-boundary properties (Ref 20), but a convincing proof is still missing.

As seen in the previous section for 2-D grain growth, it is possible to relate the growth rate of grains to their topology (Eq 15, the von Neumann-Mullins relation, Ref 62) under the assumption of isotropic grain-boundary energy, γ, and mobility, m_{gb}. For the 3-D case, several theoretical approaches have addressed this problem. Two different kinds of approaches can be distinguished:

- Approaches derived under the assumption that the volume evolution of grains within a topological class can be accurately described by an average value for the whole class
- Approaches that express the exact volume rate of change of grains as a function of their particular geometry

The only independent variable of the former models is the topological class, whereas the latter use the metrics of single grains. The models of Mullins (MU) (Ref 63), Hilgenfeldt et al. (HI) (Ref 64), and Glicksman-Rios (GR) (Ref 65) belong to the first category, whereas the models of Cahn and MacPherson-Srolovitz (CMS) (Ref 66, 67) fall into the second category. The latter approach is considered an exact 3-D extension of the von Neumann-Mullins equation and is also the analytical solution for the growth rate, in this case, of 3-D grains.

Network models lend themselves to an easy calculation of all necessary geometrical properties of the grains in a polycrystal for an evaluation of the CMS equation (Ref 24). This equation can be used to validate grain-growth models because it must be obeyed by any grain undergoing grain growth. In Fig. 24, an example with three different grains taken from a polycrystal simulation is shown. The symbols in this figure were extracted from the simulations; the solid lines were calculated with the
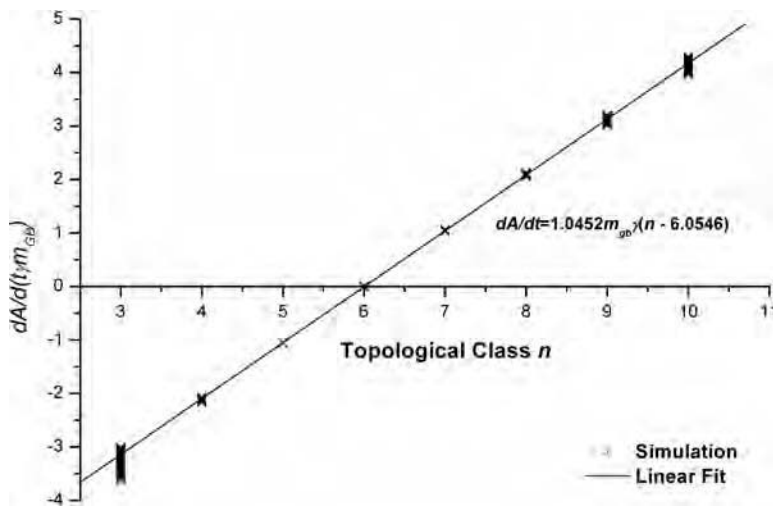
Fig. 21 Comparison of the simulation results with theoretical predictions of the von Neumann-Mullins relation taken from Ref 61

LIVE GRAPH
Click here to view

LIVE GRAPH
Click here to view
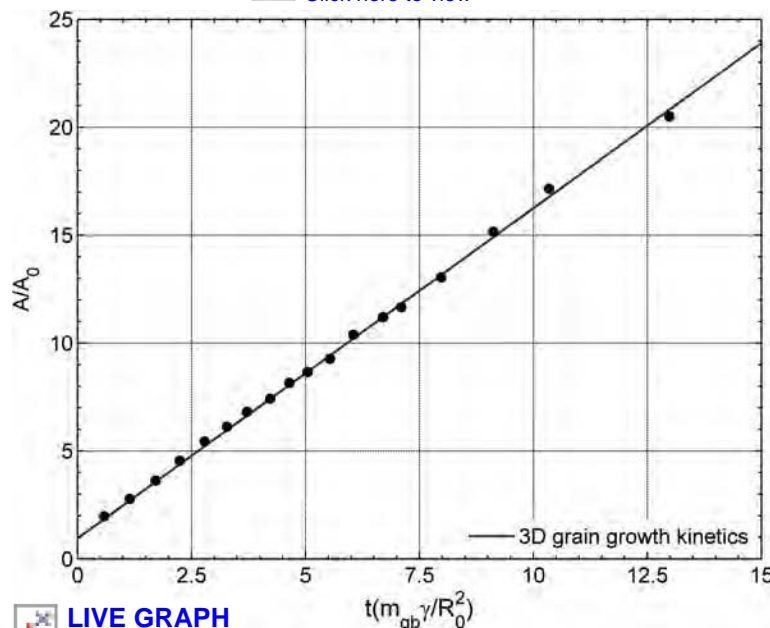
Fig. 22 Three-dimensional grain-growth kinetics with $A = \pi R^2$. Source: Ref 20
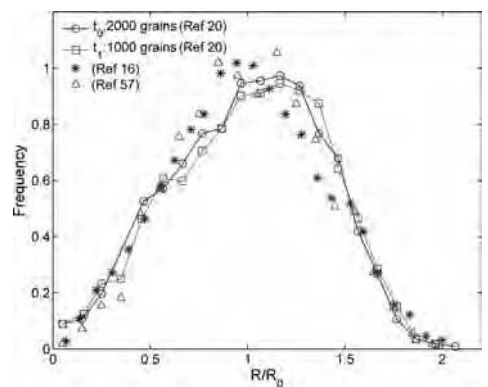
Fig. 23 Three-dimensional grain-size distributions. The distributions for different times, t_0 and t_1, demonstrate self-similarity. Three different network models are compared. All of them show a similar grain-size distribution.

LIVE GRAPH
Click here to view
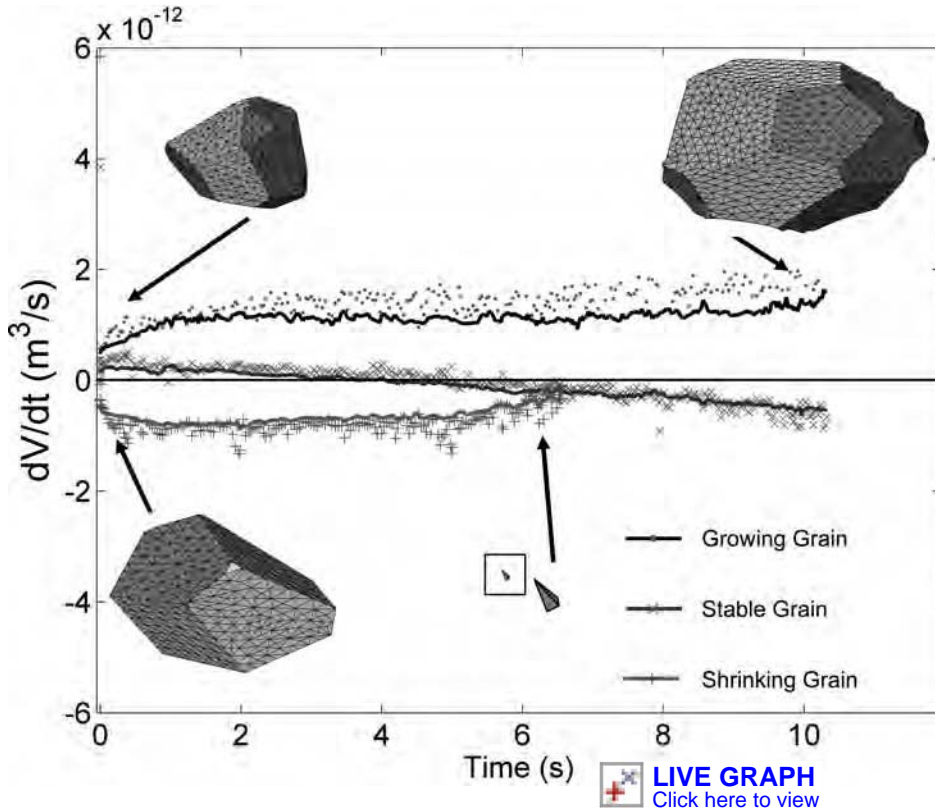
Fig. 24 Growth rate for three different grains in a polycrystal. The solid lines were calculated with the Cahn-MacPherson-Srolovitz equation. The symbols were obtained from the simulations. Source: Ref 24
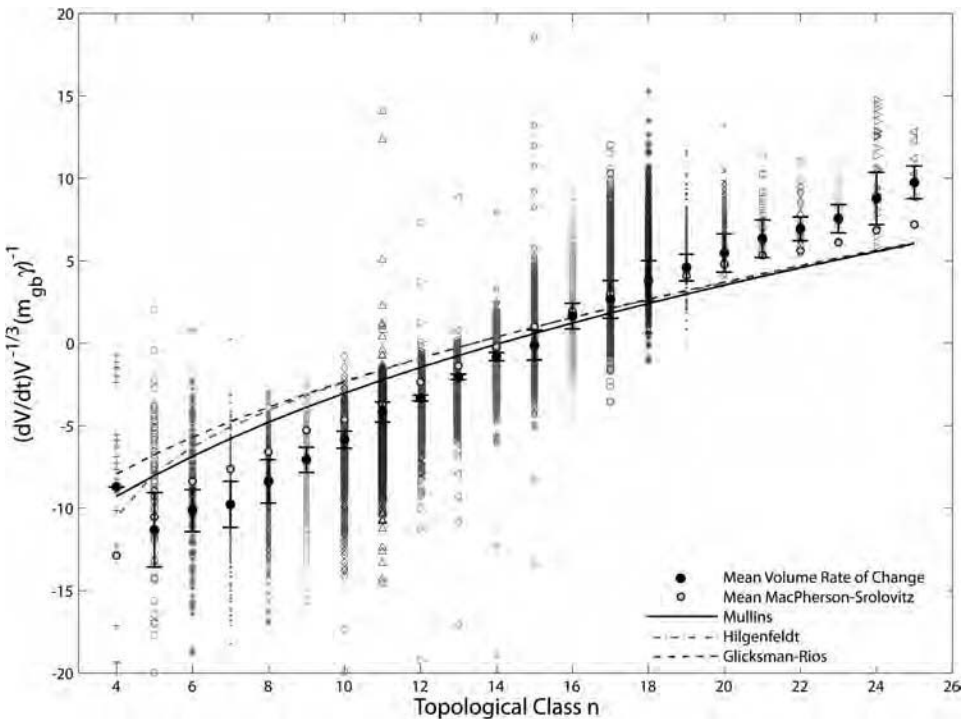
Fig. 25 Comparison of different approaches for prediction of three-dimensional growth rate. The agreement of the considered network simulation with all approaches is remarkable. Source: Ref 24

CMS relation. The simulations show a remarkably good agreement with the predictions of the equation.

The CMS equation points out that there cannot be a constant growth rate for a given topological class. Hence, grains with the same number of grain boundaries may exhibit different growth rates, depending on the difference between the mean width of the grain and the length of the triple lines, which change from grain to grain. However, other approaches indicate that the volume of grains within a topological class tends to change with an average constant rate for that class. The volume rate of change as a function of the topological class is plotted for all relations in Fig. 25 and compared to simulation results. It is evident that grains within a topological class can have various growth rates, because they may have different metrical properties in the same instant. All approaches show a very good agreement with the simulated mean value of each topological class. This has been confirmed independently by network model simulations by various authors (Ref 20, 57) and even in phase-field simulations (Ref 6).

Grain Growth under the Action of an External Force. Grain growth is intimately related to grain-boundary migration and therefore susceptible to the same factors that affect the movement of the grain boundaries. Notably, the various factors that cause grain-growth stagnation are related to some kind of impediment of grain-boundary motion. Grain-boundary motion can be caused by many different driving forces. In general form, any gradient of an intensive thermodynamic variable generates a driving force, for example, a gradient of temperature, pressure, density of defects, or energy density (Ref 68). In the context of grain growth, only capillary forces are usually considered, due to grain-boundary curvature. However, the action of other driving forces can considerably affect grain-microstructure evolution. For example, experiments on polycrystalline titanium samples (Ref 69–71) showed that annealing in a magnetic field can substantially modify crystallographic texture evolution during grain growth. In metals with low crystal symmetry, a magnetic field induces an orientation-dependent magnetic energy density, E_m, in a crystal:

$$E_m = \frac{1}{2}\mu_0 H^2 \chi \tag{Eq 16}$$

where μ_0 is the vacuum permeability of the free space, H is the magnetic field strength, and χ is the magnetic susceptibility.

A grain boundary is the interface between differently oriented crystals. Therefore, a grain boundary will experience a driving force to move toward the direction of the highest magnetic energy density:

$$p_m = \frac{1}{2}\mu_0\Delta\chi H^2\left(\cos^2\theta_2 - \cos^2\theta_1\right) \tag{Eq 17}$$
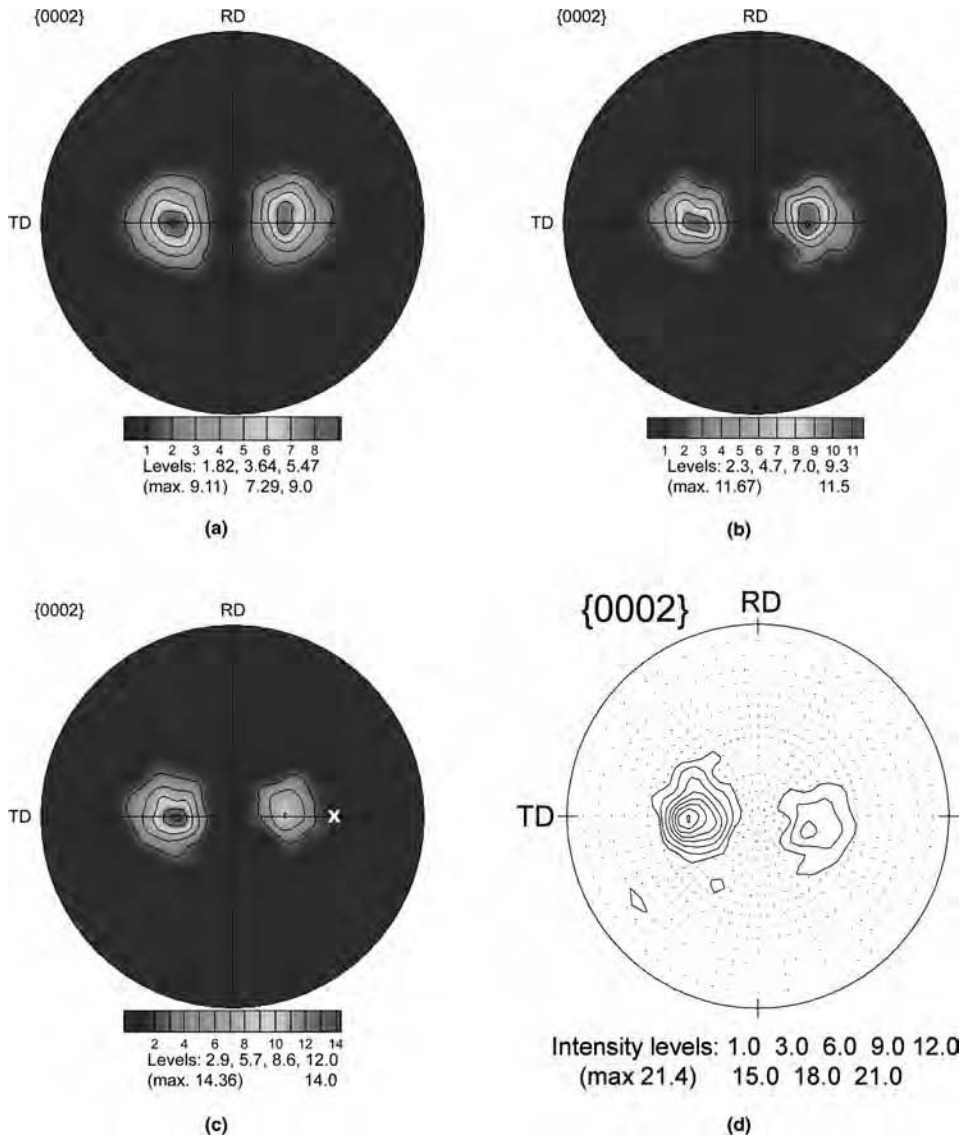
Fig. 26 Titanium annealed in a magnetic field. (a) {0002} pole figure for the titanium sheet sample corresponding to the initial texture. (b) and (c) Simulated pole figures after 10 min annealing at 750 °C for a two-dimensional titanium polycrystal. (b) without field. (c) In a magnetic field (field direction is indicated by a white cross). (d) Experimentally determined texture after magnetic annealing. RD, rolling direction; TD, transverse direction. Source: Ref 70

Here, $\Delta\chi$ is the difference of the susceptibilities parallel, $\chi_\parallel$, and perpendicular, $\chi_\perp$, to the basal planes; and θ is the angle between the c-axis and the magnetic field direction. To consider the action of a magnetic field in a network model, it is only necessary to add Eq 17 as a driving force in Eq 4.

As an example, grain growth in the presence of an extrinsic driving force was investigated experimentally on rolled titanium samples (Ref 69–71) and by network model simulations. In these experiments, the titanium samples were annealed in a magnetic field oriented in such a way that the growth of grains with certain orientations was preferred over others. The simulations started from a discretized orientation mapping of the titanium sample, and the

physical parameters were set to match the experimental conditions. The texture and grain-growth kinetics obtained from the simulations were compared with the experimental results (Fig. 26).

The initial texture (Fig. 26a) was characterized by two symmetrical peaks about the rolling direction. In terms of Euler angles, these correspond approximately to the orientations $\{\varphi_1 = 180°, \Phi = 35°, \varphi_2\}$ and $\{\varphi_1 = 0°, \Phi = 35°, \varphi_2\}$. The simulated texture evolution showed excellent agreement with the experimental results (Fig. 26c). The intensity of the component $\{\varphi_1 = 0°, \Phi = 35°, \varphi_2\}$ decreased with increasing time (Fig. 26c) because the magnetic field was directed almost parallel to this component (Fig. 26c). In the absence of a magnetic

field, the texture showed no relevant changes in its evolution with time (Fig. 26b), as expected.

The change of the texture was accompanied by specific kinetics of the various texture components. For instance, the favored component had much faster kinetics than that of the disfavored component and the overall kinetics (Fig. 27). The applied field forced the disfavored grains to grow more slowly and thus caused a size advantage of the favored grains over the disfavored ones. Also, the simulated kinetics showed good agreement with the experimental results (Ref 70).

In the example presented here, grain growth was affected by the introduction of an additional magnetic driving force. Such force can be used to drive the migration of grain boundaries, however, only in magnetically anisotropic materials. A more versatile method is the use of mechanical stresses. During materials processing, grain growth usually occurs concurrently with the application of stresses. This can and has been simulated by means of network models (Ref 72) also. For this purpose, finite-element modeling simulation was used to calculate the introduced elastic energy density of the grains and was coupled with vertex simulations to simulate boundary migration and therefore the progress of grain growth (Ref 72).

Grain Growth in Systems with Finite Mobility of the Boundary Junctions. A polycrystalline aggregate is a system composed of grains, grain boundaries, triple lines, and quadruple junctions. For a long time, it was assumed that the only element affecting grain-boundary migration was the grain boundary itself. Triple lines and quadruple junctions were not taken into account and were considered unimportant. However, in recent years, several theoretical and experimental studies (Ref 73, 74) have demonstrated that triple lines can have kinetics different from the adjoining grain boundaries; that is, triple lines can possess a finite mobility and therefore can drag grain-boundary motion. The study of the kinetic properties of a system of connected boundaries is only possible in the course of steady-state motion. Steady-state motion can be achieved in very few configurations, though (Fig. 28). This is because the driving force for the motion of the system is still the curvature of the boundaries, so that a steady-state motion requires a specific geometry of the configuration. Two steady-state geometries have been studied (Ref 75) for the 2-D case (Fig. 28a, b). Steady-state motion for the 3-D case (Fig. 28c) has recently been addressed by network model simulations (Ref 24).

The controlling quantity of connected grain-boundary systems is the dimensionless parameter Λ that characterizes the influence of a finite triple-junction mobility on the evolution of the system. For systems with triple junctions, $\Lambda = \Lambda_{tl}$ is given for the geometry in Fig. 28(a) by:
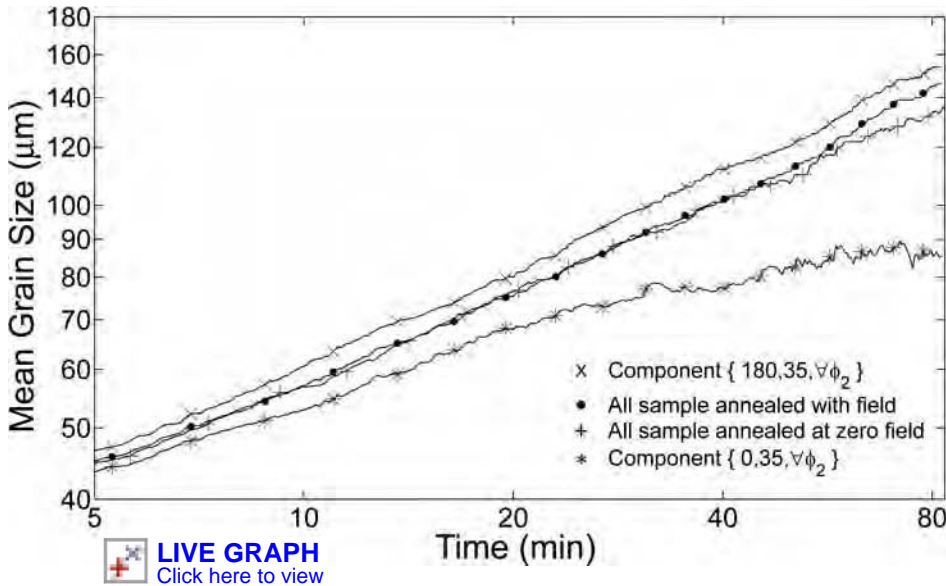
Fig. 27 The overall kinetics for conventional and magnetic annealing do not show any significant differences. This is caused by a superposition of the kinetics of the favored component with the kinetics of the disfavored component.
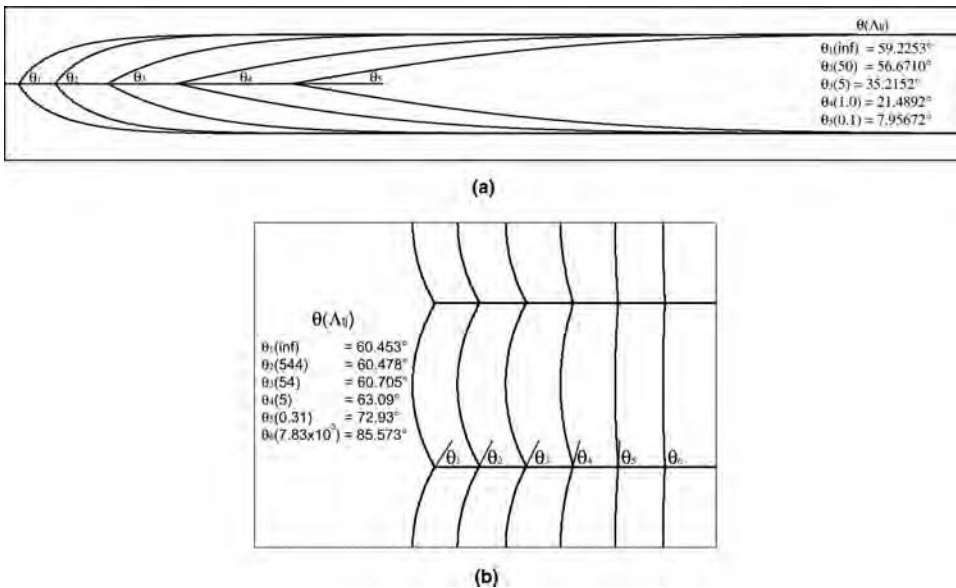
Fig. 29 Simulated grain-boundary shapes for two-dimensional systems for different values of Λ_{tl}

Fig. 28 Steady-state motion. (a) and (b) Two geometries of grain-boundary systems with triple junctions. (c) Quadruple junctions where steady-state motion can be observed

$$\Lambda_{tl} = \frac{m_{tl} \cdot r}{m_{gb}} = \frac{2\theta}{1 - 2\cos\theta} \qquad \text{(Eq 18a)}$$

and for the configuration in Fig. 28(b) by:

$$\Lambda_{tl} = \frac{m_{tl} \cdot r}{m_{gb}} = -\frac{\ln\sin\theta}{1 - 2\cos\theta} \qquad \text{(Eq 18b)}$$

The parameter Λ_{qp}, which defines the influence of a quadruple-point mobility, m_{qp}, on the migration of the grain boundaries, reads:

$$\Lambda_{qp} = \frac{m_{qp} \cdot r^2}{m_{gb}} = -\frac{\ln\sin\theta}{1 + 3\cos\theta} \qquad \text{(Eq 18c)}$$

where m_{tl} is the triple-line mobility, m_{gb} is the grain-boundary mobility, m_{qp} is the quadruple-point mobility, r represents the grain size, and θ is the contact angle at the junction (Fig. 28). A consequence of a finite triple-line and quadruple-point mobility is the change of the boundary shape, as illustrated in Fig. 29 for 2-D systems. Experimental investigations in zinc tricrystals and aluminum polycrystals have confirmed the predicted change of the grain-boundary shape (Ref 74). In the case of a finite quadruple-junction mobility, a similar shape change of the grain boundaries has been demonstrated (Ref 76).

The network models are able to reproduce these features very well (Ref 23, 50, 76, 77) and thus can be used for the simulation of grain growth in polycrystals under the influence of a finite junction mobility (Ref 76). The simulation results fully agree with the theoretical predictions for the motion of such systems (Fig. 30). Because the mobilities of grain boundaries, triple lines, and quadruple points are explicit variables in network models, such simulations easily lend themselves to a study of junction effects.

Summary

Network models are very efficient tools for the simulation of grain growth and related phenomena because they are able to essentially reproduce the physics governing grain-boundary migration. These models, unlike thermodynamic models like monte carlo or phase-field models, are fully scalable and hence offer the possibility of a direct comparison with experiments and analytical approaches. The main feature of network models is that they require as input the topological network that builds up a

(a)

LIVE GRAPH
Click here to view

(b)
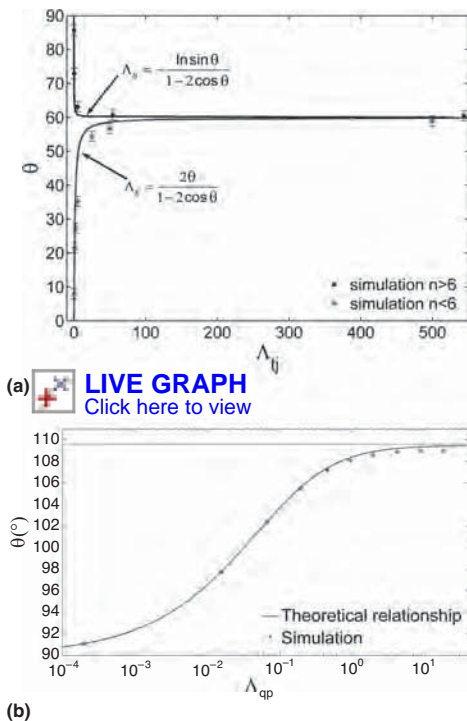
Fig. 30 Comparison of simulation results with theoretical relationships for finite (a) triple-line mobility and (b) quadruple-junction mobility

LIVE GRAPH
Click here to view

microstructure. Such input can be numerically generated by means of special algorithms for the construction of tessellations, for example, Voronoi and radical plane mosaics, or by digitization of experimental data. The use of a topological network allows deriving the equations of motion from the local geometry of the grain boundaries. However, it also introduces the necessity to implement topological transformations to correctly handle the changes of topology that take place during grain growth.

Network models are very flexible and can be used to simulate various phenomena where boundary migration plays a relevant role. While in the current contribution this was illustrated with simulations of 2-D and 3-D normal grain growth, grain growth in the presence of an extrinsic driving force, and the effect of a finite mobility of boundary junctions on grain growth, it is stressed that many more cases can be studied with these models, for example, grain growth in the presence of second-phase particles (Ref 78), columnar grain growth (Ref 79), and subgrain growth during recrystallization (Ref 80).

REFERENCES

1. M.P. Anderson, G.S. Grest, P.S. Sahni and D.J. Srolovitz, *Acta Metall.*, Vol 32, 1984, p 783
2. Z. Yang, S. Sista, J.W. Elmer, and T. DebRoy, *Acta Metall.,* Vol 48, 2000, p 4813
3. M.P. Anderson, G.S. Grest, R.D. Doherty, L. Kang, and D.J. Srolovitz, *Scr. Metall.*, Vol 23, 1989, p 753
4. P. Streitenberger and D. Zöllner, *Scr. Mater.*, Vol 55, 2006, p 461
5. L.Q. Chen, *Scr. Metall. Mater.*, Vol 32, 1995, p 115
6. S.G. Kim, D.I. Kim, W.T. Kim, and Y.B. Park, *Phys. Rev. E*, Vol 74, 2006, p 061605
7. R.L. Fullman, *Metal Interfaces*, American Society for Metals, Metals Park, OH 1952, p 179
8. D. Weaire and J.P. Kermode, *Phil. Mag. B*, Vol 48, 1983, p 245
9. D. Weaire and J.P. Kermode, *Philos. Mag. B*, Vol 50, 1984, p 379
10. A. Soares, A.C. Ferro, and M.A. Fortes, *Scr. Metall.*, Vol 19, 1985, p 1491
11. K. Kawasaki, T. Nagai, and K. Nakashima, *Philos. Mag. B*, Vol 60, 1989, p 399
12. K. Nakashima, T. Nagai, and K. Kawasaki, *J. Stat. Phys.*, Vol 57, 1989, p 759
13. J.P. Kermode and D. Weaire, *Comp. Phys. Comm.*, Vol 60, 1990, p 75
14. Y. Enomoto and R. Kato, *Acta Metall. Mater.*, Vol 38, 1990, p 765
15. K.A. Brakke, *Exp. Math.*, Vol 1, 1992, p 141
16. K. Fuchizaki, T. Kusaba, and K. Kawasaki, *Philos. Mag. B*, Vol 71, 1995, p 333
17. C. Maurice and F.J. Humphreys, Thermomechanical Processing in Theory, *Model. Prac.*, 1997, p 201
18. D. Weygand, Y. Bréchet, and J. Lepinoux, *Philos. Mag. B*, Vol 78, 1998, p 329
19. C. Maurice and F.J. Humphreys, Grain Growth in Polycrystalline Materials, *MMS*, 1998, p 81
20. D. Weygand, Y. Bréchet, J. Lepinoux, and W. Gust, *Philos. Mag. B*, Vol 79, 1999, p 703
21. D. Kinderlehrer and C. Liu, *Math. Mod. Meth. App. Sci.*, Vol 11, 2001, p 713
22. L.A. Barrales-Mora, V. Mohles, P.J. Konijnenberg, and D.A. Molodov, *Comput. Mater. Sci.*, Vol 39, 2007, p 160
23. L.A. Barrales-Mora, "2D and 3D Grain Growth Modeling and Simulation," Doctoral thesis, RWTH-Aachen, Cuvillier-Verlag, *Göttingen*, 2008
24. L.A. Barrales-Mora, G. Gottstein, and L.S. Shvindlerman, *Acta Mater.*, Vol 56, 2008, p 5915
25. K.G.F. Janssens, *Model. Simul. Mater. Sci. Eng.*, Vol 11, 2003, p 157
26. Y. Liu, T. Baudin, and R. Penelle, *Scr. Mater.*, Vol 34, 1996, p 1679
27. A. Kuprat, D. George, G. Straub, and M.C. Demirel, *Comput. Mater. Sci.*, Vol 28, 2003, p 199
28. A.J. Haslam, S.R. Phillpot, D. Wolf, D. Moldovan, and H. Gleiter, *Mater. Sci. Eng. A*, Vol 318, 2001, p 293
29. K.A. Brakke, *Exp. Math.*, Vol 1, 1992, p 141
30. H.J. Frost, C.V. Thompson, C.L. Howe, and J. Whang, *Scr. Mater.*, Vol 22, 1998, p 65
31. J. Svoboda, *Scr. Metall. Mater.*, Vol 28, 1993, p 1589
32. G. Spanos, D.J. Rowenhorst, A.C. Lewis, and A.B. Geltmacher, *MRS Bull.*, Vol 33, 2008, p 597
33. D. Kammer and P.W. Voorhees, *MRS Bull.*, Vol 33, 2008, p 603
34. J.Y. Buffière, P. Cloetens, W. Ludwig, E. Maire, and L. Salvo, *MRS Bull.*, Vol 33, 2008, p 611
35. D. Juul Jensen, S.E. Offerman, and J. Sietsma, *MRS Bull.*, Vol 33, 2008, p 621
36. M.A. Mangan, P.D. Lauren, and G.J. Shiflet, *J. Micros.*, Vol 188, 1997, p 36
37. T.L. Wolfsdorf, W.H. Bender, and P.W. Voorhees, *Acta Mater.*, Vol 45, 1997, p 2279
38. M.V. Kral and G. Spanos, *Acta Mater.*, Vol 47, 1999, p 711
39. A.C. Lewis and A.B. Geltmacher, *Scr. Mater.*, Vol 55, 2006, p 81
40. M.D. Uchic, M.A. Groeber, D.M. Dimiduk, and J.P. Simmonds, *Scr. Mater.*, Vol 55, 2006, p 23
41. D.J. Rowenhorst, A. Gupta, C.R. Feng, and G. Spanos, *Scr. Mater.*, Vol 55, 2006, p 11
42. A. Brahme, M.H. Alvi, D. Saylor, J. Fridy, and A.D. Rollett, *Scr. Mater.*, Vol 55, 2006, p 75
43. M. Zeng and G. Gottstein, *Aluminium*, Vol 78, 2002, p 878
44. E.G. Klarreich, *Am. Sci.*, Vol 88, 2000, p 152
45. J. O'Rourke, Chapter 5, in *Computational Geometry in C*, 2nd ed., Cambridge University Press, 1998
46. M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, Chapter 7, in *Computational Geometry Algorithms and Applications*, 2nd ed., Springer, 1998
47. G. Voronoi, *J. fuer die Reine Angewandte Mathematik*, Vol 133, 1907, p 97
48. B.J. Gellatly and J.L. Finney, *J. Non-Cryst. Solids*, Vol 50, 1982, p 313
49. E.W. Weisstein, Available from http://mathworld.wolfram.com/RadicalLine.html
50. Y. Ma, "Experimental Investigation and Computer Simulation of Grain Growth and Microstructure Evolution in 2D Polycrystals," Ph.D. thesis, Dortmund, Wulff-Verlag, 2008
51. V. Mohles, *Philos. Mag. A*, Vol 81, 2001, p 971
52. V. Mohles, Fundamental Dislocation Theory and 3D Dislocation Mechanics, *Multiscale Modeling of Plasticity and Fracture by Means of Dislocation Mechanics*, R. Pippan, Ed., Springer Verlag, 2009, p 1, accepted for publication
53. V. Mohles, "Thermisch Aktivierte Versetzungsbewegung in Kristallen auf der Grundlage von Simulationsrechnungen," Doctoral thesis, TU Braunschweig, Shaker, Aachen, 1997
54. C. Herring, Surface Tension as a Motivation for Sintering, *The Physics of Powder Metallurgy*, R. Gomer et al., Ed., McGraw-Hill, New York, NY, 1949
55. F.N. Rhine, *Microstructology*, Dr. Riederer-Verlag, Stuttgart, 1986, p 28–29

56. A.C. Ferro and M.A. Fortes, *Interface Sci.*, Vol 5, 1997, p 263
57. F. Wakai, N. Enomoto, and H. Ogawa, *Acta Mater.*, Vol 48, 2000, p 1297
58. N. Rivier and A. Lissowski, *J. Phys. A*, Vol 15, 1982, p L143
59. C. Maurice, Numerical Modelling of Grain Growth: Current Status, *Proceedings of the First Joint International Conference on Recrystallization and Grain Growth*, Vol 1, G. Gottstein and D.A. Molodov, Ed., Springer, Aachen, Germany, 2001, p 123
60. P.A. Beck, J.C. Kremer, L.J. Dremer, and M.L. Holzworth, *Trans. AIME*, Vol 175, 1948, p 372
61. L.A. Barrales-Mora, V. Mohles, P.J. Konijnenberg, and D.A. Molodov, *Comp. Mater. Sci.*, Vol 39, 2007, p 160
62. W.W. Mullins, *J. Appl. Phys.*, Vol 27, 1956, p 900
63. W.W. Mullins, *Acta Metall.*, Vol 37, 1989, p 2979
64. S. Hilgenfeldt, A.M. Krynik, S.S. Koehler, and H.A. Stone, *Phys. Rev. Lett.*, Vol 86, 2001, p 2685
65. M.E. Glicksman, *Philos. Mag.*, Vol 85, 2005, p 3
66. J.W. Cahn, *Trans. Metall. Soc. AIME*, Vol 239, 1967, p 610
67. R.D. MacPherson and D.J. Srolovitz, *Nature*, Vol 446, 2007, p 1053
68. G. Gottstein and L.S. Shvindlerman, *Grain Interface Migration in Metals: Thermo-Dynamics, Kinetics, Applications*, CRC Press, Boca Raton, FL, 1999
69. D.A. Molodov, P.J. Konijnenberg, N. Bozzolo, and A.D. Sheikh-Ali, *Mater. Lett.*, Vol 59, 2005, p 3209
70. D.A. Molodov, C. Bollmann, P.J. Konijnenberg, L.A. Barrales Mora, and V. Mohles, *Mater. Trans.*, Vol 48, 2007, p 2800
71. D.A. Molodov, P.J. Konijnenberg, L.A. Barrales Mora, and V. Mohles, *J. Mater. Sci.*, Vol 41, 2006, p 7853
72. C.C. Battaile, W.A. Counts, G.W. Wellman, T.E. Buchheit, and E.A. Holm, *Metall. Mater. Trans. A*, Vol 38, 2007, p 2513
73. G. Gottstein and L.S. Shvindlerman, *Scr. Mater.*, Vol 54, 2006, p 1065
74. U. Czubayko, V.G. Sursaeva, G. Gottstein, and L.S. Shvindlerman, *Acta Mater.*, Vol 46, 1998, p 5863
75. G. Gottstein, Y. Ma, and L.S. Shvindlerman, *Acta Mater.*, Vol 53, 2005, p 1535
76. L.A. Barrales Mora, V. Mohles, L.S. Shvindlerman, and G. Gosttstein, *Acta Mater.*, Vol 56, 2008, p 1151
77. D. Weygand, Y. Bréchet, and J. Lépinoux, *Interface Sci.*, Vol 7, 1999, p 285
78. D. Weygand, Y. Bréchet, and J. Lépinoux, *Acta Mater.*, Vol 47, 1999, p 961
79. A.Y. Badmos, H.J. Frost, and I. Baker, *Acta Mater.*, Vol 50, 2002, p 3347
80. F.J. Humphreys and M. Hatherly, *Recrystallization and Related Annealing Phenomena*, Pergamon Press, Oxford, 1995

Phase-Field Microstructure Modeling

Chen Shen, GE Global Research
Yunzhi Wang, The Ohio State University

Fundamentals

The collection of a hierarchy of structural and chemical nonuniformities (imperfections or defects) in solids constitutes materials microstructure that, in turn, determines materials properties. Well-known examples of the structural defects include dislocations and homo- and heterophase interfaces, while typical examples of the chemical defects include concentration variation across heterophase interfaces (such as precipitates) and impurity segregation at various structural defects. Stacking faults, twin boundaries, antiphase domain boundaries, grain boundaries, and ferromagnetic and ferroelectric domain walls belong to the category of homophase interface. The length scales at which these variations occur define the length scales of a microstructure.

Field Description of Microstructure. A field description of the structural and chemical nonuniformities seems to be an intuitive and natural way to define a microstructure. In the phase-field approach, the two types of nonuniformities are characterized by two types of continuous fields: conserved and nonconserved order parameters. Typical examples of the conserved order parameters include concentration of a chemical species, density, and molar volume in a multicomponent and multiphase system. Examples of the nonconserved order parameters include long-range order parameters for atomic ordering, inelastic displacement or inelastic strain (eigenstrain or transformation strain) for dislocations, martensitic particles and microcracks, and magnetization and polarization for ferromagnetic and ferroelectric transitions. These order parameters are continuous fields (functions of position, $\mathbf{r}$, e.g., $\phi(\mathbf{r}, t)$) and are referred to as phase fields in the phase-field method. A graphic plot of the order parameter fields produces pictures that are similar to the ones typically observed under a microscope. Because a microstructure is an ever-evolving feature toward thermodynamic equilibrium, the descriptors ϕ generally include time, t, as the variable.

Total Energy Functional and Variation. A field description of microstructure involves a large number of variables (one or several for each spatial coordinate, $\mathbf{r}$, with each one regarded as a degree of freedom). An effective way to formulate their time evolution (the kinetics of microstructure change) is to define a total energy of the system and derive the kinetics along the steepest descent path of the total energy with respect to these degrees of freedom. Formally, the total energy is written as:

$$E = E[\phi(\mathbf{r}, t)] \qquad \text{(Eq 1)}$$

Since the independent variable ϕ is itself a function (of the spatial coordinate), E is regarded as a functional. In the framework of thermodynamics, E is directly related to one of the free energies subject to given external constraints. Normally, a material is under a constant pressure and temperature; thus, E is the Gibbs free energy. The steepest descending direction of the total energy is given by derivatives of E with respect to each field variable and constitutes the thermodynamic driving force for the change of that variable at each location. When an unconstrained system is in a thermodynamic equilibrium, all the derivatives identically become zero. (The equilibrium state of a constrained system is slightly modified with the use of Lagrange multipliers. For example, by mass conservation, a solid solution reaches equilibrium when the derivatives become uniform in the space. The uniform constant is the Lagrange multiplier. However, the derivatives of the corresponding grand function, for example, Eq 31 in a later section, are identically zero.) Because E is a functional, the derivatives are regarded as functional derivatives in the calculus of variations (any calculus of variations textbook may be referred to for mathematical background). While E is a scalar quantity, its variational derivative:

$$\delta E / \delta \phi(\mathbf{r}, t) \qquad \text{(Eq 2)}$$

is a d-dimensional vector, where d is the total degrees of freedom of the system.

Kinetics of the Conserved Field and the Generalized Diffusion Equation. When a microstructure is represented by a spatial distribution of the solute concentration, the field variable can be defined as the mole fraction of the solute:

$$\phi = X(\mathbf{r}, t) \qquad \text{(Eq 3)}$$

Naturally, for a multicomponent alloy, the field can be defined as a set of $n - 1$ composition fields, $\phi = \{X_i(\mathbf{r}, t), i = 1, \ldots, n - 1\}$, where n is the total number of chemical species, among which only $n - 1$ are independent because of the constraint:

$$\sum_{i=1}^{n} X_i \equiv 1$$

According to gradient thermodynamics (Ref 1–3), when nonuniformities in composition and structure exist in a heterogeneous system, the free energy of the system depends not only on local composition and structure but also on their spatial variations. Thus, the total chemical free energy of a system of nonuniform concentration may be formally written as (Ref 3):

$$E^{\text{chem}} = E^{\text{chem}}[X(\mathbf{r}, t)] = \int f(X, \nabla X, \nabla^2 X, \ldots) d\mathbf{r} \qquad \text{(Eq 4)}$$

With symmetry considerations and taking only the leading nonvanishing term in the concentration gradient, Eq 4 reduces to:

$$E^{\text{chem}} = \int V_m^{-1} [f_m(X) + \kappa_m \nabla X \cdot \nabla X] d\mathbf{r} \qquad \text{(Eq 5)}$$

where f_m is the molar free energy of a corresponding homogeneous material of composition X. The subscript m stands for molar (per mole) quantities. The second term in the bracket takes into account the actual spatial variation (gradient) of X, where the gradient coefficient κ_m (in a unit of $\text{J} \cdot \text{m}^2 \cdot \text{mol}^{-1}$) is related to the second derivatives of f in Eq 4 (see Eq 2.7 in Ref 3). V_m is the molar volume. For brevity, the independent variables $\mathbf{r}$ and t in X in Eq 5 are dropped from here on.

If an existing spatial distribution of solute is not in thermodynamic equilibrium (an example of nonuniform solute in thermodynamic equilibrium is segregation at a grain boundary), its temporal evolution follows the Cahn-Hilliard generalized diffusion equation:

$$\frac{1}{V_m}\frac{\partial X}{\partial t} = \nabla \cdot M \nabla V_m \frac{\delta E^{\text{chem}}}{\delta X} \qquad \text{(Eq 6)}$$

where M is the mobility (in a unit of $J^{-1} \cdot \text{mol}^2 \cdot m^{-1} \cdot s^{-1}$). It is easy to verify that from Eq 5 (assuming V_m is independent of spatial location):

$$V_m \frac{\delta E^{\text{chem}}}{\delta X} = \frac{\partial f_m}{\partial X} - 2\kappa_m \nabla^2 X \qquad \text{(Eq 7)}$$

Recall that $\Delta\mu = \partial f_m/\partial X$ is commonly regarded as the diffusion potential or exchange potential (a difference between the chemical potential of solute and solvent atoms), thus:

$$\Delta\tilde{\mu} \equiv \Delta\mu - 2\kappa_m \nabla^2 X \qquad \text{(Eq 8)}$$

is written as a generalized diffusion potential, with the second term on the right as a gradient correction. With that, Eq 6 can be rewritten as:

$$\frac{1}{V_m}\frac{\partial X}{\partial t} = -\nabla \cdot \mathbf{J} \qquad \text{(Eq 9)}$$

and

$$\mathbf{J} = -M\nabla(\Delta\tilde{\mu}) \qquad \text{(Eq 10)}$$

Equations 9 and 10 give the conventional form of the diffusion equation, with $\mathbf{J}$ being the diffusion flux (in a unit of $\text{mol} \cdot m^{-2} \cdot s^{-1}$) defined in a laboratory frame of reference.

Kinetics of Nonconserved Fields and Structural Relaxation. When a microstructure is characterized by structural nonuniformities in a material, a set of nonconserved order parameters are used in phase-field models. The number of order parameters is determined by the degrees of freedom of the structure. For example, three order parameters are needed for describing the Ni_3Al intermetallic (γ') phase to produce all four antiphase domains in an $L1_2$ ordered crystal structure (Ref 4–6).

The functional formulation for a structurally nonuniform system in the phase-field method is similar to that for a chemically nonuniform system presented previously (Eq 3–5), where Eq 3 is replaced by:

$$\phi = \eta(\mathbf{r}, t) \qquad \text{(Eq 11)}$$

to distinguish the structural (nonconserved) fields from the compositional (conserved) fields. A subscript could also be assigned to η if there are multiple fields, or it could be written in a tensor form for generality. The structural order parameters usually do not satisfy conservation law the way mass and concentration do, and they follow a different type of kinetic equation, known as the time-dependent Ginzburg-Landau equation or the Allen-Cahn equation:

$$\frac{\partial\eta}{\partial t} = -L\frac{\delta E^{\text{chem}}}{\delta\eta} \qquad \text{(Eq 12)}$$

Here, the unit of the kinetic coefficient L is $J^{-1} \cdot m^3 \cdot s^{-1}$.

The Role of the Gradient Term—Interface Property and Curvature. When the diffusion potential in the phase-field model (Eq 8) is compared to the conventional Fickian diffusion equation, there is a correction term for the presence of a gradient in the composition field. The gradient term is a major feature in phase-field models where the field representation of microstructure results in a unified description of multiple phases and a smooth transition from one phase to another. (This is compared to conventional moving-boundary diffusion models in which the diffusion solution is solved separately for individual phase domains and bounded by explicit conditions along the interphase boundaries.)

Physically, the spatial (nonlocal) interaction represented by the gradient term in gradient thermodynamics of nonuniform systems arises in a continuum limit transition from its discrete counterparts (Ref 7, 8) or from microscopic theories in statistical mechanics (for example, Ref 9, 10). It accounts for the change in atomic bonding from one location to its neighboring locations. In phase-field models where the phase fields are physical order parameters, minimization of the total free energy determines the balances between the local free energy term that prefers an infinitely sharp interface and the gradient energy term that prefers an infinitely diffusion interface. Such a balance regulates the interface profile and gives interface energy (σ) and width (w):

$$\sigma = 2V_m^{-1}\int_{\phi_1}^{\phi_2}\sqrt{\kappa\Delta f(\phi)}\,d\phi \qquad \text{(Eq 13)}$$

$$w = 2(\phi_2 - \phi_1)\sqrt{\kappa/\Delta f_{\max}} \qquad \text{(Eq 14)}$$

where the quantities ϕ_1, ϕ_2, $\Delta f(\phi)$, and $\Delta f_{\max}$ are defined in Fig. 1.

In some phenomenological phase-field models, typically those evolved from solidification models, the gradient term does not appear on every defined phase-field variable. For example, the gradient terms on concentration fields are neglected without further necessity (Ref 11) or

even with a particular purpose to separate the chemical contribution to the interface property (Ref 12).

With either physically and phenomenologically defined phase fields, the gradient term provides another important feature in two- or three-dimensional space: the interface curvature. This can be seen more clearly in the expansion:

$$\nabla^2\phi = \frac{\partial^2\phi}{\partial r^2} + \frac{d-1}{r}\frac{\partial\phi}{\partial r} \qquad \text{(Eq 15)}$$

It can be seen (Ref 13) that the second term on the right contributes to local interface velocity, a term in proportion to $(d-1)/R$, where R is the local radius of curvature of the interface, and d is the dimension of the space. Note that $\nabla^2\phi$ is proportional to the variational derivative of the gradient energy term (e.g., Eq 5, 7). Equation 15 implies that the gradient term contributes to the driving force from both the variation of phase field across the interface, $\partial^2\phi/\partial r^2$, and the local radius of curvature, R. The interface curvature is the important driving force for grain growth and coarsening (Gibbs-Thomson effect).

Microscopic versus Coarse-Grained Phase-Field Models. Equations 13 and 14 show that interfaces in phase-field models have finite equilibrium widths, unique chemical and structural variations within them, and the associated interfacial energies. When applied at the natural (typically microscopic) length scales of a given defect (such as an individual dislocation, interface, or nucleating precipitate), the phase-field model has a unique advantage over the sharp-interface models in predicting the fundamental properties of the defect, such as its equilibrium size and energy (Ref 3), and the critical configuration and activation energy of a nucleus (Ref 14) rather than using them as inputs. These phase-field models may be regarded as the microscopic phase-field models.

To take advantage of the ability of a phase-field model to predict fundamental properties of defects, it is essential to choose well-defined physical quantities as the order parameters in a phase-field model (Ref 3, 14–16). This seems to be straightforward in certain cases (such as the
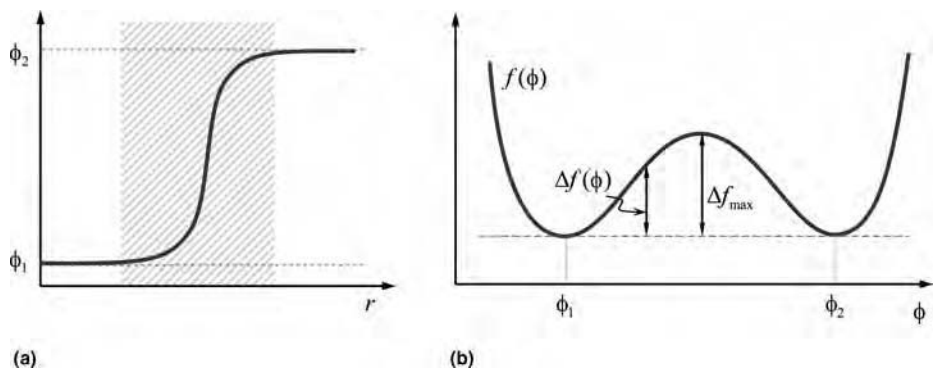
Fig. 1 Graphs describing energy function. (a) Schematic of an equilibrium phase field (ϕ) across an interface (shadowed range) along the spatial coordinate r. (b) Free energy density. The excess energy $\Delta f(\phi)$ between ϕ_1 and ϕ_2, which are the equilibrium phase-field values at the two sides of the interface, contributes to the interface energy.

examples mentioned previously) but could be nontrivial and difficult to do or even be more convenient not to do so in some other cases, such as solidification and grain growth (Ref 17, 18). In solidification, for example, an infinite number of order parameters may be required for a rigorous characterization of the structure of the liquid phase (Ref 17, 19). In this case, a simple phenomenological order parameter was introduced to distinguish liquid from solid (Ref 17). Phase-field models formulated with phenomenological order parameters are generally regarded as numerical techniques to avoid boundary tracking in describing evolving microstructures of complex geometries, although the physical origin of the order parameters may still be traced to the transition from microscopic picture to continuum limit (Ref 19). It remains a challenge to define physically rigorous but mathematically tractable order parameters in solidification and grain growth so that a physical rather than phenomenological free energy model can be formulated and the fundamental properties of the solid-liquid interfaces and grain boundaries can be predicted by using the phase-field method. Some of the order parameters introduced in atomistic simulations to identify these interfaces (Ref 20, 21) could be further explored for such purposes.

In phase-field models where phase fields are phenomenological order parameters or in coarse-grained phase-field models where the length scales of interest far exceed boundary or dislocation core width, the gradient terms still produce similar diffuse profiles of the order parameters across an interface or a dislocation core. However, these profiles no longer represent real structures of the defect. The models retain their advantages over the sharp-interface models in treating complicated geometrical and topological changes of large-defect ensembles during microstructural evolution (such as dendritic solidification, grain-growth and domain coarsening, dislocation network formation and coarsening, and various phase transformations) (Ref 17, 18, 22–24). Their applications have offered many invaluable insights into the sequence of microstructural evolutions and mechanisms of pattern formation in many material systems during various material processes. However, the models lose the intrinsic ability of a diffuse-interface model to predict the fundamental properties of these defects. Moreover, the phenomenological and coarse-grained phase-field models provide quantitative rate information only by matching phase-field model parameters to the standard sharp-interface models at the asymptotic limit of zero interface width (Ref 11) or, more recently, in the so-called thin-interface limit (Ref 25). Quantitative phase-field simulations at coarse-grained mesoscale levels have since been carried out extensively with the new thin-interface analyses (Ref 25–28). Some simple techniques (Ref 29, 30) based on physical arguments of equivalent driving forces were also developed to relax the restriction on the interface thickness. In addition, adaptive algorithms (Ref

31, 32) and new data structures (Ref 33, 34) have also been developed to increase the computational efficiency of quantitative phase-field simulations.

Long-Range Elastic Interactions. Heterogeneous phase distribution in solids can cause a long-range elastic strain field that varies at the scale of the microstructure (Ref 4, 35, 36). Even in a single-phase domain, spatial variation of the composition can also induce strains if the solute differs in atomic size from the solvent. The elastic deformation that accommodates the misfit in crystal lattices (due to lattice parameter, crystal structure, or orientation difference) of adjoining phases/domains to form coherent or semicoherent interphase boundaries, known as coherency strain, can play a significant role in solid-state phase transformations. Its effects range from precipitate morphology (e.g., equilibrium shape, habit planes) and spatial arrangement to the overall driving force for the transformation. In addition, because a nucleating phase may possibly adopt a metastable structure with low-energy coherent interfaces with the parent matrix, the final transition to the stable phase structure is controlled by the coherency strain energy and its interplay with the interface energy.

In phase-field models for solid-state phase transformations and dislocations, the coherency elasticity problems are treated in the framework of Eshelby (Ref 37, 38) using the general theory of phase-field microelasticity by Khachaturyan and Shatalov (Ref 4, 39–41), formulated upon spatial distribution of the stress-free transformation strain (SFTS) field:

$$\varepsilon_{ij}^T(\mathbf{r}) = \sum_p \varepsilon_{ij}^T(\phi_p(\mathbf{r})) \tag{Eq 16}$$

as a linear combination of individual phase fields, labeled as subscript p, that contributes to transformation strains. As a simplification, it may be approximated as:

$$\varepsilon_{ij}^T(\mathbf{r}) = \sum_p \varepsilon_{ij}^{T0}(p)\phi_p(\mathbf{r}) \tag{Eq 17}$$

where $\varepsilon_{ij}^{T0}(p)$ is the coefficient of the linear term in a Taylor expansion of $\varepsilon_{ij}^T(\phi_p(\mathbf{r}))$ with respect to ϕ_p. In the case of the concentration field ($\phi_p = X$), this is known as Vagard's law. The phase-field microelasticity theory of Khachaturyan and Shatalov treats the elasticity problem in a variational approach by formulating the total elastic energy as a functional of the SFTS fields:

$$E^{\mathrm{el}} = \frac{1}{2}\int d\mathbf{r} C_{ijkl}\varepsilon_{ij}^T(\mathbf{r})\varepsilon_{kl}^T(\mathbf{r}) + \frac{V}{2}C_{ijkl}\bar{\varepsilon}_{ij}\bar{\varepsilon}_{kl}$$
$$- \bar{\varepsilon}_{ij}\int d\mathbf{r} C_{ijkl}\varepsilon_{kl}^T(\mathbf{r})$$
$$- \frac{1}{2}\int \frac{d\mathbf{g}}{(2\pi)^3} n_i \tilde{\sigma}_{ij}^T(\mathbf{g})\Omega_{jk}(\mathbf{n})\tilde{\sigma}_{kl}^{T*}(\mathbf{g})n_l \tag{Eq 18}$$

with $\sigma_{ij}^T = C_{ijkl}\varepsilon_{kl}^T$, $[\Omega^{-1}]_{jk} = n_i C_{ijkl}n_l$, and $\mathbf{n} = \mathbf{g}/|\mathbf{g}|$. Here $\mathbf{g}$ is a reciprocal space vector.

$\tilde{\sigma}_{ij}^T(\mathbf{g}) = \int d\mathbf{r}\sigma_{ij}^T(\mathbf{r})\exp(-i\mathbf{g}\cdot\mathbf{r})$ is the Fourier transform of $\sigma_{ij}^T(\mathbf{r})$. The separation of the homogeneous strain $\bar{\varepsilon}_{ij}$ (the mean value of $\varepsilon_{ij}^T(\mathbf{r})$) gives flexibility for treating various boundary conditions of the elasticity problem (Ref 4, 42, 43).

Through Eq 16, the elastic energy becomes a sole functional of the phase fields, ϕ_p, just as the chemical free energy given in Eq 4. It is thus possible to compute the total driving force as a combination of the chemical and elastic energies:

$$\frac{\delta(E^{\mathrm{chem}} + E^{\mathrm{el}})}{\delta\phi_p} \tag{Eq 19}$$

which are the variational derivatives of the total energy with respect to the same set of phase fields $\{\phi_p\}$. This provides phase-field models with an ability for self-consistent handling of multiphase, multivariant coherent microstructures with interplay among the chemical free energy, interfacial energy, and elastic energy (for example, see review in Ref 36).

The contributing phase fields to SFTS or inelastic strain in general can be any crystalline defects, including solute clusters, precipitates, dislocations, and so on. Their mutual elastic interactions are accounted for through the coupling in Eq 16. The Khachaturyan-Shatalov microelasticity theory was also extended to treat inhomogeneous (position-dependent) elastic modulus (Ref 43–46), which further extends the applications to include cracks and voids (Ref 47), free surfaces (Ref 48), elasticity-induced rafting (Ref 49, 50), and so on.

Modeling Nucleation

Because the majority of solid-state phase transformations are first-order phase transformations, modeling microstructure evolution typically proceeds along the route of nucleation, growth, and coarsening, either focusing on one stage or treating them as concurrent processes (Ref 51). The phase-field governing equations treat both conserved (Eq 6) and nonconserved (Eq 12) fields, corresponding to incongruent (involving long-range diffusion) and congruent (such as atomic ordering, grain growth, massive and martensitic transformation, and plastic deformation) processes, respectively. Meanwhile, the curvature contribution due to the gradient term (Eq 15) gives the driving force for precipitate and domain coarsening.

Nucleation needs separate handling, because both phase-field kinetic equations (Eq 6, 12) imply a nonpositive time-derivative of the total energy, that is, $d(E^{\mathrm{chem}} + E^{\mathrm{el}})/dt \leq 0$. This excludes any activation process, which requires a temporary increase in the total energy. In the phase-field method, nucleation has been implemented mainly in two ways: (1) numerically solving the stochastic phase-field equations with additional Langevin force terms (Ref 52–58),

and (2) an explicit nucleation algorithm that stochastically seeds nuclei in an evolving microstructure according to the nucleation rates calculated as a function of local concentration, stress, and temperature (Ref 59). The Langevin force approach is self-consistent with the phase-field method, and the effect of coherency strain energy, which plays an important role in the relaxation process of the fluctuations, is accounted for automatically. Nevertheless, the quantitative use of Langevin force terms in describing first-order phase transitions requires microscopic degrees of freedom, which confines both length and time scales of the simulation (Ref 60). Thus, the present use of the Langevin approach for nucleation in mesoscale microstructural simulations is qualitative in nature, and its applications are limited to site-saturation conditions (Ref 54–58). The explicit algorithm is computationally more efficient and has been applied successfully in concurrent nucleation and growth processes under either isothermal or continuous cooling conditions (Ref 59, 61, 62).

Langevin Force Approach. The approach is an analog to the treatments for Brownian motion in statistical mechanics, which mimics the thermal perturbation forces from the environment. The Langevin force terms are incorporated directly into the phase-field dynamics equations, that is:

$$\frac{1}{V_m}\frac{\partial X}{\partial t} = \nabla \cdot M \nabla \frac{\delta(E^{\mathrm{chem}} + E^{\mathrm{el}})}{\delta X} + \xi_X(\mathbf{r}, t)$$

$$(\text{Eq } 20)$$

$$\frac{\partial \eta}{\partial t} = -L\frac{\delta(E^{\mathrm{chem}} + E^{\mathrm{el}})}{\delta \eta} + \xi_\eta(\mathbf{r}, t) \qquad (\text{Eq } 21)$$

The Langevin force terms, in their simplest forms, are assumed uncorrelated in both space and time. This requires their first and second moments to satisfy:

$$\begin{aligned} &\langle \xi_X(\mathbf{r}, t)\rangle = 0, \langle \xi_\eta(\mathbf{r}, t)\rangle = 0 \\ &\langle \xi_X(\mathbf{r}, t)\xi_X(\mathbf{r}', t')\rangle = -2k_B T M\delta(t-t')\nabla^2\delta(\mathbf{r}-\mathbf{r}') \\ &\langle \xi_\eta(\mathbf{r}, t)\xi_\eta(\mathbf{r}', t')\rangle = 2k_B T L\delta(t-t')\delta(\mathbf{r}-\mathbf{r}') \end{aligned}$$

$$(\text{Eq } 22)$$

and yields a Gaussian distribution. In discrete forms, the Langevin force terms may be emulated with a random number generator. For example, if $\{\rho: \rho_i, i = 1, \dots\}$ represents a series of computer-generated Gaussian-distributed random numbers that satisfy $\langle \rho_i \rangle = 0$ and $\langle \rho_i \rho_i \rangle = \delta_{ii}$, then (Ref 63, 64):

$$\xi_\eta(t_n, \mathbf{r}_m) = \sqrt{2k_B T L/(\lambda^d \Delta t)}\rho \qquad (\text{Eq } 23)$$

$$\xi_c(t_n, \mathbf{r}_m) = \sqrt{2k_B T M/(\lambda^d \Delta t)}\nabla \cdot \boldsymbol{\rho} \qquad (\text{Eq } 24)$$

where Δt is the time step, λ is the grid size, d is the dimensionality of the space, and $\boldsymbol{\rho}$ is a d-dimensional vector, with each component being an independent Gaussian random number ρ_i.

At a metastable state, the balance between the perturbing Langevin force, ξ, and the restoring force, $\delta(E^{\mathrm{chem}} + E^{\mathrm{el}})/\delta\phi$, determines the fluctuation amplitude of the phase fields according to the fluctuation-dissipation theorem (Ref 65). Nevertheless, unlike in second-order phase transitions where the correlation length is typically much greater than the atomic spacing, nucleation involves thermal fluctuations at atomic scales, which limits the level of coarse-graining that a phase-field model typically performs.

Explicit Nucleation Algorithm. The explicit nucleation algorithm (Ref 59) is an alternate approach that circumvents some difficulties encountered in the Langevin force approach, in particular, at a length scale much greater than the correlation length (typically, the size of a critical nucleus). By evaluating the local nucleation rate as a thermally activated process:

$$J(\mathbf{r}, t) = J_0(\mathbf{r}, t)\exp(-\Delta F^*(\mathbf{r}, t)/k_B T) \qquad (\text{Eq } 25)$$

nucleation events are explicitly introduced through a probabilistic Poisson seeding process at a probability (Ref 59):

$$P(\mathbf{r}, t) \approx 1 - \exp(-J(\mathbf{r}, t)\Delta V \Delta t) \qquad (\text{Eq } 26)$$

where P is the probability of forming a critical nucleus in a volume, ΔV, and a time interval, Δt, which are predefined according to the length and time scales of the model. Once a location is determined to be transformed, the phase fields are updated correspondingly to reflect the formation of nuclei in the matrix.

The explicit nucleation algorithm does not specify how to compute ΔF^* in Eq 25, the activation energy of nucleation. This gives the flexibility to choose any nucleation theory for the model. In the classical nucleation theory for homogeneous nucleation, for example:

$$\Delta F^* = \alpha\sigma^d/(\Delta f_V)^{d-1} \qquad (\text{Eq } 27)$$

where σ is the interface energy, and Δf_V is the per-volume driving force for nucleation, which is usually formulated as a function of phase field, $\phi(\mathbf{r})$, and temperature. $\alpha = 16\pi/3$ for a spherical nucleus in three dimensions (i.e., $d = 3$), and $\alpha = \pi$ for a circular nucleus in two dimensions ($d = 2$). Note that both σ and Δf_V are position dependent and can incorporate location-dependent interaction energy terms that constitute defect-assisted (heterogeneous) nucleation.

Incorporation of Elastic Energy. When a nucleating phase differs sufficiently in crystal lattice from the matrix phase, coherency strain energy must be incorporated into the nucleation driving force, Δf_V, in Eq 27. The chance of forming a completely incoherent interface between a critical nucleus and matrix is rare because of the usually high nucleation barrier associated with an incoherent interface. Coherency strain energy modifies the chemical driving force for nucleation and also sometimes results in formation of metastable phases in the course of precipitation. Formally:

$$\Delta f_V = \Delta f_V^{\mathrm{chem}} + \Delta E^{\mathrm{el}}[\phi^*(\mathbf{r})]/V_0 \qquad (\text{Eq } 28)$$

can be written as a sum of the chemical contribution and the elastic contribution, with V_0 as the volume of the nucleus. However, the elastic energy of a nucleus in an elastically anisotropic solid, $\Delta E^{\mathrm{el}}[\phi^*(\mathbf{r})]$, given by Eq 18, depends on configuration of the critical nucleus, $\phi^*(\mathbf{r})$, which itself is determined by minimizing the total (chemical plus elastic) energy. In other words, it is impossible to write a per-volume elastic energy, in general, that is independent of the size of the nucleus, as is done for the chemical energy. Additionally, the elastic energy also includes long-range interaction between the nucleus and pre-existing microstructure and thus cannot be formulated simply as a function of the local values of the phase fields. However, under circumstances where the pre-existing microstructure is much greater in size than the critical nucleus, it is shown (Ref 66) that the elastic energy can be approximated as a product of the nucleus volume, V_0, and a per-volume coefficient that is a function of pre-existing microstructure but independent of V_0:

$$\begin{aligned} \Delta E^{\mathrm{el}}[\phi^*(\mathbf{r})] &\equiv \Delta E^{\mathrm{el}}[\phi_0(\mathbf{r}) + \Delta\phi^*(\mathbf{r})] \\ &\approx \Delta f^{\mathrm{el}}[\phi_0(\mathbf{r})]\cdot V_0 \end{aligned} \qquad (\text{Eq } 29)$$

where $\phi_0(\mathbf{r})$ represents the configuration of the pre-existing microstructure, and $\Delta\phi^*(\mathbf{r})$ is the configuration of an isolate nucleus. This approximation retains the convenience of using Eq 27 with:

$$\Delta f_V = \Delta f_V^{\mathrm{chem}} + \Delta f^{\mathrm{el}}[\phi_0(\mathbf{r})] \qquad (\text{Eq } 30)$$

The computation of $\Delta f^{\mathrm{el}}[\phi_0(\mathbf{r})]$ is straightforward and can be made rather efficient (Ref 64, 67). As an example, Fig. 2 shows snapshots of microstructures simulated with the Langevin force approach and the explicit nucleation algorithm, using the same set of model parameters.

Calculation of Activation Energy and Critical Nucleus Configuration. The energetics of nucleation may be treated on a more general basis using the phase-field method (Ref 14). Note that the total energy functional, Eq 1, is defined not only for an arbitrary microstructure but also for a critical nucleus, or a critical nucleus in an arbitrary microstructure. As a critical nucleus, the phase-field configuration $\phi^*(\mathbf{r})$ is in an (unstable) equilibrium state and must satisfy:

$$\delta\Phi[\phi^*(\mathbf{r})] = 0 \qquad (\text{Eq } 31)$$

as a necessary condition, with $\Phi \equiv E - \lambda C$. $C = C[\{\phi_p\}]$ represents additional physical constraints on the phase fields, and $\lambda = \{\lambda_t\}$ is the Lagrange multiplier of each respective constraint. The solution for the critical nucleus is a saddle point on the high-dimensional surface of Φ and is the maximum on the minimum energy path (MEP) that connects the starting metastable state and the final transformed state. Because the independent variables are now values of a continuum field, instead of a single variable of radius in the classical nucleation
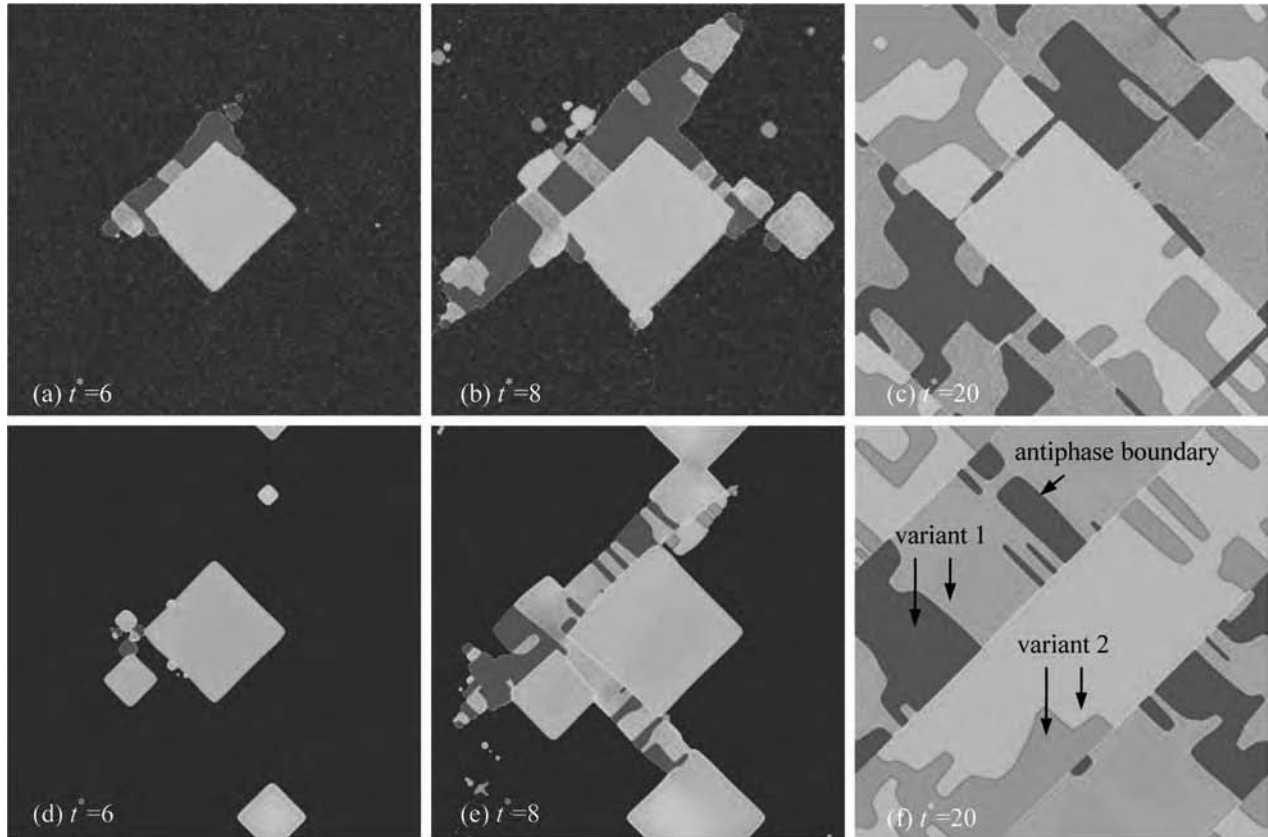
Fig. 2 Microstructures modeled with the Langevin force approach and the explicit nucleation algorithm. Simulated coherent nucleation in a cubic → tetragonal transformation in two dimensions by (a–c) the Langevin force approach and (d–f), explicit nucleation algorithm approach. Reduced time (t) is shown. (Model output images are in color.) Source: Ref 64

theory, the critical nucleus is treated in a much more general and rigorous way that also takes into account the shape, orientation, and profile of the nucleus and its long-range interactions with any stress-carrying pre-existing defects. This is the basis of the nonclassical nucleation theory of Cahn and Hilliard (Ref 14). With $\phi^*(\mathbf{r})$, the activation energy can be calculated:

$$\Delta F^* = E[\phi^*(\mathbf{r})] \qquad \text{(Eq 32)}$$

and used to compute the nucleation rate in Eq 25. This approach allows for the elimination of the assumption made in the preceding section for treating coherent nucleation; that is, the criterion that pre-existing chemical and structural nonuniformities be much greater in length scale than the critical nucleus is not required anymore.

However, the analytical solution $\phi^*(\mathbf{r})$ is rather complex to obtain, except for a few simple cases (Ref 4, 68). This is due to the very high dimensional space that Φ is usually associated with. Because $\phi^*(\mathbf{r})$ corresponds to unstable stationary points, the conventional phase-field dynamic equations governed by total free energy minimization are incapable of obtaining the exact solution, even with the use of Langevin force terms.

The nudged elastic band (NEB) method (Ref 69–71) has been a reliable approach in theoretical

chemistry and condensed matter physics for finding saddle points. To search for the solution $\phi^*(\mathbf{r})$, the NEB method generates a chain of replicas (nodes) of microstructural configurations that approximately pass the saddle point and applies the "force" on node i as (Ref 72):

$$\mathbf{F}_i = \mathbf{F}_i^s|_\parallel + \mathbf{F}_i^p|_\perp \qquad \text{(Eq 33)}$$

It is a sum of a spring-type force, $\mathbf{F}_i^s$, determined by the relative distance between adjacent nodes and a potential force, $\mathbf{F}_i^p \equiv -\delta\Phi/\delta\phi_i$, determined by the total energy surface. The subscripts $\parallel$ and $\perp$ stand, respectively, for the parallel and perpendicular component of the forces resolved on the local tangent of the node, which is evaluated from the energy Φ of the node and its immediate neighbors along the chain (Ref 72). Each node, except for the two fixed ends, is relaxed iteratively subject to the force, $\mathbf{F}_i$, which is simultaneously updated, until the chain converges. For problems of stress-driven nucleation and nucleation in phase transformations, the original NEB method was found to be inefficient, because the last node (the final transmission state) could be very far from the saddle point, and too many nodes were required to describe the downhill portion of the MEP. A so-called free-end modification to the

NEB method greatly improves the efficiency (Ref 73, 74) by allowing the last node, N, to be a nonminima, which furthermore moves to swing the band to improve its posture. The force on node N is:

$$\mathbf{F}_N = \mathbf{F}_{N,N-1}^s - \frac{\mathbf{F}_{N,N-1}^s \cdot \mathbf{F}_N^p}{\mathbf{F}_N^p \cdot \mathbf{F}_N^p} \mathbf{F}_N^p \qquad \text{(Eq 34)}$$

where $\mathbf{F}_{N,N-1}^s$ is the spring force between node $N - 1$ and node N.

The procedure for using the NEB method with the phase-field method may be illustrated in Fig. 3 with the same example of two-dimensional cubic → tetragonal transformation (Ref 16). First, run a conventional phase-field simulation with Langevin force constantly applied to the cubic matrix until the tetragonal precipitate phase emerges and becomes stable, even after the Langevin force is turned off. The amplitude of the Langevin force could be increased artificially to accelerate the nucleation process but keep the simulation stable. Second, take the configuration of the supercritical tetragonal phase as the end node of the NEB chain, and assign the first node as a homogeneous parent (cubic) phase; run the NEB relaxation subject to the force $\mathbf{F}_i$ specified in Eq 33 and 34. The combination of the two steps allows a more reliable search for the true saddle
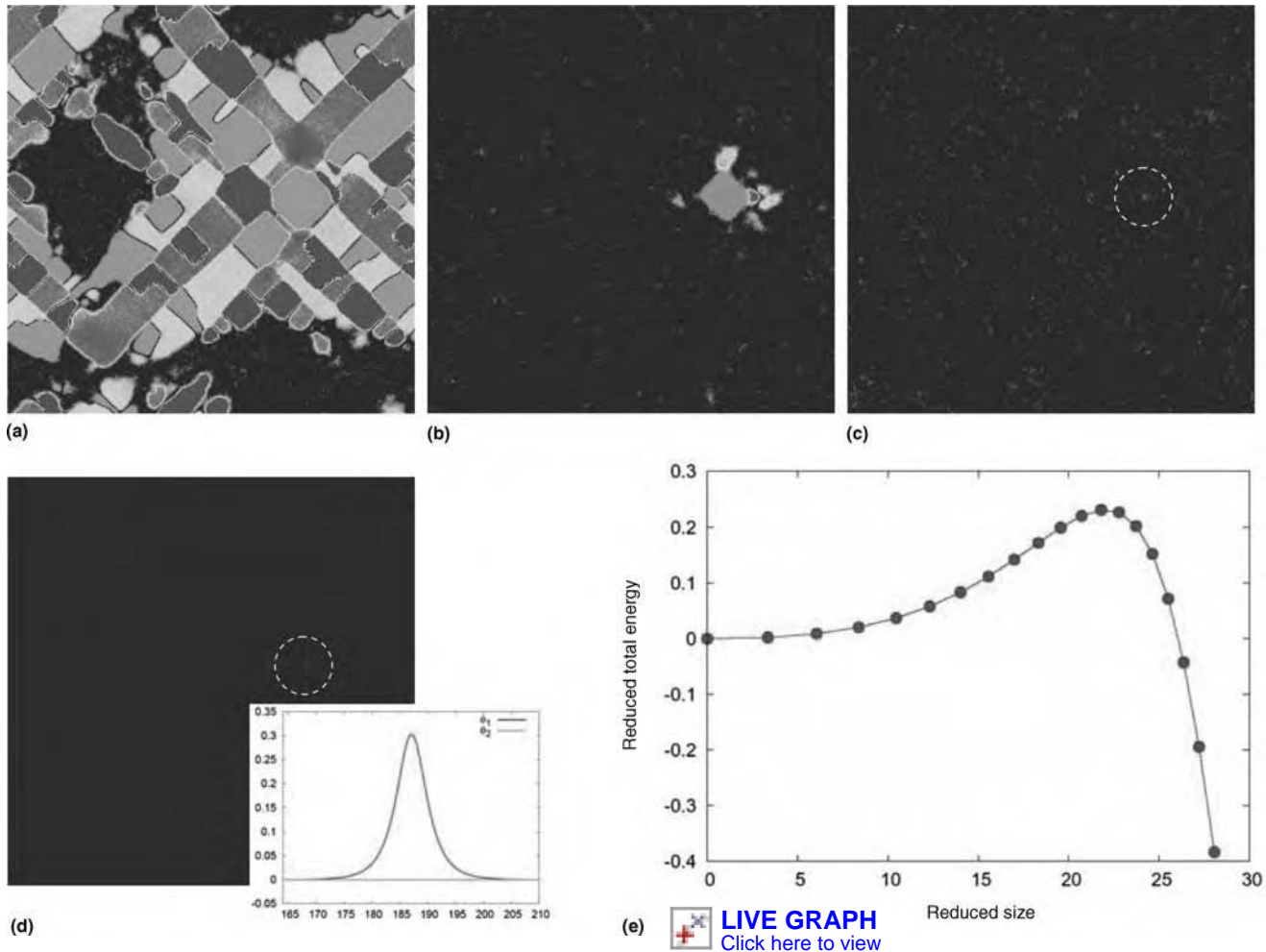
Fig. 3 Use of the nudged elastic band (NEB) method with the phase-field method. (a–c) Microstructural evolution during a cubic → tetragonal transformation in two dimension simulated by Langevin dynamics of the phase-field method. (b) Supercritical configuration at an earlier moment of (a). Using this as the end configuration in the NEB calculation gives the nucleation barrier (e) and the critical nucleus configuration (in the dashed circle in d), with the one-dimensional cross-sectional profile given in the inset, which is compared to a near-critical configuration (c) that is traced back from (b) in the Langevin dynamics. (Model output images are in color.) Source: (a–e) Ref 16

point for a critical nucleus without the artifacts due to ad hoc assumptions on the end-node configuration.

Modeling Growth and Coarsening

The deterministic phase-field kinetic Eq 6 and 12 cover both growth and coarsening, with the driving force being the functional variation of the total free energy and the kinetic rate in proportion to the driving force. These equations can be reduced to sharp-interface equations of domain (precipitate or grain) growth and coarsening (for example, see Ref 75). For a spherical domain, for example, they can be reduced, respectively, to (Ref 13):

$$\frac{dR}{dt} = \frac{2L\kappa_\eta \eta_s}{\sigma} H - \frac{L\kappa_\eta (d-1)}{R} \qquad \text{(Eq 35)}$$

$$\frac{dR}{dt} = \frac{D}{R}\left[\frac{\delta X}{\Delta X} - \frac{(d-1)d_0}{R}\right] \qquad \text{(Eq 36)}$$

where σ is the interfacial energy, $D = V_m^2 M(d^2 f_m/dX^2)$ is the diffusivity, $d_0 = V_m^2 \sigma(d-1)/(d^2 f_m/dX^2)/(\Delta X)^2$ is the correlation length, $\Delta X = X_\beta - X_\alpha$ is the miscibility gap, and δX is the concentration supersaturation. The growth rate, dR/dt, for a nonconserved field (Eq 35) reproduces a linear kinetic law under a constant external field, H, or a parabolic kinetic law for pure curvature-driven kinetics (in grain growth, for instance), depending on which of the two terms is dominant at the right-hand side (RHS) of the equation. For conserved fields such as solute composition, the growth law can be parabolic, driven by supersaturation, δX, or cubic, driven by curvature, governed respectively by the two terms on the RHS of Eq 36. The second term is what underlies coarsening kinetics.

Apart from these basic features, applications of the phase-field method for modeling microstructural evolution in alloy systems of practical interests face a number of difficult challenges. Some of them include resolving the intrinsic multiscale

nature of a microstructure (e.g., domain size and boundary width) and handling of multicomponent diffusion in multiphase systems. (The time scale is associated with the length scale.)

Simulation Length Scale. According to Eq 13 and 14, the interfacial width and energy are determined by the interplay between the local free energy (in particular, Δf) and the gradient energy. When applied at a microscopic length scale ($\sim$nm), the phase-field models predict equilibrium boundary widths that correspond to their natural values (typically of the order of nanometers, depending on particle size and temperature). When applied at coarse-grained length scales (e.g., $\sim$μm), however, the phase models yield interfacial widths far exceeding their natural values. The interfacial regions in a phase-field model must be numerically smooth (several grid-size wide, regardless of the actual grid size) to ensure the accuracy in evaluation of the gradient terms. If the local chemical free energy model is formulated based on an available thermodynamic database, such

as the Calculation of Phase Diagrams (CALPHAD) database, which fixes the Δf values in Eq 13 and 14, the interfacial energies required to keep a diffuse interface at ~μm length scales will be several orders of magnitude higher than their physical values (Ref 29). This will significantly alter the driving force for microstructure evolution, in particular at early stages of a phase transformation, where microstructural evolution is dominated by nucleation events, and at later stages, where microstructural evolution is dominated by coarsening events.

To decouple boundary width with physically meaningful boundary energy in the diffuse-interface model so that the phase-field method can be coarse-grained to an arbitrary length scale without encountering unrealistically high interfacial energies, Kim et al. (Ref 12) proposed a new way of treating interfaces. In their formulation, the interface region is treated as a homogeneous mixture of the adjacent phases (α, β,...) characterized, respectively, by phase compositions (X_α, X_β,...), and the diffusion potentials of each phase composition are required to be equal; that is, $\partial f_\alpha/\partial X_\alpha = \partial f_\beta/\partial X_\beta =...$, where $f_\alpha = f_\alpha(X_\alpha)$, $f_\beta = f_\beta(X_\beta)$..., are the free energies of the respective phases. The latter constraint corresponds to a parallel-tangent construction and gives rise to a force (diffusion potential) balance condition among constituent phases in thermodynamics. Such a treatment, combined with the absence of gradient terms on the composition field, completely removed the contribution of the composition profile across an interface to the interface energy (Fig. 4). An earlier model by Tiaden et al. (Ref 76) produces the same feature at a dilute solution limit.

Multicomponent Diffusion. The phase-field kinetic equations for conserved fields (Eq 5, 6) can be extended directly to handle multicomponent diffusion. Since, by definition, the n compositions of an n-component system are subject to the constraint:
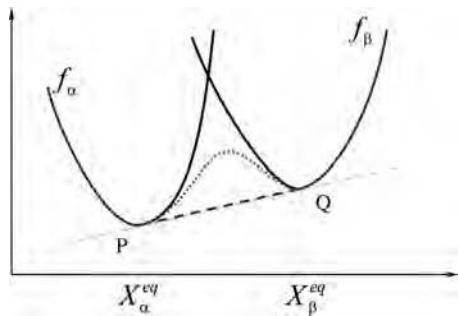
Fig. 4 Free energy curves for individual α and β phases (solid curves); Wheeler-Boettinger-McFadden model (dotted curve); and Kim-Kim-Suzuki (KKS) model (dashed curve). The chemical free energy contribution to interface energy is graphically represented by the area under the free energy curves and above the common tangent (PQ) between X_α^{eq} and X_β^{eq}. In the KKS model, the excess energy in the interface region is removed by making the free energy equal to that of a two-phase mixture (i.e., a straight line between P and Q). Source: Adapted from Ref 12

$$\sum_{i=1}^{n} X_i = 1 \qquad (Eq\ 37)$$

there are only $n - 1$ independent phase-field equations. By eliminating (any) one component, for example, X_n, using Eq 37, the molar free energy becomes:

$$f_m = \sum_{i=1}^{n} \mu_i X_i = \sum_{i=1}^{n-1} \mu_i X_i + \mu_n X_n$$
$$= \mu_n + \sum_{i=1}^{n-1} (\mu_i - \mu_n) X_i \qquad (Eq\ 38)$$

where μ_i is the partial molar free energy (chemical potential) of the i-th component. The multicomponent diffusion can thus be described by coupled $n - 1$ equations:

$$\frac{1}{V_m}\frac{\partial X_i}{\partial t} = \nabla \cdot M_{ij}\nabla V_m \frac{\delta}{\delta X_j}\int f_m dV$$
$$= V_m \nabla \cdot \sum_{j=1}^{n-1} M_{ij}\nabla(\mu_j - \mu_n) \qquad (Eq\ 39)$$

for $i = 1, ..., n - 1$. Here, the gradient energy term has been dropped, as in the solidification models. The mobility coefficient in a multicomponent equation is extended to a matrix. Both f_m and M_{ij} are now functions of the composition ($X_1, X_2,..., X_n$). Their values are typically obtained from accessed thermodynamic and mobility databases (Ref 77–83).

Multiphase-Field Method. Steinbach et al. (Ref 84, 85) generalized the phase-field method (Ref 11) to treat multiple ($N > 2$) coexisting phases, each characterized by its local fraction, $\eta_\alpha(\mathbf{r}, t)$. By this definition, the phase fields, η_α, follow an additional constraint:

$$\sum_{\alpha=1}^{N} \eta_\alpha(\mathbf{r}, t) = 1 \qquad (Eq\ 40)$$

where η_α has a value of unity in the bulk phase α and zero in other phases. An intermediate value between 0 and 1 occurs only in the boundary (interface and multiphase junctions) between the α phase and another phase(s). A general multibody phase-field equation for multiple phases was then decomposed into pair-wise dual-interaction terms:

$$\frac{\partial \eta_\alpha}{\partial t} = -\sum_{\beta=1}^{\bar{N}} \frac{L_{\alpha\beta}}{\bar{N}}\left(\frac{\delta E^{chem}}{\delta \eta_\alpha} - \frac{\delta E^{chem}}{\delta \eta_\beta}\right) \qquad (Eq\ 41)$$

where $\bar{N}$ is the number of phases (labeled by β) adjacent to phase α with values η_β between (but not including) 0 and 1. The index β in the summation only counts on those phases. The kinetic coefficient L is now taken in the form of a pair-wise coefficient between two joining phases. Equation 41 was derived (Ref 85) in a variational framework with the use of a Lagrange multiplier to account for the interdependence among η_α (Eq 40) and resolved a

previous issue of violation to interface stress balance (Ref 84).

Incorporation of compositional fields to the multiphase-field model was also developed in a solidification application (Ref 76) and later in the framework of Kim et al. (Ref 86) for modeling eutectic solidification of a ternary alloy. The methodology applies equally to solid-state phase transformations and grain growth (Ref 87).

Material-Specific Inputs—Thermodynamic and Kinetic Data

The phase-field method has made its way to industrial applications in recent years. Different from fundamental studies that are interested in generic prototype model systems, industrial applications require the consideration of actual materials systems. In general, tailoring of a phase-field model to a specific alloy involves matching a number of model parameters to material-specific properties. Referring to Eq 5, 6, and 12, these model parameters include the chemical free energy, f_m, gradient coefficient, κ_m, kinetic coefficients M and L, and molar volume, V_m. To take into account elastic behaviors, one will need additional parameters, such as elastic modulus, lattice parameters, and lattice correspondence between the precipitate and matrix phases. The molar volume is usually assumed to be constant or a linear function of solute concentrations, with a typical value of ~10^{-5} m^3/mol. The gradient coefficient, κ_m, can be determined by matching interface energy and width to experimental values in conjunction with the bulk free energy, f_m, for example, according to the relations in Eq 13 and 14. This leaves two important independent model inputs: the chemical free energy and the kinetic coefficients (mobilities).

The energy and mobility of most alloy systems possess rather complex dependence on multiple field variables. For instance, a commercial alloy may have more than ten chemical elements, and the corresponding Gibbs free energy and mobility matrix are both functions of the concentration of these elements, in addition to temperature and pressure. To present them in tractable forms for practical modeling, one generally must (1) choose appropriate base functions to represent these quantities, and (2) determine the coefficients in the base functions by fitting the free energy and mobilities to thermodynamic and mobility databases.

While the choice of an appropriate set of base functions varies from application to application, and a different choice could even be applied in a mixed way, two types of approaches have been use widely. The first approach is to adopt a polynomial base function. It has been used in solidification models to construct a free energy for liquid and solid phases. One of the original forms appears as (Ref 88):

$$f(\eta, m) = \frac{1}{4}\eta^4 - \left(\frac{1}{2} - \frac{m}{3}\right)\eta^3 + \left(\frac{1}{4} - \frac{m}{2}\right)\eta^2$$

(Eq 42)

The construction of the function is phenomenological and provides two local minima, respectively, for a solid phase (at $\eta = 1$) and a liquid phase (at $\eta = 0$) and a temperature-dependent coefficient m that switches the global minimum between the solid and liquid phases as it varies above and below zero. In particular, the coefficient m was explicitly linked to the temperature as (Ref 11) (Fig. 5):

$$m(T) = -\beta = -\frac{L\delta}{2\sigma}\frac{T - T_m}{T_m}$$

(Eq 43)

The coefficients associated with the free energy of Eq 42 or 43 and with other phase-field parameters (e.g., κ_m and L) are fitted to a combination of measurable physical parameters in the standard (sharp-interface) solidification model, for example, interface energy, latent hear, interfacial width, and the velocity coefficient, by matching the phase-field model to the standard sharp-interface model (Ref 11).

In the phase-field models of solid-state structural transformations, the polynomial form is often applied by following the Landau theory (Ref 65). It is used to expand a free energy into a power series and uses symmetry operations to eliminate and combine terms. From the Landau theory, the polynomial function typically takes a form of 2-4 power series for second-order phase transitions and 2-3-4 or 2-4-6 forms for first-order transitions. Equation 42 is actually in a 2-3-4 form. The phase-field variables treated in this framework are defined in a physical context and are regarded as (long-range) order parameters originated from order-disorder transitions. An example of such is the free energy of the disordered γ phase and ordered γ' phase of a binary nickel-aluminum alloy (Ref 6, 61):
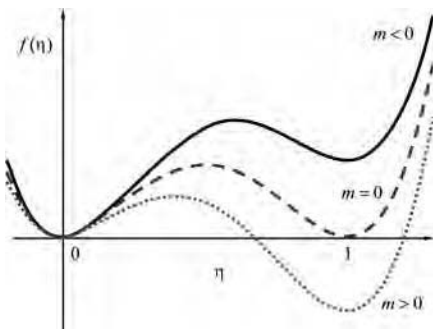
Fig. 5 Free energy curves of Eq 42. The temperature-dependent parameter m shifts the minimum (stable phase) from liquid to solid as its value changes from negative to positive.

$$f(c, \eta_1, \eta_2, \eta_3) = \frac{a_1}{2}(c - c_1)^2 + \frac{a_2}{2}(c_2 - c)\sum_{i=1}^{3}\eta_i^2$$
$$- \frac{a_3}{3}\eta_1\eta_2\eta_3 + \frac{a_4}{4}\sum_{i=1}^{3}\eta_i^4$$

(Eq 44)

where (η_1, η_2, η_3) are three long-range order parameters that account for four antiphase domains of γ' phase at (1, 1, 1), (1, −1, −1), (−1, 1, −1), (−1, −1, 1), and the disordered γ phase at (0,0,0). The coefficients are fitted to thermodynamic parameters that characterize both equilibrium (e.g., equilibrium compositions) and nonequilibrium (e.g., driving force of the phase transformation) properties.

The second type of base functions is originated from solution models in chemical thermodynamics and is typically applied to composition fields. For example, a regular solution model for a ternary alloy reads:

$$f(X_1, X_2, X_3) = \sum_{i=1}^{3}X_iG_i^0 + RT\sum_{i=1}^{3}X_i\ln X_i$$
$$+ \sum_{i=1}^{3}\sum_{j=i+1}^{3}X_iX_jL_{ij}^{00}$$

(Eq 45)

This approach has been systematically matured in the past decade within the CALPHAD framework (Ref 89, 90), with successful extension to multicomponent commercial alloys and development of more sophisticated models taking into account structural transformations (e.g., the cluster variation model, Ref 90). In the CALPHAD database, model parameters were systematically optimized by using available experimental data and, more recently, data from first-principles calculations. As an example, the chemical free energy of a binary γ/γ' nickel-aluminum system (compared to Eq 44, using a four-sublattice model) is given as (Ref 91):

$$
\begin{aligned}
f(c, \eta_1, \eta_2, \eta_3) = {}& cg_{Al}^0 + (1 - c)g_{Ni}^0 + c(1 - c)\sum_{i=0}^{3}L_i(2c - 1)^i \\
& + 4U_1c^2(\eta_1^2 + \eta_2^2 + \eta_3^2) \\
& + 12U_4(1 - 2c)c^2(\eta_1^2 + \eta_2^2 + \eta_3^2) - 48U_4c^3\eta_1\eta_2\eta_3 \\
& + 0.25RT\{c(1 + \eta_1 + \eta_2 + \eta_3)\ln[c(1 + \eta_1 + \eta_2 + \eta_3)] \\
& + [1 - c(1 + \eta_1 + \eta_2 + \eta_3)]\ln[1 - c(1 + \eta_1 + \eta_2 + \eta_3)] \\
& + c(1 + \eta_1 - \eta_2 - \eta_3)\ln[c(1 + \eta_1 - \eta_2 - \eta_3)] \\
& + [1 - c(1 + \eta_1 - \eta_2 - \eta_3)]\ln[1 - c(1 + \eta_1 - \eta_2 - \eta_3)] \\
& + c(1 - \eta_1 + \eta_2 - \eta_3)\ln[c(1 - \eta_1 + \eta_2 - \eta_3)] \\
& + [1 - c(1 - \eta_1 + \eta_2 - \eta_3)]\ln[1 - c(1 - \eta_1 + \eta_2 - \eta_3)] \\
& + c(1 - \eta_1 - \eta_2 + \eta_3)\ln[c(1 - \eta_1 - \eta_2 + \eta_3)] \\
& + [1 - c(1 - \eta_1 - \eta_2 + \eta_3)]\ln[1 - c(1 - \eta_1 - \eta_2 + \eta_3)]\}
\end{aligned}
$$

(Eq 46)

with the long-range order parameters in Eq 44 being replaced by site fractions. The model parameters g_{Al}^0, g_{Ni}^0, L_i, U_1, and U_4 are obtained by using the CALPHAD technique. The free energy (Eq 46) has been applied directly in phase-field modeling of microstructural evolution in nickel-aluminum (Ref 78). It was further extended to multicomponent systems and used in phase-field modeling by defining an individual set of order parameters for each element (see review in Ref 92).

A combination of polynomial and CALPHAD approaches has become an effective means for treating multiphase systems of both structural and chemical nonuniformities. In this approach, the CALPHAD method provides the free energy of each individual phase, f_p ($p = 1,2,\ldots,N$), and the polynomial is used to synthesize all phases in a single function:

$$f(\{X_i\}, \{\eta_p\}) = \sum_{p=1}^{N}\eta_pf_p(\{X_i\})$$
$$+ \sum_{p=1}^{N}\sum_{q=p+1}^{N}\omega_{pq}\eta_p\eta_q$$

(Eq 47)

This treatment simplifies the task for coupling different phases as well as multicomponent compositions. It inherits the flexibility of adjusting interface properties through parameter ω_{pq} and the gradient coefficient κ_m from solidification models and, in the meantime, takes advantage of accurate material-specific multicomponent free energies by the CALPHAD method. With this technique, construction of the Kim-Kim-Suzuki (KKS) model for scaling the interface width and thus the simulation length scale is also straightforward to implement (Ref 86). A similar approach was applied to nickel-aluminum (Ref 80):

$$
\begin{aligned}
f(c, \eta_1, \eta_2, \eta_3, \eta_4) = {}& h(\eta_1, \eta_2, \eta_3, \eta_4)f_{\gamma'} \\
& + [1 - h(\eta_1, \eta_2, \eta_3, \eta_4)]f_{\gamma} \\
& + \omega g(\eta_1, \eta_2, \eta_3, \eta_4)
\end{aligned}
$$

(Eq 48)

where $f_{\gamma'}$ and f_{γ} are the chemical free energies of the γ' and γ phases, respectively, and:

$$h = \sum_{i=1}^{4}[\eta_i^3(6\eta_i^2 - 15\eta_i + 10)]$$

is an interpolation function with a value from 0 to 1, and a double-well function:

$$g = \sum_{i=1}^{4}[\eta_i^2(1 - \eta_i)^2] + \alpha\sum_{i=1}^{4}\sum_{j>i}^{4}\eta_i^2\eta_j^2$$

Note that in Eq 48 there are four phenomenological order parameters to represent the four antiphase domains at (1,0,0,0), (0,1,0,0), (0,0,1,0), and (0,0,0,1), instead of three physical long-range order parameters in Eq 44 or 46.

The chemical mobility in multicomponent diffusion (Eq 39) may be given as:

$$M_{ij} = \frac{1}{V_m}\sum_{l=1}^{n}(\delta_{jl} - X_j)(\delta_{lk} - X_k)X_lM_l$$

(Eq 49)

where $M_l = M_l(\{X_i\}, \{\eta_p\})$ is the mobility of component l in a mixture of multiple phases characterized by $\{\eta_p\}$. In general, it is dependent on both composition and long-range order parameters, for example (Ref 81):

$$M_l = M_l^{\alpha} + M_l^{\beta} - (M_l^{\alpha})^{\eta}(M_l^{\beta})^{1-\eta}$$

(Eq 50)

which is a combination of the atomic mobilities, M_l^{α} and M_l^{β}, of α and β phases that can be obtained from mobility databases.

The mobility coefficient for the long-range order parameters, L, characterizes the contribution of interface kinetics (due to either the kinetic rate of atomic attachment on a physical interface or an effect due to the finite width of a mathematical-diffuse interface as compared to a sharp-interface model). For solid-state phase transformations, because the interface motion is usually under a diffusion-controlled limit, L can be determined at a vanishing kinetic coefficient condition (Ref 86).

Examples of Applications

The following four examples illustrate some aspects of phase-field modeling discussed thus far. These examples include applications of the phase-field method at the microscopic length scale, with focus on understanding and predicting fundamental properties of individual defects, as well as applications at the coarse-grained level, where collective behaviors of complex microstructures subject to external conditions are the main concerns. As the length scale increases, the strategies for incorporating changes to the definition of phase fields and energy models and coupling of multiple crystalline defects in microstructure evolution are demonstrated.

Example 1: Critical Nucleus in Cubic → Tetragonal Transformation. The transition of cubic to tetragonal phase involves changes in crystal structure and symmetry, which introduces elastic energy to the activation energy of nucleation. From crystallographic symmetry, there are three equivalent orientational variants of the tetragonal phase forming in the parent cubic phase matrix (Fig. 6), each characterized by the stress-free transformation strain (SFTS):
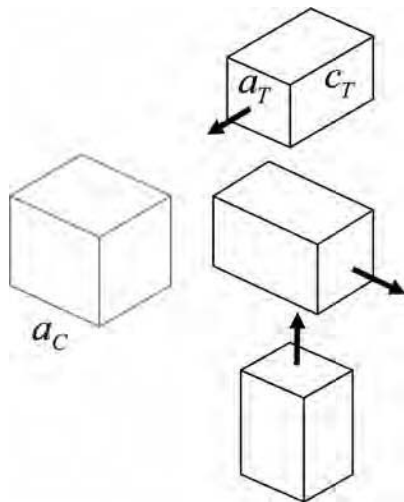
$$\varepsilon_{ij}^{T0}(1) = \varepsilon_0 \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

$$\varepsilon_{ij}^{T0}(2) = \varepsilon_0 \begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

$$\varepsilon_{ij}^{T0}(3) = \varepsilon_0 \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad \text{(Eq 51)}$$

where ε_0 is a scalar constant determined by the change in the lattice constant $\varepsilon_0 = (a_C - a_T)/a_C$. The nonisotropic SFTS associated with each variant results in a deviation of the shape of the nucleus from a sphere. The equilibrium shape is therefore determined by both interface energy and elastic energy. Moreover, a nucleating embryo may adopt a composite configuration that is formed by more than one variant. The co-existing variants reduce the overall SFTS of the embryo and therefore the distortion to the parent matrix. This provides a possible alternative to the critical nucleus configuration if the reduction in the strain energy exceeds the cost of the energy for creating extra interfaces between the variants in the embryo. To study this behavior in more detail, a chemical free energy was chosen:

$$f(\eta_1, \eta_2, \eta_3)$$
$$= \Delta f_0 \left[0.1 \sum_{p=1}^{3} \eta_p^2 - 3.2 \sum_{p=1}^{3} \eta_p^4 + 2.1 \left(\sum_{p=1}^{3} \eta_p^2 \right)^3 \right]$$
$$\text{(Eq 52)}$$

This is a typical Landau polynomial in 2-4-6 form for a first-order phase transition (see the previous section) and gives local minima at $(\eta_1, \eta_2, \eta_3) = (0,0,0)$ for the parent cubic phase and $(\pm 1,0,0)$, $(0,\pm 1,0)$, and $(0,0,\pm 1)$ for the

three variants of the tetragonal phase (Fig. 7). The same free energy was also used for the two-dimensional simulations in the section "Modeling Nucleation" in this article. The same procedure as in that section can be applied to combine the phase-field model (the energetics) and the nudged elastic band method to determine the saddle point on the total free energy surface for the critical nucleus:

$$E = E[\eta_1, \eta_2, \eta_3]$$
$$= \int \left[f + \kappa((\nabla \eta_1)^2 + (\nabla \eta_2)^2 + (\nabla \eta_3)^2) \right] dV$$
$$+ E^{el}[\eta_1, \eta_2, \eta_3]$$
$$\text{(Eq 53)}$$

The total elastic energy, E^{el}, is given by Eq 17 and 18. An isotropic elastic modulus is used so that any deviation of the critical nucleus from a spherical shape is contributed by the SFTS. The Poisson's ratio is chosen as 0.3. The isotropic gradient coefficient is $\kappa = 2.0$. To increase the contribution from the elastic energy over the chemical energy, the following dimensionless ratio is increased from 0.5 to 1.0:

$$\xi \equiv \mu \varepsilon_0^2 / \Delta f_0 \quad \text{(Eq 54)}$$

Here, μ is the shear modulus.

The configurations of the critical nucleus at various values of ξ are shown in Fig. 8. At a small elastic energy contribution ($\xi = 0.5$), the nucleus is formed by a single tetragonal variant and has a nearly spherical shape because of the (isotropic) interface energy. With increased ξ, the nucleus switches to a two-variant composite configuration as the reduction in the elastic energy becomes increasingly significant to the total energy reduction (Ref 16).

Example 2: Dislocation Core Structure. Another application of the phase-field model at the microscopic level is dislocation core

Fig. 6 Three orientational variants in a cubic → tetragonal transformation
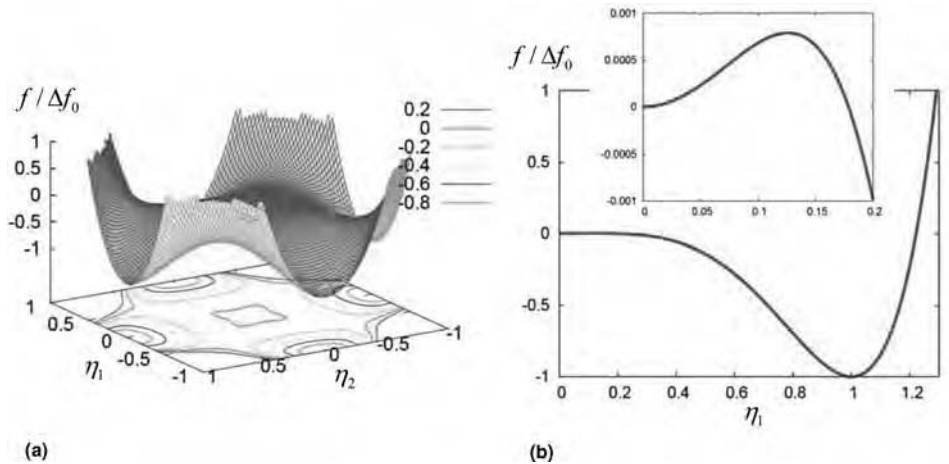
Fig. 7 Chemical free energy of cubic → tetragonal transformation (Eq 52). (a) Free energy surface showing an $\eta_1 - \eta_2$ cross section (0,0) for the cubic phase and $(\pm 1,0)$ and $(0,\pm 1)$ for the tetragonal phase. (b) Profile along η_1 showing that the cubic phase is a metastable phase (a local minimum). The amplitude of the barrier (inset) is determined by Δf_0. (Model output images are in color.)

LIVE GRAPH
Click here to view
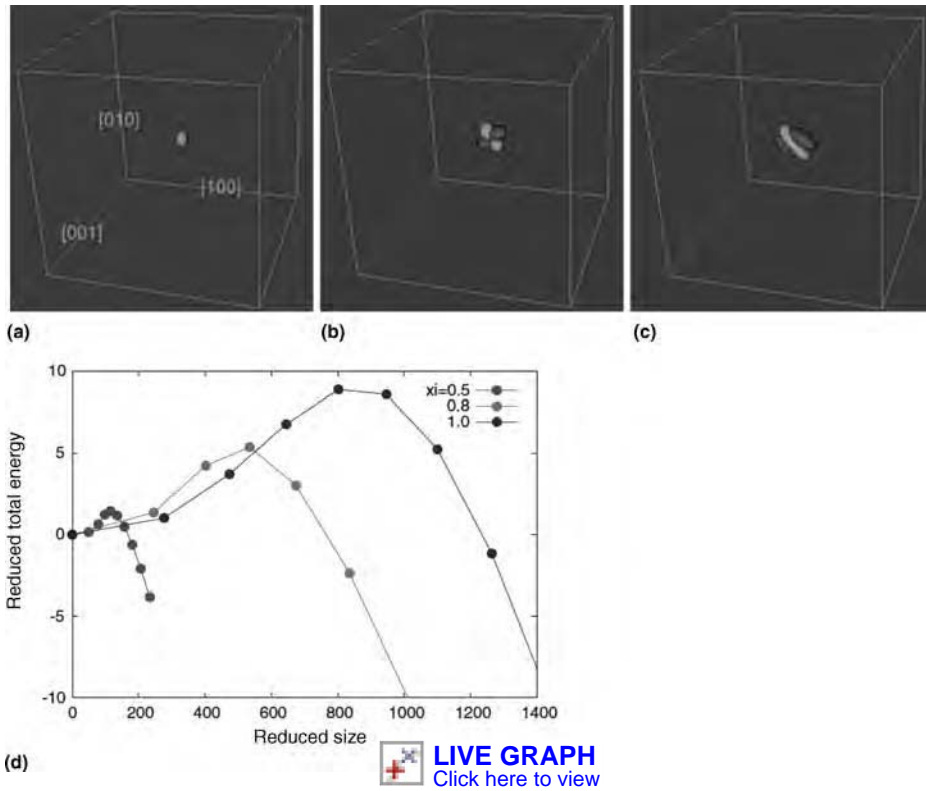
Fig. 8 Transition of the critical nucleus from a single-variant to a two-variant configuration with increasing lattice misfit. (a) ξ = 0.5. (b) ξ = 0.8. (c) ξ = 1.0. The corresponding nucleation barriers are shown in (d). (Model output images are in color.)

modeling (Ref 15, 93, 94). In the framework of the Peierls model of dislocation (Ref 95), the (inelastic) displacement field associated with a dislocation is treated as a continuous variable across the dislocation core. The energy associated with a dislocation consists of a crystalline energy (or misfit energy) from the atomic position disregistry (deviation from its location in a reference perfect crystal):

$$E^{\text{cryst}} = \int dV \frac{\gamma(\mathbf{u}(\mathbf{r}))}{d} \qquad (Eq\ 55)$$

and an elastic energy (Eq 18) with SFTS replaced by the eigenstrain of dislocations:

$$\varepsilon_{ij}^T = \varepsilon_{ij}^T(\mathbf{u}(\mathbf{r})) \equiv [\mathbf{u}(\mathbf{r}) \otimes \mathbf{n} + \mathbf{u}(\mathbf{r}) \otimes \mathbf{n}]/2d \quad (Eq\ 56)$$

In Eq 55 and 56, **u** is the inelastic displacement vector, **n** is the normal vector of the slip plane, and d is the corresponding interplanar distance. γ is the generalized stacking fault (GSF) energy (Ref 96), which is typically calculated by ab initio methods or approaches based on empirical interatomic potentials, such as the embedded atom method. In dislocation models, the crystalline energy has a role equivalent to the chemical free energy, E^{chem}, in phase transformation modeling. To be consistent with the Peierls model, the gradient term in a phase-field model is removed because the counterbalance between the elastic energy and the crystalline

energy automatically produces a diffuse dislocation core profile at the microscopic length scale considered here.

Figure 9 shows an example of the predicted core structure of a ⟨110⟩{111}-type superdislocation in Ni₃Al (γ′). In this case:

$$\mathbf{n} = [111]/\sqrt{3}$$

and:

$$d = a/\sqrt{3}$$

where a is the lattice parameter. The GSF energy, γ, and the anisotropic elastic constants are from Ref 97. With the balance between the elastic energy (which prefers an infinitely extended dislocation core) and the crystalline energy (which prefers an infinitely compact dislocation core), the dislocation core in equilibrium exhibits a fourfold extended structure, for both edge and screw type, that consists of two complex stacking faults (CSFs) and one antiphase boundary (APB), as shown in the inset in Fig. 9. The core structures calculated by the phase-field model are in good agreement with the Peierls model, with the same input of γ and elastic constants.

Dislocation and γ/γ′ Microstructure Interaction. External loads on single-crystal nickel-aluminum (γ/γ′) alloys cause activities of dislocations in the γ channels. The interaction of these dislocations with the microstructure at

elevated temperatures leads to directional coarsening (rafting) of γ′ precipitates and influences the long-term creep behavior of the alloy in its service life (Ref 98–103). Simulation of the channel dislocation and γ/γ′ microstructure is an example of integrating different defect fields in a phase-field model. In the present case, a conserved phase field is defined for the solute (aluminum)concentration that characterizes the diffusion process of γ′ coarsening, and a set of nonconserved slip displacement fields are introduced to characterize dislocations from all active slip systems.

Since the length scale relevant to coarsening is associated with the typical diffusion field and the size of precipitates ($\sim 10^2$ nm), the microscopic details such as dislocation core structures must be coarse-grained by restoring the use of the gradient term of the displacement fields (Ref 104). The choice of the gradient term coefficient is determined by making the dislocation core numerically smooth (effective core width approximately 20 nm) on the mesoscale computational grid (with grid size 6.6 nm).

The elastic interaction between microstructure and dislocations is considered here as the primary contributing factor to γ′ rafting. In particular, the transformation strain (Eq 17) is written as (Ref 105):

$$\varepsilon_{ij}^T(\mathbf{r}) = \frac{2(a_{\gamma'} - a_\gamma)}{(a_{\gamma'} + a_\gamma)(c_{\gamma'}^{\text{eq}} - c_\gamma^{\text{eq}})}[c(\mathbf{r}) - c_\gamma^{\text{eq}}]\delta_{ij}$$
$$+ \sum_q \frac{\mathbf{b}_0(q) \otimes \mathbf{n}(q) + \mathbf{n}(q) \otimes \mathbf{b}_0(q)}{2d(q)}\eta_q(\mathbf{r})$$
$$(Eq\ 57)$$

The first term on the RHS of Eq 57 is from the lattice misfit between the γ (a_γ) and γ′ ($a_{\gamma'}$) phases. δ_{ij} is the Kronecker delta and represents a dilatational strain tensor. c_γ^{eq} and $c_{\gamma'}^{\text{eq}}$ are, respectively, the equilibrium concentrations in the γ and γ′ phases. The second term, similar to Eq 56, represents the inelastic strains from individual slip systems (labeled by q), each associated with the unit Burgers vector, $\mathbf{b}_0(q)$, and characterized by η_q. $\mathbf{b}_0$ is in a form of $a/2\langle 110 \rangle$, with a as the lattice parameter. It is noted that the two SFTS terms in Eq 57 account for not only the elastic interaction among the respective defects (i.e., precipitate-precipitate, dislocation-dislocation) but also their mutual interactions through the total elastic energy (Eq 18).

Depending on the Schmidt factor, the response of each slip system of dislocations to the external load is different: Under [100] uniaxial load, only eight of the twelve primary $1/2\langle 110 \rangle\{111\}$ slip systems have a nonzero resolved stress component. Additionally, the coherency stress of the γ/γ′ microstructure further differentiates the activities of the eight slip systems in each of the three types of ⟨110⟩-oriented γ channel (for example, Fig. 10). The resultant uneven distribution of slip systems in each γ channel causes unequal diffusion
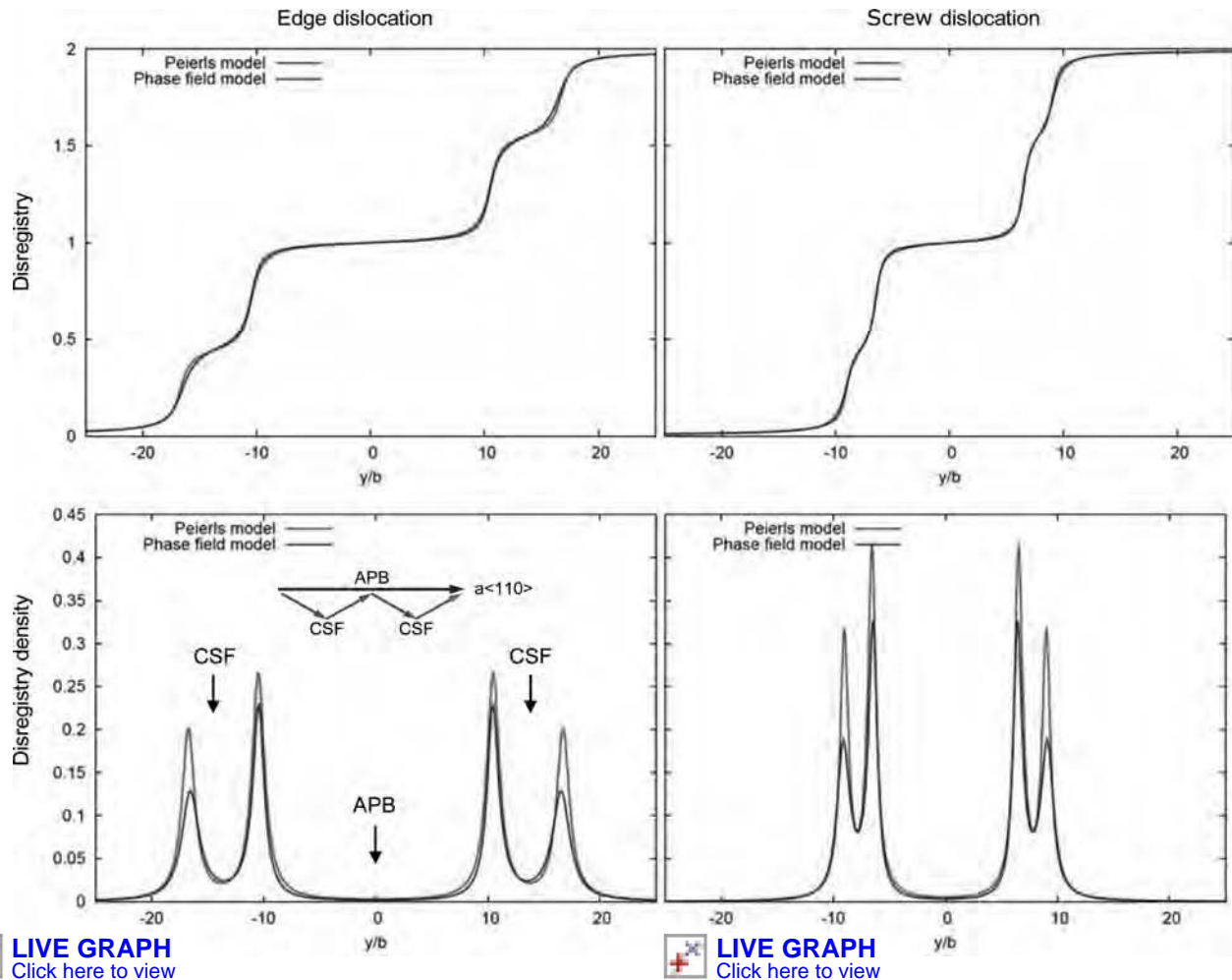
Fig. 9 Core structures of ⟨110⟩{111}edge and screw dislocations in Ni₃Al and a comparison with solutions from the Peierls model (courtesy of Professor Gunther Schoeck). Both calculations used the same input of generalized stacking fault energy, γ, and elastic moduli. $b = a/\sqrt{2}$, where a is the lattice parameter. CSF, complex stacking fault; APB, antiphase boundary. (Model output images are in color.)
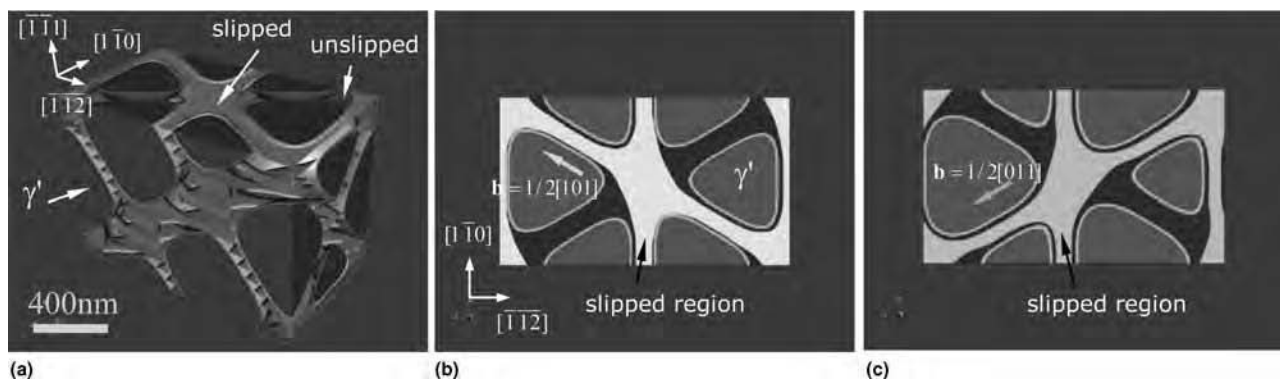
Fig. 10 Dislocations (shown as the boundaries between slipped and unslipped regions) in the γ/γ′ microstructure of −0.3% lattice misfit under 152 MPa [001] tensile stress. Cross-sectional view on the slip plane for (b)1/2[101](1̄11)dislocations and (c)1/2[011](1̄11)dislocations. Both show uneven distribution in γ channels. (Model output images are in color.) Source: Ref 105
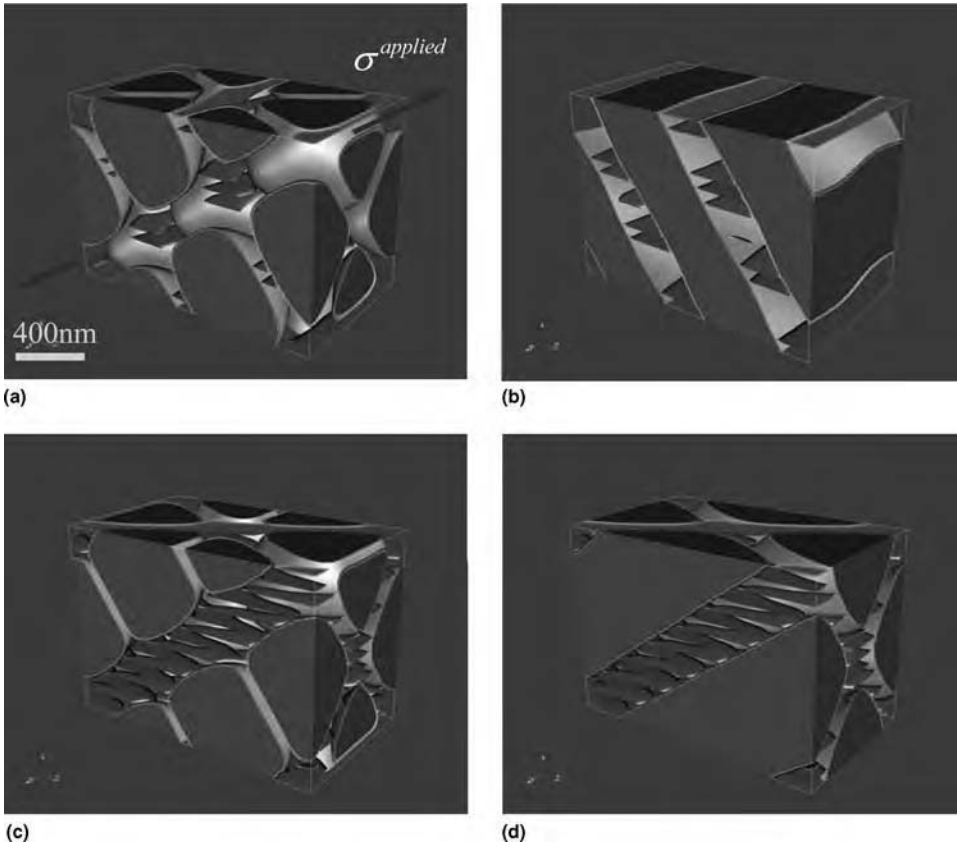
Fig. 11 Formation of N-type raft at (a) 3.6 h and (b) 10.7 h in an alloy with −0.3% lattice misfit, and formation of P-type raft at (c) 3.6 h and (d) 7.2 h with +0.3% lattice misfit under 152 MPa tensile stress (σ) along the [001] direction. The coordinate system is the same as in Fig. 10. (Model output images are in color.) Source: Ref 105
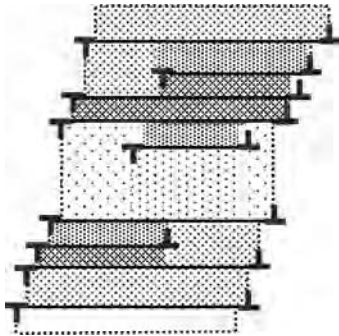
Fig. 12 Discrete dislocations and a continuum field of the inelastic (plastic) strain field (dotted and shadowed regions) that yields the same plastic deformation at a coarse-grained length scale (much greater than dislocation core size). (Model output images are in color.)

(chemical) potentials in each channel under an applied load, and thus, diffusion occurs differently. This difference results in some γ channels being closed up and others becoming wider, consequently forming the plate (N)-type and elongated (P)-type γ′ morphologies (Fig. 11).

γ′ Rafting during Creep Deformation. The model of dislocation-microstructure interaction in the last example can be extended to a larger length scale close to experimental observations of rafted γ′ microstructures to account for statistical variations in the morphology and spatial distribution of γ′ precipitates and dislocations. To do so, major modifications were made to both dislocation and γ/γ′ microstructure models (Ref 106).

First, the discrete description of dislocations is extended to a continuum description of the inelastic (plastic) deformation field. The phase fields that were originally defined to characterize the inelastic displacement field of individual dislocations on discrete slip planes are replaced by new phase fields that describe plastic strain fields distributed continuously in the space (Fig. 12). The crystalline energy that originally carries periodic interatomic potential with respect to inelastic displacement (or disregistry) is, by coarse-graining, converted to a constant averaged potential energy that only renormalizes the total energy.

Second, the γ/γ′ microstructure evolution incorporates the model of Kim et al. (Ref 12) (see the section "Modeling Growth and Coarsening" in this article) to allow treatment of

solute diffusion at an increased length scale without artificially altering the driving forces for precipitate growth and coarsening. Accordingly, the chemical free energy for γ/γ′ phases is chosen in a form of Eq 48, with four phenomenologically defined order parameters to characterize the γ and γ′ phases and the four types of antiphase domain in γ′. While the individual free energies of the γ and γ′ phases could have been imported from the CALPHAD database for specific alloys, only fitted parabolic polynomials were used in this study, because the simulated microstructure is in a coarsening stage, where solute concentration variation in γ and γ′ is nearly around the equilibrium compositions of each phase.

Figure 13 shows the simulated rafting microstructures at positive and negative lattice misfit. The simulation linear dimension is 5 μm, with a grid size of 20 nm. Besides the development of similar N- and P-type rafting morphologies under uniaxial load conditions, as in the previous example, the microstructure simulated at a larger length scale shows a remarkable resemblance to those observed in the experiment (Ref 103, 107–109) regarding both microstructure morphologies and time scale. Due to the elastic interaction, it can be also seen that the N-type microstructure appears to have more regular alignment across rafted precipitates compared to the P-type microstructure, at the same level of external load and lattice misfit.

Summary

Fundamental issues of phase-field theory and its applications, both as physical and phenomenological models, are discussed. Examples show that when used at the microscopic level, phase-field models can be applied to understand and predict fundamental properties of extended defects such as interfacial width and dislocation core size, chemical and structural variations within the defects and the associated defect energies, as well as activation energy for defect nucleation, which is in contrast to sharp-interface models where the fundamental properties of defects are model inputs rather than outputs. When applied at various coarse-grained levels, the phase-field models lose their abilities to predict defect energies and sizes but gain the ability to deal with the collective behavior of a large ensemble of both chemically and mechanically interacting defects, retaining their advantages over the sharp-interface models in treating complicated geometrical and topological changes. Some of the important issues in dealing with complex alloy systems are addressed, including linking to material databases and the limitations and technical difficulties associated with length scales. Examples show that multiscale modeling within the phase-field method may approach real-world complexity, such as rafting and creep deformation in superalloys.
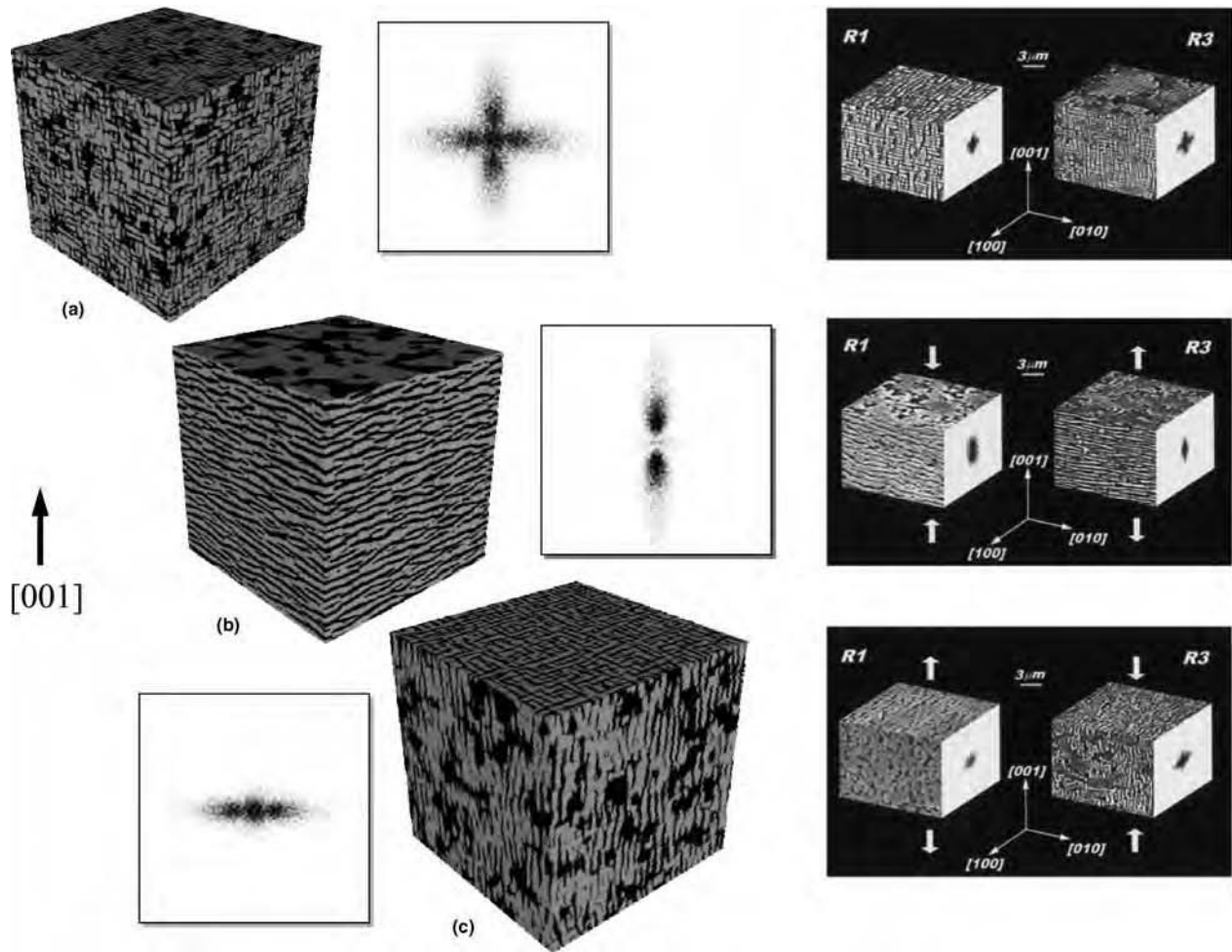
Fig. 13 (a) Simulated γ/γ' microstructures with $\pm0.3\%$ lattice misfit and no external load, aged at 1300 K for 4.7 h. After an additional 5.6 h aging under 152 MPa tensile stress along [001], rafted microstructures developed from (a) by assuming lattice misfit of (b) -0.3% and (c) $+0.3\%$. Source: Ref 106. The insets show the Fourier transform (diffraction pattern) of the corresponding microstructure. The images on the right are from experimental observations under similar conditions. Source: Ref 108

REFERENCES

1. J.D. van der Waals, *Konink. Akad. Weten. Amsterdam,* (Sect. 1) 1, 1893, p 56
2. L. Landau and E. Lifshitz, *Phys. Zeit. Sowjetunion,* Vol 8, 1935, p 153
3. J.W. Cahn and J.E. Hilliard, *J. Chem. Phys.,* Vol 28, 1958, p 258
4. A.G. Khachaturyan, *Theory of Structural Transformations in Solids,* John Wiley & Sons, New York, 1983
5. R. Poduri and L.Q. Chen, *Acta Mater.,* Vol 46, 1998, p 3915
6. Y. Wang, D. Banerjee, C.C. Su, and A.G. Khachaturyan, *Acta Mater.,* Vol 46, 1998, p 2983
7. M. Hillert, *Vol. Sc.D.,* Mass. Inst. Tech., 1956
8. Y.W. Lee and H.I. Aaronson, *Acta Metall.,* Vol 28, 1980, p 539
9. J.S. Langer, *Ann. Phys.,* Vol 54, 1969, p 258
10. J.S. Langer, *Ann. Phys.,* Vol 65, 1971, p 53
11. A.A. Wheeler, W.J. Boettinger, and G.B. McFadden, *Phys. Rev. A,* Vol 45, 1992, p 7424
12. S.G. Kim, W.T. Kim, and T. Suzuki, *Phys. Rev. E,* Vol 60, 1999, p 7186
13. J.S. Langer, *Solids Far from Equilibrium,* C. Godrèche, Ed., Cambridge University Press, New York, 1992
14. J.W. Cahn and J.E. Hilliard, *J. Chem. Phys.,* Vol 31, 1959, p 688
15. C. Shen and Y. Wang, *Acta Mater.,* Vol 52, 2004, p 683
16. C. Shen, J. Li, and Y. Wang, *Metall. Mater. Trans. A,* Vol 39, 2008, p 976
17. W.J. Boettinger, J.A. Warren, C. Beckermann, and A. Karma, *Ann. Rev. Mater. Res.,* Vol 32, 2002, p 163
18. L.Q. Chen, *Ann. Rev. Mater. Res.,* Vol 32, 2002, p 113
19. A.G. Khachaturyan, *Philos. Mag. A,* Vol 74, 1996, p 3
20. J.J. Hoyt, M. Asta, and A. Karma, *Phys. Rev. Lett.,* Vol 86, 2001, p 5530
21. S.M. Foiles and J.J. Hoyt, *Acta Mater.,* Vol 54, 2006, p 3351
22. Y. Wang and L.Q. Chen, *Methods in Material Research,* John Wiley & Sons, Inc., 2000, p 2a.3.1
23. A. Karma, *Encyclopedia of Materials: Science and Technology,* Vol 7, K.H.J. Buschow, R.W. Cahn, M.C. Flemings, B. Ilschner, E.J. Kramer, and S. Mahajian, Ed., Elsevier, Oxford, 2001, p 6873
24. *Handbook of Materials Modeling*, Vol 2, S. Yip, Ed., Springer, 2005, Chap. 7

25. A. Karma and W.-J. Rappel, *Phys. Rev. E,* Vol 57, 1998, p 4323
26. A. Karma and W.-J. Rappel, *Phys. Rev. E,* Vol 53, 1996, p 3017
27. R.F. Almgren, *SIAM J. App. Math.,* Vol 59, 1999, p 2086
28. G.B. McFadden, A.A. Wheeler, and D.M. Anderson, *Physica D,* Vol 144, 2000, p 154
29. C. Shen, Q. Chen, Y.H. Wen, J.P. Simmons, and Y. Wang, *Scr. Mater.,* Vol 50, 2004, p 1023
30. C. Shen, Q. Chen, Y.H. Wen, J.P. Simmons, and Y. Wang, *Scr. Mater.,* Vol 50, 2004, p 1029
31. N. Provatas, N. Goldenfeld, and J. Dantzig, *Phys. Rev. Lett.,* Vol 80, 1998, p 3308
32. W.M. Feng, P. Yu, S.Y. Hu, Z.K. Liu, Q. Du, and L.Q. Chen, *J. Comput. Phys.,* Vol 220, 2006, p 498
33. J. Gruber, N. Ma, A.D. Rollett, and G.S. Rohrer, *Model. Simul. Mater. Sci. Eng.,* Vol 14, 2006, p 1189
34. S. Vedantam and B.S. Patnaik, *Phys. Rev. E,* Vol 73, 2006, p 016703
35. Y. Wang, L.Q. Chen, and A.G. Khachaturyan, *Computer Simulation in Materials Science—Nano/Meso/Macroscopic Space and Time Scales,* H.O. Kirchner, K.P. Kubin, and V. Pontikis, Ed., Kluwer Academic Publishers, 1996, p 325
36. C. Shen and Y. Wang, *Handbook of Materials Modeling, Vol B: Models*, S. Yip, Ed., Springer, 2005, p 2117
37. J.D. Eshelby, *Proc. R. Soc. Lon., Series A,* Vol 241, 1957
38. J.D. Eshelby, *Proc. R. Soc. Lon., Series A,* Vol 252, 1959, p 561
39. A.G. Khachaturyan, *Fiz. Tverd. Tela,* Vol 8, 1966, p 2710
40. A.G. Khachaturyan, *Sov. Phys. Solid State,* Vol 8, 1967, p 2163
41. A.G. Khachaturyan and G.A. Shatalov, *Sov. Phys. Solid State,* Vol 11, 1969, p 118
42. D.Y. Li and L.Q. Chen, *Acta Mater.,* Vol 45, 1997, p 2435
43. Y.U. Wang, Y.M. Jin, and A.G. Khachaturyan, *J. Appl. Phys.,* Vol 92, 2002, p 1351
44. A. Onuki, *J. Phys. Soc. Jap.,* Vol 58, 1989, p 3065
45. A.G. Khachaturyan, S. Semennovskaya, and T. Tsakalakos, *Phys. Rev. B,* Vol 52, 1995, p 15909
46. S.Y. Hu and L.Q. Chen, *Acta Mater.,* Vol 49, 2001, p 1897
47. Y.U. Wang, Y.M. Jin, and A.G. Khachaturyan, *Appl. Phys. Lett.,* Vol 79, 2001, p 3071
48. Y.U. Wang, Y.M. Jin, and A.G. Khachaturyan, *Acta Mater.,* Vol 51, 2003, p 4209
49. D.Y. Li and L.Q. Chen, *Scr. Mater.,* Vol 37, 1997, p 1271
50. N. Zhou, C. Shen, M.J. Mills, and Y. Wang, *Acta Mater.,* Vol 56, 2008, p 6156

51. R. Wagner and R. Kampmann, *Materials Science and Technology: A Comprehensive Treatment,* R.W. Cahn, P. Haasen, and E.J. Kramer, Ed., VCH, Weinheim, New York, 1991
52. J.D. Gunton, M.S. Miguel, and P.S. Sahni, *Phase Transitions and Critical Phenomena,* Vol 8, C. Domb and J.L. Lebowitz, Ed., Academic Press, New York, 1983
53. K.R. Elder, *Comp. Phys.,* Vol 7, 1993, p 27
54. Y. Wang, L.Q. Chen, and A.G. Khachaturyan, *Solid-Solid Phase Transformations*, W.C. Johnson, J.M. Howe, D.E. Laughlin, and W.A. Soffa, Ed., TMS, Warrendale, PA, 1994, p 245
55. Y. Wang, H.Y. Wang, L.Q. Chen, and A.G. Khachaturyan, *J. Am. Ceram. Soc.,* Vol 78, 1995, p 657
56. Y. Wang and A.G. Khachaturyan, *Acta Metall. Mater.,* Vol 45, 1997, p 759
57. Y. Le Bouar, A. Loiseau, and A.G. Khachaturyan, *Acta Mater.,* Vol 46, 1998, p 2777
58. Y.H. Wen, Y. Wang, and L.Q. Chen, *Philos. Mag. A*, Vol 80, 2000, p 1967
59. J.P. Simmons, C. Shen, and Y. Wang, *Scr. Mater.,* Vol 43, 2000, p 935
60. K. Binder, *Rep. Prog. Phys.,* Vol 50, 1987, p 783
61. Y.H. Wen, J.P. Simmons, C. Shen, C. Woodward, and Y. Wang, *Acta Mater.,* Vol 51, 2003, p 1123
62. B. Wang, Y.H. Wen, J.P. Simmons, and Y. Wang, *Metall. Mater. Trans. A,* Vol 39, 2008, p 984
63. Y. Wang, *Materials Science and Engineering, Vol Doctor of Philosophy*, Rutgers, The State University of New Jersey, New Brunswick, NJ, 1995
64. C. Shen, J.P. Simmons, and Y. Wang, *Acta Mater.,* Vol 55, 2007, p 1457
65. L.D. Landau and E.M. Lifshitz, *Statistical Physics,* Pergamon Press, Oxford, New York, 1980
66. C. Shen, J.P. Simmons, and Y. Wang, *Acta Mater.,* Vol 54, 2006, p 5617
67. C. Shen, J.P. Simmons, and Y. Wang, *Solid-to-Solid Phase Transformations in Inorganic Materials,* J.M. Howe, D.E. Laughlin, J.K. Lee, U. Dahmen, and W.A. Soffa, Ed., The Minerals, Metals & Materials Society, Phoenix, AZ, 2005
68. J.W. Cahn, *Acta Metall.,* Vol 10, 1962, p 907
69. H. Jonsson, G. Mills, and K.W. Jacobsen, *Classical and Quantum Dynamics in Condensed Phase Simulations,* B.J. Berne, G. Ciccotti, and D.F. Coker, Ed., World Scientific, 1998, p 385
70. G. Mills and H. Jonsson, *Phys. Rev. Lett.,* Vol 72, 1994, p 1124
71. G. Mills, H. Jonsson, and G.K. Schenter, *Surf. Sci.,* Vol 324, 1995, p 305
72. G. Henkelman and H. Jonsson, *J. Chem. Phys.,* Vol 113, 2000, p 9978

73. T. Zhu, J. Li, A. Samanta, H.G. Kim, and S. Suresh, *Proc. Natl. Acad. Sci. USA,* Vol 104, 2007, p 3031
74. J. Li, P.A. Gordon, and T. Zhu, 2007
75. J.S. Langer, *Direction in Condensed Matter Physics,* G. Grinstein and G. Mazenko, Ed., World Scientific, Singapore, 1986, p 165
76. J. Tiaden, B. Nestler, H.J. Diepers, and I. Steinbach, *Physica D*, Vol 115, 1998, p 73
77. U. Grafe, B. Botteger, J. Tiaden, and S.G. Fries, *Scr. Mater.,* Vol 42, 2000
78. J.Z. Zhu, Z.K. Liu, V. Vaithyanathan, and L.Q. Chen, *Scr. Mater.,* Vol 46, 2002, p 401
79. H. Kobayashi, M. Ode, S.G. Kim, W.T. Kim, and T. Suzuki, *Scr. Mater.,* Vol 48, 2003, p 689
80. J.Z. Zhu, T. Wang, A.J. Ardell, S.H. Zhou, Z.K. Liu, and L.Q. Chen, *Acta Mater.,* Vol 52, 2004, p 2837
81. Q. Chen, N. Ma, K. Wu, and Y. Wang, *Scr. Mater.,* Vol 50, 2004, p 471
82. K. Wu, N. Zhou, X. Pan, J.E. Morral, and Y. Wang, *Acta Mater.,* Vol 56, 2008, p 3854
83. T. Kitashima and H. Harada, *Acta Mater.,* 2009, in press, doi:10.1016/j.actamat.2009.01.006
84. I. Steinbach, F. Pezzolla, B. Nestler, M. Seeβelberg, R. Prieler, G.J. Schmitz, and J.L.L. Rezende, *Physica D,* Vol 94, 1996, p 135
85. I. Steinbach and F. Pezzolla, *Physica D,* Vol 134, 1999, p 385
86. S.G. Kim, W.T. Kim, T. Suzuki, and M. Ode, *J. Cryst. Growth,* Vol 261, 2004, p 135
87. Y. Wang, N. Ma, Q. Chen, F. Zhang, S.L. Chen, and Y.A. Chang, *JOM,* September 2005, p 32
88. R. Kobayashi, *Physica D,* Vol 63, 1993, p 410
89. L. Kaufman and H. Bernstein, *Computer Calculation of Phase Diagrams with Special Reference to Refractory Metals,* Academic Press, New York, 1970
90. N. Saunders and A.P. Miodownik, *CALPHAD (Calculation of Phase Diagrams): A Comprehensive Guide,* Pergamon, Oxford, New York, 1998
91. I. Ansara, N. Dupin, H.L. Lukas, and B. Sundman, *J. Alloy. Compd.,* Vol 247, 1997, p 20
92. T. Kitashima, *Philos. Mag.,* Vol 88, 2008, p 1615
93. C. Shen, J. Li, M.J. Mills, and Y. Wang, *Integral Materials Modeling: Towards Physics-Based Through-Process Models,* G. Gottstein, Ed., Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2007
94. C. Shen, J. Li, and Y. Wang
95. R.E. Peierls, *Proc. Phys. Soc.*, Vol 52, 1940, p 34
96. V. Vitek, *Philos. Mag.,* Vol 18, 1968, p 773
97. G. Schoeck, S. Kohlhammer, and M. Fähnle, *Philos. Mag. Lett.,* Vol 79, 1999, p 849

98. J.K. Tien and S.M. Copley, *Metall. Trans.,* Vol 2, 1971, p 215
99. J.K. Tien and S.M. Copley, *Metall. Trans.,* Vol 2, 1971, p 1663
100. J.K. Tien and R.P. Gamble, *Metall. Trans.,* Vol 3, 1972, p 2157
101. T. Miyazaki, K. Nakamura, and H. Mori, *J. Mater. Sci.,* Vol 14, 1979, p 1817
102. T.M. Pollock and A.S. Argon, *Acta Metall. Mater.,* Vol 41, 1992, p 2253
103. T.M. Pollock and A.S. Argon, *Acta Metall. Mater.,* Vol 42, 1994, p 1859
104. Y.U. Wang, Y.M. Jin, A.M. Cuitino, and A.G. Khachaturyan, *Acta Mater.,* Vol 49, 2001, p 1847
105. N. Zhou, C. Shen, M.J. Mills, and Y. Wang, *Acta Mater.,* Vol 55, 2007, p 5369
106. N. Zhou, C. Shen, M.J. Mills, and Y. Wang, submitted to *Philos. Mag.*
107. O. Paris, M. Fahrmann, E. Fahrmann, T.M. Pollock, and P. Fratzl, *Acta Mater.,* Vol 45, 1997, p 1085
108. M. Fahrmann, W. Hermann, E. Fahrmann, A. Boegli, T.M. Pollock, and H.G. Sockel, *Mater. Sci. Eng., Struct. Mater. Prop. Microstruct. Proc.,* Vol 260, 1999, p 212
109. N. Matan, D.C. Cox, P. Carter, M.A. Rist, C.M.F. Rae, and R.C. Reed, *Acta Mater.,* Vol 47, 1999, p 1549

Modeling of Microstructure Evolution during Solidification Processing*

Ch.-A. Gandin, Ecole Nationale Superieure des Mines de Paris
I. Steinbach, Ruhr-University Bochum

Introduction

Modeling of structure formation in casting of alloys involves several length scales, ranging from the atomic level (10^{-10}–10^{-9} m) to the casting dimensions, i.e. the macroscopic scale, (10^{-2}–1 m). Intermediate length scales are used to define the microstructure of the growing phases (i.e., 10^{-7}–10^{-4} m, e.g., dendritic or eutectic patterns) and the grain structure (i.e., 10^{-4}–10^{-2} m, e.g., equiaxed and columnar grains). This article concentrates on these intermediate length scales, where transport phenomena govern the spatial and temporal evolution of the structure. To calculate this evolution, conservation equations are written for each individual phase (liquid or solid), with the most challenging aspect being the treatment of the interfaces between phases and the nonequilibrium condition of the liquid metal during the solidification state.

The microstructure of a casting can consist of different grains of the same thermodynamic phase, distinct by orientation, or different phases of one alloy, distinct by composition and atomic structure. Within one grain metallic alloys also show a microstructure caused by segregation patterns due to the dendritic growth instability of the growth front. Figure 1(a) presents a longitudinal cross section of a 0.17 m (7 in.) Al-7wt%Si cylindrical ingot. The grain structure is revealed by the various gray levels. Elongated columnar grains are present at the bottom of the casting, while a more isotropic grain size is observed at the top. Only magnification of the grains toward the top of the casting reveals the dendritic nature of the equiaxed microstructure made of the α-aluminum phase (Fig. 1b). Further magnification would reveal the lamellas of the eutectic microstructure, where silicon is solidified as a second thermodynamic phase together with the α-aluminum phase. It is located between dendrite arms and formed from the interdendritic liquid at the end of solidification.

As details of the microstructure cannot be resolved at the macroscopic scale of the casting, models have been developed to determine an average grain size of the equiaxed grains and the extent of a columnar zone (Ref 1, 2). These models are based on averaging procedures to write conservation equations for a two-phase mixture made of solid and liquid. These equations are solved at the scale of the casting using a control volume (CV) or finite element (FE) method. The results of such models are average grain sizes, and possibly average dendrite arm spacings, within the casting. It consequently does not directly provide a map of the structure. The typical representative elementary volume (REV) of such macroscopic scale method is shown in Fig. 1. It contains several grains or phases that are not resolved on the macroscopic scale.

The cellular automaton (CA) method was first developed to model the macrostructure in casting by resolving each individual grain. The grains are defined by envelopes, drawn around a dendritic structure, and the entire grain structure of a casting can be computed. This is possible by considering a smaller REV, whose typical size is shown by the large highlighted square in Fig. 2(a) and its magnified view in Fig. 2(b). As a consequence, the dendritic structure cannot be resolved and is approximated by microscopic models. Special algorithms have to be designed to integrate the evolution of the volume fraction of the phases present in the REV over time, as well as to integrate the kinetics of the development of envelopes defining the grains.

An even smaller REV must be used to compute the detailed dendritic or eutectic microstructure within the grains. The typical REV of such a microscopic scale method is shown by small squares in Fig. 2(b). The phase field (PF) method is now very popular (Ref 9–15) because of its profound thermodynamic basis. Although only small portions of the grain structure can be computed, the PF method offers the possibility of simulating directly the

development of the solid/liquid interface, including its kinetics, pattern formation, and microsegregation with a limited number of approximations.

While this article focuses on a presentation of the CA and PF methods that represent the state of the art for modeling macrostructure and microstructure, several other methods are under development. One could cite the level-set method (Ref 16) and the pseudo front tracking method (Ref 17) for modeling interface development, as well as other forms of the front tracking method to simulate the growth of the boundary defined between a mushy zone and the melt (Ref 18).

Before examining the technical details, it is worth noting that all models for microstructure formation on different scales rely on an indicator function that gives the phase state at a specific point in space and time. On the macroscopic scale it is usually the fraction of solid (f_s) that has the value 1 in solid, 0 in liquid and values $0 < f_s < 1$ in the two-phase region, the so-called mushy region. This indicator function can be mapped to the local temperature only under conditions close to equilibrium solidification. In most cases, however, it will be an independent variable, the time evolution of which depends on the nucleation and microscopic growth conditions (see, e.g., Ref 19). In the CA model, a discrete indicator function I_v^S is used to distinguish between the mushy zone and the liquid (for details see below) that characterize the state of one REV, that is, one cell. A changing value of the indicator function between two adjacent REVs defines the boundary between the solidifying mush and liquid. Applying a kinetic model for the motion of this boundary then determines the growth of the grain structure. The liquid outside a grain envelope has to stay in an undercooled state until the local REV is captured by a growing envelope, or an appropriate nucleation condition is fulfilled. The PF method, finally, uses the PF

(b)

(a)
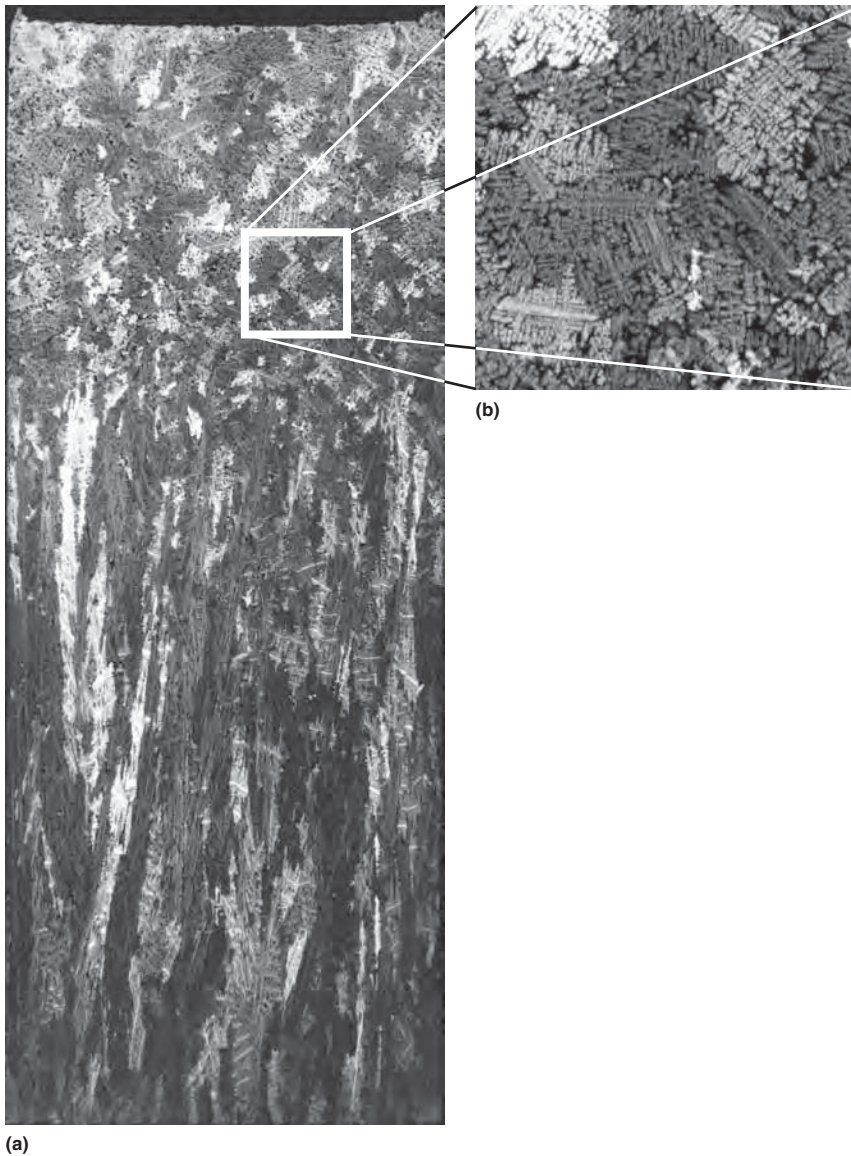
Fig. 1 Ingot structure (a) as observed in a longitudinal cross section of a 0.17 m (7 in.) height Al-7 wt%Si cylindrical casting. Distinction is possible between elongated columnar grains located in the bottom two-thirds of the casting and the isotropic equiaxed grains in the top region. Magnification in (b) reveals the dendritic microstructure of the alloy, with an interdendritic black region corresponding to a eutectic microstructure that is not resolved at this scale. The size of the highlighted square would be typical of the representative elementary volume (REV) used to solve conservation equations at the scale of the casting using a finite element method. Only one REV is represented while the number of REV is defined to fully cover the domain of the casting to be modeled.

variable $\Phi(x,t)$ as an analogue to f_s, where the intermediate values $0 < \Phi < 1$ characterize the interface between solid and liquid. Again the phase state changes only at the position of the interface if no nucleation condition is defined in the bulk liquid or solid. Variation of the PF variable through the interface reflects the nature of a diffuse interface at the atomistic scale. Using these indicator functions, all models are able, with a different level of sophistication, to account for nonequilibrium phenomena that characterize a typical solidification process and are reflected in the complex rules of microstructure formation.

This article starts from the microscopic scale with principles and applications of the PF method. Then, using the mesoscopic scale of individual grains, the CA model is introduced as a computationally efficient method to predict grain structures in castings. Finally, the coupling of the CA to macroscopic calculation of heat, flow, and mass transfers in castings and applications to realistic casting conditions is discussed.

Direct Microstructure Simulation Using the Phase Field Method

A phase field $\Phi_\alpha(x)$ is an indicator function of the real space coordinate $\boldsymbol{x}$, indicating that the material at this coordinate exists in the state of the phase α if $\Phi_\alpha(\boldsymbol{x}) = 1$, or in some other (not α) phase if $\Phi_\alpha(\boldsymbol{x}) = 0$.* The transition between one phase and another is assumed steep, but continuous. The transition region thus characterizes the interface between the phases that is "diffuse" instead of a sharp jump between the phases. During a phase transformation this interface moves by converting the thermodynamically unstable phase into a more stable phase; the description of this interface motion is the main goal of the method. The idea of a diffuse interface dates back to van der Waals (Ref 20), who intensively thought about the forces between atoms and molecules. It is well accepted today that the solid/liquid interface in a metallic system can be described as a diffuse transition on a width of several atomic distances. In numerical calculations using the PF method, however, the width of the interface η is mostly taken to be large compared to its true physical dimension, but small compared to the scale of the microstructure r that has to be resolved. The justification to do so lies in the invariance or the PF equation (for constant outer fields and to the first order in η/r) against rescaling of the interface (Ref 11). In the case of coupling the PF to diffusion fields, the effect of nonconstant diffusion fields inside the interface on the kinetics of the transformation and the redistribution of solute has to be considered. A rigorous way to treat this problem was developed by Karma and coworkers (Ref 9, 10) for binary alloys in the dilute solution limit. For the general case of multicomponent multiphase alloys this correction scheme still deserves to be adapted. The general aspects of the PF method are only briefly outlined. For a more complete review see (Ref 21).

In general, a PF model starts from an appropriate thermodynamic function. For solidification under constant pressure, the Gibbs free energy F is used. It is the integral over the free-energy density contributed by the interface f^{IF} and by the bulk thermodynamics, f^{TD}:

$$F = \int f dx = \int dx (f^{\text{IF}} + f^{\text{TD}}) \qquad \text{(Eq 1)}$$

In the solid state an elastic contribution will also be important (Ref 12) as well as magnetic or electric contributions for certain systems. From the principle of energy minimization, the evolution equation of the PF and the concentration are derived:

$$\dot{\Phi}_\alpha = -\sum_\beta \mu_{\alpha\beta} \left(\frac{\delta F}{\delta \Phi_\alpha} - \frac{\delta F}{\delta \Phi_\beta} \right) \qquad \text{(Eq 2a)}$$

$$\dot{c}^i = \sum_j \nabla M^{ij} \nabla \frac{\delta F}{\delta c^j} \qquad \text{(Eq 2b)}$$

where $\mu_{\alpha\beta}$ is the mobility of the interface between phase α and phase β, c^i is the concentration of component i and M^{ij} the chemical mobility matrix. The different PF models differ

---

*In the context of thermodynamics the expression "phase field" is also used to characterize a region of the composition space of an alloy, where one specific phase is stable. The concept therein is complementary; however, the approach to determine a "phase field" is different.

(a)

(b)
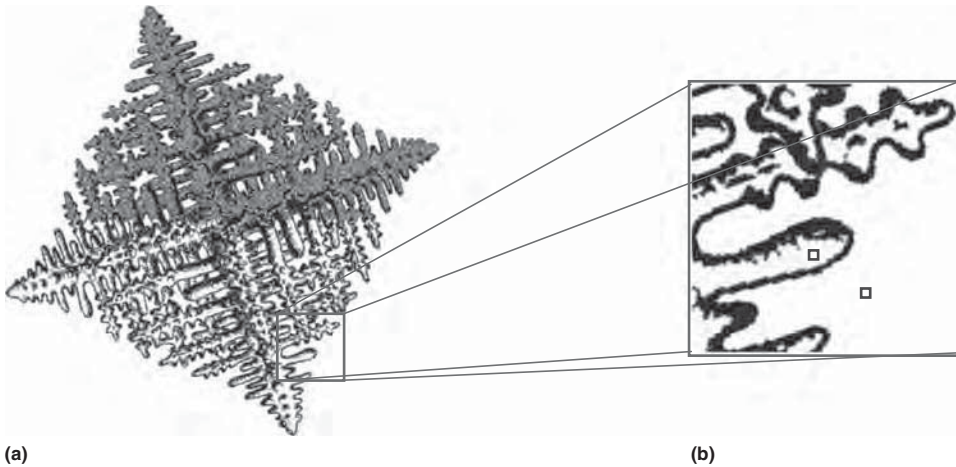
Fig. 2 Schematic of a single grain (Ref 8) growing in a uniform temperature field is shown in (a). The square highlighted in (a) shows a typical length scale for representative elementary volumes (REV) used by the cellular automaton (CA) method. The two small squares in (b) show the typical scale for REV used by the phase field (PF) method. Only one CA REV is shown although the number of CA REV is defined to fully cover the domain of the casting to be modeled. Similarly, only two PF REV are shown while the number of PF REV is defined to cover a domain of the casting representative of the microstructure to be modeled.

in the particular formulation, but agree in general principle. The following formulation given in (Ref 13) is used to find the particular form:

$$\dot{\Phi}_\alpha = \sum_\beta \mu_{\alpha\beta} \left\{ \sigma_{\alpha\beta} \left[\Phi_\beta \nabla^2 \Phi_\alpha - \Phi_\alpha \nabla^2 \Phi_\beta \right. \right.$$
$$\left. \left. - \frac{\pi^2}{2\eta_{\alpha\beta}^2} (\Phi_\alpha - \Phi_\beta) \right] + \frac{\pi}{\eta_{\alpha\beta}} \sqrt{\Phi_\alpha \Phi_\beta} \Delta G_{\alpha\beta} \right\} \quad \text{(Eq 3a)}$$

$$\Delta G_{\alpha\beta} = f_\alpha - f_\beta - \mu^i \left(c_\alpha^i - c_\beta^i \right) \quad \text{(Eq 3b)}$$

$$\dot{c}^i = \sum_j \nabla D^{ij} \left[\nabla c^j - \left(\sum_\alpha c_\alpha^j \nabla \Phi_\alpha \right) \right] \quad \text{(Eq 3c)}$$

where $\sigma_{\alpha\beta}$ is the interfacial energy, $\eta_{\alpha\beta}$ the interface width, $\Delta G_{\alpha\beta}$ the free-energy difference, and μ^i the generalized chemical potential between phase α and β. The specific free-energy density, f_α, of phase α related to the concentration c_α^i in phase α. The diffusivity matrix, D^{ij}, can be calculated from the chemical mobilities of the individual components and the thermodynamic factor $\partial^2 f / \partial c_\alpha^i \partial c_\alpha^i$. In the interface it is dependent on the properties of both phases that build the interface. For more details and a sound derivation see Ref 13.

The term in square brackets in the PF (Eq 3a) penalizes high curvature of the interface while $\Delta G_{\alpha\beta}$ drives the interface in direction of the thermodynamic stable phase. In the diffusion equation, the contribution proportional to the gradient of the PF variable accounts for solute redistribution at the interface. These equations are usually solved with standard finite element or control volume algorithms in the entire calculation domain with the advantage that no explicit tracking of the interface is necessary. As the thermodynamic driving force $\Delta G_{\alpha\beta}$ in the PF equation depends on the local concentration and a moving interface redistributes solute via the $\Delta\Phi$ term, both Eq 3a and 3b are closely coupled. The entire set of equations represents

the interaction between growth and diffusion and interface curvature that naturally result in the formation of dendritic microstructure during metal solidification. Before giving examples, it is useful to examine two further issues related to solidification in casting processes: multicomponent alloy thermodynamics and nucleation.

Multicomponent Alloy Thermodynamics

During solidification any material separates into two (or more) different phases: liquid and solid (or different solid phases). The clear reason is the minimization of the total Gibbs energy as the sum of the Gibbs energies of the individual phases weighted by the phase fractions. Neglecting surface energy, the PF and concentration equations described previously will tend to the thermodynamic equilibrium as the minimum of Gibbs energies. It is therefore natural to combine a computational method for equilibrium calculation, the so-called CALPHAD method (Ref 22), and the PF method. In CALPHAD databases the Gibbs energies of the individual bulk phases f_α are tabulated as functions of the alloy composition c_α^i. The direct coupling between PF simulation and CALPHAD calculation was first demonstrated in Ref 23 and is described in more detail in Ref 13 and 24. For alternative approaches see Ref 25, 26. It must be stated clearly that only a reliable description of the alloy thermodynamics of the bulk phases enables a realistic PF simulation in technical alloys.

Nucleation

Nucleation is not included in the general framework of PF theory. Some authors (Ref 12, 27) propose models with strong thermal noise to introduce random nucleation in a PF simulation.

However, it is well known that an interface in a well-resolved PF simulation needs a minimum of four to five numerical cells; thus the size of a nucleus in an inoculated casting, the order of which is lower than a micrometer, cannot be resolved in a calculation of solidification structures of the order of several tens of micrometers. To overcome this problem, a deterministic seed density model is incorporated, where a distribution of heterogeneous seed particles is distributed over the calculation domain without resolving them explicitly. A seed is activated, if the local supercooling of a solid phase in the liquid matrix exceeds the supercooling needed for a solid particle to grow when its size corresponds to the size of the seed particle (Ref 28). The growth of the nucleus is handled analytically, and the solid fraction is accumulated locally until its size can be treated by the PF (Ref 24).

Equiaxed Solidification of a Hypereutectic AlCuSiMg Alloy

During commercial alloy solidification, a number of different phases nucleate. Nucleation and growth are closely correlated, and the whole process is governed by the competition between external heat extraction and release of latent heat. The latter may be considered in an average over the calculation domain, as the heat diffusion is typically 3 orders of magnitude higher than the solute diffusion in the liquid (Ref 24). Figure 3 shows four stages of solidification of the commercial piston alloy KS1295, where the main components Al, Cu, Si, and Mg have been considered. The alloy composition is hypereutectic with respect to the aluminum-silicon eutectic composition. Cooling was simulated by assuming a constant heat-extraction rate of 15 J/cm³, deduced from a macroscopic casting simulation of the piston, and considering the release of latent heat during solidification (Ref 21). First the primary silicon particles nucleate from the melt (Fig. 3a) and grow by depletion of silicon from the melt. A second effect of nucleating the silicon particles first is that the melt is depleted of impurities that would serve as nuclei for the α-aluminum phase. Also it is unlikely that α-aluminum will nucleate on the silicon particles, as the crystal lattices of both phases are quite different. So, the silicon particles continue to grow and the melt is depleted of silicon far below the eutectic composition.

In the calculation shown in Fig. 3(a), the α-aluminum phase is nucleated in the corners of the simulation box that, in this calculation, represents the area of one grain. Next, a rapid dendritic growth of the aluminum phase is observed, while nearly no eutectic aluminum-silicon is formed at this stage. The solidification path is depicted schematically in Fig. 4, which deviates from the equilibrium path as a result of the suppression of nucleation of the α-aluminum phase at the eutectic point. In the melt, during the growth of the primary silicon particles, the silicon content falls
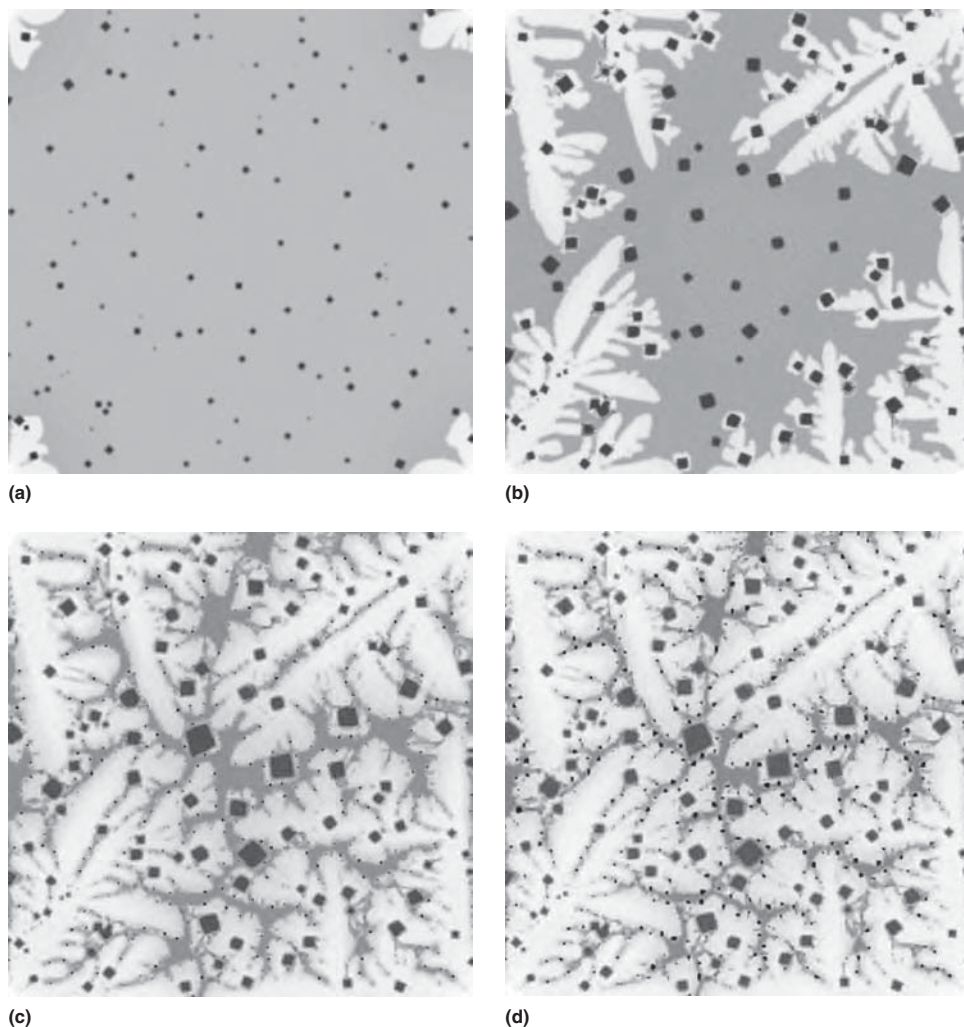
Fig. 3 Simulated solidification structure in a hypereutectic AlSiCuMg alloy in different stages. (a) Primary silicon particles and α-Al dendrite nucleated in the corners. (b) Growth of the silicon particles and the α-Al dendrite with engulfment of particles and tip splitting. (c) Precipitation of copper silicide particles. (d) Solidification structure before eutectic termination. The calculation domain is 800 by 800 μm in a two-dimensional approximation. Calculated by MICRESS

LIVE GRAPH
Click here to view
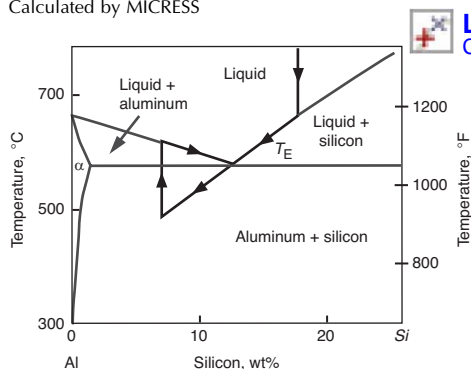
Fig. 4 Solidification path in hypereutectic Al-Si alloy

significantly below the eutectic composition. Then, after nucleation of the α-aluminum phase, recalescence is observed and the concentration close to the α-aluminum phase returns to the equilibrium line. The silicon particles, on the other hand, profit from the rejection of silicon from the α-aluminum phase and continue to grow until they are engulfed by the α-dendrite.

Also, tip splitting is observed (Fig. 3b). Figure 3 (c) shows the stage after nucleation of the copper silicides in the interdendritic region, and Fig. 3 (d) shows the last stage of the simulation shortly before termination of the solidification in the eutectic point with secondary silicon nucleating. This terminating eutectic phase, however, is much finer than the primary phases and therefore cannot be resolved in the present simulation.

Microsegregation in Steel

As a second example, consider the solidification of a low-alloyed steel Fe-0.8%Mn-0.7%Si-0.03%P-0.4%C. The emphasis of this simulation is the microsegregation profile of phosphorus after complete solidification. Phosphorus in steel today can mostly be eliminated through proper purification. However, if there is residual phosphorus left, it is known to embrittle

the material and local accumulation should be avoided. Because of the chemical interaction between silicon, carbon, and phosphorus, the full diffusion matrix including cross terms has to be taken into account. The thermodynamic and diffusion data rely on the TCFE3 and MOB2 databases by Thermo-Calc Software. The calculation starts from a single seed in the corner of the calculation domain. Secondary arms of the dendritic microstructure form with a mean spacing of about 100 μm, resulting from a high cooling rate of 0.2 K/s (0.4 °F/s).

Figure 5 shows the microsegregation profile of the alloying elements before complete solidification. The small bar in the figure shows the position of a virtual energy-dispersive x-ray spectroscopy (EDX) scan between two side branches where the relative compositions are shown for different times: immediately after the last interdendritic liquid has solidified locally, 50 s later, and 40 min later but at a temperature still above γ-α transformation. All alloying elements segregate to the melt during solidification, where phosphorous shows the strongest relative segregation, while manganese and silicon are nearly indistinguishable (Fig. 6a). Directly after the last liquid has solidified at the location of the virtual EDX scan, carbon is forced to diffuse out of the interdendritic region by the cross interaction with silicon (Fig. 6b). Phosphorus follows on a longer time-scale (Fig. 6c), showing an inverse segregation in the solid compared to the primary segregation during solidification, while silicon and manganese are nearly immobile in the solid. In the combined PF calculation of the solidification structure, solute partitioning, and chemical diffusion, a realistic picture of the distribution of alloying elements in a casting can be drawn.

Direct Grain Structure Simulation Using the Cellular Automaton Method

A cellular automaton (CA) is a theoretical concept to simulate the evolution of a complex structure by cooperative action of autonomous cells. Each cell acts according to the same rules, but is dependent on the state of the neighboring cells and the local evolutions of average quantities. The rules come from microscopic descriptions developed for the evolution of a mushy zone (e.g., microsegregation) and can be taken either from a PF simulation or from analytical models (see below). For a PF simulation, a typical domain would be the CA REV cell (highlighted area in Fig. 2a shown expanded in Fig. 2b), fully paved with PF REV (small squares shown within Fig. 2b). For a CA simulation, CA REV cells are overlaid onto the domain shown in Fig. 1(a). A PF simulation using a lattice defined to cover the domain shown in Fig. 1(a) is out of reach because of computer limitations. For that reason, coupling of structure formation with a macroscopic FE method solving mass, heat,
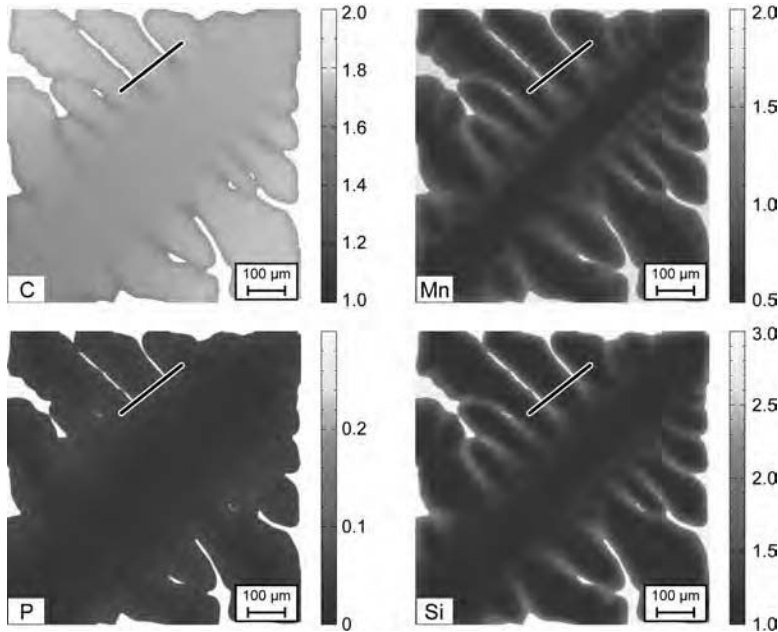
Fig. 5 Cross section of a dendritic microstructure in Fe-0.8%Mn-0.7%Si-0.03%P-0.4%C solidified at a cooling rate of 0.2 K/s (0.4 °F/s). Calculated by MICRESS

and fluid flows has been developed at the scale of the CA REV cells.

Figure 7 shows schematically the superimposition of the CA lattice on the same domain used for a macroscopic finite element (FE).* Each cell ν is uniquely defined by the coordinates of its center, $C_\nu = (x_\nu, y_\nu)$, located in a finite element mesh, F. Element F has N_n^F nodes, labeled n_i^F in Fig. 7, which are locally indexed from $i = 1$ to N_n^F. In this study, N_n^F is equal to 3 for the triangular elements used. For the purpose of exchanging information from the FE nodes to the CA cells, linear interpolation coefficients, $c_\nu^{n_i^F}$, are defined between each node n_i and a cell ν. A variable available at the nodes, ξ_{n_i}, can thus be used to calculate an interpolated value at cell ν, ξ_ν:

$$\xi_\nu = \sum_{i=1}^{N_n^F} c_\nu^{n_i^F} \xi_{n_i} \qquad \text{(Eq 4)}$$

Similarly, when information computed at the level of the cells of the CA grid, ξ_ν, is required at the level of a FE node, ξ_n, the following summation is performed:

$$\xi_n = \frac{1}{\Lambda_n} \sum_{i=1}^{N_n^n} c_{\nu_i}^n \xi_{\nu_i} \qquad \text{(Eq 5)}$$

where N_n^n is the number of cells ν seen by node n. For node n_i^F, it corresponds to all the cells drawn in Fig. 7. This averaging procedure is normalized by the summation over all interpolation coefficients:

$$\Lambda_n = \sum_{i=1}^{N_n^n} c_{\nu_i}^n \qquad \text{(Eq 6)}$$

Indices

The state index, I_ν^S, is used to characterize the phase and the neighborhood of a cell ν. It is defined as:

- Cell ν is liquid: $I_\nu^S = 0$.
- Cell ν is no longer liquid and at least one of its nearest neighboring cell, μ_i ($i \in [1, N^\nu]$), is still liquid ($I_{\mu_i}^S = 0$): $I_\nu^S = +1$.
- Cell ν is no longer liquid, and neither is its nearest neighboring cell, $\mu_i (I_{\mu_i}^S \neq 0, i \in [1, N^\nu])$: $I_\nu^S = -1$.

where N^ν is the number of cells defined in the neighborhood of cell ν. The Moore configuration considering the first and second nearest neighboring cells is used, yielding to $N^\nu = 8$.

The grain index, I_ν^G, is used to track the development of the mushy zone within cell ν. It is defined as:

- Cell ν is liquid ($I_\nu^S = 0$): $I_\nu^G = 0$.
- Cell ν is no longer liquid ($I_\nu^S \neq 0$), and the mushy zone is still developing: $I_\nu^G = +1$ with $g_\nu^m < 1$, g_ν^m being the volume fraction of the mushy zone in cell ν.
- Cell ν is no longer liquid ($I_\nu^S \neq 0$), and its mushy zone is fully developed: $I_\nu^G = -1$ for $g_\nu^m = 1$.

The CA model described hereafter is based on rules to modify the state index, I_ν^S, and grain index, I_ν^G, associated with each cell ν of

a square lattice with the goal to simulate the development of the mushy zone upon solidification.

Nucleation

It is assumed that nucleation of grains takes place on heterogeneous particles present in the melt as soon as a critical undercooling is reached. Because such nucleation behavior is only dependent on undercooling, it is considered as an instantaneous nucleation law. A Gaussian distribution is used to characterize the density of heterogeneous particles as a function of the critical undercooling at which they are activated (Ref 19).

In order to initiate the solidification process at the microscopic scale, nucleation sites are randomly distributed among the cells of the CA square lattice. A nucleation site contained in cell ν is characterized by a critical nucleation undercooling, ΔT_ν^{nucl}, and a random crystallographic orientation, θ_ν.

There are three tests performed to nucleate a new grain in a cell ν containing a nucleation site:

- The cell is still liquid: $I_\nu^S = 0$.
- The cell temperature, T_ν, falls below the local liquidus temperature: $T_\nu < T_L(w_\nu)$ with $T_L(w_\nu) = T_M + mw_\nu$, where T_M is the melting temperature of the solvent in the binary phase diagram, m is a linear approximation of the liquidus slope of the binary phase diagram defined with the solute element, and w_ν is the local composition of cell ν. Temperature T_ν is calculated from a conversion of the average enthalpy of cell ν, H_ν, considering a purely liquid cell ($H_\nu = C_p T_\nu + \Delta_s^l H_f$, C_p is the specific heat and $\Delta_s^l H_f$ is the latent heat of fusion).
- The temperature of cell ν, T_ν, deduced by interpolation from the nodes n_i^F defining element F in which cell ν is located (Eq 4), must fall below the critical temperature defined for nucleation, $T_\nu < T_L(w_\nu) - \Delta T_\nu^{nucl}$. This is of course only possible providing a nucleation site was previously ascribed to cell ν.

These tests are applied to all the cells containing a nucleation site. If the three tests are verified, the following actions are taken to initialize the growth of the grain:

- The state index of cell ν, I_ν^S, is updated following the definitions given previously. Consequently, the state index of all cells defined in the neighborhood of cell ν is also updated.
- The grain index of cell ν, I_ν^G, is updated following the definitions given previously.
- The center of the growing shape of the newly nucleated grain, G_ν, is located at the center of cell ν, C_ν. A square shape with main diagonals corresponding to the preferential <10> directions of the dendrites stems and arms are initialized with a very small size. Angle θ_ν defines the orientation of the [10] direction of the grain, $G_\nu S_\nu^{[10]}$, with respect to the **Ox** axis.

---

*It is to be noticed that in Fig. 7 the FE mesh is a triangle while it is a quadrangle in Fig. 1(a). However, both types of elements are possible when coupling with the CA method.

(a)   **(b)**   **(c)**
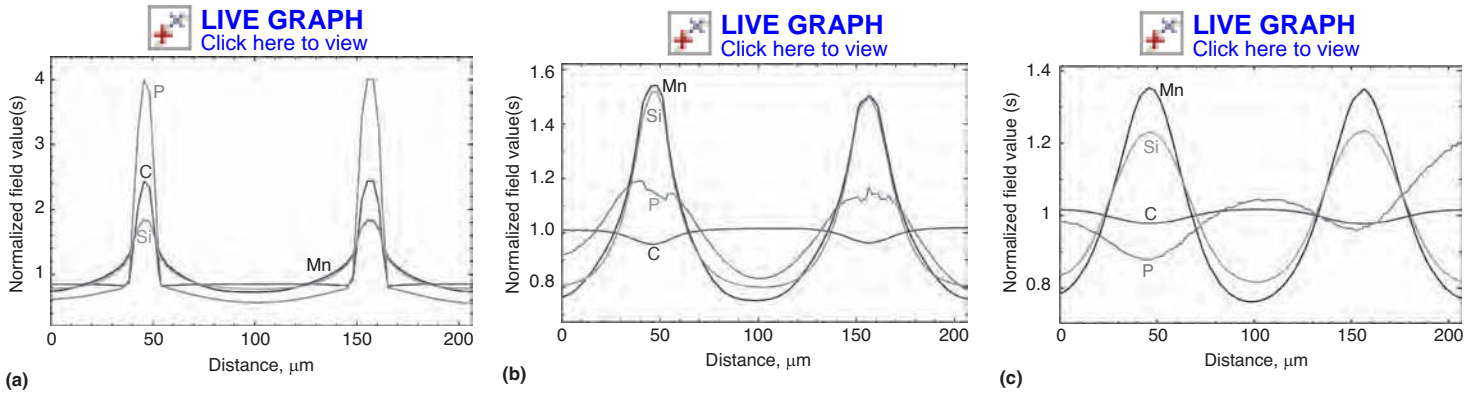
Fig. 6 Element distribution in relative units, normalized by the mean concentration, in the scan position indicated in Fig. 5. (a) Before local solidification. The elements are enriched in the interdendritic region. (b) Immediately after solidification. Phosphorus is still enriched in the position of last solidification, but carbon has already diffused away due to interaction with silicon. (c) After 40 min at a temperature of approximately 1000 °C (1830 °F). Silicon and manganese are almost immobile in the solid and prevail in the region of last solidification, while phosphorus also was forced to diffuse into the regions of first solidification, depleted from silicon. Calculated by MICRESS
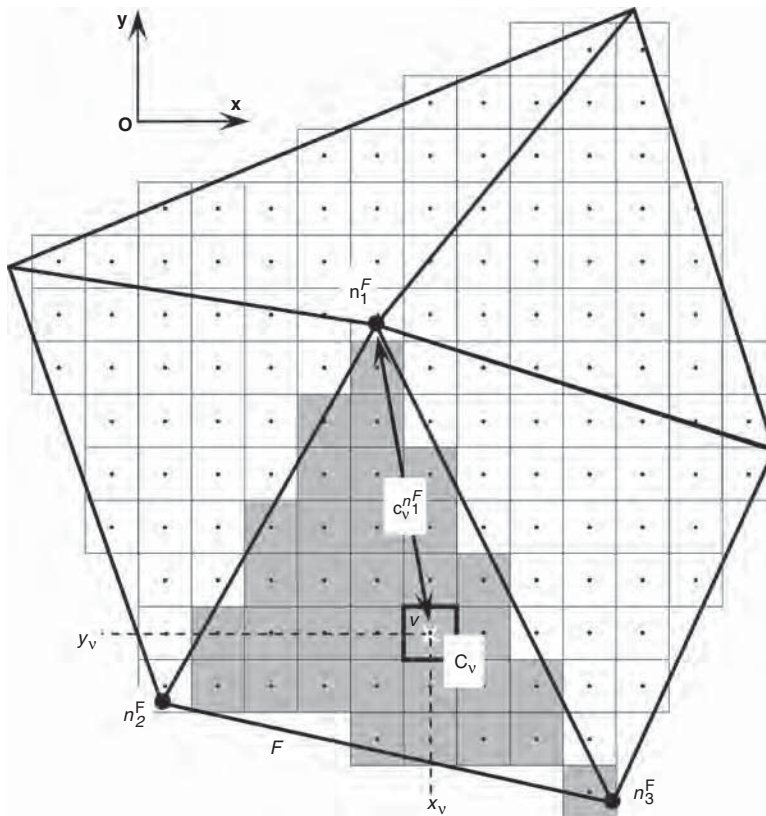
Fig. 7 Topological links between a triangular finite element mesh, F, and a square cell of the cellular automaton grid, ν. Interpolation coefficients, $c_\nu^{n_i^F}$, are defined between each cell and the N_n^F nodes of the finite element mesh, n_i^F, with $i = \left[1, N_n^F\right]$ and $N_n^F = 3$ in this figure.

Growth

Unlike with the PF method, the development of the solid/liquid interface is not directly modeled by the CA method. Instead, only the development of the grain envelope is simulated. Since the growth kinetics required for determining the velocity of the envelope of the grains is mainly based on diffusion in the liquid at a length scale given by the radius of curvature of the solid/liquid interface, it cannot be modeled with a CA model.* Similarly, while one

could also take advantage of the PF $\Phi_\alpha(x)$ indicator and compute the volume fraction of solid located within the CA REV. As stated before, coupling is yet not achievable between the PF and the CA methods. Thus, the CA method makes use of growth kinetics theories and microsegregation models to approximate growth morphology and kinetics of the grain envelopes, as well as its internal fraction of solid.

Figure 8 schematizes the growth propagation of the mushy zone from a cell ν to a cell μ. Within one time step, the four half-diagonals

associated to the shape of cell ν have grown from center G_ν. As a result, the mushy zone has extended from the polygon delimited by the white continuous thin lines to the polygon delimited by the black continuous thin lines and identified by the $S_\nu^{[ij]}$ tips, $[ij] \in [01, 10, \overline{1}0, 0\overline{1}]$. Cell μ was initially liquid ($I_\mu^S = 0, I_\mu^G = 0$), while the mushy zone has developed in cell ν ($I_\nu^S = +1, I_\nu^G = +1$). After the propagation of the mushy zone in cell ν, C_μ is now engulfed by the growing shape $S_\nu^{[ij]}$. Capture of cell μ is thus achieved by its neighboring cell ν. The initial growing shape associated to a cell μ, $S_\mu^{[ij]}$ (hashed area delimited by a continuous bold lines), as well as the position of the growth center of cell μ, G_μ, are then calculated. Two other growing shapes are schematized in Fig. 8 for cell μ. Tips $S_\mu^{[ij]\,min}$ (gray area delimited by dashed bold lines) represent the initial shape defined at a former time when edge $S_\nu^{[01]}S_\nu^{[10]}$ reached position C_μ, thus starting to engulf cell μ. Tips $S_\mu^{[ij]\,min}$ thus define the minimum size associated with the growth of cell μ. When no movement of the grain is accounted for, the center of this first growing shape, T_μ, is kept unchanged during the entire time of the simulation. If growth is maintained homothetic with respect to the velocity in all <10> directions, the maximal growing shape of cell μ can be evaluated when the mushy zone has propagated to all the neighboring cells. In Fig. 8, this maximal growing shape corresponds to tips $S_\mu^{[ij]\,max}$ defined when the center of the southeast neighbor of cell μ is reached by edge $S_\mu^{[0\overline{1}]}S_\mu^{[10]}$. All the neighboring cells of μ are then captured.

The current, minimal, and maximal growing shapes, respectively $S_\mu^{[ij]}$, $S_\mu^{[ij]\,min}$, and $S_\mu^{[ij]\,max}$, are used to define the volume fraction of the mushy zone, g_μ^m, associated to cell μ as:

---

*Some authors (Ref 29, 30) use a CA type of method to compute the development of microstructure at the scale of the solid/liquid interface. However, the size of the CA cells is then decreased to a size similar to that used for a PF simulation. Then realistic simulation of the development of the structure at the scale of the casting cannot be achieved.
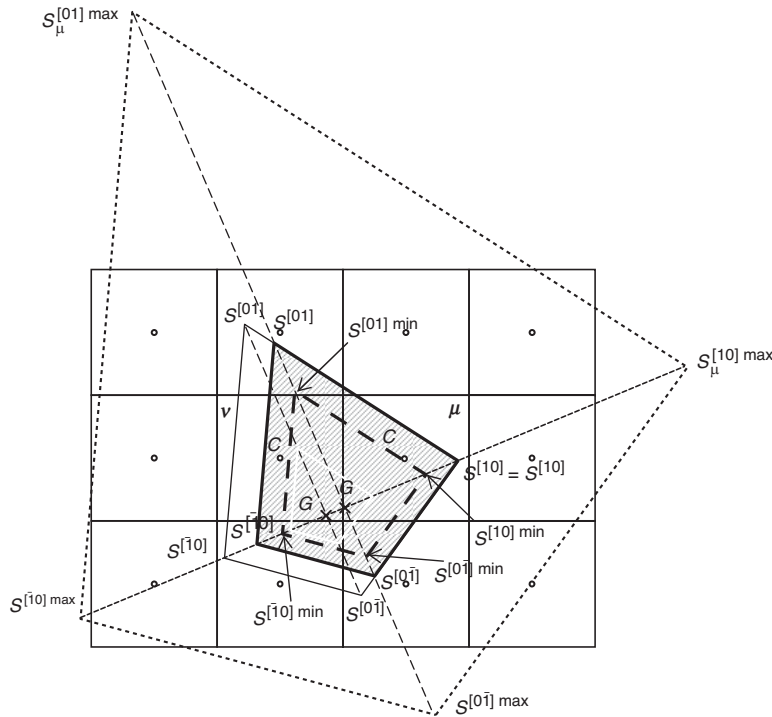
Fig. 8 The algorithm to model the growth of a cell ν identified as part of a grain is schematized, as well as the propagation of the grain to a neighboring cell μ. The growing shape associated with cell ν has propagated from the white to the black continuous thin lines to reach positions $S_\nu^{[ij]}, [ij] \in [01, 10, \overline{1}0, 0\overline{1}]$. At that time, the growing shape $S_\nu^{[ij]}$ has engulfed the center of cell μ, C_μ. The actual size of the growing shape associated to cell $\mu, S_\mu^{[ij]}$ (hashed area delimited by a continuous bold line), is calculated. The initial growing center and shape associated to cell μ, respectively G_μ and $S_\mu^{[ij] \, min}$ (grey area delimited by the dashed bold lines), as well as its maximal growing shape, $S_\mu^{[ij] \, max}$ (delimited by the dotted bold line), used to defined the volume fraction of the mushy zone associated to cell μ, g_μ^m, are also shown.

$$g_\mu^m = \text{Min}\left[\frac{A_\mu - A_\mu^{min}}{A_\mu^{max} - A_\mu^{min}}, 1 \right] \qquad \text{(Eq 7)}$$

where A_μ, A_μ^{min}, and A_μ^{max} are the areas associated with the current, initial, and maximal growing shapes, respectively. The volume fraction of solid associated to cell μ, g_μ^s, is then given by:

$$g_\mu^s = g_\mu^m g_\mu^{s \, m} \qquad \text{(Eq 8)}$$

where $g_\mu^{s \, m}$ is the volume fraction of solid located in the envelope $S_\mu^{[ij]}$. It is to be noted that the growing shape associated with a given cell ν is not a regular square as was the case in previous procedures (Ref 31, 32). Limitation to a simple square was based on the assumption that the temperature is uniform within the cell, and all four <10> directions could consequently grow at the same velocity. Since the grain envelope is sustained by a series of neighboring cells, the temperature gradient could be taken into account by the temperature difference between cells. Because of the effect of the fluid flow on the growth kinetics of the four <10> directions, each of the directions $\mathbf{G}_\nu \mathbf{S}_\nu^{[ij]}, [ij] \in [01, 10, \overline{1}0, 0\overline{1}]$, makes a different angle with the fluid-flow direction. Four different values of the growth velocities, $\nu_\nu^{[ij]}$, are

thus calculated. A similar extension was proposed by Takatani et al. for the study of grain texture formation in strip casting of thin steel sheets due to the effect of the relative fluid-flow velocity on grain structure formation (Ref 33). Details concerning the boundary layer correlation used to compute the supersaturation of the dendrite tips growing into an undercooled melt in the presence of fluid flow, and thus to deduce its growth velocity, are available in Ref 34.

The CA model has been extended to account for the transport of the grains due to fluid flow and sedimentation. This part of the model is not presented here. More information is available in Ref 35.

Coupling of Direct Structure Simulation at Macroscopic Scale

The finite element method is used to solve the average conservation equations written for the total mass, the energy, the mass of solute, and the momentum at the scale of the casting (Ref 36, 37). As outputs of a FE simulation, it is possible to predict the time evolution of the average enthalpy, the average velocity of the liquid phase (the solid being fixed), and the average solute composition. The average

enthalpy and solute composition are converted into evolutions of temperature, solute composition of the liquid, and volume fraction of solid. This can be carried out by applying a simple microsegregation model, such as the lever-rule approximation deduced from the phase diagram data, at the microscopic scale of the dendrite arm spacing. In such a purely macroscopic calculation, the presence of the mushy zone is thus a direct function of the average enthalpy and the local liquidus temperature. As soon as the enthalpy falls below the value of the enthalpy that corresponds to the local liquidus temperature, the solid starts to form and the volume fraction of solid consequently increases. The final solid is formed at the eutectic temperature. The method described previously does not consider coupling with direct simulation of structure development.

In order to account for the effect of the growth undercooling of the microstructure, as well as for the transport of equiaxed grains that can freely develop in the liquid, the mesoscopic cellular automaton (CA) method has been coupled with the macroscopic FE method (Ref 38). The output of the CA model is the grain structure. Such a coupling between the meso- and macroscopic methods makes it possible to take into account nonequilibrium solidification paths due to growth undercooling. The volume fraction of solid is not only a function of the average composition and enthalpy, but also a function of the presence of the mushy zone. The presence of the mushy zone itself becomes a function of the location of the grains. The latter depends on both the undercooling of the growing dendritic grain structure and its sedimentation and transport due to the fluid flow. The coupling with the movement of the free equiaxed grains requires modifying the drag force entering the Navier-Stokes equation. This force depends on the relative velocity of the liquid with respect to the solid (i.e., velocity of the equiaxed grains), as well as on the volume fraction of the equiaxed grains in the macroscopic REV. It is computed according to the model proposed by Wang et al. (Ref 39), and it is added as a source term in the formulation of the FE method.

In order to calculate the velocity of the grains that are free to move in the liquid, the model proposed by Ahuja et al. (Ref 40) has been reprogrammed. The trajectory of each individual grain is computed with a dedicated CA algorithm. The supersaturation entering the dendrite growth kinetics is computed using a correlation that accounts for the effect of the relative flow velocity with respect to the solid velocity as well as the orientation between the dendrite growth direction and the fluid-flow direction (Ref 34). It is worth noticing that the CAFE model presents similar objectives as the Eulerian model proposed by Wang and Beckermann (Ref 6). The main difference is due to the Lagrangian description used to track each individual grain with the CA method. Only the liquid velocity is calculated with the FE method.

Such a simplification requires the transport of average quantities due to the movement of the solid phase to be accounted for. This is achieved by the development of coupling scheme between the CA and FE method.

The experimental configuration proposed by Hebditch and Hunt (Ref 41, 42) is considered for the application of the CAFE model. It consists of a parallelepipedic cavity with dimensions 100 mm (4 in.) long, 60 mm (2 in.) high, and 13 mm (0.5 in.) thick. All faces are carefully insulated except for one of the smallest faces. The Pb-48wt%Sn binary alloy is cooled down and solidified by imposing a circulation of water in a copper chill that is maintained in contact with this face. As a consequence of this configuration, a two-dimensional Cartesian approximation of the transport phenomena can be made by neglecting the interactions of the fluid flow with the two largest faces of the mold (Ref 43). The FE mesh is refined in both the horizontal and vertical directions as shown in Fig. 9(M). The location of the measurements of the composition in the final as-cast ingot are also shown in Fig. 9(M). A detailed table of the values of the parameters is given in Ref 38.

Figure 9 presents the results of the simulations. Maps of grain structures and macrosegregation, as well as composition profiles are shown. Time evolution of the development of the grain structure shows that equiaxed grains accumulate in the bottom region of the ingot by sedimentation. The general trends of the

segregation map are the same as the one predicted by a purely FE calculation (Ref 36). However, variations known as mesosegregations are also predicted at a smaller length scale. These variations are caused by the development of equiaxed grains in preferred zones, leading to instability of the segregation field that further develops with the propagation of the mushy zone. It is interesting to observe the formation of channels in the top left side of Fig. 9(b). The mechanism that led to the formation of this channel is instability of the growth front. Equiaxed grains have moved and accumulated at a position close to the actual root of the channels, forming arms of fixed equiaxed grains. These arms served as barriers to the fluid flow that then developed preferentially on their sides. Consequently, a channel with enriched solute developed between the arms, similar to the mechanism leading to the formation of a freckle (channel segregation). The average composition profiles drawn in Fig. 9 (c) also show similar trends.

In the present configuration defined by the experiments performed by Hebditch and Hunt (Ref 41), comparison with a calculation that does not account for the development of the grain structure (purely macroscopic FE calculation) shows that macrosegregation is not directly influenced by the structural features. The effect of the transport of equiaxed grains on the macrosegregation is well demonstrated in the literature (Ref 44, 45), especially for larger ingot sizes. The limited size of the

present Pb-48wt%Sn ingot does not favor such an effect. It should also be said that the grain structure model only considers the formation of equiaxed dendritic grains: no globular grains with high inner volume fractions of solid are taken into account. Such a limitation could be detrimental for the application of the present model to the prediction of segregation induced by grain movement. Not only the morphology of the equiaxed grain structure plays a role on segregation, but also it is expected to considerably modify the fluid flow. Despite the fact that the latter effect is accounted for in the present model, the consequence on the overall fluid flow remains limited and macrosegregation still is explained mainly by the same thermosolutal considerations that lead to the segregation map computed with a purely macroscopic FE calculation.

There are fewer studies in the literature on the effect of the grain structure on mesosegregation. The fact that instability of the growth front leads to further instability of the segregation map is well established for the formation of channels that could give rise to freckle formation. However, the proposition that accumulation of transported equiaxed grains on the growth front could lead to the formation of channels and eventually freckles is new. Only in situ diagnostics by direct visualization of the formation of the grain structure could validate this prediction.

Since the grain structure is difficult to extract from Ref 41 and 42, and since the
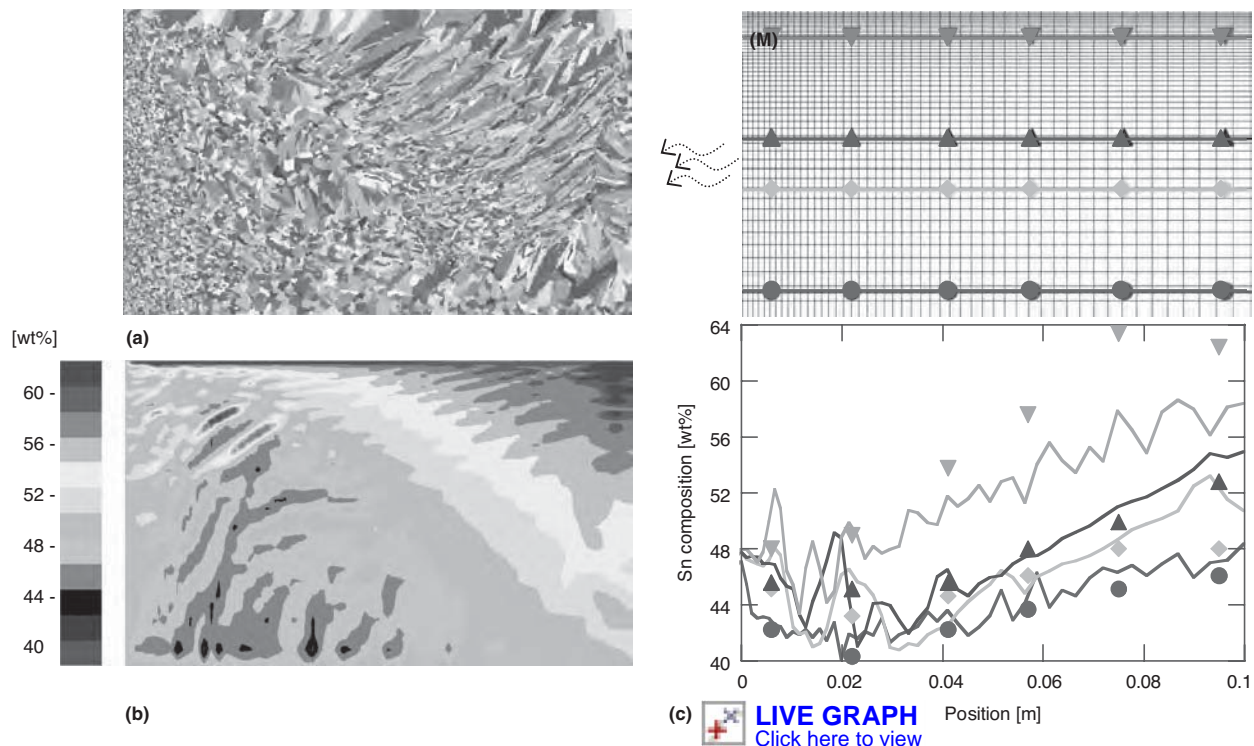
Fig. 9 Predictions from the cellular automaton finite element model. (a) Final grain structure. (b) Segregation map of tin with its composition scale. (c) Composition profiles for a Pb-48wt%Sn alloy. Equiaxed grains nucleated in the undercooled melt are free to move due to sedimentation and buoyancy-driven flows. The FE mesh (M) is drawn together with horizontal lines indicating the location of the profiles drawn in (c). The symbols in (M) indicate the location of the measurements (Ref 41). The same styles are used in (c) and (M) for curves and symbols.

measurements were carried on a very coarse grid and for a relatively large volume (4 mm, or 0.2 in., diam rods were extracted from the as-cast structure), it is not possible to directly use these experimental results to validate the predictions of the CAFE model. Similarly, no detailed measurement of the average composition has been conducted yet with the purpose of characterizing a segregation map with a definition smaller than the grain size. It is thus difficult to validate the present predictions of mesosegregation. Furthermore, the measurements are often carried out with samples of several cubic millimeters. Such measurements do not permit validation of the simulations, unless the predicted segregation maps are also drawn by averaging the composition over the same volume as the one used for the measurements. This procedure, however, has not been used here since it rubs out the effect of the grain structure.

Summary

Two models used to predict direct structure formation in casting have been presented. The development of solid/liquid interfaces or of grain envelopes can be tracked using a PF method or a CA method, respectively. While growth morphology, kinetics, and segregation are direct outputs of the PF method, its application remains limited to small portions of a casting. Computer limitations are the main problem and are not likely to be solved within the next few decades (Ref 46). For that reason, the CA method has been developed with the goal of providing a direct simulation of the structure in the entire casting while not considering the development of the solid/liquid interface. As a consequence of this approximation, only the development of grain envelopes is simulated, and microscopic models that have been developed in the literature for a long time are integrated. These microscopic models include kinetics models for the growth front (grain envelope or boundary between the mushy zone and the liquid outside the envelope) or microsegregation models. The advantage of this CA mesoscopic method is the possibility of studying the influence of the development of structure on quantities characterized at the scale of the casting (e.g., macrosegregation).

The development of quantitative models is very much related to the availability of experimental data. Such data are required to test the assumption of the models. Since the assumption applies at the scale of the structure, experimental data need to be collected in very well defined and characterized conditions. This is true for casting geometry, heat flow conditions, and alloy properties. Thus, the development of experimental facilities is still a need for further modeling activities.

Coupling of models and numerical methods could be the solution to improving the results for engineering processing of solidification structure. However, such coupling is difficult because of the interdisciplinary nature of the field. It is recognized that such challenging coupling is not simple.

REFERENCES

1. R. Trivedi and E. Sunseri, Non-Plane Front Solidification: Cellular and Dendritic Growth, and Columnar to Equiaxed Transition, *Casting*, S. Viswanathan, Ed., Vol 15, *ASM Handbook*, ASM International, 2008
2. S. Shankar, Eutectic Solidification, *Casting*, S. Viswanathan, Ed., Vol 15, *ASM Handbook*, ASM International, 2008
3. Peritectic Solidification, *Casting*, S. Viswanathan, Ed., Vol 15, *ASM Handbook*, ASM International, 2008
4. W. Kurz, Plane Front Solidification, *Casting*, S. Viswanathan, Ed., Vol 15, *ASM Handbook*, ASM International, 2008
5. M. Krane, V. Voller, and B. Li, Modeling of Transport Processes and Electromagnetics, *Casting*, S. Viswanathan, Ed., Vol 15, *ASM Handbook*, ASM International, 2008
6. C.Y. Wang and C. Beckermann, Equiaxed Dendritic Solidification with Convection: Part I: Multiscale/Multiphase Modeling, *Metall. Mater. Trans.*, Vol 27A, 1996, p 2754–2764
7. M.A. Martorano, C. Beckermann, and Ch.-A. Gandin, *Metall. Mater. Trans. A*, Vol 34, 2003, p 1657
8. W. Kurz and D.J. Fisher, *Fundamentals of Solidification*, 3rd ed., Trans Tech Publications, Switzerland, 1992
9. A. Karma, Phase-Field Formulation for Quantitative Modeling of Alloy Solidification, *Phys. Rev. Lett.*, Vol 87, 2001, p 115701–1–4
10. B. Echebarria, R. Folch, A. Karma, and M. Plapp, Quantitative Phase-Field Model for Alloy Solidification, *Phys. Rev. E*, Vol 70, 2004, p 061604–1–22
11. G. Caginalp and P. Fife, Phase-Field Methods for Interfacial Boundaries, *Phys. Rev. B*, Vol 33, 1986, p 7792–7794
12. L.Q. Chen, Phase-Field Models for Microstructure Evolution, *Ann. Rev. Mater. Res.*, Vol 32, 2002, p 113–140
13. J. Eiken, B. Böttger, and I. Steinbach, Multiphase Field Approach for Multicomponent Alloys with Extrapolation Scheme for Numerical Application, *Phys. Rev. E*, Vol 73, 2006, p 066122–1–9
14. A.A. Wheeler, W.J. Boettinger, and G.B. McFadden, Phase-Field Model for Isothermal Phase Transitions in Binary Alloys, *Phys. Rev. A*, Vol 45, 1993, p 7424–7439
15. I. Steinbach, F. Pezzolla, B. Nestler, M. Seeßelberg, R. Prieler, G.J. Schmitz, and J.L.L. Rezende, A Phase Field Concept for Multiphase Systems, *Phys. D*, Vol 94, 1996, p 135–147
16. L. Tan and N. Zabaras, A Level Set Simulation of Dendritic Solidification with Combined Features of Front Tracking and Fixed-Domain Methods, *J. Comput. Phys.*, Vol 211, 2006, p 36–63
17. A. Jacot and M. Rappaz, A Pseudo-Front Tracking Technique for the Modelling of Solidification Microstructures in Multi-Component Alloys, *Acta Mater.*, Vol 50, 2002, p 1909–1926
18. D.J Browne and J.D. Hunt, A Fixed Grid Front-Tracking Model of the Growth of a Columnar Front and an Equiaxed Grain During Solidification of an Alloy, *Numer. Heat Transfer, Part B: Fundamentals*, Vol 45, 2004, p 395–419
19. M. Rappaz, Modeling of Microstructure Formation in Solidification Processes, *Int. Mater. Rev.*, Vol 34, 1989, p 93–123
20. J.D. van der Waals, "The Thermodynamic Theory of Capillarity under the Hypothesis of a Continuous Variation of the Density," Amsterdam; English Trans. J.S. Robinson, *J. Stat. Phys.*, 1979
21. W.J. Boettinger, J.A. Warren, and C. Beckermann, Phase-Field Simulation of Solidification, *Ann. Rev. Mater. Res.*, Vol 32, 2002, p 163–194
22. L. Kaufman and H. Bernstein, *Computer Calculation of Phase Diagrams with Specific Reference to Refractory Materials*, Academic Press, 1970
23. U. Grafe, B. Böttger, J. Tiaden, and S.G. Fries, Coupling of Multicomponent Thermodynamic Databases to a PF Model: Application to Solidification and Solid State Transformations of Superalloys, *Scr. Mater.*, Vol 42, 2000, p 1179–1186
24. B. Böttger, J. Eiken, and I. Steinbach, Phase-Field Simulation of Equiaxed Solidification in Technical Alloys, *Acta Mater.*, Vol 54, 2006, p 2697–2704
25. R.S. Qin and E.R. Wallach, A Phase-Field Model Coupled with a Thermodynamic Database, *Acta Mater.*, Vol 51, 2003, p 6199–6210
26. H. Kobayashi, M. Ode, S.G. Kim, W.T. Kim, and T. Suzuki, Phase-Field Model for Solidification of Ternary Alloys Coupled with Thermodynamic Database, *Scr. Mater.*, Vol 48, 2004, p 689–694
27. L. Granacy, T. Pustai, T. Borzsonyi, J.A. Warren, and J.F. Douglas, A General Mechanism of Polycrystalline Growth, *Nature Mater.*, Vol 3, 2004, p 645–650
28. A.L. Greer, P.S. Cooper, M.W. Meredith, W. Schneider, P. Schumacher, J.A. Spittle, and A. Tronche Grain Refinement of Aluminium Alloys by Inoculation, *Adv. Eng. Mater.*, Vol 5, 2003, p 81–91
29. V. Pavlyk and U. Dilthey, Simulation of Weld Solidification Microstructure and Its Coupling to the Macroscopic Heat and Fluid Flow Modelling, *Mod. Sim. Mater. Sci. Eng.*, Vol 12, 2004, p 33–45
30. L. Beltran-Sanchez and D.M. Stefanescu, Growth of Solutal Dendrites: A Cellular Automaton Model and Its Quantitative Capabilities, *Mater. Trans. A*, Vol 34, 2003, p 367–382

31. Ch.-A. Gandin, R.J. Schaefer, and M. Rappaz, Analytical and Numerical Predictions of Dendritic Grain Envelopes, *Acta Mater.*, Vol 44, 1996, p 3339–3347

32. Ch.-A. Gandin and M. Rappaz, A 3D Cellular Automaton Algorithm for the Prediction of Dendritic Grain Growth, *Acta Mater.*, Vol 45, 1997, p 2187–2195

33. H. Takatani, Ch.-A. Gandin, and M. Rappaz, EBSD Characterisation and Modelling of Columnar Dendritic Grains Growing in the Presence of Fluid Flow, *Acta Mater.*, Vol 48, 2000, p 675–688

34. Ch.-A. Gandin, G. Guillemot, B. Appolaire, and N.T. Niane, Boundary Layer Correlation for Dendrite Tip Growth with Fluid Flow, *Mater. Sci. Eng.*, Vol A342, 2003, p 44–50

35. G. Guillemot, Ch.-A. Gandin, and H. Combeau, Modeling of Macrosegregation and Solidification Grain Structures with a Coupled Cellular Automaton-Finite Element Model, *ISIJ Int.*, Vol 46, 2006, p 880–895

36. N. Ahmad, H. Combeau, J.-L. Desbiolles, T. Jalanti, G. Lesoult, J. Rappaz, M. Rappaz, and C. Stomp, Numerical Simulation of Macrosegregation: A Comparison between Finite Volume Method and Finite Element Method Predictions and a Confrontation with Experiments, *Metall. Mater. Trans.*, 1998, Vol 29A, p 617–630

37. M. Bellet, V.D. Fachinotti, S. Gouttebroze, W. Liu, and H. Combeau, A 3D-FEM Model Solving Thermomechanics and Macrosegregation in Binary Alloys Solidification, in *Solidification Processes and Microstructures—A Symposium in Honor of Wilfried Kurz*, M. Rappaz, C. Beckermann, and R. Trivedi, Ed., The Minerals, Metals and Materials Society, Warrendale, 2004, p 41–46

38. G. Guillemot, Ch.-A. Gandin, and H. Combeau, Modeling of Macrosegregation and Solidification Grain Structures with a Coupled Cellular Automaton—Finite Element Method, *Solidification Processes and Microstructures: A Symposium in Honor of Prof. W. Kurz*, PA, 2004, p 157–163

39. C.Y. Wang, A. Ahuja, C. Beckermann, and H.C. de Groh III, Multiparticle Interfacial Drag in Equiaxed Solidification, *Metall. Mater. Trans.*, Vol 26B, 1995, p 111–119

40. C.Y. Wang, A. Ahuja, C. Beckermann, R. Zakhem, P.D. Weidman, and H.C. de Groh III, Drag Coefficient of an Equiaxed Dendrite Settling in an Infinite Medium, in *Micro/Macro Scale Phenomena in Solidification*, HTD-218, AMD-139, ASME, 1992, p 85–91

41. D.J. Hebditch and J.D. Hunt, Observations of Ingot Macrosegregation on Model Systems, *Metall. Trans.* Vol 5, 1974, p 1557–1564

42. D.J. Hebditch, "Segregation in Castings," Ph.D. Thesis, Oxford University, UK, 1973

43. J.-L. Desbiolles, Ph. Thévoz, and M. Rappaz, Micro-/Macrosegregation Modeling in Casting: A Fully Coupled 3D Model, in *Modeling of Casting, Welding and Advanced Solidification Processes X*, D. M. Stefanescu, J. Warren, M. Jolly, and M. Krane, Ed., The Minerals, Metals and Materials Society, Warrendale, PA, 2003, p 245–252

44. C. Beckermann, Modelling of Macrosegregation: Applications and Future Needs, *Int. Mater. Rev.*, Vol 47, 2002, p 243–261

45. G. Lesoult, V. Albert, H. Combeau, D. Daloz, A. Joly, C. Stomp, G.U. Grün, and P. Jarry, Equiaxed Growth and Related Segregations in Cast Metallic Alloys, *Conference IUMRS-ICAM'99*, International Union of Materials Research Societies 1999, Beijing, China

46. V.R. Voller and F. Porte-Agel, Moore's Law and Numerical Modeling, *J. Comput. Phys.*, Vol 179, 2002, p 698–703

Fundamentals of the Modeling of Damage Evolution and Defect Generation

Modeling and Simulation of Cavitation during Hot Working

P.D. Nicolaou, UES Inc.
A.K. Ghosh, University of Michigan
S.L. Semiatin, Air Force Research Laboratory

MANY METALLIC MATERIALS develop internal cavities when subjected to large uniaxial or multiaxial tensile strains at elevated temperatures. These materials include conventional alloys of aluminum, titanium, copper, lead, and iron as well as emerging intermetallic materials such as titanium aluminide alloys (Ref 1–3). For a given material with a given microstructure, the extent of cavitation depends on the specific deformation conditions (e.g., strain rate, temperature, and stress state). In most cases, cavitation may lead to premature failure at levels of deformation far less than those at which flow-localization-controlled failure would occur. Cavitation is a very important phenomenon in hot working of materials because it may yield inferior properties in the final part, let alone lead to premature failure during forming.

Cavity formation usually comprises three distinct but simultaneously occurring stages, that is, nucleation, growth, and coalescence. An important requirement for cavitation during deformation under either conventional hot working (high-strain-rate) conditions or superplastic forming is the presence of a tensile stress. Under conditions of homogeneous compression, on the other hand, cavitation is typically not observed. In fact, cavities produced during tensile flow may often be closed during subsequent compressive flow. Similarly, it has been demonstrated that the superposition of a hydrostatic pressure during hot forming can reduce or even eliminate cavitation (Ref 4).

This article deals with the modeling and simulation of cavitation phenomena. It is divided into six sections. Experimental observations of cavitation are briefly summarized first. The next two sections review the modeling of cavity nucleation and cavity growth; the former topic is treated more extensively in a companion article in this Volume. The discussion of cavity growth focuses on both mesoscale and microscale models under uniaxial versus multiaxial tensile-stress conditions; mesoscale models incorporate the influence of local microstructure and texture on cavitation. Descriptions of cavity coalescence and shrinkage are summarized in the following two sections. The last part of this article deals with the simulation of the tension test to predict tensile ductility and to construct failure-mechanism maps. Table 1 includes a list of symbols used in this article.

Cavitation Observations

Optical and scanning electron microscopy (SEM) are the usual techniques applied to quantify cavity formation within metallic materials. Specifically, samples are sectioned along one or more directions and are prepared for metallographic analysis. Low-magnification optical microscopy permits a gross examination of the sample, while SEM investigation enables the observation of cavities in greater detail. Optical micrographs with magnifications between 50 and 200× are typically analyzed using commercial software packages (e.g., NIH, Image J) to obtain measurements of several important features of cavities, for example, average radius, shape, and angle with respect to the principal directions. Further analysis of such measurements is used to determine the cavity area fraction, cavity volume, cavity density, and cavity shape (Ref 5, 6).

Table 1   List of symbols

a, size of major axis of an elliptical cavity	k_D, constant dependent on the geometry of the deformation
A_o, area of the nominally uniform slice of the specimen	L_i, cavity length along principal direction "i"
A_{lb}, load-bearing area	λ, cavity spacing
A_m, macroscopic area of a tension specimen	m, strain-rate sensitivity
b, size of minor axis of an elliptical cavity	M, Taylor factor
γ, surface energy of the cavity	M_h, Taylor factor of hard phase
γ_p, surface energy of a particle	M_s, Taylor factor of soft phase
γ_i, surface energy of the particle-matrix interface	n, stress exponent
C_v, cavity volume fraction	P, applied hydrostatic pressure
C_{v_o}, initial cavity volume fraction	r_c, critical cavity radius
D, diffusion coefficient	r, cavity radius
δ, grain-boundary thickness	r_o, radius at which a cavity becomes stable
δ_{ij}, Kronecker delta	R, average of the three semiaxes of the cavity
$\bar{\varepsilon}$, effective strain	$\bar{r}$, average cavity radius
$\dot{\bar{\varepsilon}}$, effective strain rate	ρ, relative density
ε_o, strain at which a cavity becomes stable	U, profile radius at the neck of a tension specimen
ε, axial strain	w, half-diameter or half-width of a tension specimen
$\dot{\varepsilon}_1, \dot{\varepsilon}_2, \dot{\varepsilon}_3$, strain rate along the principal directions 1, 2, 3	σ, external normal stress
$\dot{\varepsilon}_{ij}$, strain-rate tensor	$\bar{\sigma}$, effective stress
ε_N, axial strain at the neck of a specimen	$\bar{\sigma}_s$, effective stress of soft phase
ε_i, strain in the defect portion of the specimen	$\bar{\sigma}_h$, effective stress of hard phase
ε_u, strain in the uniform region of the specimen	σ_M, mean stress
f_h, volume fraction of hard phase	σ_z, principal stress along the longitudinal direction
ϕ, stress-intensification factor	σ_r, principal stress along the radial direction
F, size of initial geometric defect (taper)	σ_θ, principal stress along the hoop direction
F_T, stress triaxiality factor	σ_z^{mac}, macroscopic longitudinal stress
η_i, individual cavity-growth rate parameter in uniaxial tension	σ_r^{mac}, macroscopic radial stress
η_{APP}, apparent cavity-growth rate	σ_θ^{mac}, macroscopic hoop stress
η^{ts}, individual cavity-growth rate parameter in complex stress state	σ'_{ij}, deviatoric stress tensor
η_i, cavity-growth rate along principal direction "i"	s, center-to-center cavity spacing
k, Boltzmann constant	V, cavity volume
K, strength coefficient	V_o, volume at which a cavity becomes stable
	Ω, atomic volume

Higher-magnification SEM analysis is often conducted in selected areas of a specimen to establish cavity nucleation sites, nucleation strain, onset of coalescence, coalescence mechanisms, and so on relative to microstructural features (triple points, second-phase particles) that influence cavitation behavior. As an example, Fig. 1(a) shows a typical optical micrograph used to determine the cavity features in Ti-6Al-4V following hot tension testing (Ref 5); a corresponding high-magnification SEM micrograph (Fig. 1b) reveals the onset of coalescence of cavities.

Experimental measurements and observations of cavitation are summarized in Table 2, which summarizes the broad test conditions and major conclusions related to cavitation behavior for a number of different alloys (Ref 7–18). This list is not exhaustive but provides an overview of the extent of the phenomenon. Observations for aluminum alloys are not included here, inasmuch as they are discussed in the companion article by A.K. Ghosh in this Volume.

As microscopic examinations reveal, cavitation comprises three distinct stages that usually occur simultaneously: nucleation, growth, and coalescence. The modeling of each of these stages is briefly described in the following sections.

Modeling of Cavity Nucleation

Nucleation represents the first stage of cavitation. Cavities nucleate by:

- Intersection of matrix slip bands with nondeformable second-phase particles or with grain boundaries
- Sliding along grain boundaries, giving rise to stress concentrations at triple junctions that are not relaxed by diffusional transport
- Condensation of vacancies at grain boundaries

Once a cavity is generated, the propensity for it to grow (or shrink) is determined from the local stress equilibrium (Ref 19, 20). Specifically, a cavity is stable when its size exceeds a critical radius, r_c, which is given by the following relationship:

$$r_c = 2(\gamma + \gamma_p - \gamma_i)/\sigma \qquad \text{(Eq 1)}$$

in which γ, γ_p, and γ_i denote the surface energies of the cavity/void, the particle, and the particle-matrix interface; σ is the applied stress. This criterion implies that flow hardening is required to continuously nucleate and grow cavities, which usually does not occur during conventional hot working or superplastic flow (except in cases of significant grain growth). The equation also leads to required stresses for initiation and early growth that are unrealistically high. Therefore, other methods based on nucleation and growth from inhomogeneities/regions of high local stress triaxiality have been developed. For example, Ghosh et al. (Ref 21) have developed a constrained-plasticity model in which nanocavities nucleated by slip intersections with nondeformable second-phase particles or grain boundaries grow due to stresses normal and/or parallel to the interface by local plasticity at the tip of a cracklike defect. In this case, the stress state is highly dilatational, and the early rate of growth is thus extremely rapid. The growth rate of a cavity decreases significantly as its size increases and becomes comparable to the microstructural feature that produces the constraint. Such models are described in more detail in the companion article on cavity nucleation in this Volume.

Modeling of Cavity Growth

Cavity growth follows nucleation. Cavity-growth mechanisms can be classified into two broad categories: diffusion-controlled growth and plasticity-controlled growth. Diffusional growth dominates when the cavity size is very small. As the cavity size increases, diffusional growth decreases very quickly, and plastic flow of the surrounding matrix becomes the controlling mechanism (Ref 19, 22).

Diffusion-controlled growth has been modeled (Ref 22–24) assuming spherical, widely spaced, and noninteracting cavities. The variation of the cavity radius, r, with effective strain, $\bar{\varepsilon}$, is given by the following expression:

$$\frac{dr}{d\bar{\varepsilon}} = \frac{D\delta}{2kT\dot{\bar{\varepsilon}}}\Omega\frac{\bar{\sigma}}{r^2}\left(\frac{k_D}{3} + \frac{P}{\sigma} + \frac{2\gamma}{r\bar{\sigma}}\right)\frac{1}{\ln(\lambda/2r) - 3/4} \qquad \text{(Eq 2)}$$

Here, D is the diffusion coefficient, δ is the grain-boundary thickness, k is the Boltzmann constant, Ω is the atomic volume, $\dot{\bar{\varepsilon}}$ is the effective strain rate, $\bar{\sigma}$ is the effective stress, λ is the cavity spacing, P is the applied hydrostatic pressure, γ is the surface energy, and k_D is a constant dependent on the geometry of the deformation.

The diffusional-growth mechanism is of limited engineering importance because the great majority of cavity growth occurs by a plasticity-controlled mechanism under the conventional hot working conditions employed in industrial forming processes. In such cases, the cavity-growth rate depends on the deformation temperature, the strain rate, and the stress state. The remainder of this section, therefore, deals with the modeling of cavity growth under both uniaxial and multiaxial stress conditions.

Plasticity-Controlled Growth under Uniaxial-Tension Conditions. The plasticity-

(a)

(b)
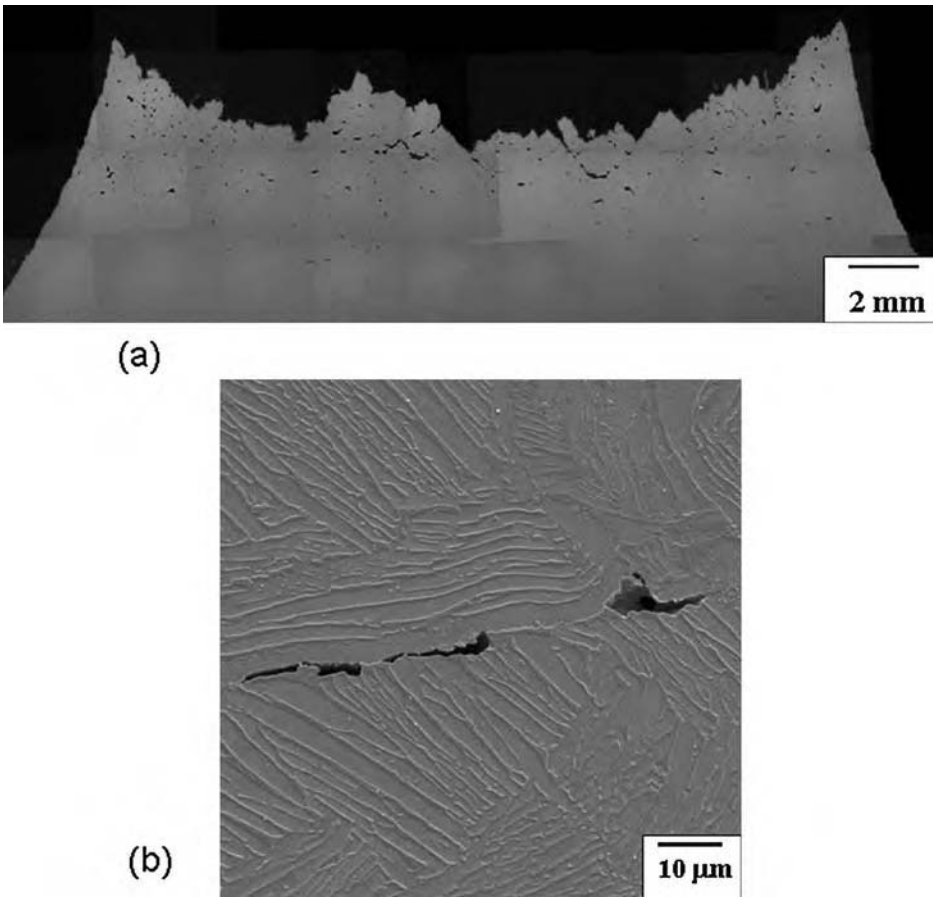
Fig. 1 Cavitation observations for Ti-6Al-4V with a colony-alpha microstructure. (a) Optical micrograph. (b) SEM micrograph

Table 2 Experimental observations of cavitation

Investigator(s)	Reference	Material	Testing	Observations
Humphries and Ridley	7	α/β brass	Tensile, $T = 540–640°$ C $\dot{\varepsilon} = 1.67 \times 10^{-4}$ to 3.33×10^{-3} s^{-1}	1. Cavities appeared to be nucleated at triple points. 2. Cavitation increased with the increase of strain, strain rate, and grain size as well as the decrease of temperature.
Caceres and Wilkinson	8, 9	Ultrafine Coronze 638 (Cu-2.5%Al-1.8%Si-0.4%Co)	Tensile	1. The damage process is essentially independent of the strain rate at regions I and II of the σ-$\dot{\varepsilon}$ curve. 2. At high strain rates, the level of cavitation was low, and failure occurred by necking. 3. Extensive cavity coalescence was observed up to strains of 1.5.
Belzunce and Suery	10	Superplastic α/β brasses	Tensile	Cavities are preferentially located at junctions between two α-grains.
Chandra et al.	11	Superplastic α/β brass	Tensile	1. Cavitation is associated with grain-boundary sliding. 2. Nucleation occurs at points of stress concentrations in the sliding interfaces. 3. α-β interfaces are predominant on nucleation due to unbalanced strain accommodation between the two grains.
Chokshi	12	Cu-2.8Al-1.8Si-0.4Co	Tensile, $T = 550°$ C, $\dot{\varepsilon} = 1.3 \times 10^{-3}$ s^{-1}	1. Stringers form parallel to the tensile axis. 2. Cavities nucleate continuously during deformation. 3. Cavity coalescence occurred in a direction perpendicular to the tensile axis.
Sagat and Taplin	13	Ternary brass	Tensile, $T = 400 – 800°$ C, $\dot{\varepsilon} = 1.7 \times 10^{-3}$ to 1.7×10^{-5} s^{-1}	The plastic flow was accompanied by simultaneous nucleation and growth of intergranular and interface cavities; failure occurred by fracture without macroscopic necking.
Livesey and Ridley	14	Microduplex Pb-Sn eutectic	Tensile	1. Cavities led to a brittle-type mode of fracture. 2. The level of cavitation was enhanced as the volume fraction, hardness, and size of the intermetallic particles were increased.
Livesey and Ridley	15	Cu-Zn-Ni microduplex alloys	Tensile	Cavities with radius over 0.5 μm grew by plasticity, while the growth of cavities with radius below 0.5 μm is controlled by diffusional processes.
Ahmed et al.	16	Superplastic Zn-22Al	Tensile, $T = 230°$C, $\dot{\varepsilon} = 1.3 \times 10^{-3}$ s^{-1}	Cavitation occurs up to strain rates of 3×10^{-2} s^{-1} and is most extensive in region I of the σ-$\dot{\varepsilon}$ curve.
Kim et al.	17	Al$_2$O$_3$-0vol%ZrO$_2$	…	Cavities changed their shape during deformation; the cavity aspect ratio decreased from ~1.0 at the initial stage to ~0.5 prior to failure.
Semiatin et al.	2	Ti-6Al-4V	Tensile, $T = 875–950$ ^{o}C	1. Cavity initiation occurred at very low strains (~0.2) regardless of strain rate for $T < 900$ ^{o}C. 2. Above 900 ^{o}C, substantially higher cavity-initiation strains and lower cavity-growth rates were observed.
Lombard et al.	18	γ-titanium aluminide	Tensile, $T = 900 – 1200°$ C $\dot{\varepsilon} = 10^{-2}$ to 10^{-4} s^{-1}	1. Cavity growth was plasticity controlled. 2. The largest cavity size and cavity density increased with decreasing strain or strain rate and decreasing temperature. 3. At 1200 ^{o}C, cavity-growth rates were lowest and tensile ductilities were highest.

controlled growth of an isolated, noninteracting cavity during uniaxial tensile deformation is given by the following equation:

$$V = V_o \exp(\eta(\varepsilon - \varepsilon_o)) \text{ or } r = r_o \exp\left(\frac{\eta}{3}(\varepsilon - \varepsilon_o)\right)$$
(Eq 3a)

or

$$\frac{dV}{d\varepsilon} = \eta V$$
$$\frac{dr}{d\varepsilon} = \frac{\eta r}{3}$$
(Eq 3b)

In Eq 3, V and r are the cavity volume and radius, respectively; V_o and r_o are the volume and radius of the cavity at a strain ε_o at which it becomes stable; ε denotes axial strain; and η is the individual cavity-growth rate parameter (Ref 18, 19, 25). The cavity-growth rate, η, is related to material parameters such as the strain-rate sensitivity, m, which itself is a function of deformation parameters (strain rate and temperature). For a planar array of spherical, noninteracting, grain-boundary cavities under tensile straining conditions, the following relationship between η and m has been derived (Ref 26):

$$\eta = 1.5\left(\frac{m+1}{m}\right) \sinh\left[\frac{2}{3}\frac{(2-m)}{(2+m)}\right]$$
(Eq 4)

This theoretical relationship has been found to describe well experimental observations for a wide range of cavitating metals and alloys.

The measurement of the cavity-growth rate of an individual cavity is difficult because continuous nucleation and cavity coalescence takes place in parallel with growth. To overcome this problem, the cavity volume fraction, C_v, is measured; its variation with strain is described by a relationship similar to Eq 3:

$$C_v = C_{v_o} \exp[\eta_{APP}(\varepsilon - \varepsilon_o)]$$
(Eq 5)

in which the apparent cavity-growth rate, η_{APP}, replaces the individual cavity-growth rate, η. The parameter η_{APP} is an average incorporating the growth of individual cavities, continuous cavity nucleation, and coalescence; η_{APP} is readily determined from semilog plots of C_v versus ε (Ref 18, 27) (e.g., Fig. 2). The data in Fig. 2 correspond to a gamma titanium aluminide alloy tested in tension at a temperature of 1000 °C and a strain rate of 10^{-4} s^{-1}. The slope of the fitted line equals η_{APP} (Ref 18).

Simulations of cavity growth with continuous nucleation have led to the delineation of the difference between η_{APP} and η. Typical simulation results (Fig. 3) show the ratio η_{APP}/η as a function of the cavity-nucleation rate, N (defined as the number of cavities that initiate per unit volume and per unit strain), and the strain-rate sensitivity. This ratio is close to unity when the cavity-nucleation rate is low; however, it deviates from unity as N increases. In addition, it has been found that the
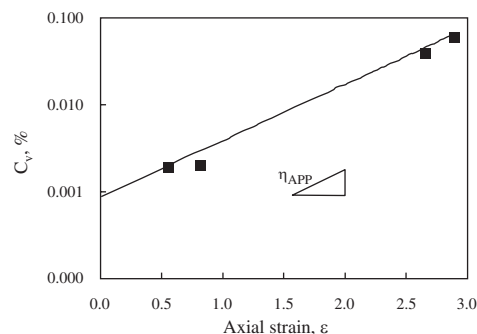
Fig. 2 Measured cavity volume fraction (C_v) as a function of axial strain from tension testing of a gamma titanium aluminide alloy. The test temperature was 1000 °C, and the strain rate was 10^{-4} s^{-1}. Source: Ref 18
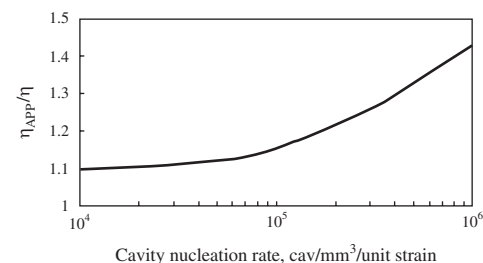
Fig. 3 Simulation results for the ratio of the apparent to individual cavity-growth rates as a function of the cavity-nucleation rate. Source: Ref 27

dependence of η_{APP}/η on the strain-rate sensitivity, m, is relatively weak (Ref 27).

The applicability of simple relationships such as Eq 3 and 5 has been established by the modeling work conducted by Hancock (Ref 28) and Stowell (Ref 29).

Hancock Model. This plasticity-controlled growth analysis considers a single cylindrical hole in an incompressible viscous solid (Ref 28). The model gives rise to an additional surface-tension term in Eq 3(b), that is, γ/r, in which γ is the surface energy and is typically of the order of 1 J/m^2. The term γ/r may be viewed as a radial stress introduced on the surface of the cylindrical void. For such a void, the resulting equation is:

$$\frac{dr}{d\varepsilon} = \frac{\eta}{3}\left(r - \frac{3\gamma}{2\sigma}\right) \qquad \text{(Eq 6)}$$

The previous expression reveals that the surface-tension term is negligible when the applied stresses are high, which is typically the case under hot working conditions.

Stowell Model. The model by Stowell for plasticity-controlled cavity growth was developed for superplastic alloys. The analysis focuses on a cylindrical representative volume element with a diameter equal to the cavity spacing and a length equal to the grain diameter (Ref 29). The model accounts for the increase in strain rate in the region where the cavity is present relative to that in the uncavitated region. An initial cavity elongates with strain, leading to the definition of an equivalent radius $r = (\frac{1}{3})(a + 2b))$, in which a and b denote the major and minor axes of the elliptical cavity, respectively. This approximation for the equivalent cavity radius becomes less accurate as the aspect ratio $a{:}b$ increases. Assuming that the cavity volume (V) is proportional to r^3 and the cavity density (number of cavities per initial unit volume, V_o) is N, the Stowell analysis leads to an expression similar to Eq 3.

Stowell's model has some limitations, mostly related to the assumptions that a fixed number of cavities grow from the beginning of the deformation, and there is no interaction between neighboring cavities. In addition, it assumes the same law for the different directions of cavity growth.

Plasticity-Controlled Growth under Multiaxial Stress States. The stress state developed during industrial metalworking operations is usually multiaxial. To describe cavity growth, modifications must therefore be made to relationships such as Eq 3(a, b), which is strictly applicable for deformation under uniaxial-tension conditions. For this purpose, the simplest approach consists of the application of a relationship similar to Eq 3 in which the uniaxial-tension cavity-growth parameter η is replaced by a cavity-growth parameter for the complex stress state, that is, η^{ts}, and the uniaxial strain, ε, is replaced by the effective strain, $\bar{\varepsilon}$. A complex stress state is quantified by considering the stress ratio $\sigma_M/\bar{\sigma}$, that is, the ratio of the mean stress

to effective stress. The cavity-growth rate parameter η^{ts} is simply a function of $\sigma_M/\bar{\sigma}$, that is:

$$\frac{\eta^{ts}}{\eta} = F\left(\frac{\sigma_M}{\bar{\sigma}}\right) \qquad \text{(Eq 7)}$$

Thus, the multiaxial stress analog of Eq 3(a) is:

$$r = r_o \exp\left[\frac{\eta^{ts}}{3}(\bar{\varepsilon} - \bar{\varepsilon}_o)\right] \text{ or}$$
$$r = r_o \exp\left[\frac{\eta}{3}F\left(\frac{\sigma_M}{\bar{\sigma}}\right)(\bar{\varepsilon} - \bar{\varepsilon}_o)\right] \qquad \text{(Eq 8)}$$

The function $F(\sigma_M/\bar{\sigma})$ is therefore required to describe cavity growth under multiaxial stress states. As described next, the function $F(\sigma_M/\bar{\sigma})$ has been determined using continuum mechanics approaches (Ref 30–33) as well as semiempirical methods (Ref 34).

McClintock Model. McClintock used a continuum approach to obtain equations that describe the growth of voids of different geometries under different stress conditions (Ref 32, 33, 35). The initial focus was on the description of the growth of cylindrical voids in a linear-hardening material subject to an axisymmetric stress state or in a rigid, nonhardening material under axisymmetric deformation conditions. This was followed by an assessment of the growth of voids within a moderately work-hardening material. The overall result was a law that describes the growth, within a shear band, of a cylindrical void of elliptical cross section in a linear-hardening matrix under arbitrary biaxial loading (Ref 36, 37):

$$\frac{1}{r}\frac{dr}{d\bar{\varepsilon}} = \frac{1}{2(1-n)}\sinh\left[(1-n)\frac{\sigma}{\tau}\right] \qquad \text{(Eq 9)}$$

in which τ is the shear stress within the shear band, σ is the normal stress across it, $\bar{\varepsilon}$ is the effective strain, and n is the strain-hardening exponent. A simple modification of Eq 9 provides an order-of-magnitude estimate for the growth of elliptical cavities by taking an average of the three principal stresses:

$$\frac{dR}{d\bar{\varepsilon}} = \left\{\frac{\alpha}{1-n}\sinh\left[(1-n)\frac{\sigma_M}{\tau_o}\right]\right\}R \qquad \text{(Eq 10)}$$

Here, R denotes the average of the three semi-axes of the cavity, α is a constant of the order of unity, and σ_M is the mean stress. This model shows that the void growth decreases as n increases. By analogy with Eq 3(b), the term within the angular brackets in Eq 10 equals $\eta^{ts}/3$. The value of η^{ts} can be readily calculated from a relationship derived by Rice and Tracey, which is described next.

Rice-and-Tracey Model. A continuum-mechanics approach was also employed by Rice and Tracey (Ref 30) to describe the growth of a spherical cavity within a plastic, nonhardening material obeying the von Mises yield criterion. The strain field was assumed to comprise three contributions:

- A uniform strain field due to plastic deformation of the matrix
- A spherically symmetric strain field resulting from the change of the cavity volume but involving no shape change
- A strain field (decaying at remote distances) that arises from changes of the void shape but not its volume

The analysis of Rice and Tracey revealed that the contribution of the change of the cavity shape to the strain is minimal; on the other hand, the other two factors had a much more potent effect. In fact, they showed that by neglecting the cavity-shape-change strain field, the error introduced was less than 1%. This analysis led to the determination of the following dependence of η^{ts} on the stress ratio $\sigma_M/\bar{\sigma}$:

$$\eta^{ts} = 0.558 \sinh(3\sigma_M/2\bar{\sigma}) + 0.008 \, \nu \cosh(3\sigma_M/2\bar{\sigma}) \qquad \text{(Eq 11)}$$

in which ν is a function of the principal strain rates, that is, $\nu = -3\dot{\varepsilon}_2/(\dot{\varepsilon}_1 - \dot{\varepsilon}_3)$, with $\dot{\varepsilon}_1 \geq \dot{\varepsilon}_2 \geq \dot{\varepsilon}_3$.

Pilling-and-Ridley Model. An alternate, semiempirical approach (Ref 34) has also been used to quantify the effect of the hydrostatic pressure, P, on the cavity-growth rate, η^{ts}, for superplastic alloys, which often exhibit extensive cavitation. The empirical relationship between η^{ts} and η is:

$$\eta^{ts} = \eta\left(1 + 2\frac{P}{\bar{\sigma}}\right) \qquad \text{(Eq 12)}$$

where P is taken to be negative if it is compressive and positive if it is tensile. Because $\sigma_M = P + \bar{\sigma}/3$, $P/\bar{\sigma} = \sigma_M/\bar{\sigma} - \frac{1}{3}$, Eq 12 can then be rewritten in terms of the stress ratio:

$$\eta^{ts} = \eta\left(\frac{1}{3} + 2\frac{\sigma_M}{\bar{\sigma}}\right) \qquad \text{(Eq 13)}$$

Therefore, by combining Eq 8 and 13, the cavity size for a multiaxial stress state can be calculated.

Application of Models. The applicability of relationships for η^{ts}/η, such as Eq 10 to 13 (Ref 38), has been determined by comparing model predictions to experimental measurements. The collected measurements correspond to a variety of materials deformed under different stress states (e.g., equibiaxial tension, plane-strain tension, uniaxial tension of notched specimens, etc.) and processing conditions (i.e., temperature, strain rate). The experimental data fall within a broad scatter band (Fig. 4). Model predictions are also shown in this figure. A comparison of the scatter band and trend lines for the various equations indicates that none of the models provides precise predictions of the measurements. Nevertheless, the semiempirical model described by Eq 13 does mirror the observed trend better than the others. In fact, the upper and lower limits of the experimental scatter band follow this relationship (Eq 13) multiplied by a factor Q equal to 0.75

Fig. 4 Dependence of the ratio of the cavity-growth parameter under triaxial and uniaxial states of stress (η^{ts}/η) on the stress triaxiality ($\sigma_M/\bar{\sigma}$). The measurements (shaded area) are compared to various model predictions. Source: Ref 38

or 1.25, respectively. Thus, the general correlation of cavity growth and stress state is described by the following relationship:

$$r = r_o \exp\left[\frac{\eta}{3}Q\left(\frac{1}{3}+2\frac{\sigma_M}{\bar{\sigma}}\right)(\bar{\varepsilon}-\bar{\varepsilon}_o)\right] \qquad \text{(Eq 14)}$$

The usefulness of the correlation described by Eq 14 has been confirmed using two different sets of cavity size measurements (Fig. 5). The first set came from experiments involving a range of stress triaxialities developed during notched-tension testing of a titanium alloy (Ref 39), while the second focused on results from the equibiaxial-tension testing of an aluminum alloy (Ref 40).

Measurements of the average diameter of cavities developed during notched-tension testing of Ti-6Al-4V (with a colony-alpha microstructure) at 815 °C and a nominal strain rate of 0.1 s^{-1} (Ref 39) are plotted as a function of the stress ratio $\sigma_M/\bar{\sigma}$ in Fig. 5(a). The individual data points correspond to two different cavity-growth strain ranges, $\varepsilon-\varepsilon_o$. In addition, predictions for $\eta = 6.5$, values of the constant Q between 0.75 and 1.25, and the same levels of $\varepsilon-\varepsilon_o$ (i.e., ~0.14 and ~0.22) are also shown in the graph as shaded areas. It can be seen that the correlation embodied in Eq 14 bounds the experimental measurements well.

Measurements (data points) of the effective-strain dependence of the average diameter of cavities developed during the equibiaxial-tension testing ($\sigma_M/\bar{\sigma} = 0.66$) of a fine-grained, modified 5083 aluminum alloy at two different strain rates and a temperature in the superplastic range are shown in Fig. 5(b) (Ref 40). Applying Eq 14 with $Q = 0.75$ and 1.25, a cavity-growth rate in uniaxial tension of $\eta = 3.2$, and an assumed nucleation strain, ε_o, of zero (solid lines), it is seen that model predictions bound the observations for this material as well.

Cavity-Growth Rate along the Principal Directions. In many cases, cavities do not grow as simple spheres whose radii increase with strain; rather, they grow as ellipsoidal voids. Hence, the use of a single cavity-growth parameter η (or η^{ts}) is not justified. In such cases, it is often useful to quantify the cavity-growth rate along each of the principal directions of the ellipsoidal cavity. In particular, the change of the size of a cavity along direction i may be described by the relationship:

$$L_i = L_{io}\exp(\eta_i\bar{\varepsilon}) \qquad \text{(Eq 15)}$$

in which L_i and η_i denote the size and cavity-growth rate parameter along direction i.

As an example, directional cavity growth has been observed during the hot torsion testing of Ti-6Al-4V (Ref 41). The cavity-growth rates along the three principal specimen directions (z, r, θ) were determined by measuring the cavity sizes, L_z, L_r, and L_θ via optical microscopy on the z-r and z-θ cross sections. The measured average cavity lengths (L_z, L_r, and L_θ) as a function of effective strain (Fig. 6) indicated that the

Fig. 5 Comparison of predictions from Eq 14 of the average cavity size as a function of the stress triaxiality ($\sigma_M/\bar{\sigma}$) with measured cavity sizes (data points) developed during (a) notched-tension testing of Ti-6Al-4V (Source: Ref 39) or (b) equibiaxial-tension testing of a fine-grained 5083 modified aluminum alloy (Source: Ref 39, 40)
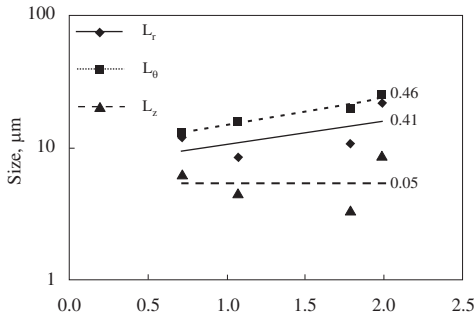
LIVE GRAPH
Click here to view

Fig. 6 Strain and direction dependence of the size of cavities developed during hot torsion testing of Ti-6Al-4V with a colony-alpha microstructure at 815 °C and a surface effective strain rate of 0.04 s^{-1}. Source: Ref 41

cavity-growth parameter is indeed directional in nature $\{\eta\}$:

$$\{\eta\} = \begin{pmatrix} \eta_z & & \\ & \eta_r & \\ & & \eta_\theta \end{pmatrix} = \begin{pmatrix} 0.05 & & \\ & 0.41 & \\ & & 0.46 \end{pmatrix}$$

The volumetric cavity-growth rate η equals the sum (trace) of the components of the previous matrix: $\eta = \eta_z + \eta_r + \eta_\theta$ or $\eta = 0.92$. The value $\eta = 0.92$ was found to be in broad agreement with the value of η if the cavities had been treated as spherical in shape.

Micromechanical Models of Plasticity-Controlled Growth. Classical plasticity-controlled models, such as those discussed previously, neglect the effect of local texture and microstructure on cavity growth. As such, their principal use is for the prediction of the average cavity size and volume fraction developed during hot working or superplastic forming. However, it is the size of the largest cavities that play the most important role with respect to the properties and performance of a finished product. To this end, so-called micromechanical models that enable such predictions have been developed to fill this gap.

Micromechanical models (Ref 42, 43) focus on quantifying local stresses and the partitioning of strain between adjacent grains (or colonies in the case of a lamellar microstructure) due to differences in crystallographic orientation. The first analyses were developed for alpha/beta titanium alloys with a colony-alpha microstructure in which cavitation is quite severe. The plastic anisotropy of the alpha (hexagonal close-packed) phase leads to colonies with markedly higher or lower Taylor factors, which therefore behave differently during deformation.

Micromechanical cavity-growth models comprise three elements:

- An analysis to estimate the approximate stress ratio within adjacent hard and soft grains or colonies
- A self-consistent calculation for the partitioning of the macroscopic strain/strain rate to differently oriented grains/colonies
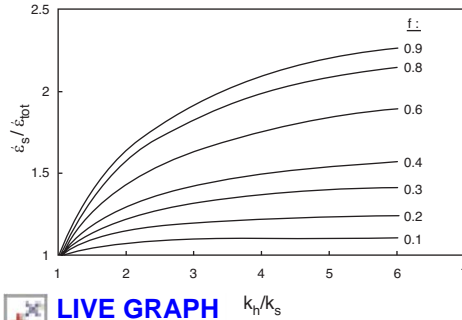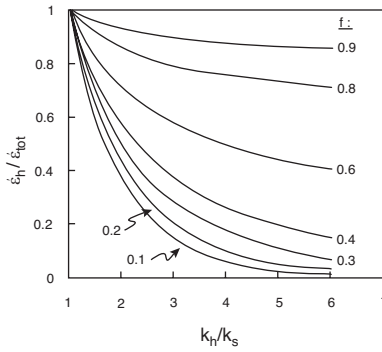
LIVE GRAPH
Click here to view

Fig. 7 Self-consistent model predictions of the average strain rates developed in the individual phases of a two-phase (hard, h, and soft, s) mixture as a function of k_h/k_s, f, and the macroscopic (imposed) strain rate. The m value for both phases was 0.23. Source: Ref 44

- A plasticity-controlled model incorporating stress-state effects to estimate the actual cavity-growth kinetics (e.g., Eq 8).

The basic assumptions of these models are:

- The cavity formed at the boundary between a hard grain/colony and a soft grain/colony is small relative to the grain size and hence has little influence on the stresses and strains that are developed in the absence of the cavity.
- Cavities grow into softer grains/colonies.

Uniaxial Tension. For uniaxial-tension deformation, the stress ratio $(\sigma_M/\bar{\sigma})$ controlling cavity growth within the soft colony is given by the following expression (Ref 42):

$$\frac{\sigma_M}{\bar{\sigma}} = f_h \left[\frac{M_h}{M_s} \left(\frac{\dot{\bar{\varepsilon}}_h}{\dot{\bar{\varepsilon}}_s} \right)^m \left(\frac{\bar{\varepsilon}_h}{\bar{\varepsilon}_s} \right)^n - 1 \right] + \frac{1}{3} \quad \text{(Eq 16)}$$

in which M, $\dot{\bar{\varepsilon}}$, and $\bar{\varepsilon}$ denote the Taylor factor, strain rate, and strain of a hard (subscript "h") or soft (subscript "s") grain/colony, respectively; f_h is the volume fraction of the hard grains/colonies; and m and n denote the strain-rate sensitivity and the strain-hardening exponent, respectively. The partitioning of strain/strain rate between hard and soft grains/colonies, which is needed as input for both the stress-state (Eq 16) and the cavity-growth (Eq 14 for $Q = 1$) calculations, may be estimated from the self-consistent calculations (Ref 44), in which the relative strain rates are a function of the volume fraction of the hard grains/colonies and the ratio of strength coeffi-

cients/Taylor factors of the hard and soft grains/colonies. In Fig. 7, model results of the ratio of the strain rate of higher-flow-stress grain/colony to the macroscopic (imposed) strain rate $(\dot{\varepsilon}_h/\dot{\varepsilon}_{tot})$ and the corresponding ratio for the lower-flow-stress grain/colony $(\dot{\varepsilon}_s/\dot{\varepsilon}_{tot})$ are given as a function of the ratio of the strength coefficients, k_h/k_s, and the different volume fraction, f, of the hard grain/colony. The strain rate in the higher-flow-stress (hard) phase is less than in the lower-flow-stress phase. The difference increases as either f decreases and/or k_h/k_s increases.

A typical application of the model for cavitation during hot tension of Ti-6Al-4V with a colony-alpha microstructure is shown in Fig. 8. Here, measured radii, r (for cavities larger than 3 μm) (data points), are plotted as a function of the Taylor factor ratio, M_h/M_s. The solid lines represent model predictions; the strain at which each measurement corresponds is shown in the legend. The shaded bands correspond to a range of the fraction of the hard phase, f_h, between 0.4 and 0.8 (increasing from right to left, as shown by the arrow in the graph) and three different cavity-growth strain intervals, that is, $\varepsilon-\varepsilon_o$ = 0.18, 0.35, and 0.55, in which ε_o denotes the cavity-nucleation strain. The line within the shaded area corresponds to $f_h = 0.6$. The model reveals that for a given M_h/M_s, the cavity radius increases as the fraction of the hard phase increases at the expense of the soft one, largely because the strain accommodated by the soft phase increases. However, the cavity radius has a stronger dependence on the strain at nucleation $(\varepsilon-\varepsilon_o)$ than f_h. Overall, the experimental observations in Fig. 8 are well bounded by the model predictions for a cavity-growth strain range between 0.18 and 0.55, with the majority of cavity measurements lying close to the $\varepsilon-\varepsilon_o = 0.18$ condition (Ref 42).

Multiaxial Stress States. The principal aspects of the micromechanical analysis of cavity growth under a macroscopic multiaxial stress state also comprise estimates of the local stress state, the partitioning of strain/strain rate between hard and soft grains/colonies using a strain-partitioning model (Ref 44), and the application of the cavity-growth equation (Eq 14 for $Q = 1$) to the soft colony.

The macroscopic stress state is assumed to comprise three principal stresses (e.g., axial, σ_z; radial, σ_r; and hoop, σ_θ, for axisymmetric deformation), which can be determined from finite-element method (FEM) analysis. The microscopic local stresses in the hard and soft grains/colonies are determined by taking into account the yield functions of the hard and soft colonies and the relevant local-equilibrium equations. For a material obeying the von Mises yield criterion, the yield-function relationships are:

$$\bar{\sigma}_h = \frac{1}{\sqrt{2}} \left\{ (\sigma_{\theta_h} - \sigma_{z_h})^2 + (\sigma_{z_h} - \sigma_{r_h})^2 + (\sigma_{r_h} - \sigma_{\theta_h})^2 \right\}$$

$$\text{(Eq 17)}$$
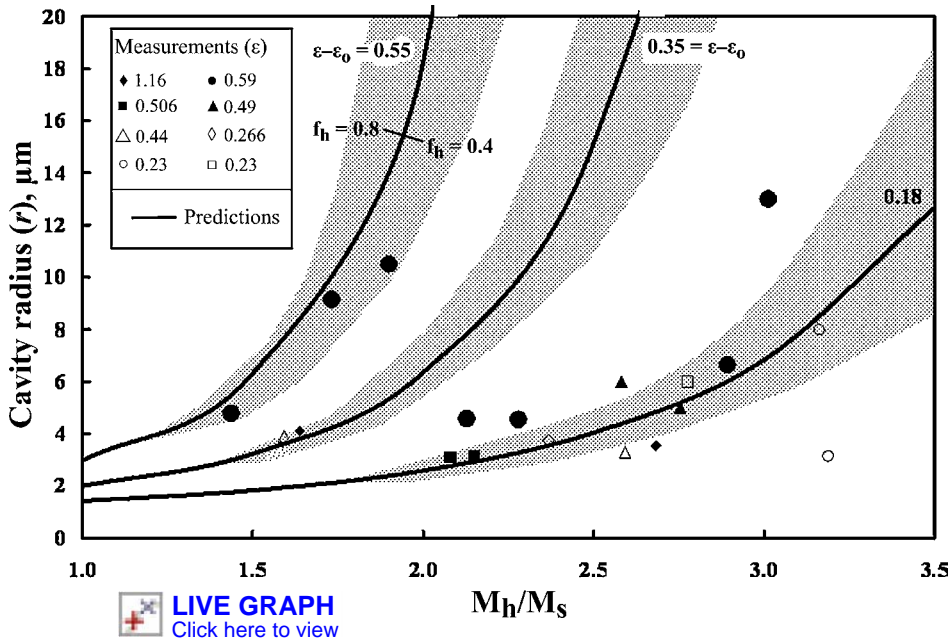
Fig. 8 Comparison of measurements and predictions of the cavity radius as a function of the Taylor-factor ratio M_h/M_s for cavity radii greater than 3 μm. The solid lines represent predictions for $f_h = 0.6$ for three different ranges of the cavity-growth strain. The local fraction of hard orientations, f_h, increases in the shaded region from 0.4 to 0.8 in the direction shown. Source: Ref 42
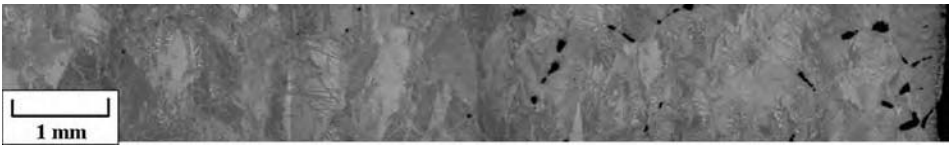
Fig. 9 Montage of micrographs along the equatorial plane showing the variation of cavitation in the radial direction developed during hot pancake forging of Ti-6Al-4V with a colony-alpha microstructure. Source: Ref 43

$$\bar{\sigma}_s = \frac{1}{\sqrt{2}}\left\{(\sigma_{\theta_s} - \sigma_{z_s})^2 + (\sigma_{z_s} - \sigma_{r_s})^2 + (\sigma_{r_s} - \sigma_{\theta_s})^2\right\}$$
(Eq 18)

The subscripts θ, r, and z refer to the specific principal stress components; the subscripts h and s refer to the hard and soft colonies, respectively; and $\bar{\sigma}_h$ and $\bar{\sigma}_s$ denote the flow stress of the soft and hard colonies, respectively.

The macroscopic axial (σ_z^{mac}) and radial (σ_r^{mac}) stresses that are determined from FEM analysis are each assumed to be a rule-of-mixtures average of the corresponding stress components in the hard and soft colonies, thus satisfying the load equilibrium considerations:

$$\sigma_z^{mac} = f_h\sigma_{z_h} + (1 - f_h)\sigma_{z_s}$$
(Eq 19)

$$\sigma_r^{mac} = f_h\sigma_{r_h} + (1 - f_h)\sigma_{r_s}$$
(Eq 20)

Furthermore, the hoop stress is usually taken to be essentially the same in the soft and hard grains/colonies and is equal to the macroscopic (FEM) stress σ_θ^{mac}:

$$\sigma_\theta^{mac} = \sigma_{\theta_h} = \sigma_{\theta_s}$$
(Eq 21)

If the flow stresses of the hard ($\bar{\sigma}_h$) and the soft ($\bar{\sigma}_s$) colonies are known, the axial and radial stresses in each colony can be determined by solving Eq 17 to 21. More typically, $\bar{\sigma}_s$ and $\bar{\sigma}_h$ are determined by applying the self-consistent model to estimate the strain rates and hence the flow stresses in the hard and soft colonies (Ref 43, 44).

Having determined the stress components, the stress ratio $\sigma_M/\bar{\sigma}$ in both the hard and soft colonies can be calculated. Because the soft colony undergoes more strain than the hard colony, micromechanical cavity-growth calculations are typically based on the strain in the soft colony and the value of η^{ts} corresponding to its stress ratio. The values of the strain and η^{ts} in the softer colony are then used to estimate the size of the largest colonies in accordance with Eq 8.

The micromechanical model for the size of the largest cavities developed under a state of multiaxial stress has been validated via observations for the hot pancake forging of cylindrical Ti-6Al-4V preforms with a colony-alpha structure (Ref 43) (Fig. 9). Typical calculations of the stress ratio ($\sigma_M/\bar{\sigma}$) developed within the hard and soft colonies at the equatorial free

surface reveal that the stress ratio in the softer colony is very sensitive to the Taylor factor ratio M_h/M_s, whereas it exhibits a much weaker dependence on M_h/M_s and has smaller values for the harder colonies. Model predictions for the size of the largest cavities (based on the calculated stress ratio, the corresponding values of η^{ts}, and the strain in the softer colony from the strain-partitioning model) are compared to measured radii of the large cavities (data points) as a function of distance form the free surface in Fig. 10. This comparison shows that the model mimics the observed behavior, with the best fit obtained for M_h/M_s between 1.5 and 3.

Modeling of Cavity Coalescence

Cavity coalescence refers to the stage at which voids link together, leading to final fracture. In most models, cavity coalescence/linkage is assumed to occur between first- and second-nearest-neighbor cavities only. If it is postulated that there is no nucleation of new cavities, then two coalescence mechanisms are possible (Ref 45–47):

- *Impingement:* This mechanism refers to the exhaustion of the ligament between the two cavities. As two cavities grow, the ligament between them is reduced until the outer surfaces of the cavities come into contact. This condition is described by the following relationship:

$$s - (r_i + r_j) = 0$$
(Eq 22)

in which s denotes the center-to-center cavity spacing, and r_i and r_j are the instantaneous radii of cavities i and j, respectively. It is important to note that a small initial intercavity spacing does not necessarily imply that coalescence will occur, because the horizontal and vertical spacing as well as the cavity-growth rate are both important.

- *Internal necking:* Rupture of matrix material between two cavities occurs as a result of plastic instability and flow localization, much as though the cavities and matrix between them form a microscopic tension specimen. Various criteria for describing coalescence via such an internal necking process have been based on (1) a critical matrix-ligament strain, ε_{cr}, (2) a critical stress, or (3) a critical void volume fraction. The first criterion is the one applied in most cases.

The critical matrix ligament strain, ε_{cr}, depends on the initial cavity spacing and the individual cavity-growth rate, η. This strain can be determined by considering the deformation of a microspecimen containing two cavities along its midplane, which act like a geometric defect. Details of such simulations are given in the section on tensile ductility in this article. Here, typical results for such a mechanism are summarized in terms of the critical strain as a
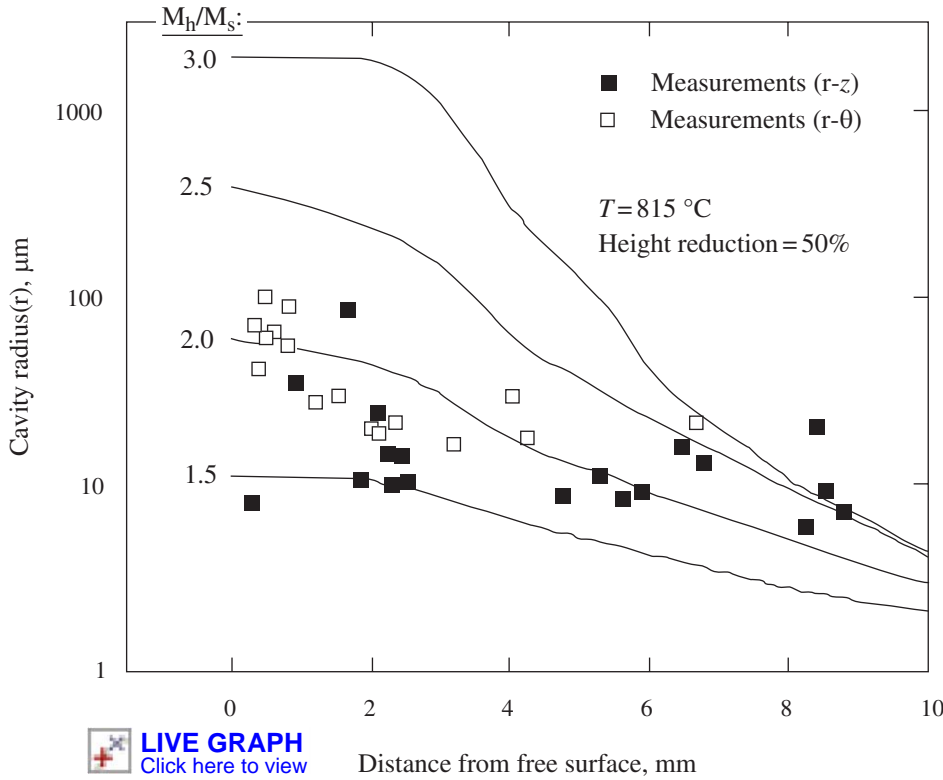
Fig. 10 Comparison of measurements and model predictions of the variation of the size of the largest cavities as a function of the distance from the free surface for a Ti-6Al-4V pancake hot forged to a 50% height reduction following preheating at 815 °C. Cavity-size measurements were from both axial (r-z) and in-plane (r-θ) sections. Source: Ref 43

Fig. 12 Simulation predictions of the average cavity radius for cavities that do or do not coalesce (Eq 23). The cavity volume fraction (C_v) is also shown on the graph.

Fig. 11 Dependence of the critical strain for ligament failure on the center-to-center cavity spacing and the individual cavity-growth rate, η. Source: Ref 47

in which C_v is the instantaneous volume fraction of cavities, η is the uniaxial-tension cavity-growth parameter (Eq 3), $\delta\bar\varepsilon(= d\bar\varepsilon)$ is a small increment of strain, $(dr/d\bar\varepsilon)_i$ is the rate of growth per unit strain of an isolated cavity ($= \eta r/3$ from Eq 3), and Φ is given by:

$$\Phi = (1 + \eta\bar\varepsilon\delta\bar\varepsilon/3 + (\eta\delta\;\bar\varepsilon)^2/27 \qquad \text{(Eq 24)}$$

Equation 23 can be easily applied to predict average cavity size by a simple spreadsheet analysis, in which strain is the independent variable, and the average cavity size is the dependent variable. An example of the application of Eq 23 is shown in Fig. 12 for a material with an individual cavity-growth rate (η) equal to 3. The cavity size for coalescence or no coalescence, along with the corresponding cavity volume fraction, is shown in this figure. These results show that the effect of coalescence on the average cavity radius becomes noticeable when the cavity volume fraction exceeds ∼7%.

In spite of prior research in this area, coalescence mechanisms are still not fully understood due to the experimental challenge in obtaining reliable data for the population of cavities during the process. These difficulties are associated primarily with the complex shapes of cavities. However, the recent application of novel experimental techniques has partially overcome such challenges. For example, Dupuy and Blandin (Ref 22) and Martin et al. (Ref 49) have conducted x-ray microtomography to obtain three-dimensional images of cavity populations in a nondestructive manner (Fig. 13). From this work, a coalescence or interlinkage parameter (IP) has been introduced. This parameter is the ratio of the volume of the largest cavity to the total volume of cavities in the specific region investigated. The microtomography results suggest that coalescence occurs over relatively large domains of strain and cavity volume fractions (Ref 22, 49). For example, Fig. 14 shows the variation of IP with the cavity volume fraction (which itself increases with strain) for aluminum alloy 5083 tested at 525 °C and a strain rate of 10^{-4} s^{-1}. The value of IP is relatively

function of the center-to-center cavity distance, d, and the cavity-growth rate, η (Fig. 11) (Ref 47); the critical strain increases as d increases and/or as η decreases.

Cavity coalescence may also be regarded as a process that, in effect, increases the average cavity-growth rate. In particular, the effect of

pairwise coalescence on the average cavity-growth rate, $d\bar r/d\bar\varepsilon$, can be estimated from the following equation (Ref 48):

$$\frac{d\bar r}{d\bar\varepsilon} = \frac{8C_v\Phi\eta\left(0.13r - 0.37(dr/d\bar\varepsilon)_i\delta\bar\varepsilon\right) + (dr/d\bar\varepsilon)_i}{1 - 4C_v\Phi\eta\delta\bar\varepsilon}$$

$$\text{(Eq 23)}$$

Fig. 13 Three-dimensional view of the largest coalesced cavities obtained via x-ray tomography of a sample of aluminum alloy 5083 following deformation in tension at 525 °C and 10^{-4} s^{-1}. Source: Ref 22
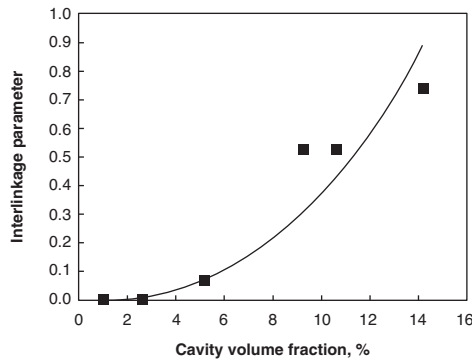
Fig. 14 X-ray tomography measurements of the cavity interlinkage parameter as a function of cavity volume fraction in a sample of aluminum alloy 5083 following deformation in tension at 525 °C and 10^{-4} s^{-1}. Source: Ref 49

LIVE GRAPH
Click here to view
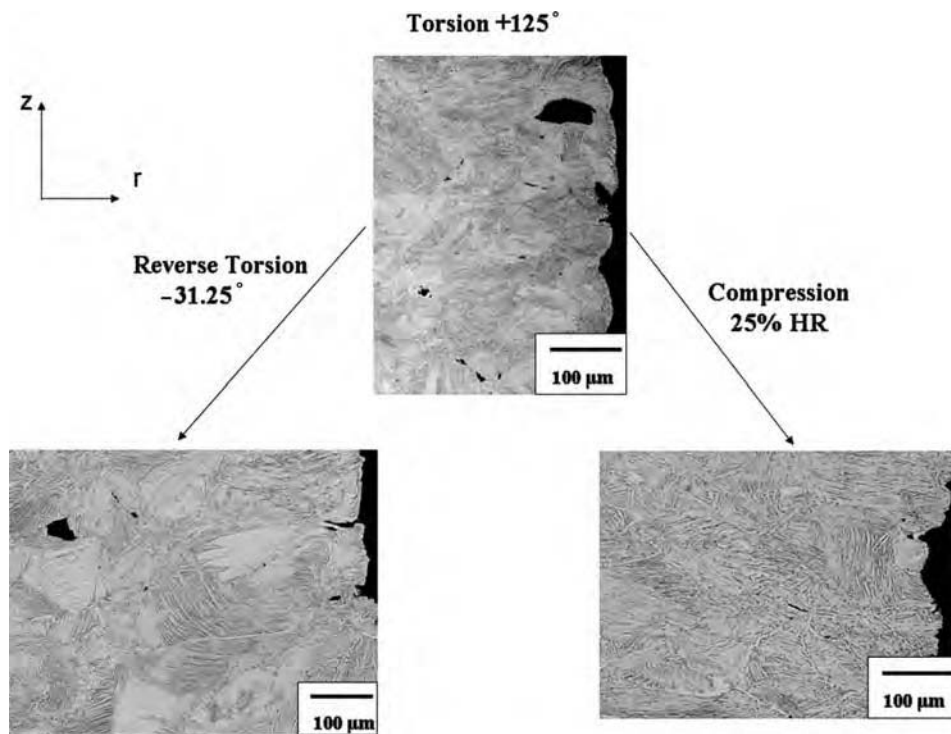
Fig. 15 Micrographs showing the effect of strain path on cavitation in Ti-6Al-4V samples with a colony-alpha microstructure deformed at 815 °C and an effective strain rate of 0.04 s^{-1} via monotonic torsion, reversed torsion, and compression following monotonic torsion. Source: Ref 51

low (<0.05) for a cavity volume fraction ≤5% (corresponding to a strain of ∼1.2), and then it increases sharply.

Modeling of Cavity Shrinkage

Cavities that nucleate and grow under a tensile state of stress may be reduced in size during a change in strain path, during which the stress state becomes compressive. A micromechanical model that describes cavity-shrinkage kinetics has been developed and applied (Ref 50). In this approach, the stresses developed within the soft grains/colonies during a strain-path change that gives rise to compressive triaxiality are estimated in the same way as in the case of open-die forging, that is, by applying the yield conditions and load-equilibrium equations. To estimate the strain-rate components and thus the densification rate $\dot{\rho} = \dot{\varepsilon}_{11} + \dot{\varepsilon}_{22} + \dot{\varepsilon}_{33}$, the stresses are inserted into a model for the consolidation of porous media (Ref 62):

$$\dot{\varepsilon}_{ij} = \left(\frac{K(\rho)\phi^2 \dot{\bar{\varepsilon}}}{\bar{\sigma}}\right)\left[(1+\nu)\sigma'_{ij} + (1-2\nu)\sigma_M \delta_{ij}\right]$$

(Eq 25)

in which $\dot{\varepsilon}_{ij}$ denotes the strain rate tensor, σ'_{ij} is the deviatoric stress tensor, σ_M is the mean stress, $\bar{\sigma}$ is the effective stress, ν is the Poisson's ratio of the porous body, ϕ is the stress-intensification factor, and δ_{ij} is the Kronecker delta. The function $K(\rho)$ is associated with the relative density ρ; it is commonly assumed to be equal to ρ when the latter is greater than 90%.

The applicability of Eq 25 to model cavity shrinkage has been demonstrated for Ti-6Al-4V with a colony-alpha microstructure subjected to torsion followed by reversed torsion or uniaxial compression (Fig. 15). Cavity-shrinkage predictions (Ref 50, 51) showed good agreement with such observations (Fig. 16a, b). The more rapid closure kinetics during compression compared to reversed torsion (as shown in the micrographs in Fig. 15) were quantified correctly. Such results were explained on the basis of the higher levels of compressive stress triaxiality in compression compared to those in reversed torsion, the orientation of cavities in torsion relative to the applied stresses, and dynamic spheroidization of the microstructure, which affected the stress triaxiality in torsion to a greater extent than in compression.

Modeling and Simulation Approaches to Predict Tensile Ductility and Develop Failure-Mode Maps

Cavitation may lead to premature failure and thus to a significant reduction in the tensile ductility or forming limit relative to that in materials that do not cavitate but fail instead by strain localization prior to fracture (Ref 52). For a given value of the strain-rate sensitivity (m), the decrease in elongation for fracture-controlled failures due to cavitation depends on the rates of cavity nucleation and growth, the cavity shape and size distribution, and the spatial distribution of the cavities.

Early models of the effect of cavitation on failure (e.g., Ref 53) focused only on the effect of cavity generation on flow stability in uniaxial tension. In these approaches, it was assumed that all cavities were spherical and uniformly distributed across a given cross section; triaxiality and stress-concentration effects were neglected. The models postulated that cavitation increases the three dimensions of the specimen (length, width, thickness) in proportion to the level of cavitation. Key results indicated that the progress of cavity generation renders deformation less stable by reducing the effective values of both the strain-hardening exponent (which primarily impacts the uniform elongation) and the strain-rate sensitivity (which affects the rate of flow localization following the onset of instability).

More recent models quantify the effect of cavitation on tensile ductility in more detail. These approaches include those that do or do not incorporate the effect of stress triaxiality in the neck on cavitation.

(a) LIVE GRAPH
Click here to view

(b) LIVE GRAPH
Click here to view   **Effective strain**

Fig. 17 Microspecimen geometry used in cavitation analysis by Zaki. Source: Ref 58
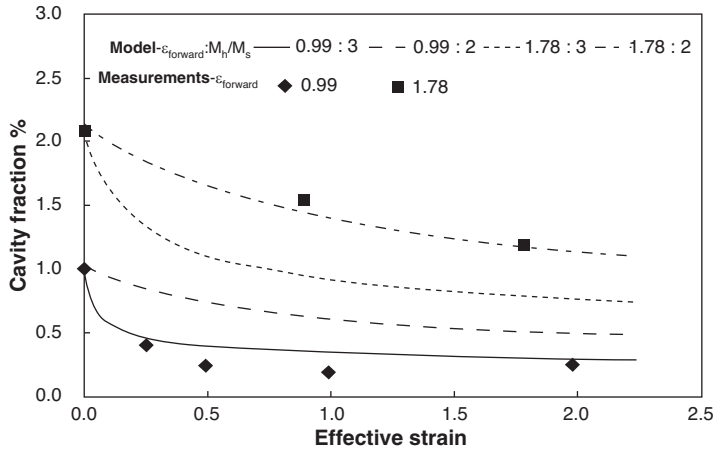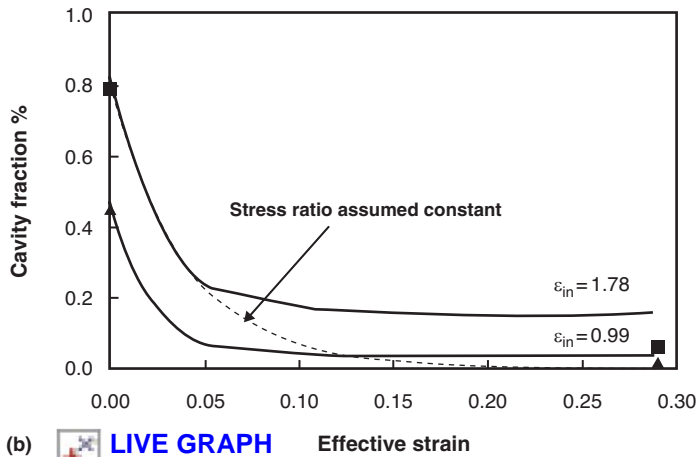
Fig. 16 (a) Comparison of model predictions (lines) and experimental measurements (data points) of the cavity fraction as a function of effective strain during reversed torsion at 815 °C of Ti-6Al-4V samples with two different levels of the initial cavity fraction. The model predictions are plotted for two different ratios of the Taylor factor between the hard and the soft colonies (M_h/M_s). Source: Ref 50 (b) Comparison of model predictions (lines) and measurements (data points) of the cavity fraction as a function of the macroscopic effective strain during compression of Ti-6Al-4V samples following torsional predeformation to a strain level denoted ε_{in}. Both prestraining and compression were conducted at 815 °C and an effective strain rate of 0.04 s^{-1}. Source: Ref 51

Long-wavelength models of the tension test seek to quantify cavitation and flow-localization behavior by assuming that the neck which developed after the onset of instability has a very long wavelength, and thus the stress triaxiality due to necking can be neglected (Ref 54, 55). In such situations, cavitation and strain-localization behavior are quantified by determining the relative deformation of two material slices (one with an initial defect of some sort) via a simple load-equilibrium approach. In particular, the analysis is based on the principle that the load (= axial stress multiplied by the cross-sectional area) transmitted through the two slices representing the specimen is constant. For a noncavitating material, deformed in uniaxial tension, the following equilibrium expression is obtained:

$$\exp(-\varepsilon/m)\varepsilon^{n/m}d\varepsilon = (1-F)^{1/m}\exp(-\varepsilon_N/m)\varepsilon_N d\varepsilon_N \tag{Eq 26}$$

In Eq 26, the subscript N refers to an initial geometric defect, F; and $F = (A_o - A_{oN})/A_o$, in which A_o and A_{oN} denote the area of the uniform and defected slice of the specimen.

For a cavitating material, cavity growth decreases the effective cross-sectional area of the specimen. Combining Eq 5 and 26 and assuming a uniform distribution of cavities, the following equilibrium relationship is obtained in long-wavelength two-slice models:

$$\exp(-\varepsilon/m)\varepsilon^{n/m}\{1 - C_{v_o}\exp[\eta_{APP}(\varepsilon - \varepsilon_o)]\}^{1/m}d\varepsilon = (1-F)^{1/m}\exp(-\varepsilon_N/m)\varepsilon_N\{1 - C_{v_o}\exp[\eta_{APP}(\varepsilon_N - \varepsilon_o)]\}^{1/m}d\varepsilon_N \tag{Eq 27}$$

Lian and Suery (Ref 56) were the first to develop and apply an expression of the form of Eq 27. In their simulations, they assumed an initial cavity volume fraction (C_{v_o}) between 10^{-3} and 10^{-4}. Cavitation failure was assumed to occur when the cavity volume fraction in the
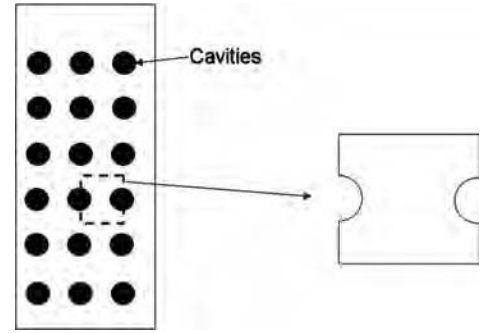
defect region reached 30%. From such analyses, maps showing the failure mode as a function of m and η can be derived. The maps are divided into three regimes:

- A regime in which the material fails by flow localization/necking
- A regime in which cavity growth dominates and leads to failure without pronounced flow localization
- An intermediate mixed-mode regime in which both cavitation and necking contribute to failure

Model predictions typically show good first-order agreement with observations in the literature, despite the fact that a precise description of cavity nucleation and coalescence is not included in the analysis.

The macroscopic, two-slice, long-wavelength model can be extended to establish the effect of superimposed hydrostatic pressure on the failure mode (Ref 57). For given values of m and η, the predicted failure mode changes when the superimposed pressure is greater than a certain value. It has been found that materials that fail without neck formation under zero-pressure conditions do indeed undergo flow localization prior to fracture when high pressures that suppress cavity growth are imposed.

Analogous to the macroscopic, long-wavelength analysis to determine ductility and failure mode, microscopic load-equilibrium approaches have been developed to investigate the internal necking and failure of the ligament between two cavities and hence the cavity-coalescence phenomenon. These latter techniques typically rely on a representative microspecimen (lying within the material) with a uniform rectangular cavity array (Fig. 17). In this case, a load-equilibrium analysis, such as that developed by Zaki (Ref 58), can be used to treat the straining process within the uniform and necked/defect portions of the specimen, the latter representing the ligament between two cavities. The load-equilibrium equation is:

$$\varepsilon_u^{n/m}d\varepsilon_u = \varepsilon_i^{n/m}\{1 - r_o\exp[(\eta/3)\varepsilon_i]\}^{2/m}d\varepsilon_i \tag{Eq 28}$$

in which n is the strain-hardening exponent, m is the strain-rate sensitivity, ε_u is the strain in

the uniform region, and ε_i is the strain in the defect portion of the specimen. Failure is assumed to occur when the ratio $\varepsilon_i/\varepsilon_u$ tends toward infinity. It has been found that the strain-rate sensitivity strongly affects the total strain when the individual cavity-growth rate is low (<3).

Macroscopic Load-Equilibrium Approach with Stress Triaxiality. A load-equilibrium approach that includes the influence of stress tri-axiality developed within the neck on flow localization and failure during tension testing provides more physically realistic estimates of ductility and failure mode than long-wavelength models for both cavitating and noncavitating materials (Ref 59, 60). For a sheet or round-bar tension test, the formulation is based on:

- A discretized sample geometry rather than merely two slices (e.g., Fig. 18)
- A description of the material flow (stress-strain) behavior
- An appropriate load-equilibrium equation (to describe the relationship between the strain rate in each slice to that in the middle slice)
- A specified boundary condition (e.g., constant crosshead velocity)

For a cavity-free material flow behavior, the simple engineering power-law formulation is often used:

$$\bar{\sigma} = K\bar{\varepsilon}^n \dot{\bar{\varepsilon}}^m \qquad \text{(Eq 29)}$$

in which $\bar{\sigma}$, $\bar{\varepsilon}$, and $\dot{\bar{\varepsilon}}$ denote the effective stress, effective strain, and effective strain rate, respectively, and K, n, and m represent the strength coefficient, strain-hardening exponent, and the strain-rate-sensitivity index, respectively.
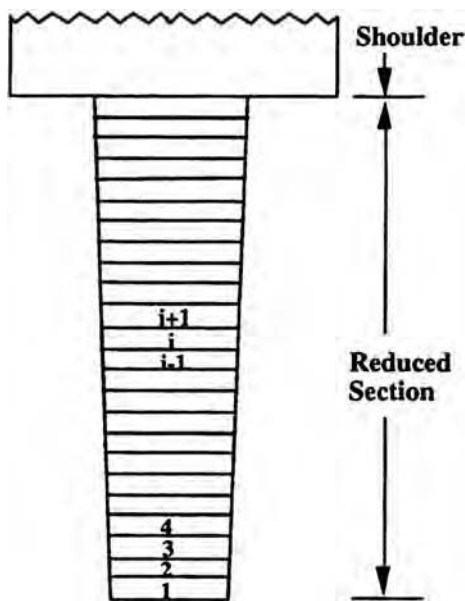
At any instant of deformation, the axial load is the same in each element in order to maintain equilibrium. When cavitation occurs, the load-bearing area, A_{lb}, differs from the apparent, or macroscopic area, A_m. Assuming spherical and uniformly spaced cavities, the relationship between the two quantities (denoting the initial, or uncavitated, area as A_o^{sp}) is:

$$A_m = A_{lb}/(1 - C_v)^{2/3} \text{ and } A_{lb} = A_o^{sp}\exp(-\bar{\varepsilon}) \qquad \text{(Eq 30)}$$

The load borne by each slice is equal to the product of its load-bearing cross-sectional area and axial stress; the axial stress is equal to the flow stress corrected for stress triaxiality due to necking and is evaluated at a strain rate corresponding to that which the material elements experience. The load equilibrium condition is thus described by:

$$\bar{\sigma}_i A_{lb}^i/F_T^i = \bar{\sigma}_j A_{lb}^j/F_T^j \qquad \text{(Eq 31)}$$

in which the subscripts and/or superscripts i and j denote the corresponding parameters for elements i and j, respectively; F_T represents the tri-axiality factor; and A_{lb} is the load-bearing area. The triaxiality factor at the symmetry plane of the neck (as well as away from this plane, at least to a first order) is given by the following equations for round-bar (F_T^r) and sheet (F_T^s) specimen geometries:

$$F_T^r = \left\{ \left[1 + \left(2\frac{U}{w}\right)\right] \ln\left[1 + \left(\frac{w}{2U}\right)\right] \right\}^{-1} \qquad \text{(Eq 32a)}$$

$$F_T^s = \left\{ \left(1 + 2\frac{U}{w}\right)^{1/2} \right.$$
$$\left. \times \ln\left[1 + \frac{w}{U} + \left(\frac{2w}{U}\right)^{1/2}\left(1 + \frac{w}{2U}\right)^{1/2}\right] - 1 \right\}^{-1} \qquad \text{(Eq 32b)}$$
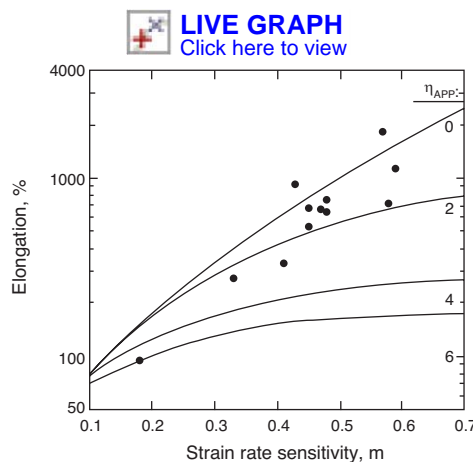
in which w represents the specimen half-diameter or half-width, and U is the profile radius of the neck (Ref 61).

The matrix strain rate, $\dot{\bar{\varepsilon}}$, that is, the strain rate of the cavity-free material, can also be related to the macroscopic sample strain rate, $\dot{\bar{\varepsilon}}_m$ (measured from the rate of change of the specimen dimensions). Using power dissipation arguments, the relationship between the two strain rates is (Ref 57)

$$\dot{\bar{\varepsilon}} = (1/\varphi \, \rho)\dot{\bar{\varepsilon}}_m \qquad \text{(Eq 33)}$$

in which ρ is the relative density of the specimen ($\rho = 1 - C_v$), C_v is the cavity volume fraction, and φ is the stress-intensification factor, which, for spherical and uniformly distributed cavities, is given by $\varphi = 1/\rho^{2/3}$.

Equations 29 to 33 can be used to simulate the isothermal, hot tension test of a cavitating material. Briefly, such simulations comprise the following steps:

1. Specify the initial specimen geometry (width, thickness, length, and initial cavity volume fraction in each element) as well as the overall nominal (or true) strain rate for the tension test. Calculate F_T for each element using Eq 32(a, b).
2. Based on the geometry and local cavity volume fraction, determine the initial strain-rate distribution for each element in the specimen relative to that in the central (reference) element using Eq 31. Determine absolute values of strain rate using the boundary condition. Here, use is made of the material constitutive relationship (Eq 29) with the strain rate from Eq 33.
3. Impose an increment of deformation (based on the strain rate calculated in step 2) and cavitation (based on Eq 5) for each element, and determine the new specimen geometry, F_T distribution, and local cavity volume fractions.
4. Calculate the engineering stress and strain.
5. Calculate a new strain-rate distribution using Eq 31 as in step 2.
6. Repeat steps 3 to 5 until failure

Simulation outputs include predictions of engineering stress-strain curves and the total elongation at failure. For example, typical predictions of tensile elongation for sheet materials that exhibit strain-rate hardening (Ref 60) (with a rate-sensitivity index of m) and no strain hardening are shown in Fig.19. The figure illustrates the increase of ductility with increasing m and decreasing η_{APP}. In addition, by incorporating a fracture criterion (e.g., fracture occurs when the cavity volume fraction in an element reaches 30%), the competition between flow-localization- and cavitation-controlled failure can be quantified and used to construct maps showing the failure mode as a function of m and η_{APP}. For example, a failure-mode map for non-strain-hardening materials is plotted in Fig. 20 (Ref 60). For deformation under superplastic

Fig. 18 Discretization of the sheet specimen for simulation of flow localization and cavitation during isothermal hot tension testing. Source: Ref 60

Fig. 19 Macroscopic model predictions of total elongation as a function of m and η_{APP} for sheet tension testing of samples with a 2% taper and strain-hardening exponent $n = 0$. The individual data points represent measurements taken from the literature. Source: Ref 60

Fig. 20 Failure-mode map developed from macroscopic model simulations of the sheet tension test. Model predictions of the failure mode are compared to experimental observations (data points). Source: Ref 60

LIVE GRAPH
Click here to view

Fig. 21 Micrographs of (a) an orthorhombic titanium aluminide alloy that failed in tension by flow localization (Source: Ref 63) and (b) a near-gamma titanium aluminide alloy that failed in tension by fracture (cavitation) (Source: Ref 64)

conditions ($m > 0.3$) and $\eta_{APP} > 2$, the map shows that failure is fracture/cavitation controlled. On the other hand, flow-localization-controlled failure is seen to predominate only for small values of the apparent cavity-growth rate.

Failure observations for several titanium aluminide alloys (Fig.21) illustrate the types of behavior that can be modeled by this approach. For example, a flow-localization type of failure (Fig. 21a) has been observed for the orthorhombic titanium aluminide alloy Ti-21Al-22Nb deformed at 980 °C and a nominal strain rate of 1.6×10^{-3} s^{-1} (Ref 63). On the other hand, the gamma titanium aluminide alloy Ti-45.5Al-2Cr-2Nb exhibits fracture prior to flow localization during tension testing at 1200 °C and a nominal strain rate of 10^{-3} s^{-1} (Fig. 21b) (Ref 64). The values of m, η_{APP}, and the corresponding failure modes that pertain to these (and similar) experimental observations are cross plotted in Fig. 20. The open data points represent fracture-controlled failures, while the solid data points are flow-localization-controlled failures. The predicted failure modes based on the values of m and η_{APP} show good agreement with the observations.

Hybrid Macro-Micro Modeling. The various features of the macroscopic load-equilibrium model with stress triaxiality and the microspecimen (cavity-coalescence) model described in the previous two sections can be combined to develop a hybrid macro-micro model for the tensile behavior of cavitating materials (Ref 47). In such a hybrid model, the critical strain (ε_{cr}) for intercavity ligament rupture due to internal necking at a microscale is determined from simulations of the deformation of a microtension specimen (Ref 58); these simulations provide ε_{cr} as a function of cavity spacing, d, and material parameters such as the strain-rate-sensitivity index (m) and the cavity-growth rate (η). The steps of this hybrid model are similar to the ones of the macroscopic load-equilibrium approach. The major difference lies in the fact that the load-bearing area is determined from the macroscopic area of each slice of the specimen and the projection of the void area in the horizontal plane of this slice. Therefore, the load-bearing area is not necessarily constant within a particular slice at a given level of tensile elongation. Rather, it is smallest in regions heavily populated with cavities or in regions in which extensive coalescence has occurred. In addition, the location of the minimum load-bearing area within a given slice is not the same at two different levels of deformation because of cavity coalescence. For example, the minimum load-bearing area may have been located in a region with the highest local cavity density during an initial deformation stage. However, because not all cavity coalescences take place simultaneously but rather progress with strain, the minimum load-bearing area at a later stage of deformation may have then been located where the maximum number of coalescences had occurred.
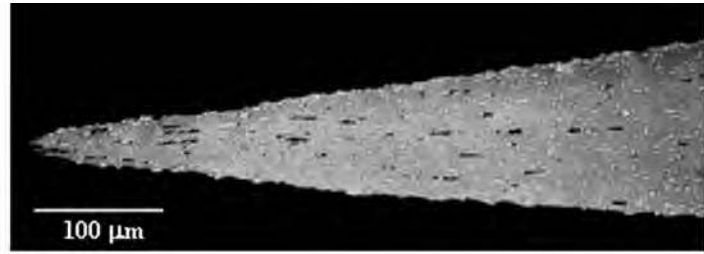
Model predictions of engineering stress-strain curves from both a macroscopic model and a hybrid macro-micro model are presented in Fig. 22 for a non-strain-hardening material, assuming a range of individual cavity-growth rates, η, and a constant (i.e., no new nucleation) cavity density of N = 100 cav/mm^2. The hybrid model predicts that flow localization occurs earlier during the deformation/cavitation process compared to the macroscopic model. Furthermore, flow localization appears to be more predominant in the hybrid model for the cavity-growth rates η considered. In fact, for $\eta = 2$, the shape of the engineering stress-strain curve suggests noticeable flow localization for the hybrid model but a fracture-controlled failure mode for the macroscopic model. The predicted trends are similar for higher initial-cavity densities as well. Such behavior can be ascribed to the fact that the hybrid analysis considers randomly, rather than uniformly, dispersed cavities within the specimen, and the cavity population and coalescence in a particular area affect the load-bearing area (Ref 47). Therefore, the load-bearing area is constant within a slice for the macroscopic model. On the other hand, this area varies even within the same slice for the hybrid model, inasmuch as it depends on the local cavity concentration and/or coalescences that occur in a particular (local) region of the slice. As a result, as deformation proceeds, the load-bearing area is lower in the hybrid model. Hence, the load-bearing area decreases more rapidly with strain in the hybrid model, and the engineering stress is lower and decreases more rapidly with engineering strain.

Summary

The parameters that govern the cavitation process during the hot working of metals have been summarized. Models of plasticity-controlled cavity growth, both for uniaxial and complex (multiaxial) states of stress, are relatively advanced compared to those for the nucleation and coalescence of cavities. In many cases, mesoscale models of plasticity-controlled growth can provide reasonable estimates of the average cavity size. Microscale analyses enable the prediction of the size of the largest cavities by taking into account local microstructure and texture effects. Nevertheless, models of cavity nucleation and coalescence are important with respect to predicting overall tensile ductility. Last, models are being developed to treat the shrinkage of cavities that result from changes in stress state/strain path during complex industrial hot working processes.
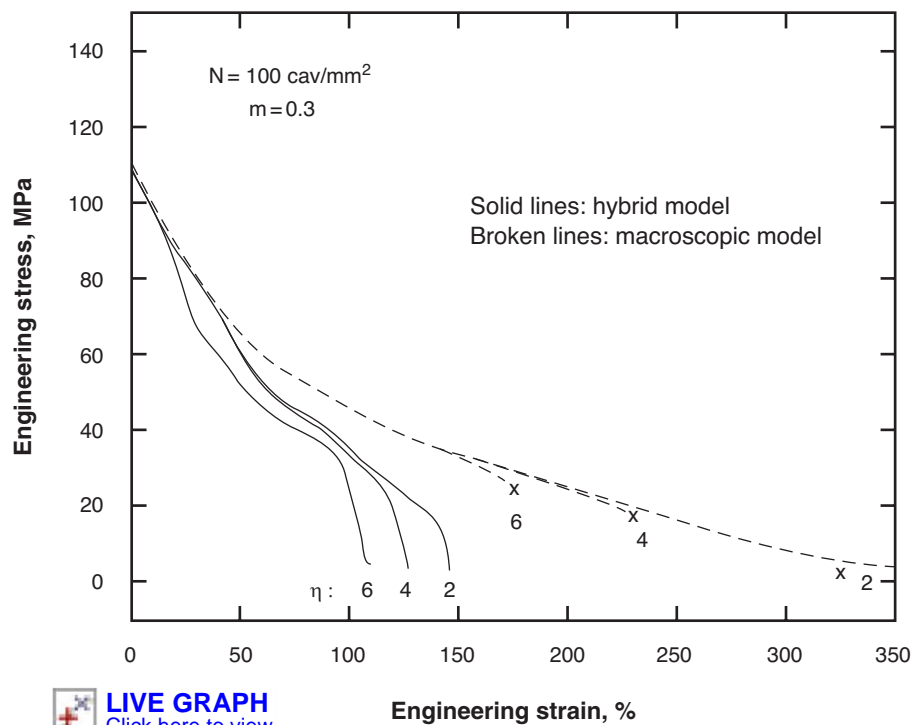
Fig. 22 Hybrid (macro-micro) and macroscopic model predictions of engineering stress-strain curves for $m = 3$, $N = 100$ cav/mm², and various values of η. Source: Ref 47

REFERENCES

1. M.I. Ahmed and T.G. Langdon, *Metall. Trans. A,* Vol 8A, 1977, pp. 1832–1833.
2. S.L. Semiatin, V. Sheetharaman, A.K. Ghosh, E.B. Shell, and P.N. Fagan, *Mater. Sci. Eng. A,* Vol A256, 1998, pp. 92–110.
3. C.C. Bampton and J.W. Edington, *J. Eng. Mater. Tech.,* Vol 105, 1983, pp. 55–60.
4. J. Pilling and N. Ridley, *Superplasticity in Aerospace,* H.C. Heikkenen and T.R. McNelly, Ed., TMS, Warrendale, PA, 1988, pp. 183–197.
5. P.D. Nicolaou, R.L. Goetz, J.D. Miller, and S.L. Semiatin, *Metall. Mater. Trans. A,* Vol 34A, 2003, pp. 2397–2400.
6. P.D. Nicolaou and S.L. Semiatin, *Acta Mater.,* Vol 51, 2003, pp. 613–623.
7. C.W. Humphries and N. Ridley, *J. Mater. Sci.,* Vol 13, 1978, pp. 2477–2482.
8. C.H. Caceres and D.S. Wilkinson, *Acta Metall.,* Vol 32, 1984, pp. 423–434.
9. C.H. Caceres and D.S. Wilkinson, *Scripta Metall.,* Vol 16, 1982, pp. 1363–1365.
10. J. Belzunce and M. Suery, *Acta Metall.,* Vol 31, 1983, pp. 1497–1504.
11. T. Chandra, J.J. Jonas, and D.M.R. Taplin, *J. Mater. Sci.,* Vol 13, 1978, pp. 2380–2384.
12. A.H. Chokshi, *Metall. Trans. A,* Vol 18A, 1987, pp. 63–67.
13. S. Sagat and D.M.R. Taplin, *Acta Metall.,* Vol 24, 1976, pp. 307–315.
14. D.W. Livesey and N. Ridley, *J. Mater. Sci.,* Vol 13, 1978, pp. 825–832.
15. D.W. Livesey and N. Ridley, *Metall. Trans. A,* Vol 13A, 1982, pp. 1619–1626.
16. M.M.I. Ahmed, F.A. Mohamed, and T.G. Langdon, *J. Mater. Sci.,* Vol 14, 1979, pp. 2913–2918.
17. B.-N. Kim, K. Hiraga, Y. Sakka, and B.-K. Jang, *Scripta Mater.,* Vol 45, 2001, pp. 61–67.
18. C.M. Lombard, A.K. Ghosh, and S.L. Semiatin, *Metall. Mater. Trans. A.,* Vol 32A, 2001, pp. 2769–2779.
19. B.P. Kashyap and A.K. Mukherjee, *Res. Mechanica,* Vol 17, 1986, pp. 293–355.
20. G.H. Edward and M.F. Ashby, *Acta Metall.,* Vol 27, 1979, pp. 1505–1518.
21. A.K. Ghosh, D.-H. Bae, and S.L. Semiatin, *Mater. Sci. Forum,* Vol 304–306, 1999, pp. 609–616.
22. L. Dupuy and J.-J. Blandin, *Acta Mater.,* Vol 50, 2002, pp. 3251–3264.
23. M.V. Speight and W. Beere, *Metal Sci.,* Vol 9, 1975, pp. 190–191.
24. A. Needleman and J.R. Rice, *Acta Metall.,* Vol 28, 1980, pp. 1315–1332.
25. A.K. Chokshi, *J. Mater. Sci.,* Vol 21, 1986, pp. 2073–2082.
26. A.C.F. Cocks and M.F. Ashby, *Metal Sci.,* Vol 16, 1982, pp. 465–478.
27. P.D. Nicolaou, S.L. Semiatin, and A.K. Ghosh, *Metall. Mater. Trans. A,* Vol 31A, 2000, pp. 1425–1434.
28. J.W. Hancock, *Metal Sci.,* 1976, pp. 319–325.
29. M.J. Stowell, *Metal Sci.,* 1980, pp. 267–272.
30. J.R. Rice and D.M. Tracey, *J. Mech. Phys. Solids,* Vol 17, 1969, pp. 200–217.
31. P.F. Thomason, *Ductile Fracture of Metals,* Pergamon Press, 1990.
32. F.A. McClintock, *Ductility,* American Society for Metals, 1968, pp. 255–277.
33. F.A. McClintock, S.M. Kaplan, and C.A. Berg, *Int. J. Fract. Mech.,* Vol 2, 1966, pp. 614–627.
34. J. Pilling and N. Ridley, *Acta Metall.,* Vol 34, 1986, pp. 669–679.
35. F.A. McClintock, *J. Appl. Mech., Trans. ASME,* Vol 90, 1969, pp. 363–371.
36. C.A. Berg, in *Proc. Fourth U.S. Natl. Cong. Appl. Mech.,* Vol 2, ASME, New York, 1962, pp. 855–880.
37. W.M. Garrison and N.R. Moody, *J. Phys. Chem. Solids,* Vol 48, 1987, pp. 1035–1074
38. P.D. Nicolaou, A.K. Ghosh, and S.L. Semiatin, *Metall. Mater. Trans. A,* Vol 35A, 2004, pp. 2187–2190.
39. P.D. Nicolaou, R.L. Goetz, and S.L. Semiatin, *Metall. Mater. Trans. A,* Vol 35A, 2004, pp. 655–663.
40. D.H. Bae, A.K. Ghosh, and J.R. Bradley, *Metall. Mater. Trans. A,* 2003, Vol 34A, pp. 2449–2463.
41. P.D. Nicolaou, J.D. Miller, and S.L. Semiatin, *Metall. Mater. Trans. A,* Vol 36A, 2005, pp. 3461–3470.
42. T.R. Bieler, P.D. Nicolaou, and S.L. Semiatin, *Metall. Mater. Trans. A,* Vol 36A, 2005, pp. 129–140.
43. P.D. Nicolaou and S.L. Semiatin, *Metall. Mater. Trans. A,* Vol 7A, 2006, pp. 1567–1574.
44. S.L. Semiatin, F. Montheillet, G. Shen, and J.J. Jonas, *Metall. Mater. Trans. A,* Vol 33A, 2002, pp. 2719–2727.
45. J.J. Blandin, B. Hong, A. Varloteaux, M. Suery, and D. L'Esperance, *Acta Mater.,* Vol 66, 1996, pp. 2317–2326.
46. M.J. Stowell, *Superplastic Forming of Structural Alloys,* N.E. Paton and C.H. Hamilton, Ed., TMS-AIME, Warrendale, PA, 1982, pp. 321–326.
47. P.D. Nicolaou and S.L. Semiatin, *Metall. Mater. Trans. A,* Vol 35A, 2004, pp. 1141–1149.
48. J. Pilling, *Mater. Sci. Techn.,* Vol 1, 1985, pp. 461–466.
49. C.F. Martin, C. Josserond, L. Salvo, J.J. Blandin, P. Cloetens, and E. Boller, *Scripta Mater.,* Vol 42, 2000, pp. 375–381.
50. P.D. Nicolaou and S.L. Semiatin, *Metall. Mater. Trans. A,* Vol 37A, 2006, pp. 3697–3705
51. P.D. Nicolaou, R.L. Goetz, and S.L. Semiatin, *Metall. Mater. Trans. A,* Vol 39A, 2008, pp. 659–665.
52. D.A. Woodford, *Trans. ASM,* Vol 26, 1969, pp. 291–293.
53. J.J. Jonas and B. Baudelet, *Acta Metall.,* Vol 25, 1977, pp. 43–50.
54. F.A. Nichols, *Acta Metall.,* Vol 28, 1980, pp. 663–673.
55. A.K. Ghosh and R.A. Ayres, *Metall. Trans.,* Vol 7A, 1976, pp. 1589–1591.
56. J. Lian and M. Suery, *Mater. Sci. Techn.,* Vol 2, 1986, pp. 1093–1098.

57. D.-J. Zhou, J. Lian, and M. Suery, *Mater. Sci. Techn.,* Vol 4, 1988, pp. 348–353.
58. M. Zaki, *Metall. Mater. Trans. A,* Vol 27A, 1996, pp. 1043–1046.
59. C.M. Lombard, R.L. Goetz, and S.L. Semiatin, *Metall. Trans. A,* Vol 24A, 1993, pp. 2039–2047.
60. P.D. Nicolaou, S.L. Semiatin, and C.M. Lombard, *Metall. Mater. Trans. A,* Vol 27A, 1996, pp. 3112–3119.

61. P.W. Bridgman, *Studies in Large Plastic Flow and Fracture,* McGraw-Hill, New York, NY, 1952.
62. S.L. Semiatin, R.E. Dutton, and S. Shamasundar, *Processing and Fabrication of Advanced Materials IV,* T.S. Srivatsan and J.J. Moore, Ed., TMS, Warrendale, PA, 1996, pp. 39–52.
63. P.D. Nicolaou and S.L. Semiatin, *Metall. Mater. Trans. A.,* Vol 28A, 1997, pp. 885–893.

64. C.M. Lombard, "Superplasticity in Near-Gamma Titanium Aluminides," Ph.D. Thesis, University of Michigan, Ann Arbor, 2001.

Modeling of Cavity Initiation and Early Growth during Superplastic and Hot Deformation

A.K. Ghosh and D.-H. Bae, University of Michigan
S.L. Semiatin, Air Force Research Laboratory

THE FORMATION AND GROWTH of internal voids in metallic alloys are of considerable concern in components produced by superplastic and hot forming processes. Grain-boundary cavitation, observed under temperature and strain-rate conditions for which the boundaries are weaker than the grain interiors, often results from grain-boundary sliding and the concentration of strain around nondeformable particles and hard second phases located at grain boundaries. Strain incompatibility between phases is the primary mode of cavity initiation during elevated-temperature deformation, but sometimes, cavity nuclei pre-exist in multicomponent systems or are generated by the intersection of slip within materials. The pre-existence of cavities is a function of processing conditions to which structural materials are subjected and are thus often dependent on:

- Residual stresses developed during cooling from a high temperature due to the crystallographic anisotropy of misoriented grains/phases, with such phases invariably exhibiting different coefficients of thermal expansion
- Gradients in deformation and temperature that are impossible to avoid during solidification and deformation processing

Pre-existing cavities are usually very small in size. Nevertheless, such cavity nuclei can enlarge during subsequent deformation. Grain boundaries represent a region with material properties that are different from those of the grain interior, and this can create a tendency for grain-boundary shear and sliding under the traction between grains. This condition is illustrated in Fig. 1, which shows that triple-junction regions can develop voids more easily than other regions if grains slide relative to each other like nondeforming blocks. At lower deformation temperatures, at which grain-

boundary regions are not weaker than the grain interiors, the intersection of slip with particles and the intersection of slip from different regions of a crystal can open up voids. These concepts are instructive but somewhat oversimplified, because the deformation of metallic alloys is not uniform everywhere. For example Fig. 2 shows the results of finite-element analysis calculations for the strain distribution in polycrystalline metals composed of strain-hardening grains of different orientations and cohesive grain-boundary layers that exhibit a viscous response. The strain in the material is seen to be nonuniform, even in the absence of nondeformable particles. Strain incompatibility can thus be generated easily, and this too can lead to the nucleation of voids.

A careful examination of the microstructures in engineering alloys, such as those based on aluminum and titanium, has shown that many voids are nucleated due to incompatible deformation between phases (in accordance with the mechanisms suggested by Fig. 1 and 2), although some cavities may pre-exist after primary processing and enlarge during subsequent hot deformation. Even when cavities are very small (e.g., nanoscopic in size), increased levels of local hydrostatic stress may arise from plastic constraint and lead to rapid dilatation of interfacial defects, followed by a slower growth rate due to reduced constraint as the voids expand further. In the remainder of this article, this broad understanding is elaborated.

Early Concepts of Creep Cavitation

In early theories of cavitation in metallic materials, several key ideas were developed for creep deformation and then applied to superplastic and conventional hot deformation. In these approaches, creep at low stresses and the

diffusional contribution to cavity enlargement were considered the most important issues. These metallurgical processes were considered to be slow, permitting rounding of voids. The essential idea was that a minimum stress equal to $2\gamma/r$ (γ is surface tension; r is void radius) must be applied to a solid to keep a void open; otherwise, surface diffusion would close it. Subsequently, cavity-growth models were developed; these analyses focused on stress concentrations along the grain boundaries, with diffusional flow relaxing tensile stresses at cavity tips on the grain boundary and higher tensile stresses developing in regions between cavities. This leads to higher
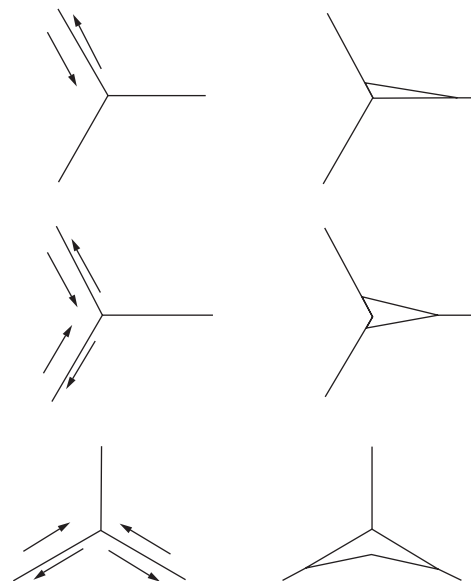
Fig. 1 Schematic illustrations of grain-boundary sliding events that open voids at grain-boundary triple junctions. The simplicity of these sliding events ignores the distribution of strains and stresses that can occur in various parts of a polycrystalline solid.
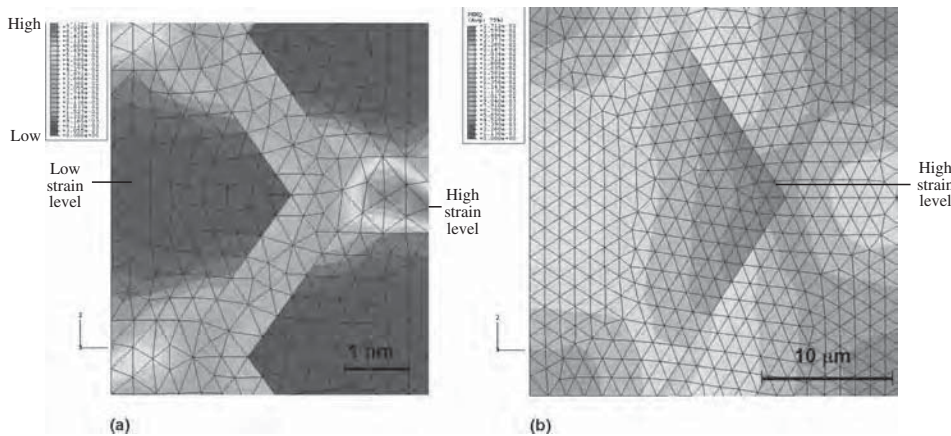
Fig. 2 Calculated distribution of effective strain in polycrystals of different grain sizes under uniaxial tension. (a) Grain size = 4 nm. (b) Grain size = 25 μm. A smaller grain size shows more strain near the grain boundaries, including lobes of shear strain, but the coarser grain size produces higher overall strain within the grain, concentrated more in the vicinity of grain boundaries. Acknowledgment: Calculations performed by Xiang Li
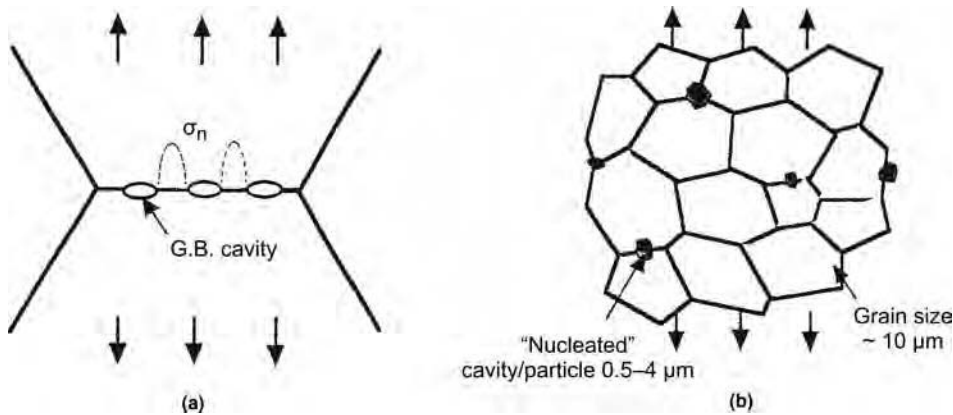
Fig. 3 Schematic illustration of morphologies of grain-boundary (GB) cavity sources. (a) Hull-Rimmer cavitation in creeping solids. (b) Nondeformable particles in a fine-grained superplastic matrix

vacancy concentration between the cavities and vacancy transport from the regions between cavities toward the cavities themselves. The enlargement of the cavities was believed to be primarily due to such vacancy transport, and the rounding of the cavities to be due to surface diffusion. Thus, the overall cavitation process was thought to be dominated strongly by diffusional mechanisms.

Two subsequent developments necessitated re-examination of these ideas for higher-melting-point alloys that were superplastic and alloys that were found to be superplastic at high strain rates at which the diffusional approaches do not produce meaningful results. In particular, hot working and superplastic forming processes take advantage of the enhanced strain-rate sensitivity and formability of materials at elevated temperatures. In forging and forming processes, during which a desired shape is imparted to a workpiece, internal damage in the form of voids often develops in many engineering alloys. High-temperature cavitation models developed initially for the prediction of

creep life in limited-ductility materials with creep-resistant microstructures (Ref 1–6) were extended to material systems exhibiting several hundred to thousand percent tensile elongations at higher strain rates and temperatures suitable for forming applications. Re-examination of the microstructural and phenomenological details of cavitation for the formable alloys (e.g., cavity sources, distribution, size, and volume fractions) show, however, a lack of correspondence with existing creep models, such as that of Hull and Rimmer (Ref 1). For example, in a diffusional void-growth model, a regular distribution of small voids is assumed to be situated on grain boundaries normal to the maximum tensile stress direction, and growth occurs by vacancy transport along the boundary. Voids in formable alloys, however, seldom form as arrays on single boundaries but rather initiate sporadically and are associated with certain nondeformable particles or hard phases whose spacing is of the order of two to five grains or more (Fig. 3). In many alloys, single particles intersect one or two grain-boundary planes.

An unsettling issue with early (diffusion-based) models is the significant underprediction of void size and volume fraction observed with increasing deformation during hot forming, particularly at modest levels of strain (<0.5), for which diffusional cavity-growth processes have been generally accepted to operate (Ref 7). Furthermore, diffusional growth, driven by an effective stress of $\sigma - 2\gamma/r$ (in which σ denotes the applied stress), is unsuitable for nanometer-scale voids, because this stress predicts unrealistic values of the initial void-growth rate, and a physically implausible adjustment of the value of r then becomes necessary.

To overcome the deficiencies of early cavitation models, an alternate approach based on interface-constrained plasticity has been developed and validated. This new model, summarized as follows, explains why the cavity-growth rate is enhanced by increased strain rate, larger particle size, and decreasing forming temperature/lower m-value (strain-rate sensitivity parameter) in metals, features that cannot be quantified by the diffusional-growth mechanism. The predictions of the model are capable of explaining experimental observations, including what appears to be continuous nucleation of cavities during deformation.

Cavitation Observations during Hot Working

To place the basis for the interface-constrained plasticity cavitation model into proper perspective, the microstructural details of the early stages of cavitation are summarized in this section. In particular, careful observations have been made for two materials representing a wide range of microstructure and mechanical properties:

- A modified 5083 aluminum alloy (composition in weight percent: Al-4.5Mg-0.4Cu-0.7Mn) processed to fine-grained, recrystallized sheet (grain size ~7 μm) for superplastic forming applications
- A beta-annealed Ti-6Al-4V plate containing large colonies (>100 μm) of α and β lamellae

The aluminum alloy contained nondeformable hard phases, such as Al_6Mn particles and iron-aluminum intermetallics, at which cavities were found to nucleate (Fig. 4). For the titanium alloy, large α-layers lying along the boundaries of lamellar colonies, having higher strength than the colony structure, create incompatibility for shear. The thin layer of softer β-phase constrained between this layer and the colony-α plates has been found to generate cavitation. The two alloys were tested under uniaxial tension at constant true strain rate (Ref 8) at 450 to 550 °C and 10^{-4} to 10^{-2} s^{-1} (aluminum alloy) or 600 to 850 °C and 10^{-2} to 3 s^{-1} (the titanium alloy), representing superplastic
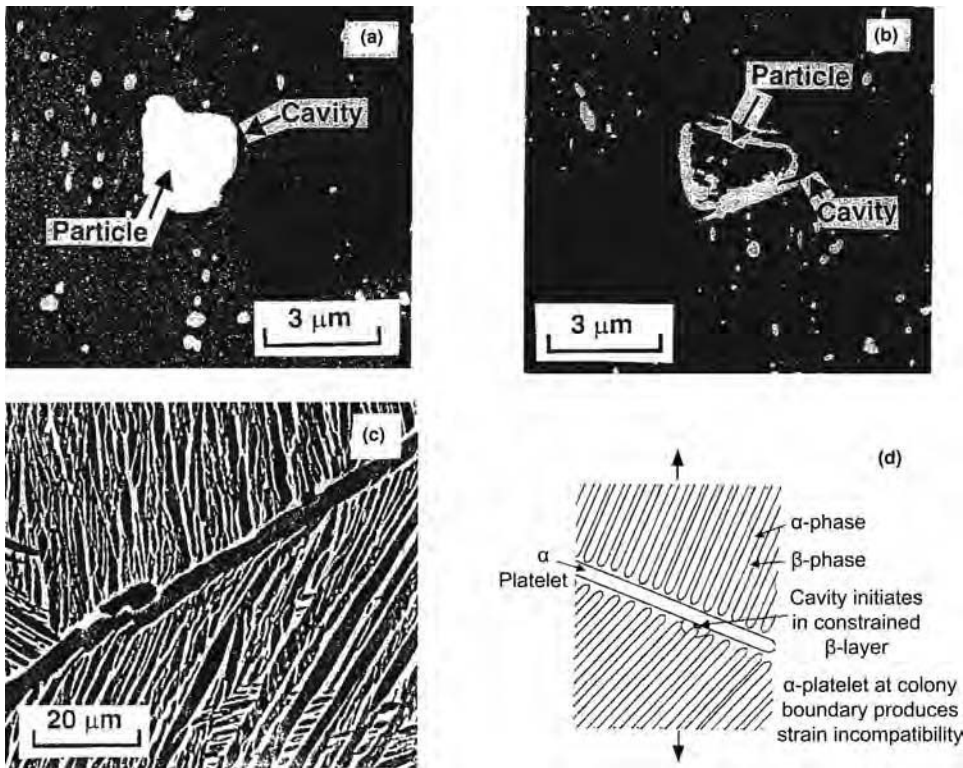
Fig. 4 Cavity initiation at the interface of nondeformable second phases. Al-Mg-Cu-Mn alloy showing (a) a cracklike cavity at the particle interface and (b) complete debonding around a particle (embedded). Ti-6A1-4V showing (c) a cavity at the interface of grain-boundary α-phase and (d) a schematic of the constrained β-layer in which the cavity nucleates

forming and hot forming conditions, respectively. Sectioning of samples after interrupted tests (three orthogonal planes per test condition) revealed the evolution of cavitation with strain, as determined by optical, scanning, and transmission electron microscopy (TEM), that is, at various magnification levels. Quantitative image analysis techniques were used to examine and monitor cavitation with strain. A problem with particle disintegration during sample preparation and debonding from the matrix was found for the aluminum alloy but was minimized by using special polishing procedures.

An analysis of the morphology of voids in the aluminum alloy indicated that they were associated with the particle-matrix interfaces and were cracklike in geometry (i.e., penny-shaped) when they were very small. Also, at inception, the void sizes found by TEM were a few nanometers, but they did not self-sinter/ close even after an exposure at 550 °C. These findings rule out the occurrence of Hull-Rimmer creep in these materials, except perhaps at low strain rates. Furthermore, the aluminum sheet contained some initial voids before recrystallization. The total number was found to decrease somewhat with annealing at 550 °C. However, even after 3 h of annealing, some voids remained, although sintering theory suggested that only 7 to 8 min would be needed for void closure by surface tension at this temperature. While the pre-existing voids before superplastic deformation were few in number, these observations question the effectiveness of surface tension in creating healed and well-bonded interfaces. Partial sintering leaves weak (defective) regions. Such interface defects, as well as constrained, thin regions of matrix between harder phases, are likely sites for subsequent void nucleation. It is, however, possible in some materials to generate good chemical wetting and essentially defect-free interfaces, whereas different sites may contain unsealed nanovoids that can grow in size.

Based on the experimental observations, initial void growth thus appears to occur along particle-matrix interfaces in a manner similar to crack propagation, with only slight extension normal to the interface. Occasionally, near grain boundaries, slight extension along the grain boundary is also possible due to intersection of an interface crack with the boundaries, possibly due to grain-boundary sliding effects, as illustrated in Fig. 1. As shown in Fig. 4 (a and b), when the entire interface is debonded, a void larger than the debonded particle is created; that is, a step jump in void size occurs. The particle that created this larger void easily falls out during specimen preparation. Freshly debonded voids are often faceted and, in rare situations, still hold the particle that generated them. This is understandable from the statistics of the metallographic sectioning process, which

seldom provides an opportunity for the retention of a fully debonded particle within its corresponding cavity.

Voids have also been found to nucleate *continuously* as a function of strain in single- and two-phase metallic alloys, intermetallics, and ceramics (Ref 9–12); that is, more cavities become visible with increasing strain. Figure 5 shows this behavior for the aluminum alloy. Not only do more cavities nucleate with increasing strain, but the rate of continuous nucleation also increases with increasing strain rate and decreasing temperature, both of which can be attributed to the higher flow stress of the matrix under these conditions. Thus, the nucleation of new cavities is believed to be stress-dependent, which is expected on the basis of the stress necessary to break bonds in the regions of interfacial defects.

Any model purporting to describe the early stage of cavitation must therefore address experimental observations of continuous nucleation, cracklike interface cavities, cavity growth from nanometer-scale sizes, and debonding at particle interfaces and formation of large-faceted cavities.

Analysis of Cavitation under Constrained Conditions

Model Formulation. Based on the phenomenology of cavitation described previously, a model has been developed to describe cavity growth from nanometer-scale defects pre-existing at the particle-matrix interface assumed to be strongly bonded, except in a region of the defect that is assumed to be penny-shaped and serve as the cavity nucleus. Enlargement of such cavity nuclei occurs under stresses normal (and/or parallel) to the interface by local plasticity at the tip of the cracklike defect, because a debond cannot support a tensile traction across it. In the well-bonded region away from such a defect, the normal stresses and hydrostatic tension are high (Ref 13), leading to low rates of deformation here, but at the free surface of the defect, stresses are high and *deviatoric* in nature, thus producing high local strain rates. For a stress normal to the interface, the constrained region around the particle is schematically illustrated in Fig. 6(a). The analysis is performed by applying a known stress to the matrix away from the particle and then computing the higher net stress carried by the region containing the particle due to load transfer from the matrix, as in a composite material. This calculation is based on the size of the particle and that of the representative cell surrounding the particle, both of which can be estimated from microstructural observations.

For a thin layer of matrix attached to the particle surface containing the defect, the matrix is assumed to be perfectly bonded to the particle on its bottom edge along an exterior circle (radius = R) (Fig. 6b). The circumferential
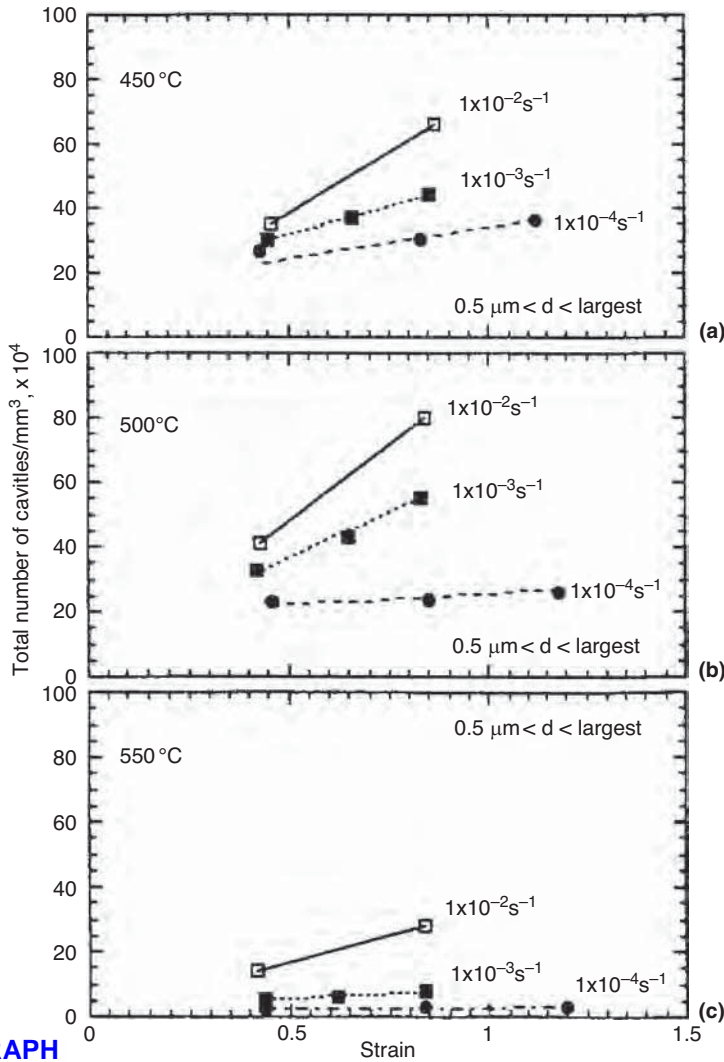
Fig. 5 Total number density of all voids in an Al-Mg-Cu-Mn superplastic alloy as a function of strain at various temperatures and strain rates, showing that voids nucleate continuously during deformation
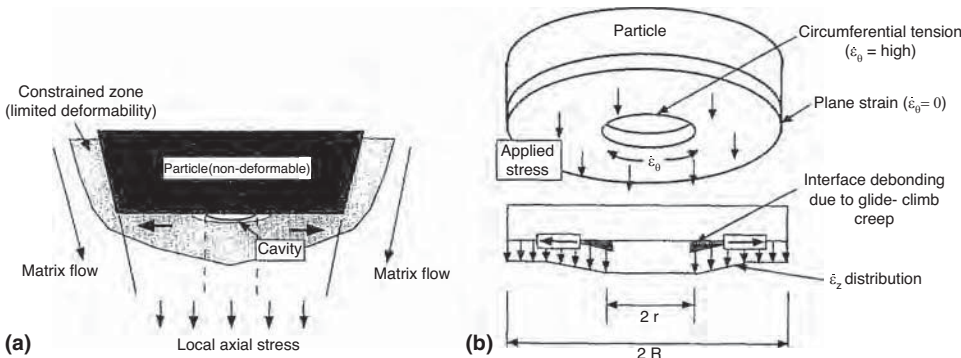
Fig. 6 Cavity growth at matrix/hard particle interface, (a) Flow incompatibility around hard phase, and the development of a constrained matrix zone. (b) Intensified deformation at the tip of a penny-shaped cavity with large circumferential tension

strain rate ($\dot{\varepsilon}_\theta$) along this outer circle is fixed at zero (plane strain), because the particle is assumed to be nondeformable. The metal deforms, and at the free surface of the void, the

strain rates are higher due to a lack of constraint, that is, the radial stress $\sigma_r = 0$, which produces deviatoric stresses and plastic flow. The metal matrix below the thin region of

disbond is larger in area than the penny-shaped cavity that is shown, and hence, its deformation rate is of little consequence to the enlargement of the penny-shaped cavity and is therefore neglected. The deformation of the ring-shaped slab of matrix attached to the particle can be analyzed with the assumption of volume constancy. The tensile strain normal to the interface (ε_z) leads to outward expansion of the hole interior, described therefore as interface-constrained dilatational growth of the cavity. The state of strain at the inner diameter of the cavity (radius = r) is primarily one of circumferential tension, whose magnitude is proportional to $(R/r)^2$; that is, it increases rapidly with particle size R. Even under elastic loading of the overall material, localized plastic strain at the void tip could initiate dislocation glide-and-climb activity that leads to debonding and hence cavity nucleation and growth. There is no impediment to plasticity and growth of the nanocavity on the particle interface for the orientation shown, and its growth rate is predicted by geometry and volume conservation as long as the cavity size is small relative to the particle size R.

The numerical solution of the incremental cavity-extension problem involves a displacement-based approach. The initial height of the slab is assumed to be h_0. After that, height increases due to the axial (z) strain increment, $\Delta\varepsilon_z$, volume constancy, leads to the following relation:

$$\pi(R^2 - r^2)h_0 = \pi(R^2 - r_1^2)h_0(1 + \Delta\varepsilon_z) \quad \text{(Eq 1)}$$

Here, r_1 is the enlarged diameter of the cavity after the axial strain increment. Furthermore, $r_1 = r(1 + \Delta\varepsilon_\theta)$, in which $\Delta\varepsilon_\theta$ is the corresponding increment in circumferential strain at the inner surface of the cavity when the increment $\Delta\varepsilon_z$ is imposed on the thin slab. Combining this expression with Eq 1, one obtains:

$$(1 + \Delta\varepsilon_z) = \frac{(R^2/r^2) - 1}{(R^2/r^2) - (1 + \Delta\varepsilon_\theta)^2} \quad \text{(Eq 2)}$$

After some further algebra, a simplified representation of the cavity-growth description in terms of time derivatives can be derived:

$$\rho = \text{strain-rate ratio} = \dot{\varepsilon}_\theta/\dot{\varepsilon}_z$$
$$= 0.5((R/r)^2 - 1)/(1 + 2\varepsilon_z) \quad \text{(Eq 3)}$$

The detailed strain-rate distribution around the penny-shaped cavity can be estimated using a numerical procedure in which the interface slab is first discretized into a number of elements (Fig. 7). At the edge of the particle ($x_n = R$), $\dot{\varepsilon}_\theta = 0$, and hence, $\rho = 0$. The strain-rate ratio, ρ, increases toward the hole in the thin slab of matrix. For n elements chosen within the matrix, a small $\Delta\varepsilon_z$ is applied at $x_n = R$, and the change in height of this element is calculated, thereby dictating the value of x_{n-1} in the element to its right. The corresponding

stress distribution is calculated by first computing the effective strain rate from ρ as:

$$\dot{\bar{\varepsilon}} = (2/\sqrt{3})(1 + \rho + \rho^2)^{0.5}\dot{\varepsilon}_z$$

and then using a material constitutive relation of the form $\bar{\sigma} = k\dot{\bar{\varepsilon}}^m$ (where K is the strength coefficient, and m is the strain-rate sensitivity exponent) to calculate the effective-stress distribution. The σ_z distribution can then be determined by applying the definition of the von Mises effective stress, which gives $\sigma_z = 1.414$ $\bar{\sigma}[(\beta - \alpha)^2 + (\alpha - 1)^2 + (1 - \beta)^2]^{-0.5}$, where the stress ratios α and β, defined as $\alpha = \sigma_\theta/\sigma_z$ and $\beta = \sigma_r/\sigma_z$, are related to the strain-rate ratio through the associated flow rule, $\rho = (2\alpha - \beta - 1)/(2 - \alpha - \beta)$.

When the cavity grows to a certain size (i.e., is completely debonded from the particle),

continued growth is assumed to occur under unconstrained conditions and is then modeled in accordance with the classical Hancock relation, $d(\ln R)/d\varepsilon \sim \eta/3$, in which η denotes the cavity-growth parameter.

Model Predictions. Figure 8 shows typical model predictions for the growth of a penny-shaped cavity under interfacial constraint. Figure 8(a) shows how the growth of a nanometer-scale defect is strongly influenced by the particle size, R, assuming that the same-sized initial defect exists in each case. If optical microscopy is used to detect voids and its resolution is 0.8 μm, it would be expected that the voids produced by the larger particles would be detectable at strain levels of 3 to 4 % for $m = 0.3$, but cavities resulting from a 0.1 μm diameter particle would not be visible until a much larger strain. Figure 8(b) shows

constrained-growth predictions for materials with $m = 0.3$ and 0.5. Constrained growth is faster for materials with lower m-values. Furthermore, larger initial defect size can cause early visibility (using optical microscopy) and early debonding of the particle. The dashed lines in Fig. 8(b), indicative of *unconstrained* cavity growth, reveal a slower growth rate after debonding (Ref 14) and suggest that voids may not be visible until the imposed strain becomes very large. Thus, depending on the particle size, initial defect size, and dependence of m versus ε curve, particles will debond and enter the unconstrained growth phase at different strains. In other words, an increasing number of voids will become visible, with increasing strain giving rise to what appears to be continuous nucleation.

Figure 9 shows the results of similar calculations for Ti-6A1-4V (with a colony-alpha microstructure) as a function of temperature. The m-value of this material increases with temperature. A smooth transition to unconstrained growth is incorporated through progressive debonding and a gradual loss of load-carrying ability. The predicted strain for visible cavities increases with increasing temperature, in broad agreement with experimental observations.

Model Implications. The continuum analysis described previously is justified even for small cavities (e.g., nanoscale voids), because while flow begins with a few dislocations at the void tip, the dislocation density must be high enough to accommodate the high local strain rates that are developed during cavity growth. The number of participating dislocations increases with increasing void size as the
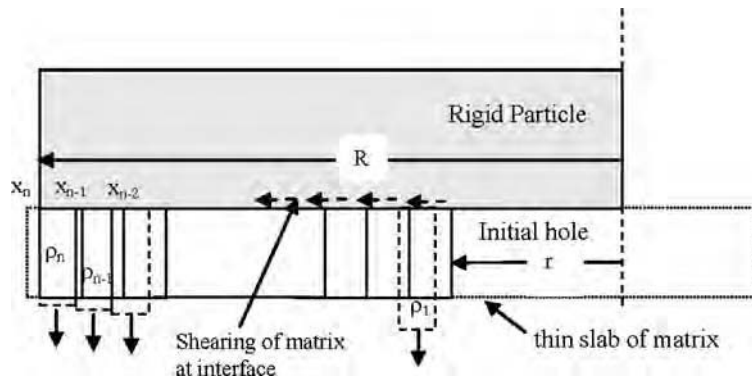
Fig. 7 Schematic illustration of matrix discretization used to estimate the kinetics of cavity growth under constrained-plasticity conditions
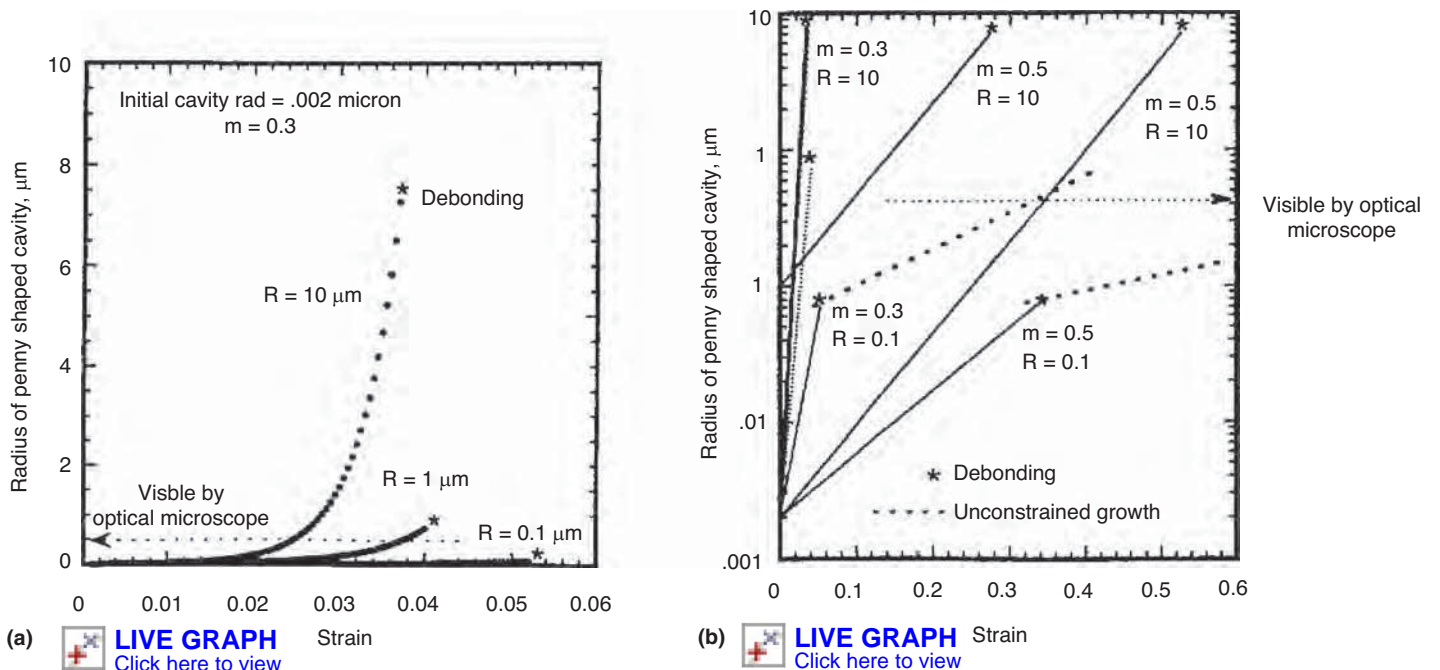
Fig. 8 Predictions of interface-constrained plasticity growth of cavities. (a) Effect of particle size (R) on growth and debonding for m = 0.3. (b) Effect of particle size, initial defect size (1 or 100 nm), and m-value on constrained growth, debonding, and subsequent unconstrained growth

Fig. 9 Predicted rate growth of nanocavities in Ti-6A1-4V titanium alloy as a function of temperature under both constrained and unconstrained conditions

referred to as nucleation, during hot working and superplastic forming of metals is emerging. The key components of this description are:

- Interface defects between a hard phase and the deformable matrix initiate cracklike growth during the early stages of deformation due to local plastic constraint.
- The model associated with the new approach indicates that early growth by interface-constrained plasticity growth is extremely rapid for submicron-sized voids and likely overshadows diffusional growth in superplastic (and nonsuperplastic) alloys.
- Constrained growth through the submicron range is governed by the size of the hard phase, the strain-rate sensitivity of the alloy, and the strain near the hard phase.
- As the interface debonds, the constraint is lost, and general plasticity begins at void sizes in the range of 0.5 to 5 μm. Because voids with sizes below ~1 μm are not readily observable, submicron growth of cavities is not readily observed, and when such cavities do grow to observable size, nucleation is said to occur. So-called continuous nucleation occurs because smaller particles have smaller constrained zones and require more strain to produce complete debonding. Hence, the small voids need more strain to become visible.
- Continuous nucleation effects have been combined with a model of the growth of individual cavities. The predictions for an aluminum alloy are in reasonable agreement with cavitation data obtained from controlled tests. Cavitation rates can increase 2 to 5 times due to continuous nucleation.
- Pre-existing cavities of observable size are few in number. C_{vo}, a measure of initial cavity volume fraction often used in the literature, does not adequately capture the physics of the early stage of the cavitation process.

movement of dislocations produces surface steps at the interface, and the interface delamination process continues (Ref 15). Figure 2 shows another reason to rely upon the continuum-type analysis, inasmuch as it captures and distinguishes between the strain distribution on the nano- *and* micrometer scales. The complexity of molecular dynamics analysis, which does not simulate realistic strain rates for viscous processes, is thus avoided.

Because only one grain-boundary plane intersects a particle in most cases, and the interfacial cavity is not always at a grain boundary, vacancy transport along grain boundaries cannot be a major contributor to the interface-constrained plasticity-growth process or its crack like geometry. In a few instances when the cavity does intersect a grain-boundary plane, a slight extension along the grain boundary may be seen, but not nearly of the magnitude as that which occurs along the interface due to plasticity. It should be borne in mind, however, that while vacancy transport and grain-boundary sliding are unable to accommodate the large circumferential tension near the defect, the far-field stress, strain, and strain rate are still influenced by the sliding contribution inherent in the superplastic constitutive law. The *m*-value ($\delta\log\sigma/\delta\log\dot\varepsilon$) corresponding to the applied strain rate directly influences the force balance across the interface and produces a strong effect on the rate of growth of the penny-shaped void (Ref 14).

The aforementioned model of interface-controlled growth of nanoscopic voids followed by unconstrained growth at larger void sizes avoids the uncertainty of earlier, diffusion-based approaches with regard to the choice of the initial cavity size and cavity-stability considerations using surface tension rules. The new model provides good estimates of cavity nucleation and growth behavior during hot forming and superplastic forming of materials that are consistent with measurements.

Summary

A new outlook on the mechanism and kinetics of early-stage cavity growth, heretofore

REFERENCES

1. D. Hull and D.E. Rimmer, *Philos. Mag.,* Vol 4, 1959, p 673
2. J.E. Harris, *Trans. Metall. Soc.,* Vol 233, 1965, p 1509
3. M.V. Speight and W. Beere, *Met. Sci.,* Vol 9, 1975, p 190
4. J.W. Hancock, *Met. Sci.,* Vol 10, 1976, p 319
5. M.J. Stowell, *Met. Sci.,* Vol 17, 1983, p 1
6. A.H. Chokshi and T.G. Langdon, *Metall. Mater. Trans. A,* Vol 35, 1987, p 1089
7. J. Pilling and N. Ridley, *Superplasticity in Crystalline Solids,* The Inst. Met., 1989, p 131
8. P.A. Friedman and A.K. Ghosh, *Mater. Trans. A,* Vol 27, 1996, p 3030
9. A.K. Ghosh, *Deformation of Polycrystals,* Proc. Second Riso Conference, 1981, p 277
10. A.K. Ghosh and D.-H. Bae, *Mater. Sci. Forum,* Vol 243–245, 1997, p 89

11. K. Hiraga and K. Nakano, *Mater. Sci. Forum,* Vol 243–245, 1997, p 387
12. C.M. Lombard, A.K. Ghosh, and S.L. Semiatin, *Proc. of Conf. on Gamma TiAl,* TMS, 1997
13. J. Orr and D.K. Brown, *Eng. Frac. Mech,* Vol 6, 1974, p 261
14. A.K. Ghosh, D.-H. Bae, and S.L. Semiatin, *Mater. Sci. Forum,* Vol 304–306, 1999, p 609
15. S.X. Mao and A.G. Evans, *Acta Mater.,* Vol 45, 1997, p 4263

SELECTED REFERENCES

- W. Beere, Stresses and Deformation at Grain Boundaries, *Philos. Trans. R. Soc. (London) A,* Vol 288, 1978, p 177–196
- J.W. Edington, K.N. Melton, and C.P. Cutler, Superplasticity, *Prog. Mater. Sci.,* Vol 21, 1976, p 63–170
- G.H. Edward and M.F. Ashby, Intergranular Fracture during Power-Law Creep, *Acta Metall.,* Vol 27, 1979, p 1505–1518
- M.V. Speight, The Role of Grain-Boundary Sliding in the Creep of Polycrystals, *Acta Metall.,* Vol 24, 1976, p 725–729
- M.J. Stowell, D.W. Livesey, and N. Ridley, Cavity Coalescence in Superplastic Deformation, *Acta Metall.,* Vol 32, 1984, p 35–42

Models for Fracture during Deformation Processing

Howard Kuhn, The Ex One Company

IN VERY BROAD TERMS, MODELING OF FRACTURE DURING DEFORMATION PROCESSES has been a major focal point for production engineers since cracking first occurred as a source of scrap in manufacturing. G. Sachs, in his 1954 compilation of articles in *Fundamentals of the Working of Metals* (Ref 1), pointed out that "Two characteristics of any metal or alloy are of particular interest to the engineer engaged in the working of metals. The one is the ability of the metal to withstand permanent or plastic changes in shape or deformations, and it is called ductility... The other metal property of interest to the processing engineer is the resistance of the metal to deformation... For any given material to be processed, both the ductility and the deformation resistance may differ with the variables of processing within very wide limits." In other words, it was recognized early on that the plastic deformation and fracture behavior of materials were functions not only of the material but the processing conditions as well.

Background

At that time, ductility was primarily expressed through measurements of elongation and reduction of area in a tension test. Ongoing testing and everyday experience illustrated that the ability of a metal to tolerate deformation without fracture was usually enhanced by increasing temperature and by decreasing rate of deformation. It was also recognized that ductility was strongly dependent on the size and amount of particles or precipitates dispersed throughout the material. Increasing volume fraction of particles drastically reduces tensile ductility (Ref 2) by providing sources for void initiation, which then grow and coalesce into cracks with increasing plastic deformation.

Process engineers further understood that tensile stresses, or better yet, hydrostatic tension, played a strong role in fracture during metalworking processes. Pioneering work by Bridgman (Ref 3) during WWII showed that deforming materials under high superimposed

hydrostatic pressure dramatically enhanced their ductility, so it was presumed that hydrostatic tension would have the opposite effect. It was further recognized that secondary tensile stresses occur in generally compressive processes, such as the hoop stress due to barreling in axial compression of a cylinder. However, plasticity theory was in its infancy and not yet available in convenient form for everyday practitioners to determine the detailed stress and strain conditions throughout a workpiece undergoing practical metalworking operations.

Early "models" of fracture in metalworking processes assembled these data for a given material, which was expressed through:

- Tabulations (both mental and on paper) of the ranges of temperature and deformation rate at which cracking was likely
- Sketches of crack locations in various processes (Fig. 1), with guidelines (again, both mental and on paper) on their prevention

Experience-based knowledge was the foundation of predictive models for process design to avoid defects. Today (2009), such experience-based knowledge is supplemented by analytical methods using models of fracture to provide a more reliable approach to troubleshooting fracture problems in processes, or to design a process to avoid fracture.

Physical modeling of processes, using soft materials such as plasticene (nonhardening clay) or lead, became popular in the evolution of plasticity analysis. Grid lines were placed on the model material, which was then formed by low-cost tooling that represented the actual process of interest. The deformed grid lines then revealed the pattern of deformation and strains occurring at various points throughout the workpiece. Compressed plasticene cylinders with internally gridded meridian planes were used to determine the degree of inhomogeneity due to various lubrication conditions (Ref 4, 5), and lead billets with internally gridded planes were used to understand the inhomogeneity of
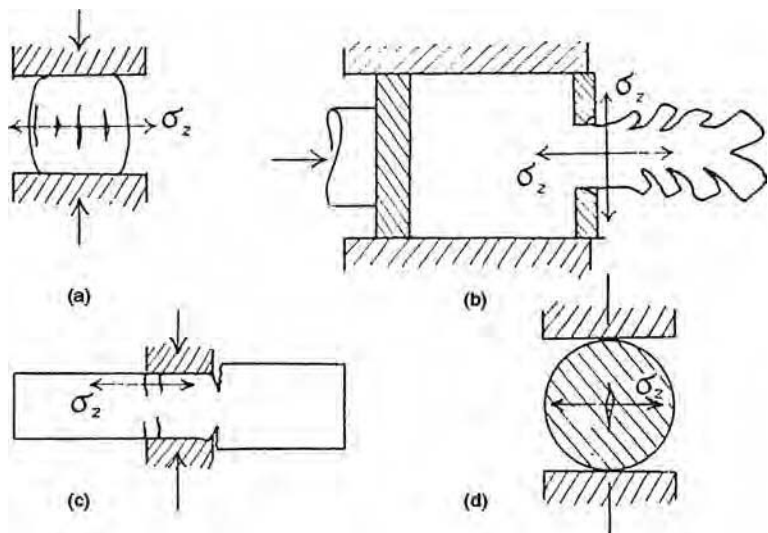
Fig. 1 Sketches of processing defects form a basis for early models. (a) Forging or compression test specimen. (b) Extrusion. (c) Drawing. (d) Forging with no horizontal constraint at die surface Source: Ref 1

plastic deformation during open-die forging and rolling (Ref 6) and extrusion (Ref 7). Plasticene and copper specimens with grid lines were used to interpret the complex distribution of strains in shear spinning (Ref 8). Plasticene with small, embedded metal cylinders has also been used to visualize the localized deformation around inclusions leading to void initiation, growth, and coalescence in ductile fracture (Ref 9). Translation of the strain measurements into stress distributions was accomplished by development of the visioplasticity method. Measured grid displacements were first transformed into strain-field calculations, and the Levy-Mises

plasticity equations were then used to calculate the accompanying stress fields (Ref 10). This method proved tedious but did provide some important insights into the critical stresses in the vicinity of crack formation. Modern tools for digitizing images would streamline the visioplasticity method today, but its use is superseded by computerized finite-element analytical methods, as discussed later.

The slab method of analysis is of little help in modeling fracture during plastic deformation processes because it considers deformation to be homogeneous through the thickness of the workpiece, and it is clear that nonuniformity of

deformation is a key ingredient leading to stress states favorable to fracture. Nevertheless, the method has been used to interpret fracture conditions during compression of fiber-reinforced composite plates (Ref 11). As shown in Fig. 2, compression of a composite plate (25 vol% stainless steel wire in a 6061 aluminum matrix) perpendicular to the fiber direction leads to void formation around the fibers near the plate ends. During compression of a plate, the lateral pressure, q, and normal pressure, p, increase toward the lateral central axis of the plate (Fig. 3); the rate of increase depends on the lumped parameter $\mu L/h$, where μ is the coefficient of friction, and L and h are the instantaneous length and height of the plate. Comparing the lateral positions of the voids with the local stresses calculated by the slab method, the ratio of lateral pressure to vertical pressure, q/p, required to prevent void formation was calculated. Figure 4 shows the
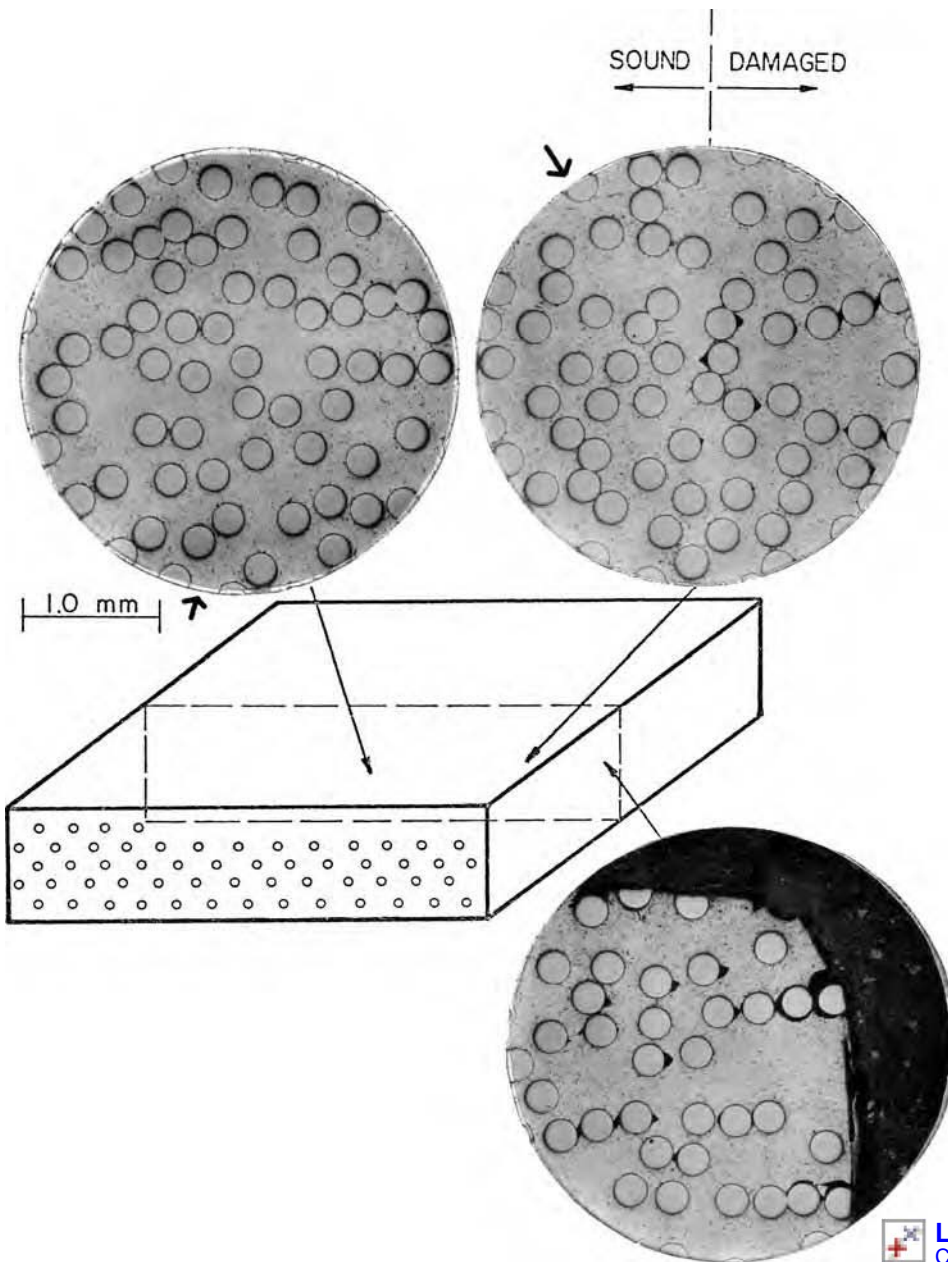
Fig. 2 Macrographs of the polished mid-plane of a composite plate compressed to 50% reduction in height. The innermost evidence of void formation around reinforcing fibers delineates between sound and damaged material. Matrix is 2024 aluminum alloy with 25 vol% NS355 stainless steel wire (0.23 mm diameter) reinforcement. Plates are 6.4 mm thick, 12.7 mm wide and variable length from 6.4 to 19.0 mm. Source: Ref 11
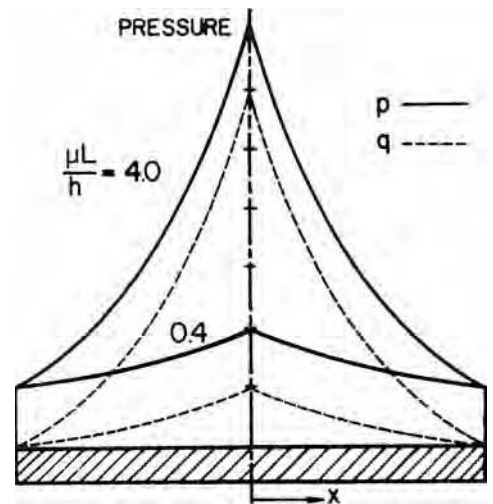
Fig. 3 Distribution of vertical forming pressure, p, and lateral internal pressure, q, in the lateral direction for two different values of $\mu L/h$. Source: Ref 11
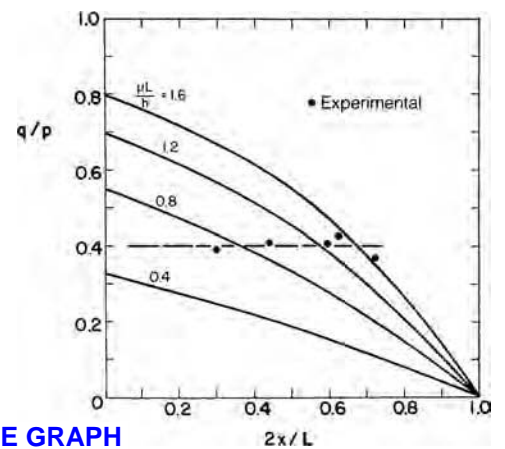
Fig. 4 Values of calculated local lateral pressure to axial pressure ratio (q/p) at measured positions of the transition between sound and damaged material in compressed composite plates of various geometries at 20% reduction in height. Source: Ref 11

experimental results for 20% compression and various forging parameters μL/h. Repeating the tests for various degrees of compression, a forming limit was determined for the composite (Fig. 5).

The upper-bound method of analysis was developed primarily to provide a quick determination of the pressures required in various metalforming operations (Ref 12). Based on equating the external work done to the energy dissipated by an assumed flow field in the workpiece, its first manifestation considered rigid rectangular, triangular, or trapezoidal blocks separated by lines of velocity discontinuity. The method involves searching for the pattern of blocks to represent metal flow that minimizes the overall deformation energy and thus the forming load. While the method is not useful in determining localized stresses for prediction of crack formation, it can be used to determine the process conditions likely to produce defects. Johnson and Kudo (Ref 13) showed that in a double forging/extrusion operation, a flow field that assumed cavity formation at the centerline (known as central burst) would lead to the lowest forming load under certain geometric conditions (Fig. 6). Avitzur (Ref 14) later extended the upper-bound method by using continuous functions to
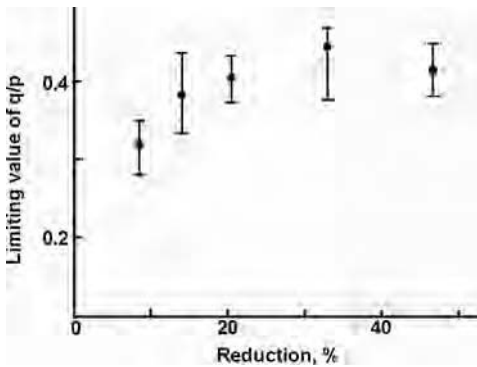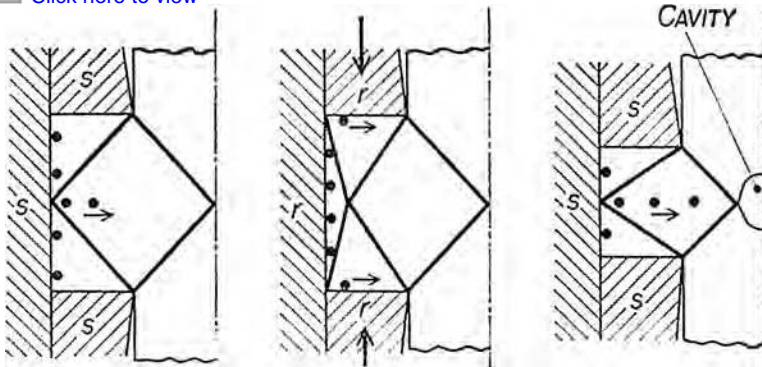
represent the flow field and applying optimization techniques to determine the flow functions that minimize the energy of dissipation. Perturbations can be incorporated in the flow functions to simulate defect formation. In some situations, the external work required to create the flow field with defects is lower than the work required for sound flow, as shown in Fig. 7 for wiredrawing. The calculated external applied load (drawing force) is plotted along the vertical axis as a function of die half-angle. For angles less than α_1, sound flow occurs, and for angles greater than α_2, sound flow occurs with a dead-metal zone forming at the die. Between these two limits, however, a flow field allowing for central burst formation requires a lower load than either of the sound flow modes. Therefore, in this range of die angles, a necessary condition for central burst cracking is satisfied. Repeated calculations using the
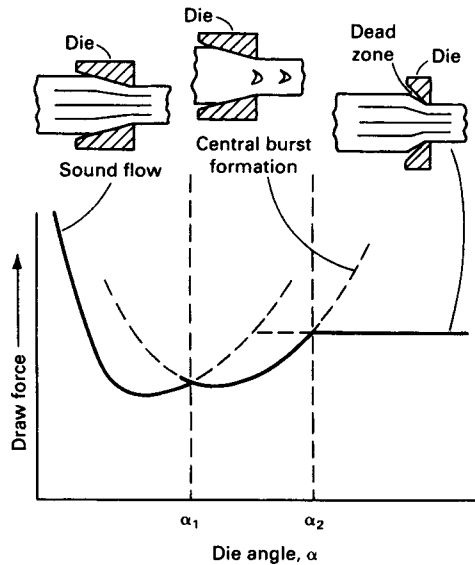
upper-bound method provide the combinations of die angle and reduction that cause central burst (Fig. 8) for various m factors ($0 < m < 1.0$) representing friction along the die surfaces. For a given value of friction, internal cracking is predicted to occur for die angle and reduction combinations below the line, and no defects occur above the line. Note that, for a given die angle, one approach to avoid central burst is to increase the reduction! This example is a clear illustration of the role of process parameters (in this case, geometric conditions) in the occurrence of fracture. It should be repeated that the upper-bound method for defect prediction gives only a necessary condition. The strain-hardening and strain-rate-hardening characteristics of the material can be included in the analysis, but the material microstructural characteristics are not included. Therefore, when operating in the central burst range illustrated in Fig. 8, fracture can occur; whether or not it will occur depends on the material structure (voids, inclusions, segregation, etc.). Experimental validation of the forming limits in Fig. 8 showed that, of 500 extrusions in the central burst zone, 4.5% of the parts had central burst, while none of 500 extrusions in the safe zone showed central burst (Ref 17).

Slip-line field analysis provides a geometric method for calculation of stress states in plastic deformation processes, particularly the hydrostatic stress state. While strictly applicable only to non-work-hardening material and plane-strain deformation, the method provides useful insight into the role of deformation process geometric conditions on the potential for fracture and its location in the workpiece. For example, double indentation by flat punches is a classical problem in slip-line field analysis, illustrated in Fig. 9 for various ratios of

Fig. 5 Forming limit in terms of the limiting pressure ratio q/p to prevent void formation as a function of height reduction. Source: Ref 11

Fig. 7 Drawing force as a function of die half-angle. Central burst formation requires lower external applied load than sound flow in the region between α_1 and α_2. Source: Ref 25

Fig. 6 Upper-bound analysis of a double extrusion-forging process showing sound flow on the left and cavity formation at the centerline on the right. Source: Ref 13

Fig. 8 Upper-bound prediction of central burst in wire drawing in terms of reduction and die half-angle. Increasing friction, expressed by the friction factor, m, increases the defect region of the map. Source: Ref 25

workpiece height to punch width, h/b. Boundaries of the deformation zone change as h/b increases. For $h/b < 1$, the deformation region is spread over an area nearly as large as the punch width. When $h/b > 1$, the deformation zone meets the centerline at a point. For any configuration, the punch pressure and the hydrostatic stress at the centerline can be calculated by traditional slip-line field analysis (Ref 19). Figure 10 shows that both punch pressure and centerline hydrostatic stress increase with increasing h/b. The critical geometry is $h/b = 1.8$ because the hydrostatic stress becomes tensile for h/b ratios greater than this value. The tooling arrangement and deformation geometry in Fig. 9 approximates several other metalworking processes, including rolling, drawing/extrusion, and open-die forging (Fig. 11). For similar h/b ratios in these processes, the stresses throughout the deformation zone can be approximated by those calculated

from slip-line analysis for double indentation. The results of slip-line field analysis for double indentation can be applied to prediction of central burst in drawing and extrusion. For example, in extrusion or drawing, h/b is approximated by:

$$h/b = \alpha[1 + (1 - R)^{1/2}]^2 / R \qquad \text{(Eq 1)}$$

where α is the die half-angle, and R is the area reduction. Taking $h/b = 1.8$ as the critical ratio at which the centerline hydrostatic stress becomes tensile, the relationship between α and R can be calculated. The result is shown in Fig. 12 along with the similar relationship predicted by the upper-bound analysis (Fig. 8). The correlation is remarkable in view of the dissimilarity in die shape between extrusion and double indentation. The similarity in results shown in Fig. 12 indicates that the flow mode for defect formation in the upper-bound method

is physically equivalent to the development of tensile hydrostatic stress at the centerline. The use of slip-line field analysis for fracture modeling was also used in the classic study by Coffin and Rogers (Ref 20). Slip-line field analysis of sheet drawing was used in combination with very carefully executed experiments on

Fig. 10 Variation of the normalized indentation pressure (P/Y, where Y is the yield strength) and the normalized centerline hydrostatic stress (σ_h/Y) with h/b ratio as calculated from slip-line field analysis. Source: Ref 25
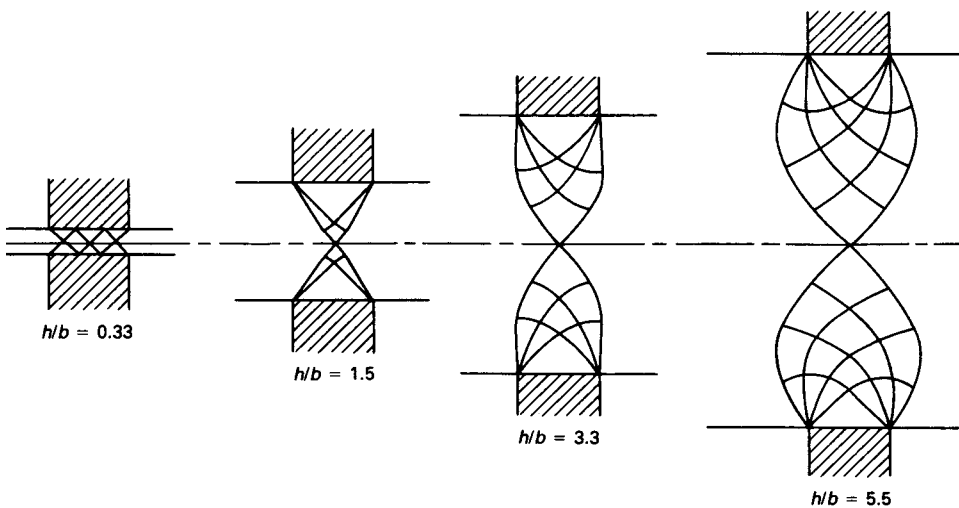
Fig. 9 Slip-line fields for double indentation at various ratios of workpiece height to punch width, h/b. Source: Ref 25
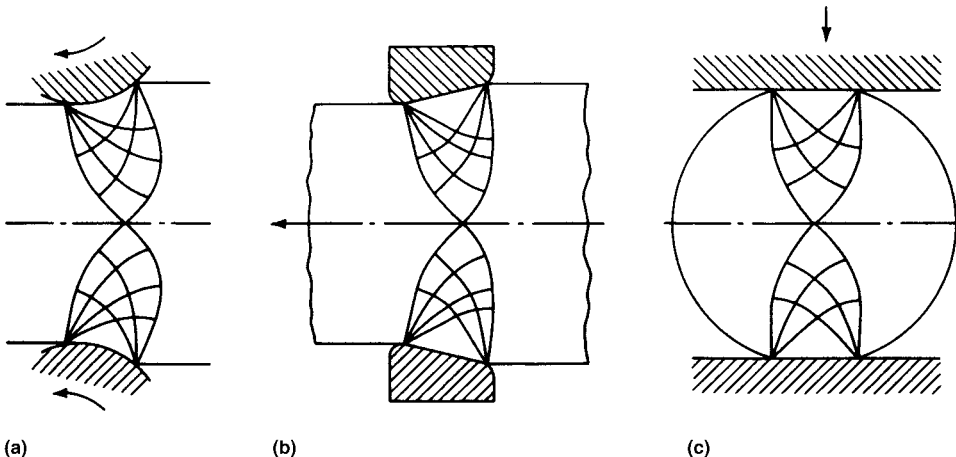
Fig. 11 Slip-line fields for (a) rolling, (b) drawing, and (c) side pressing. These fields are similar to those for double indentation of a thick slab shown in Fig. 9.
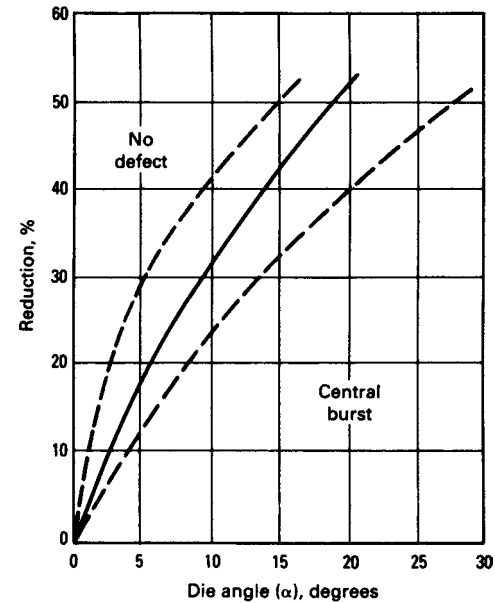
Fig. 12 Prediction of central burst in wire drawing by the tensile stress criterion and slip-line field analysis (solid line) of double indentation. The range of predictions by upper-bound analysis (Fig. 8) is shown by dashed lines. Source: Ref 25

various copper alloys, followed by extensive optical metallography. Results showed that hydrostatic tension at the centerline exclusively led to void formation and slight density reduction, while hydrostatic compression led to zero or slight density increase (Fig. 13).

Fracture Criteria

The fracture models described previously, while providing useful insights, are approximate in their determination of the process or material
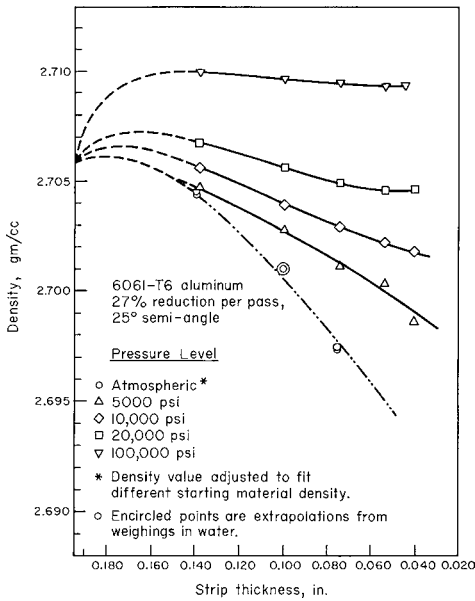
Fig. 13 Density of 6061-T6 aluminum alloy drawn under different externally applied hydrostatic pressures as a function of thickness. Nominal 27% reduction per pass with die half-angle of 25°. Source: Ref 17

(a)

(b)
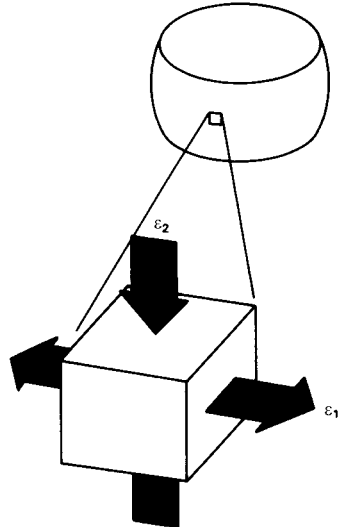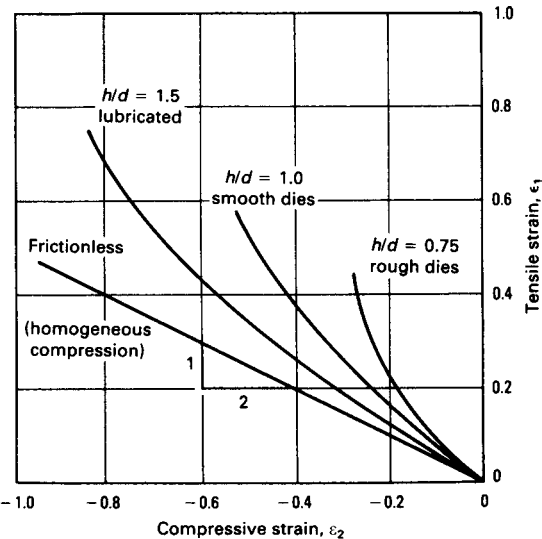
Fig. 14 Localized strains on (a) the bulging cylindrical surface of an upset test and (b) their variation with aspect ratio and friction conditions. Source: Ref 34

conditions leading to fracture. They rely primarily on the expectation that hydrostatic tension is the major culprit leading to fracture but do not address the role of any material characteristics.

Modern finite-element analytical techniques have advanced in four decades from simple linear structures to nonlinear rigid-plastic analysis to elastic-plastic analysis of anisotropic, work-hardening, strain-rate-sensitive materials in complex three-dimensional, large-strain processes. As a result, they are capable of providing detailed, fine-granularity maps of all components of the stress and strain tensors and their combinations, as well as scalar temperatures and their vector gradients, throughout the history of a deformation process. Accurate prediction of fracture during metalworking processes, then, should be accomplishable by inserting a criterion of fracture that faithfully represents the behavior of the material of interest.

General observations of ductile fracture in metalworking can be distilled to two universal conclusions:

- Tensile stress and plastic deformation constitute necessary ingredients for fracture, and control of these localized conditions through modification of the global process parameters (die geometry, workpiece geometry, and friction) can restrain the occurrence of fracture (Ref 1, 3, 11, 13, 20–24).
- Ductile fracture involves the generation, growth, and coalescence of voids, and this sequence of events is dependent on the microstructure and deformation mechanisms prevalent in the material, as well as the localized conditions of stress and deformation (Ref 1, 2, 25–27).

A criterion of fracture for use in finite-element analyses should include these effects implicitly, if not explicitly.

A suitable validation standard should be established for evaluation of the various criteria. Tension testing is not suitable because it produces one basic stress state rather than the variety experienced in metalforming operations. The upset test on cylinders, on the other hand, can generate a variety of stress states on its cylindrical surface through changes in workpiece geometry and interface friction conditions. In addition, a wealth of information is available on this test through research by a number of investigators (Ref 28–33).

Upset compression testing has become a standard for workability evaluation. As shown in Fig. 14, a range of strain combinations can be developed at the cylindrical free surface simply by altering friction and geometry conditions. The influence of friction and consequent bulging on circumferential tensile stress development is clearly shown in Fig. 15. Compression with friction produces circumferential tension that leads to fracture, while frictionless compression prevents barreling, tension, and cracking. Stresses at the cylindrical surface of the upset cylinder are related to the surface strains by:

$$\sigma_1 = {}^4\!/_3\lambda[\varepsilon_1 + {}^1\!/_2\,\varepsilon_2]$$
$$\sigma_2 = {}^4\!/_3\lambda[\varepsilon_2 + {}^1\!/_2\,\varepsilon_1] \tag{Eq 2}$$

where $\lambda = \bar{\varepsilon}/\bar{\sigma}$ is the ratio of effective strain to effective stress. For frictionless (homogeneous) compression, $\varepsilon_1 = -{}^1\!/_2\,\varepsilon_2$, so the circumferential stress σ_1 is zero. That is, for frictionless compression, the stress state is purely compressive (i.e., $\sigma_2 = -Y$, the material flow stress), and the hydrostatic stress is $-Y/3$. For larger absolute ratios of strain, the localized circumferential stress becomes more tensile, while the axial compressive stress becomes less compressive.

Alterations of the compression test geometry have been devised to extend the range of surface strains available toward the vertical, tensile test specimens. Test specimens are essentially prebulged by machining a taper or a flange on the cylinders (Fig. 16). Compression then causes lateral spread of the interior material, which expands the rim circumferentially while applying little axial compression to the rim. Therefore, the tapered and flanged upset test specimens provide strain states consisting of small compressive strain components. Each combination of height, *h*, and thickness, *t*, gives

Fig. 15 Compression tests on 2024-T35 aluminum alloy. Left to right: Undeformed specimen, compression with friction (cracked), compression without friction (no cracks)

a different ratio of tensile to compressive strain. The strain states developed at the surfaces of straight, tapered, and flanged compression test specimens are summarized in Fig. 17.

The variety of strain combinations available in compression tests offers the possibility for material testing over most of the strain combinations that occur in actual metalworking processes. A number of samples of the same material and condition are tested, each one under different friction and geometry parameters. Tests are carried out until fracture is observed, and the local axial and circumferential strains are measured at fracture. Figures 18 to 20 give some examples of results for AISI and SAE 1045 carbon steel, 2024-T351 aluminum alloy at room temperature, and 2024-T4 alloy at a hot-working temperature (Ref 28). In some cases, the fracture strains fit a straight line of slope −½; in others, the data fit a dual-slope line with slope −½ over most of the range and slope −1 near the tensile strain axis. Similar data have been obtained for a wide variety

of materials. In each case, the straight-line behavior (single or dual slope) appears to be characteristic of all materials, but the height of the line varies with the material, its microstructure, test temperature, and strain rate. Similar results have been found, initially by Kudo and Aoi (Ref 29), Kobayashi (Ref 30), Thomason (Ref 31), Ganser et al. (Ref 32), and Bao and Wierzbicki (Ref 33).

The nature of the fracture loci shown in Fig. 18 to 20 suggests an empirical fracture criterion representing the material aspect of workability. The strain paths at potential fracture sites in material undergoing deformation processing (determined by measurement or mathematical analysis) can then be compared to the fracture strain loci. Such strains can be altered by process parameter adjustment, and they represent the process input to workability. If the

process strains exceed the fracture limit lines of the material of interest, fracture is likely.

Fracture models have been devised in an attempt to derive a fracture criterion. Most models are based on the concept of void nucleation and growth to coalescence along bands of high shear stress, as depicted in Fig. 21, originally developed by McClintock (Ref 35) and later by Rice and Tracey (Ref 36). Figure 22 shows a plot of the predicted strain locus at fracture by the McClintock model. The predicted fracture strain line has a slope of −½ over most of its length, matching that of the experimental fracture line. Near the tensile strain axis, the slope of the predicted line is −1, matching that of actual material results shown in Fig. 19 and 20.

In sheet forming, the observation that a neck forms before fracture, even under biaxial stress conditions in which localized instability cannot occur, has prompted consideration of the effects of inhomogeneities in the material. For example, a model of localized thinning due to a small inhomogeneity has been devised (Ref 37). Beginning with the model depicted in Fig. 23, plasticity mechanics is applied to determine the rate of thinning of the constricted region, t_B, in relation to that of the thicker surrounding material, t_A, in region A. When the rate of thinning reaches a critical value, the limiting strains are considered to have been reached, and a forming-limit diagram can be constructed. The analysis includes the effects of crystallographic anisotropy, work-hardening rate, and inhomogeneity size, t_B/t_A.

The R-Z- Models. The Marciniak Kuczynski model (Ref 37) was applied to free surface fracture in bulk-forming processes because of evidence that localized instability and thinning also precede this type of ductile fracture (Ref 38). Two model geometries were considered, one having a groove in the axial direction (Z-model) and the other having a groove in the radial direction (R-model), as shown in Fig. 23. Applying plasticity mechanics to each model, fracture is considered to have occurred when the thin region B, t_B, reduces to zero thickness. When the fracture strains are plotted for different applied stress ratios, a fracture strain line can be constructed. As shown in Fig. 24, the predicted fracture line matches the essential features of the experimental fracture lines (Fig.18 to 20). The slope is −½ over most of the strain range and approximately −1 near the tensile strain axis.

For both the R- and Z-models shown in Fig. 23, the applied strain ratio during any increment of deformation will be defined as:

$$K = -d\varepsilon_z/d\varepsilon_\theta \qquad (Eq\ 3)$$

This gives an increment in equivalent strain in region A, the matrix material:

$$d\bar{\varepsilon}_A = d\varepsilon_\theta(2/\sqrt{3})(1 - K + K^2)^{1/2} \qquad (Eq\ 4)$$

and an applied stress ratio, from the Levy-Mises equations, of:
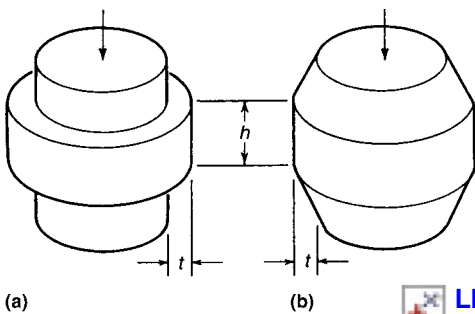
Fig. 16 Depiction of (a) flanged and (b) tapered prebulged compression test specimens. Lateral spread of interior material under compression expands the rim circumferentially while little axial compression is applied (see Fig. 17).

LIVE GRAPH
Click here to view
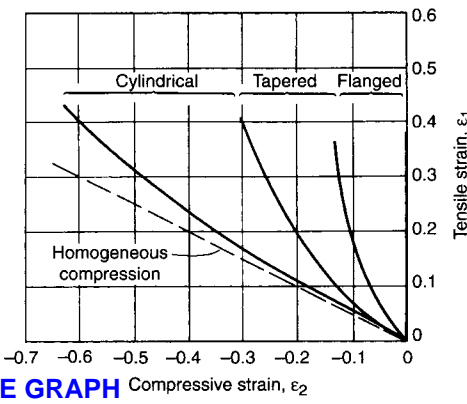
Fig. 17 Range of free surface strain combinations for compression tests having cylindrical (Fig. 14), tapered, and flanged (Fig. 16) edge profiles. The ranges shown are approximate, and they may overlap a small amount.

LIVE GRAPH
Click here to view
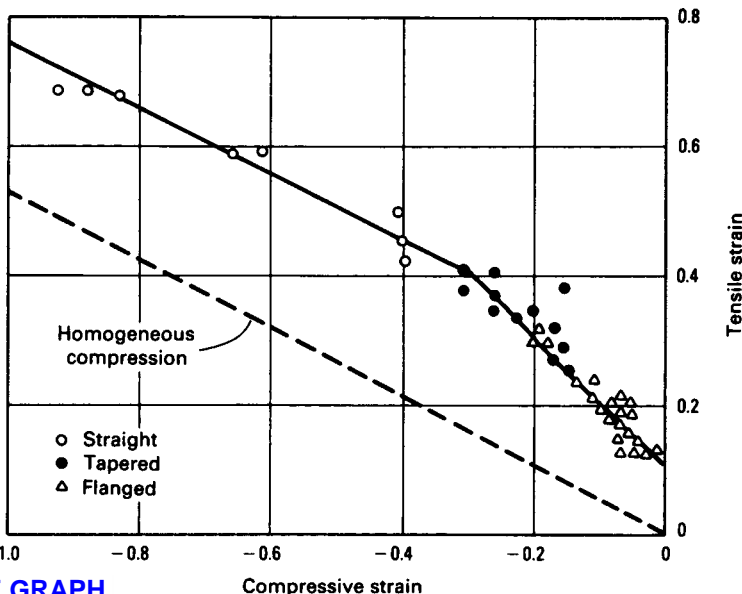
Fig. 18 Fracture locus for AISI 1045 cold-drawn steel

$$r = \sigma_{z_A}/\sigma_{\theta_A} = (2K - 1)/(K - 2) \qquad \text{(Eq 5)}$$

The equivalent stress in region A is then:

$$\bar{\sigma}_A = \sigma_{\theta_A}(1 - r + r^2)^{1/2} \qquad \text{(Eq 6)}$$

The radial stress is taken to be zero because each model applies to an element at or near the surface.

In region B, the groove, equilibrium gives:

$$\sigma_{\theta_B} = \sigma_{\theta_A}(t_A/t_B) = \sigma_{\theta_A}/f \qquad \text{(Eq 7)}$$

where $f = t_B/t_A$

Assuming σ_z is the same in regions A and B:

$$\sigma_{z_B}/\sigma_{\theta_B} = (\sigma_{z_A}/\sigma_{\theta_A})f = rf \qquad \text{(Eq 8)}$$

then the equivalent stress in region B is:

$$\bar{\sigma}_B = \sigma_{\theta_B}(1 - rf + r^2f^2)^{1/2} \qquad \text{(Eq 9)}$$

Now combining Eq 6, 7, and 9:

$$\bar{\sigma}_B = \bar{\sigma}_A[(1 - rf + r^2f^2)/(1 - r + r^2)]^{1/2}/f \qquad \text{(Eq 10)}$$

Using the relation between equivalent stress and equivalent strain, $\bar{\sigma} = C\bar{\varepsilon}^n$, the equivalent stress in Eq 10 is eliminated, giving:

$$\bar{\varepsilon}_B = \bar{\varepsilon}_A\{[(1 - rf + r^2f^2)/(1 - r + r^2)]/f^2\}^{1/2n} \qquad \text{(Eq 11)}$$

Evaluating Eq 11 for each increment in deformation gives $d\bar{\varepsilon}_B$, the increment of effective strain in region B. Then, from the Levy-Mises equations, the incremental strains in region B are:

$$d\varepsilon_{\theta_B} = d\bar{\varepsilon}_B(1 - rf/2)/(1 - rf + r^2f^2)^{1/2} \qquad \text{(Eq 12)}$$

$$d\varepsilon_{z_B} = d\bar{\varepsilon}_B(rf - \tfrac{1}{2})/(1 - rf + r^2f^2)^{1/2} \qquad \text{(Eq 13)}$$

$$d\varepsilon_{r_B} = d\bar{\varepsilon}_B(-rf/2 - \tfrac{1}{2})/(1 - rf + r^2f^2)^{1/2} \qquad \text{(Eq 14)}$$

The change in thickness ratio, $f = t_B/t_A$, for the R-model is given by:

$$df = f(d\varepsilon_{z_B} - d\varepsilon_{z_A}) \qquad \text{(Eq 15)}$$

and for the Z-model:

$$df = f(d\varepsilon_{r_B} = d\varepsilon_{r_A}) \qquad \text{(Eq 16)}$$

This prescribes a new value of f, which, along with the appropriate value of K, is used to calculate the strains for a new increment of deformation using Eq 12 to 14.

It can be expected that the R-model will lead to thinning of the groove for any value of $K < 2.0$, because the resulting tensile stress, σ_θ, will produce greater thinning than would naturally occur under the axial compressive stress, σ_z. At $K = 2.0$, no circumferential tensile stress occurs, and both regions A and B thin the same amount under the axial stress ($df = 0$).

The Z-model will lead to thinning for $0 \leq K < 1.0$. In this domain, σ_θ is greater than $|\sigma_z|$, and the Poisson contraction in the radial direction in the groove due to tensile stress, σ_θ, will be greater than the Poisson extension due to compressive stress, σ_z. At $K = 1.0$, these effects balance so that no thinning occurs, and for $K > 1.0$, the groove increases in thickness more rapidly than the matrix ($df > 0$).

The calculation procedure for the model of localized thinning is relatively simple and can be used to give the strains and f-value explicitly. Calculations in this study were performed on a programmable desk-top electronic calculator. A typical result of the variation of f during deformation is given in Fig. 25.

In an attempt to use these models to predict the forming limits in upsetting of cylinders, calculations were performed on both the R-model and Z-model for values of K ranging between 0 and 2.0. Prestrain (ε_0) was taken as 0.1, and two different n-values (0.1 and 0.25) were used. Initial f_0 values of 0.8, 0.95, and 0.99 were also used. Calculations were terminated when f became zero, which is taken as the limiting condition of fracture (Ref 39).

The Cockcroft-Latham criterion of fracture (Ref 40) is not based on a micromechanical model of fracture but simply recognizes that tensile stress and plastic deformation are the essential macroscopic ingredients of fracture. The Cockcroft-Latham criterion suggests that fracture occurs when the accumulated tensile strain energy reaches a critical value:

$$\int_0^\varepsilon \sigma^* d\varepsilon = C \qquad \text{(Eq 17)}$$

where σ^* is the maximum tensile stress; ε is the equivalent strain; and C is a constant determined experimentally for a given material, temperature, and strain rate. This criterion is an outgrowth of the more general postulate by Freudenthal (Ref 41) that fracture occurs when the plastic work per unit volume reaches a critical value. Other variations of the basic concept have been proposed by Oh et al. (Ref 42), Brozzo et al. (Ref 43), and Bao and Weirzbicki (Ref 44). Gouveia et al. (Ref 45) showed by experiment and finite-element study that the Cockcroft-Latham criterion provided the best predictive capability. The Cockcroft-Latham criterion, by integrating the tensile plastic work over the entire metalworking process, accumulates the damage done

LIVE GRAPH
Click here to view
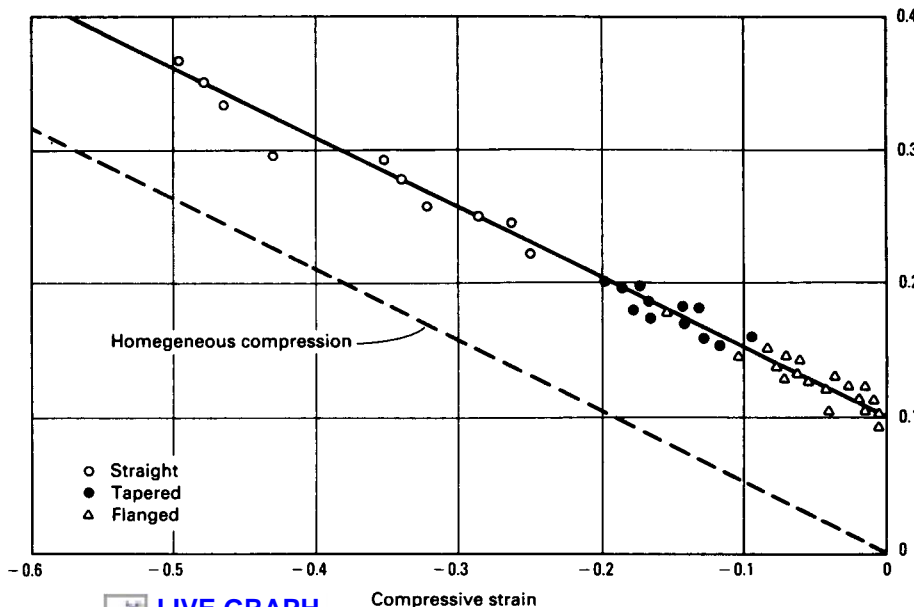
Fig. 19 Fracture locus for aluminum alloy 2024-T351 at room temperature
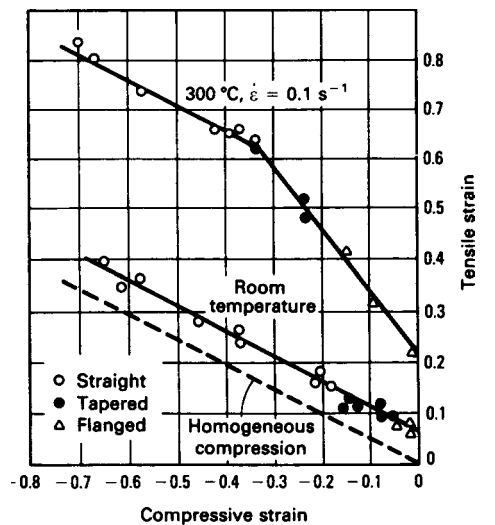
Fig. 20 Fractue locus for aluminum alloy 2024-T4 at room temperature and at 300 °C (570 °F). $\dot{\varepsilon} = 0.1\ \text{s}^{-1}$

LIVE GRAPH
Click here to view
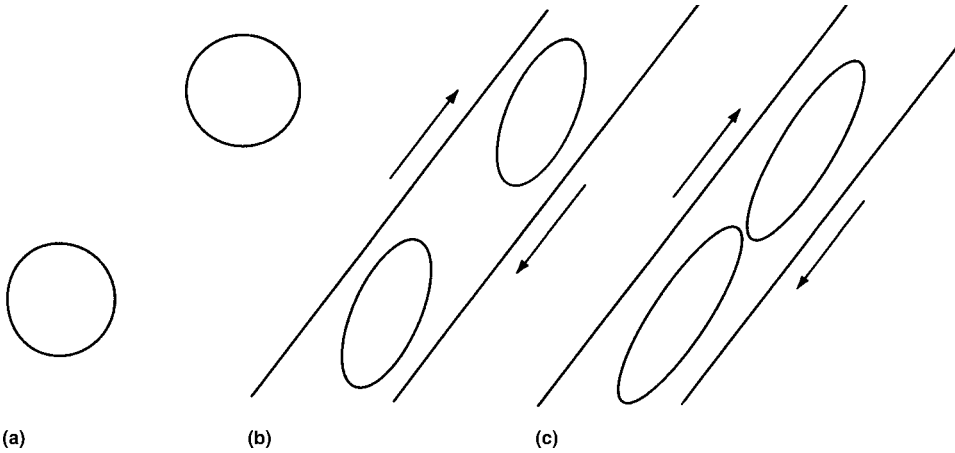
Fig. 21 McClintock model of void coalescence by shear from (a) initial circular voids, through (b) growth, and (c) void contact or coalescence.
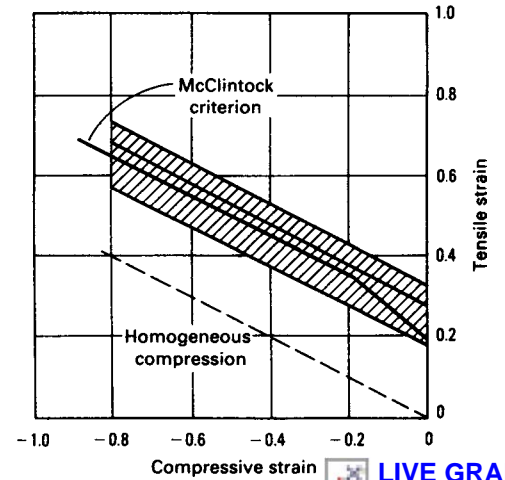
Fig. 22 Fracture strain locus predicted by the McClintock model of void growth. The shaded area represents typical experimental fracture loci such as Fig. 18 to 20
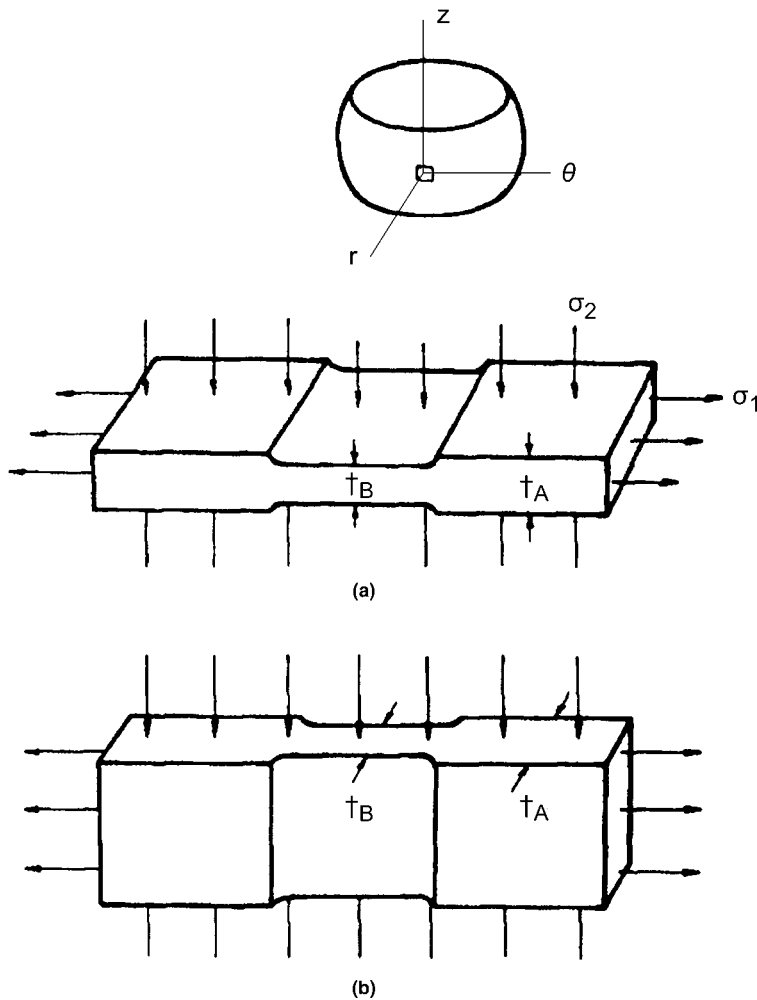
Fig. 23 Material elements used as models for plasticity analysis of localized thinning and fracture of materials. (a) R-model (b) Z-model
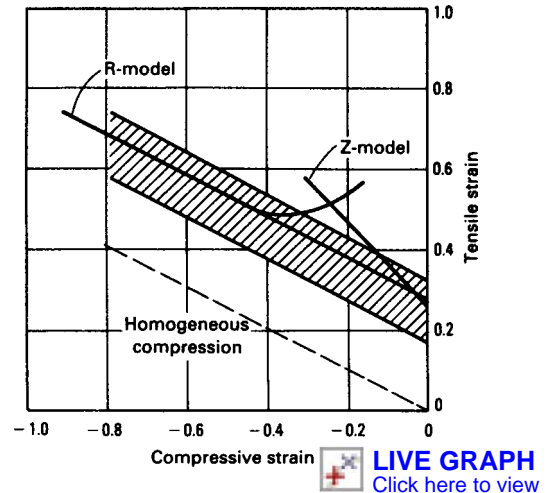
Fig. 24 Fracture strain locus predicted by the model of localized thinning. The shaded area represents typical experimental fracture loci, such as Figs 18 to 20

to the microstructure (microvoid formation, void growth and coalescence) by tensile stresses and plastic deformation so that when it reaches a critical value, C, fracture occurs.

Extension of the Criterion to the Fracture Locus for Compression. The Cockcroft-Latham criterion has been reformulated to provide a predicted fracture line for comparison with the experimental fracture strain line.

This criterion was extended to cover the case of a simple compression. The strains involved in upsetting a cylinder—the basic test used in experimental determination of fracture lines—are illustrated in Fig. 26. In this case, the circumferential stress, σ_θ, is the maximum tensile stress, that is, σ^* in Eq 17.

If the stress ratio is defined as:

$$r = \sigma_z/\sigma_\theta \tag{Eq 18}$$

and the imposed strain ratio is:

$$\alpha = -d\varepsilon_z/d\varepsilon_\theta \tag{Eq 19}$$

then from the Levy-Mises equations:

$$\alpha = (1 - 2r)/(2 - r) \tag{Eq 20}$$

$$\sigma^* = \sigma_\theta = \left[\frac{\overline{\sigma}}{d\overline{\varepsilon}}\right]\frac{d\varepsilon_\theta}{(1 - r/2)} \qquad \text{(Eq 21)}$$

Assuming the power strain-hardening law is valid during the deformation:

$$\overline{\sigma} = K\overline{\varepsilon}^n \qquad \text{(Eq 22)}$$

where $\overline{\sigma}$ is the equivalent stress, $\overline{\varepsilon}$ is the equivalent strain, K is the stress at $\overline{\varepsilon} = 1.0$, and n is the strain-hardening exponent.

The relationship between equivalent strain and the individual strains can be written as:

$$\overline{\varepsilon} = \left[(\varepsilon_r - \varepsilon_z)^2 + (\varepsilon_z - \varepsilon_\theta)^2 + (\varepsilon_\theta - \varepsilon_r)^2\right]^{1/2}\left[\frac{\sqrt{2}}{3}\right] \qquad \text{(Eq 23)}$$

From this equation, and Eq 19, $\overline{\varepsilon}$ can be written in terms of ε_θ and α. The expression involving $\overline{\varepsilon}$, ε_θ, and $\underline{\alpha}$ is:

$$\overline{\varepsilon} = 2\varepsilon_\theta[(1 - \alpha + \alpha^2)/3]^{1/2} \qquad \text{(Eq 24)}$$

From Eq 17, 21, and 24, the following equation was obtained:

$$\int_0^{\overline{\varepsilon}} \sigma^* d\overline{\varepsilon} = C = \left[\frac{2k}{\sqrt{3}}\right]\left[\frac{2}{\sqrt{3}}\right]^n$$

$$\int_0^{\varepsilon_{\theta f}} \left[(1 - \alpha + \alpha^2)^{1/2}\varepsilon_\theta\right]^n (2 - \alpha)d\varepsilon_\theta \qquad \text{(Eq 25)}$$

The value of C can be determined from Eq 25 after integrating and solving for ε_{θ_f}. For strain ratio $\alpha = 0$, the plane-strain case, ε_{θ_f} is equal to the intercept of the fracture line, a. Then:

$$\varepsilon_{\theta_f}(\alpha = 0) = a = \left[\frac{C}{k}\right]^{1/1+n} \cdot 3^{(n+2)/2(n+1)}$$

$$\cdot \left[\frac{1}{2}\right][(n+1)/2]^{1/(1+n)} \qquad \text{(Eq 26)}$$

where a is the intercept on the ε_θ axis of the fracture line, k is the stress at $\overline{\varepsilon} = 1.0$, and n is the strain-hardening exponent.
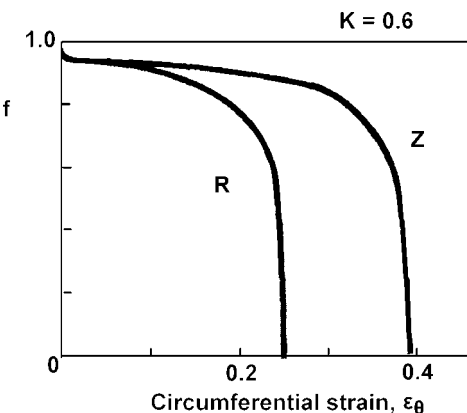
The value of the intercept, a, for various materials was determined experimentally. For 1020 steel and 1045 steel, a was found to be 0.32 and 0.29, respectively. Figure 27 shows that the fracture strain line predicted by the Cockroft–Latham criterion, Eq 17, also is in reasonable agreement with the experimental results. The height of the predicted fracture line is determined by experiment, such as a tension test, rather than by speculation on a material inhomogeneity strength.

Figure 27 shows that the fracture strain line predicted by the Cockroft-Latham criterion is also in reasonable agreement with experimental results. This criterion does not show the dual-slope behavior of the previous models and some actual materials. However, the correspondence between the Cockroft-Latham criterion, the fracture model-based criteria, and the experimental data shows that all of the criteria are generally in agreement with the overall nature of surface strains at fracture in upset compression tests. In addition, the Cockroft-Latham criterion (Eq 17) is the easiest to implement in a finite-element analysis code.

To be a truly useful fracture model, however, the criterion must be applicable to all conditions in metalworking processes, not just the surface fracture in compression tests. Mathematically, the criterion can be extended to three-dimensional stress states by imposing a pressure, P, on the free surface. This has the effect of applying a hydrostatic pressure, P, to all directions because, to maintain yielding and the same strain state, the other two stresses must increase by the magnitude P as well. Then, the forming-limit line becomes modified, as illustrated in Fig. 28. For superimposed pressure ($P > 0$), the fracture line increases in height and its slope increases slightly. For superimposed tension ($P < 0$), the height of the fracture line decreases and curves downward.

The effect of hydrostatic pressure on fracture has been studied extensively.

It has been shown that superimposed pressure increases fracture strain for various materials and test methods. Under high hydrostatic pressure, even brittle materials such as marble have been made to flow plastically.

The fracture model of Lee and Kuhn (Ref 39) can be modified to include hydrostatic pressure and thus determine its effect on the fracture strain line.

The Cockroft–Latham criterion is modified to include the effects of superimposed hydrostatic pressure, because it has been shown that this criterion agrees well with experimental data for free surface fracture. The Cockroft–Latham criterion (Eq 17) proposes that fracture will occur when the critical value, C, is reached.

Modifications to the criterion are made as follows. The deformation strain path for the region where fracture occurs is taken to be linear:

$$\alpha = \frac{d\varepsilon_\theta}{d\varepsilon_r} \qquad \text{(Eq 27)}$$

where $d\varepsilon_r$ and $d\varepsilon_\theta$ are increments of strain in the radial and circumferential directions, respectively. For each increment of deformation, the equivalent strain increment is:

$$d\overline{\varepsilon} = (2/\sqrt{3})(\alpha^2 - \alpha + 1)^{1/2}d\varepsilon_r \qquad \text{(Eq 28)}$$

Individual strain increments can be expressed in terms of stresses by means of the Lèvy–Mises equations:

$$d\varepsilon_\theta = \frac{d\overline{\varepsilon}}{\overline{\sigma}}[\sigma_\theta - \tfrac{1}{2}(\sigma_z + \sigma_r)] \qquad \text{(Eq 29)}$$

$$d\varepsilon_r = \frac{d\overline{\varepsilon}}{\overline{\sigma}}[\sigma_r - \tfrac{1}{2}(\sigma_z + \sigma_\theta)] \qquad \text{(Eq 30)}$$

Equations 27, 29, and 30 are combined to form:

$$d\varepsilon_r(2 - \alpha)/d\overline{\varepsilon} = 3(\sigma_r - \sigma_z)/2\overline{\sigma} \qquad \text{(Eq 31)}$$

$$d\varepsilon_r(1 - 2\alpha)/d\overline{\varepsilon} = 3(\sigma_\theta - \sigma_z)/2\overline{\sigma} \qquad \text{(Eq 32)}$$

Substituting Eq 28 for the value of $d\overline{\varepsilon}$ yields:

$$\sigma_r = \overline{\sigma}(2 - \alpha)/[\sqrt{3}(\alpha^2 - \alpha + 1)^{1/2}] + \sigma_z \qquad \text{(Eq 33)}$$

$$\sigma_\theta = \overline{\sigma}(1 - 2\alpha)/[\sqrt{3}(\alpha^2 - \alpha - 1)^{1/2}] + \sigma_z \qquad \text{(Eq 34)}$$

Fig. 25 Variation of $f = t_A/t_B$ for the R-model and Z-model. $K = 0.6$, $f_0 = 0.99$, $n = 0.25$

AXIAL STRAIN $\varepsilon_z = \ln(h/h_o)$
HOOP STRAIN $\varepsilon_\theta = \ln(w/w_o)$ or $\ln(D/D_o)$

Fig. 26 Schematic diagram of upset test on cylindrical specimens. Surface strains at the equator are measured as shown.
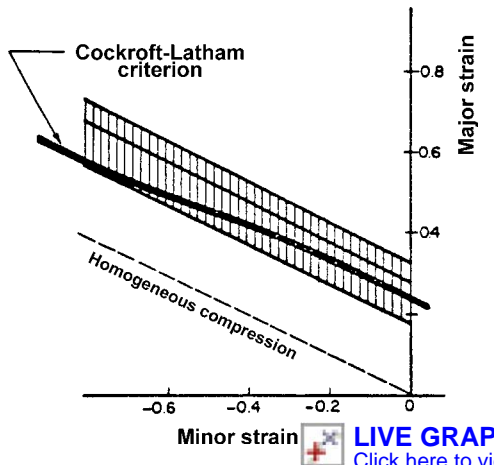
Fig. 27 Fracture strain locus predicted by the Cockroft–Lathan criterion. The shaded region represents the range of fracture lines determined experimentally.

It can be assumed that, as shown in Fig. 28, the z-plane is a free surface, and normal pressure P is applied to it. The effect of normal pressure P is to increase the pressure in all three directions; in other words, it increases the hydrostatic pressure level. Note that the first terms of Eq 33 and 34 are the same as those for deformations without an applied pressure for given value of α, and σ_z appears as an addition in each equation.

From Eq 27 and 33, the Cockcroft-Latham criterion can be modified for superimposed hydrostatic pressure, P, by taking $\sigma^* = \sigma_r$ as:

$$C = \int_0^{\overline{\varepsilon}_f} \sigma^* d\overline{\varepsilon} = \int_0^{\overline{\varepsilon}_f} (\sigma^* - P) d\overline{\varepsilon}$$

$$= \int_0^{\overline{\varepsilon}_f} \left\{ \overline{\sigma}(2-\alpha)/[\sqrt{3}(\alpha^2 - \alpha + 1)^{1/2}] - P \right\} d\overline{\varepsilon}$$

$$= \int_0^{\overline{\varepsilon}_f} \left\{ \overline{\sigma}(2-\alpha)/[\sqrt{3}(\alpha^2 - \alpha + 1)^{1/2}] \right\} d\overline{\varepsilon}$$

$$- \int_0^{\overline{\varepsilon}_f} P \, d\overline{\varepsilon}$$

$$= \int_0^{\varepsilon_{rf}} \left(\frac{2K}{3}\right) \left(\frac{2}{\sqrt{3}}\right)^n (2-\alpha)[(\alpha^2 - \alpha + 1)^{1/2}]^n \varepsilon r^n$$

$$- d\varepsilon r - P \int_0^{\varepsilon_{rf}} \frac{2}{\sqrt{3}} (\alpha^2 - \alpha + 1) d\varepsilon_r$$

$$= \left(\frac{2K}{3}\right) \left(\frac{2}{\sqrt{3}}\right)^n (2-\alpha)(\alpha^2 - \alpha + 1)^{n/2} \frac{1}{n+1} \varepsilon_{rf}^{n+1}$$

$$- P \frac{2}{\sqrt{3}} (\alpha^2 - \alpha + 1)^{1/2} \varepsilon_{rf}$$

$$\text{(Eq 35)}$$

where ε_{rf} is the radial strain at fracture.

For the model material under study, from previous experimental data for K, n, and ε_{rf} at $\alpha = 0$ (plane strain), C was calculated. Then, ε_{rf} was calculated for various values of α between 0 and 2 and P by an iterative method.
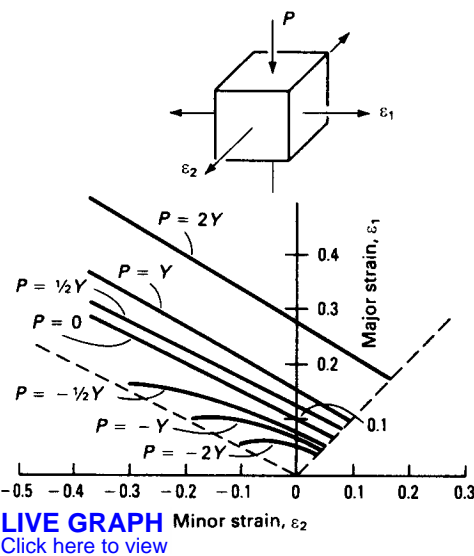
The results are shown in Fig. 29 for various hydrostatic pressure levels (in relation to yield strength, Y). As the level of hydrostatic pressure increases, the intercept increases and the slope of the fracture line increases above one-half. This is consistent with emperical results.

The measured surface strains and hydrostatic pressure levels determined by the visioplasticity method were combined with theoretical forming-limit diagrams developed from the Cockcroft–Latham criterion.

As shown in Fig. 30, at 45% height strain for unlubricated specimens, strains are below the fracture line, and no fracture, in fact, was observed. For the lubricated specimens, the strains are above the fracture line, and fractures were observed.

To evaluate the Cockcroft-Latham criterion for nonfree surface fractures, the double-extrusion forging process (recall Fig. 6) was carried out with a split billet containing a deformation grid (Fig. 31). Measurements of the grid displacements on this midplane were made at several increments of deformation, and the visioplasticity method (Ref 10) was used to calculate the strains and hydrostatic pressure at the midpoint. The hydrostatic stress state at the center of the specimen is initially compressive and then reverses, becoming tensile as the flange thickness is reduced and metal flows into the opposing hubs. Meanwhile, the strains at the center are increasing monotonically as deformation progresses. This is illustrated in Fig. 32 by the steps 0, 1, 2, and 3. As deformation proceeds, the strains at the center increase, but the hydrostatic pressure is also increasing, so the fracture line moves upward. Then, as the flange thickness approaches one-half of the hub base diameter (die orifice diameter), the hydrostatic stress becomes tensile. At this point, the fracture line decreases in height, but the strains at the center continue to rise and cross the fracture line, coinciding with the formation of the central burst. The calculated hydrostatic tension at fracture was 0.3 yield stress (Y). This approach could be used for predicting central burst in drawing and extrusion to provide a material-dependent criterion, as opposed to the more simplistic upper-bound and tensile stress criteria described previously.
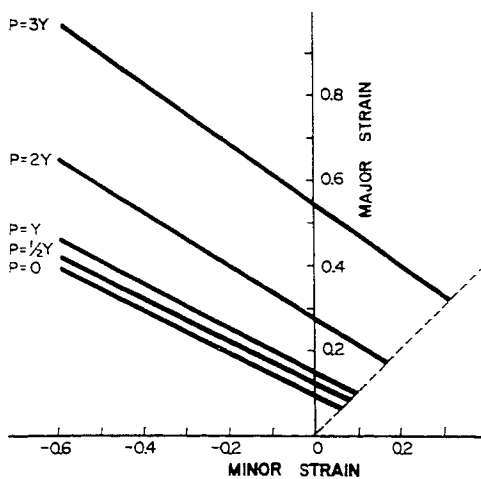
Frequently, cracks occur during forging on surfaces that are in contact with the dies and then move around a die corner (Fig. 33). From the observation of a variety of such defects, it appears that a common characteristic is an abrupt change in frictional shear traction distribution in the region of the crack.

A technique for studying die contact surface cracks was developed by means of a disk compression test using dies having a rough surface in the central region and a smooth surface in the outer region. Figure 34 shows the top view of a 6061 aluminum alloy disk compressed between such dies. In the transition region between the rough central die surface and the smooth outer region, radial cracks initiate and propagate outward. Such cracks occurred at approximately 30% reduction when the smooth outer region was lubricated with a synthetic fluorine-containing resin. The cracks occurred at approximately 45% reduction when grease lubrication was used in the outer smooth region. No cracks occurred even for very large reductions when the smooth outer region was not lubricated.

Grid marks placed on the die contact surface of the disks were used to measure the distribution of surface strains in the radial direction. Figure 35 gives an example of such measurements. In the rough central region, the strains are zero, while in the smooth outer region, the strains are equal and constant. In the transition, however, the circumferential strain, ε_θ,

Fig. 28 Movement of the fracture strain line due to superimposed hydrostatic stress. Applied stress is represented in terms of multiples of the yield strength, Y. Negative values of P indicate hydrostatic tension. Calculations are based on a modification of the Cockcroft–Lathan criterion.

Fig. 29 Fracture strain lines as predicted by the Cockcroft–Lathan criterion in quadrants one and two of the strain plane; P is the superimposed pressure.

Fig. 30 Resulting strains at the transition at 45% deformation and fracture lines for each case of friction condition. The resulting strain crossed the fracture lines for synthetic fluorine-v containing resin- (a) and grease-lubricated (b) cases but is below the fracture line for the unlubricated case (c).
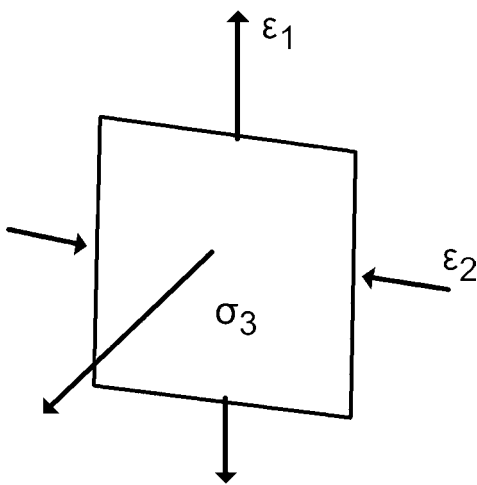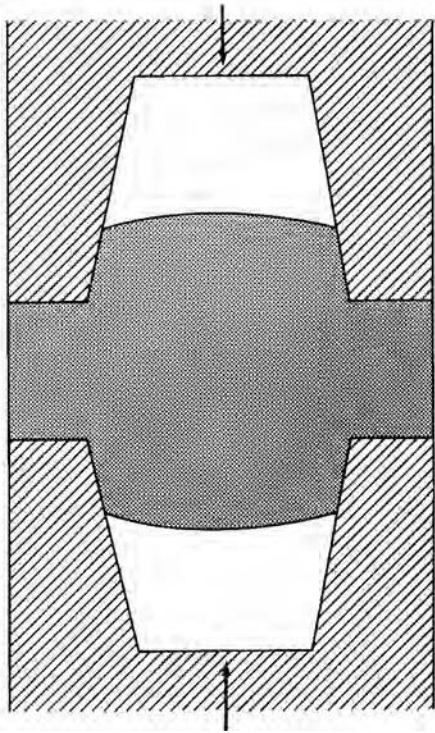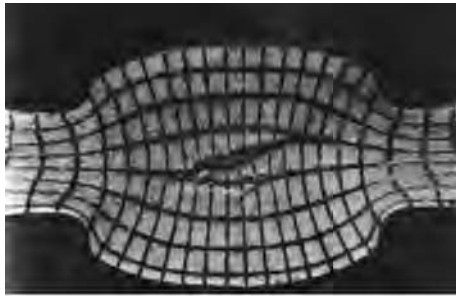
Fig. 31 Internal fracture during the double-extrusion forging of aluminum alloy 6061. Grid deformations on the middle longitudinal plane are shown. The stress-strain states are defined by the insert, σ_3 is perpendicular to the plane of the schematic bulging the free surface.

LIVE GRAPH
Click here to view
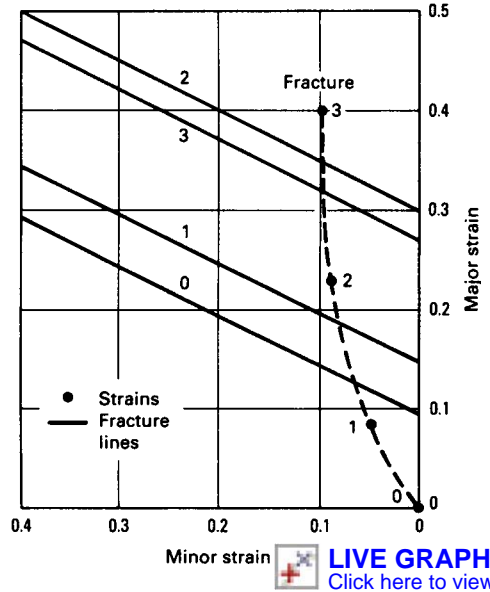
Fig. 32 Progression of surface strains and fracture line at the central internal location of the double-extrusion forging shown in Fig. 31. The fracture line rises from 0 to 1 to 2 as internal pressure increases and then falls to point 3 as internal stress becomes tensile. Meanwhile, the strains at the center continue to rise and cross the fracture line at the deformation stage in which internal fracture occurs.

Fig. 33 Die contact surface cracking during forging extrusion of aluminum alloy powder compact. (a) Cross section. (b) Normal to vertical rib surface. Note also the cracks at the top free surface. The cracks form as material flows from the flange area beneath the punch into the die cavity. Stress-strain states are defined in insert, σ_3 is normal to the free surface.

jumps abruptly from zero to its constant value in the outer region, and the radial strain, ε_r, overshoots to a very high value before returning to its constant value in the smooth outer region. The strains shown in Fig. 35 were the same regardless of the friction condition in the smooth outer region. Therefore, fractures in the transition region occur because of the combination of large tensile surface strains and low hydrostatic stress state. This explains the occurrence of cracks at low reduction when a synthetic fluorine-containing resin is used, and no occurrence of fracture when no lubricant is used. The synthetic fluorine-containing resin, having a near-zero friction coefficient, results in very low radial back pressure on the transition region, while grease and no lubricant provide progressively larger back pressures.

By means of visioplasticity analysis (Ref 10), the stresses were determined at the contact surface in the vicinity of the transition region. The resulting normal die pressure plus the surface radial and circumferential strains define the stress and strain states in the transition region and can be illustrated on a forming-limit diagram. Figure 36 shows the change in surface strains and the increase in the fracture line due to increasing normal pressure during compres-

Fig. 34 (a) Top view of aluminum alloy 6061 disk compressed between dies. (b) Cracks form at the transition region between rough and smooth areas of the die.
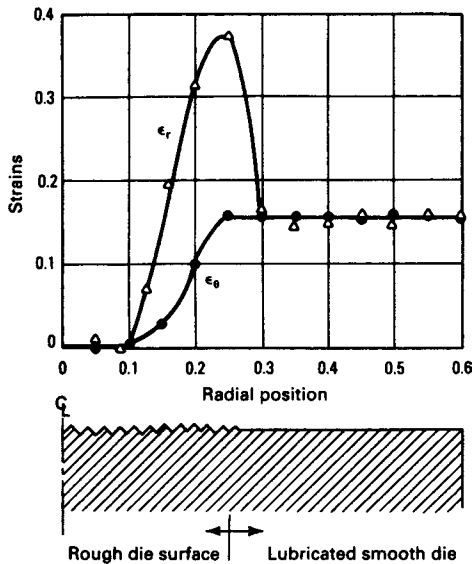
Fig. 35 Radial variation of contact surface strains after 30% compression of the disk shown in Fig. 34

LIVE GRAPH
Click here to view

sion of a disk with grease lubricant in the outer region (indicated by the increments of reduction to 45%). The fracture line increases due to increasing pressure at a slower rate than the strains increase, and at 45% reduction, the strain path exceeds the fracture line, and cracks are observed. For a synthetic fluorine-containing resin lubricant, the crossover occurs at approximately 30% reduction, while in the case
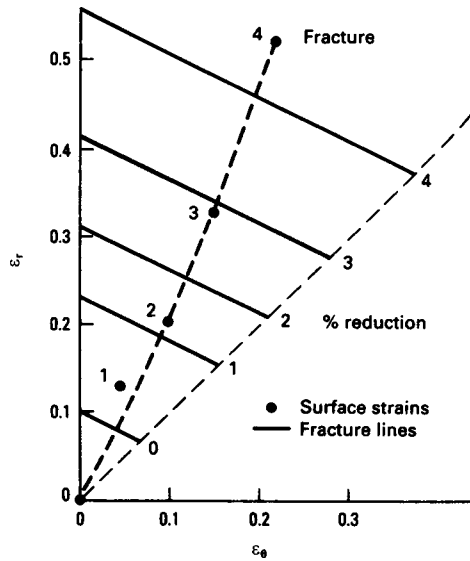
Fig. 36 Progression of surface strains and fracture line at the transition region between rough and smooth zones of the compressed disk shown in Fig. 35. Points 1, 2, 3, and 4 represent 10, 20, 30, and 45% reduction, respectively.

LIVE GRAPH
Click here to view

of no lubricant, the fracture line moves progressively away from the strain path.

While these are crude measurements taken at only a few deformation steps, Fig. 32 and 36 do provide some experimental validation of the Cockcroft-Latham fracture line (Fig. 28), and it would be expected to provide accurate predictions of fracture at any location in a workpiece. A true test of the criterion is to embed it in finite-element codes for prediction of fracture under a variety of metalworking processes. The criterion has become integral to large deformation codes such as DEFORM and MSC.MARC. As a simple example, Fig. 37(a) shows the prediction of fracture at the centerline of a round billet compressed between parallel flat dies (Ref 46, 47), similar to the geometry shown in Fig. 38. The damage value is highest at the center and reaches the critical value for fracture. Figure 37(b) shows the same billet between a three-die arrangement. The new deformation geometry reduces the damage level below the critical value for the material.

A further example of application of the Cockcroft-Latham criterion, the DEFORM code for simulation of plastic deformation processes, was applied to axisymmetric extrusion, as shown in Fig. 39, in which three separate reductions are taken (Ref 48). The Cockcroft-Latham criterion was incorporated with the detailed calculation of stresses and strains throughout the extrusion process to indicate regions in which high microstructural damage occurred. It is clear that as the damage increases, central burst is predicted to occur after the third reduction, which is commonly observed.

Another example, shown in Fig. 40, involves formation of an axisymmetric part formed by

back extrusion. Circumferential cracks form on the inside surface of the part at the location identified in Fig. 41 by point P2. Note that P2 is originally under the punch but then moves around the punch radius and up the sidewall of the part, similar to the example shown in Fig. 33. The calculated damage, according to the Cockcroft-Latham criterion embedded in DEFORM, is shown for P2 in Fig. 42. Note that the damage is very low as long as P2 is under the punch but increases rapidly as soon as it goes around the punch corner and separates from the punch. This is similar to the experimental measurements and calculations carried out in Fig. 34 through 36.

Testing precautions must be used in applying the Cockcroft-Latham criterion through finite-element codes. While the codes are capable of producing faithful reproductions of the distributions of stress and strain throughout the workpiece during a metalworking process, accurate prediction of fracture-occurrence location and time still depends on a faithful representation of the fracture phenomenon through the material factor, C. One approach is to use the upset compression test to establish a fracture line for the material, as in Fig. 18–20, and then perform finite-element analyses of the tests to establish the value of tensile plastic work at which the fractures occur, that is, the value of C.

When considering workability tests, it is important to recognize that fractures initiate in localized regions where interaction between the stress and strain states and the material structure reaches a critical level. Orientation, shape, and volume fraction of inclusions and other microstructural inhomogeneities have a dominant effect on the fracture process. Therefore, it is critically important that workability test specimens contain material having the same microstructural features as the material in the localized, potential fracture regions of the actual process.

Specifically, when evaluating a workpiece for surface fractures, the test specimens used to evaluate C must have surfaces that contain the as-received surface of the workpiece under consideration. The as-received surface may contain laps, seams, a decarburized layer, and so on, which affect fracture initiation. By the same argument, evaluation of material for internal fractures such as central burst must involve test specimens taken from the middle of the workpiece, where, for example, segregation of second phases may have occurred. In this case, bending tests wherein the convex bending surface coincides with the centerline of the workpiece may be the preferred test geometry. Furthermore, because of possible anisotropy effects, orientation of the critical stresses with respect to any inclusion alignment must be the same in the test specimens as it is in the actual process and material of interest. All of these considerations place great importance on correct selection of the test specimen material
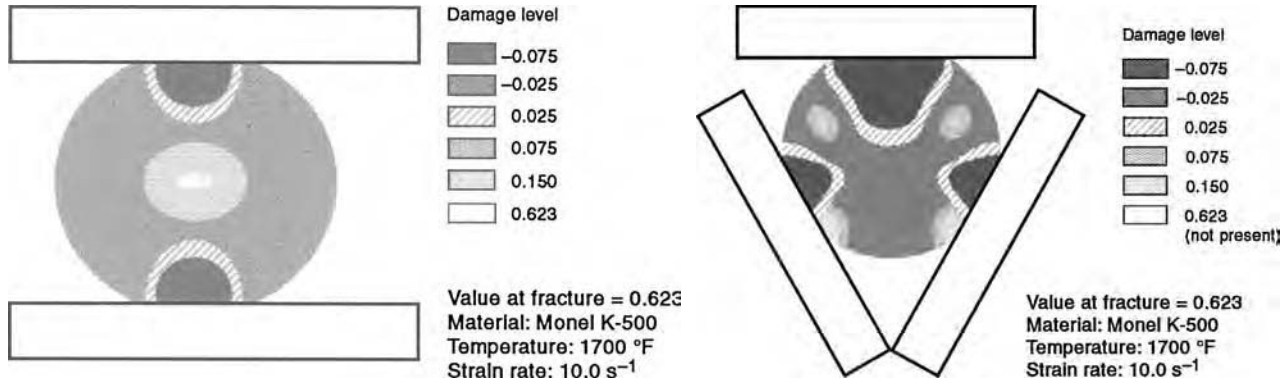
Fig. 37 Damage level modeled for round billet. (a) Model showing damage values corresponding to the process used to manufacture the part shown in Fig. 38. (b) Model showing damage values obtained after modification to the platen used to manufacture the part shown in Fig. 38

Fig. 38 Plasticine ingot representing side pressing of a K-Monel ingot. Hard beads have been embedded to illustrate the presence of tensile stresses at the center of the ingot. Damage seen is directly related to the tensile stresses.

Fig. 39 Prediction of internal damage and central burst by the Cockcroft criterion in the DEFORM large deformation finite element program. Source: Ref 43

and geometry to provide a true representation of the tensile plastic energy to fracture. For this purpose, bending as well as compression tests must be considered in preference to tensile testing, because the failure in a tension test occurs by necking instability and involves material that is at the center of the specimen rather than at the surface.

Fundamental Fracture Models

In the current state of fracture prediction in metalworking processes, computer models provide robust simulations of the deformation, stresses, and strains throughout the workpiece, but the material fracture characteristics must be input from data gathered through properly selected test specimens and procedures. New efforts are being made to provide a prediction of the fracture characteristics of a material from knowledge of its microstructural and crystallographic details.

One approach involves the application of finite-element analysis in a multiscaled approach (Ref 49). Unit cell models are used at each scale to discern the micromechanics of the fracture process. The three scales (Fig. 43) include the atomic distance to assess debonding of particles, the nanometer scale of microvoid-nucleating secondary particles, and the micrometer scale of primary void-forming inclusions. One difficulty with the approach involves the distribution of particles. Fracture generally begins at the weakest link of the material, which would be the location of closely spaced particles. For this reason, modeling is performed around clusters of particles having a higher density of particles than the average throughout the material. Another problem involves accurate representation of debonding at particle-matrix interfaces. A first-principles calculation can be used to determine the bonding energy at the interfaces. Then, the particle-matrix interface is modeled by the finite-element code as one-dimensional cohesive elements. Figure 44 shows an example of microvoid formation and coalescence in two-dimensional shear deformation. A more long-range approach involves application of first-principles-based molecular dynamics calculations to simulate the generation of dislocations around particles, leading to separation of the matrix from the particle and void formation (Ref 50). Figure 45 illustrates a typical sequence of events in the vicinity of a particle.

REFERENCES

1. G. Sachs, *Fundamentals of the Working of Metals,* Pergamon Press, 1954
2. B.I. Edelson and W. M. Baldwin, *Trans. ASM,* Vol 55, 1962, p 230–250
3. P.W. Bridgman, *Studies in Large Plastic Flow and Fracture,* Harvard Press, 1964
4. G.W. Pearsall and W.A. Backofen, "Frictional Boundary Conditions in Plastic Compression," Paper 62-Prod-2, ASME, 1962

Fig. 40 Photograph of a back etruded part. Note the severe circumferential fracture (dark area) and die underfill (below dark area) on the inside diameter (ID).
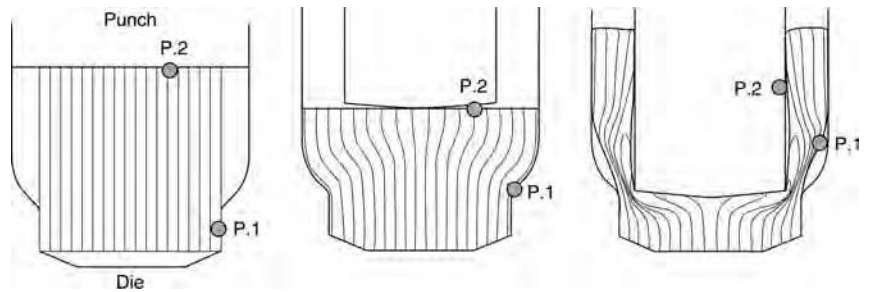
Fig. 41 Flowlines for the forming of the component. Note that point P2 moves from beneath the punch into the wall of the part.
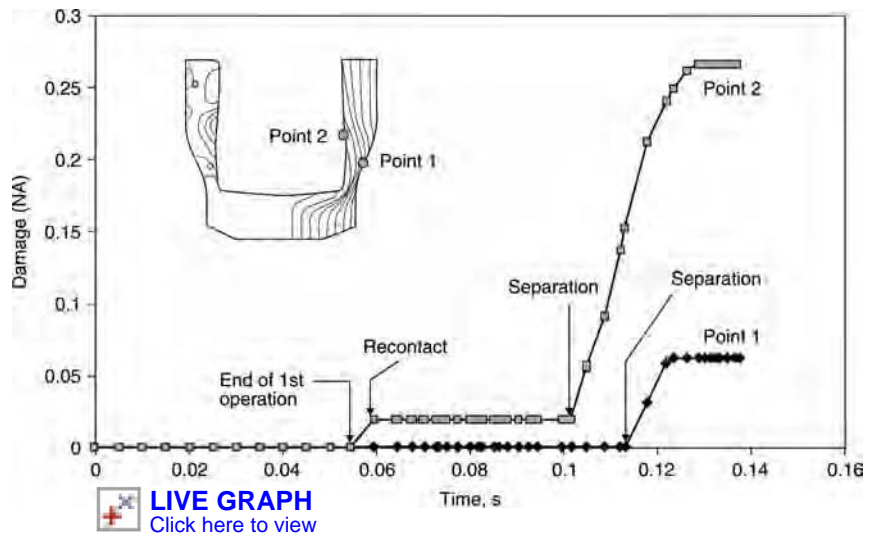
Fig. 42 Point tracking the damage factor of two points on the surface of the component using the Cockroft criterion in the DEFORM large deformation finite element program.

5. M.C. Shaw, Surface Conditions in Deformation Processes, *Fundamentals of Deformation Processing,* Ninth Sagamore Army Materials Research Conference, Syracuse University Press, 1964, p 107–130
6. I.Y. Tarnovskii, A.A. Pozdeyev, and V.B. Lyashkov, *Deformation of Metals during Rolling,* Pergamon Press, 1965
7. C.T. Yang and E.G. Thomsen, Plastic Flow in a Lead Extrusion, *Trans. ASME,* Vol 75, 1953, p 575–579
8. S. Kalpakcioglu, An Experimental Study of Plastic Deformation in Power Spinning, *CIRP Annales,* Vol 10, 1962
9. A.S. Kao and H.A. Kuhn, Physical Modeling of Ductile Fracture in Metalforming Process, *J. Eng. Mater. Technol.,* Vol 112, 1990, p 302–308
10. E.G. Thomsen, C.T. Yang, and S. Kobayashi, *Mechanics of Plastic Deformation in Metal Processing,* The Macmillan Company, 1965, p 178
11. T. Erturk, H.A. Kuhn, and A. Lawley, Forging of Metal-Matrix Composites—Forming Limit Criteria, *Met. Trans.,* Vol 5, 1974, p 2295–2303
12. H. Kudo, An Upper-Bound Approach to Plane-Strain Forging and Extrusion, *Int. J. Mech. Sci.,* Vol 1, 1960, p 57
13. W. Johnson and H. Kudo, *The Mechanics of Metal Extrusion,* Manchester University Press, 1962
14. B. Avitzur, *Metal Forming—Processes and Analysis,* McGraw-Hill, 1968
15. W.M. Evans and B. Avitzur, "Die Design for Drawing Extrusion," Paper MF67-582, Society of Manufacturing Engineers, 1967
16. B. Avitzur, "Strain Hardening and Strain Rate Effects in Plastic Flow Through Conical Converging Dies," Paper 66-Prod-17, American Society of Mechanical Engineers, 1966

17. Z. Zimmerman, H. Darlington, and E.H. Kottcamp, Jr., Selection of Operating Parameters to Prevent Central Bursting during Cold Extrusion, *Mechanical Working and Steel Processing,* TMS, 1970, p 405
18. R. Hill, On the Inhomogeneous Deformation of a Plastic Lamina in a Compression Test, *Philos. Mag.,* Vol 41, 1950, p 733
19. R. Hill, *Mathematical Theory of Plasticity,* Oxford Press, 1950, p 128
20. H.C. Rogers and L.F. Coffin, "Investigation of the Nature of Structural Damage in Metal-Forming Processes," Final report, Contract NOw-66-0546-d, Naval Air Systems Command, June 1967
21. R. Wright, Workability and Process Design in Extrusion and Drawing, *Handbook of Workability and Process Design,* ASM International, 2003, p 316–321
22. A. Chaudhary and S. Vaze, Design Optimization for Dies and Preforms, *Metalworking: Bulk Forming,* Vol 14A, *ASM Handbook,* ASM International, 2005, p 640–644
23. W.A. Backofen, Some Unifying Aspects of Deformation Processing, *Fundamentals of Deformation Processing,* Ninth Sagamore Army Materials Research Conference, Syracuse University Press, 1964, p 1–10

24. H.A. Kuhn, G.E. Dieter, and S.L. Semiatin, Workability and Process Design—An Introduction, *Handbook of Workability and Process Design,* ASM International, 2003, p 3–20
25. K.E. Puttick, Ductile Fracture in Metals, *Philos. Mag.,* Vol 4, 1959, p 964
26. G.E. Dieter, Introduction to Ductility, *Ductility,* ASM International, 1967, p1–27
27. H.C. Rogers, The Mechanism of Crack Propagation in Ductile Metals, *Acta Metall.,* Vol 7, 1959, p 750
28. H.A. Kuhn, Workability Theory and Application in Bulk Forming Processes, *Handbook of Workability and Process Design,* ASM International, 2003, p 172–187
29. H. Kudo and K. Aoi, Effect of Compression Test Condition upon Fracturing of a Medium Carbon Steel—Study on Cold-Forgeability Test: Part II, *J. Jpn. Soc. Technol. Plast.,* Vol 8, 1967, p 17–27
30. S. Kobayashi, Deformation Characteristics and Ductile Fracture of 1040 Steel in Simple Upsetting of Solid Cylinders and Rings, *J. Eng. Ind. (Trans. ASME),* 1970, p 391–399
31. P.F. Thomason, Tensile Plastic Instability and Ductile Fracture in Uniaxial Compression Tests, *Int. J. Mech. Sci.,* Vol 11, 1968, p 187–202
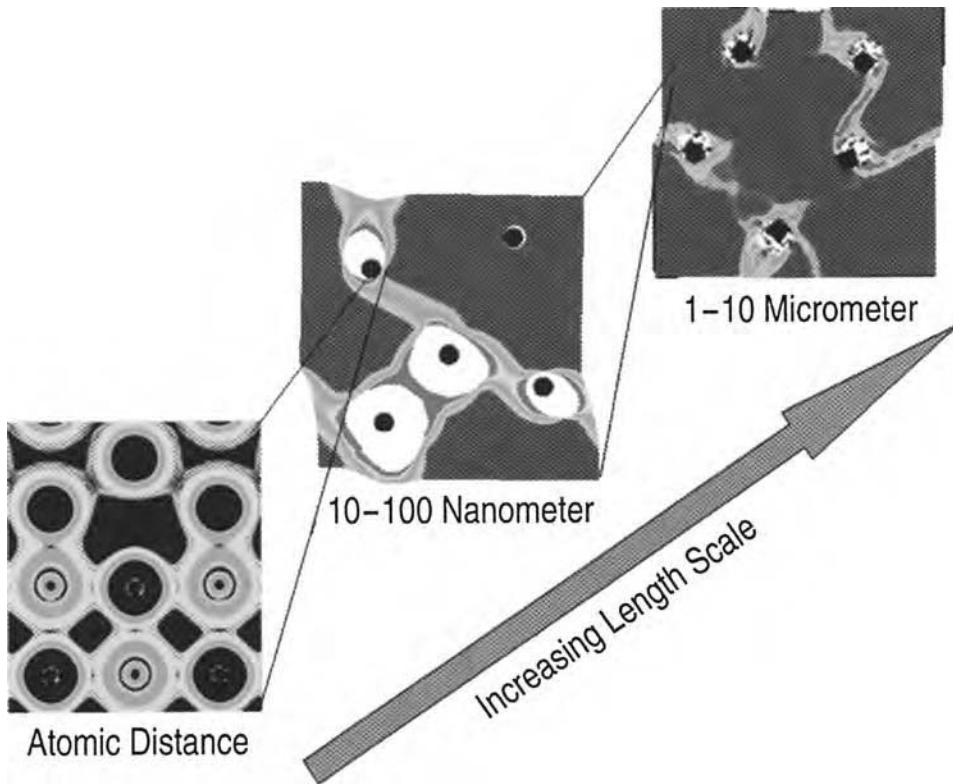
Fig. 43 Three scales of fracture models in a comprehensive approach to simulation of ductile fracture by void formation, growth, and coalescence. Source: Ref 44
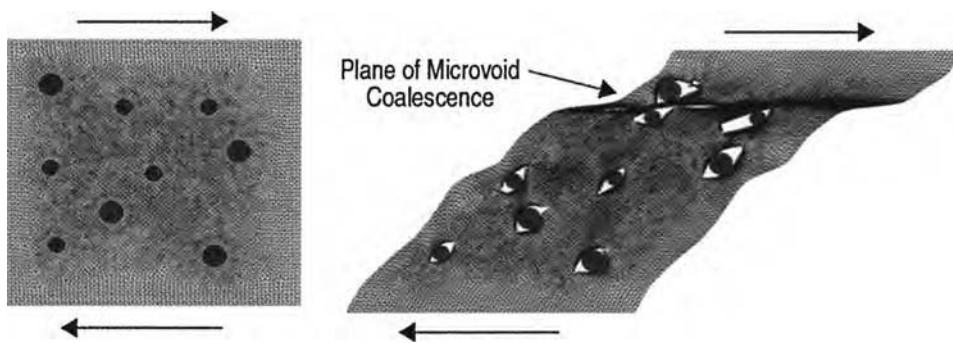
Fig. 44 Simulated formation of microvoids and their growth and coalescence. Source: Ref 44

Fig. 45 Molecular dynamics modeling of dislocation motion and void formation in the vicinity of a hard particle. Source: Ref 45

32. H.P. Ganser, A.G. Atkins, O. Kolednik, F.D. Fischer, and O. Richard, Upsetting of Cylinders: A Comparison of Two Different Damage Indicators, *J. Eng. Mater. Technol.,* Vol 123, 2001, p 94–99

33. Y. Bao and T. Wierzbicki, A Comparative Study on Various Ductile Crack Formation Criteria, *J. Eng. Mater. Technol.,*Vol 126, 2004, p 314–324

34. H.A. Kuhn, P.W. Lee, and T. Erturk, A Fracture Criterion of Cold Forging, *J. Eng. Mater. Technol. (Trans. ASME),* Vol 95, 1973, p 213–218

35. F.A. McClintock, A Criterion for Ductile Fracture by the Growth of Holes, *J. Appl. Mech.,* Vol 33, 1968, p 363–371

36. J. Rice and D. Tracey, On the Ductile Enlargement of Voids in Triaxial Stress Fields, *J. Mech. Phys. Solids,* Vol 17, 1969, p 201–217

37. Z. Marciniak and K. Kuczynski, A Model of Localized Thinning in Sheet Metalforming, *Int. J. Mech. Sci.,* Vol 9, 1967, p 609

38. H.A. Kuhn and P.W. Lee, Strain Instability Leading to Fracture at the Free Surface of Upset Cylinders, *Metall. Trans.,* Vol 2, 1971, p 3197–3202

39. P.W. Lee and H.A. Kuhn, Fracture in Cold Upset Forging—A Criterion and Model, *Metall. Trans.,* Vol 4, April 1973, p 969–974

40. M.G. Cockcroft and D.J. Latham, Ductility and the Workability of Metals, *J. Inst. Met.,* Vol 96, 1968, p 33–39

41. A.M. Freudenthal, *The Inelastic Behavior of Solids,* John Wiley & Sons, 1950

42. S.I. Oh., C.C. Chen, and. S. Kobayashi, Ductile Fracture in Axisymmetric Extrusion and Wire Drawing: Part 2—Workability in

Extrusion and Drawing, *J. Eng. Ind. (Trans. ASME),* Vol 101 (No. 1), p 3644

43. P. Brozzo, B. Deluca, and R.A. Rendina, A New Method for the Prediction of the Formability Limits of Metal Sheets, *Proceedings of the Seventh Biennial Conference on Sheet Metal Forming and Formability,* International Deep Drawing Research Group, Amsterdam, 1972

44. Y. Bao and T. Wierzbicki, On the Cut-Off Value of Negative Triaxiality for Fracture, *Eng. Fract. Mech.,* Vol 72, 2005, p 1049–1069

45. B. Gouveia, J. Rodrigues, and P. Martins, Ductile Fracture in Metalworking

Experimental and Theoretical Research, *J. Mater. Process. Technol.,* Vol 101, 2000, p 52–63

46. D. Zhao, J.P. Bandstra, and H.A. Kuhn, A New Fracture Criterion for Metalworking Processes, *Concurrent Engineering Approach to Materials Processing,* S.N. Dwivedi, A.J. Paul, and F.R. Dax, Ed., TMS, 1992, p 107–119

47. M.L. Tims, J.D. Ryan, W.L. Otto, and M.E. Natishan, Crack Susceptibility of Nickel-Copper Alloy K-500 Bars during Forging and Quenching, *Proc. First Int. Conf. on Quenching and Control of Distortion,* ASM International, 1992, p 243–250

48. S.I. Oh, J. Walters, and W.T. Wu, Finite-Element Method Applications in Bulk Forming, *Metalworking: Bulk Forming,* Vol 14A, *ASM Handbook,* ASM International, 2005, p 615–639

49. F.J. Vernerey, C. McVeigh, W.K. Liu, B. Moran, D. Tewari, D.M. Parks, and G. Olson, The 3-D Computational Modeling of Shear-Dominated Ductile Failure in Steel, *J. Met.,* Dec 2006, p 45–51

50. M.A. Myers, S. Traiviratana, V.A. Lubarda, D.J. Benson, and E.M. Bringa, The Role of Dislocations in the Growth of Nanosized Voids in Ductile Failure of Metals, *J. Met.,* Feb 2009, p 35–41

Modeling of Hot Tearing and Other Defects in Casting Processes

Brian G. Thomas, University of Illinois (UIUC)

AS COMPUTATIONAL MODELS MATURE, their practical benefit to improving casting processes is growing. Accurate calculation of fluid velocities, temperature, microstructure, and stress evolution is just the first step. Achieving tangible improvements to casting processes requires the accurate prediction of actual casting defects and product properties. Defects that form during solidification are important not only to the casting engineer but also to engineers involved in subsequent manufacturing processes and evaluation. Solidification defects are responsible for many of the defects in final manufactured products and failures in service. They originate from inclusion entrapment, segregation, shrinkage cavities, porosity, mold-wall interactions, cracks, and many other sources that are process-specific. Casting defects can be modeled by extending the results of casting simulations through postprocessing and/or by solving further coupled equations that govern these phenomena. The prediction of defect formation is made difficult by the staggering complexity of the phenomena that arise during commercial casting processes. This article introduces some of the concepts involved in modeling some of these solidification defects and focuses in more detail on hot tearing.

Inclusions

Inclusions are responsible for many serious surface defects and internal quality problems in cast products. They arise from foreign particles, such as eroded sand particles, and impurities remaining in the liquid metal after upstream refining (Ref 1). Nonmetallic inclusion particles act as sites of stress concentration and hydrogen gas nucleation, leading to lower fatigue life, hydrogen embrittlement, surface defects, and other problems in the final product. Predicting their damage requires knowledge of the number, size distribution, composition, and morphology of the inclusions coming from upstream processing prior to casting. Obtaining this knowledge ideally involves modeling the multiphase fluid flow, turbulent mixing and diffusion, species transport, chemical reactions, and particle interactions that create the inclusions in upstream processes.

Considerable modeling of these phenomena has been addressed in previous simulations of vacuum degassers, R-H degassers (a type of recirculating degasser), ladles, tundishes, and other refining vessels and transfer operations used in metallurgical processing (Ref 2). These models solve the multiphase Navier-Stokes equations for turbulent fluid flow, using software such as FLUENT (Ref 3), and provide the flow field for subsequent simulation of inclusion particle transport. The first challenge is to properly incorporate the phenomena that drive the flow, which usually include the buoyancy of injected gas bubbles (Ref 4), which depends on the shape of the bubbles, ranging from spherical caps to spheres. Other effects important to accurately computing the flow field may include natural convection, which requires a coupled heat-transport solution for the temperature field. When electromagnetic stirring is used, these forces require modeling the applied magnetic field. Another challenge is to incorporate the effects of turbulence. Computationally-efficient choices include simple "mixing-length" models, the two-equation models such as k-ε to simulate the time-average flow pattern. Large-eddy simulation (LES) models can simulate the details of the time-evolving turbulent vortices, but at great computational expense. These methods have been compared with each other and with measurements of fluid flow in continuous casting (Ref 5–7).

Formation. Modeling the thermodynamics and kinetics of particle formation, transport, collisions, and removal or entrapment in the molten metal during upstream refining processes is the next crucial step. Thermodynamic reactions to quantify the precipitates that form in these multicomponent alloy systems can be predicted by simultaneous solution of chemical equilibrium equations, where the biggest challenge is to find accurate activity coefficients. Equilibrium compositions can also be found by comparing free-energy functions, such as used in Thermo-Calc (Ref 8), FACT-Sage (Ref 9), MTDATA (Ref 10), Gemini (Ref 11), and other thermodynamic modeling software. The kinetics of nonmetallic inclusion formation are generally controlled by species transport in the liquid and at reaction interfaces, such as the slag-metal surface, where droplets of the different liquid phases, solid particles, and gas bubbles interact. The physical entrainment of slag particles into the molten metal is another important source of inclusions (Ref 12), which requires transient multiphase modeling of the free surface, considering its breakup into droplets and surface tension effects, and pushes current modeling capabilities to their limit.

Another important source of inclusions is reoxidation of the molten metal by exposure to air. Oxygen absorbs rapidly from the atmosphere into any exposed molten metal and combines to form precipitates, which has been predicted in molten steel from the alloy content (Ref 13). Predictions are limited by understanding of the entrainment of oxygen from the atmosphere, the turbulent flow of the liquid steel during pouring, which determines the gas-metal interface shape, and the internal transport and reactions of chemical species in the molten metal.

Transport. The transport of particles through the flowing metal is the next crucial step to determine the inclusion distribution in the final product and can be modeled in several ways (Ref 14). Although the effect of bubbles on the flow pattern can be modeled effectively using Eulerian-Eulerian multiphase models, the fate of inclusion distributions is best modeled via Lagrangian particle tracking. In this method, the trajectories of many particles are integrated from the local velocity field, based on previous solution of the fluid velocities of a mold-filling simulation. The effect of turbulence on the chaotic particle paths is very important and is best

modeled with the transient turbulent velocity field, using LES (Ref 15). In more computationally-efficient time-averaged simulations of the turbulent flow field, the effect of turbulence on particle motion can be approximated using methods such as "random walk," where the velocity at each time increment is given a randomly-generated component with magnitude proportional to the local turbulence level (Ref 3). This method has been applied successfully to simulate particle motion in continuous casting molds (Ref 16).

Inclusion particle size distributions evolve during transport due to collisions with each other and by their attachment to the surface of bubbles. Collisions can be modeled by tracking the evolution in the number distribution of particles in each size range, including the local effects of Brownian motion, turbulence, and diffusion, which is aided by size-grouping models to cover the large range of particle sizes (Ref 17). Attachment and removal by bubbles can be modeled by computing the attachment rates of different particle sizes to different bubble sizes and shapes in computational models of these microscale phenomena (Ref 16, 18). These attachment rates can then be incorporated into the macroscale models of fluid and particle trajectories (Ref 16, 18). In the extreme, inclusions may agglomerate into large clogs, which can restrict the flow of molten metal, cause detrimental changes in the downstream flow pattern, and can lead to catastrophic defects in the final product. Modeling and analysis of clogging is a complex subject that has been reviewed elsewhere (Ref 19).

Capture. Particle capture into the solidification front is a critical step during the modeling of inclusion transport. Small particles flow between the dendrites, so they can be modeled as entrapped when they touch a domain wall. Larger particles may be pushed by the interface or engulfed by a fast-moving planar front (Ref 20). More often, they are entrapped when they are suspended in front of the solidification front long enough for the dendrites to surround them. Entrapment is greatly lessened by tangential flow across the solidification front (Ref 21). A criterion for entrapment has been developed based on balancing the many forces that act on a particle suspended at the interface (Ref 21). Particles that never touch the interface, or escape capture, eventually may be removed if the flow pattern transports them to the casting boundaries, such as the top surface of some processes, where they can enter the slag layer.

The final step is to predict the property changes caused by the entrapped inclusions, which is a challenging modeling task and depends on downstream processing, such as rolling and heat treatment. Even with simply cooling to ambient, precipitation continues in the solid state, where the inclusion distribution is greatly affected by kinetic delays due to nucleation and solid-state diffusion (Ref 22). This is further complicated by preferential precipitation at grain boundaries and compatible existing inclusions and is affected by strains, local microsegregation, and many other phenomena. Clearly, the modeling of inclusions is a challenging task.

Segregation

Segregation is caused by the partitioning of alloying elements between the liquid and solid phases during solidification. Because species diffusion in the solid is very slow, this phenomenon is usually manifested by small-scale composition differences, called microsegregation, which explains how the spaces between dendrites are enriched in alloy relative to the dendrite centers. Although it contributes greatly to macrosegregation, porosity, inclusions, and other defects, microsegregation alone is not usually considered a defect, and it can be removed by homogenization heat treatment. When fluid flow is present, however, large-scale species transport leads to macrosegregation, where the composition differences arise over large distances, such as between the center and surface of a casting. This serious defect cannot be removed. It is extremely difficult to predict, because it involves so many different coupled phenomena, and at vastly different length and time scales. In addition to predicting fluid flow, species transport, and solidification, segregation requires prediction of the dendrite morphology and microstructure and the complete stress state, including deformation of the spongy mushy zone (Ref 23) and mechanical bulging and bending of the casting surface (Ref 24). Moreover, the fluid flow must be accurately characterized at both the microscopic scale between dendrite arms and at the macroscopic scale of the entire casting. Each of these modeling tasks is a large discipline that has received significant effort over several decades.

Segregation is the main phenomenon responsible for many different kinds of special defects that only affect particular casting processes. For example, "freckle" defects can arise during the directional solidification of turbine blades when buoyancy-driven flow allows winding vertical channels to penetrate between dendrites and become filled with segregated liquid near the end of solidification (Ref 25). Inverse segregation or surface exudation in direct-chill continuous casting of aluminum ingots arises during the initial stages of solidification when thermal stress pushes out droplets of enriched interdendritic liquid through pores in the spongy mushy zone where it extends to the ingot surface (Ref 26). A comprehensive summary of the modeling of this important class of defects is beyond the scope of this article, and reviews of various aspects of this complex subject can be found elsewhere (Ref 23, 27–29).

Shrinkage Cavities, Gas Porosity, and Casting Shape

Shrinkage cavities are voids in a casting that form due to the thermal contraction of liquid pockets after they become surrounded by solid that prevents the feeding of additional liquid. Porosity is the name for small voids that form due to the evolution and entrapment of gas bubbles. These two important classes of defects are related. Both involve the entrapment of liquid pockets, a criterion for the nucleation of gas bubbles, and depend on the overall shrinkage of the casting, which requires a complete thermomechanical stress calculation, in addition to accurate prediction of fluid flow and solidification. A rough estimate of shrinkage cavity potential is possible from postprocessing analysis of the results of a simple solidification heat-transfer analysis, looking for regions where solid surrounds the liquid. This simple analysis can be automated by tracking parameters that represent shrinkage potential, such as the Niyama criterion (Ref 30, 31) More accurate prediction of shrinkage requires complete modeling of fluid flow, heat transfer, and thermal-stress analysis. The fluid flow analysis is further complicated by the need for accurate characterization of the permeability of the porous dendritic network, which also depends on the microstructure and alloy segregation. The stress analysis depends on the evolving strength of the solid, in addition to the mushy zone, interaction with the mold, and other phenomena that are discussed further in the section on hot tearing.

In addition to the phenomena that govern shrinkage cavity formation, gas porosity prediction also requires modeling the transport of dissolved gases, the nucleation of bubbles or gas pockets, and their possible transport after they form. This modeling also involves the same complications discussed in the prediction of inclusions, including nonequilibrium thermodynamics, chemical reactions, nucleation, precipitate formation, and growth kinetics. Indeed, precipitation reactions are alternative ways for the dissolved gases to be consumed. Finally, gas bubbles that float during solidification can collide and coalesce, depending on surface tension. When combined with improper venting, this can lead to the creation of a defect found at the top of foundry castings, known as a surface blow hole.

Shrinkage and porosity defects are related to the final shape of the casting. When the solid metal shell is strong enough to resist shrinkage and retain its external dimensions, internal shrinkage and porosity may be more problematic. In contrast, practices that lessen shrinkage and porosity may involve more external shrinkage of the exterior. Inaccurate final dimensions is another casting defect. Because comprehensive modeling of these defects requires the simultaneous solution of so many different

equation systems, with so many uncertain fundamental properties, this class of defects is difficult to predict and is the subject of intense ongoing research. The art of modeling these defects involves how to make simplifying assumptions with the least loss of accuracy. Further details on the current state of the art in modeling of this important class of defects is given elsewhere (Ref 28, 32, 33).

Mold-Wall Erosion

Feeding molten metal into the casting cavity is a critical operation where defects may arise. Excessive turbulence and velocity impingement of the molten metal can erode the surface of the mold wall, especially near the in-gate. In sand molds, this can dislodge sand particles to act as another source of inclusions in the final casting. Even with permanent metal molds such as used in pressure die casting, excessive velocity against the metal walls can locally erode the metal, enlarging the casting cavity and creating surface defects.

Erosion rate has been related to the metal velocity and other parameters in a few previous studies, based mainly on empirical correlations (Ref 34). For example, in die casting, erosion strength has been characterized by integrating the instantaneous velocity over the time of the injection cycle for each local portion of the mold-wall surface (Ref 35). The resulting contours over the mold surface can be correlated with erosion damage.

Erosion of the mold wall due to fluid flow also may remove protective surface coatings and allow chemical reactions between the mold and the exposed mold metal. Thus, the mechanical erosion may be accompanied by chemical erosion and/or metallurgical corrosion, which often act together to wear away the mold surface. Analysis of the chemical component requires consideration of the thermodynamic reactions and their kinetics. The interdiffusion of elements in the molten metal to contaminate the mold walls can lower the local melting temperature. This is responsible for the problem of soldering in aluminum die casting in steel molds (Ref 36, 37).

Mold-Wall Cracks

Cracks in the mold wall are another source of defects in the casting, in addition to lowering the lifetime of permanent molds. Mold cracks decrease the local heat-transfer rate, allowing local strain concentration in the adjacent solidifying metal and causing hot-tear cracks at the casting surface that mirror those in the mold. In water-cooled molds, mold cracks also pose a safety hazard, from the chance of molten metal contacting the cooling water. Mold cracks, or heat checks, are caused by repeated rapid and severe fluctuations in the mold surface temperature. They can be predicted from the results of a transient thermal-stress analysis of the mold itself, by combining the calculated inelastic strain (due to plasticity and creep) with measurements of cycles to failure from thermal-fatigue experiments. For example, surface cracks in copper molds used from continuous casting were predicted by comparing the results of transient three-dimensional finite-element analysis of the copper mold and its support structure during cyclic loading with measured fatigue cycle-to-failure data (Ref 38). In addition to adopting practices to lower the maximum surface temperature, the mold lifetime was predicted to increase by lessening constraint of the mold by loosening bolts (Ref 38, 39). Often, the prediction of mold cracks requires consideration of the chemical interaction of the liquid metal with the mold, such as formation of brass in copper molds by the preferential absorption of zinc from the molten metal.

Other Defects

Many other casting defects arise due to problems specific to individual processes. Grain defects, such as unwanted grain boundaries, are important in directional solidification processes, such as the casting of single-crystal turbine blades, where high-temperature creep resistance is the most important property. In the Czochralski process, where single crystals are slowly pulled from doped melts to cast rods for making semiconductor wafers, even dislocations are serious defects that must be minimized. Examples in foundry sand casting include cold shut, blow holes, liquid metal penetration into the sand grains, and other surface defects. Some insight into these defects can be found from the results of a solidification heat-transfer analysis. For example, problems related to cold shut can be estimated from a simulation if the molten metal freezes before the casting cavity is completely filled, leaving voids or seams at the junction where two streams met. Crystal defects depend on the temperature gradient across the solidification front. Further insight can be gained from direct modeling of the microstructure (Ref 40) and molecular dynamics or quantum-mechanics models of dislocations and other phenomena at the atomic scale (Ref 41). Many important process-specific defects have received little attention by the modeling community.

A final category of defects may be termed "goofups" because their cause is so obvious, and the solution involves, at most, only basic calculations. For example, a short pour occurs when the volume of metal poured is less than the volume of the casting cavity. Unsightly mismatch seams arise when the two halves of the foundry casting mold are not aligned due to poor maintenance of the hinges and pins. Although obvious, avoiding such defects requires careful and diligent operations. Here, expert-system-type software may help, aided in these examples by embedding simple volume calculations and tracking of maintenance schedules. The rest of this article focuses on the important defect of hot-tear crack formation.

Hot-Tear Cracks

Crack formation is caused by a combination of tensile stress and metallurgical embrittlement. Although solidifying metal is subject to embrittlement due to a number of different mechanisms at different temperature ranges, hot-tear cracks form near the solidus temperature. Embrittlement is so severe near this temperature that hot-tear cracks form at strains on the order of only 1%, making them responsible for most of the cracks observed in cast products. Hot-tear cracks form because thin liquid films between the dendrites at grain boundaries are susceptible to strain concentration, causing separation of the dendrites and intergranular cracks. The prediction of these cracks presents a formidable challenge to modellers, owing to the many complex, interacting phenomena that govern stress and embrittlement, some of which are not yet fully understood:

- Predicting temperature, strain, and stress during solidification requires calculation of the history of the cast product and its environment over huge temperature intervals. Characterizing the heat-transfer coefficients at the boundaries and interfaces is one of many difficulties.
- The mechanical problem is highly nonlinear, involving liquid-solid interaction and complex constitutive equations. Stress arises primarily from the mismatch of strains caused by large temperature gradients and depends on the time- and microstructure-dependent inelastic flow of the material. Even identifying the numerous metallurgical parameters involved in these relationships is a daunting task.
- The coupling between the thermal and the mechanical problems is an additional difficulty. This coupling comes from the mechanical interaction between the casting and the mold components, through gap formation or the buildup of contact pressure, locally modifying the heat exchange.
- Accounting for the mold and its interaction with the casting makes the problem multidomain, usually involving numerous deformable components with coupled interactions and contact analysis.
- Cast parts usually have very complex three-dimensional shapes, which puts great demands on the interface between computer-aided design and the mechanical solvers and on computational resources.
- The main cause of embrittlement is the segregation of solute impurities and alloying elements to the interdendritic liquid between primary grains, which lowers the solidus temperature locally. Segregation is most severe, and thus most important, at the grain boundaries, owing to the greater local

interdenritic spacing locally, which allows the liquid to persist longer between grain boundaries.

- Larger primary grain size increases strain concentration and embrittlement, so it must also be predicted. Because the grain size evolves with time, the grain size in the final cooled microstructure differs from the primary grain size, so grain size measurements for model validation should be inferred from analysis of the microsegregation pattern.
- Stress on the liquid films depends on the ability of liquid to flow through the dendritic structure to feed the volumetric shrinkage, relative to the strength of the surrounding dendritic skeleton. Thus, accurate permeability models are required for the mushy zone, which, in turn, require accurate prediction of the microstructure, including the dendrite arm shapes, especially at the grain boundaries.
- Crack prediction requires modeling the distribution of supersaturated dissolved gas and its nucleation into pores or crack surfaces.
- The formation of solid precipitates tends to pin the primary grain boundaries, enhancing strain concentration. The interfering precipitates also act as nucleation sites for both gas bubbles and voids, both of which increase embrittlement. Modeling precipitation is difficult, owing to the multicomponent nature of commercial alloys and the importance of kinetic delays.
- The subsequent refilling of hot tears with segregated liquid alloy can cause internal defects that are just as serious as exposed surface cracks, which oxidize. This again requires accurate prediction of both interdendritic and intergranular solute flow.
- The most important parameters to hot tearing— the stress-tensor field, which acts to concentrate tensile strain in liquid regions of the mushy zone, and the fluid-velocity vector field, which acts to fill the voids— are both

three-dimensional time-varying quantities that depend greatly on the orientation and shape of the microstructure. Thus, even empirical criteria to predict hot tears depend on conducting experiments with the proper load orientation, rates, and microstructures.

- The important length scales range from micrometers (dendrite arm shapes) to tens of meters (metallurgical length of a continuous caster), with a similar huge order-of-magnitude range in time scales.

Heat-Transfer Modeling

Accurate calculation of the evolving temperature distribution during the casting process is the first and most important step in the analysis of hot tears. In addition to solving the transient heat-transport equation with phase change, this critical task usually requires coupling with turbulent fluid flow during mold filling and interaction with the mold walls, with particular attention to the interfacial gap.

Heat transfer across the mold-casting interface depends on the size of the gap (if open) or the contact pressure (if closed), so coupling with results from a mechanical analysis is often needed. Figure 1 shows the changes in interfacial heat transfer for these two cases. When a gap opens between the casting and the mold due to their relative deformation, the heat transfer drops in proportion to the size of the gap. Heat flows across the interface, q_{gap}, by conduction through the gas within the gap and by radiation between the two parallel surfaces:

$$q_{gap} = \frac{k_{gas}}{g}(T_c - T_m) + \frac{\sigma(T_c^4 - T_m^4)}{\frac{1}{\varepsilon_c} + \frac{1}{\varepsilon_m} - 1} \qquad \text{(Eq 1)}$$

where $k_{gas}(T)$ is the thermal conductivity of the gap; g is the gap thickness; T_c and T_m are the local surface temperature of the casting and

mold, respectively; ε_c and ε_m are emissivities; and σ is the Stefan-Boltzmann constant. To avoid numerical problems at small gap sizes, this function should be truncated to a finite value, h_0, which corresponds to the closed-gap case and depends on the average roughness. More sophisticated functions can be applied to account for mold coating layers, different material layers, radiation conduction, contact resistances to incorporate surface roughness, and other phenomena. Specific examples of these gap heat-transfer laws are provided elsewhere for continuous casting with oil lubrication (Ref 42) and mold flux (Ref 43).

When contact between the mold and casting is good, the interfacial heat flux increases with contact pressure according to a power law (Ref 44), such as:

$$q_{contact} = (h_0 + Ap_c^B)(T_c - T_m) \qquad \text{(Eq 2)}$$

where p_c is the contact pressure, and A and B are fitting parameters that depend on the materials, lubricants, roughnesses, and temperature. After removal from the mold, heat transfer is given by uncoupled surface convection coefficients. Accurate characterization of the surface heat flux for all of these conditions requires careful calibration and validation with experimental measurements and is a critical step in modeling.

Thermomechanical Modeling

Prediction of the displacements, strains, and stresses during the casting process is the next step in predicting residual stress, the distorted shape, and crack defects, including hot tears. As previously mentioned, stress analysis is also important in the prediction of porosity and segregation. The modeling of mechanical behavior requires solution of the equilibrium or momentum equations relating force and stress, the compatibility equations relating strain and displacement, and the constitutive equations relating stress and strain. This is because the boundary conditions specify either force or displacement at different regions of the domain boundaries.

Governing Equations. The conservation of force (steady-state equilibrium) or momentum (transient conditions) can be expressed by:

$$\rho\left(\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v} \cdot \nabla \boldsymbol{v}\right) = \nabla \cdot \boldsymbol{\sigma} + \rho\mathbf{g} \qquad \text{(Eq 3)}$$

where $\boldsymbol{\sigma}$ is the stress tensor, ρ is the density, $\mathbf{g}$ is the gravitational acceleration, $\mathbf{v}$ is the velocity field, ∇ is the gradient operator and matrices (vectors and tensors) are denoted in bold. Once solidified, the velocity terms that comprise the left side of Eq 3 can be neglected.

The strains that dominate thermomechanical behavior during solidification are on the order of only a few percent, prior to crack formation. With small gradients of spatial displacement,
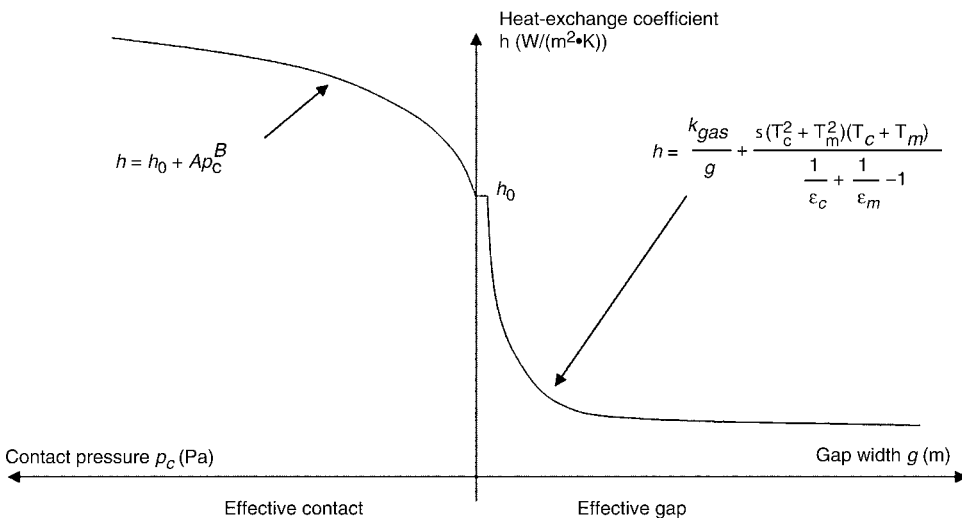
Fig. 1 Modeling heat-transfer coefficient across the mold-casting gap

The figure shows a plot with Heat-exchange coefficient h (W/(m²·K)) on the vertical axis, Contact pressure p_c (Pa) to the left and Gap width g (m) to the right on the horizontal axis. The left branch is labeled $h = h_0 + Ap_c^B$ and the right branch is labeled:

$$h = \frac{k_{gas}}{g} + \frac{s(T_c^2 + T_m^2)(T_c + T_m)}{\frac{1}{\varepsilon_c} + \frac{1}{\varepsilon_m} - 1}$$

with h_0 marked at the transition, "Effective contact" on the left and "Effective gap" on the right.

$\nabla \mathbf{u} = \partial \mathbf{u} / \partial \mathbf{x}$, the compatibility equations simplify to (Ref 45):

$$\mathbf{\varepsilon} = \frac{1}{2} \left(\nabla \mathbf{u} + (\nabla \mathbf{u})^{\mathrm{T}} \right) \qquad \text{(Eq 4)}$$

where $\mathbf{\varepsilon}$ is the strain tensor, $\mathbf{u}$ is the displacement vector, and $^{\mathrm{T}}$ denotes transpose. This small-strain assumption simplifies the analysis considerably. The compatibility equations can also be expressed as a rate formulation, where strains become strain rates and displacements become velocities. This formulation is more convenient for a transient computation with time integration involving fluid flow and/or large deformation.

Choosing constitutive models to relate stress and strain is a very challenging aspect of stress analysis of solidification, because it depends on accurately capturing the highly nonlinear evolution of the material microstructure with numerical parameters. Traditionally, this is accomplished with a family of elastic-plastic stress-strain curves at the appropriate temperatures and strain rate(s) and perhaps by adding a separate strain-rate function of temperature, stress, and time to account for the time-dependent softening effects of creep.

However, the state variables of strain and time are not enough to quantify the strength of the material, especially during loading reversals. Furthermore, the effects of plastic strain and creep-strain rate are not independent. Thus, unified models have been developed that combine the different microstructural mechanisms together in terms of state variables that relate more closely to fundamental microstructural parameters such as dislocation density. Many models of different complexity can be found in the literature (Ref 46, 47). In their simplest form, these constitutive equations for metals are often expressed in terms of the state variables of temperature and inelastic strain:

$$\dot{\mathbf{\varepsilon}} = \dot{\mathbf{\varepsilon}}^{\mathrm{el}} + \dot{\mathbf{\varepsilon}}^{\mathrm{in}} + \dot{\mathbf{\varepsilon}}^{\mathrm{th}} \qquad \text{(Eq 5)}$$

$$\dot{\mathbf{\varepsilon}}^{\mathrm{el}} = \frac{1+\nu}{E} \dot{\mathbf{\sigma}} - \frac{\nu}{E} \mathrm{tr}(\dot{\mathbf{\sigma}})\mathbf{I} + \dot{T} \frac{\partial}{\partial T} \left(\frac{1+\nu}{E} \right) \mathbf{\sigma} - \dot{T} \frac{\partial}{\partial T} \left(\frac{\nu}{E} \right) \mathrm{tr}(\mathbf{\sigma})\mathbf{I} \qquad \text{(Eq 6)}$$

$$\dot{\mathbf{\varepsilon}}^{\mathrm{in}} = f(\sigma, T, \text{structure}) \qquad \text{(Eq 7)}$$

$$\dot{\mathbf{\varepsilon}}^{\mathrm{th}} = \left[\sqrt[3]{\frac{\rho(T_0)}{\rho}} - 1 \right] \mathbf{I} \qquad \text{(Eq 8)}$$

These tensor equations are expressed in terms of rates, where an overdot is the time derivative, tr is the trace of a matrix, $\mathbf{I}$ is the identity tensor, and every variable should depend on temperature, T. The strain-rate tensor, $\dot{\mathbf{\varepsilon}}$, is split into an elastic component $(\dot{\mathbf{\varepsilon}}^{\mathrm{el}})$,

an inelastic (nonreversible) component $(\dot{\mathbf{\varepsilon}}^{\mathrm{in}})$, and a thermal component $(\dot{\mathbf{\varepsilon}}^{\mathrm{th}})$. Equation 6 is the hypoelastic Hooke's law, where E is Young's modulus, ν is the Poisson's coefficient, and $\dot{\mathbf{\sigma}}$ is the time derivative of the stress tensor $\mathbf{\sigma}$. Equation 7 gives a framework for evolving the inelastic strain tensor, $\mathbf{\varepsilon}^{\mathrm{in}}$, which is often used as the only parameter to characterize material structure. The thermal strains (Eq 8) include the solidification shrinkage and are based on the temperature field solved with the heat-transfer model. Care should be taken in choosing a consistent reference temperature, T_0, and in differentiating to extract the thermal strain rate, which can be accomplished numerically. Finding suitable constitutive equations to characterize the material mechanical response for the wide range of conditions experienced during solidification is a formidable task that requires careful experiments under different loading conditions, a reasonable form for the theoretical model, and advanced fitting procedures to extract the model coefficients.

Solution Strategies. Thermomechanical analysis of casting processes poses special difficulties due to the simultaneous presence of liquid, mushy, and solid regions that move with time as solidification progresses, the highly nonlinear constitutive equations, complex three-dimensional geometries, coupling with the thermal analysis, interaction with the mold, and many other reasons. Several different strategies have been developed, according to the process and model objectives:

- A first strategy is to perform a small-strain thermomechanical analysis on just the solidified portion of the casting domain, extracted from the thermal analysis results. This strategy is convenient when the solidification front is stationary, such as the continuous casting of aluminum (Ref 48) and steel (Ref 42, 49). For transient problems, such as the prediction of residual stress and shape (butt-curl) during startup of the direct-chill and electromagnetic continuous casting processes for aluminum ingots, the domain can be extended in time by adding layers (Ref 48).
- A second popular strategy considers the entire casting as a continuum, modifying the parameters in the constitutive equations for the liquid, mushy, and solid regions according to the temperature and phase fraction. For example, liquid can be treated by setting the strains to zero when the temperature is above the solidus temperature. The primary unknowns are the displacements or displacement increments. To facilitate the tracking of state variables, a Lagrangian formulation is adopted, where the domain follows the material. This popular approach can be used with structural finite-element codes, such as MARC (Ref 50) or ABAQUS (Ref 51), and with commercial solidification codes or special-purpose software, such as ALSIM (Ref 52)/ALSPEN (Ref 53), CASTS (Ref 54), CON2D (Ref 55, 56), Magmasoft

(Ref 57), and Procast (Ref 58, 59). It has been applied successfully to simulate deformation and residual stress in shape castings (Ref 60, 61), direct chill casting of aluminum (Ref 48, 52, 53, 60, 62, 63), and continuous casting of steel (Ref 55, 64). Time integration of the highly nonlinear constitutive equations can benefit from special local-global integration numerical methods (Ref 56) or recent explicit methods (Ref 65). Assuming small strain and avoiding Poisson's ratio close to 0.5 for stability reasons (Ref 66, 67) means that the liquid phase is not modeled accurately. Thus, some phenomena must be incorporated from other models, such as heat transfer from impinging liquid jets (Ref 68) and fluid feeding into the mushy zone (Ref 55).

- A third strategy simulates the entire casting, treating the mass and momentum equations of the liquid and mushy regions with a mixed velocity-pressure formulation. The primary unknowns are the velocity (time derivative of displacement) and pressure fields, which makes it easier to impose the incompressibility constraints and to handle hydrostatic pressure loading. Indeed, the velocity-pressure formulation is also applied to the equilibrium of the solid regions, in order to provide a single continuum framework for the global numerical solution. This strategy has been implemented into codes dedicated to casting analysis, such as THERCAST (Ref 64, 69, 70) and VULCAN (Ref 71). If stress prediction is not important so that elastic strains can be ignored, then this formulation simplifies to a standard fluid-flow analysis, which is useful in the prediction of bulging and shape in large-strain processes. For problems involving large strain, such as squeeze casting, this strategy is suited to an arbitrary Lagrangian Eulerian (ALE) formulation. In a Eulerian formulation, material moves through the computational grid, which remains stationary in the laboratory frame of reference and requires careful updating of the state variables. In ALE, mesh updating is partially independent of the material velocity to maintain the quality of the computational grid. Further details are provided elsewhere (Ref 69, 72).

Hot-Tearing Criteria

The next step is to quantify embrittlement and to incorporate it with the thermal-stress analysis to predict hot-tear cracks. Hot-tearing phenomena are too complex, too small-scale, and insufficiently understood to model in detail as part of the macroscale thermomechanical analysis. Thus, several different criteria and approaches have been developed to predict hot tears from the results of such analyses. This topic is the focus of many ongoing research efforts, and although many of these criteria

reproduce observed trends, much more work is needed before quantitative predictions are reliable.

Different approaches are needed for different microstructures and metals, according to the most important phenomena that govern crack formation. Hot-tear cracks forming within large networks of mushy, equiaxed grains require accurate constitutive models to quantify the rheology of the mushy region. Cracks between columnar grains require models that incorporate the balance between liquid feeding between dendrites and tensile deformation perpendicular to the direction of dendrite growth. The hot tearing of aluminum alloys additionally depends on the critical stress to nucleate a gas bubble. In steel, dissolved gas contents are usually low, so hot tears usually refill with segregated liquid without opening into cracks. This macrosegregation is very damaging, so it becomes a very important phenomenon to model accurately. Every criterion depends on experimental measurements and how best to incorporate them.

Thermal-Analysis-Based Criteria. The results of the solidification heat-transfer analysis alone can provide some important insights into hot tearing. As illustrated in Fig. 2, the location of hot-tear cracks observed in a casting can be
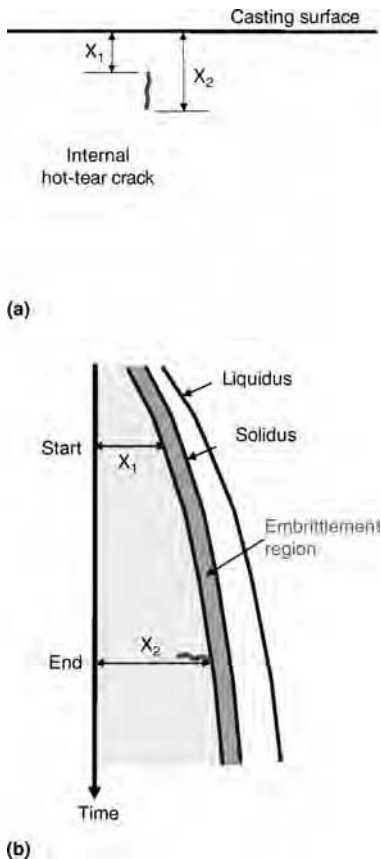
Fig. 2 Relating the location of hot-tear crack formation to results of a transient thermal simulation. (a) Measure crack location in casting. (b) Predict shell thickness history

related to their time of formation. Cracks tend to initiate near the casting surface (x_1) and propagate toward the center of casting (x_2) as solidification progresses. Figure 2(b) shows the progress of the mushy zone and important isotherms with time, based on the results of a solidification heat-transfer model. In the case of continuous casting, the time axis also corresponds to distance in the casting direction, so the figure depicts the actual shape of the solidification fronts in the real caster.

Casting conditions that produce faster solidification and alloys with wider freezing ranges are more prone to hot tears. Thus, many criteria to indicate hot-tear cracking susceptibility (HCS) are solely based on thermal analysis. One (Ref 73) simply compares the local time spent between two critical solid fractions, g_{s1} and g_{s2} (typically 0.9 and 0.99, respectively), with the total local solidification time (or a reference solidification time):

$$HCS_{clyne} = \frac{t_{0.99} - t_{0.90}}{t_{0.90} - t_{0.40}} \qquad (Eq\ 9)$$

Mechanical-Analysis-Based Criteria. Many different criteria have been developed to predict hot-tear cracks from the results of a mechanical analysis. Regardless of the model formulation, developing an accurate criterion function to predict hot tears relies on measurements, such as the submerged split-chill tensile test (Ref 74–76). This experiment applies and measures a tensile load on the solidifying shell, perpendicular to the growth direction, so it matches the conditions present in hot tearing between columnar grains. Other experiments, such as the Gleeble, apply a tensile load to remelted metal that is held in place by surface tension. Care must be taken in the interpretation of such measurements because the load is generally applied in the same direction as solidification-front growth. Proper interpretation of any hot-tearing experiment requires detailed modeling of the experiment itself, because conditions are never constant, and, at best, only raw data such as temperature, displacement, and force can be measured. The parameters of greatest interest must be extracted using models.

Criteria based on classical mechanics often assume cracks will form when a critical stress is exceeded, and they are popular for predicting cracks at lower temperatures (Ref 77–80). Tensile stress is also a requirement for hot-tear formation (Ref 81). This critical stress depends greatly on the local temperature and strain rate. The maximum tensile stress occurs just before formation of a critical flaw (Ref 82).

Measurements often correlate hot-tear formation with the accumulation of a critical level of mechanical strain while applying tensile loading within a critical solid fraction where liquid feeding is difficult. This has formed the basis for many hot-tearing criteria. One such model (Ref 81) accumulates inelastic deformation over a brittleness temperature range, which is defined, for example, as $g_s \in [0.85, 0.99]$ for a

Fe-0.15wt%C steel grade. The local condition for fracture initiation is then:

$$\sum_{g_{s1}}^{g_{s2}} \Delta \varepsilon^{in} \geq \varepsilon_{cr} \qquad (Eq\ 10)$$

in which the critical strain, ε_{cr}, is 1.6% at a typical strain rate of 3×10^{-4} s^{-1}. Careful measurements during bending of solidifying steel ingots have revealed critical strains ranging from 1 to 3.8% (Ref 81, 83). The lowest values were found at high strain rate and in crack-sensitive grades (e.g., high-sulfur peritectic steel) (Ref 81). In aluminum-rich copper alloys, critical strains were reported from 0.09 to 1.6% and were relatively independent of strain rate (Ref 82).

The critical strain decreases with increasing strain rate, presumably because less time is available for liquid feeding, and also decreases for alloys with wider freezing ranges. The following empirical equation for the critical strain in steel, ε_{cr}, was based on fitting measurements from many bend tests (Ref 84):

$$\varepsilon_{cr} = \frac{0.02821}{\dot{\varepsilon}^{0.3131} \Delta T_B^{0.8638}} \qquad (Eq\ 11)$$

where $\dot{\varepsilon}$ is the strain rate (s^{-1}) and ΔT_B is the brittle temperature range (°C) defined between the temperatures corresponding to solid fractions of 0.9 and 0.99.

An elegant analytical-criterion model has been derived to predict hot tearing, based on when the local liquid feeding rate along the interdendritic spaces between the primary columnar dendrites is insufficient to balance the rate of tensile strain increase in the perpendicular direction across the mushy zone (Ref 85, 86). Specifically, gas pores cavitate to separate the residual liquid film between the dendrites when the tensile strain rate exceeds a critical value:

$$\dot{\varepsilon} \geq \frac{1}{R} \left[\frac{\lambda_2^2 \|\nabla T\|}{180 \mu_l} \frac{\rho_L}{\rho_S} (p_m - p_C) - v_T \frac{\rho_S - \rho_L}{\rho_S} H \right] \qquad (Eq\ 12)$$

in which μ_l is the dynamic liquid viscosity, λ_2 is the secondary dendrite arm spacing, p_m is the local pressure in the liquid ahead of the mushy zone, p_C is the cavitation pressure, v_T is the velocity of the solidification front, and $\|\nabla T\|$ is the magnitude of the temperature gradient across the mushy zone. The quantities R and H depend on the solidification path of the alloy:

$$R = \int_{T_2}^{T_1} \frac{g_s^2 F(T)}{g_l^3} dT \quad H = \int_{T_2}^{T_1} \frac{g_s^2}{g_l^2} dT$$

$$F(T) = \frac{1}{\|\nabla T\|} \int_{T_2}^{T} g_s dT \qquad (Eq\ 13)$$

where the integration limits are calibration parameters (Ref 87). The upper limit T_1 may be the liquidus or the coherency temperature, while the lower limit T_2 typically is within the solid fraction range of 0.95 to 0.99 (Ref 88).

This criterion model has been applied to hot tearing of aluminum microstructures (Ref 87).

For hot tearing within large mushy regions, typically equiaxed microstructures, constitutive behavior of the mushy zone to predict the local fluid flow and deformation of the dendritic network presents an important additional challenge. Other criterion models that focus more on this aspect of hot tearing have recently been developed (Ref 88–90). Further details on hot tearing of aluminum alloys are reviewed elsewhere (Ref 91).

Microscale Model-Based Criteria. Detailed computational models can be developed based on temperature, fluid flow, stress, and strain in the mushy zone during solidification. For example, a finite-element model of an equiaxed mushy zone of aluminum has been applied to investigate constitutive behavior and to quantify strain concentration in the liquid films for a few specific sets of conditions (Ref 92). Once such models are more mature, their results can be incorporated into better criteria for hot tearing. A final difficult task is extracting results from the macroscale model results to compare with the criterion models, owing to the sensitivity of numerical estimates of parameters such as strain rate to numerical oscillations and mesh refinement effects. Thus, coupling difficulties between the macro- and microscale models is another reason that hot-tear crack prediction is an ongoing challenge.

Microsegregation Modeling

Quantifying the relationship between temperature and phase fractions is an essential part of each model involved in the prediction of hot tearing, including the heat transfer, the mechanical, and the hot-tear criterion models. This relationship determines how latent heat is evolved in the heat-transfer model and how to switch between constitutive models in the mechanical model. Although simple linear, lever-rule, or Scheil-based relations are usually sufficient for these macroscale models, microsegregation is an essential aspect of embrittlement and greatly affects the phase-fraction temperature relationship involved in any hot-tearing criterion. Better relationships use the results of microsegregation models that consider partial diffusion of multiple solute elements in the solid phase, using simple analytical solutions (Ref 93), or one-dimensional models of a single secondary dendrite arm (Ref 94). More advanced models couple this calculation together with the macroscale models and allow the relationship to evolve to incorporate nucleation undercooling and other phenomena (Ref 27). Ideally, the relationships applied between dendrites and at grain boundaries should be different, and they should vary with location in the casting, to account for macrosegregation and other phenomena. An important concept, which is often overlooked, is that the same (or very close) relationship must be used in each model of the analysis. Inconsistency

between microsegregation models is one of the main reasons why different researchers have proposed different critical temperatures in their hot-tear criteria. Experiments conducted to quantify the parameters in hot-tearing models should fully report both the raw data and the models used to extract hot-tearing parameters, including the microsegregation model.

Model Validation

Model validation is a crucial step in any computational analysis. Analytical solutions are needed to prove internal consistency of the model and to control discretization errors. Comparison with experiments is needed to prove the model assumptions, property data, and boundary conditions. Weiner and Boley (Ref 95) derived an analytical solution for unidirectional solidification of an unconstrained plate, which serves as an ideal benchmark problem to validate thermal and mechanical models. The plate is subjected to sudden surface quench from a uniform initial temperature to a constant mold temperature, with a unique solidification temperature, an elastic-perfectly-plastic constitutive law, and constant properties.

This benchmark problem can be solved with a simple mesh of one row of elements extending from the casting surface into the liquid, as shown in Fig. 3. Numerical predictions should match with acceptable precision using the same element type, mesh refinement, and time steps planned for the real problem. For example, the solidification stress analysis code CON2D (Ref 55) and the commercial code ABAQUS (Ref 51) were applied for typical conditions of steel casting (Ref 56).

Figures 4 and 5 compare the temperature and stress profiles in the plate at two times. The temperature profile through the solidifying shell is almost linear. Because the interior cools relative to the fixed surface temperature, its shrinkage generates internal tensile stress, which induces compressive stress at the surface. With no applied external pressure, the average stress through the thickness must naturally equal zero, and stress must decrease to zero in the liquid. Stresses and strains in both transverse directions are equal for this symmetrical problem. The close agreement demonstrates that the computational model is numerically consistent and has an acceptable mesh resolution. Such studies reveal that a relatively fine mesh is needed to achieve reasonable accuracy, and that results from many thermomechanical models reported in previous literature had insufficient mesh refinement. Comparison with experimental measurements is also required, to validate that the modeling assumptions and input data are reasonable.
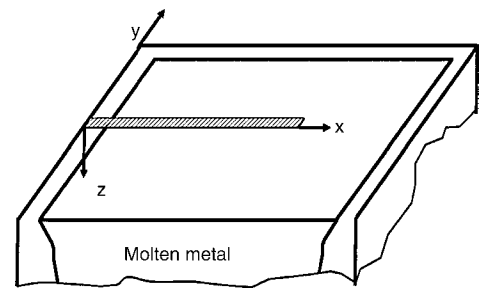
Fig. 3 One-dimensional slice domain for modeling solidifying plate
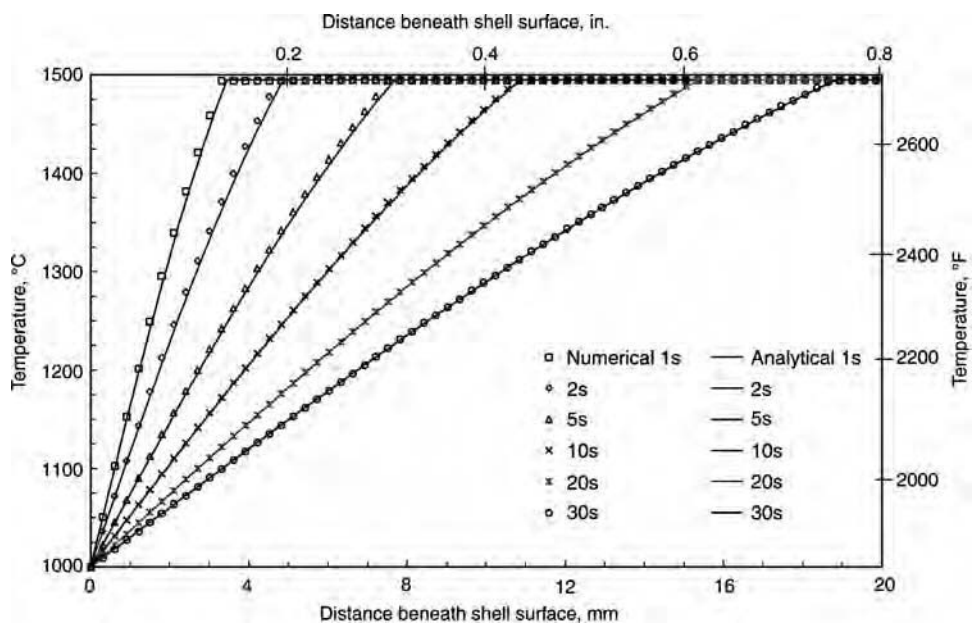
Fig. 4 Temperatures through solidifying plate at different times, comparing analytical solution and numerical predictions
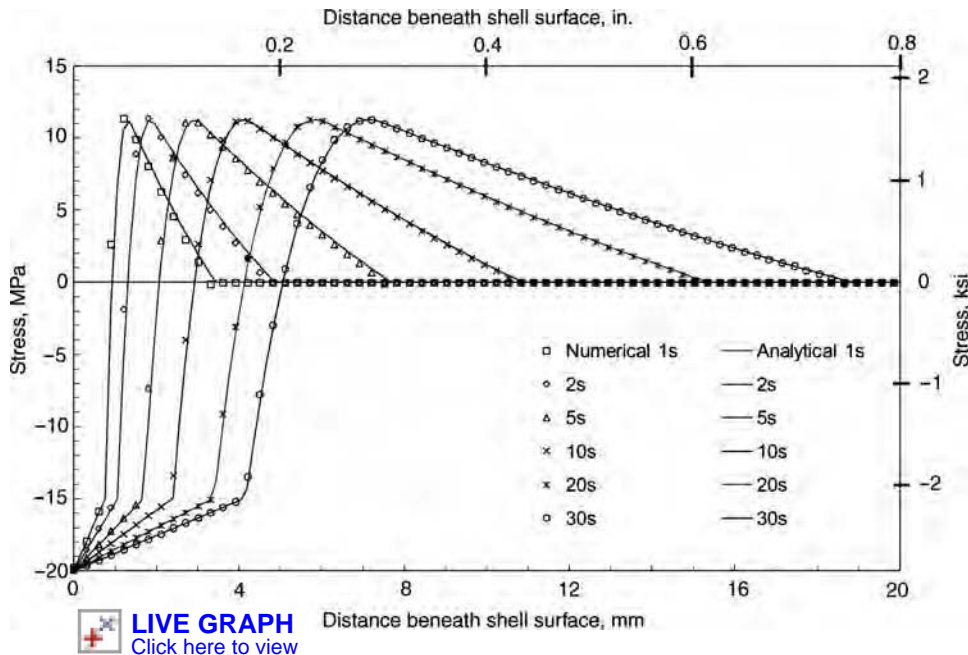
Fig. 5 Lateral (*y* and *z*) stress through solidifying plate at different times, comparing analytical solution and numerical predictions
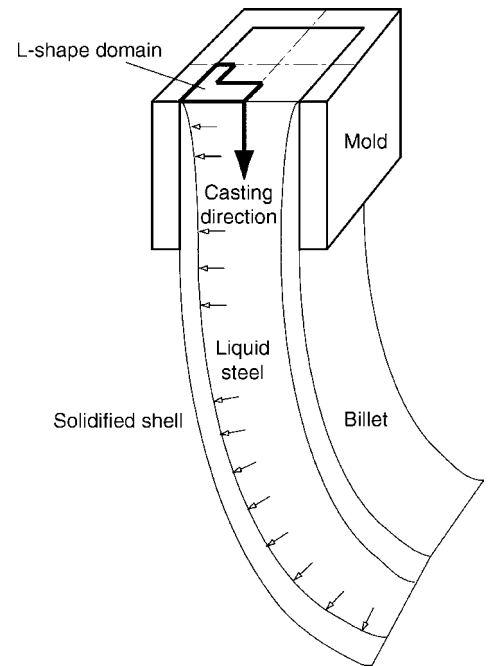
Fig. 6 Model domain

Case Study—Billet Casting Speed Optimization

A Lagrangian model of temperature, distortion, strain, stress, and hot tearing has been applied to predict the maximum speed for continuous casting of steel billets without forming off-corner internal hot-tear cracks. The two-dimensional transient finite-element thermomechanical model, CON2D (Ref 55, 56), has been used to track a transverse slice through the solidifying steel strand as it moves downward at the casting speed to reveal the entire three-dimensional stress state. The two-dimensional assumption produces reasonable temperature predictions, because axial (*z*-direction) conduction is negligible relative to axial advection (Ref 43). In-plane mechanical predictions are also reasonable, because bulging effects are small, and the undiscretized casting direction is modeled with the appropriate condition of generalized plain strain. Other applications with this model include the prediction of ideal taper of the mold walls (Ref 96) and quantifying the effect of steel grade on oscillation mark severity during level fluctuations (Ref 97).

The model domain is an L-shaped region of a two-dimensional transverse section, shown in Fig. 6. Removing the central liquid region saves computation and lessens stability problems related to element locking. Physically, this "trick" is important in two-dimensional domains because it allows the liquid volume to change without generating stress, which mimics the effect of fluid flow into and out of the domain that occurs in the actual open-topped casting process. Simulations start at the meniscus, 100 mm below the mold top, and extend through the 800 mm long mold and below, for a caster with no submold support. The instantaneous heat flux, given in Eq 14, was based on plant measurements (Ref 98). It was assumed to be uniform around the perimeter of the billet surface in order to simulate ideal taper and perfect contact between the shell and mold. Below the mold, the billet surface temperature was kept constant at its circumferential profile at mold exit. This eliminates the effect of spray cooling practice imperfections on submold reheating or cooling and the associated complication for the stress/strain development. A typical plain carbon steel was studied (0.27% C, 1.52% Mn, 0.34% Si) with 1500.7 °C liquidus temperature and 1411.8 °C solidus temperature:

$$q(\text{MW/m}^2) = \begin{cases} 5 - 0.2444t(\text{s}) & t \le 1.0 \text{ s} \\ 4.7556t(\text{s})^{-0.504} & t > 1.0 \text{ s} \end{cases}$$

(Eq 14)

Different constitutive models were used for each phase of the solidifying steel. The following elastic-visco-plastic constitutive equation was developed for the austenite phase (Ref 99) as a function of percent carbon content (%C) by fitting constant strain-rate tensile tests (Ref 100, 101) and constant-load creep tests (Ref 102) to the form in Eq 5 and 7:

$$\dot{\varepsilon}_{eq} = f_{\%C}\langle\sigma_{eq} - \sigma_0\rangle^{1/m}\exp\left(-\frac{4.465\times10^4}{T}\right)$$

where

$$f_{\%C} = 4.655 \times 10^4 + 7.14 \times 10^4(\%\text{C}) + 1.2 \times 10^4(\%\text{C})^2$$

$$\sigma_0 = (130.5 - 5.128 \times 10^{-3}T)\varepsilon_{eq}^{f_2}$$

$$f_2 = -0.6289 + 1.114 \times 10^{-3}T$$

$$1/m = 8.132 - 1.54 \times 10^{-3}T$$

(Eq 15)

with *T* in kelvin, and σ_{eq} and σ_0 in MPa

Further equations, such as the associated flow rule, are needed to transform this scalar equation into tensor form and to account for reversals in loading conditions. Equation 15 and a similar one for delta-ferrite have been implemented into the finite-element codes CON2D (Ref 55) and THERCAST (Ref 103) and applied to investigate several problems involving mechanical behavior during continuous casting.

Elastic modulus is a crucial property that decreases with increasing temperature. It is difficult to measure at the high temperatures important to casting, owing to the susceptibility of the material to creep and thermal strain during a standard tensile test, which results in excessively low values. Higher values are obtained from high-strain-rate tests, such as ultrasonic measurements (Ref 104). Elastic modulus measurements in steels near the solidus temperature range from ~1 (Ref 105) to 44 GPa (Ref 106), with typical modulus values of ~10 GPa near the solidus (Ref 98, 107, 108).

The density needed to compute thermal strain in Eq 8 can be found from a weighted average of the values of the different solid and liquid phases, based on the local phase fractions. For the example of plain low-carbon steel, the following equations were compiled (Ref 55) based on the phase fractions of alpha-ferrite (f_α), austenite (f_γ), delta-ferrite (f_δ) (Ref 109, 110), and liquid (f_l) measurements (Ref 111):

$$\rho(kg/m^3) = \rho_\alpha f_\alpha + \rho_\gamma f_\gamma + \rho_\delta f_\delta + \rho_l f_l$$
$$\rho_\alpha = 7881 - 0.324T(^\circ C) - 3 \times 10^{-5} T(^\circ C)^2$$
$$\rho_\gamma = \frac{100[8106 - 0.51T(^\circ C)]}{[100 - (\%C)][1 + 0.008(\%C)]^3}$$
$$\rho_\delta = \frac{100[8011 - 0.47T(^\circ C)]}{[100 - (\%C)][1 + 0.013(\%C)]^3}$$
$$\rho_l = 7100 - 73(\%C) - [0.8 - 0.09(\%C)]$$
$$[T(^\circ C) - 1550]$$

(Eq 16)

Sample results are presented here for one-quarter of a 120 mm square billet cast at speeds of 2.0 and 5.0 m/min. The latter is the critical speed at which hot-tear crack failure of the shell is just predicted to occur. The temperature and axial (z) stress distributions in a typical section through the wideface of the steel shell cast at 2.0 m/min are shown in Fig. 7 and Fig. 8 at four different times during cooling in the mold. Unlike the analytical solution in Fig. 4, the surface temperature drops as time progresses. The corresponding stress distributions are qualitatively similar to the analytical solution (Fig. 5). The stresses increase with time, however, as solidification progresses. The realistic constitutive equations produce a large region of tension near the solidification front. The magnitude of these stresses (and the corresponding strains) are not predicted to be enough to cause hot tearing in the mold, however. The results from two different codes reasonably match, demonstrating that the formulations are accurately implemented, convergence has been achieved, and that the mesh and time-step refinement are sufficient.

Figure 9(a) shows the distorted temperature contours near the strand corner at 200 mm below the mold exit for a casting speed of 5.0 m/min. The corner region is coldest, owing to two-dimensional cooling. The shell becomes hotter and thinner with increasing casting speed, owing to less time in the mold. This weakens the shell, allowing it to bulge more under the ferrostatic pressure below the mold.

Figure 9(b) shows contours of hoop stress constructed by taking the stress component acting perpendicular to the dendrite growth direction, which simplifies to σ_x in the lower right portion of the domain and σ_y in the upper left portion. High values appear at the off-corner subsurface region, due to a hinging effect that the ferrostatic pressure over the entire face exerts around the corner. This bends the shell around the corner and generates high subsurface tensile stress at the weak solidification front in the off-corner subsurface location. This tensile stress peak increases slightly and moves toward the surface at higher casting speed.

Stress concentration is less and the surface hoop stress is compressive at the lower casting speed. This indicates no possibility of surface cracking. However, tensile surface hoop stress is generated below the mold at high speed in Fig. 9(b) at the face center due to excessive bulging. This tensile stress, and the accompanying hot-tear strain, may contribute to longitudinal cracks that penetrate the surface.

Hot tearing was predicted using the criterion in Eq 10 with the critical strain given in Eq 11. Inelastic strain was accumulated for the component oriented normal to the dendrite growth direction, because that is the weakest direction and corresponds to the measurements used to obtain Eq 11. Figure 9(c) shows contours of hot-tear strain in the hoop direction. The highest values appear at the off-corner subsurface region in the hoop direction. Moreover, significantly higher values are found at higher casting speeds. For this particular example, hot-tear strain exceeds the threshold at 12 nodes, all located near the off-corner subsurface region. This is caused by the hinging mechanism
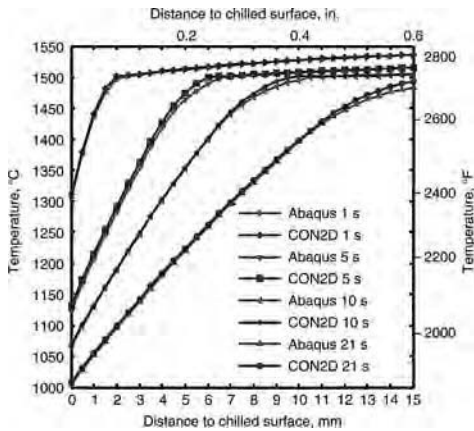
Fig. 7 Temperature distribution along the solidifying slice in continuous casting mold
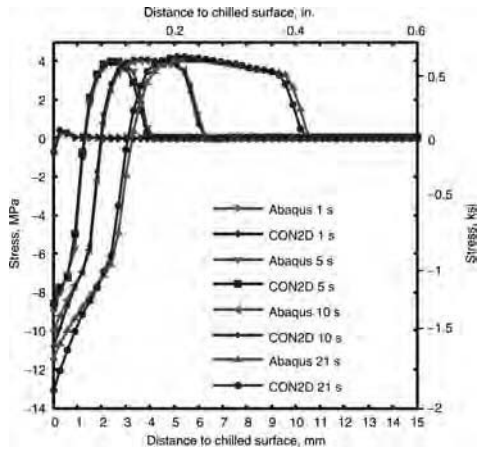
Fig. 8 Lateral (y and z) stress distribution along the solidifying slice in continuous casting mold
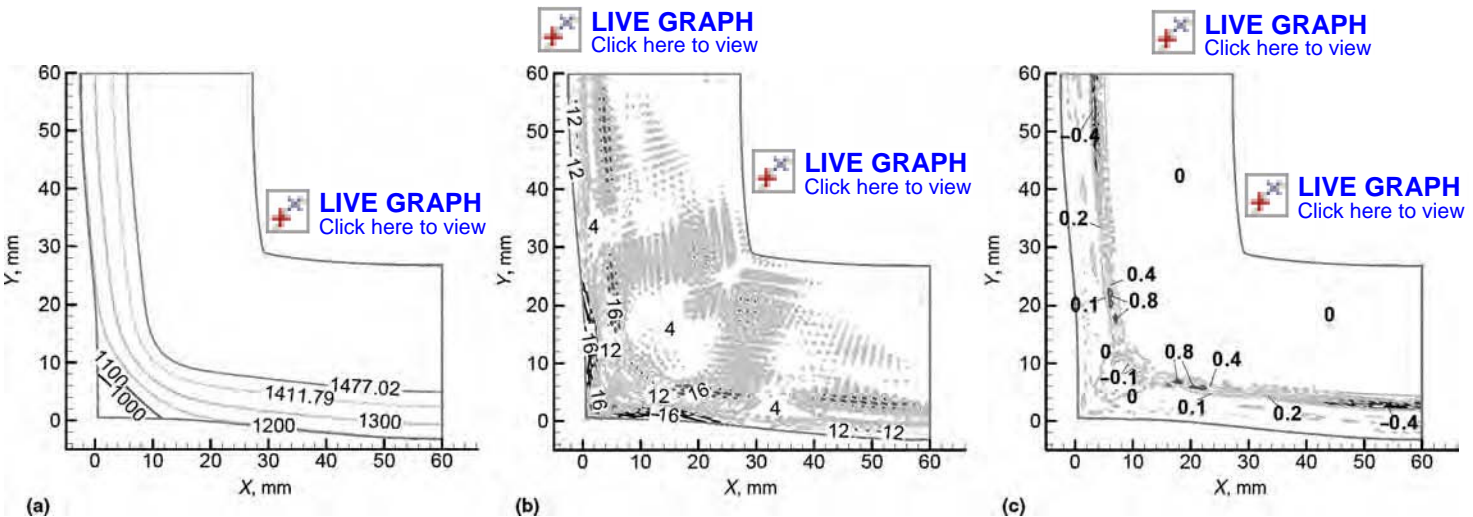
Fig. 9 Distorted contours at 200 mm below mold exit. (a) Temperature. (b) Hoop stress. (c) Hot-tear strain
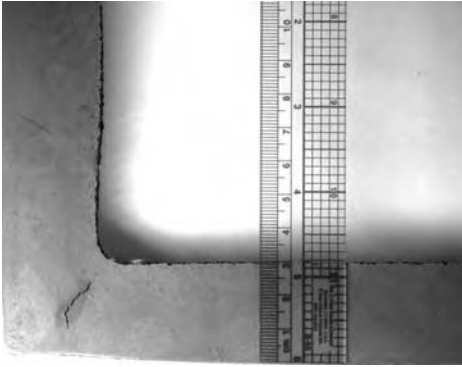
Fig. 10 Off-corner internal hot-tear crack in break-out shell from a 175 mm square bloom

LIVE GRAPH
Click here to view

Fig. 11 Comparison of critical casting speeds, based on hot-tearing criterion and typical plant practice. Source: Ref 112

around the corner. No nodes fail at the center surface, in spite of the high tensile stress there. The predicted hot-tearing region matches the location of off-corner longitudinal cracks observed in sections through real solidifying shells, such as the one pictured in Fig. 10. The bulged shape is also similar.

Results from many computations were used to find the critical speed to avoid hot-tear cracks as a function of section size and working mold length, as presented in Fig. 11 (Ref 108). These predictions slightly exceed plant practice, which is generally chosen by empirical trial and error. This suggests that plant conditions such as mold taper are less than ideal, that other factors limit casting speed, or those speeds in practice could be increased. The qualitative trends are the same.

This quantitative model of hot tearing provides many useful insights into the continuous casting process. Larger section sizes are more susceptible to bending around the corner and thus have a lower critical speed, resulting in less productivity increase than expected. The trend toward longer molds over the past three decades enables a higher casting speed without cracks by producing a thicker, stronger shell at mold exit.

Conclusions

The prediction of defects represents the culmination of solidification modeling. It enables models to make practical contributions to real commercial processes, but it requires incorporating together and augmenting the models of almost every other aspect of casting simulation. Hot-tear crack prediction requires accurate thermal and mechanical analysis, combined with criteria for embrittlement. As computing power and software tools for computational mechanics advance, it is becoming increasingly possible to perform useful analysis of fluid flow, temperature, deformation, strain, stress, and related phenomena in real casting processes. Computations are still hampered by the limits of mesh resolution and computational speed, especially
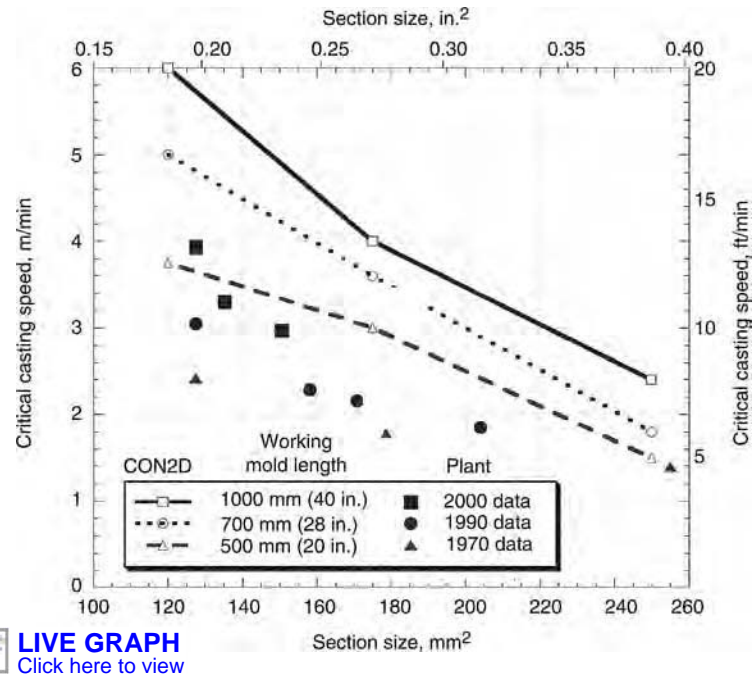
for realistic three-dimensional geometries and defect analysis. The modeling of defects such as hot tears is still in its infancy, and there is much work to be done.

REFERENCES

1. L. Zhang, B.G. Thomas, X. Wang, and K. Cai, Evaluation and Control of Steel Cleanliness—A Review, *Steelmaking Conf. Proc.,* Vol 85, March 10–13, 2002 (Nashville, TN), ISS, Warrendale, PA, 2002, p 431–452.
2. A. Ashrafian and S.T. Johansen, Special Issue: CFD Applications in Metallurgical and Process Industries—Preface, *Progress in Computational Fluid Dynamics,* Vol 7 (2–4), 2007, p 69.
3. FLUENT6.2, Ansys Inc., 10 Cavendish Court, Lebanon, NH, 2009.
4. S.T. Johansen and F. Boysan, Fluid-Dynamics in Bubble Stirred Ladles. 2. Mathematical Modeling, *Metall. Mater. Trans. B—Process Metallurgy,* Vol 19 (5), 1988, p 755–764.
5. B.G. Thomas, Q. Yuan, S. Sivaramakrishnan, T. Shi, S.P. Vanka, and M.B. Assar, Comparison of Four Methods to Evaluate Fluid Velocities in a Continuous Slab Casting Mold, *ISIJ Int. (Japan),* Vol 41 (10), 2001, p 1262–1271.
6. Q. Yuan, B. Zhao, S.P. Vanka, and B.G. Thomas, Study of Computational Issues in Simulation of Transient Flow in Continuous Casting, *Steel Res. Int.,* Vol 76 (1, Special Issue: Simulation of Fluid Flow in Metallurgy), 2005, p 33–43.
7. Q. Yuan, S. Sivaramakrishnan, S.P. Vanka, and B.G. Thomas, Computational and Experimental Study of Turbulent Flow in a 0.4-Scale Water Model of a Continuous Steel Caster, *Metall. Mater. Trans.,* Vol 35B (5), 2004, p 967–982.
8. B. Sundman, B. Jansson, and J.-O. Andersson, The Thermo-Calc Databank System, *Calphad,* Vol 9, 1985, p 153–190.
9. A.D. Pelton, C.W. Bale, G. Eriksson, K. Hack, S. Petersen, and P. Koukkari, Demonstration of FACTSage and ChemApp at HTMC-X, *High Temperature Materials Chemistry, Abstracts of the Tenth International IUPAC Conference*, K. Hilpert, F.W. Froben, and L. Singheiser, Ed., April 10–14, 2000 (Jülich, Germany), 2000, p 77.
10. A.T. Dinsdale, S.M. Hodson, T.I. Barry, and J.R. Taylor, Computations Using MTDATA of Metal-Matte-Slag-Gas

Equilibria, *Computer Software in Chemical and Extractive Metallurgy,* Pergamon Press, New York, NY, 1989, p 59–74.

11. B. Cheynet, Complex Chemical Equilibria Calculations with the THERMODATA System, *Computer Software in Chemical and Extractive Metallurgy,* Pergamon Press, New York, NY, 1989, p 31–44.

12. A.W. Cramb, Y. Chung, J. Harman, A. Sharan, and I. Jimbo, The Slag/Metal Interface and Associated Phenomena. A History of Pneumatic Steelmaking, *Iron Steelmaker (USA),* Vol 24 (3), 1997, p 77–83.

13. L. Wang and C. Beckermann, Prediction of Reoxidation Inclusion Composition in Casting of Steel, *Metall. Mater. Trans. B—Process Metallurgy and Materials Processing Science,* Vol 37 (4), 2006, p 571–588.

14. B.G. Thomas and L. Zhang, Review: Mathematical Modeling of Fluid Flow in Continuous Casting, *ISIJ Int.,* Vol 41 (10), 2001, p 1181–1193.

15. Q. Yuan, B.G. Thomas, and S.P. Vanka, Study of Transient Flow and Particle Transport during Continuous Casting of Steel Slabs, Part 2: Particle Transport, *Metall. Mater. Trans. B.,* Vol 35B (4), 2004, p 703–714.

16. L. Zhang, J. Aoki, and B.G. Thomas, Inclusion Removal by Bubble Flotation in a Continuous Casting Mold, *Metall. Mater. Trans. B,* Vol 37B, 2006, p 361–379.

17. T. Nakaoka, S. Taniguchi, K. Matsumoto, and S.T. Johansen, Particle-Size-Grouping Method of Inclusion Agglomeration and Its Application to Water Model Experiments, *ISIJ Int.,* Vol 41 (10), 2001, p 1103–1111.

18. L. Zhang and S. Taniguchi, Fundamentals of Inclusions Removal from Liquid Steel by Bubbles Flotation, *Int. Mater. Rev.,* Vol 45 (2), 2000, p 59–82.

19. K.G. Rackers and B.G. Thomas, Clogging in Continuous Casting Nozzles, *Continuous Casting,* Vol 10, *Tundish Operations,* Iron and Steel Society, Warrendale, PA, 2003, p 264–274.

20. A.V. Catalina, S. Mukherjee, and D.M. Stefanescu, A Dynamic Model for the Interaction between a Solid Particle and an Advancing Solid/Liquid Interface, *Metall. Mater. Trans. A,* Vol 31A (10), 2000, p 2559–2568.

21. Q. Yuan and B.G. Thomas, Transport and Entrapment of Particles in Continuous Casting of Steel, *Third Int. Congress on Science and Technology of Steelmaking,* May 9–11, 2005 (Charlotte, NC), Association for Iron and Steel Technology, Warrendale, PA, 2005, p 745–762.

22. M. Serriere, C.A. Gandin, E. Gautier, P. Archambault, and M. Dehmas, Modeling of Precipitation Coupled with Thermodynamic Calculations, *Aluminum Alloys 2002: Their Physical and Mechanical Properties Pts 1–3,* Vol 396-4, 2002, p 747–752.

23. G. Lesoult, C.A. Gandin, and N.T. Niane, Segregation during Solidification with Spongy Deformation of the Mushy Zone, *Acta Mater.,* Vol 51, 2003, p 5263–5283.

24. K. Miyazawa and K. Schwerdtfeger, Macrosegregation in Continuously Cast Steel Slabs: Preliminary Theoretical Investigation on the Effect of Steady State Bulging, *Arch. Eisenhutten,* Vol 52 (11), 1981, p 415–422.

25. A.F. Giamei and B.H. Kear, On the Nature of Freckles in Nickel Base Superalloys, *Metall. Trans.,* Vol 1, 1970, p 2185–2192.

26. H.J. Thevik, A. Mo, and T. Rusten, A Mathematical Model for Surface Segregation in Aluminum Direct Chill Casting, *Metall. Mater. Trans. B—Process Metallurgy and Materials Processing Science,* Vol 30 (1), 1999, p 135–142.

27. C. Beckermann, Modelling of Macrosegregation: Applications and Future Needs, *Int. Mater. Rev.,* Vol 47 (5), 2002, p 243–262.

28. J.A. Dantzig and M. Rappaz, *Solidification,* Taylor and Francis, EPFL Publishing, 2009, p 250.

29. M.C. Flemings and G.E. Nereo, Macrosegregation, *Trans. AIME,* Vol 239, 1967, p 1449–1461.

30. E. Niyama, T. Uchida, M. Morikawa, and S. Saito, Method of Shrinkage Prediction and Its Application to Steel Casting Practice, *AFS Int. Cast Metals J.,* Vol 7, 1982, p 52–63.

31. K.D. Carlson and C. Beckermann, Prediction of Shrinkage Pore Volume Fraction Using a Dimensionless Niyama Criterion, *Metall. Mater. Trans. A—Physical Metallurgy and Materials Science,* Vol 40A (1), 2009, p 163–175.

32. K.D. Carlson, Z.P. Lin, C. Beckermann, G. Mazurkevich, and M. Schneider, Modeling of Porosity Formation in Aluminum Alloys, *Aluminium Alloys 2006, Pts 1 and 2,* Vol 519–521, 2006, p 1699–1706.

33. P.D. Lee, A. Chirazi, and D. See, Modeling Microporosity in Aluminum-Silicon Alloys: A Review, *J. Light Met.,* Vol 1 (1), 2001, p 15–30.

34. K. Venkatesan and R. Shivpuri, Experimental and Numerical Investigation of the Effect of Process Parameters on the Erosive Wear of Die-Casting Dies, *J. Mater. Eng. Perform.,* Vol 4 (2), 1995, p 166–174.

35. K. Domkin, J. Thorborg, and J. Hattel, Modeling of High-Temperature and Diffusion-Controlled Die Soldering in Aluminum High-Pressure Die Casting, *J. Mater. Process. Technol.,* 2008, in press.

36. Q. Han and S. Viswanathan, Analysis of the Mechanism of Die Soldering in Aluminum Die Casting, *Metallurg. Mater. Trans. A,* Vol 34 (1), 2003, p 139–146.

37. S. Shankar and D. Apelian, Die Soldering: Mechanism of the Interface Reaction between Molten Aluminum Alloy and Tool Steel, *Metall. Mater. Trans. B,* Vol 33 (3), 2002, p 465–476.

38. J.-K. Park, B.G. Thomas, I.V. Samarasekera, and U.-S. Yoon, Thermal and Mechanical Behavior of Copper Moulds during Thin Slab Casting (II): Mould Crack Formation, *Metall. Mater. Trans.,* Vol 33B, June 2002, p 437–449.

39. T. O'Conner and J. Dantzig, Modeling the Thin Slab Continuous Casting Mold, *Metall. Mater. Trans. B,* Vol 25B (4), 1994, p 443–457.

40. M. Asta, C. Beckermann, A. Karma, W. Kurz, R. Napolitano, M. Plapp, G. Purdy, M. Rappaz, and R. Trivedi, Solidification Microstructures and Solid-State Parallels: Recent Developments, Future Directions, *Acta Mater.,* Vol 57 (4), 2009, p 941–971.

41. W.A. Curtin and R.E. Miller, Atomistic/Continuum Coupling in Computational Materials Science, *Model. Simul. Mater. Sci. Eng.,* Vol 11 (3), 2003, p R33–R68.

42. J.E. Kelly, K.P. Michalek, T.G. O'Connor, B.G. Thomas, and J.A. Dantzig, Initial Development of Thermal and Stress Fields in Continuously Cast Steel Billets, *Metall. Trans. A,* Vol 19A (10), 1988, p 2589–2602.

43. Y. Meng and B.G. Thomas, Heat Transfer and Solidification Model of Continuous Slab Casting: CON1D, *Metall. Mater. Trans.,* Vol 34B (5), 2003, p 685–705.

44. C.V. Madhusudana and L.S. Fletcher, Contact Heat Transfer—The Last Decade, *AIAA J.,* Vol 24, 1985, p 510–523.

45. G.E. Mase and G.T. Mase, *Continuum Mechanics for Engineers*, 2nd ed., CRC Press, 1999.

46. Y. Estrin, A Versatile Unified Constitutive Model Based on Dislocation Density Evolution, *Constitutive Modelling— Theory and Application*, Vol MD-Vol 26/AMD-Vol 121, ASME, New York, 1991, p 65–75.

47. C. Agelet de Saracibar, M. Cervera, and M. Chiumenti, On the Constitutive Modeling of Coupled Thermomechanical Phase-Change Problems, *Int. J. Plast.,* Vol 17, 2001, p 1565–1622.

48. J. Sengupta, S.L. Cockcroft, D.M. Maijer, and A. Larouche, Quantification of Temperature, Stress, and Strain Fields during the Start-Up Phase of Direct Chill Casting Process by Using a 3-D Fully Coupled Thermal and Stress Model for AA5182 Ingots, *Mater. Sci. Eng. A,* Vol 397, 2005, p 157–177.

49. A.E. Huespe, A. Cardona, and V. Fachinotti, Thermomechanical Model of a Continuous Casting Process, *Computer Methods in Appl. Mech. Engr.,* Vol 182 (3), 2000, p 439–455.

50. *MARC Users Manual,* MARC Analysis Research Corp., Palo Alto, CA, 1991.

51. "ABAQUS Theory Manual v6.0," Abaqus, Inc., Pawtucket, RI, 2004.

52. A. Hakonsen and D. Mortensen, *Modeling of Casting, Welding and Advanced Solidification Processes VII,* The Minerals, Metals and Materials Society, Warrendale, PA, 1995, p 963–970.

53. H.G. Fjaer and A. Mo, ALSPEN—A Mathematical Model for Thermal Stresses in DC-Cast Al Billets, *Metall. Trans.,* Vol 21B (6), 1990, p 1049–1061.

54. G. Laschet, J. Jakumeit, and S. Benke, Thermo-Mechanical Analysis of Cast/ Mould Interaction in Casting Processes, *Z. Metallkd.,* Vol 95, 2004, p 1087–1096.

55. C. Li and B.G. Thomas, Thermo-Mechanical Finite-Element Model of Shell Behavior in Continuous Casting of Steel, *Metall. Mater. Trans. B.,* Vol 35B (6), 2004, p 1151–1172.

56. S. Koric and B.G. Thomas, Efficient Thermo-Mechanical Model for Solidification Processes, *Int. J. Num. Meths. Eng.,* Vol 66 (12), 2006, p 1955–1989.

57. MAGMASOFT, http://www.magmasoft. com, 2007.

58. PROCAST, Procast, http://www.esi-group. com/SimulationSoftware/Die_Casting_So- lution/Products/Casting_simulation/, 2007.

59. M. Samonds and J.Z. Zhu, Coupled Thermal-Fluid-Stress Analysis of Castings, *Proc. MCWASP IX, Ninth Int. Conf. on Modeling of Casting, Welding and Advanced Solidification Processes,* P.R. Sahm, P.N. Hansen, and J.G. Conley, Ed., August 20–25, 2000 (Aachen Germany), Shaker Verlag, Aachen, 2000, p 80–87.

60. J.M. Drezet, A. Bughardt, H.G. Fjaer, and B. Magnin, *Mater. Sci. Forum 2000,* Vol 329–330, 2000, p 493–500.

61. H.G. Fjaer and A. Hakonsen, The Mechanism of Pull-In during DC-Casting of Aluminum Sheet Ingots, *Light Metals 1997,* The Minerals, Metals and Materials Society, Warrendale, PA, 1997, p 683–690.

62. A. Mo, M. Rappaz, and L.L. Martin, *Aluminum,* Vol 78 (10), 2002, p 856–864.

63. A. Henry, *Light Met. Age,* Vol 58 (7–8), 2000, p 66–67.

64. M. Bellet, O. Jaouen, and I. Poitrault, An ALE-FEM Approach to the Thermomechanics of Solidification Processes with Application to the Prediction of Pipe Shrinkage, *Int. J. Num. Meth. Heat Fluid Flow,* Vol 15, 2005, p 120–142.

65. S. Koric, L.C. Hibbeler, and B.G. Thomas, Explicit Coupled Thermo-Mechanical Finite Element Model of Steel Solidification, *Int. J. Num. Meths. Engineering,* 2009, submitted.

66. O.C. Zienkiewicz and R.L. Taylor, *The Finite Element Method,* 4th ed., McGraw Hill, New York, NY, 1988.

67. J.-M. Drezet, B. Commet, H.G. Fjaer, and B. Magnin, Stress-Strain Computations of the DC Casting Process of Aluminum Alloy: A Sensitivity Study on Material Properties, *Proc. MCWASP IX, Ninth Int.*

Conf. on Modeling of Casting, Welding and Advanced Solidification Processes, P.R. Sahm, P.N. Hansen, and J.G. Conley, Ed., August 20–25, 2000 (Aachen Germany), Shaker Verlag, Aachen, 2000, p 33–40.

68. X. Huang, B.G. Thomas, and F.M. Najjar, Modeling Superheat Removal during Continuous Casting of Steel Slabs, *Metall. Trans. B,* Vol 23B (6), 1992, p 339–356.

69. M. Bellet and V.D. Fachinotti, ALE Method for Solidification Modelling, *Comput. Methods Appl. Mech. Eng.,* Vol 193, 2004, p 4355–4381.

70. Thercast presentation, www.transvalor. com and www.scconsultants.com, 2007.

71. M. Chiumenti, M. Cervera, and C.A.D. Saracibar, Coupled Thermomechanical Simulation of Solidification and Cooling Phases in Casting Processes, *Proc. MCWASP XI, 11th Int. Conf. on Modeling of Casting, Welding and Advanced Solidification Processes,* C.-A. Gandin and M. Bellet, Ed., The Minerals, Metals and Materials Society, Warrendale, PA, 2006, p 201–208.

72. B.G. Thomas and M. Bellet, Modeling of Stress, Distortion, and Hot Tearing, *Casting,* Vol 15, *ASM Handbook,* S. Viswanathan and E. DeGuire, Ed., ASM International, 2008, p 449–461.

73. T.W. Clyne and G.J. Davies, Comparison between Experimental Data and Theoretical Predictions Relating to Dependence of Solidification Cracking on Composition, *Solidification and Casting of Metals,* The Metals Society, London, 1979, p 275–278.

74. P. Ackermann, W. Kurz, and W. Heinemann, In Situ Tensile Testing of Solidifying Aluminum and Al-Mg Shells, *Mater. Sci. Eng.,* Vol 75, 1985, p 79–86.

75. C. Bernhard, H. Hiebert, and M.M. Wolf, Simulation of Shell Strength Properties by the SSCT Test, *ISIJ Int. (Japan),* Vol 36 (Suppl. Science and Technology of Steelmaking), 1996, p S163–S166.

76. M. Suzuki, C. Yu, and T. Emi, In-Situ Measurement of Tensile Strength of Solidifying Steel Shells to Predict Upper Limit of Casting Speed in Continuous Caster with Oscillating Mold, *ISIJ Int., Iron Steel Inst. Japan,* Vol 37 (4), 1997, p 375–382.

77. K. Kinoshita, T. Emi, and M. Kasai, Thermal Elasto-Plastic Stress Analysis of Solidifying Shell in Continuous Casting Mold, *Tetsu-to-Hagane,* Vol 65 (14), 1979, p 2022–2031.

78. J.O. Kristiansson, Thermal Stresses in the Early Stage of Solidification of Steel, *J. Therm. Stresses,* Vol 5, 1982, p 315–330.

79. B.G. Thomas, I.V. Samarasekera, and J.K. Brimacombe, Mathematical Model of the Thermal Processing of Steel Ingots, Part II: Stress Model, *Metall. Trans. B,* Vol 18B (1), 1987, p 131–147.

80. K. Okamura and H. Kawashima, Calculation of Bulging Strain and Its Application to Prediction of Internal Cracks in Continuously Cast Slabs, *Proc. Int. Conf. Comp. Assoc. Mat. Design Proc. Simul.,* ISIJ, Tokyo, 1993, p 129–134.

81. A. Yamanaka, K. Nakajima, K. Yasumoto, H. Kawashima, and K. Nakai, Measurement of Critical Strain for Solidification Cracking, *Modelling of Casting, Welding, and Advanced Solidification Processes—V,* M. Rappaz, M.R. Ozgu, and K.W. Mahin, Ed. (Davos, Switzerland), TMS, Warrendale, PA, 1990, p 279–284.

82. P. Wisniewski and H.D. Brody, Tensile Behavior of Solidifying Aluminum Alloys, *Modelling of Casting, Welding, and Advanced Solidification Processes— V,* M. Rappaz, M.R. Ozgu, and K.W. Mahin, Ed. (Davos, Switzerland), TMS, Warrendale, PA, 1990, p 273–278.

83. A. Yamanaka, K. Nakajima, and K. Okamura, Critical Strain for Internal Crack Formation in Continuous Casting, *Ironmaking Steelmaking,* Vol 22 (6), 1995, p 508–512.

84. Y.-M. Won, T.J. Yeo, D.J. Seol, and K.H. Oh, A New Criterion for Internal Crack Formation in Continuously Cast Steels, *Metall. Mater. Trans. B,* Vol 31B, 2000, p 779–794.

85. U. Feurer, Mathematisches Modell der Warmrissneigung von Binären Aluminium Legierungen, *Giessereiforschung,* Vol 28, 1976, p 75–80.

86. M. Rappaz, J.-M. Drezet, and M. Gremaud, A New Hot-Tearing Criterion, *Metall. Mater. Trans. A,* Vol 30A (2), 1999, p 449–455.

87. J.M. Drezet and M. Rappaz, Prediction of Hot Tears in DC-Cast Aluminum Billets, *Light Metals,* J.L. Anjier, Ed. TMS, Warrendale, PA, 2001, p 887–893.

88. M. M'Hamdi, S. Benum, D. Mortensen, H. G. Fjaer, and J.M. Drezet, The Importance of Viscoplastic Strain Rate in the Formation of Center Cracks during the Start-Up Phase of Direct-Chill Cast Aluminium Extrusion Ingots, *Metall. Mater. Trans. A,* Vol 34, 2003, p 1941–1952.

89. M. M'Hamdi, A. Mo, and H.G. Fjaer, *Metall. Mater. Trans. A,* Vol 37, 2006, p 3069–3083.

90. V. Mathier, S. Vernede, P. Jarry, and M. Rappaz, Two-Phase Modeling of Hot Tearing in Aluminum Alloys: Applications of a Semicoupled Method, *Metall. Mater. Trans. A—Physical Metallurgy and Materials Science,* Vol 40A (4), 2009, p 943–957.

91. D.G. Eskin, Suyitno, and L. Katgerman, Mechanical Properties in the Semi-Solid State and Hot Tearing of Aluminum Alloys, *Prog. Mat. Sci.,* Vol 49, 2004, p 629–711.

92. A.B. Phillion, S.L. Cockcroft, and P.D. Lee, A Three-Phase Simulation of the Effect of Microstructural Features on Semi-Solid Tensile Deformation, *Acta Mater.,* Vol 56 (16), 2008, p 4328–4338.

93. Y.M. Won and B.G. Thomas, Simple Model of Microsegregation during Solidification of Steels, *Metall. Mater. Trans. A (USA),* Vol 32A (7), 2001, p 1755–1767.

94. T. Matsumiya, H. Kajioka, S. Mizoguchi, Y. Ueshima, and H. Esaka, *Trans. Iron Steel Inst. Jpn.,* Vol 24, 1984, p 873–882.

95. J.H. Weiner and B.A. Boley, Elasto-Plastic Thermal Stresses in a Solidifying Body, *J. Mech. Phys. Solids,* Vol 11, 1963, p 145–154.

96. B.G. Thomas and C. Ojeda, Ideal Taper Prediction for Slab Casting, *ISSTech Steelmaking Conference,* Vol 86, April 27–30, 2003 (Indianapolis, IN), 2003, p 396–308.

97. J. Sengupta and B.G. Thomas, Effect of a Sudden Level Fluctuation on Hook Formation during Continuous Casting of Ultra-Low Carbon Steel Slabs, *Modeling of Casting, Welding, and Advanced Solidification Processes XI (MCWASP XI) Conference*, M. Bellet, C.Z. Gandin, and J.E. Allision, Ed., May 28–June 2, 2006 (Opio, France), 2006, p 727–736.

98. C. Li and B.G. Thomas, Maximum Casting Speed for Continuous Cast Steel Billets Based on Sub-Mold Bulging Computation, *Steelmaking Conf. Proc.,* Vol 85, March 10–13, 2002 (Nashville, TN), ISS, Warrendale, PA, 2002, p 109–130.

99. P. Kozlowski, B.G. Thomas, J. Azzi, and H. Wang, Simple Constitutive Equations for Steel at High Temperature, *Metall. Trans. A,* Vol 23A (3), 1992, p 903–918.

100. P.J. Wray, Plastic Deformation of Delta-Ferritic Iron at Intermediate Strain Rates, *Metall. Trans. A,* Vol 7A, Nov 1976, p 1621–1627.

101. P.J. Wray, Effect of Carbon Content on the Plastic Flow of Plain Carbon Steels at Elevated Temperatures, *Metall. Trans. A,* Vol 13A (1), 1982, p 125–134.

102. T. Suzuki, K.H. Tacke, K. Wunnenberg, and K. Schwerdtfeger, Creep Properties of Steel at Continuous Casting Temperatures, *Ironmaking Steelmaking,* Vol 15 (2), 1988, p 90–100.

103. F. Costes, A. Heinrich, and M. Bellet, 3-D Thermomechanical Simulation of the Secondary Cooling Zone of Steel Continuous Casting, *Proc. MCWASP X, Tenth Int. Conf. on Modeling of Casting, Welding and Advanced Solidification Processes*, D.M. Stefanescu, J.A. Warren, M.R.J. and M.J.M. Krane, Ed., The Minerals, Metals and Materials Society, Warrendale, PA, 2003, p 393–400.

104. D.L. Donsbach and M.W. Moyer, Ultrasonic Measurement of Elastic Constants at Temperatures from 20 to 1100 °C, *Ultrasonic Materials Characterization,* H. Berger and M. Linzer, Ed., June 7–9, 1978 (Gaithersburg, MD), National Bureau of Standards, Special Pub. 596, Nov 1980.

105. O.M. Puhringer, Strand Mechanics for Continuous Slab Casting Plants, *Stahl Eisen,* Vol 96 (6), 1976, p 279–284.

106. D.R. Hub, "Measurement of Velocity and Attenuation of Sound in Iron up to the Melting Point," paper 551, Proc. IVth Int. Vong. Acoustics (Copenhagen), 1962.

107. H. Mizukami, K. Murakami, and Y. Miyashita, Mechanical Properties of Continuously Cast Steels at High Temperatures, *Tetsu-to-Hagane,* Vol 63 (146), 1977, p S 652.

108. C. Li and B.G. Thomas, Thermo-Mechanical Finite Element Model of Bulging and Hot Tearing during Continuous Casting of Steel Billets, *Modeling of Casting, Welding, and Advanced Solidification Processes*, Vol X, D. Stefanescu, J. Warren, M. Jolly, and M. Krane, Ed., May 25–30, 2003 (San Destin, FL), TMS, Warrendale, PA, 2003, p 385–392.

109. K. Harste, A. Jablonka, and K. Schwerdtfeger, Shrinkage and Formation of Mechanical Stresses during Solidification of Round Steel Strands, *Fourth Int. Conf. on Continuous Casting* (Centres de Recherches Metallurgiques and Verein Deutscher Eisenhuttenleute), Stahl und Eisen, Brussels, 1988, p 633–644.

110. K. Harste, "Investigation of the Shrinkage and the Origin of Mechanical Tension during the Solidification and Successive Cooling of Cylindrical Bars of Fe-C Alloys," Ph.D. dissertation thesis, Technical University of Clausthal, 1989.

111. I. Jimbo and A. Cramb, The Density of Liquid Iron-Carbon Alloys, *Metall. Trans. B,* Vol 24B, 1993, p 5–10.

112. E. Howard and D. Lorento, Development of High Speed Casting, *1996 Electric Furnace Conference Proceedings,* Dec 9–12, 1996 (Dallas, TX), ISS, Warrendale, PA, 1996.

Phenomenological or Mechanistic Models for Mechanical Properties

Modeling of Tensile Properties

Peter C. Collins and Hamish L. Fraser, The Ohio State University

---

TENSILE PROPERTIES OF METALLIC MATERIALS are characterized by the elastic modulus (i.e., Young's modulus) exhibited by a material prior to the formation of a permanent set, the yield stress, ultimate tensile stress (UTS), and a measure of ductility prior to failure (usually either the elongation to failure and/or the reduction in area). While considerable progress has been made regarding the prediction of the elastic modulus of metals and fairly simple alloys using ab initio, quantum mechanical calculations, the plastic properties are considerably more complex. Thus, in the case of these latter properties, there are a number of approaches, including phenomenological, mechanistic (including atomistic), and rules-based, that are the subject of considerable current study. Considering the three plastic properties, yield stress is difficult to predict, UTS very difficult, and ductility extremely difficult. The reason why these properties present such difficulties regarding computation is because in any given (useful) alloy containing a number of alloying elements, a number of strengthening mechanisms, for example, can contribute. In general, it is not known which of these mechanisms is contributing significantly to the value of the given property, and the specific roles of each of the alloying elements are often not well understood.

In metallic materials, plastic deformation generally occurs by dislocation glide, twinning, phase transformations, climb of dislocations, grain-boundary sliding, and diffusion of point defects. Which of these deformation processes is activated depends largely on intrinsic properties, for example, the Peierls stress, and extrinsic factors such as temperature and imposed strain rate. Generally, the first three processes are associated with tensile properties, and the latter three with creep. Strengthening mechanisms are those that inhibit these deformation processes. For example, if a material deforms by dislocation glide (by far the most common process), then the material can be strengthened if glide is made more difficult, because a larger tensile stress must be applied to effect dislocation motion and hence plastic strain. The most common strengthening mechanisms are:

- Solid-solution strengthening
- Age/precipitation hardening
- Dispersion strengthening
- Grain size reduction
- Strengthening from cold work
- Strengthening from interfaces

Regarding ductility, in general, the more effectively the given strengthening mechanism inhibits the deformation processes, the greater will be the reduction in ductility. Simply considered, this is because the increased strengthening results in an increase in the applied stress needed to produce plastic strain. For a given material, the applied stress will eventually exceed a critical value where fracture processes become active, leading to failure. The situation is complicated by significant differences between the global state of stress in a given specimen compared with local states of stress, which can be rather different. Reference is made to the occurrence of stress concentrations that can occur because of extrinsic factors, such as at small, nondeforming particles, particularly those with an elliptical shape with sharp major radii of curvature, for example, impurity inclusions, and intrinsic processes, such as the formation of a dislocation pileup at an obstacle, for example, at a grain boundary. In this latter case, the stress concentration is proportional to the number of dislocations in the pileup, which can be a significant number, leading to a substantial stress concentration. Thus, while on average, the applied stress may be at a value somewhat lower than the critical value for activation of failure mechanisms, locally, stress concentrations may be such as to increase the local stress above that critical value, leading, for example, to local crack initiation. The occurrence and influence of these stress concentrations cause significant difficulties in terms of developing predictive models.

It would be very advantageous if materials engineers and designers would have a set of computational tools for the prediction of the interrelationships between microstructure and properties, and for this article, specifically tensile properties, for a wide range of materials. As noted previously, in any given metallic material, there are a number of contributing strengthening mechanisms. A computational tool would require the contribution of each of the strengthening mechanisms to be predicted and then summed in an appropriate way to derive an estimate of the tensile properties. As seen in the following, where each of the strengthening mechanisms are described and the current state of modeling assessed, while general understandings have been developed, accurate predictions are not typically available. Because of space limitations, the focus of this article is on deformation mechanisms pertinent to structural materials; the extremely interesting case of deformation in nanoscaled materials, such as multilayered systems, is not discussed.

Current State of Understanding and Modeling of Strengthening Mechanisms

Solid-Solution Strengthening

The intentional (and sometimes unintentional) tailoring of interstitial and substitutional solid solutions is arguably one of the most important methods by which the strength of alloys may be affected. Consider the effects of solid-solution strengthening on:

- Steels, where it is well established that interstitial carbon and nitrogen have the largest effect on yield strength and hardness (e.g., $\Delta\sigma_{ys} \sim 3000$ MPa/wt%C)
- Titanium-base alloys, where the effect of interstitial oxygen has also been known to markedly increase the yield strength (Ref 1), as is reflected in the differences in yield strength among the grades of commercially pure titanium (e.g., $\sigma_{ys}^{\text{Grade 1, } \sim0.18O} \sim 220$ MPa; $\sigma_{ys}^{\text{Grade 4, } \sim0.4O} \sim 560$ MPa)
- Al-Si-Cu alloys, where the supersaturation of copper (and early stages of precipitation, Guinier-Preston-Bagaryatskii zones) have been identified as the significant contributor to strength (Ref 2)
- Die-cast magnesium alloys, where solid-solution strengthening is one of the key design considerations in their development (Ref 3)
- Cast magnesium alloys, where both aluminum and zinc are considered to exhibit rather potent solid-solution strengthening effects (Ref 3, 4)

Clearly, solid-solution strengthening is an important mechanism for inclusion in models for the prediction of tensile properties of an alloy.

Legacy Understanding. The strengthening effect of solid solutions is due to the interaction of dislocations with solute atoms. The origin of this effect has been attributed to either the interactions between strain fields of dislocations with those of solute atoms and/or the difference in elastic modulus between the solvent and solute atoms. In both cases, the interaction can be one where the energy of the dislocation is reduced (an attractive interaction) or increased (a repulsive interaction) when interacting with solute atoms. Thus, when the interaction of strain fields results in a decrease in the overall amount of strain (e.g., when a large substitutional solute occupies a lattice site below the extra half-plane of atoms in an edge dislocation, i.e., where the strains are tensile in nature), the line energy of the dislocation will be reduced, as will the degree of misfit about the solute. Likewise, when the elastic modulus of the solute is lower than that of the solvent atoms, a decrease in the line energy will also result, because the strain energy (E_s) of a dislocation is proportional to $G \cdot \mathbf{b}^2$, where G is the shear modulus, and $\mathbf{b}$ is the magnitude of the Burgers vector. In the case of an attractive interaction, the solute atoms will strongly interact with dislocations, and will tend to "lock" the dislocations. Motion of such dislocations requires the defects to break away from the interacting solute atoms, requiring an increased applied shear stress; that is, the material is strengthened. For repulsive interactions, the increase in strain energy of the dislocations if these defects are moved close to such solute atoms will also require an increase in the applied shear stress, and so, the material will appear to be stronger.

Regarding the interaction of strain fields, it is necessary for the strain fields to be of the same kind, that is, hydrostatic (or dilatational) or shear. In principle, the strain fields associated with screw dislocations are shear in nature, whereas for edge dislocations there are both shear and dilatational distortions, the latter arising from displacements parallel to $\pm[\mathbf{b} \times \mathbf{u}]$, where $\mathbf{b}$ is the Burgers vector and $\mathbf{u}$ the dislocation line direction. Mixed dislocations will also have components of both types of strain field, with their respective magnitudes depending on the angle

between their Burgers vector and line direction. Regarding the strain fields about solute atoms, for metallic materials with close-packed crystal structures, substitutional and interstitial solutes generally exhibit hydrostatic strain fields, as do solutes in non-close-packed systems (e.g., body-centered cubic, or bcc), whereas interstitial solutes in these latter systems tend to exhibit asymmetrical distortions, such that both dilatational and shear strains exist. Hence, it can be expected that reasonable interactions will exist between, on the one hand, edge and mixed dislocations and, on the other hand, substitutional and interstitial solutes, whereas, in principle, strong interactions are expected between interstitial solutes and screw dislocations only in bcc materials. These various interactions are summarized in Table 1. In fact, in close-packed systems, dissociation into partial dislocations occurs, for example, Shockley partials in face-centered cubic (fcc) materials. Because the Burgers vectors of these defects are inclined to one another, it is not possible for both of them to be in the screw orientation simultaneously, and so, a weak interaction between at least one of the partials and both interstitial and substitutional solutes in fcc will occur.

Modeling of Solid-Solution Strengthening. Based on the spatial and temporal nature of the interactions between solutes and dislocations, two separate approaches have been developed, which may be classified as:

- Those that attempt to model the interactions of a moving dislocation in a random fixed dispersion of solute atoms
- Those that attempt to model the nucleation of slip at a barrier of solute atoms

A third type of dislocation/solute interaction, solute drag, can be used to explain a solute atmosphere that diffuses at a rate similar to the velocity of the dislocation. However, because this dislocation/solute atom interaction requires either elevated temperatures or low strain rates, it is not expected to be relevant for most tensile property models. For example, it may be important to consider solute drag if the models to be developed are of flow stresses of the material at certain forming conditions.

Model Type 1: A Moving Dislocation in a Random Fixed Dispersion of Solute Atoms. For solute-rich, homogeneously distributed substitutional-solid solutions, the original model of Mott and Nabarro (Ref 5) included as a primary variable the size-misfit factor, ε_b ($=(1/b)(\mathrm{d}b/\mathrm{d}c)$) in their final derivation, given by:

$$\tau = G\varepsilon_b^2 c^{5/3} (\ln c)^2$$

where, in addition to ε_b, the other variables are the modulus (G) and the solute concentration (c). The fundamental limitations of the original derivation include its inability to explain the apparent effect that valencies have on the strengthening effect (Ref 6, 7), the pure shear strain fields associated with screw dislocations,

temperature effects, and the nonhomogeneous nature of the solute distribution that is experimentally observed. The limitations regarding the effect of valence state and the experimentally observed nonhomogeneous distribution of solute were addressed by Labusch, who incorporated more realistic approximations of solute clusters and a modulus factor into the size-misfit factor to obtain the experimentally observed and often cited $c^{2/3}$ dependency:

$$\tau = \frac{A}{b}\left(f_m^4 c^2 w/E\right)^{1/3}$$

where E is the line energy of the dislocation, A is a function of the interaction of forces between the dislocation and obstacle (e.g., the solute cluster) located at a distance y from the dislocation, and w is the summation of the strengths of the individual solute atoms in the cluster. In addition,

$$f_m(= (G\mathbf{b}^2/120)\varepsilon_L)$$

is dependent upon: ε_L which is a misfit parameter given as the square root of a combination of squares of the contributions of the size and modulus factors. This inclusion of a modulus factor will account for the effect of modulus differences in solute content to a first approximation. Nabarro later revisited the problem and derived two forms, one whose functional form is identical to that obtained by Labusch, where A is ½. However, he also obtained a relation to describe isolated pinning points for lean solid solutions (e.g., less than the Friedel limit of ~100 to 200 ppm), given as:

$$\tau = \frac{1}{b}\left(f_m^2 c/2^{5/2} E\right)^{1/2}$$

where the variables have the previously described meanings. These models still neglect to account for any thermal effects or the interaction between screw dislocations and types of solute. The thermal effects neglected are related to the experimentally observed drop in yield strength with decreasing yield strength up to ~0.3 to 0.4 T_m (e.g., the homologous temperature), where the yield strength appears to plateau with further increases in temperature. This plateau has been attributed to thermally activated recovery processes in the material.

Finally, because these models consider the overall solid-solution strengthening effect in binary alloy systems, it is relevant to consider the extension of such models to materials containing a number of different solutes. Labusch first theorized (Ref 8) and Gypen and Deruyttere (Ref 9) experimentally verified that the total combination of the individual solid-solution strengthening effects, for solute elements A and B, may be taken as:

$$\Delta\tau_{SS}^{total} = \left(\Delta\tau_A^{2/3} + \Delta\tau_B^{2/3}\right)^{3/2}$$

Model Type 2: Slip Nucleation. Feltham (Ref 10) was the first to propose a model in which

Table 1 Summary of the expected interactions between the strain fields of dislocations and solute atoms in the given crystals

		Dislocation character		
Crystal structure(a)	Type of solute	Edge	Mixed	Screw
Close-packed (e.g., fcc, hcp)	Substitutional	√√	√	⋯
	Interstitial	√√	√	⋯
bcc	Substitutional	√√	√	⋯
	Interstitial	√√	√√	√√

(a) fcc, face-centered cubic; hcp, hexagonal close-packed; bcc, body-centered cubic. √√ indicates the possibility of strong interactions, whereas √ indicates the possibility of weak interactions.

the mechanism is based on slip nucleation from barriers. Although the first model assumed a fairly concentrated solid solution and was meant to account for certain aspects of fcc materials, it was soon discovered to describe accurately the break-away phenomenon in dilute alloys ($c \sim 0.01$ at.%) as well as solid-solution strengthening in hexagonal close-packed (hcp) and bcc materials (0.005 to 45 at.%). The nature of the model allows a first approximation of the temperature dependencies of solid-solution strengthening. The model is derived by considering the work required to move an edge dislocation from one linear segment of pinning solutes to the next linear segment of pinning solutes. It may be shown that the work is the sum of the energy required to unpin the dislocation from the solute atoms, the increase in the dislocation line length, and the work done by the applied shear stress. Notably, the critical resolved shear stress is a temperature-dependent term, thereby partly overcoming a limitation of model type 1. A further advantage of this approach is that by relating various equations for work, the shear stress function can be solved in closed form and then verified. The relevant equations for this approach are:

$$W = U_c^{1/2}\frac{L}{b} + n^2 Gb^3 \frac{b}{L} - \frac{1}{2}n\tau b^3 \frac{L}{b}$$
$$\frac{L}{b} = (4Gn/\tau)^{1/2}$$
$$W(\tau) = mkT$$

where, in addition to c, b, G, τ, k, and T, which have their meanings as given previously, U is the energy expended per solute atom to reach the saddle point between two linear segments of pinning particles, n is related to the distance between linear segments (typically 4 to 6 for fcc and hcp, 1 to 2 for bcc), L is the length of

the dislocation segment remaining pinned, and $m(= \ln(v_o/v))$, where v_o and v are the original and instantaneous dislocation velocities, and m is typically $25 + 2.3$. Further details of the temperature dependence of the critical resolved shear stress, the temperature dependency on the deformation activation volume, and the low-temperature anomalies are adequately described in the review by Butt and Feltham (Ref 10). It should be noted that in their treatment, the contribution of other pinning points, such as dislocation networks, is considered as part of a modified solute content, c^*. While they do not directly consider the temperature effects of c^*, they do consider the temperature effects in the work term that may include c^*.

Additional Resources. For a more detailed review, the reader is referred to the numerous review articles of solid-solution hardening, notably the most recent review of Butt and Feltham (Ref 10). A rather simplified $c^{1/2}$ dependency has also been derived and shown elsewhere (Ref 11–13). Research into stress relaxation (typically <10% of the load in aluminum-base alloys) has attempted to relate stress relaxation (and a corresponding increase in strength upon reload) to the rearrangement of dislocations and solute atoms during the unload (Ref 14, 15).

Age/Precipitation Hardening

In a number of alloys, it is possible to solution heat treat a material in a single-phase field, quench to room temperature, and then subsequently age at an intermediate temperature, leading to the precipitation of second-phase particles. Often, the precipitation sequence involves nonequilibrium zones and precipitates, such that the nucleation frequency is very high, leading to extremely refined distributions of particles. Also, solute content can influence

the transformation paths and types of precipitates formed. These refined distributions of particles can lead to very significant increases in strength. While the classical examples of alloys heat treated in this way are the widely used aluminum-copper alloys (Fig. 1a), Al-Mg-Si alloys, and nickel-base superalloys, heat treatments may also be used to optimize strengths in β-titanium alloys, some copper-base alloys (copper-titanium, copper-beryllium, copper-cobalt), some magnesium-base alloys (magnesium-zinc, magnesium-rare earths, Mg-Zn-Ca) (Fig. 1b), and iron-base alloys. An example of the variation in the property of a material that may occur as a result of different aging conditions (both time and temperature) is shown in Fig. 1(a, b). In this figure, the variation is associated with the time to and magnitude of the peak yield strengths and reflects the material response to heat treatment. Further, it has been shown that for many of these heat treatable alloys, the alloy content can significantly affect the peak properties as well (Fig. 2). This type of material response is typically called age (or precipitation) hardening.

In general, the ability to age harden a material is dependent on two criteria. The first is the requirement that a large concentration of solute be quenched into a supersaturated solid solution, so that the resulting volume fraction of finely dispersed precipitates is significant. This requirement is met for alloy systems where the equilibrium solvus exhibits a significant curvature, such that the maximum solid solubility will be relatively large. For example, consider the aluminum-copper binary system, of which the relevant portion appears in Fig. 3. In this figure, it can be observed that the equilibrium solute concentration at temperatures below 250 °C is less than 0.1 wt% (0.04 at.%) Cu. However, at 548 °C, the solid solubility is ~5.7 wt% (2.5 at.%) Cu. The second
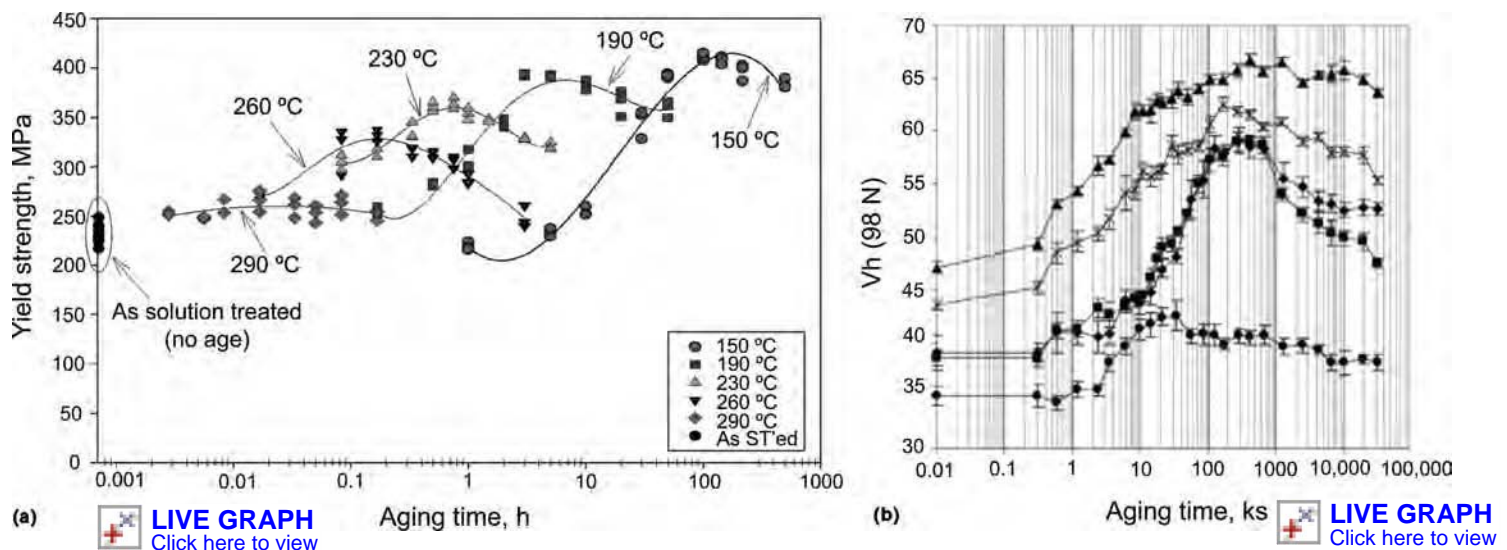
Fig. 1 Age-hardening response. (a) Aluminum alloys aged at various temperatures measured by tensile strength. Source: Ref 16. (b) Magnesium alloys as measured by aging time versus Vicker's hardness. Source: Ref 17, adapted from Ref 18

criterion involves the nature of the precipitation processes. Thus, to permit a very refined distribution of particles to form on aging, either (near) homogeneous nucleation or copious heterogeneous nucleation of particles on defects and interfaces is required. This latter criterion is often accomplished through the precipitation of nonequilibrium zones and phases where the activation energy for nucleation is reduced, largely through reductions in interfacial energies and elastic misfits.

Legacy Understanding. Age hardening initiates when a supersaturated solid solution is aged and subsequently decomposes through the formation of zones and second phases. The modeling of deformation in the presence of such phase transitions is normally based on a distribution of spherical particles, obviously an approximation because the zones and precipitates are often disk, ellipsoidal, or rod shaped as well as spherical. It is usually considered that this precipitation of zones and second phases rapidly reduces the matrix solute concentration from a supersaturation to an equilibrium value. Upon further aging, the distribution of particles coarsens by Ostwald ripening (particle coarsening), such that with a fixed volume fraction of second phase, the average particle size (with radius r) will increase, as will the interparticle spacing (L). At short aging times and at relatively lower temperatures, the particles are relatively coherent and characterized mainly by misfit strain fields, especially if they are GPB zones (i.e., compositionally different from the matrix but having the same fcc crystal structure). As intermediate phases form, the crystal structure of the particles is different from that of the matrix, and this will contribute to strengthening; that is, the difference in Peierls

stress (equivalently the critical resolved shear stress for dislocation motion) between the matrix and particles will contribute to strengthening. This is particularly the case if the particles are ordered intermetallic compounds (such as γ'-Ni$_3$Al in nickel, as in γ-nickel-base superalloys), where resistance to dislocation motion arises not only from differences in Peierls stresses but also from formation of antiphase boundaries.

The typical age-hardening response, as shown in Fig. 1, exhibits an initial increase in yield strength (or hardness, as was measured in the early literature) to a peak strength, followed by a decrease in the value of this property. There are two competing deformation mechanisms that can account for these variations (Fig. 4): the first operative as the value of the yield strength increases where the particles are cut, or sheared, by dislocations, and the second, which is responsible for the decreasing part of the aging response where dislocations can bypass the particles (the Orowan mechanism). These mechanisms are considered in turn.

When precipitation occurs in an alloy where age hardening is effective, the zones or particles (the precipitation sequence depending on the alloy system) are extremely finely dispersed and are mainly characterized by their coherency strain fields. The very small interparticle spacing is such that a given dislocation will lie both in regions between the zones or particles as well as through them because of the line tension (i.e., the energy/unit length) associated with the dislocation. Therefore, further motion of the defect requires some segments to glide completely through the obstacles (i.e., mainly the strain fields) and for others to glide from

regions of high strain to those of lower-strain fields between the zones or particles. Consequently, the strengthening effect is not particularly significant. As the particles coarsen, resulting in increased interparticle spacings (a fixed volume fraction of second phase), more segments of the dislocations lie between the particles, such that an increased fraction of the length of the dislocations must do work against the strain fields of the obstacles. Also, with aging time, the particles become more effective obstacles, because they adopt crystal structures different from the matrix. Hence, the strengthening effect becomes more significant. Further aging results in increased particle sizes and spacings, such that essentially all segments of dislocations can lie in regions between the particles, and motion of dislocations requires that all segments of the defects must do work cutting through the obstacles; this would correspond to the maximum strengthening effect.

As the interparticle spacing increases with continued aging, it becomes possible for the dislocation line to bow between the particles. This is depicted in Fig. 5(a–d), where a dislocation line with the line direction shown is assumed to glide to the right on a slip plane parallel to the plane of the page. Provided the particles are sufficiently widely spaced, the dislocations will adopt the configuration shown in Fig. 5(b, c). Note that at the points marked "A" (Fig. 5c), the dislocation segments have opposite signs and so pinch off to form a loop around each particle and a dislocation line that has effectively bypassed the particles. This is the Orowan mechanism. As shown subsequently, the shear stress required to operate the mechanism is proportional to $G\mathbf{b}/L$ (G being the shear modulus, $\mathbf{b}$ the magnitude of the Burgers vector, and L the interparticle spacing), and so with continued aging and coarsening of the particle dispersion, bypassing of obstacles becomes easier, and hence, the strengthening effect will decrease, as is observed.
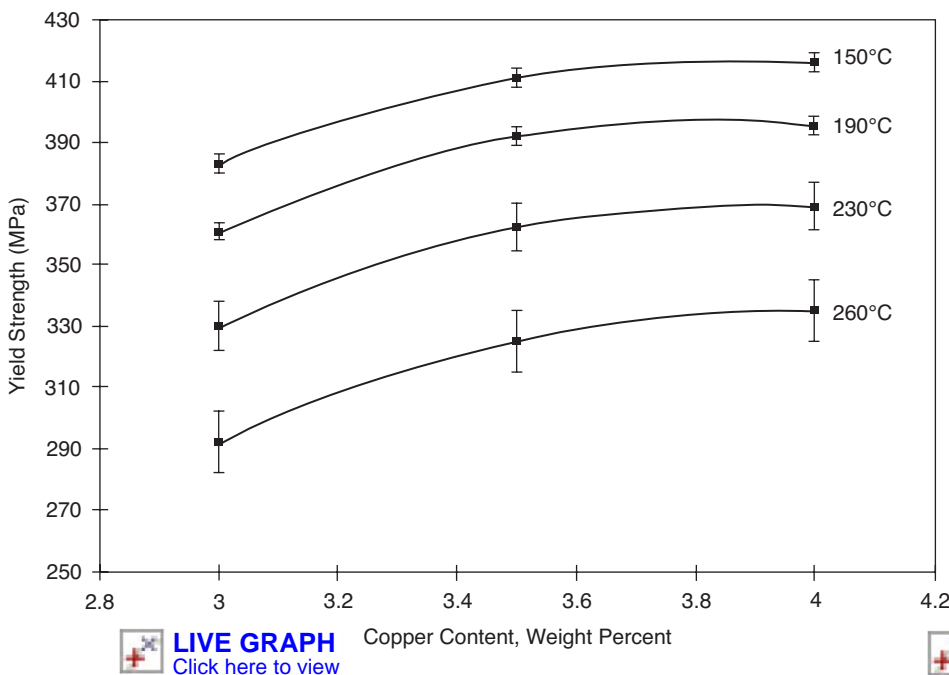
Fig. 2 Effect of solute content on peak yield strength. Source: Ref 16
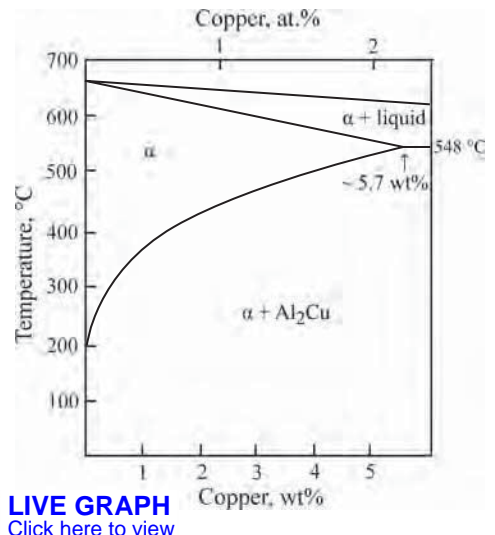
Fig. 3 Portion of the aluminum-copper system
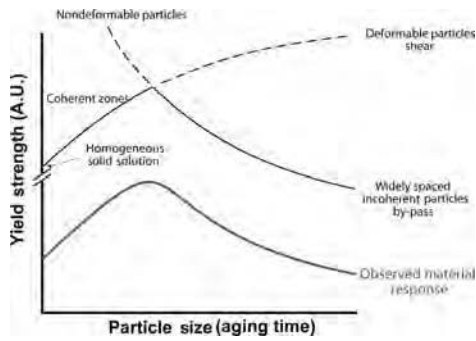
Fig. 4 Effect of particle size on yield strength with operating mechanisms involving dislocation particle interactions. Time and yield strength in arbitrary units (A.U.)

Fig. 5 Schematic of the Orowan mechanism. The dislocation line moves left to right through steps a, b, c, and d. **b** is the Burgers vector. Other symbols are self-evident.

Fig. 6 Schematic of particle cutting, showing multiple dislocations (A, B) operating from source S_1.

In addition to influencing the yield strength, these two mechanisms, dislocation cutting and particle bypass (Orowan mechanism), can also influence the ultimate tensile strength and ductility. In the case of dislocation cutting, each time a particle is cut, the area of the particle subtended on the slip plane decreases, as shown in Fig. 6. Therefore, that particle effectively becomes a weaker obstacle, such that it will require less work to be done by the next dislocation emitted by the given source to cut the particle; of course, this new dislocation will also reduce the effective size of the particle, again reducing its effectiveness as an obstacle. Such successive particle cutting results in slip localization, where, in principle, it becomes increasingly easier for the dislocation source to operate, implying possible work softening. Of course, there are other deformation processes at work that complicate the situation, such that rarely is work softening observed, but with dislocation cutting as the predominant strengthening mechanism, there is limited scope to realize significant work hardening. Such strain localization will lead to significant pileups of dislocations at grain boundaries, producing very significant stress concentrations, which may have a marked influence on the initiation of intergranular fracture, particularly if grain-boundary phases exist at the boundaries, which is often the case in commercial alloys. In contrast, in the case of dislocation bypass, debris in the form of dislocation loops is left around each particle that is bypassed.

The next dislocation to attempt to bypass the given particles will experience a repulsive force, because its line direction is parallel to that of the segments of the loops with which it first comes into proximity, and so the passage of the second dislocation is more difficult than the first, requiring an increase in the applied stress to continue dislocation motion; that is, the sample would work harden. If the dislocation is screw in character, then it could possibly undergo a double cross-slip mechanism of bypass, depicted in Fig. 7, but this would also require an increase in applied stress, because the cross-slip plane is most likely to be less favorably oriented than the primary system. The third dislocation would experience an even greater repulsive interaction. These increases in applied stress would lead to the operation of other dislocation sources, such that the slip would be spread throughout the grain, rather than become localized (as with cutting).

As can be seen, there are some very important differences between these two deformation mechanisms. While in heat treatable aluminum alloys, for example, peak hardness usually is achieved in the dislocation-cutting regime, there are problems associated with slip localization and properties such as ductility and stress-corrosion cracking. Interestingly, the commercial aluminum-lithium alloys were made viable by introducing a very refined dispersion of the phase δ'-Al_3Li for strengthening. To reduce the extent of the resulting slip localization, a second series of nondeformable particles was also introduced to disperse slip.

Modeling Precipitation Strengthening. This section discusses legacy and current understanding of dislocation cutting and dislocation bypass.

Dislocation Cutting. As discussed previously, dislocation cutting occurs when the distribution of particles is extremely refined, most usually consisting of either zones or second phases with coherent interfaces with attendant misfit strains, semicoherent interfaces with misfit dislocations, and particles with crystal structures that are somewhat different from that of the matrix, leading to significant differences in Peierls stresses and choice of slip planes. Clearly, it is a somewhat complex situation, with contributions to hardening changing as a function of aging time as the distribution of particles develops and changes structurally. A prediction on the basis of small coherent particles yields a flow stress variation given by (Ref 19):

$$\tau = 4.1 G \varepsilon^{3/2} f^{1/2} \left(\frac{r}{\mathbf{b}} \right)^{1/2}$$

where G is the shear modulus, ε is the coherency misfit strain, f is the volume fraction, r is the particle radius, and $\mathbf{b}$ is the magnitude of the Burgers vector. As can be seen, the yield stress will vary as a function of $r^{1/2}$, which is a reasonable approximation to the expected behavior. A different formulation (Ref 20) incorporates the various contributions to the strengthening, discussed previously, and yields an equation to estimate the increase in shear stress arising from the presence of the distribution of particles and is written as:

$$\Delta\tau_c \cong \frac{\gamma^{3/2}}{b} f^{1/2} \frac{r^{1/2}}{(6E_v)^{1/2}}$$

where γ is an effective interfacial energy, E_v is the strain energy of a dislocation (taken as $\frac{1}{2}G\mathbf{b}^2$), and the other symbols have their usual meaning. Again, the strengthening effect is predicted to vary as a function of $r^{1/2}$. Both of these expressions are useful in that they predict the trend in strength variation as a function of aging time. However, they do not return accurate predictions for strengthening. For example, consider the age-hardening curve shown previously in Fig. 1. As can be seen, the strength increase does not vary uniformly, that is, simply as a function of $r^{1/2}$, because the nature of the zones and particles is changing with aging time. Each form of the zones and intermediate phases gives rise to a particular strengthening effect.
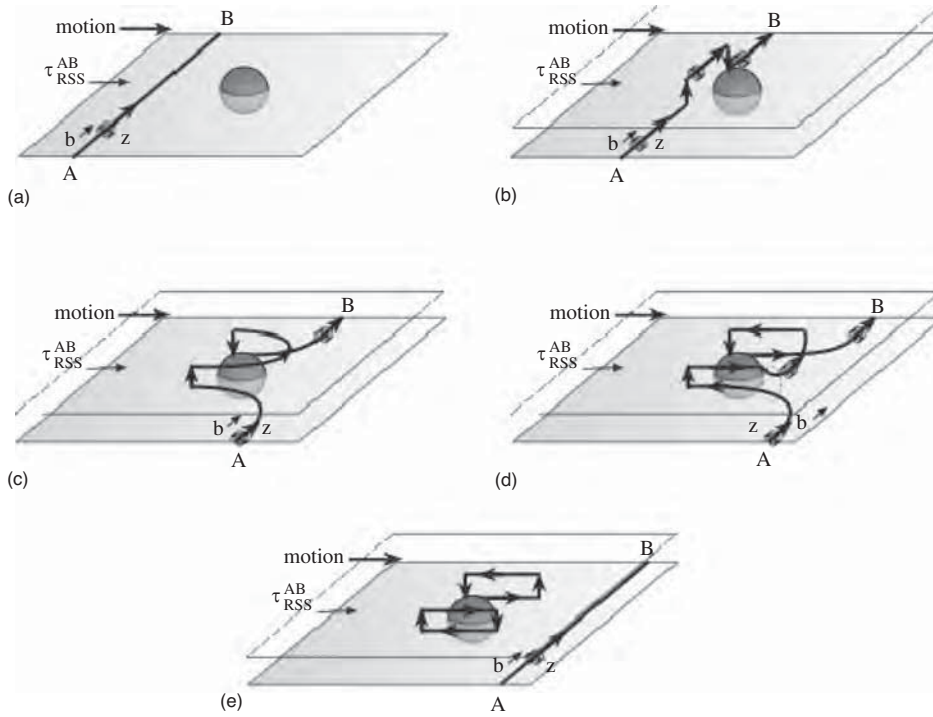
Fig. 7 Schematic of the double cross-slip mechanism around a particle

The Orowan Mechanism (Dislocation Bypass). As noted previously, if the interparticle particle spacings are sufficiently large, dislocations bypass the particles by the Orowan mechanism. The effective increase in the flow stress for this particular case is given by:

$$\tau_{\text{Orowan}} \approx \frac{\alpha G\mathbf{b}}{L}(\cos\phi_c)^{3/2}$$

where α is a factor, L is the particle spacing, ϕ_c is shown in Fig. 5(b), and G and $\mathbf{b}$ have their usual meanings. When the particles are sufficiently far apart, ϕ_c approaches 0, and the $\cos\phi_c$ term can be ignored. Further, α may be taken to be 0.84 for dislocations in a random array of strong spherical obstacles, where cos $(\phi_c/2) > 0.70$ (i.e., most cases). It should be noted that this approximates the local shear stress when the dislocation is interacting with the particles and is not applicable when the dislocation is not interacting with the particle (e.g., when it is moving to the next array of barriers).

It may also be shown that the spacing L and Burgers vector $\mathbf{b}$ may be related to the size of the particles and the concentration of solute and combined in the following form:

$$\tau_{\text{Orowan}} \approx \frac{0.84G\mathbf{b}}{L} = \frac{0.84Gf^{1/2}}{r}$$

Emerging Concepts. The simplified models for bypass often assume a random array of precipitates and tend to exclude the precipitate morphology and/or its orientation relationship with the matrix. Therefore, recent research has

focused on deriving more accurate models of shear stress using a variety of approaches (Ref 21–23), including computational dislocation-precipitate simulations to study various precipitate morphologies and distributions. For very thin platelike precipitates on the {100} planes, a few different analytical expressions for the shear strength have been proposed, including (Ref 22):

$$\tau_{\theta'} = 0.13\left(\frac{G\mathbf{b}}{\sqrt{dw}}\right)\left(\sqrt{f} + 0.75\sqrt{\frac{d}{w}}f + 0.14\frac{d}{w}f^{3/2}\right)$$
$$\times \left(\ln\frac{0.87\sqrt{dw}}{r_0}\right)$$

where d is the diameter of the precipitate, w is the thickness, f is the volume fraction, and r_0 is the inner cutoff radius, which is set to $\mathbf{b}$, or (Ref 21):

$$\sigma_{\text{ppt}} = M\left(\left(\frac{G\mathbf{b}}{2\pi\sqrt{1-\nu}}\right)\left(0.931\sqrt{\frac{0.306\pi dw}{f}}\right.\right.$$
$$\left.\left. -\frac{\pi d}{8} - 1.061w\right)\left(\ln\frac{1.225w}{r_0}\right)\right)$$

For computational efforts, the terms for the size and fraction of the precipitates may be modified to account for time and temperature effects as the nature and type of particles evolve. Based on such modifications, Weakley-Bollin et al. analyzed these models (Ref 2) and found that for Al-Si-Cu alloys, each predict peak hardness to occur at a time much longer than experimentally observed. However, both models are more sensitive to the thickness (w) of the precipitate than to the diameter. Unfortunately, the true

distribution of precipitate thicknesses is extremely difficult to quantify, because the thickness of the strengthening precipitates in these systems is often less than 5 nm. Other researchers have used more complex dislocation-particle models to simulate complex particle distributions, where the variation in both size and spacing may be studied (Ref 24), which may help overcome may of the assumptions required to derive the aforementioned models. Lastly, stress relaxation is an important element when attempting to understand the ductility of precipitation-hardened systems as a function of temperature (Ref 25).

It is beneficial to engineer the shape of the precipitate to maximize their interaction with dislocations. For example, in the magnesium-tin system, where more than one morphology is possible, the strengthening afforded by the rods is far greater than for the thin plates, because the rods cross many ($\sim10^2$ to 10^3) slip planes rather than only a few (~10 to 50 for the thin plates). Fundamental experimental and modeling work is being performed on these types of magnesium-base alloys (Ref 17) to study how changes in alloy content affect lattice parameters and therefore precipitation morphologies and sequences.

It may also be advantageous to form ordered precipitates, because this increases the effectiveness of the strengthening afforded by a dispersion of particles. This increases not only the magnitude of the Burgers vector but also results in an increase in strengthening due to the influence of antiphase-boundary (APB) energy, γ_{APB}. Thus, shearing of the ordered particles is affected by glide of superdislocations, and, in general, these defects are dissociated into superpartials, bounding a ribbon of APB. This is illustrated by reference to the model system, namely γ/γ' nickel-base superalloys, where the ordered phase γ' (Ni$_3$Al with the L12 crystal structure) is dispersed in the γ-matrix (fcc nickel solid solution). The orientation relationship between the γ and γ' phases is described by $\{001\}_\gamma//\{001\}_{\gamma'}$, $<100>_\gamma//<100>_{\gamma'}$, and the difference between the lattice parameters of the two phases is very small; that is, there exists a small but significant elastic misfit, the magnitudes of which depend on the chemical composition of the given superalloy. In the matrix, dislocations with $\mathbf{b} = \frac{1}{2} <110>$ glide on $\{111\}$, and in the ordered phase, dislocations with $\mathbf{b} = <110>$ glide on $\{111\}$ planes parallel to those in the matrix. In the ordered particles, the perfect dislocations dissociate into superpartial dislocations, such that, for example:

$$\mathbf{b} = [101] \rightarrow \frac{1}{2}[101] + \text{APB} + \frac{1}{2}[101]$$

Each of the superpartial dislocations (with $\mathbf{b} = \frac{1}{2} <110>$) can further dissociate into pairs of Shockley-like partials, each with $\mathbf{b} = \frac{1}{6} <112>$, bounding a complex stacking fault. This latter dissociation is not shown here, because it is the influence of the APB on strength that is of interest. The shear of the

particles may be described by a strengthening term, taking the form:

$$\tau = 0.5\alpha\left(\frac{\gamma_{APB}}{\mathbf{b}}\right)^{1.5}\left(\frac{\mathbf{b}df}{\Gamma}\right)^{1.5} - 0.5\frac{\gamma_{APB}}{\mathbf{b}}f$$

where α is a numerical factor that is dependent on the shape (for example, 0.72 for spherical particles, Ref 26, and $1/\pi$, Ref 19, without a reported shape).

In addition to the contribution to strengthening from the presence of the APB in these ordered particles, there is also a contribution that arises from the source of the superdislocations. Rather than nucleate new dislocations within the particles, the gliding matrix dislocations transmit through the matrix/particle interfaces. Of course, it requires that two matrix dislocations with $\mathbf{b} = \frac{1}{2}<110>$ are transmitted to form one superdislocation, with $\mathbf{b} = <110>$. These two matrix dislocations have the same sign, with parallel Burgers vectors, and so they will repel one another. This force of repulsion must be overcome to enable slip transmission to occur. This provides a contribution to strengthening. When the leading and trailing dislocations are within the same particle, the strengthening may be given as:

$$\tau_c = \frac{0.84\Gamma f^{0.5}w}{d\mathbf{b}}\left(1.28\frac{d\gamma_{APB}}{w\Gamma} - 1\right)^{0.5}$$

where w accounts for the elastic repulsion between the leading and trailing dislocations and is of the order unity (Ref 27). The limitations of the two preceding equations is that they are valid only for low volume fractions of the precipitates (i.e., less than 0.1).

Dispersion Strengthening

Dispersion strengthening is an active mechanism in alloys intended for use at intermediate and elevated temperatures. It is, in many ways, very similar to age hardening, described previously, except that the particles are chosen because of their thermal stability, that is, their resistance to Ostwald ripening (particle coarsening). From coarsening theory, the stability of particles, when exposed to elevated temperatures, depends largely on the product of three terms, namely the interfacial energy, the diffusivity of the slowest elemental species, and the solubility of solute in the matrix. For the age-hardening systems, for example, the heat treatable aluminum alloys (i.e., taken from the 2*xxx*, 6*xxx*, 7*xxx*, and 8*xxx* aluminum alloy series), the dispersions are not particularly stable thermally because of the relatively large solubilities of the solute elements (an advantage in terms of increased volume fraction of precipitates) and their reasonable rates of diffusion. Dispersion-strengthened systems are usually those that contain a dispersion of particles where the solubilities of the solute are extremely low and the diffusion rates are also very slow, for example, rare earth oxides (producing oxide-dispersion-strengthened or ODS, systems). In these ODS alloys, the solubilities of both the rare earths and oxygen tend to be very low in typical matrices (e.g., nickel and aluminum), and the diffusivities of the rare earth elements are also low. Figure 8 shows the retention of a beneficial strengthening effect at higher temperatures offered by an ODS aluminum alloy with respect to both pure aluminum and a precipitation-hardened Al-Mg-Si system, 6061. In this figure, although the precipitation-hardened system exhibits a similar hardness at low temperatures, when the particles fully dissolve ($\sim$225 °C), the primary strengthening mechanism is lost, resulting in a significant reduction in hardness. However, the ODS alloy remains stronger at elevated temperatures (Ref 28).

Modeling of Dispersion Strengthening. With such low solubilities, the heat treat method for producing age-hardening systems is not appropriate, and methods for such a mechanical alloying must be used to effect reasonable volume fractions of these thermally stable particles. The resulting particles are essentially nondeformable, and the interparticle spacings are such that the Orowan mechanism operates, and the strengthening is given by:

$$\tau_{loc}(\text{strong}) = \tau_{Orowan} \approx \frac{0.84G\mathbf{b}}{\lambda} = \frac{0.84Gf^{1/2}}{N^{1/3}}$$
$$= \frac{0.84Gf^{1/2}}{r}$$

To maximize the increase in strength due to dispersoids, it is desirable to minimize the radius of the dispersoid and maximize the volume fraction. However, in addition to this strengthening mechanism, the debris from the formation of dislocation loops about the particles from the Orowan mechanism act to work harden the material, leading to an often-observed high work-hardening rate for dispersion-strengthened materials. In addition, researchers have explored the possibility of the generation of geometrically necessary dislocations (e.g., prismatic loops that intersect the glide plane) at the matrix-precipitate boundaries due to the presence of hard particles in the plastically deforming matrix (Ref 28, 29). These dislocations will also contribute to the work-hardening component of the strength of the material. As the temperatures increases, it is important to consider the effects of thermally activated dislocation motion, such as cross slip around the dispersoids, and climb. In addition, dynamic recrystallization and recovery may reduce the work-hardening effects at elevated temperatures.
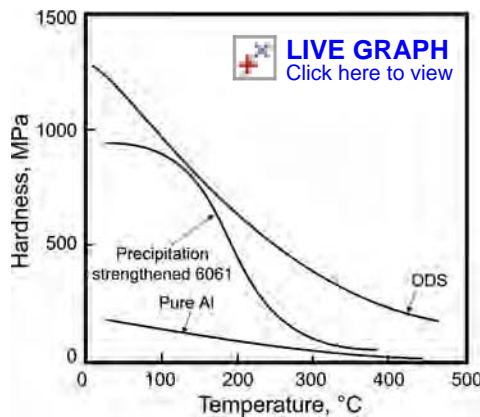
Grain-Size Strengthening/Hardening

The independent observations and ground-breaking work of E.O. Hall and N.J. Petch (Ref 30–32) in the early 1950s led to the development of the Hall-Petch relationship, which relates the effect of grain size with the corresponding strength of a material. The experimentally observed relationship is given as:

$$\sigma_y = \sigma_i + kd^{-1/2}$$

where σ_i is an intrinsic grain lattice friction stress, k is taken as a material constant, and d is a measure of the grain size. Thus, $kd^{-1/2}$ may be considered as the increase in yield stress ($\Delta\sigma_y$) resulting from a grain-size effect (i.e., $\Delta\sigma_{HP}$). Figure 9 schematically illustrates the often-observed grain-size dependence for the three common metallic crystal structures. In this case, the intrinsic strength is that of the matrix and, in principle, includes the influence of other strengthening mechanisms.

Legacy Understanding. Consider a polycrystalline material subjected to a tensile load. As the applied stress is increased, slip is assumed to initiate in those grains that are most favorably oriented for dislocation glide, that is, where the Schmid factor (= $\cos\phi\cos\lambda$, where λ is the angle between the slip direction and the

Fig. 8. Effect of dispersions on the hardness of an oxide-dispersion-strengthened (ODS) material as a function of temperature as compared to pure aluminum and 6061 aluminum alloy. Source: Adapted from Ref 28
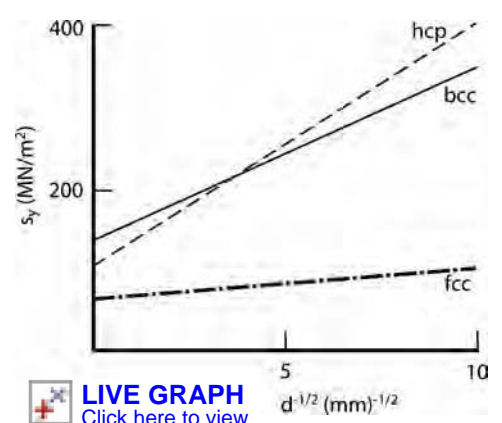
Fig. 9 Effect of grain size on the yield strength for different crystal structures. hcp, hexagonal close-packed; bcc, body-centered cubic; fcc, face-centered cubic. Source: Adapted from Ref 19, Fig. 7.32

tensile axis, and ϕ is the angle between the slip plane normal and the tensile axis) is maximized. When the critical resolved shear stress is exceeded in one of these favorably oriented grains, a dislocation source (marked S_1 in Fig. 10) operates, emitting dislocation loops. In the absence of obstacles, the leading dislocation loop will expand across the slip plane until encountering the grain boundary, for example, at "A" in the figure. Because of the discontinuity represented by the grain boundary, the dislocation will initially remain at the boundary. The next dislocation to be emitted by the source will also glide to the region adjacent to the boundary but will experience a repulsive force from the first dislocation, because these defects have like signs. This repulsive force promotes a backstress that subtracts from the resolved shear stress (τ_{rss}). Hence, the applied stress must be increased for the original dislocation source to continue operation. After several dislocations have been emitted, these defects will continue to form a dislocation pileup (Fig. 10) that exhibits a stress concentration at its head, the magnitude of which is proportional to the number of dislocations contained in the pileup. In the original understanding of this mechanism, this stress concentration may help activate another dislocation source in the neighboring, less-well-oriented grain, marked S_2 in Fig. 10. The grain-size effect arises from the fact that when the grain is large, a significant number of dislocations may join the pileup before the backstress effectively stops the operation of the source. In this case, the stress concentration can be very significant, because the number of dislocations in the pileup is large. Only modest increases in the applied stress would be necessary to activate source S_2 in the neighboring grain. In contrast, when the grain size is small, then, because of the proximity of the dislocation source to the grain boundary, the number of dislocations in the pileup will be limited, because the backstress will be more effective in inhibiting the operation of the source. In this case, the stress concentration will be relatively small. Consequently, the
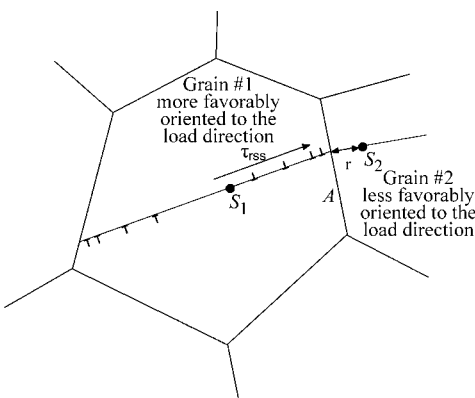
applied shear stress must be increased significantly to initiate operation of the source at S_2; that is, the sample has been strengthened.

More recent considerations have examined the possibility of transmission of slip across grain boundaries as an alternative event to operating a dislocation source in the neighboring grain. Because it is usually most unlikely that slip planes in the neighboring grain will be parallel to that in the original grain, transmission of dislocations often requires that a residual defect with an irrational Burgers vector be left in the boundary, and these sessile defects may cause problems to further transmission. Because of the multiplicity of slip systems in most metallic alloys, operation of new sources, rather than slip transmission, is expected to be the predominant event.

Modeling grain-boundary strengthening represents a difficult task. For example, accurate descriptions of the stress field ahead of a crack tip in a neighboring grain are highly complex (Ref 33). In consequence, fairly simple approximations of this stress field are used to assess the influence of the pileup in activating a potential dislocation source in a neighboring, less favorably oriented grain; here, the stress concentration ahead of the pileup is approximated, varying simply as $(d/r)^{1/2}$ (where $2d$ is the grain diameter, and r is the location of the potential source in the neighboring grain). In the absence of a grain boundary, the shear stress acting on the active slip plane would be τ_i, given by $\tau_i \approx \tau_{crss}/(\cos\phi\cos\lambda)$, where τ_{crss} is the critical resolved shear stress for activation of dislocation motion. In the presence of the grain boundary acting as an obstacle, the shear stress is increased (i.e., the strengthening effect) to τ, so that the effective resistance of the boundary is given by $(\tau - \tau_i)$. It is this stress that is enhanced by the stress concentration at the head of the dislocation pileup. For a potential source in the neighboring grain at a distance r from the pileup (Fig. 10), the stress, τ_r, would be given approximately by:

$$\tau_r = (\tau - \tau_i)\left[\frac{d}{r}\right]^{1/2}$$

The shear stress τ_r operates in a plane parallel to the pileup; to operate a source in the neighboring grain at r, its value must be τ_{max}. Substituting this value into the previous equation and rearranging yields:

$$\tau = \tau_i + \tau_{max} \cdot r^{1/2} \cdot d^{-1/2}$$

However, the shear stress must act on the slip plane of the new source in the neighboring grain, so an orientation factor, m', is introduced, such that $\tau_{max} = m' \cdot \tau_{crit}$, where τ_{crit} is the shear stress required to operate the new source. It follows that:

$$\tau = \tau_i + m' \cdot \tau_{crit} \cdot r^{1/2} \cdot d^{-1/2}$$

To write this in terms of the applied stress and hence assess the contribution of this mechanism

to increasing the yield stress, a second orientation factor is introduced, that is, relating applied stress to resolved shear stress, namely $\sigma = m \cdot \tau$. Also, a further simplifying assumption is made by equating the two orientation factors, such that $m = m'$. Hence, the yield stress is expected to vary with grain size in the following way:

$$\sigma_y = \sigma_i + m^2 \cdot \tau_{crit} \cdot r^{1/2} \cdot d^{-1/2}$$

Let $k_y = m^2 \cdot \tau_{crit} \cdot r^{1/2}$, so that:

$$\sigma_y = \sigma_i + k_y \cdot d^{-1/2}$$

This is the well-known Hall-Petch equation, where k_y is known as the Hall-Petch parameter. This relationship has been validated experimentally for many metallic materials, and so, despite the simplifying assumptions, the form appears to be correct. The degree to which grain-size reduction increases the yield stress depends on the magnitude of the Hall-Petch parameter.

There are important materials parameters that influence the value of k_y, one being the probability of a slip system being reasonably oriented in the adjacent grain (the orientation factor), and the other being the value of τ_{crit}, that is, the shear stress necessary to activate glide in the material, the effective critical resolved shear stress. Based on these, an assessment of the effectiveness of grain size as a strengthening mechanism can be made. Thus, if the material has a limited number of slip systems available and an intrinsically high value of the critical resolved shear stress or strong locking of dislocations (e.g., by solute), then a reasonably large value of k_y would be expected, and grain size should contribute significantly to strengthening. Alloys based on fcc metals have a large number of slip systems available (i.e., ½<110>{111}) and generally fairly low values of τ_{crit}, and hence, strengthening from a reduction in grain size is not expected to be significant. For bcc materials, although the number of available slip systems is large (1/2<111> on {110}, {112}, and {123}), they often exhibit large values of the critical resolved shear stress, and also, there tends to be very strong locking of dislocations. Hence, grain-size reduction in bcc alloys is often an effective strengthening mechanism. For hcp materials, the number of available slip systems tends to be low, while values of the critical resolved shear stress vary considerably (e.g., low for magnesium and high for α-titanium). Hence, the effectiveness of this strengthening mechanism depends on the specific hcp metal upon which the given alloy is based and the degree to which other strengthening mechanisms have been exploited in alloy design. It is possible to derive a ranking of the constant k based on these two factors, which are similar to what is shown in Table 1. These results explain the trends observed in Fig. 9. Experimentally, it has been shown that the parameter k is approximately 1 order of

Fig. 10 Schematic showing the important variables that contribute to the observed Hall-Petch effect. See details in the text.

magnitude greater for bcc alloys when compared with fcc alloys (bcc: 0.5 to 1.6 MN/m$^{-3/2}$; fcc: 0.06 to 0.15 MN/m$^{-3/2}$) (Ref 32–40). Values for hcp magnesium have been reported in the range of 0.11 to 0.54 MN/m$^{-3/2}$ (Ref 41–52) and in the same range ($\sim$0.61 MN/m$^{-3/2}$) for hcp titanium (Ref 53).

In addition to the higher-angle grain boundaries, efforts have been made to model similar strengthening effects observed in subgrain structures. These models attempt to directly incorporate the dislocation densities of the subgrain boundaries and consider them to exhibit different boundary strengths (e.g., $\Delta\tau_{\text{Subgrain boundary}} = \tau - \tau_i$) than those afforded by high-angle grain boundaries. Thus, it is possible to show that the grain-size-dependent term is equivalent to a dislocation-density-dependent term, as shown on the right side in:

$$kd^{-1/2} = (m^2\tau_{\text{crit}}r^{1/2})d^{-1/2} \equiv \alpha mG\mathbf{b}\rho_T^{1/2}$$

This dislocation density term must be separated into the accumulation of dislocations associated with subgrain boundaries, which have been termed statistically stored dislocations (SSDs), and those associated with grain boundaries, which have been termed as geometrically necessary dislocations (GNDs). Thus, the equation becomes:

$$\alpha mG\mathbf{b}\rho_T^{1/2} = \alpha mG\mathbf{b}(\rho_{\text{SSD}} + \rho_{\text{GND}})^{1/2}$$

Thus, the total effect becomes:

$$\sigma_y = \sigma_i + k_{\text{tot}}d^{-1/2} = \sigma_i + \alpha mG\mathbf{b}(\rho_{\text{SSD}} + \rho_{\text{GND}})^{1/2}$$

It is then possible to approximate the ρ_{SSD} either as a mixed tilt/twist boundary with $\rho_{\text{tilt/twist}} = \rho_{\text{SSD}} = 1.5 \, S_v/\mathbf{b}$, where S_v is the area of boundaries per unit volume, is the misorientation angle, and $\mathbf{b}$ is the Burgers vector; or as Ashby's approximation that ρ_{SSD} is linearly dependent on strain, $\rho_{\text{SSD}} \approx \varepsilon/(\mathbf{b}\lambda_s)$, where λ_s is the free slip distance. Similarly ρ_{GND} may either be neglected and assumed to be adequately incorporated in the original grain-size factor, $kd^{-1/2}$, or approximated by $\rho_{\text{GND}} \approx \varepsilon/(4\mathbf{b}d)$, where d is the grain size. Thus, two equivalent types of equations emerge:

$$\sigma_y = \sigma_i + \alpha mG\sqrt{1.5\mathbf{b}(S_v\theta)} + kd^{-1/2}$$
$$= \sigma_i + \alpha mG\mathbf{b}\left(\frac{\varepsilon}{\mathbf{b}\,\lambda_s} + \frac{\varepsilon}{4\mathbf{b}d}\right)^{1/2}$$

These models have been used to adequately describe the behavior of aluminum-magnesium alloys (Ref 54) and Copper, Nickel, and aluminum (Ref 55). While these latter variations of the Hall-Petch relationship do include a polygonized substructure from previous cold work as well as a strain term, the strain hardening of the grains is not considered in the Hall-Petch relationship, which is typically reported relative to the yield strength at 0.2% strain. While some have reported the relationship to be valid at larger strains as well, the Hall-Petch parameters are a function of work-hardening rates (Ref 56). It has also been shown that the intrinsic strength will change as a function of strain and cold work history (i.e., $\sigma_{i,\text{tot}} = \sigma_i + \Delta\sigma_i(\varepsilon)$), because the multiple slip systems and dislocations within a grain will interact and create pinning points.

This strengthening mechanism is not necessarily limited to the effects of only grain size but may be extended to other second phases. For example, some of the early work on steels correlated the martensite packet size, rather than the grain size, against the yield strength and arrived at similar observations (Fig. 11) (Ref 57–60). For the original data shown in this figure, it was postulated that the difference in slope for the various compositions was attributed to the carbon buildup at the martensite packet size boundaries. However, as has been shown, the k is also dependent on the critical resolved shear stress of the crystal, and the carbon content would change the critical resolved shear stress for the two types of martensite. Data taken from recent work on the effect of α-colony sizes in $\alpha+\beta$ titanium alloys show a deviation from the $d^{-1/2}$ relationship (Ref 61), suggesting that, in this case, it is not a simple Hall-Petch relationship. More importantly, an analysis of the titanium data shows a minimal k factor ($\sim$0.1 MN/m$^{-3/2}$). Thus, it is important to consider whether other interfacial strengthening mechanisms are present in order to predict and differentiate whether these other domain sizes (e.g., martensite packets and colonies of lamellae) will exhibit a Hall-Petch relationship. Such work has been conducted in the study of $\gamma + \alpha2$ polysynthetically twinned crystals, where the colony-like structure exhibits a Hall-Petch factor of 5 MN/m$^{-3/2}$ and an unrealistic negative value of σ_0, whereas the Hall-Petch parameter for equiaxed α is $\sim$0.9 to 1.4 MPa/m$^{-1/2}$ (equivalent with MN/m$^{-3/2}$). Therefore, researchers have drawn conclusions that the material response of this type of colony microstructure may not be adequately described by a Hall-Petch phenomenon, and another mechanism (or competing mechanisms) likely dominates (see the section "Strengthening from Interfacial Barriers to Slip Transmission in this article) (Ref 62–65). In addition to considering other domain types, research into stress-activated strain mechanisms at boundaries other than grain boundaries (e.g., twin boundaries) has been conducted. In this research, twin boundaries act as the obstacles, and either dislocation or twinning is then activated in the neighboring domain, which does result in an apparent Hall-Petch effect (Ref 66) with a larger apparent k-factor ($>$30 MN/m$^{-3/2}$), which can be attributed to the critical stress component rather than the m-factor.

Strengthening from Cold Work

It has long been recognized that cold work can impart a significant increase in the strength of a material. The implication is that the structural changes accompanying cold work impede dislocation motion in a given sample on the subsequent application of stress in, for example, a tension test, such that an increase in applied stress is required to cause the sample under consideration to yield. The most common change accompanying cold work is a significant increase in dislocation density, often referred to as debris, and much of the strengthening effect derived from cold work is due to the interaction of mobile dislocations with that increased dislocation density. In fact, essentially the same processes that lead to work hardening during tensile testing are responsible for the increased strength in samples of cold-worked alloys.
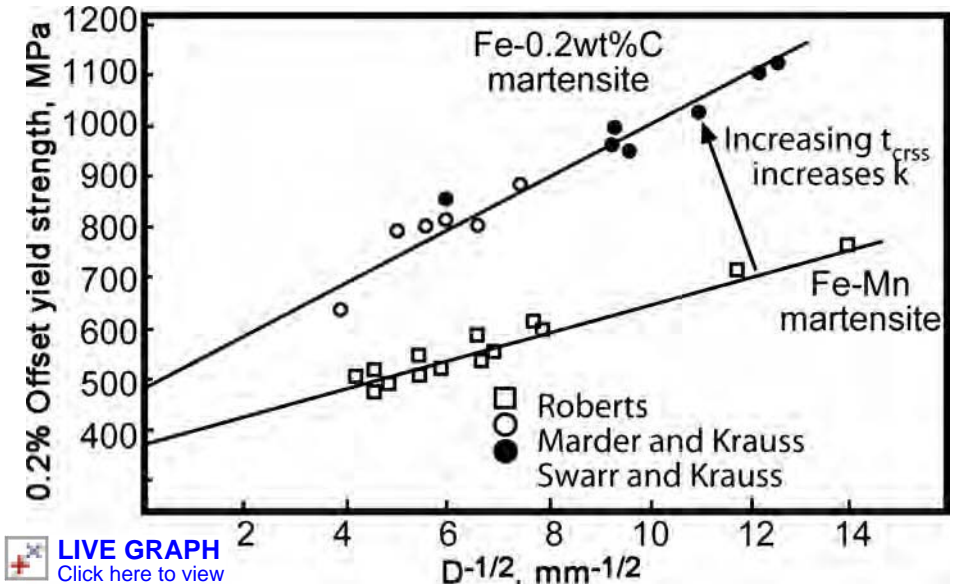
Fig. 11 Effect of martensite packet size on the yield strength. Source: Ref 58–60

Legacy Understanding. As stated previously, it is the interaction of mobile dislocations with increased dislocation density following cold work of a material that results in strengthening. There are a number of interactions between dislocations that can lead to impediment of further dislocation glide. To illustrate these processes, two simple interactions are described.

The first of these involves the intersection of two dislocations gliding on two inclined slip planes, as depicted in Fig. 12. In Fig. 12(a), the edge dislocation *AB* intersects the screw dislocation *CD*. As a result, a jog is formed where the jogged segment *PS* lies parallel to the direction of the Burgers vector of the dislocation *CD*, as shown in Fig. 12(b). As can be seen, the jogged segment no longer lies in the primary slip plane. Further motion of the dislocation would be achieved by edge segments on either side of the jog bowing out in the primary slip plane, and the jogged segment would be able to follow by gliding on the "slip plane" defined by the Burgers vector and jog line direction, that is, the plane marked *PQRS*. In this way, the intersection impedes the motion of the edge dislocation to some degree. Consider the situation in Fig. 12(c), where the dislocation is now screw in character, as shown. The resulting jog (Fig. 12d) is again in the direction of the Burgers vector of the

intersected dislocation (*CD*), but in this circumstance, the jog is edge in nature, along a screw dislocation. Upon further motion of the dislocation *AB*, the jogged segment is essentially sessile, because motion in the direction of the dislocation *AB* can only be achieved nonconservatively, that is, by the formation of either vacancies or interstitials. Depending on the local state of stress, it may be possible for the jogged segment to glide along the dislocation line until it encounters the edge component, and then glide could occur as discussed previously. In this second case (Fig. 12c, d), the formation of an essentially sessile jog would be a significant impediment to dislocation motion. Of course, if this were to happen frequently, for example, the interaction of a mobile dislocation with forest dislocations, then significant hardening would result.

The second type of interaction to be described leads to dipole formation. Consider the operation of two similar sources, S_1 and S_2, producing dislocations on parallel slip planes. As the first dislocation with positive sign from S_1 (marked "A") glides to the right in Fig. 13(a), the dislocation of opposite sign from S_2 (marked "B") glides to the left. These dislocations interact with one another, the forces acting between them being shown in Fig. 13(b). As can be seen, when the distances *x* and *y* are equal, the force between them is

zero. If, for example, dislocation "B" attempts to move further to the left, it experiences a positive force (i.e., in the positive *x*-direction) resisting such motion. Likewise, if "B" attempts to glide to the right, it experiences a restoring force in the negative *x*-direction. The dislocation "A" experiences the same set of forces, and, in essence, these two dislocations are locked in position; because they are of opposite sign, they form a dislocation dipole. The strength of the interaction depends on the separation of the slip planes, *y*.

It should be noted that a comprehensive treatment of stress- strain or cold-work behavior is extremely complicated, because this would require a sufficiently complex integration of individual elements beyond the treatment of a single dislocation/barrier interaction. However, the general treatment, first conducted by Taylor (Ref 67), allows for a rudimentary understanding of the observed phenomenon when considered together with a sufficient understanding of the corresponding dependency limitations (e.g., temperature, solute content, stacking fault energies, etc.). In Taylor's treatment, it is possible to account for the barriers to dislocation motion due to the presence of other interacting dislocations rather than solute atoms, precipitates, or dispersoids. The interaction may occur in a variety of ways, including, but not limited to the original treatment of interacting edge dislocations on parallel slip planes by Taylor (Ref 67), the treatment of dislocations approaching a pileup with a resulting backstress (e.g., at grain boundaries or dislocation loops left around particles or dispersoids), and the treatment of dislocations approaching an array of forest dislocations, which has both long-range (field and bowing) and short-range (individual dislocation) components.

It is useful to first consider the case of deformation of single crystals before attempting to understand the behavior of polycrystalline materials. Figure 14 shows schematically the observed stress-strain curve for a crystal oriented such that slip occurs on only one slip system. It is clear that three distinct stages occur that correspond to a marked increase in strain (e.g., excluding the prestrain elastic region). In stage I (easy glide), dislocations moving on the slip system with the highest resolved shear stress account for plastic strain. Without the presence of other obstacles (e.g., boundaries, precipitates), it is possible for the dislocations to move over large distances (up to the crystal diameter), accounting for the considerable shear strain. In this stage, the deformation has been observed to be characterized largely by edge dislocation dipoles. Screw dislocations, if originally present, easily cross slip and annihilate with those of the opposite sign. It is generally accepted that this behavior occurs for fcc and hcp materials and high-purity bcc materials (Ref 68–70). The work-hardening rate ($_I/G \sim 10^{-4}$) is equivalent for these three systems. The degree of strain over which stage I occurs is a function of temperature, composition, size,
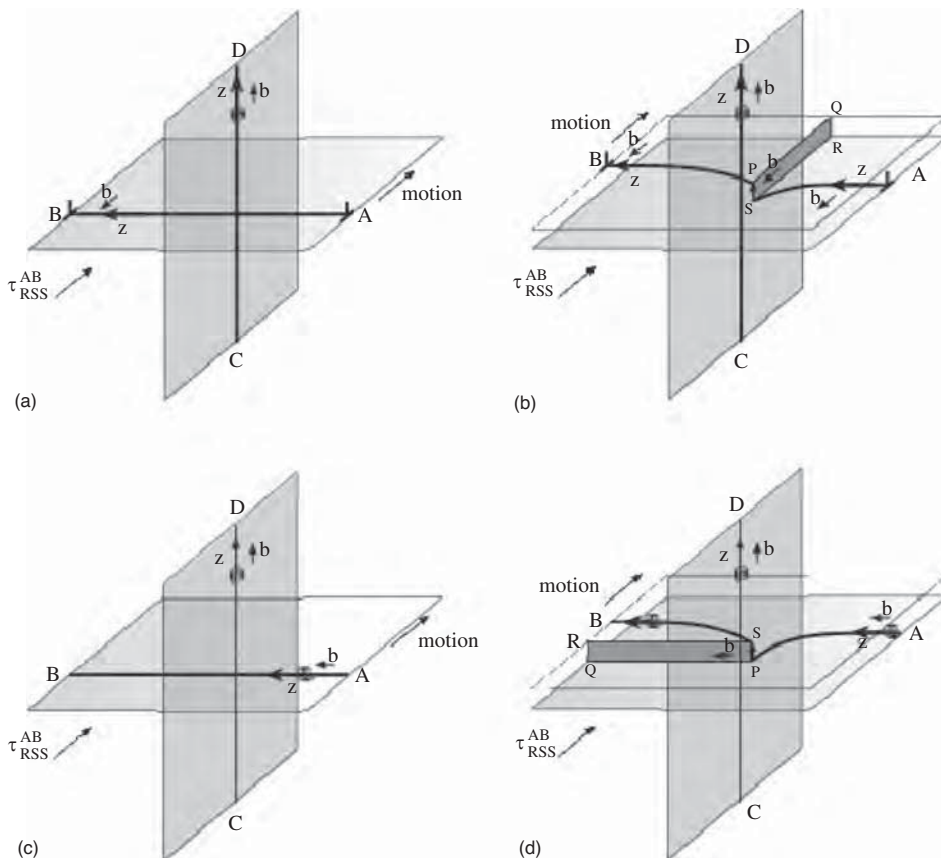
Fig. 12 Two discocation reactions (a, c) and their products (b, d). See text for description
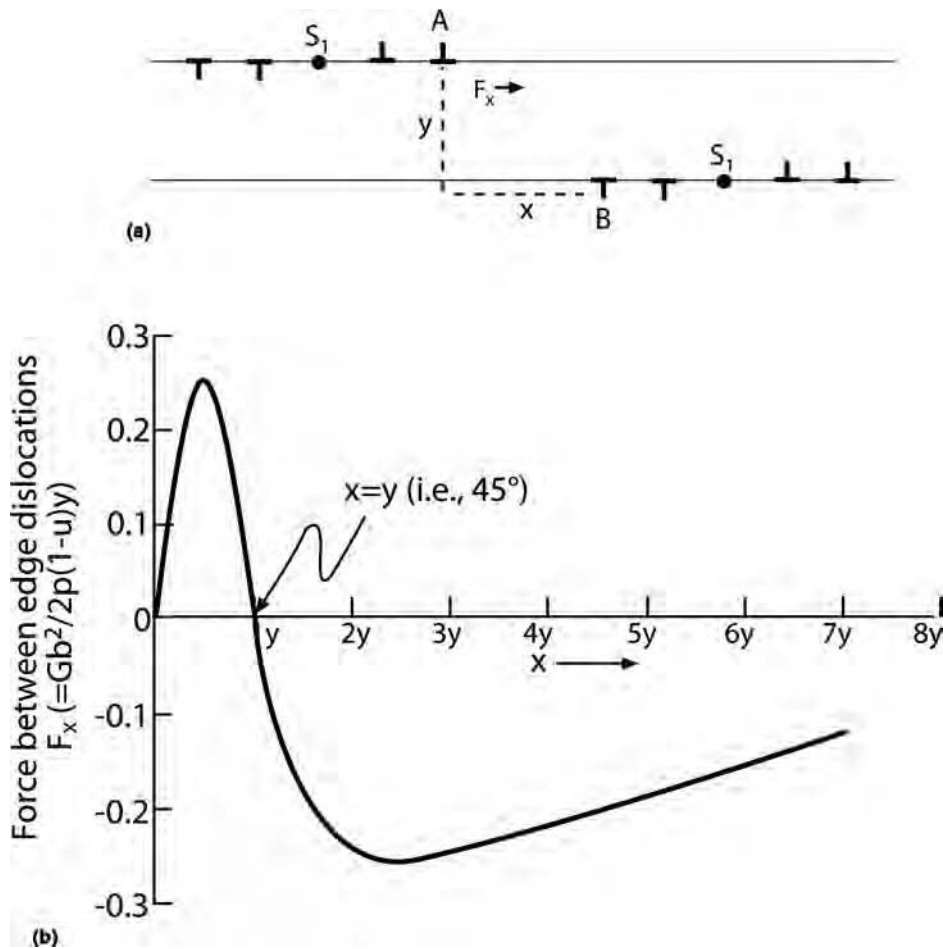
Fig. 13 Dislocation dipole formation and forces

Fig. 14 Regions of single-crystal slip

crystal orientation, and, for hcp crystal structures, the dominant slip system. For example, while the degree of easy-glide deformation hcp systems that deform by prism slip (e.g., titanium, zirconium) is equivalent to fcc and high-purity bcc (i.e., ~20%), hcp systems that deform by basal slip (e.g., magnesium, cobalt, zinc) can exhibit easy-glide strains out to failure (~ 100 to 500%). It is notable that for lower-purity bcc materials, stage I does not necessarily occur, and the material may exhibit a stress-strain curve similar to that of polycrystalline bcc material (Ref 68). It is important to consider the role of stacking-fault energy. It has also been observed that, with regard to the dependence on crystal orientation, stage I easy glide is not as distinguishable in single crystals oriented such that multiple slip systems are possible (e.g., toward the corner of the stereographic triangle) (Ref 68). In such systems, the multiple slip systems lead to dislocation/dislocation interactions other than dipoles of stage I.

The nature of these interactions between multiple slip systems is similar to that which occurs in stage II behavior. Stage II is characterized by a significant increase in the work-hardening rate of the material (θ_{II}/G ~3 ×

10^{-3}, for most metals), which has been attributed to the multiplication and movement of dislocations on secondary slip systems. Barriers to dislocation motion, such as forest dislocations, jogs, or Lomer-Cottrell barriers (or sessile, nonmobile, stair-rod dislocations), can readily form as products of the multiple slip systems. Thus, the slip-line length decreases, requiring a greater stress to achieve the same strain. Because the activation of such secondary slip systems is dependent upon orientation, it can be shown that the strain at which stage II starts is dependent upon this parameter. In stage III, dynamic recovery begins, and the observed curve is parabolic. Because recovery is also dependent upon composition and temperature, the onset of stage III is also proportional to temperature.

The behavior of the polycrystalline material is significantly different, complicated by:

- The multiple orientations of the individual crystal grains
- The necessary activation of multiple slip systems within a given grain to satisfy the von Mises condition of five independent slip systems to accommodate an arbitrary shape change, which arises from the six
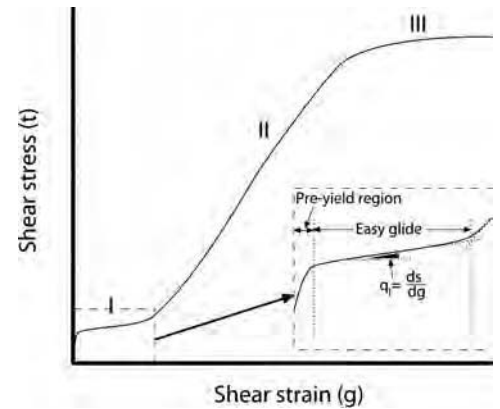
independent strain components, less one for the condition of constant volume (Ref 71, 72)
- The smaller size of the individual grains with the associated hard barrier of the boundaries, which reduces the slip length

Notably, the need to operate five independent slip systems to accommodate an arbitrary shape change in polycrystalline materials may be used to explain why certain hcp metals (e.g., zinc, tin) that deform by basal slip (only two independent slip systems) have such remarkably low ductility in the polycrystalline form, while they exhibit such a long stage I easy-glide region for single crystals. Indeed, their stage I easy-glide strain is far greater than that of other materials.

Modeling Strengthening by Cold Work. Because the products of dislocation interaction on multiple slip systems provide the barriers to the motion of other active dislocations, a hardening model that includes the spacing (l) of the dislocation- product barriers is given as:

$$\tau \approx \frac{\alpha G \mathbf{b}}{l}$$

which is identical in form to those of the bypass models of precipitation and dispersion hardening. In the original treatment by Taylor, which considers the work hardening due to an elastic interaction, the variable l was considered to be the mean distance between the dislocations. This may be approximated as $\rho^{-1/2}$, where ρ is the dislocation density. In this case, the constant α is given as:

$$\alpha = \frac{1}{8\pi(1 - \upsilon)}$$

and:

$$\tau \approx \alpha G \mathbf{b} \sqrt{\rho}$$

In experimental evaluations of these equations, it has been determined that α ranges from the low values (~0.05 to 0.1) up to 1. Additionally, it has been observed that the form of the

dislocation density varies depending upon which elements of the dislocation networks act as barriers. If the dislocation density is sufficiently small, or the primary dislocations have yet to rearrange into cell walls, then the density may be considered as ρ_p (primary), where the main work-hardening component is from primary dislocation interactions. However, if there has been arrangement of the dislocations into a walled or forest configuration, the dislocations included in the density should include those in the wall or forest configuration. In this case, the stress is derived from a similar argument to Orowan looping. In either case, the form remains the same.

Because the shear strain may be related to the dislocation density, it is possible to derive an elementary form of the equation showing the parabolic nature of the stress-strain curve:

$$\tau = C\mathbf{b}^{1/2}\gamma^{1/2}$$

This form requires that one make several assumptions about the shape of the dislocation loops, which will change the form of the constant C (including the α term). A more generalized form may be taken as:

$$\tau = \alpha G\left(\frac{\mathbf{b}}{x}\right)^{1/2}\gamma^n$$

where n is the strain-hardening coefficient, and x is the average distance a dislocation moves before an obstacle arrests its motion. It should be noted that these equations may be modified to include a temperature-independent term (typically τ_G) and a temperature-dependent term (τ_S), accounting for the temperature effect on modulus. It has been noted that both of these terms are important in pure fcc materials, whereas in alloys and bcc materials, the solid-solution effects and Peierls stress dominate τ_S, which is greater than τ_G. These equations may be used to describe the phenomenon observed in either stage II or that observed in polycrystalline materials, where multiple slip and barriers dominate the hardening term.

It is possible to relate the polycrystalline work-hardening rate ($d\sigma/d\varepsilon$) to the single-crystal work-hardening rate ($d\tau/d\gamma$) by:

$$\frac{\partial\sigma}{\partial\varepsilon} = m^2\frac{\partial\tau}{\partial\gamma}$$

where m, in polycrystalline materials, is equivalent to the Taylor factor (M) and $\sim$2 to 3 (bcc) (Ref 19, 69), $\sim$3 (fcc), and $\sim$6.5 (hcp absent of twinning).

For dispersion-strengthened materials, as described previously, there is also a work-hardening component to dispersion strengthening, where the multiplicity of dislocations around the nondeformable dispersoids (either debris loops or GND prismatic loops) further strengthen the material. At small strain (e.g., <1%, prior to a significant effect of dislocation multiplication), the yield strength may be given as:

$$\sigma = \sigma_i + \alpha G f^{3/2}\varepsilon$$

where f is the volume fraction of dispersoid particles. At larger strains, when the Orowan loops and GNDs play a role, the form may be given as:

$$\sigma = \sigma_i + \alpha G\left(\frac{f\mathbf{b}}{d}\right)^{1/2}\varepsilon^{1/2}$$

where d is the particle diameter. In principle, the same phenomenon would occur in overaged precipitation-hardened systems, but the typical size of those precipitates is much coarser than the typical effective dispersoid size. Thus, the second component of the equation approaches zero ($d >> \mathbf{b}$).

Strengthening from Interfacial Barriers to Slip Transmission

For many of the relevant engineering alloys, including iron- and titanium-base alloys, strengthening may be derived from the presence of domains of two-phase lamellar microstructures. This strengthening is derived due to the presence of the interphase interfaces, which act as barriers to slip transmission into and across a second phase. Thus, for grain boundaries in single-phase materials, the ease of slip transmission depends on how nearly parallel the slip plane in the adjacent grain is to that of the deforming grain. Generally, transmission results in a residual defect, with an irrational Burgers vector being left in the interface. The magnitude of the Burgers vector of the residual defect depends on factors such as slip-plane alignment, but its presence will complicate further slip transmission by the active slip system in the primary deforming grain. In a two-phase material, there is an added complication from the fact that slip is transmitted from one crystal structure to a different one across the given interphase boundary. As can be appreciated, the situation for modeling the transmission event is extremely complicated, and well-developed models are not yet commonly available.

Examples of Predictive Models

As was noted at the beginning of this article, the prediction of tensile properties in engineering alloys is an extremely complicated task. As seen in the earlier discussion, the individual strengthening mechanisms have been recognized and basic mechanisms have been identified to various degrees of confidence. However, the modeling of the contributions of these mechanisms to tensile strength usually involves a number of simplifying assumptions, leading to a parametric or phenomenological relationship describing increases in yield strength. There is an essential absence of models describing contributions of strengthening mechanisms to ultimate tensile strength, and there are no models to predict tensile ductility.

Consequently, in the recent past, there has been an emphasis on the development of more accurate models to predict contributions from the various strengthening mechanisms, as well as development of finite-element method/crystal-plasticity models. Additionally, rules-based approaches have been employed to yield quantitative predictions of tensile properties as a function of both heat treatment (Ref 73–75) and microstructural and compositional parameters (Ref 76–79). Some examples of these efforts are described in the following.

Atomistic Modeling of Dislocation Structures and Slip Transmission

Many of the emerging techniques aim to develop and apply physics-based multiscale modeling tools for predicting strengthening contributions to mechanical properties. Among the major modeling and simulation approaches currently being employed are ab initio (Ref 80–83) atomistic (Ref 84, 85), and phase-field (Ref 86–89) techniques.

Ab Initio and Atomistic Methods

Ab initio modeling holds the promise of being a highly accurate and predictive method. A drawback is computational complexity, because models often contain only a few hundred atoms. However, it may be used to extract fundamental quantities such as phase stability (Ref 90), elastic properties (Ref 80, 82), diffusion barriers, electronic and optical responses (Ref 91, 92), and so on. Using interatomic potentials, atomistic calculations can be used to probe defect processes. In particular, it can be applied to study dislocation processes (Ref 93) (Fig. 15). Recently, researchers have made progress in applying the nudged elastic band (Ref 93, 95) method to locate the saddle-point configurations of stressed systems consisting of a large number (10^2 to 10^3) of nontrivially activated atoms. It is now possible to compute the three-dimensional atomistic activation processes at crack tips, such as dislocation bowout (Ref 95) (Fig. 16) and crack-front kink nucleation and migration (Ref 93), as functions of the stress-intensity factor. While it is possible to handle more complex geometries, atomistic calculations in such idealized, well-controlled geometries using the empirical potential of a pure element are meant to serve as calibrations for semianalytical theories, such as that attributed to Rice (Ref 96–98), which, when calibrated, can use ab initio energetics directly for more complex alloys.

Li et al. have probed the details of the thermally activated slip transmission across various interfaces (copper twin boundary, α/β interfaces in titanium alloy) with atomistic reaction-path calculations and microscopic phase-field models (Fig. 17). The athermal stress threshold and activation volume of these interfacial/bulk
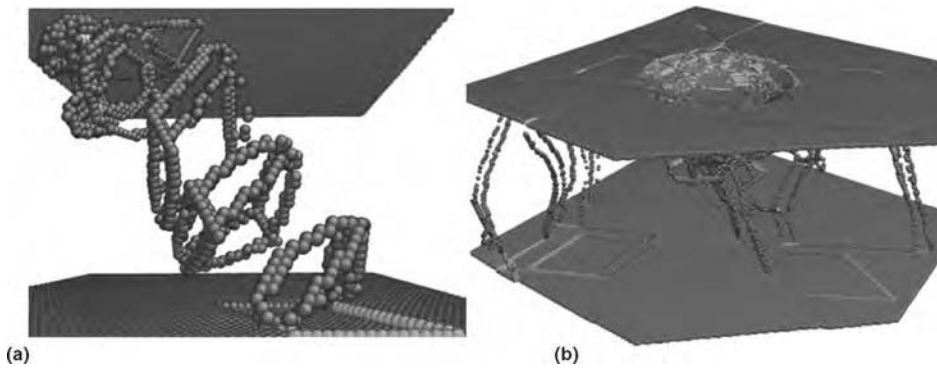
Fig. 15 Dislocation process. (a) Concatenated snapshot of a heterogeneously nucleated prismatic dislocation loop gliding through an aluminum thin film. Source: Ref 94. (b) Microstructural development in a two million atom molecular dynamics simulation of (111) surface indention of aluminum thin film. Operations of a threefold symmetric prismatic dislocation loop source result in the parallelepiped slip lines on the bottom. Highly mobile, through-thickness threading dislocations are nucleated in the nanoindention as well. Courtesy of J. Li
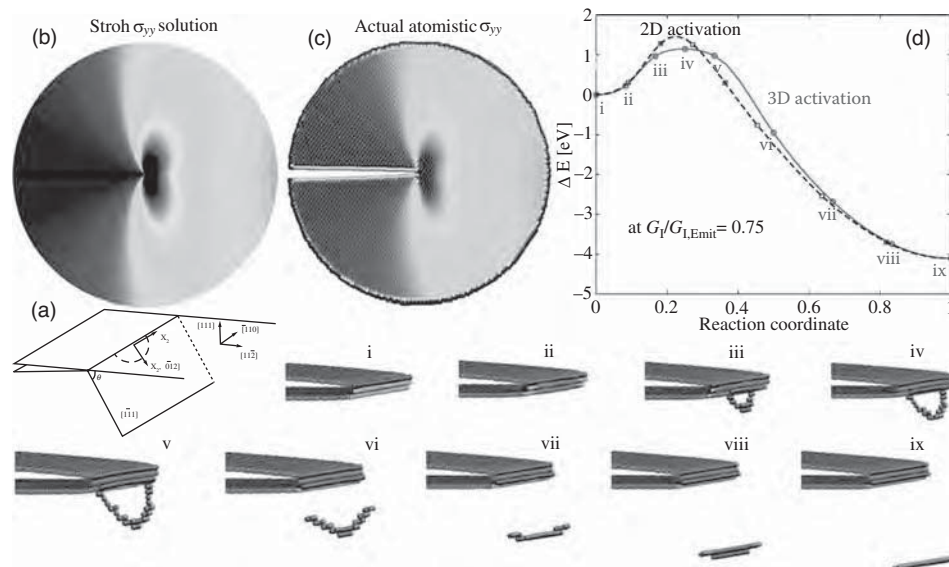
Fig. 16 Three-dimensional (3-D) atomistic activation process at crack tip. (a) Geometry of a mode I crack containing 24 unit cells (61 Å) in $x2$ and 103,920 copper atoms in a $R = 80$ Å cylinder. Source: Ref 95. (b) Continuum Stroh solution and (c) the actual atomistic local stress distribution of σyy at G_I/G_I, emit = 0.75. (d) 3-D activation minimum energy pathway (MEP) (solid line) of partial dislocation emission by bowout and its competing two-dimensional (2-D) pathway (dashed line). i to ix show the sequential nine nudged elastic band images on MEP, with iv being the saddle point; atoms whose coordination number differs from 12 are not shown. Courtesy of J. Li

dislocation reactions are calculated and found to be in reasonable agreement with experiments. They have computed the strain-rate dependence of deformation in metals whose yield strength is derived mainly from interfacial resistance. The accumulated interfacial dislocations will, in turn, influence the activation energy due to dislocation-dislocation interactions. This newly proposed interfacial hardening law is distinct from the conventional bulk work hardening.

Phase-Field Modeling

Although the phase-field method is most commonly used to describe the spatial and temporal details of microstructural development, it has also been used to describe the evolution of dislocation structures (i.e., dislocation dynamics) of a system. Phase-field modeling uses gradient thermodynamics of nonuniform systems in which two spatially continuous field variables are solved simultaneously. These field variables often describe energy and order of the system. For example, for the case of microstructural evolution, the field variables describe the compositional variations (energy) and structural symmetry (order). Such an approach can, in principle, describe the evolution of features at the mesoscopic length scale of any nonuniform system. The spatial and temporal evolution of dislocation structures is of interest in the prediction of tensile properties. Many of the efforts on increasing the maturity of the

phase-field method, including applications to dislocation dynamics, have been conducted by Y. Wang, whose companion article, "Phase-Field Microstructure Modeling," in this Volume more fully describes the use of such tools. Two examples of dislocation dynamics are shown in Fig. 18(a–d). Fig. 18(a) shows a description of dislocation interactions using the phase-field method. Here, two dislocation loops on different slip planes in an fcc material have reacted, resulting in a sessile dislocation segment. Figure 18(b) shows dislocation multiplication under increased applied uniaxial loading. Figure 18(c) shows the equilibrium dislocation structure in a γ/γ' microstructure, where the dislocations are the boundaries between the slipped and unslipped regions. Figure 18(d) shows a snapshot of a dislocation dynamics phase-field model where the dislocations are cutting the γ' precipitates and considers a newly identified reordering phenomenon (Ref 99–101).

Finite-Element Modeling

It is common to model the continuum response of the material using finite-element modeling (FEM) approaches. Efforts by S. Ghosh et al. (Ref 102–107) have focused on using adaptive meshing and accurate three-dimensional mesoscopic microstructural features (i.e., grains) to predict the properties of a material. The accurate grain reconstructions are based on experimental datasets. Following three-dimensional rendering, rigorous analysis of a suite of statistical grain descriptors have been developed, including those which describe the distributions of grain size and shape from the orientation and misorientation distributions obtained using electron backscattered diffraction. In addition, the correlation between statistical descriptors was measured, which was made possible through the ability to examine each feature individually. When analyzed, these descriptors have been used to produce statistically equivalent synthetic microstructures (Fig. 19a–c), which have been included in FEM. While the determination and inclusion of accurate descriptions of grain shape and distribution represent a significant breakthrough in such models, it remains computationally challenging to include the details of features that may occur within the grains. For example, in the case of nickel-based superalloys, the γ' distribution is often considered as part of the constitutive behavior of the individual grains rather than as a separate phase. While this may be a reasonable approximation for systems with such symmetry between the matrix and precipitate phase, it is less so for the more complex two-phase microstructures such as titanium-base alloys, which also provide challenges in determining a predictable constitutive response. Thus, a challenge these methods still face is the ability to include multiple phases that span across length scales in a computationally inexpensive fashion.
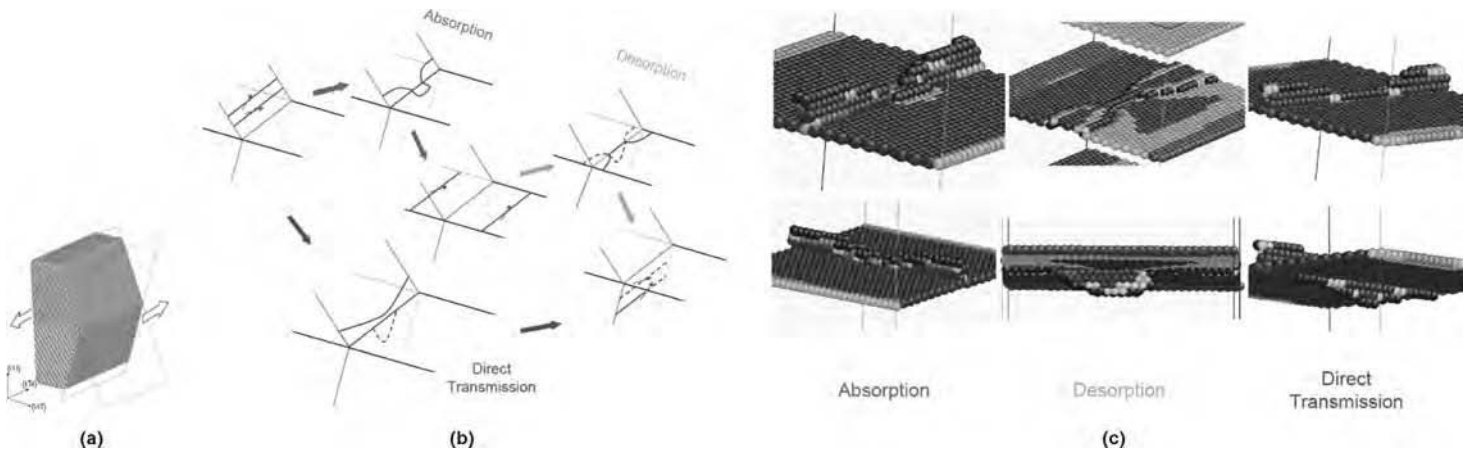
Fig. 17 Thermally activated slip transmission. (a) Coherent Σ3 {111} twin boundary in copper. (b) Schematic of two competing pathways of slip-transfer reaction discovered from nudged elastic band calculations. (c) Atomic configurations of absorption, desorption, and direct transmission. Courtesy of J. Li
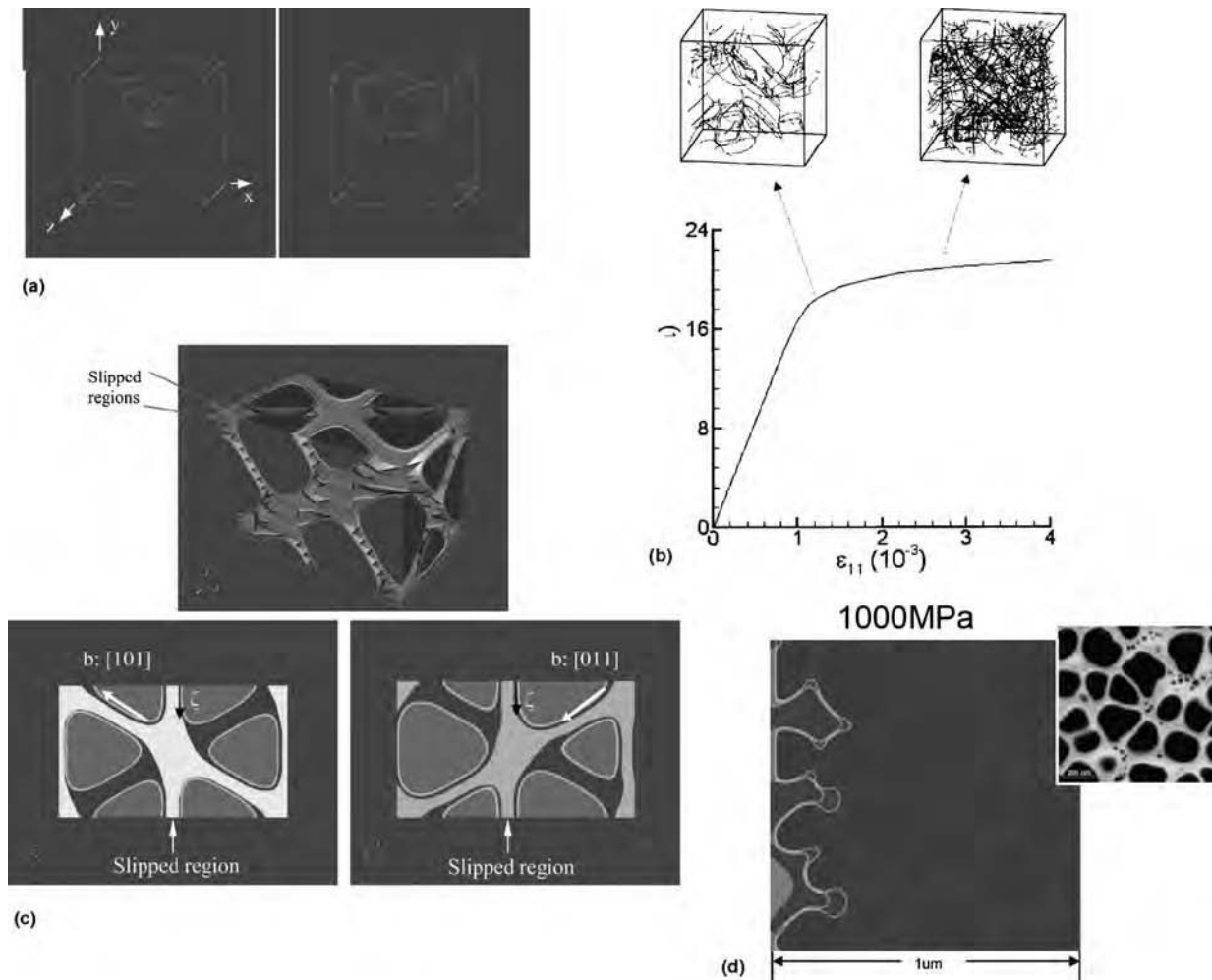
Fig. 18 Phase field modeling techniques applied to show (a) reaction of two dislocation loops resulting in a sessile component in the stair-rod configuration; (b) dislocation content at two locations of a stress-strain curve; (c) dislocations in a nickel-based superalloy of various types showing the slipped regions; and (d) dislocation activity, including the separation of partials, in a nickel-based superalloy whose microstructure has been informed from experimental data (inset). Courtesy of Y. Wang
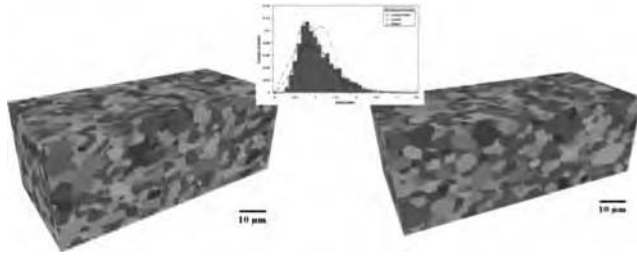
Fig. 19 Finite-element modeling. (a) Three-dimensional reconstruction of IN-100 nickel-based superalloy of a volume with dimensions 96 × 36 × 46 μm. (b) Example of grain descriptor, the distribution of grain volume. (c) Synthetic volume containing 5269 grains. Courtesy of S. Ghosh and M. Groeber
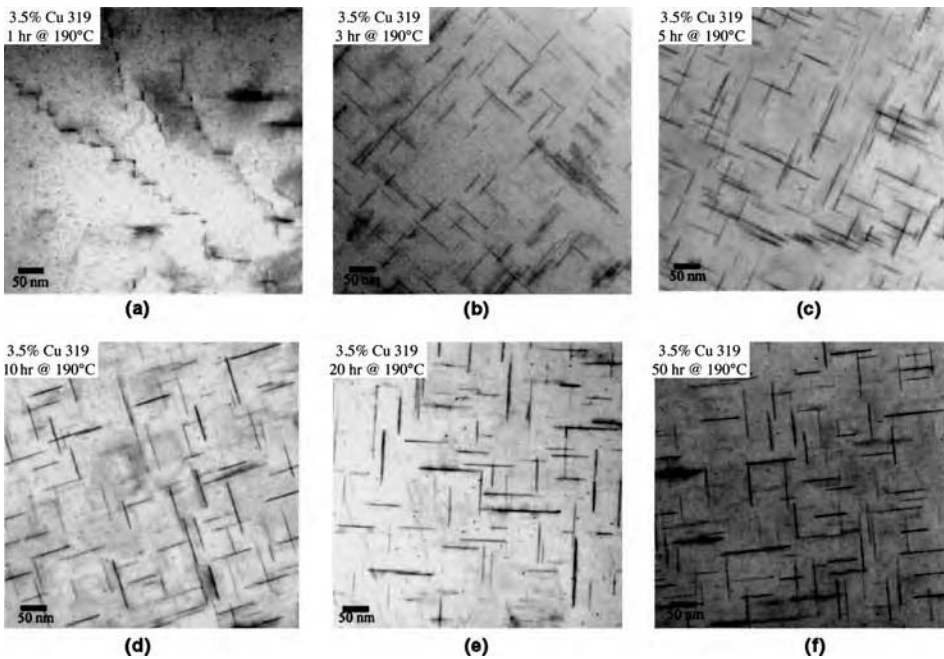
Fig. 20 Effect of aging on precipitation in an aluminum W319 alloy. Source: Adapted from Ref 2

Cast Aluminum Property Prediction

For aluminum alloys, one of the most detailed mechanistic models currently available includes contributions from the various strengthening mechanisms in order to describe the yield strength in a cast aluminum alloy 319 (Ref 2) as a function of aging conditions. The individual strengthening contributions present in the model are from the intrinsic strength of the matrix (including both grain-boundary strengthening and strengthening from the eutectic phase), solid-solution strengthening, and precipitation strengthening. Thus, the model may be written as:

$$\sigma_{tot} = \sigma_i + \Delta\sigma_{ss} + \Delta\sigma_{ppt}$$

Because variations in strength due to microstructural evolution occur in aluminum alloys such as aluminum W319, the model considers the effect of temperature on the strengthening terms. Because these alloys are solution treated and quenched, the solid solution is initially supersaturated with solute. With aging, the degree of supersaturation will decrease until an equilibrium level is reached, which is itself a function of temperature. For this model, the Guinier-Preston-Bagaryatskii zones have been included in the solid-solution effects, because they are difficult to quantify. The precipitation-strengthening effect includes both looping and shearing and is based on some of the emerging models of precipitation distributions and their hardening effects. The combination of looping and shearing may be written as:

$$\Delta\sigma_{ppt} = \left[\frac{1}{\Delta\sigma_{shear}} + \frac{1}{\Delta\sigma_{bypass}} \right]^{-1}$$

where the change in strength due to shear and bypass is appropriately indicated. The form of this equation allows for the combination of the two competing mechanisms, while ensuring that the active mechanism dominates the term,

because the other mechanism will drop significantly in magnitude upon reaching a critical particle size, assuming uniform size and distribution. As discussed in the section "*Emerging Concepts*" in this article, the simplified models for bypass often assume a random array of precipitates and tend to exclude the precipitate morphology and/or its orientation relationship with the matrix. For this model, the two functions describing the change in stress with particle geometries were evaluated. These are (Ref 22):

$$\tau_{\theta'} = 0.13 \left(\frac{G\mathbf{b}}{\sqrt{dw}} \right) \left(\sqrt{f} + 0.75\sqrt{\frac{d}{w}}f + 0.14\frac{d}{w}f^{3/2} \right) \times \left(\ln \frac{0.87\sqrt{dw}}{r_0} \right)$$

where d is the diameter of the precipitate, w is the thickness, f is the volume fraction, and r_0 is the inner cutoff radius, which is set to $\mathbf{b}$, or (Ref 21):

$$\sigma_{ppt} = M\left(\left(\frac{G\mathbf{b}}{2\pi\sqrt{1-\nu}} \right) \left(0.931\sqrt{\frac{0.306\pi dw}{f}} - \frac{\pi d}{8} - 1.061w \right) \left(\ln \frac{1.225w}{r_0} \right) \right)$$

For this model, the initial precipitate variables (i.e., size and fraction) have been determined by careful measurements using the transmission electron microscope (Fig. 20). For computational efforts, the terms for the size and fraction of the precipitates may be modified to account for the experimentally determined effects of thermal exposure on their spatial evolution.

After development and comparison with other mechanism-based models (Nie, Shercliff, and Ashby), the author (Ref 2) extracted the predicted contributions of the effects of the individual strengthening mechanisms as a function of time (Fig. 21). Weakley-Bollin et al. analyzed these models (Ref 2) and found that for Al-Si-Cu alloys, each predicted peak hardnesses that occur too late compared with experiment, which was attributed to the fact that both models are more sensitive to the thickness (w) of the precipitate than to the diameter. Unfortunately, the true distribution of precipitate thickness is extremely difficult to quantify, because the thickness of the strengthening precipitates in these systems is often less than 5 nm.

The limitations identified in their research include an absence of a fundamental model to predict the growth of the strengthening phases (θ') as a function of time, temperature, and solute content. The current model is based on extensive characterization of samples taken from each condition. Therefore, it is clearly necessary to develop more accurate models of the initial stages of the precipitates as well as details of their evolution. State-of-the-art characterization techniques, such as the local area atom probe (Ref 108, 109) and aberration-corrected transmission electron microscopes
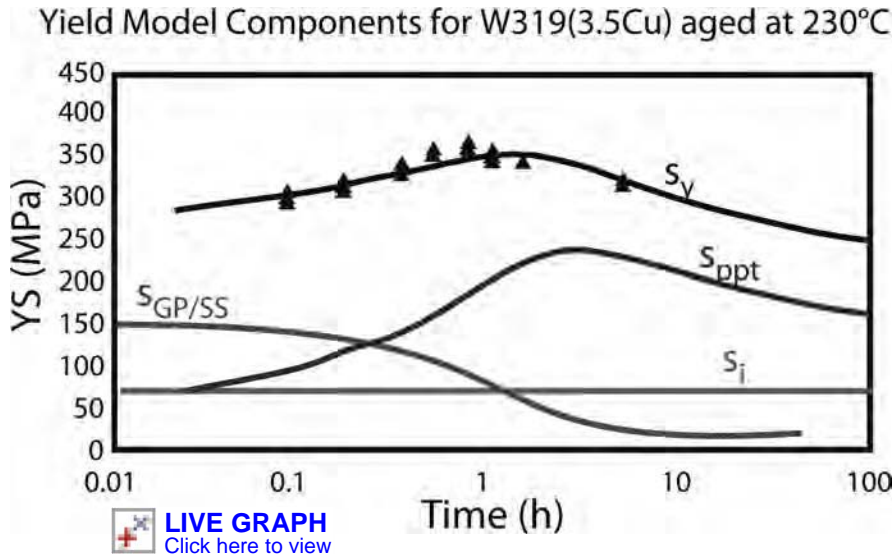
Fig. 21 Calculated effects of the individual components of the aging response. Source: Adapted from Ref 2
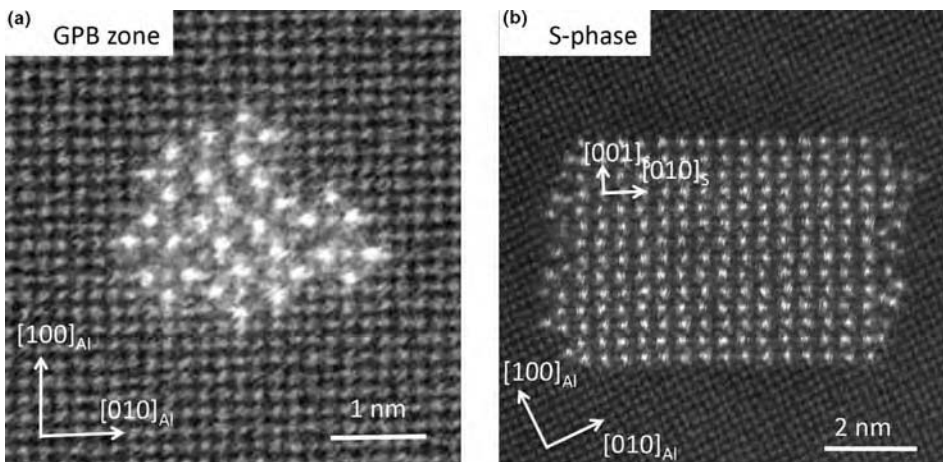
Fig. 22 High-resolution scanning transmission microscope image of (a) Guinier-Preston-Bagaryatskii (GPB) zone and (b) S-phase. Source: Courtesy of Libor Kovarik

(Ref 109, 110), are making such detailed understanding of the precipitate structures, morphologies, relationship with the matrix, and sequences possible (Fig. 22). Rigorous studies of the phase relationships and precipitation sequences (Ref 108, 111–113) are also underway. The combination of these efforts will lead to more robust thermodynamic-based predictive models and should lead to next-generation advanced precipitation-strengthened systems.

Use of Rules-Based Approaches: Neural Networks

Rules-based approaches and data-mining techniques may be used to provide quantitative predictions of tensile properties. In general, a database relating properties to various parameters (e.g., microstructural features) is divided into two parts; one part is used to train a model, and the other is used to test its reliability. Here, *prediction* corresponds to interpolation within the ranges of values of the parameters contained in the given database used to develop the model. These models are not based on physical realities or mechanistic information but provide engineers with a way of predicting properties.

One such rules-based approach involves neural networks. These networks use a hyperbolic tangent-fitting function to perform nonlinear regression to develop relationships between input parameters and the given output (here, tensile properties). The resultant models may be used in two ways. First, they may be used to provide a quantitative (interpolative) prediction of properties. Second, they may be used to perform virtual experiments. This latter application of a developed neural network provides a way to perform controlled experiments

that cannot be done in the laboratory, as shown subsequently. Clearly, the development of neural-network models requires an accurate and well-populated database relating input and output parameters. These are often not generally available. In the following, an example of the application of neural-network modeling to the prediction of properties in α/β titanium alloys is given.

Example: Neural Networks for α/β Titanium Alloys. Over a period of several decades, α/β titanium alloys have been studied intensively due to their important technological applications in a wide range of fields, especially in aerospace (Ref 114–116). Interestingly, the factors that influence and determine their mechanical properties are not well understood. It has been claimed that they depend on the characteristics of their microstructural features (Ref 117–119). However, the microstructures of such real titanium alloys are rather complex and can exhibit a range of spatial distributions and morphologies of the two primary phases (hcp α and bcc β), which vary according to the thermomechanical processing. For example, as shown in Fig. 23(a–c), the α-phase can manifest as either a colony-type morphology (Fig. 23a) or basketweave Widmanstätten morphology (Fig. 23b), depending on the cooling rate from the β-phase field, or as equiaxed α-particles (Fig. 23c) if worked in the α+β phase field. Because the spatial and morphological distributions of the individual microstructural features are themselves interdependent, it is experimentally not possible to isolate the effect of the individual features. Thus, while it is reasonable to assume the individual strengthening mechanisms described previously in this article act in concert to dictate the mechanical properties, it is very difficult to isolate the individual strengthening mechanisms. Indeed, establishing correlations between microstructural features and a given property has proven to be an elusive goal for previous researchers. For example, while Boyer et al. (Ref 120) suggested that the tensile properties of β-processed microstructures must be correlated with the colony size, prior-β grain size, and α-plate width, no such correlation could be established through experimental techniques. A similar observation was made by Rhodes et al. (Ref 121) while trying to establish the correlation between microstructural features and fatigue crack growth rate. It was suggested by these authors that the relationship between properties and the microstructural features is determined by the complex combination of the parameters. The previous state-of-the-art for relating microstructural features to the properties of these alloys is that proposed by Lütjering et al. (Ref 122), where the experimentally observed microstructural influence on properties is indicated qualitatively by + (positive), − (negative), or 0 (no influence).

To overcome the challenges of the complex and interrelated microstructures, artificial neural networks have been used to develop a
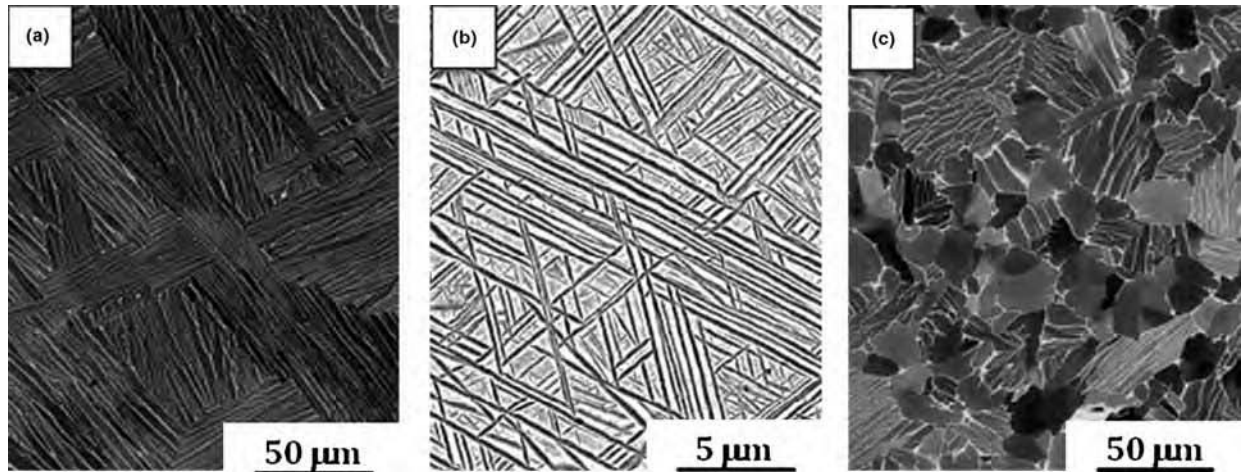
Fig. 23 Examples of microstructures that occur in titanium-based alloys. (a) Colony, resulting from slowly cooling the material, (b) basketweave, resulting from a more rapid cooling of the material, (c) equiaxed alpha particles, which result from working the material in the two-phase region.

microstructurally based quantitative (interpolative) predictive model for mechanical properties of α/β titanium alloys. Artificial neural networks are a convenient modeling tool for developing a predictive model in the short term, because they can handle a large number of variables with complex interdependencies and do not need any prior knowledge of the physics of the problem. The authors of this article have used this approach to develop models to predict both the yield and the ultimate tensile strength of β-processed and α+β-processed Ti-6Al-4V at room temperature (Ref 76, 77, 123).

β-Processed Ti-6Al-4V. The database for the β-processed Ti-6Al-4V contains the values of the thickness of α-lath, volume fraction of total α, colony size factor and prior-β grain factor as input parameters, and the tensile properties (yield strength and ultimate tensile strength) as output parameters. The neural network was trained separately for each of the two different output variables. Figure 24(a,b) shows the predictions of yield strength and ultimate tensile strength for 16 test data points (which were not used in training). The dashed line is the zero error line, and the solid lines are $+5\%$ error lines. It is seen that the trained neural network is able to predict the yield and ultimate tensile strength values of Ti-6Al-4V from the microstructural features within $+3$ and $+4\%$ error, respectively. It is of interest to identify the factors that limit the accuracy of the predictions. Because the approach used here essentially involves nonlinear regression, it is immediately obvious that the accuracy may be improved by the use of a larger population of the database. In addition, it is important to consider whether important microstructural features have not been included in the analysis. For example, the effects of local texture differences have not been quantified and included in the model. Also, the oxygen concentration, thought to harden α-titanium, has not been measured.

In addition to this promising predictive capability, neural networks can also be used to

extract the functional dependencies of the tensile properties on individual microstructural features. These dependencies can guide the development of a physically based model for the property-microstructure relationship. Thus, experimentally it is impossible to vary one particular microstructural feature independent of others; virtual experiments can be performed using trained neural networks whereby each individual microstructural parameter can be varied without changing the values of the others. For example, while the variation in yield strength as a function of α-lath thickness is plotted in Fig. 25(a) for various samples, it should be noted that the values of other microstructural features, such as colony size factor, prior-β grain factor, and volume fraction of total α, are also simultaneously changing due to their interdependent response to the given heat treatment. As a result, it is not possible to obtain an independent trend of variation in yield stress with thickness of α-lath. A solution involves the use of the trained neural network to perform virtual experiments. Thus, properties may be predicted as a function of the value of the microstructural feature of interest while keeping the other microstructural input parameters at constant values within their experimental range. Such a trend plot has been demonstrated in Fig. 25(b) for the variation of yield stress with thickness of α-lath, where the values of the other microstructural parameters (volume fraction of total α, colony size factor, and prior-β grain factor) are kept constant at their mean values within their range in the training database. From this trend plot, it is expected that the yield stress would decrease monotonically with increasing thickness of α-lath. Similarly, the trend plot in Fig. 26 shows a monotonic decrease in yield stress with increase in colony size, but the effect is not as large as that of α-lath thickness.

These trend plots of strength variation on individual microstructural parameters are consistent with the expected behavior. For

example, with decreasing thickness of α-laths, the effective slip length decreases, and so, based on restricted lengths of possible dislocation pileups, the yield strength should inversely vary with thickness of α-lath. The trend plot produced from the neural network shown in Fig. 25(b) correctly predicts this behavior. Similarly, the neural network also predicts an inverse trend for yield strength with colony size. In this case, the strengthening associated with deformation of colony laths may also be rationalized based on the slip length within the alpha phase (Ref 124–126). Thus, the Burgers orientation relationship that exists between the α and β phases within a single colony allows for an easy slip transfer across the α/β boundary, and, because the α-laths within a single colony are all similarly oriented, the slip often traverses the entire colony. This should make the colony size factor a significant microstructural dimension affecting the yield strength in this alloy. Once again, the trend plot predicted by the neural network shown in Fig. 26 correctly predicts this behavior, although the trend is less significant compared with the effect of lath thickness on yield strength. This may be due to the increased number of α/β boundaries afforded by decreasing α-lath thickness for the same colony size. Thus, although it has been assumed (Ref 126) that slip transfer across α/β boundaries is relatively straightforward, these boundaries may offer more resistance to slip transfer than previously considered. It is also necessary to consider the spatial distribution of the individual features that occur within the grains when deriving constitutive laws for the mesoscopic behavior of the grains. This was clear when considering the effect of the prior-β grain size on yield stress. Although the Hall-Petch effect would suggest that a decrease in grain size would result in stronger material, the opposite was observed. Upon further study, it was discovered that the smaller grains often contained only colony-type microstructures, while the larger grains would contain a stronger

Fig. 24 Predictions versus experimental data for (a) yield strength (YS) and (b) ultimate tensile strength (UTS) in beta-processed Ti-6Al-4V

basketweave microstructure in the grain centers. Thus, the strengthening afforded by the basketweave more than offset any possible Hall-Petch effect that may exist.

$\alpha+\beta$-*Processed Ti-6Al-4V*. For this database, the microstructural features include the equiaxed alpha size (equiaxed α size, μm), the volume fraction of equiaxed alpha ($F_v^{equiaxed-\alpha}$), the volume fraction of total alpha ($F_v^{total-\alpha}$), and the width of the alpha-laths in the transformed β-v regions (α-lath width, μm) (Ref 127, 128). In addition to the microstructural variations, this database included designed variations in the alloy content to explore their effects on the resulting mechanical properties. The alloys studied, while within the specification range of Ti-6Al-4V, may be described as Ti-xAl-yV (4.76 < x < 6.55; 3.30 < y < 4.45) with controlled variations in interstitial oxygen and iron (0.07 < wt %O < 0.20; 0.11 < wt%Fe < 0.41). The results of the virtual experiments showed a similar effect of volume fraction total alpha and the thickness of the alpha-laths on the resulting

properties as was observed for the β-processed Ti-6Al-4V. In addition, it was established that the inclusion of the basketweave microstructure was important. A feature that is not present in β-processed microstructure is the equiaxed α-particles, which may be described according to their size and volume fraction. The quality of the model is shown in Fig. 27, where the predicted yield stress is plotted against the actual experimental value. As can be seen, the model displays an uncertainty of approximately $\pm2\%$.

For these microstructures, it was determined that the volume fraction of equiaxed alpha particles is the microstructural feature that has the greatest influence on the resulting tensile properties (Fig. 28a), largely due to the extended range of volume fraction of equiaxed alpha that can be affected in these microstructures, resulting in a variation in properties of almost 60 MPa. Thus, not only is it important to consider the normalized influence of a particular variable on the properties but also the extent of the range that a particular variable can exhibit. For example, for the two microstructural features related to scale, it is also important to note that while the thickness of the Widmanstätten alpha-laths may yield a more significant negative influence on properties per micrometer increase, the range of lath thickness that can be affected by thermomechanical processing is smaller than the range of equiaxed alpha–particle sizes, $\sim$0.345 < t_{lath} < 0.667 and 4.79 < $d_{equiaxed}$ < 8.39, respectively. The trend in volume fraction of equiaxed alpha (Fig. 28b) most likely represents a result of the difference in scale of the microstructural features being considered. That is, one may speculate that the early dislocation activity may be more easily accommodated in the equiaxed alpha-particles rather than the Widmanstätten alpha-laths, and that a higher fraction of equiaxed alpha would result in a lowering of the yield strength. A second effect that will occur upon an increase in the volume fraction of equiaxed alpha particles that may affect the yield strength is the effect of partitioning of both aluminum and oxygen to the equiaxed alpha, decreasing the effectiveness of the Widmanstätten alpha-laths to strengthen the overall microstructure through a modification to the critical resolved shear stress.

A similar analysis of the dependency of properties on alloy composition yields stronger influences, particularly with respect to the α-stabilizing elements, aluminum and oxygen. The influence of oxygen on the yield strength is clear (Fig. 29a), with an effect of $\sim$940 MPa/wt%. Aluminum has less of an effect on strengthening the hcp phase ($\sim$50 MPa/wt%). Vanadium appears to have a negligible effect, while iron has a greater strengthening effect than Aluminum ($\sim$75 MPa/wt%). Based on this model, the increases in strength afforded by changes in composition are more significant than the changes in strength offered by modifications to the microstructure; the use of neural networks has identified solid-solution

Fig. 25 Example of a virtual experiment showing the effect that the thickness of the α-lath has on yield strength (YS)

LIVE GRAPH
Click here to view

LIVE GRAPH
Click here to view

Fig. 26 Example of a virtual experiment showing the effect that the colony scale factor has on yield strength (YS)

LIVE GRAPH
Click here to view

LIVE GRAPH
Click here to view

Fig. 27 Example of predicted versus experimental data for yield strength (YS) in α+β-processed Ti-6Al-4V

strengthening as an influential strengthening mechanism in the alloy Ti-6Al-4V.

REFERENCES

1. S. Ankem, G.K. Scarr, I.L. Caplan, J.C. Williams, S.R. Seagle, and H.B. Bomberger, *Multiple Regression Analysis of the Effects of Various Alloying Elements on the Properties of Titanium Alloys,* Sixth World Conference on Titanium (France), Les Editions de Physique, 1988

2. S.C. Weakley-Bollin, W. Donlon, C. Wolverton, J.W. Jones, and J.E. Allison, Modeling the Age-Hardening Behavior of Al-Si-Cu Alloys, *Metall. Mater. Trans. A,* Vol 35, 2004, p 2407

3. I.P. Moreno, T.K. Nandy, J.W. Jones, J.E. Allison, and T.M. Pollock, Microstructural Stability and Creep of Rare-Earth Containing Magnesium Alloys, *Scr. Mater.,* Vol 48, 2003, p 1029–1034

4. M. Marya, L.G. Hector, R. Verma, and W. Tong, Microstructural Effects of AZ31 Magnesium Alloy on Its Tensile Deformation and Failure Behaviors, *Mater. Sci. Eng. A,* Vol 418, 2006, p 341–356

5. N.F. Mott and F.R.N. Nabarro, *Report of a Conference on the Strength of Solids,* The Physical Society, London, 1948, p l

6. R.L. Fleischer, *Acta Metall.,* Vol 11, 1963, p 203

7. R.L. Fleischer and W.R. Hibbard, *Conference on Relation of Structure to Mechanical Properties of Metals,* National Physical Laboratories, Teddington, 1964, p 262

8. R. Labusch, G. Grange, J. Ahearn, and P. Haasen, Rate Processes in Plastic Deformation of Materials, *Proceedings from the John E. Dorn Symposium,* J.C.M. Li and A.K. Muknergee, Ed., *ASM Handbook,* American Society for Metals, 1975, p 26

9. L.A. Gypen and A. Deruyttere, Multi-Component Solid Solution Hardening, Part 2: Agreement with Experimental Results, *J. Mater. Sci.,* Vol 12, 1977, p 1034–1038

10. M.Z. Butt and P. Feltham, Review Solid-Solution Hardening, *J. Mater. Sci.,* Vol 28, 1993, p 2557–2576

11. D. Hull and D.J. Bacon, *Introduction to Dislocations,* Pergamon Press, Oxford, 1984

12. B.Q. Han and E.J. Lavernia, Deformation Mechanisms of Nanostructured Al Alloys, *Adv. Eng. Mater.,* Vol 7 (No. 6), 2005, p 457

13. C.J. Bettles and M.A. Gibson, Microstructural Design for Enhanced Elevated Temperature Properties in Sand-Castable Magnesium Alloys, *Adv. Eng. Mater.,* Vol 5 (No. 12), 2003, p 859

14. M.Z. Butt, M. Zubair, and I. Ul-Haq, A Comparative Study of the Stress
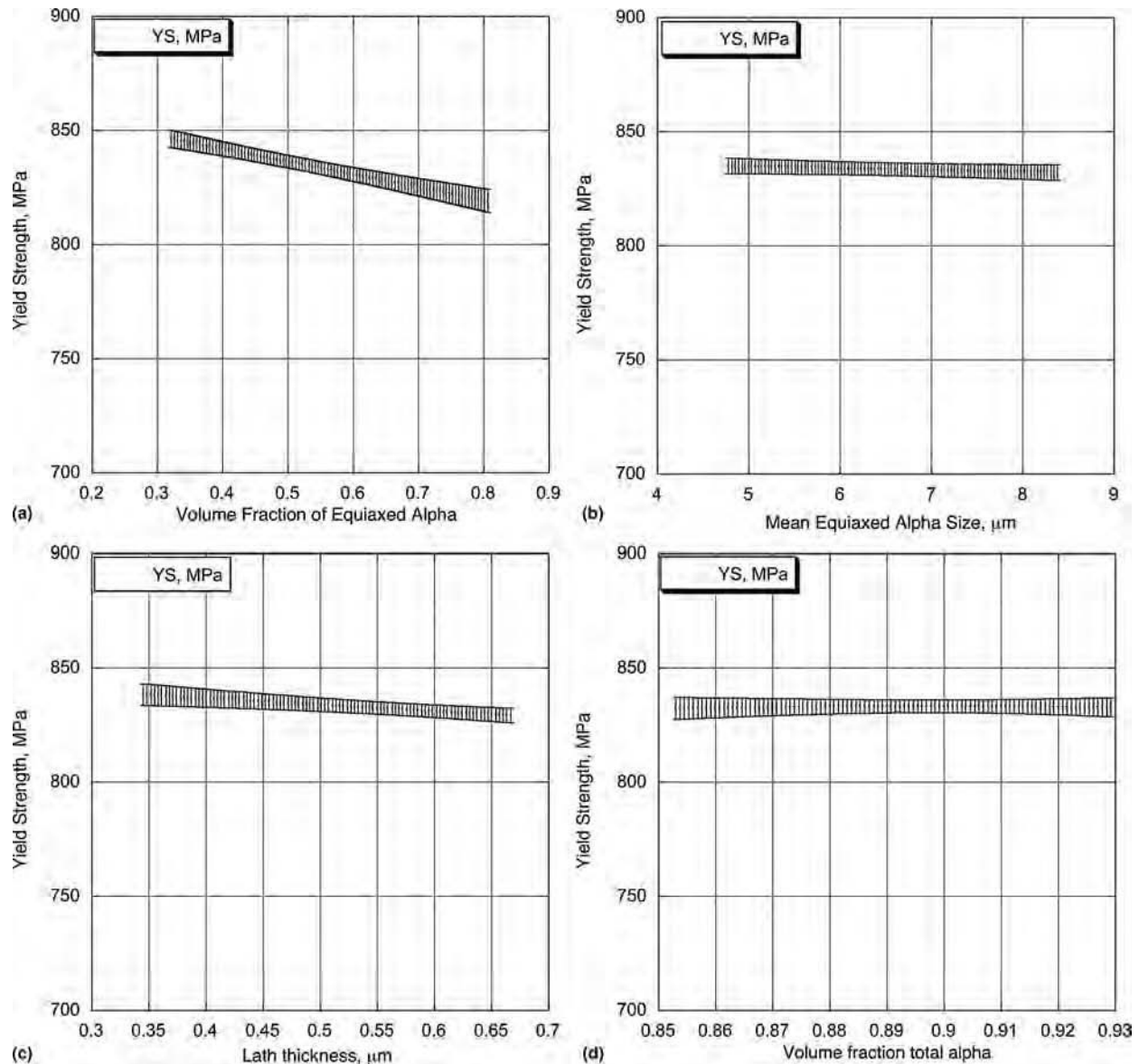
Fig. 28 Virtual experiments showing the isolated effects of the individual microstructural parameters in α+β-processed Ti-6Al-4V

Relaxation in Aged and Un-Aged High-Purity Aluminium Polycrystals, *J. Mater. Sci.,* Vol 35, 2000, p 139–144

15. M. Arshad Choudhry and M. Ashraf, Effect of Heat Treatment and Stress Relaxation in 7075 Aluminum Alloy, *J. Alloy. Compd.,* Vol 437, 2007, p 113–116

16. C. Cloutier, J.W. Jones, and J.E. Allison, "The Effect of Copper Level and Solidification Rate on the Aging Behavior of a 319-Type Cast Aluminum Alloy," SAE Technical Paper Series, 2000-01-0759

17. M. Bamberger and G. Dehm, Trends in the Development of New Mg Alloys, *Ann. Rev. Mater. Res.,* Vol 38, 2008, p 505–533

18. C.J. Bettles, M.A. Gibson, and K. Venkatesan, Enhanced Age-Hardening Behaviour in Mg-4 wt.% Zn Microalloyed with Ca, *Scr. Mater.,* Vol 51, 2004, p 193–197

19. R.E. Smallman and R.J. Bishop, *Modern Physical Metallurgy and Materials Engineering,* 6th ed., Butterworth-Heinemann, Elsevier Science, 1999

20. G. Gottstein, *Physical Foundations of Materials Science,* Springer-Verlag, Berlin Heidelberg, 2004

21. J.F. Nie and B.C. Muddle, *J. Phase Equilibria,* Vol 19, 1998, p 543–551

22. A.W. Zhu and E.A.J. Starke, *Acta Mater.,* Vol 47, 1999, p 3263–3269

23. A.W. Zhu, A.W. Chen, and E.A.J. Starke, Jr., *Acta Mater.,* Vol 48, 2000, p 2239–2246

24. V. Mohles, D. Rönnpagel, and E. Nembach, Simulation of Dislocation Glide in Precipitation Hardened Materials, *Comput. Mater. Sci.,* Vol 16, 1999, p 144–150

25. E. Rincón, H.F. López, M.M. Cisneros, H. Mancha, and M.A. Cisneros, Effect of Temperature on the Tensile Properties of an As-Cast Aluminum Alloy A319, *Mater. Sci. Eng. A,* Vol 452–453, 2007, p 682–687

26. S.I. Rao, T.A. Parthasarathy, D.M. Dimiduk, and P.M. Hazzledine, Discrete Dislocation Simulations of Precipitation Hardening in Superalloys, *Philos. Mag.,* Vol 84 (No. 30), 2004, p 3195–3215

27. B. Reppich, in *Materials Science and Technology, A Comprehensive Treatment,* Vol 6, H. Mughrabi, R.W. Cahn, P. Hassen, E. J. Kramer, Ed., Weinheim VCH, p 113

28. A.M. Redsten, E.M. Klier, A.M. Brown, and D.C. Dunand, Mechanical Properties and Microstructure of Cast Oxide-Dispersion-Strengthened Alumina, *Mater. Sci. Eng. A,* Vol 201, 1995, p 88–102
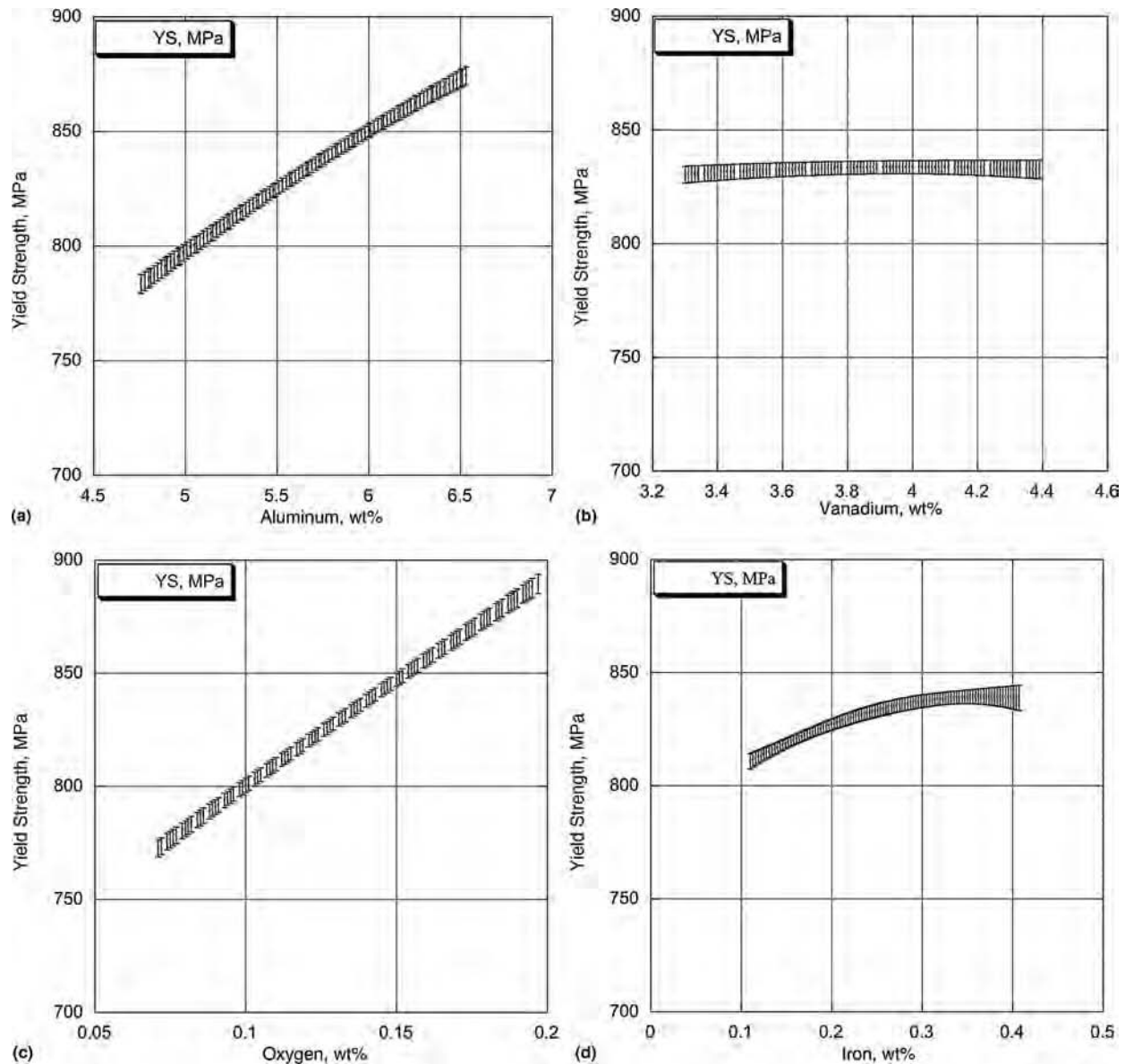
Fig. 29 Virtual experiments showing the isolated effects of the individual compositional parameters in α+β-processed Ti-6Al-4V

LIVE GRAPH
Click here to view

29. M.F. Ashby, *Philos. Mag.,* Vol 21, 1970, p 399
30. E.O. Hall, *Proc. R. Soc. B,* Vol 64, 1951, p 747
31. N.J. Petch, *J. Iron Steel Inst.,* Vol 174, 1953, p 25
32. E.O. Hall, *Yield Point Phenomena in Metals and Alloys,* Plenum Press, New York, NY, 1970
33. J. Hirth and J. Lothe, *Theory of Dislocations,* John Wiley & Sons, p 776
34. N. Hansen, *Acta Metall.,* Vol 25, 1977, p 863
35. A.W. Thompson and M.I. Baskes, *Philos. Mag.,* Vol 28, 1973, p 301
36. D.J. Lloyd, *Met. Sci.,* Vol 14, 1980, p 193
37. J.T. Al-Haidary, N.J. Petch, and E.R. De Los Rios, *Philos. Mag. A,* Vol 47, 1983, p 869

38. A.W. Thompson, *Acta Metall.,* Vol 23, 1975, p 1337
39. A.W. Thompson and W.A. Backofen, *Metall. Trans.,* Vol 2, 1971, p 2004
40. N. Hansen and B. Ralph, *Acta Metall.,* Vol 30, 1982, p 41
41. F.E. Hauser, P.R. Landon, and J.E. Dorn, *Trans. AIME,* Vol 206, 1956, p 589
42. R.W. Armstrong, The Yield and Flow Stress Dependence on Polycrystal Grain Size, *Yield, Flow and Fracture of Polycrystals,* T.N. Baker, Ed., London, 1983, p 1–31
43. G. Nussbaum, P. Sainfort, G. Regazzoni, and H. Gjestland, *Scr. Metall.,* Vol 23, 1989, p 1079
44. M. Mabuchi, Y. Yamada, K. Shimojima, C.E. Wen, Y. Chino, M. Nakamura, T.

Asahina, H. Iwasaki, T. Aizawa, and K. Higashi, The Grain Size Dependence of Strength in the Extruded AZ91 Mg Alloy, *Magnesium Alloys and Their Applications,* K.U. Kainer, Ed., Munich, Wiley-VCH, New York, 2000, p 280–284
45. W.J. Kim, C.W. An, Y.S. Kim, and S.I. Hong, *Scr. Mater.,* Vol 47, 2002, p 39
46. N. Ono, K. Nakamura, and S. Miura, *Mater. Sci. Forum,* Vol 419–422, 2003, p 195
47. H. Takahashi, Y. Oishi, K. Wakamatsu, and N. Kawabe, *Mater. Sci. Forum,* Vol 419–422, 2003, p 345
48. D.V. Wilson, *J. Inst. Met.,* Vol 98, 1966, p 133
49. P. Andersson, C.H. Cáceres, and J. Koike, *Mater. Sci. Forum,* Vol 419–422, 2003, p 123

50. D.V. Wilson, *J. Inst. Met.,* Vol 94, 1966, p 84

51. G. Mann, J.R. Griffiths, and C.H. Cáceres, Hall-Petch Parameters in Tension and Compression in Cast Mg-2Zn Alloys, *J. Alloy. Compd.,* Vol 378, 2004, p 188–191

52. B.Q. Han and E.J. Lavernia, Deformation Mechanisms of Nanostructured Al Alloys, *Adv. Eng. Mater.,* Vol 7 (No. 6), 2005, p 457

53. R. Mythili, S. Saroja, and M. Vijaya-lakshmi, Study of Mechanical Behavior and Deformation Mechanism in an α-β Ti-4.4Ta-1.9Nb Alloy, *Mater. Sci. Eng. A,* Vol 454–455, 2007, p 43–51

54. D.J. Lloyd and S.A. Court, Influence of Grain Size on Tensile Properties of Al-Mg Alloys, *Mater. Sci. Technol.,* Vol 19, Oct 2003, p 1349

55. N. Hansen, Hall-Petch Relation and Boundary Strengthening, *Scr. Mater.,* Vol 51, 2004, p 801–806

56. X. Feaugas and H. Haddou, Effects of Grain Size on Dislocation Organization and Internal Stresses Developed under Tensile Loading in FCC Metals, *Philos. Mag.,* Vol 87 (No. 7), March 1, 2007, p 989–1018

57. G. Krauss, *Steels: Heat Treatment and Processing Principles,* ASM International, 1990

58. T.E. Swarr and G. Krauss, The Effect of Structure on the Deformation of As-Quenched and Tempered Martensite in an Fe-0.2%C Alloy, *Metall. Trans. A,* Vol 7, 1976, p 41–48

59. A.R. Marder and G. Krauss, The Effect of Morphology on the Strength of Lath Mar-tensite, *Proceedings of the Second International Conference on the Strength of Metals and Alloys,* Vol III, American Society for Metals, 1970, p 822

60. M.J. Roberts, Effect of Transformation Substructure on the Strength and Tough-ness of Fe-Mn Alloys, *Metall. Trans.,* Vol 1, 1970, p 3287–3294

61. S. Kar, T. Searles, E. Lee, G.B. Viswa-nathan, J. Tiley, R. Banerjee, and H.L. Fraser, Modeling the Tensile Properties in β-Processed α/β Ti Alloys, *Metall. Trans. A,* Vol 37, 2006, p 559

62. F. Appel and R. Wagner, Microstructure and Deformation of Two-Phase γ-Tita-nium Aluminides, *Mater. Sci. Eng.,* Vol R22, 1998, p 187–268

63. Y.-W. Kim and D.M. Dimiduk, *JOM,* Vol 43, 1991, p 40

64. Y.-W. Kim, *JOM,* Vol 46, 1994, p 30

65. K. Wurzwallner, H. Clemens, P. Schretter, A. Bartels, and C. Koeppe, High-Tempera-ture Ordered Intermetallic Alloys V, *Mater. Res. Soc. Symp. Proc.,* Vol 288, I. Baker, R. Darolia, J.D. Whittenberger, and M.H. Yoo, Ed. (Pittsburgh, PA), 1993, p 867

66. I.J. Beyerlein and C.N. Tomé, A Disloca-tion-Based Constitutive Law for Pure Zr Including Temperature Effects, *Int. J. Plast.,* Vol 24, 2008, p 867–895

67. G.I. Taylor, Mechanism of Plastic Defor-mation of Crystals, *Proc. Roy. Soc. A,* Vol 145, 1934, p 362

68. F.A. McClintock and A.S. Argon, *Mechanical Behavior of Materials,* Ceramic Book and Literature Service, India, 1999

69. D. Hull and D.J. Bacon, *Introduction to Dislocations,* 3rd ed., Pergamon Press, New York, 1989

70. G. de With, *Structure, Deformation, and Integrity of Material,* Wiley-VCH, 2006

71. A.E.H. Love, *Mathematical Theory of Elasticity*, CUP

72. G.I. Taylor, *J. Inst. Met.,* Vol 62, 1938, p 307

73. H.K.D.H. Bhadeshia, "Neural Networks in Materials Science," University of Cam-bridge, http://www.msm.cam.ac.uk/phase-trans/abstracts/neural.review.html

74. H.K.D.H. Bhadeshia, *ISIJ Int.,* Vol 39, 1999, p 966–979

75. H.K.D.H. Bhadeshia, R.C. Dimitriu, S. Forsik, J.H. Pak, and J.H. Ryu, Perfor-mance of Neural Networks in Materials Science, *Mater. Sci. Technol.,* Vol 25, 2009, p 504–510

76. S. Kar, T. Searles, E. Lee, G.B. Viswa-nathan, J. Tiley, R. Banerjee, and H.L. Fraser, *Metall. Mater. Trans. A,* Vol 37 (No. 3), 2006, p 559–566

77. J.S. Tiley, "Modeling of Microstructure Property Relationships in Ti-6Al-4V," The Ohio State University, Columbus, OH, 2002

78. P.C. Collins, S. Kar, T. Searles, S. Koduri, G.B. Viswanathan, J. Tiley, R. Banerjee, and H.L. Fraser, *FDM Proceedings*

79. P.C. Collins, S. Connors, R. Banerjee, and H.L. Fraser, A Combinatorial Approach to the Development of Neural Networks for the Prediction of Composition-Micro-structure-Property Relationships in α/β Ti Alloys, *Proceedings of Ti-2003 Sci-ence and Technology,* Vol 3, 2003, p 1389–1396

80. S. Ogata, J. Li, and S. Yip, Ideal Pure Shear Strength of Aluminum and Copper, *Sci.,* Vol 298, 2002, p 807–811

81. V.V. Bulatov, W. Cai, J. Fier, M. Hira-tani, G. Hommes, T. Pierce, M. Tang, M. Rhee, K. Yates, and T. Arsenlis, Para-DiS, 2004

82. S. Ogata, J. Li, N. Hirosaki, Y. Shibutani, and S. Yip, Ideal Shear Strain of Metals and Ceramics, *Phys. Rev. B,* Vol 70, 2004, p 104104

83. S. Ogata, J. Li, and S. Yip, Twinning Pathway in BCC Molybdenum, *Europhys. Lett.,* Vol 68, 2004, p 405–411

84. Y.L. Chiu and A.H.W. Ngan, A TEM Investigation on Indentation Plastic Zones in Ni₃Al(Cr,B) Single Crystals, *Acta Mater.,* Vol 50, 2002, p 2677–2691

85. Y. Choi, K.J. Van Vliet, J. Li, and S. Sur-esh, Size Effects on the Onset of Plastic Deformation during Nanoindentation of Thin Films and Patterned Lines, *J. Appl. Phys.,* Vol 94, 2003, p 6050–6058

86. H.B. Wang, M. Dembo, S.K. Hanks, and Y.L. Wang, Focal Adhesion Kinase is Involved in Mechanosensing during Fibroblast Migration, *Proc. Natl. Acad. Sci. USA,* Vol 98, 2001, p 11295–11300

87. C. Shen and Y. Wang, Phase Field Model of Dislocation Networks, *Acta Mater.,* Vol 51, 2003, p 2595–2610

88. C. Shen and Y. Wang, Incorporation of Gamma-Surface to Phase Field Model of Dislocations: Simulating Dislocation Dis-sociation in FCC Crystals, *Acta Mater.,* Vol 52, 2004, p 683–691

89. J. Li, T. Zhu, S. Yip, K.J. Van Vliet, and S. Suresh, Elastic Criterion for Disloca-tion Nucleation, *Mater. Sci. Eng. A—Struct. Mater. Prop. Microstruct. Pro-cess.,* Vol 365, 2004, p 25–30

90. J. Li, Transformation Strain by Chemical Disordering in Silicon Carbide, *J. Appl. Phys.,* Vol 95, 2004, p 6466–6469

91. X.F. Qian, J. Li, X. Lin, and S. Yip, Time-Dependent Density Functional Theory with Ultrasoft Pseudopotentials: Real-Time Electron Propagation Across a Molecular Junction, *Phys. Rev. B,* Vol 73, 2006, p 035408

92. X. Lin, J. Li, C.J. Forst, and S. Yip, Mul-tiple Self-Localized Electronic States in Trans-Polyacetylene, *Proc. Natl. Acad. Sci. USA,* Vol 103, 2006, p 8943–8946

93. T. Zhu, J. Li, K.J. Van Vliet, S. Ogata, S. Yip, and S. Suresh, Predictive Modeling of Nanoindentation-Induced Homogeneous Dislocation Nucleation in Copper, *J. Mech. Phys. Solids,* Vol 52, 2004, p 691–724

94. J. Li, K.J. Van Vliet, T. Zhu, S. Yip, and S. Suresh, Atomistic Mechanisms Gov-erning Elastic Limit and Incipient Plastic-ity in Crystals, *Nature,* Vol 418, 2002, p 307–310

95. T. Zhu, J. Li, and S. Yip, Atomistic Study of Dislocation Loop Emission from a Crack Tip, *Phys. Rev. Lett.,* Vol 93, 2004, p 025503

96. J.R. Rice, and R. Thomson, Ductile Versus Brittle Behavior of Crystals, *Philos. Mag.,* 1974;29:73

97. J.R. Rice, Dislocation Nucleation from a Crack Tip – An Analysis Based on the Peierls Concept, *J. Mech. Phys. Solids*, 1992;40:239

98. J.R. Rice and G.E. Beltz, The Activation-Energy for Dislocation Nucleation at a Crack, *J. Mech. Phys. Solids*, 1994; 42:333

99. C. Shen and Y. Wang, Phase Field Model of Dislocation Networks, *Acta Mater.,* Vol 51, 2003, p 2595–2610

100. Y.U. Wang, Y.M. Jin, A.M. Cuitiño, and A.G. Khachaturyan, Application of Phase Field Microelasticity Theory of Phase Transformations to Dislocation Dynam-ics: Model and Three-Dimensional

Simulations in a Single Crystal, *Philos. Mag. Lett.,* Vol 81 (No. 6), 2001, p 385–393

101. N. Zhou, C. Shen, M.J. Mills, and Y. Wang, Phase Field Modeling of Channel Dislocation Activity and γ¢ Rafting in Single Crystal Ni-Al, *Acta Mater.,* Vol 55, 2007, p 5369–5381

102. M. Groeber, B. Haley, M. Uchic, and S. Ghosh, Microstructural Characterization Using 3-D Orientation Data Collected by an Automated FIB-EBSD System, *Numiform 8 Proceedings,* 2004

103. M. Groeber, B. Haley, M. Uchic, D. Dimiduk, and S. Ghosh, Towards 3D Reconstruction and Characterization of Polycrystalline Microstructures Using a FIB-SEM System, *Mater. Charact.,* 2006

104. M. Uchic, M. Groeber, D. Dimiduk, and J. Simmons, 3D Microstructural Characterization of Nickel Superalloys via Serial-Sectioning Using a Dual Beam FIB-SEM, *Scr. Mater.,* 2006

105. M. Groeber, Y. Bhandari, M. Uchic, D. Dimiduk, and S. Ghosh, An Automated Framework for Microstructure Characterization and Representation, *MMM 2006 Proceedings,* 2006

106. M. Groeber, M. Uchic, D. Dimiduk, and S. Ghosh, Development of a Robust 3D Characterization-Representation Framework for Modeling Polycrystalline Materials, *JOM,* 2007

107. Y. Bhandari, S. Sarkar, M. Groeber, M. Uchic, D. Dimiduk, and S. Ghosh, 3D Polycrystalline Microstructure Reconstruction from FIB Generated Serial Sections for FE Analysis, *Comput. Mater. Sci.,* Vol 41, 2007, p 222–235

108. S.P. Ringer and K. Hono, Microstructural Evolution and Age Hardening in Aluminium Alloys: Atom Probe Field-Ion Microscopy and Transmission Electron Microscopy Studies, *Mater. Charact.,* Vol 44 (No. 101–131), 2000

109. L. Kovarik, M.K. Miller, S.A. Court, and M.J. Mills, Origin of the Modified Orientation Relationship for S(S00)-Phase in Al-Mg-Cu Alloys, *Acta Mater.,* Vol 54, 2006, p 1731–1740

110. L. Kovarik, S.A. Court, H.L. Fraser, and M.J. Mills, GPB Zones and Composite GPB/GPBII Zones in Al-Cu-Mg Alloys, *Acta Mater.,* Vol 56 (No. 17), Oct 2008, p 4804–4815

111. S.P. Ringer, K. Hono, and T. Sakurai, The Effect of Trace Additions of Sn on Precipitation in Al-Cu Alloys: An Atom Probe Field Ion Microscopy Study, *Metall. Mater. Trans. A,* Vol 26, 1995, p 2207–2217

112. G.B. Winkelman, K. Raviprasad, and B.C. Muddle, Orientation Relationships and Lattice Matching for the S Phase in Al-Cu-Mg Alloys, *Acta Mater.,* Vol 55, 2007, p 3213–3228

113. J.F. Nie and B.C. Muddle, Strengthening of an Al-Cu-Sn Alloy by Deformation-Resistant Precipitate Plates, *Acta Mater.,* Vol 56, 2008, p 3490–3501

114. J.C. Williams and M.J. Blackburn, *Trans. Q.,* Vol 60 (No. 3), 1967, p 373–383

115. J.C. Williams and E.A. Starke, The Role of Thermomechanical Processing in Tailoring the Properties of Aluminum and Titanium Alloys*, ASM Metals/Materials Technology Series,* 1982

116. E.W. Collings, The Physical Metallurgy of Titanium Alloys, *ASM Series in Metal Processing,* H.L. Gegel, Ed., American Society for Metals, 1984

117. G. Lütjering, *Proceedings, Ninth World Titanium Conference* (China), 1998, p 1–19

118. R. Boyer, G. Welsch, and E.W. Collings, *Materials Properties Handbook: Titanium Alloys,* ASM International, 1994

119. P.A. Blenkinshop, W.J. Evans, and H.M. Flower, Ed., *Titanium '95: Science and Technology, Proceedings of the Eighth World Conference on Titanium,* The Institute of Materials, London, U.K., 1996

120. R.R. Boyer and D.R. Wallem, *Microstructure/Property Relationships of Titanium Alloys,* 1994, p 125–132

121. C.G. Rhodes, J.C. Chesnutt, and J.A. Wert, *Microstructure, Fracture Toughness and Fatigue Crack Growth Rate in Titanium Alloys,* A.K. Chakrabarti and J. C. Chesnutt, Ed., TMS, 1987, p 39–54

122. G. Lütjering, *Mater. Sci. Eng. A,* Vol 263, 1999, p 117–126

123. P.C. Collins, S. Koduri, B. Welk, J. Tiley, and H.L. Fraser, On the Influence of Alloy Content on the Tensile Properties in α-β Processed α/β Ti Alloys, in preparation

124. J. Williams and G. Lütjering, *Titanium 1980 Science and Technology,* AIME, Warrendale, PA, 1981, p 671–681

125. W.H. Miller, Jr., R.T. Chen, and E.A. Starke, Jr., *Metall. Trans. A,* Vol 18, 1987, p 1451–1468

126. S. Suri, G.B. Viswanathan, T. Neeraj, D. H. Hou, and M.J. Mills, *Acta Mater.,* Vol 47, 1999, p 1019–1034

127. J. Russ and R. Dehoff, *Practical Stereology,* Kluwer Academic Publishers/Plenum Publishers, Dordrecht/New York, 2000

128. H.J.G. Gundersen, E.B. Jenson, and R. Osterby, Distribution of Membrane Thickness Determined by Lineal Analysis, *J. Microsc.,* Vol 113 (No. 27), 1978

Modeling of Creep

Sammy Tin, Illinois Institute of Technology

FOR MANY ENGINEERING MATERI-ALS, deformation is highly sensitive to both temperature and strain rate. Under these conditions, deformation is considered to be thermally activated, that is, assisted by thermal energy. In order to develop an understanding of the underlying mechanisms governing deformation at elevated temperatures, the phenomenological effects resulting from temperature-induced thermodynamic and kinetic changes are discussed.

Fundamentals of Deformation

Ideally, the deformation behavior of all engineering materials can be described using expressions known as constitutive equations, which relate the dependence of stress, temperature, and microstructure on deformation. Constitutive equations have the general functional form:

$$\dot{\gamma} = f(\sigma, T, S_i) \qquad \text{(Eq 1)}$$

where the strain rate, $\dot{\gamma}$, can be defined as a function of stress, σ; temperature, T; and a structure parameter, S_i. This structure parameter is a complex variable that may be comprised of a number of dynamic microstructural variables, such as material composition, grain size, dislocation density, precipitate size, and morphology, that evolve during the course of deformation. Thus, the time-dependent behavior of the structure parameter is also dependent on the stress, temperature, and the current structure:

$$\frac{dS_i}{dt} = g(\sigma, T, S_i) \qquad \text{(Eq 2)}$$

For simplicity, the structure variable is often designated as being a fixed parameter that does not change over a range of imposed deformation conditions. The assumption of a constant structure variable is valid over some limited ranges of deformation conditions and can be used to describe the deformation behavior of a variety of different engineering materials fairly accurately. However, in reality, it is difficult to maintain a constant structure parameter over a

wide range of temperature and deformation in practical engineering materials. Due to the fact that structure variables are highly sensitive to temperature and stress, the goal of formulating a coupled set of constitutive equations that are capable of describing deformation in all materials remains a major challenge, particularly during deformation at elevated temperatures.

At temperatures approaching absolute zero, deformation occurs at an extremely low rate until a critical value of σ_s/μ is reached. Above this critical stress, dislocation glide mechanisms are activated, and dislocations are able to glide freely until they encounter an obstacle or exit the crystal. The critical value of σ_s/μ corresponds to a threshold stress, τ_{cri}, that is required for dislocations to overcome the shear resistance of the crystal lattice and initiate plastic flow. At stresses approaching the threshold stress, thermal fluctuations in the crystal lattice may serve to enhance the ease by which plastic flow may occur as well as dislocation activity. Thermally activated deformation can be expressed in the form:

$$\dot{\gamma}_i = \dot{\gamma}_o \exp\left(\frac{-G(\sigma, S_i)}{kT}\right) \qquad \text{(Eq 3)}$$

where k is the Boltzmann constant, $\dot{\gamma}_o$ is a material constant, and G is the Gibbs free energy for activation of dislocation glide. Depending on the dominant mode of deformation, the exact form of this relationship may change, because changes in temperature and stress may alter the manner by which dislocations and defects contribute to plastic flow.

For many materials, the pronounced effect of temperature on deformation can be attributed to different plasticity mechanisms operating at low and high temperatures. At low temperatures, T, relative to the melting temperature (T_m), where $T/T_m < 0.4$, plasticity tends to be glide limited or obstacle limited. Glide limited refers to the ability of individual dislocations to become active and move against the shear resistance of the lattice. Once active, these mobile, gliding dislocations may encounter obstacles such as solute atoms, grain discontinuities, other dislocations, or precipitates that limit their movement. Obstacles affect the movement of gliding dislocations as they influence the stress state of the crystal through which they are

moving. There are two types of stresses or interactions that need to be considered when evaluating the effect of obstacles on dislocation movement: long range, which are slowly varying stresses due to other dislocations or groups of dislocations; and short-range interactions resulting from localized changes in the resistance of the crystal lattice to shearing and dragging of jogs. At these low temperatures, thermal activation primarily aids the dislocations in penetrating these short-range barriers. The application of a shear stress, τ_a, sufficiently high in magnitude to move a dislocation forward by a small increment, dx, results in a measure of work, dW, being performed on the crystal:

$$dW = \tau_a \mathbf{b} l dx \qquad \text{(Eq 4)}$$

where $\mathbf{b}$ is the Burgers vector, and ldx is the area on the slip plane being swept out by the mobile dislocation segment.

For small plastic strains, the relationship between the glide resistance, τ, and the shear strain, τ_a, and Helmholtz free energy, F, is defined as (Ref 1):

$$\tau = \frac{1}{V}\frac{\partial \Delta F}{\partial \gamma} \qquad \text{(Eq 5)}$$

As a result, the activation free energy that must be supplied to activate the dislocation segment over the obstacle or barrier is:

$$\Delta G = \int_{\tau_a}^{\tau_{line}} (\mathbf{b}\Delta A) d\tau_{line} \qquad \text{(Eq 6)}$$

where τ_{line} is the resistance to the dislocation line segment opposing its motion. The quantity of $\mathbf{b}\Delta A$ is commonly referred to as the activation volume, which is the product of the area on the slip plane and its corresponding Burgers vector. With thermal activation, the probability of the thermal fluctuation providing sufficient energy that enables the dislocation segment to overcome the obstacle is proportional to $\exp(-\Delta G/kT)$. The resulting strain rate, $\dot{\gamma}$, can thereby be expressed as:

$$\dot{\gamma} = \dot{\gamma}_o \exp\frac{-\Delta G}{kT} \qquad \text{(Eq 7)}$$

Combining this expression with that of Eq 6 allows one to express the activation volume as:

$$b\Delta A = \frac{\partial \Delta G}{\partial \tau} = kT\frac{\partial \ln \dot{\gamma}}{\partial \tau} \qquad \text{(Eq 8)}$$

This expression suggests that variations in strain rate can reveal whether the rate-limiting obstacles are long or short range in nature.

Creep Characteristics

At high temperatures, $T > 0.4T_m$, time-dependent plasticity may occur at stresses well below the flow stress of the material, also referred to as creep deformation. Creep deformation may occur via a number of mechanisms, but all of them include an element of thermal activation that allows for plastic deformation to occur at stresses below the yield or flow stress. Under a constant stress or load at elevated temperature, the initial response of the material will be to undergo an initial reversible elastic deformation. Following this, the material will gradually extend plastically, which would be considered to be permanent or irreversible deformation. With respect to engineering applications, the main concern regarding creep deformation is the loss in dimensional stability, although prolonged creep may also lead to rupture.

Stages. The typical response of a material to a constant stress at elevated temperature can be expressed in the form of a creep curve (Fig. 1). Creep deformation is commonly classified as being comprised of three distinct stages:

- *I, or primary creep:* Following an instantaneous elastic and plastic strain upon loading, this stage is highly transient because the strain rate decreases rapidly due to the development of dislocation structures within the microstructure. The strain rate decreases until a minimum is reached.
- *II, or secondary creep:* During this stage of creep, the strain rate is at a minimum, and this is often attributed to the establishment of equilibrium between deformation and recovery mechanisms. The minimum or steady-state creep rate, $\dot{\varepsilon}_{min}$ or $\dot{\varepsilon}_{ss}$, respectively, is maintained for an extended period of time during secondary creep.
- *III, or tertiary creep:* Toward the conclusion of secondary creep, the strain rate begins to accelerate as a result of internal damage accumulation within the material and microstructure. Damage accumulation reduces the effective load-bearing cross-sectional area and may lead to catastrophic failure of the material.

Modeling Steady-State Range. Increasing the stress and temperature during creep deformation will tend to extend the primary and tertiary creep regimes and reduce the duration of steady-state, secondary creep. Since tertiary creep often results in significant dimensional nonconformance and localized deformation, creep-limited components are often removed
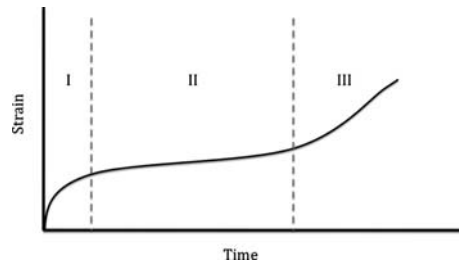
Fig. 1 Schematic of a creep curve exhibiting the three regimes: I, primary creep; II, secondary or steady-state creep; and III, tertiary creep

from service near the end of secondary creep, before the transition into tertiary. Thus, it is desirable to model the magnitude of the creep strain into the steady-state range (Ref 2):

$$\varepsilon = \varepsilon_{elastic} + \varepsilon_{in} + at^{1/3} + \dot{\varepsilon}_{SS}t \qquad \text{(Eq 9)}$$

where the total creep strain, ε, is the sum of the elastic, $\varepsilon_{elastic}$; instantaneous, ε_{in}; primary $at^{1/3}$; and steady-state creep, $\dot{\varepsilon}_{SS}t$, strains. Deformation occurring within the material during primary creep can be quite complex, particularly during the initial transient, where the dislocation density in the material increases rapidly. Thermal activation enhances dislocation mobility that allows for rapid multiplication and hardening to occur within the microstructure. As primary creep progresses, these dislocations organize themselves into low-energy subgrain structures that are retained through the secondary creep regime. The walls of the subgrains are comprised of high densities of dislocations, while the cell interiors are relatively free of dislocations (Ref 3). This process is shown schematically for pure iron being crept at 600 °C in Fig. 2. The change in dislocation density as a function of creep strain for the subgrain walls and cell interiors is shown in Fig. 3. These changes in the strain rate and accompanying changes in subgrain structure are often described by the Bailey-Orowan equation (Ref 4, 5). This expression accounts for both the hardening and recovery processes that occur during primary creep, before the attainment of steady state. If the flow stress is expressed as a function of time and strain:

$$\sigma = \sigma(\varepsilon, t) \qquad \text{(Eq 10)}$$

$$d\sigma = \left(\frac{d\sigma}{d\varepsilon}\right)_t d\varepsilon + \left(\frac{d\sigma}{dt}\right)_\varepsilon dt \qquad \text{(Eq 11)}$$

Since creep tests are conducted at constant stress, the change in stress, $d\sigma = 0$:

$$\left(\frac{d\sigma}{d\varepsilon}\right)_t d\varepsilon = -\left(\frac{d\sigma}{dt}\right)_\varepsilon dt \qquad \text{(Eq 12)}$$

$$\dot{\varepsilon} = \frac{d\varepsilon}{dt} = -\left(\frac{d\sigma}{dt}\right)_e \bigg/ \left(\frac{d\sigma}{d\varepsilon}\right)_t = \frac{r}{h} \qquad \text{(Eq 13)}$$

where r is the recovery function, and h is the hardening function. Recovery or softening
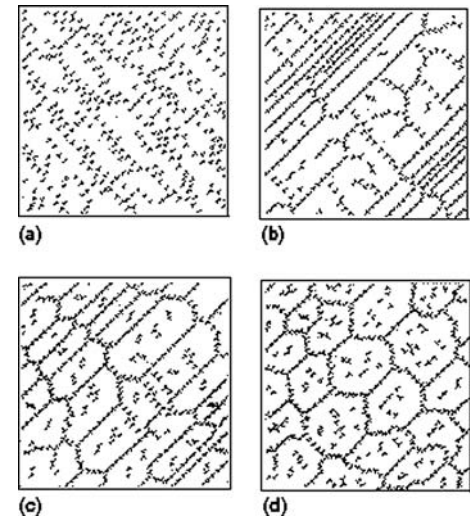
Fig. 2 Schematic showing the rearrangement of dislocations into low-energy configurations during the initial stages of creep. Source: Ref 3
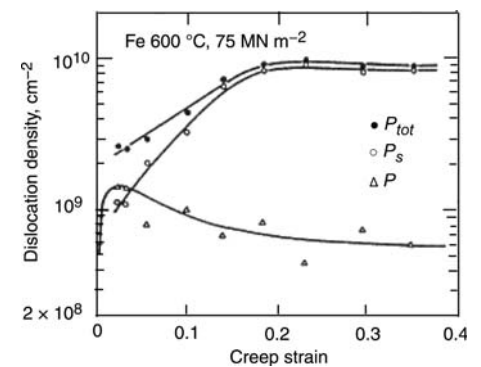
Fig. 3 Plot showing the overall changes in dislocation density, ρ_{tot}, as a function of creep strain for pure iron at 600 °C. Dislocation densities are different within the cell interiors, ρ_s, and along the cell walls, ρ. Source: Ref 3

during creep occurs as the overall dislocation density is being reduced due to dislocation annihilation or rearrangement of the dislocations from the cell interiors into low-energy configurations within the subgrain walls. The hardening function is a measure of the increase in flow stress as straining occurs. The relationship between the dislocation density during steady-state creep and the applied stress is:

$$\tau = \alpha\mu b\sqrt{\rho} \qquad \text{(Eq 14)}$$

where τ is the shear stress, α is a material constant, μ is the shear modulus, and ρ is the dislocation density. This equation suggests that the shear stress increases as the dislocation density increases. During steady-state creep, however, there is a balance between the mechanisms resulting in dislocation multiplication or hardening and the mechanisms resulting in dislocation recovery and softening. A number of theories have been established over the years to model

creep deformation based on the evolution of dislocation substructures using the rates of hardening and recovery, with limited degrees of success (Ref 6, 7). The major difficulty associated with this approach is the ability to quantitatively model inhomogeneous dislocation structures as they develop. The formation of subgrains and low-energy dislocation configurations is an extremely complex process that involves understanding of plasticity across multiple-length scales. Moreover, the detailed processes by which dislocations are generated within the cell interiors and annihilated within the cell walls are not entirely understood. Despite this, a number of creep models have been developed to describe the time-dependent deformation behavior of engineering materials.

Thermal activation assists many of the mechanisms resulting in plastic deformation as well as aiding in recovery. Recovery mechanisms are strongly influenced by temperature and diffusive processes. By plotting the temperature dependence of creep rate as a function of temperature, the activation energy can be determined. The stress dependence of the creep rate can be used to calculate the stress exponent. The steady-state creep rate can be expressed as:

$$\dot{\varepsilon}_{SS} = A\sigma^n \exp\left(\frac{-Q}{RT}\right) \qquad \text{(Eq 15)}$$

where A is a material constant, R is the ideal gas constant, n is the creep stress exponent, and Q is the activation energy for creep. As demonstrated in a number of different engineering materials, the activation energy for creep is strongly related to the activation energy for lattice self-diffusion Fig. 4. At lower temperatures, however, the activation energy for creep correlates better with the lower value of the self-diffusion coefficient for grain-boundary diffusion. Moreover, a dependence of the steady-state creep rate on stacking-fault energy has also been established (Fig. 5) (Ref 8). Based on these relationships, the steady-state creep can also be expressed as (Ref 9):

$$\dot{\varepsilon}_{SS} = A\left(\frac{\sigma}{\mu}\right)^n D_L\left(\frac{\mu b}{kT}\right) \qquad \text{(Eq 16)}$$

This is referred to as the Dorn equation (Ref 7), where D_L is the lattice self-diffusion coefficient.

Prediction of Deformation. For predicting creep deformation, most materials abide by the Monkman-Grant relationship (Ref 10):

$$\dot{\varepsilon}t_r = C_{MG} \qquad \text{(Eq 17)}$$

where t_r is the time to rupture, and C_{MG} is the Monkman-Grant constant. This function shows a direct relationship between the minimum creep rate and the time to rupture (Fig. 6) (Ref 11). Creep damage accumulates mostly in the form of cavitation and cracking along the grain boundaries during tertiary creep. This Monkman-Grant relationship demonstrates that the rate of damage accumulation along the grain
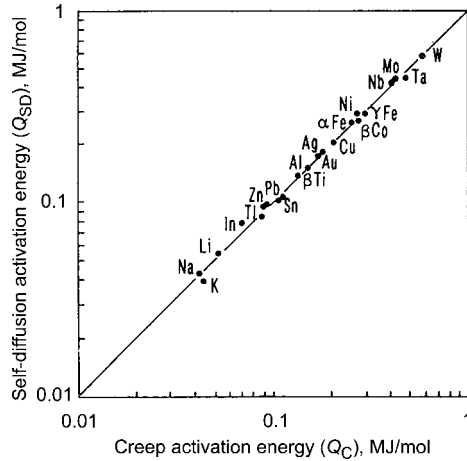
Fig. 4 Comparison of the activation energy for self-diffusion versus the creep activation energy for a variety of metals. Source: Ref 8

LIVE GRAPH
Click here to view

boundaries during tertiary creep is dependent on the rate of deformation within the grains during steady-state creep.

The ability to accurately predict the magnitude of creep deformation is an important attribute that is required for structural components that are used in creep-limited applications. For many creep-limited engineering structures, the time scales of interest are often on the order of ~10,000 to >100,000 h. With 8766 h in a year, creep deformation of actual components may occur over a period of years or even decades. In laboratory experiments, however, creep tests are conducted under much higher temperatures or stresses than those seen in service. This is done in order to expedite the acquisition of creep data, so that the tests can be performed within a reasonable period of time. As a result of the test conditions that accelerate creep deformation, it is necessary to extrapolate results obtained at high temperature and stresses to predict the creep performance at lower temperatures and stresses.

Larson-Miller parameter, P, is a commonly used engineering approach that aims to unify creep data generated over a range of temperatures and stresses. This parameter combines time to failure and temperature into a single parameter that can be expressed as:

$$P = T(\ln t + C) = \frac{Q}{R} \qquad \text{(Eq 18)}$$

where t is the time to rupture or to a defined strain, and C is a material constant. C is typically assumed to possess a value of 20, unless otherwise specified. The Larson-Miller parameter also correlated well to the ratio of the activation energy for creep divided by the gas constant. When the Larson-Miller parameter is plotted against stress, the results for a single material should all lie approximately on single line (Fig. 7). This reflects the effect of the stress in lowering the activation energy of the creep
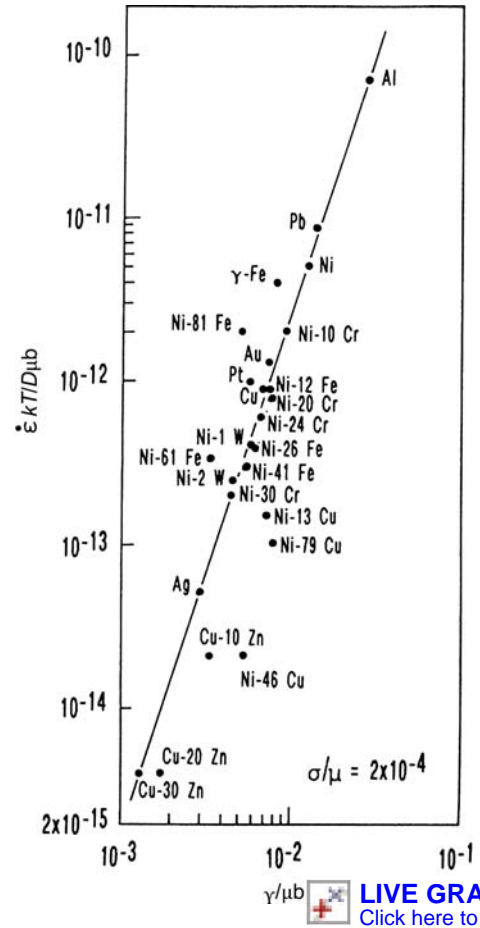
Fig. 5 Plot of the normalized creep strain rate versus the normalized stacking-fault energy for a number of alloys creeping at the same normalized stress. The linear relationship indicates that the creep rate is proportional to the stacking-fault energy.

LIVE GRAPH
Click here to view
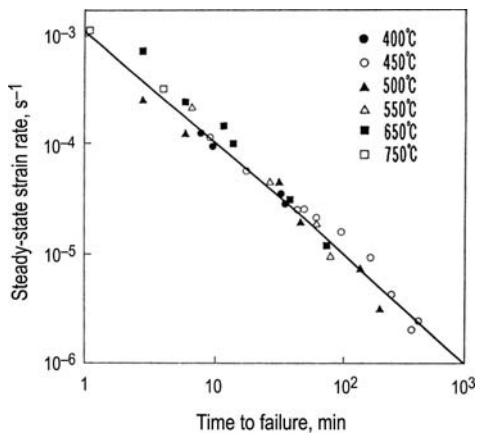
Fig. 6 Monkman-Grant relationship for copper alloys tested at a range of creep temperatures and stresses

LIVE GRAPH
Click here to view

process. Tests conducted at the same stress but at different temperatures should yield similar Larson-Miller parameters within the experimental spread and vice versa.
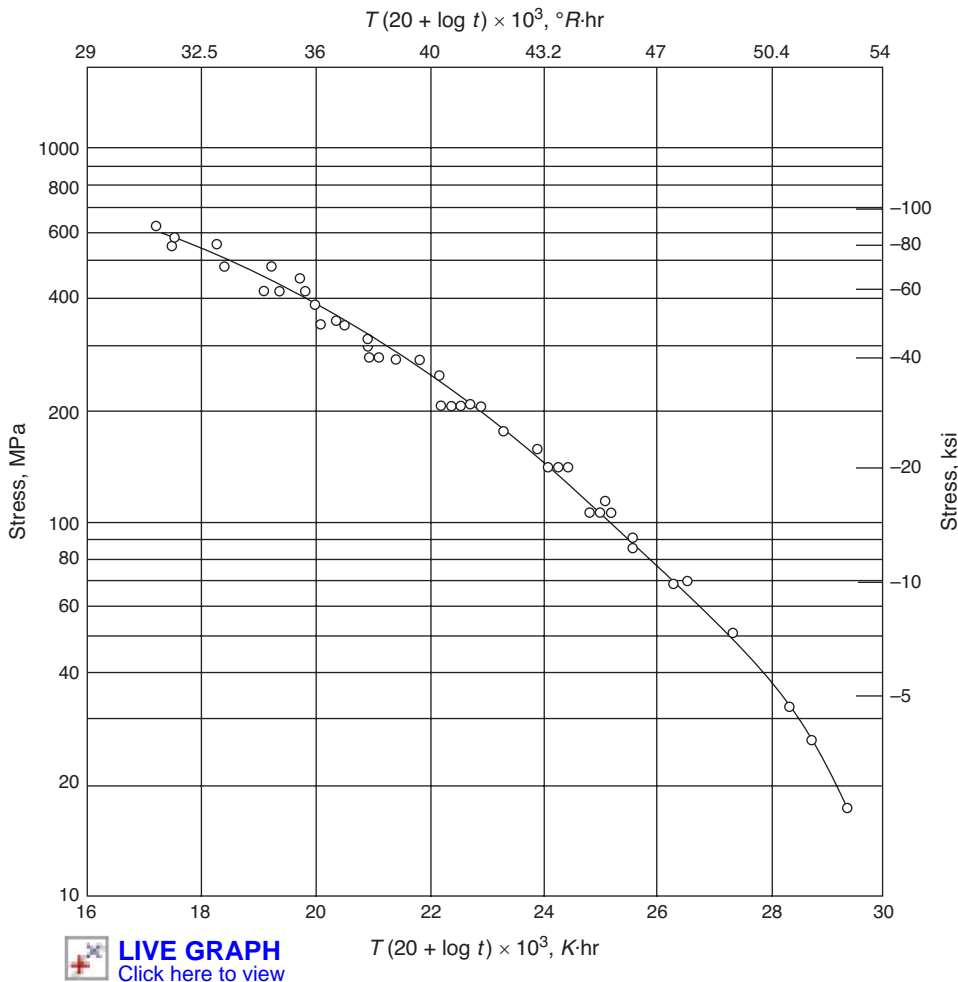
Fig. 7 Larson-Miller plot for S-590 iron

Other parameters that are commonly used for unifying creep data are the Zener-Hollomon parameter, Z, and the Sherby-Dorn parameter, θ:

$$Z = \dot{\varepsilon}\exp\left(\frac{Q}{RT}\right) \qquad \text{(Eq 19)}$$

$$\theta = t\exp\left(\frac{-Q}{RT}\right) \qquad \text{(Eq 20)}$$

The Zener-Hollomon parameter is often used for predicting the material flow behavior during hot working operations, while the Sherby-Dorn is a general creep parameter that is used in a similar manner as the Larson-Miller parameter.

Creep Mechanisms

Power Law Creep. Although effective, many of the constitutive models developed for predicting the creep performance of materials are engineering-based models that rely heavily on fitting mathematical expressions to existing databases. In order to develop truly predictive models capable of simulating the creep response as a function of temperature and stress, much effort has been dedicated to the development of physics-based models that aim to capture the rate-limiting mechanisms occurring during deformation. The difficulties of this approach, however, stem from the fact that the rate-limiting mechanism tends to vary over the wide ranges of stress and temperature encountered during service conditions.

At high stresses and at temperatures $>0.5T_\mathrm{m}$, dislocation power-law creep dominates. The strain rate is related to stress by the following equation:

$$\dot{\varepsilon} \propto \left(\frac{\sigma}{\mu}\right)^n \qquad \text{(Eq 21)}$$

For power-law creep, the strain rate (γ) is extremely sensitive to the applied stress (σ), since the stress exponent (n) tends to range between 3 and 10; μ is the shear modulus. This mode of creep involves glide of dislocations but is limited by climb of the dislocations over obstacles that inhibit further plastic flow. The obstacles may be precipitates or dislocation locks that impede their ability to glide along the slip plane. As temperatures increase to $>0.5T_\mathrm{m}$, dislocations may begin to escape from obstacles by climbing out of their slip planes and continuing to glide. In this instance, the rate-limiting mechanism for power-law creep is kinetics of the climb process (Ref 10). The rate of creep is thus determined by the rate at which dislocations move and the density of dislocations:

$$\dot{\gamma} = \rho\mathrm{b}\nu \qquad \text{(Eq 22)}$$

where ν is the average dislocation velocity, b is the Burgers vector, and ρ is the dislocation density. During plastic deformation, the density of dislocations in a metallic material is generally assumed to be proportional to the stress squared. This is due to the fact that when dislocations are pinned by interactions with other dislocations of average spacing r, the bypass stress $\sigma = \mu\mathrm{b}/r$. Since the dislocation density is inversely proportional to r^2, it is also proportional to the applied stress squared:

$$\rho = \frac{1}{r^2} \qquad \text{(Eq 23)}$$

$$\rho = \left(\frac{\sigma}{\mu\mathrm{b}}\right)^2 \qquad \text{(Eq 24)}$$

The velocity of dislocations is proportional to the applied stress. Dislocations glide within the structure in order to accommodate creep strains but can be impeded by obstacles such as kinks or jogs that form obstacles that restrict mobility. In order for deformation to continue, activation processes are required to enable climb of the dislocation over the obstacle. Under these conditions, the climb velocity, ν_c, can be expressed as (Ref 11):

$$\nu_\mathrm{c} = \frac{D_\mathrm{v}\sigma_\mathrm{n}\mathrm{b}^2}{kT} \qquad \text{(Eq 25)}$$

where D_v is the vacancy diffusion coefficient, and σ_n is the stress component normal to the slip plane. The creep strain rate can then be expressed as:

$$\dot{\gamma} = \alpha\left(\frac{\sigma}{\mu\mathrm{b}}\right)^2\frac{D_\mathrm{v}\sigma_\mathrm{n}\mathrm{b}^3}{kT} \qquad \text{(Eq 26)}$$

$$\dot{\gamma} = A\frac{D_\mathrm{v}\mu\mathrm{b}}{kT}\left(\frac{\sigma}{\mu}\right)^3 \qquad \text{(Eq 27)}$$

where A is a dimensionless constant that takes into account that the stress causing climb is proportional to the applied stress. This equation suggests that the stress exponent for steady-state creep should be three for power-law creep conditions where the rate-limiting mechanisms are associated with glide and climb of dislocations. However, this only applies for a limited number of materials, such as ice and some ceramics. Creep stress exponents will vary depending on the rate-limiting mechanisms of creep and are influenced by the particular nature of the dislocations as well as the barriers to their motion (Ref 12–15). The more general value

of the stress exponent is approximately five, but values of up to ten are commonly measured for engineering alloys.

Low Temperature Creep—Pipe Diffusion. When creep occurs at low temperatures, the measured activation energy for creep tends to fall while the stress exponent rises. These observations are attributed to the fact that lattice diffusion is no longer the dominant mechanism governing the kinetics of the system. Pipe diffusion, or diffusion of solutes or vacancies along the cores of the dislocations, becomes more pertinent than lattice diffusion. The activation energy and diffusion rates along the dislocation cores are difficult to measure experimentally but have been approximated to be similar to that of grain-boundary diffusion. With the dislocation cores serving as pathways for diffusion, the dislocation density contained within the material will have a significant impact on the pipe diffusion rates. Since the dislocation density is proportional to the stress squared, the power-law creep equation can be written as:

$$\dot{\gamma} = A \frac{D_{DC}\mu b}{kT}\left(\frac{\sigma}{\mu}\right)^{n+2} \qquad (Eq\ 28)$$

where D_{DC} is the diffusivity for pipe diffusion.

Power-Law Breakdown. When materials are subjected to high stresses, breakdown of the power law may occur as strain rates rapidly increase. The dependence of the creep strain rate on the stress can be represented with the following empirical relationship:

$$\dot{\gamma} = C \frac{D\mu b}{kT}\left[\sin h\left(\frac{\alpha\sigma}{\mu}\right)\right]^{n'} \qquad (Eq\ 29)$$

where $C\alpha^{n'} = A$ is the dimensionless constant. Under these conditions, the deformation process is predominantly governed by glide of dislocations rather than the sequential climb followed by glide characterized by the power-law creep equations. Since the applied stresses are at a level where they can readily overcome obstacles without having to first climb over them, the deformation process is no longer limited by the kinetics of dislocation climb. The high applied stresses reduce the activation barriers for glide, and the deformation mechanisms begin to resemble those of traditional plasticity.

Diffusional Creep. At low stresses, dislocation generation may be difficult. In this instance, the creep rates are proportional to the first power of stress in the power-law creep equation. Under these conditions, the creep strain rates are less sensitive to applied stresses and are strongly influenced by parameters that impact the diffusivity of the system, namely temperature and microstructure. Depending on the dominant mode of diffusion, grain boundary or lattice, the mechanism by which creep deformation occurs may vary.

In instances where the applied stresses are low and temperatures are $>0.7T_m$, creep deformation is likely to be controlled by lattice diffusion. The problem of diffusional creep was first addressed

by Nabarro and Herring (Ref 16). Creep deformation occurs as a result of coordinated atom movement between different interfaces that results in a macroscopic shape change. Intrinsically, this is an extremely slow deformation process that is limited to instances where the applied stresses are insufficient to initiate dislocation motion, while the temperature is high enough to create a gradient of vacancies that drives mass flow. Schematically, the process can be seen in Fig. 8. The creep rate resulting from diffusional flow can be derived by first considering the equilibrium concentration of vacancies, C, within a material:

$$C = C_o \exp\left(\frac{-Q_f}{kT}\right) \qquad (Eq\ 30)$$

where Q_f is the vacancy formation energy, and C_o is the equilibrium vacancy concentration coefficient. Upon the application of a stress, work is required to create vacancies. Vacancy formation, however, is influenced by the nature of the stress, because tensile stresses are more likely to result in vacancy formation than compressive stresses. Thus, at internal interfaces within the microstructure, such as grain boundaries, precipitate interfaces, and so on, the equilibrium concentration of vacancies will vary depending on whether the stresses at that particular interface are tensile or compressive. This gradient of vacancies provides a driving force for the flow of vacancies from regions of the microstructure where the stresses are tensile to the regions where compressive stresses exist. Counter to the flow of vacancies is the flow of atoms in the opposite direction. The work required to create vacancies at the tensile (C^T) and compressive (C^C) boundaries can be calculated using:

$$C^T = C_o \exp\left(\frac{-Q_f}{kT}\right)\exp\left(\frac{\sigma\Omega}{kT}\right) \qquad (Eq\ 31)$$

$$C^C = C_o \exp\left(\frac{-Q_f}{kT}\right)\exp\left(-\frac{\sigma\Omega}{kT}\right) \qquad (Eq\ 32)$$

where Ω is the atomic volume. The flux of vacancies, J_v, is then:

$$\begin{aligned} J_v &= -\frac{D_v C_o}{d}\exp\left(\frac{-Q_f}{kT}\right)\left[\exp\left(\frac{\sigma\Omega}{kT}\right) - \exp\left(-\frac{\sigma\Omega}{kT}\right)\right] \\ &\cong \frac{2D_v C_o}{d}\left(\frac{\sigma\Omega}{kT}\right) \end{aligned}$$
$$(Eq\ 33)$$

where D_v is the vacancy diffusion coefficient, and d is the grain diameter, which also represents the maximum diffusion distance of the vacancies. The strain rate caused by this diffusional flow is related to the flux of atoms deposited on the tension boundary:

$$\dot{\varepsilon} = \frac{J_v b^3}{d} = \frac{2D_v C_o}{d}\left(\frac{\sigma\Omega}{kT}\right)\frac{b^3}{d} = \frac{2D_{sd}}{d^2}\left(\frac{\sigma\Omega}{kT}\right) \qquad (Eq\ 34)$$

where D_{sd} is the self-diffusion coefficient (including the formation and migration energies
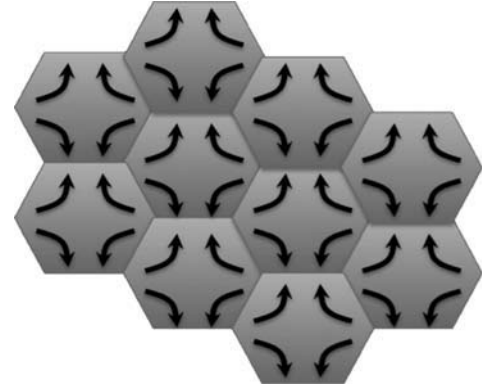
Fig. 8 Schematic showing the flow of atoms from compressive boundaries to tensile boundaries during diffusional Nabarro-Herring creep. Vacancy flow occurs in the direction opposite to that of the atom flow.

for vacancy formation). In this expression, the creep strain rate is proportional to the stress and inversely proportional to the square of the grain size.

Because diffusion is extremely sensitive to temperature, at lower temperatures lattice diffusion becomes sluggish, and diffusional flow along grain boundaries is often the dominant mass-transport mechanism. This is due to the activation energy for bulk diffusion being higher than that of grain-boundary diffusion. The effect of diffusional flow along grain boundaries, D_{gb}, was considered by Coble (Ref 17):

$$\dot{\varepsilon} = \frac{\delta D_{gb}\sigma b^4}{kTd^3} \qquad (Eq\ 35)$$

Here, the self-diffusion coefficient is replaced by the grain-boundary diffusion coefficient, and the cross-sectional area through which diffusion occurs can be assumed to be δbd, where δb is a measure of the thickness of the grain boundary. In most cases, δ is a dimensionless constant with a value of ~1. As is expected, the creep strain rates for Coble creep are highly sensitive to the grain size of the material. Small grain sizes allow for large grain-boundary areas to exist within the microstructure and are susceptible to comparatively high creep strain rates. When considering diffusional creep, it is important to note that both lattice diffusion and grain-boundary diffusion occur simultaneously. However, as the temperature increases and the activation energy for vacancy formation $Q_f = kT$, the high sensitivity of the creep strain rate to diffusivity means that there is an abrupt change from grain-boundary diffusion (Coble) to bulk lattice diffusion (Nabarro-Herring) creep kinetics.

Twinning during Creep Deformation. Mechanisms that accommodate creep deformation do not necessarily need to involve power-law creep-type dislocation activity or long-range diffusive processes. Twinning or localized microscale reordering of the crystalline structure during creep deformation has been reported in alloys

with a limited number of operative slip systems, such as hexagonal close-packed metals (Ref 18, 19), ordered intermetallics (Ref 20, 21), and even precipitation-strengthened nickel-base superalloys (Ref 22, 23). Generally considered a deformation mechanism that is limited to relatively low temperatures not indicative of creep conditions, twinning involves the dissociation of existing dislocation configurations into Shockley partials that form complex, multilayered stacking-fault structures that nucleate twins. Recently, microtwinning has been reported to serve as the dominant deformation mechanism during creep of polycrystalline nickel-base superalloys within a narrow range of temperatures and stresses. Efforts to model this particular deformation mechanism and assess the impact of this mechanism on creep are currently in progress (Ref 24).

Deformation Mechanism Maps. In principle, more than one creep mechanism may operate in a material, depending on the temperature and stress. However, for a particular combination of temperature and stress, a single creep mechanism usually dominates and serves to limit the kinetics of deformation. Since the boundaries between the different regimes tend to be relatively sharp, deformation mechanism maps, or

Ashby maps (Ref 25), may be used to define the dominant mechanism as a function of composition, grain size, temperature, and stress (Fig. 9). In their most common form, the axes of the maps are expressed in terms of the normalized shear stress, σ_s/μ, and the homologous temperature, T/T_m. Although each map is specific to a particular material and microstructure, they allow for the identification of the temperature and stress dependence based on the rate equations presented in this article. Strain-rate contours on these deformation mechanism maps are determined using the rate equations presented in this article. The spacing and gradients of these contours show that they are roughly parallel to one another. The shape and position of the various creep regimes on the deformation mechanism map is consistent from alloy to alloy, but microstructure plays an important role in shifting the boundaries and defining transitions between different mechanisms.

Creep-Strengthening Mechanisms

Most structural engineering components are fabricated from alloys rather than pure metals,

due to the need to impart high levels of strength in order to resist deformation. Strengthening arises through a wide variety of mechanisms, including substitutional and interstitial solution strengthening, dispersion and/or precipitation strengthening, and even grain-size strengthening. Depending on how these mechanisms interact with the ability of dislocations to glide within the structure, the effectiveness of these mechanisms will vary, particularly as a function of temperature. Interactions between dislocations and obstacles may be classified as being either strong or weak, depending on the magnitude of the interaction (Ref 26). When dislocations encounter strong obstacles, such as nonshearable precipitates, along its slip plane, they must bow through large angles in order to break away and bypass the obstacles and continue gliding. Weak obstacles, such as solute atoms or small, shearable precipitates, on the other hand, can be bypassed relatively easily without undergoing a large deviation in the local orientation of the dislocation line. Thus, strong and weak interactions can be characterized on the basis of the critical angle, φ_c, for breakaway of the dislocation from the obstacle.

As dislocations are forced to bow or climb over obstacles, the net increase in energy associated with the change in the dislocation line length can be expressed by defining the line tension, Γ. This tension force represents a resistance to lengthening or bowing of the dislocation. For a mixed screw-edge dislocation, the line tension is a function of the angle, θ, between the dislocation and the corresponding Burgers vector:

$$\Gamma = E_{el}(\theta) + \frac{d^2 E_{el}(\theta)}{d\theta^2} \qquad \text{(Eq 36)}$$

where $E_{el}(\theta)$ is the dislocation line energy per unit length as a function of the angle. The line tension is also commonly simplified and expressed in terms of the shear modulus, G, and the Burgers vector, b, as:

$$\Gamma = \frac{Gb^2}{2} \qquad \text{(Eq 37)}$$

The change in the length of the dislocation segment bowing in between two obstacles, dS, can also be expressed in terms of the bow radius, R, and θ as $dS = Rd\theta$. The application of a force balance between the outward force on the dislocation caused by the application of a shear stress, τ, results in:

$$\tau b dS = \tau b R d\theta = 2\Gamma \sin\left(\frac{d\theta}{2}\right) \qquad \text{(Eq 38)}$$

For small angles of θ, $\sin\theta \approx \theta$, so:

$$\tau b R d\theta = 2\Gamma\left(\frac{d\theta}{2}\right) \qquad \text{(Eq 39)}$$

$$\tau b R = \Gamma = \frac{Gb^2}{2} \qquad \text{(Eq 40)}$$
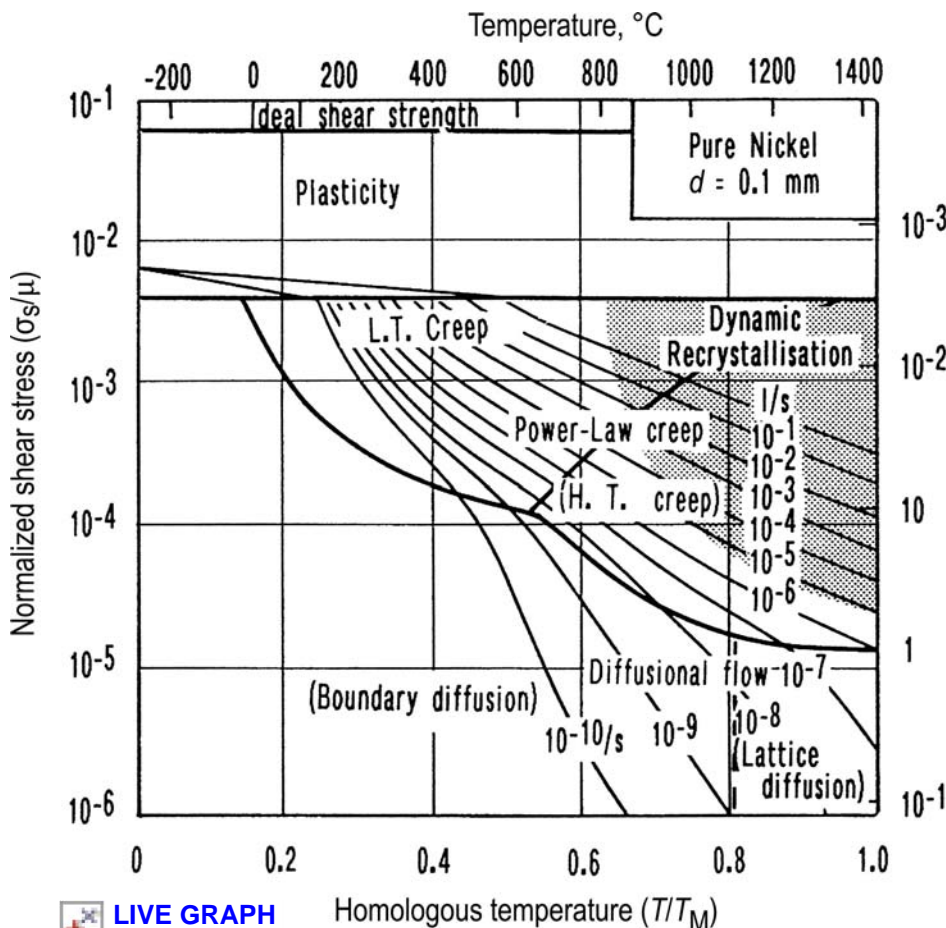
Fig. 9 Ashby deformation mechanism map for pure nickel with a grain size of 0.1 mm. L.T., low temperature; H.T., high temperature

$$\tau = \frac{Gb}{2R} \qquad \text{(Eq 41)}$$

The magnitude of the shear stress is inversely proportional to the bow radius of the dislocation. Since the particle/obstacle spacing, L, is also related to the bow radius, $R = L/2$, the critical configuration for escape of the dislocation from strong, nonshearable obstacles is:

$$\tau = \frac{Gb}{L} \qquad \text{(Eq 42)}$$

This relation is also commonly referred to as the Orowan stress and represents the upper limit of strengthening for nonshearable obstacles where $\varphi_c = 0$. In instances where the obstacles may be sheared by the dislocation, deformation may occur at stresses well below those defined by the Orowan stress. If one considers the force, F, exerted on the obstacle by the dislocation as it bows out:

$$F = 2\Gamma \cos\left(\frac{\varphi}{2}\right) \qquad \text{(Eq 43)}$$

$$\tau = \frac{Gb}{L} \cos\left(\frac{\varphi}{2}\right) \qquad \text{(Eq 44)}$$

This expression clearly suggests that strong obstacles are associated with dislocations bowing through a small critical angle, φ_c, while weak obstacles result in high bowing angles.

Creep in Engineering Alloys—Microstructural Modeling

In multicomponent materials that contain multiple phases or elements, the creep deformation mechanisms may differ from those already discussed. Particularly in alloys that are subject to microstructural changes during creep, the steady-state equations do not necessarily hold true. A constitutive description of the strain-rate-dependent behavior of engineering alloys must take into account the intrinsic causes of damage accumulation in the form of cavitation, dislocation evolution, and grain-boundary sliding. Moreover, because creep occurs at elevated temperatures, seldom does the microstructure of the alloy remain constant. Diffusive processes aid thermodynamic transformations that result in grain growth, precipitation of undesirable secondary phases, along with coarsening and dissolution of existing precipitates. As a result of these microstructural instabilities occurring during creep deformation, empirical or theoretical steady-state approaches based on the power-law creep equation (Ref 27–35) are usually insufficient for modeling creep deformation in engineering alloys (Ref 36).

To be able to accurately capture both the damage evolution as well as the microstructural changes that occur during creep, significant effort has gone into the development of constitutive equations that incorporate kinetic terms that reflect the microstructural changes that occur during creep. Consider a multicomponent

engineering alloy that is strengthened by both precipitates and solutes in a solid solution. If the precipitates are randomly distributed within the microstructure, one can define the microstructure in terms of kinetic parameters, where r_s is the mean plane section radius of the precipitates, ϕ_p is the volume fraction of precipitates, and ρ_{tot} is the total dislocation density. As a result of the precipitates and other obstacles, dislocations possess limited mobility during creep because they become trapped by obstacles. Similar to many of the basic theoretical models, thermal activation allows some of these dislocations to climb over the obstacles and continue viscous glide. By incorporating the geometrically random nature of the dislocation particle intersections into a climb/glide model, however, Dyson and McLean determined that the ability of the dislocations to separate themselves and escape from the climbing network was based on statistics and dependent on the stress, temperature, and particle-dispersion parameters. The shear creep rate, $\dot{\gamma}$ can be expressed as:

$$\dot{\gamma} = 2\rho\phi_p{}^{0.5}\left(\left[\frac{\pi}{4}\right]^{0.5} - \phi_p{}^{0.5}\right)c_j D_v \sinh\left(\frac{\tau_m b^2 \lambda_p}{kT}\right) \qquad \text{(Eq 45)}$$

where τ_m is the shear stress within the matrix, D_v is the matrix diffusivity, c_j is the dislocation jog density, ρ is the dislocation density that is able to glide, and the interparticle spacing is λ_p:

$$\lambda_p = 2r_s\left(\frac{\left[\frac{\pi}{4}\right]^{0.5} - \phi_p{}^{0.5}}{\phi_p{}^{0.5}}\right) \qquad \text{(Eq 46)}$$

With knowledge of the shear creep rate, the uniaxial creep strain rate can be expressed as:

$$\dot{\varepsilon} = \frac{2\rho\phi_p{}^{0.5}}{\bar{M}}\left(\left[\frac{\pi}{4}\right]^{0.5} - \phi_p{}^{0.5}\right)c_j D_v \sinh\left(\frac{\sigma_m b^2 \lambda_p}{MkT}\right) \qquad \text{(Eq 47)}$$

where $\bar{M}$ is the Taylor factor, and σ_m is the uniaxial matrix stress. This uniaxial matrix stress is related to the applied tensile stress, σ, by assuming that the precipitates deform elastically during creep and that the matrix is the only phase that deforms irreversibly. Since $\sigma_m = (\sigma - \sigma_k)$, where σ_k is the kinematic internal stress, the kinematic creep equation can be expressed as:

$$\dot{\varepsilon} = \frac{2\rho\phi_p{}^{0.5}}{\bar{M}}\left(\left[\frac{\pi}{4}\right]^{0.5} - \phi_p{}^{0.5}\right)c_j D_v \sinh\left(\frac{(\sigma - \sigma_k)b^2 \lambda_p}{MkT}\right) \qquad \text{(Eq 48)}$$

Since the kinematic internal stress varies as a function of the microstructure, this term corresponds to a physical representation of the microstructure, such as changes in dislocation density, grain size, precipitate size, morphology, and volume fraction.

To effectively model the creep behavior of engineering alloys using these kinematic creep equations, one can use the continuum damage mechanics (CDM) approach. This model was

developed to deal with the kinetics of microstructural changes by defining the creep strain rates as a function of hardening and softening parameters that correspond to phenomenological events (Ref 24–26). This is a powerful creep-modeling technique that can be used to identify the dominant mechanisms governing deformation as well as to provide insight into the microstructural changes that occur during creep. For the internal-state variable approach, creep deformation can be expressed as a set of linear differential equations:

$$\dot{\varepsilon} = \dot{\varepsilon}(\sigma, T, H, D_i)$$
$$\dot{H} = \dot{H}(\sigma, T, H, D_i)$$
$$\dot{D}_i = \dot{D}_i(\sigma, T, H, D_i) \qquad \text{(Eq 49)}$$

where T is temperature, H is a dynamic hardening parameter that dominates primary creep behavior, and D_i is a dimensionless microstructural damage parameter that evolves during secondary and tertiary creep. $\dot{H}$ is the rate of dynamic hardening, and $\dot{D}_i$ is the rate at which damage accumulates within the microstructure. In most instances, the primary damage mechanisms resulting in creep failures are associated with the following microstructural features: dislocation activity, particulates/precipitates, and grain-boundary cavitation. It is important to note that the creep damage mechanisms may be occurring simultaneously or in isolation from one another. The explicit form of the CDM creep model may be expressed as:

$$\dot{\varepsilon} = \frac{\dot{\varepsilon}_o}{1 - D_d}\sinh\left(\frac{\sigma(1 - H)}{\sigma_o(1 - D_p)(1 - D_c)}\right) \qquad \text{(Eq 50)}$$

In this formulation, parameters from the kinematic creep equation derived previously can be integrated into the CDM model:

$$\dot{\varepsilon}_o = \frac{2\rho\phi_p{}^{0.5}}{\bar{M}}\left(\left[\frac{\pi}{4}\right]^{0.5} - \phi_p{}^{0.5}\right)c_j D_v \qquad \text{(Eq 51)}$$

$$\sigma_o = \frac{\bar{M}kT}{b^2 \lambda_{p,i}} \qquad \text{(Eq 52)}$$

The hardening parameter, $H = \sigma_k/\sigma$, and damage parameters corresponding to dislocation activity, particles, and cavitation are defined as D_d, D_p, and D_c, respectively. Since the hardening parameter also evolves as a function of time or deformation, it can also be expressed as a rate equation, $\dot{H}$, where:

$$\dot{H} = \frac{\phi_p E}{\sigma}\left(1 - \frac{H}{H^*}\right)\dot{\varepsilon} \qquad \text{(Eq 53)}$$

The volume fraction of precipitates is ϕ, the elastic modulus of the precipitates is E, and H^* is the normalized precipitate volume fraction:

$$H^* = \frac{2\phi_p}{2\phi_p + 1} \qquad \text{(Eq 54)}$$

In addition to the hardening response, materials also undergo a softening behavior as damage accumulates within the microstructure. Damage

parameters can be attributed to gradual changes in dislocation density, (D_d), particles (D_p), and grain-boundary cavitation (D_c):

$$D_d = 1 - \frac{\rho_i}{\rho} \qquad \text{(Eq 55)}$$

$$D_p = 1 - \frac{\lambda_{p,i}}{\lambda} \qquad \text{(Eq 56)}$$

$$D_c = 1 - \frac{A_{c,i}}{A} \qquad \text{(Eq 57)}$$

where ρ, λ_p, and A_c are the dislocation density, interparticle spacing, and area fraction of the cavitated grain boundary. The subscript, i, denotes the initial for these parameters. The numerical value for each of these parameters ranges from zero to unity. Expressed in terms of rate equations, these damage parameters can be written as:

$$\dot{D}_d = \frac{\phi_p \rho^{0.5}}{r_s \rho_i} (1 - D_d)^2 \dot{\varepsilon} \qquad \text{(Eq 58)}$$

$$\dot{D}_p = \frac{K_p}{3} (1 - D_p)^4 \qquad \text{(Eq 59)}$$

$$\dot{D}_c = \frac{1}{3\varepsilon_{f,u}} \dot{\varepsilon} \qquad \text{(Eq 60)}$$

where r_s corresponds to the dislocation spacing, K_p is the particle coarsening rate, and $\varepsilon_{f,u}$ is the strain at creep rupture. Coupling of all of these rate expressions into the CDM model allows for both the prediction of the creep response as well as the evolution of the microstructure during deformation. This is an extremely useful model that has been successfully applied to simulate the deformation response of a number of engineering alloys in commercial applications.

REFERENCES

1. U.F. Kocks, A.S. Argon, and M.F. Ashby, *Prog. Matls. Sci.*, Vol. 19, 1975
2. A.H. Cottrell and V. Aytekin, *J. Inst. Metals*, Vol 77, p. 389, 1950
3. T. Takeuchi and A.S. Argon, *J. Matls Sci.*, Vol 11, p. 1542, 1976
4. R.W. Bailey, *J. Inst. Metals*, Vol 35, p. 27, 1926
5. E. Orowan, *J. Iron Steel Inst.*, Vol 54, p. 45, 1946
6. F. Prinz and A.S. Argon, *Acta Metall.*, Vol 32, p. 7, 1984
7. J.E. Bird, A.K. Murkerjee, and J.E. Dorn, *Quantitative Relation between Properties and Microstructure*, Israel Univ. Press, Jerusalem, 1969
8. F.A. Mohamed and T.G. Langdon, *Acta. Metall.*, Vol 30, p. 779, 1974
9. O.D. Sherby and J. Weertman, *Acta. Metall.*, Vol 27, p. 387, 1979
10. F.C. Monkman and N.J. Grant, *Proc. Ameri. Soc. Test. Mater.*, Vol 56, p. 593, 1956
11. J.P. Hirth and J. Lothe, "Theory of Dislocations"
12. J. Weertman, *J. Appl. Phys.,* Vol 28, p. 362, 1957
13. G. Gottstein and A.S. Argon, *Acta Metall.*, Vol 35, p. 1261, 1987
14. W.D. Nix, J.C. Gibeling, and D.A. Hughes, *Metall. Trans.,* Vol 16A, p. 2215, 1985
15. J. Weertman, *Creep and Fracture of Engineering Materials and Structures*, B. Wilshire and D. Owens, Ed., Inst. Metals, London, Vol.1, 1986
16. F.R.N. Nabarro, *Report on Conf. of Strength of Solids,* The Physical Society, London, Vol 1, 1948
17. R.S. Coble, *J. Appl. Phys.,* Vol 34, p. 1679, 1963
18. G.E. Lucas and R.M.N. Pelloux, *Metall. Mater. Trans.*, Vol 12, No. 7, p. 1321, 1981
19. M.A. Morris and M. Leboeuf, *Intermetallics*, Vol 5, No. 5, p. 339, 1997
20. L. Min and K.J. Hemker, *Acta Materialia*, Vol 45, No. 9, p. 3573, 1997
21. D.M. Knowles and Q.Z. Chen, *Mater. Sci. Eng. A.,* Vol 340, No. 1–2, p. 88, 2003
22. M. Ardakani, M. McLean, and B.A. Shollock, *Acta Materialia*, Vol 47, No. 9, p. 2593, 1999
23. G. Viswanathan, P. Sarosi, M. Henry, D. Whitis, W. Milligan, and M. Mills, *Acta Materialia*, Vol 53, No. 10, p. 3041, 2005
24. S. Karthikeyan, R. Unocic, P. Sarosi, G. Viswanathan, D. Whitis, and M. Mills, *Scripta Materialia,* Vol 54, No. 6, p. 1157, 2006
25. H. Frost and M.F. Ashby, *Deformation Mechanism Maps*, Pergamon Press, 1980
26. A.J.E. Foreman and M.J. Makin, *Phil. Mag.,* Vol 14, No. 131, p. 911, 1966
27. F.A. Leckie and D.R. Hayhurst, *Proc. Roy. Soc. Lond.,* Vol A340, p. 1689, 1974
28. A. Merzer and S.R. Bodner, *ASME J. Engr. Mater. Tech.,* Vol 101, p. 388, 1979
29. J.L. Chaboche and G. Rousselier, *J. Press. Vess. Tech.,* Vol 105, p. 153, 1983
30. J. Lemaitre and J.L. Chaboche, *Mechanics of Materials*, Cambridge University Press, 1990
31. O.D. Sherby and P.M. Burke, *Prog. in Mater. Sci.,* Vol 13, p. 325, 1967
32. G.S. Ansell and J. Weertman, *Trans AIME,* Vol 215, p. 838, 1959
33. B.F. Dyson and M. McLean, *Acta. Metall.,* Vol 31, p.17, 1983
34. Y.N. Rabontnov, *Proc. XII IUTAM Congress,* Stamford, M. Hetenyi and W.G. Vincenti, Ed., Springer, p. 137, 1969
35. L.M. Kachanov, *Izc. Ak., Nauk SSSR Otedel. Tekh. Nauk,* Vol 8, p. 26, 1958
36. B.F. Dyson and M. McLean, in *Microstructural Stability in Creep Resistant Alloys for High Temperature Plant Applications*, Institute of Metals, London, p. 371, 1998

Microstructure-Sensitive Modeling and Simulation of Fatigue

David L. McDowell, Georgia Institute of Technology

HISTORICALLY, local stress-based criteria for high-cycle fatigue (HCF) or plastic strain-based criteria for low-cycle fatigue (LCF) have been applied based on transfer of results from tests on relatively small-scale unnotched laboratory specimens (Ref 1) to structural components to estimate crack initiation life. Similarly, test results for crack propagation in precracked specimens have been employed to estimate propagation behavior of components. Explicit consideration of microstructure in these relations has been limited, and mechanisms are not completely understood (Ref 2). In this approach, it is typically assumed that the life is decomposed between initiation and propagation components, that is, $N_T = N_i + N_p$, with the latter based on fracture mechanics correlations for crack growth, for example, $da/dN = f(\Delta K, K_{max})$. The cyclic crack growth rate is a function of of the stress-intensity range and maximum stress intensity. The crack size at the end of the initiation process, a_i, is based either on consistency with crack lengths measured in smooth specimens to define the strain-life relations, typically 0.5 to 1 mm, or some nonarbitrary length corresponding to a transition from a regime of growth that is affected by the notch root stress and strain field to influence of only the remote nominal fields, incorporating the notch as part of the crack.

The treatment of fatigue crack initiation, involving the formation and early growth of cracks to some detectable size, has traditionally been distinguished from damage-tolerant analysis, which considers propagation of pre-existing cracks. In reality, both approaches are relevant to modeling total fatigue lifetime of components. Different industry sectors conventionally design components for fatigue resistance with a focus either on crack initiation or crack propagation. Figure 1 distinguishes initiation-based design applications involving high-frequency reciprocating or rotating components from propagation-based design applications, such as high-performance aircraft or building infrastructure. The distinction is primarily rooted in the practicality and cost of detecting and mitigating catastrophic failure scenarios via inspection and maintenance. The propagation life of initiated cracks in high-strength components that rapidly accumulate fatigue cycles in the HCF regime is typically short compared to initiation life, and these cracks are often difficult to detect under service conditions. Accordingly, initiation-based modeling approaches are used that account for size effects and scatter in fatigue to estimate minimum life. On the other hand, this approach is often too conservative in practice, and it is sometimes desired to "retire for cause" rather than on the basis of these estimates; this means that cracks should be detected and monitored with respect to location and size in order to remove or replace the component. This is only a viable strategy if a reliable and repeatable detection capability exists, as is the case with inspection of hot spot regions of large structures such as aircraft, building columns, bridge supports and beams, and so forth. This latter strategy of "damage tolerance" is obviously a costly investment if manual inspections are performed, or if the cost of remote sensing is high. It drives the field of material prognosis for damage-state awareness to develop low-cost, embedded sensors that can reliably detect cracks; moreover, as the detection of incipient damage (initiation stage) becomes realized, it will enable shifting of more applications toward the top in Fig. 1, increasing the importance of combined initiation and propagation strategies.

Purposes and methods of fatigue modeling and simulation differ among initiation- and propagation-dominated scenarios. Typically, the purpose of the former is to design either failsafe components or components with a finite life; the latter seeks to quantify remaining life of components with pre-existing cracks using fracture mechanics, with the intent of monitoring via an inspection scheme. Considerable variability of fatigue response is observed even for smooth specimens under controlled testing conditions, especially in HCF regime. This has been addressed by employing statistical approaches for minimum life design, typically requiring large sets of experiments. Particularly in the crack-initiation regime, the role of microstructure can be pronounced, both in terms of mean fatigue lifetime and scatter in fatigue. Moreover, especially in the HCF regime, component and notch size effects are manifested with dependence on underlying microstructure. In view of the intimate role of microstructure, computational modeling in the HCF regime is of most clear relevance and is emphasized in this article.

Stages of the Fatigue Damage Process

The advent of microstructure-sensitive micromechanics modeling of cyclic plastic deformation and fatigue processes in the last decade that employ emerging tools in computational materials science and mechanics has opened
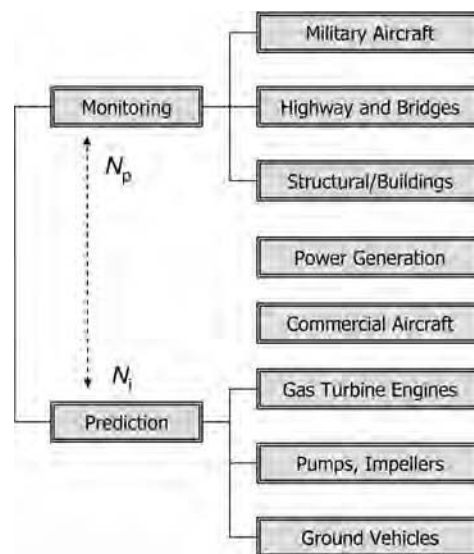
Fig. 1 Relative weighting of initiation and propagation approaches to fatigue-life estimation of components by industry sector

new horizons for fatigue modeling. Specifically, it is envisioned that microstructure-sensitive fatigue modeling can assist in tailoring of material microstructure for fatigue resistance, addressing the hierarchical structure of the material in concurrent design of microstructure and component geometry. Second, it can serve to augment costly experiments that characterize dependence of fatigue response on material microstructure, thereby reducing cycle time for insertion of modified or new materials.

Microstructure-sensitive modeling must address key aspects of physical processes of fatigue from the scales of submicron microstructure to components. As a foundation for the present article, the reader is referred to already well-established introductory textbooks and monographs (Ref 1, 2) regarding elementary concepts and definitions of stress-life, plastic strain-life, and strain-life relations for fatigue crack initiation, along with state-of-the-art approaches for correlating growth of long fatigue cracks. Early experimental work (Ref 1, 2) indicated the causal influence of cyclic plastic strain in LCF, and it has become clear that even HCF failure is governed by distributed microplasticity at the scale of individual grains. Moreover, small cracks with size on the order of microstructure heterogeneity (grains, phases) propagate below driving-force levels corresponding to long crack thresholds and can be arrested by interaction with strong microstructure barriers. Such microstructurally small cracks are subject to the varying anisotropy and heterogeneity of microstructure, giving rise to scatter in fatigue crack intiation and a breakdown of applicability of long crack fracture mechanics (Ref 3, 4). It is necessary to clarify the definition of various stages of the fatigue process. Consider the decomposition (Ref 5, 6):

$$N_t = N_i + N_p$$
$$N_i = N_{nucl} + N_{MSC} + N_{PSC'}$$
$$N_p = N_{PSC''} + N_{LC} \qquad \text{(Eq 1)}$$

where N_t is the total fatigue life (cycles to failure); initiation life, N_i, corresponds to development of cracks that are substantially longer than scales of intrinsic microstructure barriers such as grains or phases; and N_p is the number of cycles to propagate the crack(s) to failure. In Eq 1, N_{nucl} is the number of cycles to nucleate a crack (formation of a crack embryo and subcritical propagation to size of stable nucleus), and N_{MSC}, N_{PSC}, and N_{LC} represent the number of cycles to propagate the crack(s) in the regimes of microstructurally small, physically small, and long crack growth, respectively. The microstructurally small regime is normally on the order of 3 to 10 times the grain or phase size/spacing that affects retardation of the crack driving force; it is characterized by an explicit dependence of the fatigue crack growth rate on microstructure attributes such as grain size, second-phase particle size and spacing, and so on. It also depends on amplitude of applied stress, since such explicit dependence is manifested in cases where the size of the cyclic crack tip plastic zone depend on the order of the damage process zone (e.g., grain size, interphase spacing, etc.). Physically small cracks are sufficiently large such that individual microstructure attributes do not affect the crack growth rate; in other words, they are large enough such that the damage process zone at the crack tip serves as a representative volume element with regard to the fatigue crack growth rate (FCGR). Accordingly, conventional homogeneous fracture mechanics approaches apply. N_{PSC} is further decomposed as $N_{PSC} = N_{PSC'} + N_{PSC''}$, reflecting the fact that a given definition of initiation often implies an artificial partition of the growth history into parts associated with initiation and propagation regimes. These parts are often distinguished by definition rather than physics. In most engineering applications, crack initiation is defined corresponding to crack lengths beyond the microstructurally small crack (MSC) regime. For example, definition of an initiation crack length of 500 μm in a polycrystalline material with grain size of 20 μm is well beyond the regime of strong microstructure influence, the microstructurally small crack regime. A long crack-growth model, recognizing the crack shape and aspect ratio, can be applied from this point up to 500 μm. Fatigue crack initation approaches that embed $N_{PSC'}$ into some parametric law (e.g., Coffin-Manson) based on plastic strain range or stress range implicitly parameterize the growth relations and assume similitude among geometries. This is not a valid assumption in cases of macronotches with different acuities or varying intensity of stress field gradients, for example.

The relative proportion of the components of fatigue life in Eq 1 depend both on the applied stress amplitude as well as whether or not the specimens or components are notched. The ratio of the nucleation fatigue life to N_t for smooth uniaxial specimens, for example, qualitatively has the sigmoidal shape shown in Fig. 2 (a), approaching 0 in the limit at $N_t \rightarrow 1$, and 1 as $N_t \rightarrow$ infinity. The precise dependence of this ratio on N_i depends on material, with higher-strength materials resulting in a shift of the curve to the left in accordance with a reduced transition fatigue life. Lower-strength materials are considered to have greater fatigue ductility, with the ability to accommodate deformation with more extensive cyclic slip, and exhibit a more pronounced resistance to crack propagation.

For a given arbitrary fatigue life, a material may be loosely categorized as fatigue ductile or fatigue brittle. Figure 2(b) schematically compares characteristic fractions of nucleation, small crack propagation, and long crack propagation for these two distinct classes, the former typically pertaining to low-strength and the latter typically to high-strength materials. In the LCF regime, the nucleation life can be negligible for many materials (particularly alloys with extrinsic defects such as inclusions), so Fig. 2 (b) pertains more to an intermediate transition fatigue life. On the other hand, nucleation and MSC growth often dominate in HCF, and therefore, the design of fatigue-resistant materials in this regime would focus on fine scales of cracks relative to microstructure. It is emphasized that the mechanisms for high-strength materials are not really brittle in nature; rather, the scale of the process zone for inelastic deformation and evolving damage is typically much more refined than that of fatigue ductile materials. As a result, crack growth in fatigue brittle materials near the onset of growth can appear to follow a stage II trajectory (see inset at lower right in Fig. 3) (Ref 7) normal to the direction of the maximum applied principal stress, even in the MSC regime. Microstructure-sensitive MSC growth laws are often more relevant to component life estimates for fatigue ductile materials. Moreover, multiaxial fatigue parameters should distinguish the two classes of behavior (Ref 8, 9). For either class, the differences are not as significant in the LCF regime as in the transition fatigue and HCF regimes. Materials can also exhibit transitions of behavior from fatigue ductile to brittle in the HCF and very high-cycle fatigue (VHCF) regimes.

(a) LIVE GRAPH
Click here to view
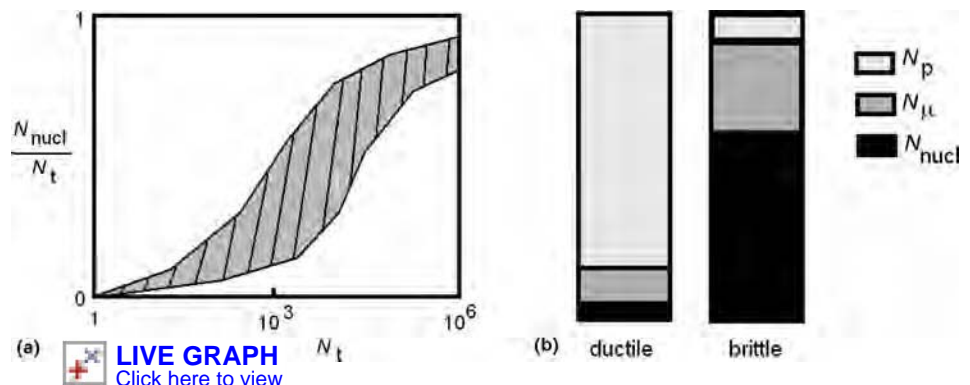
Fig. 2 (a) Crosshatched regime for typical ratios of N_{nucl}/N_t as a function of total fatigue life. (b) Schematic showing relative ratios of nucleation, small crack lives ($N_\mu = N_{MSC} + N_{PSC'}$), and long crack lives ($N_p = N_{PSC''} + N_{LC}$) for fatigue ductile and fatigue brittle materials for intermediate fatigue lives
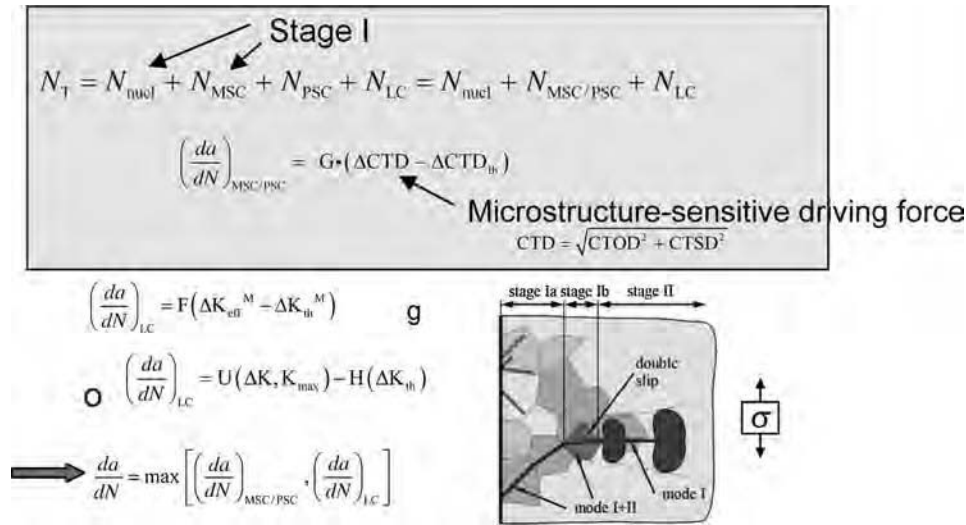
Fig. 3 Heuristic multistage modeling framework for nucleation and growth of fatigue cracks. Cycles N_{MSC} and N_{PSC}, respectively, represent the number of cycles to propagate the crack(s) in the regimes of microstructurally small crack (MSC) growth (normally 3–10 times the grain or second-phase size/spacing that affects retardation of the crack driving force; see inset figure at lower right. Source: Ref 7), physically small crack (PSC) growth, and long crack (LC) growth. In the PSC regime, the crack is suitably long to be treated using conventional linear elastic fracture mechanics (LEFM) but is still below the size considered amenable to a definition of the initial crack length for propagation analyses using LEFM. Crack tip displacement (CTD) is comprised of crack tip opening (CTOD) and sliding (CTSD) components. Constants F and M govern the long crack growth rate for a given R-ratio, and $U(*)$ and $H(*)$ are functions of their arguments.

Table 1 Relative emphasis on fatigue micromechanisms in the low-cycle (LCF) and high-cycle (HCF) regimes

Mechanisms	LCF—Extensive plasticity	HCF—Microstructure-scale (constrained) plasticity
Crack formation	Propagation-dominated: largest grains or inclusions establish initial crack length in propagation analysis.	Initiation-dominated: largest grains or inclusions control number of cycles to form a crack or to propagate past arrest limits.
Microstructurally small crack growth	Multiple cracks grow in elastic-plastic field with less microstructure influence.	First few microstructural barriers control fatigue limit and scatter of lifetime.
Physically small and long crack growth	Elastic-plastic growth persists well into crack growth history; coalescence of multisite cracks can occur.	Transition to linear elastic fracture mechanics-dominated homogeneous crack growth; single dominant crack is common.

Hierarchical Multistage Fatigue Modeling

Each regime of fatigue crack formation and growth outlined in Eq 1 should be treated with a distinct set of characteristic governing relations. Hence, they may be regarded as multistage relations. McDowell and co-workers introduced a multistage approach to microstructure-sensitive fatigue crack formation and growth that employs fatigue indicator parameters for crack nucleation at the subgrain scale (to be discussed later), along with heuristic relations for growth based on the cyclic crack tip displacement (ΔCTD), outlined in Fig. 3 (Ref 5, 6, 10). As is the case with the nucleation driving force, it is presumed that the ΔCTD is computed from microstructure-sensitive crystal plasticity relations for cracks of various scales relative to microstructure.

In the da/dN versus ΔCTD relation, G is a constant for a given microstructure (Ref 11). It is directly related to irreversibility of slip in the damage process zone. Micromechanical studies can be introduced to capture ΔCTD interactions with microstructure (Ref 4, 12–19). In general, due to lack of similitude and shear localization at the tips of cracks on the order of grain size, macroscopic, homogeneous fracture mechanics approaches to estimate ΔCTD are generally insufficient (Ref 4). The form shown in Fig. 3 does not rely on applicability of small-scale yielding or conventional linear elastic fracture mechanical (LEFM) concepts. As the crack lengthens, conventional da/dN versus ΔK_{eff} relations may be used, with the crossover defined by the maximum of MSC/PSC and long crack growth rates, as indicated in Fig. 3.

Previous work (Ref 10) has suggested an alternative definition of HCF, LCF, and transition fatigue regimes for heterogeneous microstructures compared to the standard ASTM Interantional definition based on the ratio of macroscopic elastic and plastic strain ranges

(Ref 1). Specifically, the HCF regime corresponds to stress amplitudes below macroscopic yielding for which cyclic plastic strain occurs within isolated regions of microstructure stress concentration (inclusions, favorably oriented grains, etc.). This is the regime of constrained cyclic microplasticity. Transition to LCF conditions occurs at higher applied stress amplitudes for which the cyclic plasticity becomes widespread and more homogeneously distributed. Although the focus in all stages is on dislocation plasticity, each stage demands different characteristic computational micromechanics treatment of the role of plasticity in crack formation and crack growth, as may be expected based on distinct empirical relations that have emerged for HCF and LCF.

In the HCF regime, crack nucleation and microstructurally small growth dominate the fatigue lifetime. Under HCF conditions, the cyclic plastic deformation is highly heterogeneous within the microstructure; accordingly, this is the regime in which variability and size effects are most pronounced. A strategy for computational HCF modeling of components that must last millions of cycles, such as shafts, bearings, and gears, for example, should focus on extreme value statistics of potential sites for microplastic strain localization and fracture that drive crack formation and early growth. Moreover, the issue of arrest of small cracks that form at isolated sites of cyclic plastic strain intensification is pertinent to estimation of a fatigue limit.

Table 1 outlines some of the considerations of dominant regimes of crack formation and propagation for HCF and LCF regimes defined in this manner. Microstructure-sensitive, simulation-assisted fatigue analysis is particularly useful as a means of gaining insight into fatigue responses in the HCF regime, since experimental results are often limited or inconclusive in this regime. In the VHCF regime, crack initiation can shift from the surface to the subsurface in view of multiple competing failure modes.

Small Crack Formation and Early Growth in Fatigue

There are several key points regarding the physics of crack nucleation and MSC propagation in the HCF regime. First, the slip system activity is often limited to one or two active systems, since compatibility requirements are not too demanding (elastic and plastic strains are comparable). This point can be problematic for conventional crystal plasticity models that do not focus on slip system interactions. Second, MSC growth is subject to attenuation and arrest at microstructure barriers such as grain and/or phase boundaries. In the MSC regime, cracks grow at apparent stress-intensity factor levels well below that of the long crack LEFM threshold, and considerable nonmonotonic variation (Fig. 4, 5) in crack growth rate

occurs for each crack, resulting in considerable scatter for a population of specimens. This point is clearly demonstrated for two different materials in Fig. 4 and 5. For an Fe-2 wt%Si alloy in Fig. 4 (Ref 20), certain grain boundaries correlate directly with periodic crack

growth deceleration. For Ti-6Al-4V (Ref 21, 22), as shown in Fig. 5, it was shown that oscillations of the FCGR were observed with wavelengths on the order of 7.5 μm for the cracks smaller than $a = 50$ μm, and average wavelengths of 22 μm for the cracks between $a = 50$ and 100 μm. This suggests that both grain size and microtexture (the α- and β-rich bands) are the characteristic microstructure attributes in the thickness direction affecting growth rates.

Figure 6 schematically compiles these MSC growth rate effects for a population of specimens with naturally occurring fatigue cracks in terms of resulting scatter when plotted against the conventional LEFM driving force, ΔK. Beyond the MSC regime, compounding effects of plasticity and roughness-induced closure are often considered. In the MSC regime, particularly in the transition from stage I to II growth, crack path tortuousity and roughness effects can play a role, although wake contact effects due to plastic deformation are typically quite limited. More important in the MSC regime is the anisotropy of slip and its planar nature.

Miller (Ref 23, 24) argued for recognition of multiple thresholds for fatigue, depending on crack length relative to microstructure, load history, and environment. For example, Fig. 7 shows simple fatigue crack growth relations proposed by Wang and Miller (Ref 25) for carbon steel, as supported experimentally under cyclic shear loading for crack lengths up to approximately 1 mm. Here, τ_a is the maximum shear stress amplitude; thresholds indicated by $B\tau_a^\beta a \to D$ and $(d - a) \to 0$ reflect physically small crack and microstructurally small crack thresholds, respectively, where d is the limiting microstructure barrier spacing. Figure 7 compares the characteristic predictions of MSC growth models based on continuously distributed dislocation theory (solid line) as developed by Navarro and de los Rios (Ref 26) with results of this kind of simplified model (dashed line). The key aspects of attenuation of MSC growth, elevated growth rates below the LEFM threshold, and a fatigue limit for constant amplitude loading are reproduced by the simplified model. McDowell (Ref 27, 28) further introduced extensions of these kinds of simplified models to achieve consistency with multiaxial fatigue parameters (Ref 9) and strain-life relations in tension-compression and cyclic shear, introducing multistage growth relations that eventually merge with long crack FCGR behavior.

This notion of a cascade of distinct thresholds is very important, because it arises naturally in consideration of microstructure effects. In fact, a hierarchy of thresholds can be identified. In ascending order of stress amplitude (and, to a large extent, length scale), they are listed as:

- Onset of microplasticity (or elastic shakedown)
- Crack nucleation (cyclic slip irreversibility that forms crack embryos)

Fig. 4 Variation of da/dN versus distance to initial crack tip as dictated by interaction with microstructure for a coarse-grained Fe-2wt%Si alloy. Source: Ref 20

LIVE GRAPH
Click here to view

LIVE GRAPH
Click here to view

LIVE GRAPH
Click here to view

Fig. 5 (Top) Comparisons of experimentally measured room-temperature da/dN for two different maximum stress levels and predictions using a model based on discrete dislocation mechanics. Source: Ref 21. (Bottom) Micrograph showing α- and β-rich bands in microtextured duplex Ti-6Al-4V. Source: Ref 22

Fig. 6 Characteristic variability of fatigue crack growth rate for microstructurally small cracks (MSC) in a population of specimens when plotted in terms of linear elastic fracture mechanical driving force.

- Nonpropagating MSC crack due to interaction with microstructure barriers
- Nonpropagating crack at a macronotch root
- LEFM threshold

Note that the latter two thresholds are often measured and stated within the context of LEFM analyses that consider plasticity-induced crack closure effects as well. Figure 8 illustrates these various thresholds on a Kitagawa diagram for completely reversed loading. Systematically mapping these thresholds as a function of microstructure is one of the major tasks of a hierarchical fatigue modeling strategy based on computational micromechanics. Miller (Ref 23)

further indicated the importance of considering environmental effects in overcoming or eradicating MSC thresholds, although they are not indicated on Fig. 8.

Design Against Fatigue Crack Initiation

Experiments are typically used to characterize both mean fatigue resistance and scatter in fatigue response as a function of microstructure to facilitate tailoring of microstructure to improve component-level fatigue resistance. Recently, there has been an emphasis on

applying computational micromechanics (Ref 5, 6, Ref 29–33) to hierarchical microstructures (phases, grains, inclusions, etc.) to characterize cyclic plasticity that drives crack formation at microstructure scales. Goals of such studies are varied and include:

- Reduction of the number of experiments required to assess scatter in fatigue via augmentation with computational simulations
- Parametric exploration of various fatigue damage mechanisms to support probabilistic approaches or worse-case scenarios for fatigue-life prediction schemes
- Sorting out behaviors associated with crack formation, MSC growth, and long crack growth to support the design of fatigue-resistant microstructures and process parameters (Ref 5, 6). This relies more on obtaining proper trends of models rather than absolute accuracy of correlations/predictions.
- Informing microstructure-sensitive parameters in macroscopic cyclic plasticity and fatigue models for structural applications
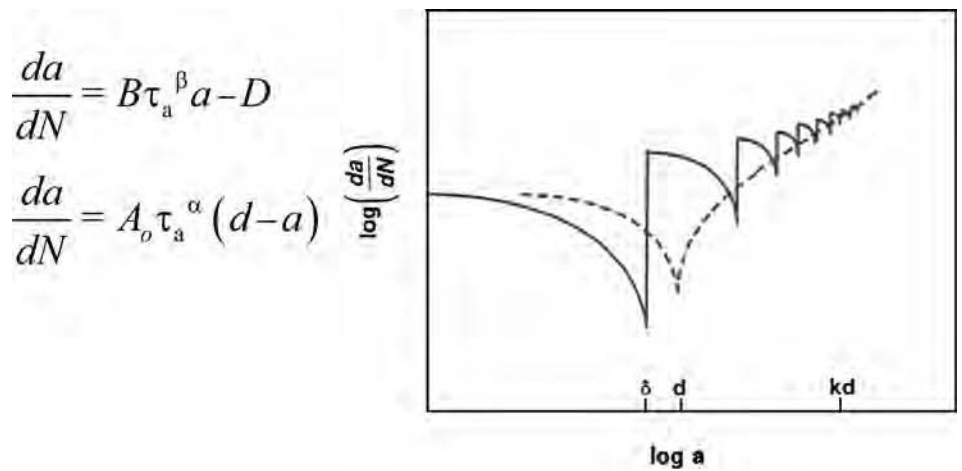
One may pursue multiple pathways in designing HCF-resistant microstructures. One way is to reduce nucleation probability though refinement and strengthening of microstructure (or nanostructure) or to decrease the number density of nucleation sites by controlling morphology. Another is to increase the threshold for MSC propagation by reducing barrier spacing, promoting extension of stage I propagation, or increasing barrier strength.

From a top-down perspective, the methodology of quantifying the effects of microstructure morphology on the mean fatigue response or variability thereof is a logical component of microstructure-sensitive design and prognosis strategies. Figure 9 provides a philosophical construct for using computational micromechanics to relate variation of microstructure (within a single sample or a population) to variability in fatigue response (or properties). We note that although a given simulation is deterministic, statistics are compiled by simulating a population of microstructures and related attributes. In the absence of robust and reliable damage-state detection methods capable of sensing and distinguishing small fatigue cracks with dimension on the order of microstructure (e.g., grain size, phase size), this kind of scheme is able to deliver information related to the expected distribution of initial cracks that would propagate to detectable lengths. Moreover, such approaches can be pursued to explore effects of microstructure attributes that contribute to extreme-value fatigue responses that populate the tails of probability distributions of surface and subsurface fatigue crack initiation sites, including transitions between mechanisms of crack formation, surface-to-subsurface transitions, and so on (Ref 34–36).

Recent research has advanced the computational modeling of detailed, specific microstructures to identify and characterize the existence

$$\frac{da}{dN} = B\tau_a^{\beta}a - D$$

$$\frac{da}{dN} = A_o\tau_a^{\alpha}(d-a)$$

Fig. 7 Simplified microstructurally small crack (MSC) and physically small crack (PSC) growth relations proposed by Wang and Miller (Source: Ref 25) addressing distinct MSC and PSC thresholds, and comparison of resulting fatigue crack growth rate behavior (dashed line) with characteristic results of continuously distributed dislocation models (solid line). Source: Ref 26
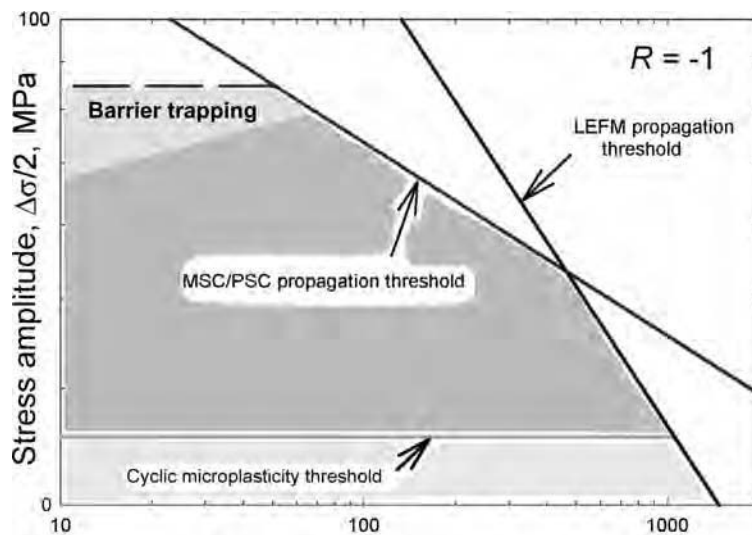
Fig. 8 Hierarchical thresholds for propagating cracks in a plot of completely reversed applied stress amplitude versus maximum inclusion size (thresholds for nucleation not shown). MSC, microstructurally small crack; PSC, physically small crack; LEFM, linear elastic fracture mechanics
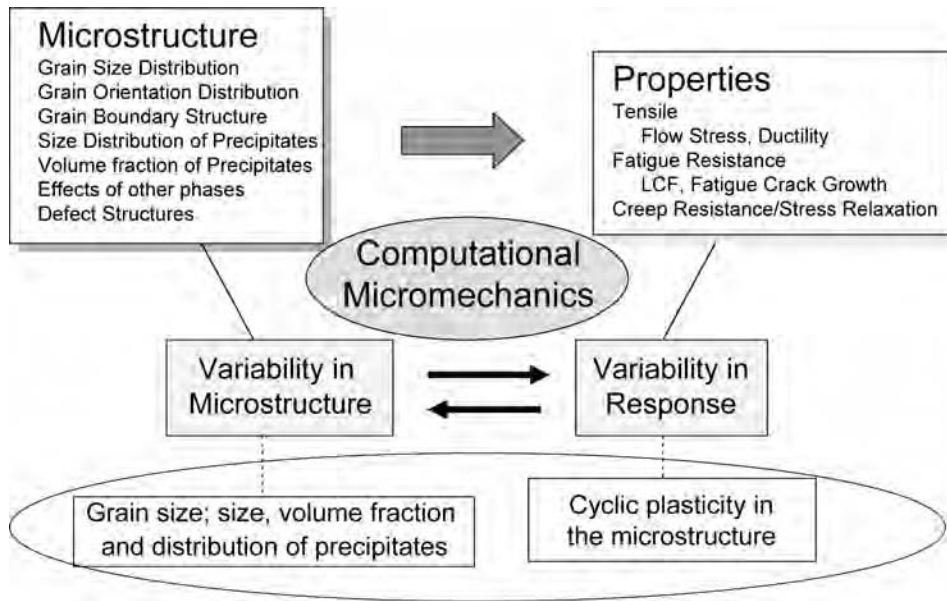
Fig. 9 Philosophical construct for simulation-assisted fatigue analysis as focused on relating variability of structure to properties or responses, with emphasis on distributions and probabilistic approaches. LCF, low-cycle fatigue

of local material hot spots controlled by intrinsic (e.g., local grain orientations) or extrinsic (distribution of inclusions or precipitates) attributes. For example, McDowell and colleagues have developed schemes by which microstructure-sensitive polycrystal plasticity models for duplex Ti-6Al-4V (Ref 37–39) and γ-γ' nickel-base superalloys (Ref 40–42) are analyzed in terms of distributions of slip and various fatigue indicator parameters (FIPs). These FIPs enable assessment of the relative potency of fatigue crack formation based on continuum-level quantities associated with each relevant mechanism. For example, the Fatemi-Socie (Ref 43, 44) shear-dominated parameter can be assessed by postprocessing numerical simulation results and has been shown to correlate very well the multiaxial fatigue crack initiation data in both LCF and HCF regimes (Ref 8). It is defined by:

$$P_{FS} = \frac{\Delta \gamma_{max}^{p^*}}{2}\left(1 + K' \frac{\sigma_{max}^n}{\sigma_y}\right) \qquad \text{(Eq 2)}$$

where $\Delta \gamma_{max}^{p^*}/2$ is the nonlocal maximum cyclic plastic shear strain averaged over a finite volume of material, σ_{max}^n and σ_y are the normal stress normal to the plane of $\Delta \gamma_{max}^{p^*}/2$ and cyclic yield strength, respectively, and K' is a constant that mediates the influence of normal stress (Ref 8, 44). Goh (Ref 45) has cast this parameter in terms of a modified Coffin-Manson law for crack formation and early growth, that is, $P_{FS} = \tilde{\gamma}_f'(2N_{inc})^c$, where $\tilde{\gamma}_f'$ is a coefficient applicable to the microstructure scale considered, and c is the Coffin-Manson exponent. The spatial volume for nonlocal averaging of the driving force is the lesser of approximately 1 μm^3 or ℓ^3, where ℓ is the scale of the transition crack length at which the crack grows out

of the influence of the microstructure notch root field (plastic strain intensification) at which it forms. The initial crack size for propagation analysis then corresponds to this scale. The precise form of the nucleation relation is system-specific in terms of slip character and slip irreversibility, comprising a topic of basic research in its own right (Ref 46–49).

Of course, classical to-and-fro slip is not responsible for all crack formation and propagation mechanisms at the microstructure scale. Cyclic plastic strain behavior is generally classified into three regimes: elastic shakedown, reversed cyclic plasticity, and plastic ratcheting (Ref 45, 50). Progressive pileup of dislocations in slip bands (Zener mechanism) that impinge on grain or phase boundaries, or at oxidized inclusion interfaces, can lead to formation and propagation of small cracks in the microstructure. In fretting fatigue, for example, progressive plastic deformation of surface layers has been shown to contribute significantly to formation and early growth (Ref 51, 52) of cracks on the order of grain size under ostensibly HCF conditions. An appropriate measure of plastic strain to reflect this sort of driving force is a nonlocal measure of the increment of cyclic rate of ratchet strain accumulation, $\Delta \gamma_{max,ratch}^{p^*}$, or its cumulative value. Manonukul and Dunne (Ref 53) and Dunne et al. (Ref 32, 54) have adopted a similar approach, with cumulative plastic strain as a measure of driving force to form subgrain size cracks in HCF of nickel-base superalloys; they have found reasonably good correlation with the sites for crack nucleation in sets of surface grains characterized by electron backscatter diffraction to render local orientation. Figure 10 illustrates FIPs that may be employed to parametrically explore driving forces for slip band cracking and slip band

impingement on grain or phase boundaries, respectively. Depending on the micromechanisms of fatigue crack nucleation and early growth, reversed plasticity or ratcheting may dominate. Accordingly, continuum measures of plastic strain to support fatigue design must be consistent with understanding of mechanisms.

Using FIPs, different microstructures can be compared in terms of their expected values and distributions of driving force(s) to form and grow small fatigue cracks. The primary objective of this kind of modeling strategy is to estimate the sensitivity of fatigue responses to microstructure variation (either within a given microstructure or among comparison microstructures), thereby using modeling and simulation to substantially leverage costly experimental characterization of the variability of fatigue life or fatigue strength in smooth and notched specimens.

It is noted that more mechanistic forms for crack formation and growth relations in the microstructure-scale regime are rapidly developing. However, the basic multistage framework in Fig. 3, when combined with mechanism-based simulations to describe cyclic plasticity in heterogeneous microstructures, can explore effects of microstructure morphology at various scales on variability in fatigue response, providing a means of rank-ordering different microstructures with regard to resistance to LCF or HCF. At present, the philosophy adopted in most work is that the formulation should be calibrated to mean fatigue behavior for selected microstructures, thereafter providing capability to assess sensitivity of fatigue-life predictions to variations of microstructure within the range of microstructures calibrated. Truly predictive methods for mean fatigue life remain as a substantial challenge for future work.

Examples of Microstructure-Sensitive Fatigue Modeling

In addition to these microstructure attributes, loading parameters such as the amplitude of the applied strain, the R-ratio, and multiaxiality can each have a significant effect on various stages of small fatigue crack formation and growth. Microstructure-sensitive modeling has recently been applied to model fatigue of several classes of advanced engineering alloys, including:

- Cast A356-T6 aluminum alloys (Ref 10, 58–60)
- γ-γ' nickel-base superalloys (Ref 40–42, 61)
- α-β titanium alloys (duplex Ti-6Al-4V) (Ref 37–39, 45, 51, 52)
- Martensitic gear steels strengthened by carburization and shot peening (Ref 62, 63)

As a specific example, the Fatemi-Socie parameter (Ref 43) P_{FS} has proven successful in multiaxial fatigue crack initiation (typically,

Prismatic slip bands of Ti-6Al

Crack formation due to intense shear along slip bands in Ti-6Al-4V

Slip band impingement on grain boundary of polycrystalline nickel

Grain interior

Normal stress

Intensive shear along slip band

$$P_{band} = \frac{\Delta \gamma_{band}^{p*}}{2} \left(1 + K_1 \frac{\sigma_n^{band*}}{\sigma_y}\right)$$

Grain boundary

Normal stress

Slip band impingement

$$P_{impinge} = \frac{\Delta \gamma_{band}^{p*}}{2} \left(1 + K_2 \frac{\sigma_n^{GB*}}{\sigma_y}\right)$$
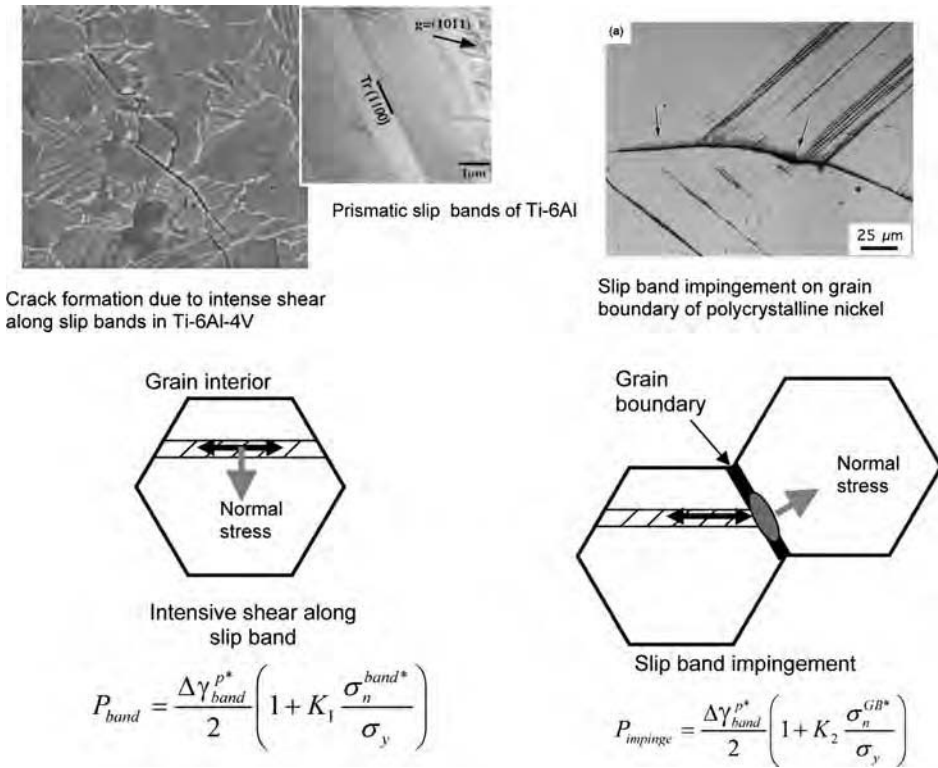
Fig. 10 Candidate fatigue indicator parameters with combined plastic shear strain and normal stress effects on candidate planes for crack nucleation and early growth along slip bands or grain boundaries for Ti-6Al and nickel. Source: Ref 55–57

Fig. 11 (Top) Pole figures for crystallographic orientation distribution. RD, rolling direction; TD, transverse direction. (Bottom) Distribution of Fatemi-Socie parameter (P_{FS}) for these textures, with loading along the transverse direction

crack lengths on the order of hundreds of microns) correlations for materials that undergo shear-dominated cracking (Ref 8, 9) and has been examined in terms of its association with mixed sliding and opening modes of small crack growth under single slip-dominated stage I fatigue (Ref 64, 65). Such measures are computed as averages over some process zone selected as representative of the scale of a critical embryo for crack nucleation, typically on the order of grain size. The parameters are assessed by postprocessing numerical simulation results. For example, the distribution of P_{FS} is shown in Fig. 11 for completely reversed cyclic straining of four different dominant textures of duplex Ti-6Al-4V alloy; note that the basal texture has a heavier tail of the distribution at higher values of P_{FS} for the same applied loading condition, indicating increased susceptiblity to fatigue crack formation with the basal texture. A tendency toward basal plane cracking in grains that are oriented for basal slip is experimentally observed (Ref 66), although there is a compounding influence of soft orientation of neighboring primary α or lamellar regions for this to occur in lieu of cracking on prismatic slip planes.

Such FIPs have been employed in modified Coffin-Manson laws for crack formation nucleation within a finite process zone. In view of its association with mixed-mode driving force for shear-dominated propagation (Ref 9, 64), it is clear that an expression such as

$P_{FS} = \tilde{\gamma}'_f (2N_{nucl})^c$ relates over scales on the order of microns to the formation of a crack embryo at much finer scales, with subcritical propagation to a length at which crack growth is amenable to fracture mechanics concepts such as the cyclic crack tip displacement range, ΔCTD (Ref 12, 13, 26, 67–69).

Prasannavenkatesan et al. (Ref 62) have found that the depth of subsurface crack nucleation at Al_2O_3 inclusion clusters in carburized and shot-peened martensitic gear steels is more accurately predicted using the P_{FS} parameter than the cyclic plastic shear strain range or cumulative shear strain. Bridier et al. (Ref 66) have found that parameters that combine effects of plastic shear strain range with normal stresses on such planes are necessary to describe individual grains where cracks nucleate and grow in duplex Ti-6Al-4V. Standard crystal plasticity models typically predict too many operative slip systems for this class of alloys, so provision must be made for preferred softening of initially activated systems due to breakdown of short-range order (Ref 39, 70); this demonstrates the importance of constitutive modeling in applying these types of FIP-based algorithms. Research on this alloy (Ref 54) has shown that nonintuitive results are sometimes observed in terms of fatigue crack initiation.

Bennett and McDowell (Ref 71) showed that martensitic-ferritic steel data for number density of lengths of surface fatigue cracks (Ref 72) up
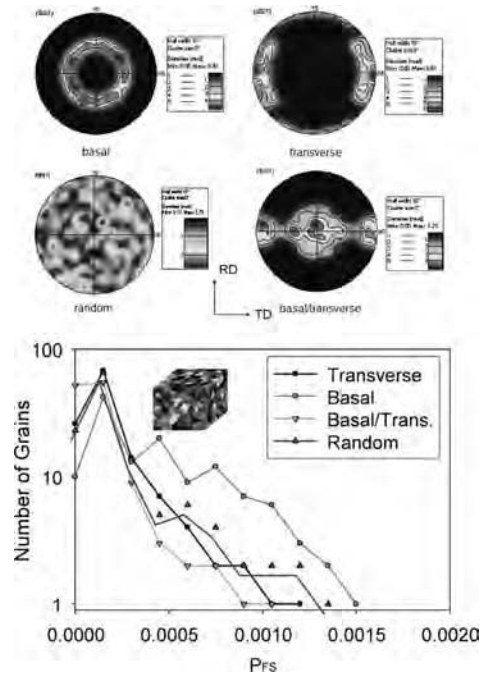
to approximately 250 μm were well described by an MSC growth law based on the P_{FS} FIP, assuming an initial crack size of 10 μm for all cracks (Fig. 12); this was in the HCF-LCF transition regime where MSC/PSC crack growth is significant.

A multistage approach has recently been used along with hierarchical polycrystal plasticity simulations for γ-γ′ nickel-base superalloys (Ref 40–42) to estimate the nucleation and growth of distributed small fatigue cracks for a range of powder metallurgy-processed IN 100 microstructures as a function of applied strain amplitude to support design of heat treatment (and associated precipitate distribution) for HCF resistance. Figure 13 illustrates the application of a combined crack nucleation and MSC growth algorithm for IN 100 (Ref 41) to estimate the fraction of grains with cracks that reach a length three times mean grain size (40 μm) as a function of applied loading cycles (lower right) for multiple statistical realizations of equiaxed polycrystalline IN 100 microstructures with log normal grain size distribution.

The field of microstructure-sensitive crystal-plasticity relations is rapidly evolving. Indeed, efforts at the Air Force Research Laboratory (Ref 74, 75) and elsewhere (Ref 30) have high value in leading toward next-generation models. Moreover, other nonlocal FIPs can be employed that account for effects of directional strain accumulation (e.g., pileups) at grain or phase boundaries (Ref 45), as shown in Fig. 10. A number of authors in the literature
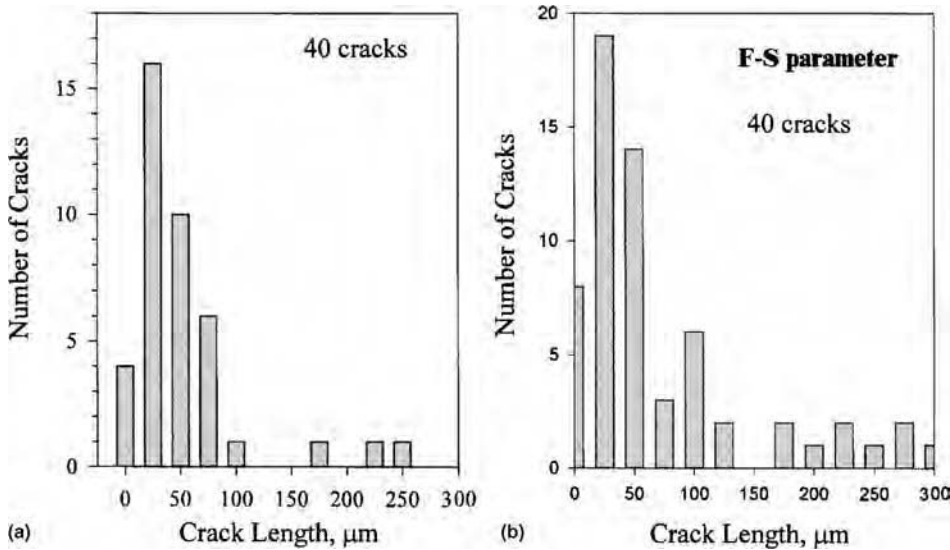
Fig. 12 Comparison of (a) measured number density of surface cracks of different lengths in low-cycle fatigue of polycrystalline martensitic-ferritic steel (Source: Ref 72) with (b) predicted distribution, assuming initial crack size of 10 μm in each grain, propagated using the fatigue indicator parameter relation $da/dN = C(\text{av } P_{FS})a$
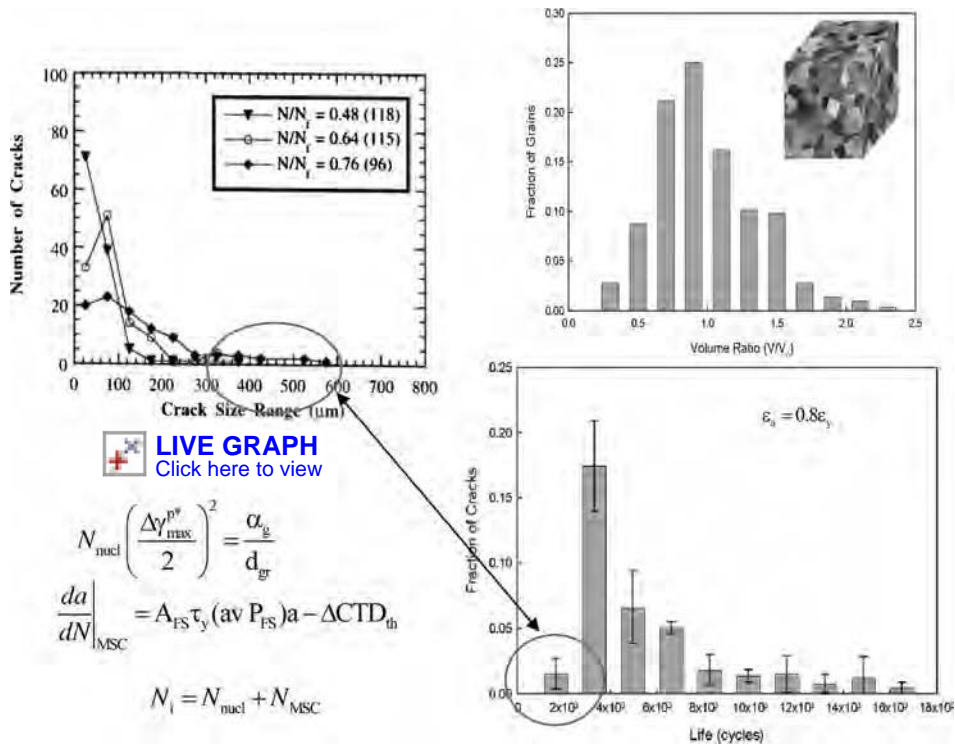
Fig. 13 Estimated distribution for fraction of grains with cracks that reach a length three times mean grain size (40 μm) as a function of applied loading cycles (lower right) for multiple statistical realizations of equiaxed polycrystalline IN 100 microstructures with log normal grain size distribution (top right) subjected to completely reversed strain-controlled cycling ($R_\varepsilon = -1$) at $T = 650$ °C (Source: Ref 41). Ellipses encapsulate the tails of the distributions that correspond to the worst-case cracking scenarios; the data at upper left (Source: Ref 73) are for another nickel-base alloy system, René 88, and pertain to the number of cracks observed as a fraction of the total fatigue life.

have appealed to FIPs that reflect cumulative plastic strain, regardless of sign, as a measure of progression to the point of crack nucleation in grains in polycrystals (Ref 32, 53, 54). In fact, a number of parameters suitable for multiaxial loading have been proposed and employed (Ref 76, 77) at the grain scale, some with more intimate potential link to fine-scale driving forces for crack nucleation and early propagation. Following McDowell and Berard

(Ref 64), McDowell and Bennett (Ref 65) made the connection more strictly to the cyclic crack tip displacement range, ΔCTD, as suggested by the growth law used in the MSC regime for IN 100 in Fig. 13. Moreover, numerous experiments and modeling efforts in the literature (Ref 4) point to the efficacy of ΔCTD (opening and sliding) to correlate crack advance (Ref 12, 13, 67–69).

Figure 14 summarizes some interesting recent work of Kunkler et al. (Ref 7) in which two-dimensional (2-D) grain-level simulations of duplex steel are modeled using strip yield-zone models based on the continuously distributed dislocation theory coupled with the crack tip sliding displacement range (ΔCTSD) to model mixed transgranular-intergranular MSC growth in polycrystals; the prediction of high growth rates in the MSC regime well below the LEFM threshold for long cracks is reasonably accurate. They assume that crack advance occurs by virtue of unbalanced dislocation nucleation and migration at the crack tip during forward and reverse loading (dislocations of opposite sign are involved). As in Fig. 3, the exponent on ΔCTSD in the growth law is unity. Arbitrary planes of crack extension are allowed by the use of sensor elements. Of course, such analysis schemes are constrained by the assumption of continuously distributed dislocation theory, essentially an extension of the simple model of Navarro and de los Rios (Ref 26) to a more complex 2-D microstructure. Moreover, the elements of the MSC and PSC regimes in Fig. 14 are essentially the same as those of the heuristic framework outlined in Fig. 3.

Closure—Challenges for Microstructure-Sensitive Fatigue Modeling

With increased emphasis on reducing materials development and insertion cycle times and design of reliable fatigue-resistant materials, modeling and simulation come to the forefront. Achieving enhanced predictive capability is a major current research driver for simulation-based design. There are clearly a number of approximate assumptions invoked in the foregoing relations for crack nucleation and subcritical growth up to the scale of a microstructurally small crack. Several compelling avenues for future development are:

- Crack nucleation is material-specific and predictive relations are lacking.
- Crack growth is related to irreversibility of slip arising from nucleation, migration, and interaction of point and line defects, yet continuum crystal-plasticity slip measures do not reflect this irreversibility, nor do they employ any predictive means to account for it.
- Grain boundaries are treated largely as compatibility surfaces (mechanical effects of misorientation) without acknowledging
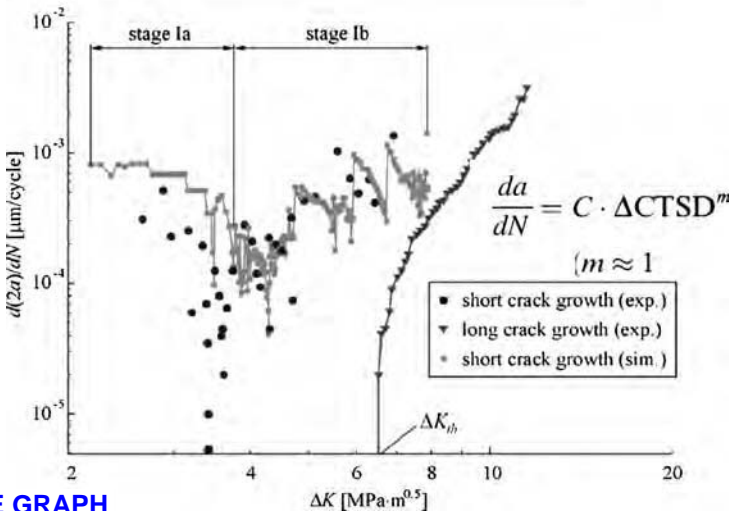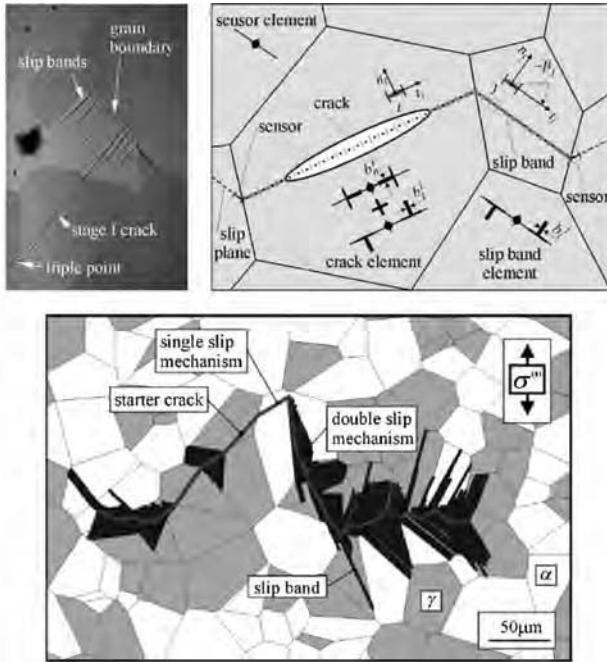
Fig. 14 Use of slip-system yield-strip models based on continuously distributed dislocation theory to model microstructurally small crack growth in austenitic/ferritic (*c*/*a*) duplex steel (*X*2CrNiMoN 22 5 3) (Source: Ref 7), reproducing key phenomena of the stage I–II transition and including slip-band plasticity and roughness-induced closure effects as the crack lengthens

Fig. 15 Domain decomposition of fatigue damage process zone at tip of a fatigue crack, indicating domains for atomistics, discrete dislocations, and continuum models

effects of their structure on dislocation mediation. Grain-boundary engineering is appealing as a context for design using microstructure-sensitive fatigue. Gao et al. (Ref 78) have shown, for example, that grain-boundary-engineered polycrystals of ME3 nickel-base superalloy can achieve reduced average crack growth rates in the MSC regime compared to as-received material. Existing methods to account for grain boundaries are of simplified nature.

- Strain localization processes associated with dislocation substructure formation (e.g., slip bands) are not treated explicitly in the continuum representation of slip. The relative ratios of transgranular and intergranular

crack nuclei cannot presently be predicted nor can the fractions of transgranular and intergranular growth paths in the MSC regime. Calibration is necessary.

- Criteria for growth of microstructurally small fatigue cracks, including dependence on stress range and *R*-ratio, are of phenomenological character and reflect influence of microstructure morphology rather than incorporating material-specific details of irreversible processes of dislocation and point defect nucleation and migration near the tip of slip bands and cracks, as well as discrete dislocation effects.
- HCF and particularly VHCF are characterized as extreme value problems in terms of

the relevant statistics that govern reliability. Multivariate extreme value statistics are necessary to address these issues and are not presently well developed.

- Crack nucleation and growth from micron-scale inclusions or pores within the microstructure do not address the important role of discrete dislocations at such scales.

A challenging, high-payoff objective for multiscale modeling is the growth of small cracks, likely requiring linkage of atomistic and discrete dislocation scales. Multiscale modeling approaches that employ domain decomposition of models with resolution ranging from atomic-level fidelity to representative volumes of microstructure are necessary to shed light on these issues (Fig. 15). Atomistic models can shed light on nucleation relations (with an important role of atomistics) and computation of the irreversibility factor, ϕ, in the microscopic MSC growth relation:

$$\frac{da}{dN}\bigg|_{\text{MSC}} = \phi \, \Delta\text{CTD} - \eta b \qquad \text{(Eq 3)}$$

which has generally eluded treatment from a continuum perspective. Parameter ϕ carries the "imprint" of atomic-scale processes of new crack surface creation, owing to slip irreversibilities, and can affect interactions with precipitates and impurities. The threshold ηb here is a conceptual form that characterizes the level of irreversible range of the ΔCTD below which the crack should not extend, where b is the magnitude of the Burgers vector—we reckon η to be on the order of unity (Ref 10). This is a rich problem domain for computational materials science and mechanics.

ACKNOWLEDGMENTS

contributions along the way by graduate students J.-Y. Berard, V. Bennett, C.-H. Goh, M. Shenoy, M. Zhang, and R. Prasannavenkatesan, as well as postdoctoral students A. Wang, R. Kumar, J. Zhang, and F. Bridier.

REFERENCES

1. J.A. Bannantine, J.J. Comer, and J.L. Handrock, *Fundamentals of Metal Fatigue Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey, 1990

2. S. Suresh, *Fatigue of Materials*, Cambridge University Press, 2nd Ed., Cambridge, UK, 1998

3. D. Taylor and J.F. Knott, *Fatigue Crack Propagation Behaviour of Short Cracks: The Effect of Microstructure,* Fatigue Fract. Engng. Mater. Struct., Vol 4, No. 2, 1981, p. 147–155

4. D.L. McDowell, Basic Issues in the Mechanics of High Cycle Metal Fatigue, *Int. J. Fracture*, Vol 80, 1996, p. 103–145

5. D.L. McDowell, Microstructure-Sensitive Computational Fatigue Analysis, *Handbook of Materials Modeling*, Part A: Methods, Sidney Yip and M.F. Horstemeyer, eds., Springer, The Netherlands, 2005, p.1193–1214

6. D.L. McDowell, Simulation-Based Strategies for Microstructure-Sensitive Fatigue Modeling, *Mater. Sci. Eng. A*, Vol 468–470, 2007, p. 4–14

7. B. Kunkler, O. Duber, P. Koster, U. Krupp, C.-P. Fritzen, and H.-J. Christ, Modelling of Short Crack Propagation – Transition from Stage I to Stage II, *Engng Fract Mech.,* Vol 75, 2008, p. 715–725

8. D.F. Socie, Critical Plane Approaches for Multiaxial Fatigue Damage Assessment, *Advances in Multiaxial Fatigue*, D.L. McDowell and R. Ellis, eds., ASTM STP 1191, ASTM, Philadelphia, 1993, p. 7–36

9. D.L. McDowell, Multiaxial Fatigue Strength, *ASM Handbook*, Vol 19 on Fatigue and Fracture, ASM International, 1996, p. 263–273

10. D.L. McDowell, K. Gall, M.F. Horstemeyer, and J. Fan, Microstructure-Based Fatigue Modeling of Cast A356-T6 Alloy, *Eng. Fract. Mech.*, Vol 70, 2003, p. 49–80

11. F.A. McClintock, Considerations for Fatigue Crack Growth Relative to Crack Tip Displacement, in: J.H. Beynon, M.W. Brown, T.C. Lindley, R.A. Smith, B. Tomkins (Eds.), *Engineering Against Fatigue*, Balkema Press, Chapter 24, 1999, p. 227–241

12. C. Li, On the Interaction Among Stage I Short Crack, Slip Band and Grain Boundary: A FEM Analysis, *Int. J. Fracture,* Vol 43, 1990, p. 227–239

13. C. Li, A Three-Dimensional Finite Element Analysis for Crystallographic Crack Near the Interface of an Incompatible Bicrystal, *Fatigue Fract. Engng. Mater. Struct.*, Vol 16, No. 1, 1992, p. 21–35

14. K. Tanaka, Y. Akiniwa, Y. Nakai, and R.P. Wei, Modeling of Small Fatigue Crack-Growth Interacting with a Grain-Boundary, *Engng. Fract. Mech.* Vol 24, No. 6, 1986, p. 803–819

15. Z. Sun, E.R. de los Rios, and K.J. Miller, Modeling Small Fatigue Cracks Interacting with Grain Boundaries, *Fatigue Fract. Engng. Mater. Struct.* Vol 14, No. 2/3, 1991, p. 277–291

16. K. Gall, H. Sehitoglu, and Y. Kadioglu, A Methodology for Predicting Variability in Microstructurally Short Fatigue Crack Growth Rates, *ASME J. Engng.. Mater. Techn.,* Vol 119, 1997, p. 171–179

17. V.P. Bennett and D.L. McDowell, Crystal Plasticity Analyses of Stationary, Microstructurally Small Surface Cracks in Ductile Single Phase Polycrystals, *Fatigue Fract. Engng. Mater. Struct.,* Vol 25, No. 7, 2002, p. 677–693

18. V.P. Bennett and D.L. McDowell, Crack Tip Displacements of Microstructurally Small Surface Cracks in Single Phase Ductile Polycrystals, *Engng. Fract. Mech.* 70, 2003, p. 185–207

19. V. Bennett and D.L. McDowell, Polycrystal Orientation Effects on Microslip and Mixed Mode Behavior of Microstructurally Small Cracks, in: K.J. Miller and D.L. McDowell (Eds.), Mixed-Mode Crack Behavior, ASTM STP 1359, ASTM, 1999, p. 203–228

20. S.S. Chakravarthula and Y. Qiao, Fatigue Crack Growth in a Coarse-Grained Iron-Silicon Alloy, *Int. J. Fatigue*, Vol 27, 2005, p. 1210–1214

21. R.W. Hamm and W.S. Johnson, A Unification of Small- and Large-Crack Growth Laws, *Fatigue Fract.Engng. Mater. Struct.*, Vol 22, No. 8, 1999, p. 711–722

22. O. Jin, R.W. Hamm, and W.S. Johnson, Microstructural Influences on the Growth of Small Cracks in Ti-6Al-4V, *Fatigue Fract.Engng. Mater. Struct.*, Vol 25, No. 6, 2002, p. 563–574

23. K.J. Miller, Materials Science Perspective of Metal Fatigue Resistance, *Mater Sci Techn.*, Vol 9, 1993, p. 453–462

24. K.J. Miller, The Two Thresholds of Fatigue Behaviour, *Fatigue Fract. Eng. Mat. Struct.*, Vol 16, No. 9, 1993, p. 931–939

25. C.H. Wang and K.J. Miller, The Effects of Mean and Alternating Shear Stresses on Short Fatigue Crack Growth Rates, *Fatigue Fract. Engng. Mater. Struct.,* Vol 15, No. 12, 1992, p. 1223–1236

26. A. Navarro and E.R. de los Rios, An Alternative Model of the Blocking of Dislocations at Grain Boundaries, *Phil. Mag.*, Vol 57, 1988, p. 37–50

27. D.L. McDowell, Multiaxial Small Fatigue Crack Growth in Metals, *Int. J. Fatigue*, Vol 19, No. 1, 1997, p. S127–S135

28. D.L. McDowell, An Engineering Model for Propagation of Small Cracks in Fatigue, *Eng. Fract. Mech.*, Vol 56, No. 3, 1997, p. 357–377

29. T. Hoshide and D.F. Socie, Crack Nucleation and Growth Modeling in Biaxial Fatigue, *Engng Fract Mech*, Vol 29, No. 3, 1988, p. 287–299

30. A. Manonukul, F.P.W. Dunne, and D. Knowles, Physically-Based Model for Creep in Nickel-Base Superalloy C263 Both Above and Below the Gamma Solvus, *Acta Mater.*, Vol 50, 2002, p. 2917–2931

31. R. Morrissey, C.-H. Goh, and D.L. McDowell, Microstructure-Scale Modeling of HCF Deformation, *Mech. Materials*, Vol 35, Nos. 3–6, 2003, p. 295–311

32. F.P.E. Dunne, A.J. Wilkinson, and R. Allen, Experimental and Computational Studies of Low Cycle Fatigue Crack Nucleation in a Polycrystal, *Int. J. Plasticity*, Vol 23, 2007, p. 273–295

33. D.L. McDowell, Materials Design: A Useful Research Focus for Inelastic Behavior of Structural Metals, in: G.C. Sih and V.E. Panin (Eds.), Special Issue of the Theoretical and Applied Fracture Mechanics, Prospects of Mesomechanics in the 21st Century: Current Thinking on Multiscale Mechanics Problems, Vol 37, 2001, p. 245–259

34. S.K. Jha, M.J. Caton, J.M. Larsen, A.H. Rosenberger, K. Li, and W.J. Porter, Superimposing Mechanisms and Their Effect on the Variability in Fatigue Lives of a Nickel-Based Superalloy, Materials Damage Prognosis, TMS Publications, 2004, p. 343–350

35. S.K. Jha, M.J. Caton, and J.M. Larsen, A New Paradigm of Fatigue Variability Behavior and Implications for Life Prediction, *Mater. Sci. Eng. A*, Vol 468–470, 2007, p. 23–32

36. S.K. Jha and J.M. Larsen, Random Heterogeneity Scales and Probabilistic Description of Long-Lifetime Regime of Fatigue, VHCF-4, Proceedings of the Fourth International Conference on Very High Cycle Fatigue, 2007, p. 385–396

37. J.R. Mayeur, Three Dimensional Modeling of Ti-Al Alloys with Application to Attachment Fatigue, M.S. Thesis, GWW School of Mechanical Engineering, Georgia Institute of Technology, 2004

38. J.R. Mayeur and D.L. McDowell, A Three-Dimensional Crystal Plasticity Model for Duplex Ti-6Al-4V, *Int. J. Plasticity*, Vol 23, No. 9, 2007, p. 1457–1485

39. M. Zhang, J. Zhang, and D.L. McDowell, Microstructure-Based Crystal Plasticity Modeling of Cyclic Deformation of Ti-6Al-4V, *Int. J. of Plasticity*, Vol 23, No. 8, 2007, p. 1328–1348

40. R.S. Kumar, A.-J. Wang, and D.L. McDowell, Effects of Microstructure Variability on Intrinsic Fatigue Resistance of Nickel-Base Superalloys – A Computational Micromechanics Approach, *Int. J. Fracture*, Vol 137, 2006, p. 173–210

41. M. Shenoy, J. Zhang, and D.L. McDowell, Estimating Fatigue Sensitivity to Polycrystalline Ni-Base Superalloy Microstructures

Using a Computational Approach, *Fatigue Fract. Eng. Mat. Struct.*, Vol 30, No. 10, 2007, p. 889–904

42. M. Shenoy, Y. Tjiptowidjojo, and D.L. McDowell, Microstructure-Sensitive Modeling of Polycrystalline IN 100, *Int. J. Plasticity*, Vol 24, 2008, p 1694–1730

43. A. Fatemi and D.F. Socie, Critical Plane Approach to Multiaxial Fatigue Damage Including Out-of-Phase Loading, *Fatigue Fract Engng Mater Struct.,* Vol 11, No. 3, 1988, p. 149–165

44. A. Fatemi, and P. Kurath, Multiaxial Fatigue Life Predictions Under the Influence of Mean-Stresses, *ASME J. Engng. Mater. Techn.*, Vol 110, 1988, p. 380–388

45. C.-H. Goh, Crystallographic Plasticity in Fretting of Ti-6Al-4V, Ph.D. Thesis, Georgia Institute of Technology, 2002

46. G. Venkataraman, Y.W. Chung, and T. Mura, Application of Minimum Energy Formalism in a Multiple Slip Band Model for Fatigue—II. Crack Nucleation and Derivation of a Generalized Coffin-Manson Law, *Acta Metall. Mater.* Vol 39, 1991, p. 2631–2638

47. K. Tanaka and T. Mura, A Dislocation Model for Fatigue Crack Initiation, *ASME J. Appl. Mech.,* Vol 48, 1981, p. 97–103

48. L. Rémy, H. Bernard, J.L. Malpertu, and F. Rezai-Aria, Fatigue Life Prediction Under Thermal-Mechanical Loading in a Nickel-Base Superalloy, in *Thermomechanical Fatigue Behavior of Materials*, ASTM STP 1186, H Sehitoglu, Ed., ASTM, West Conshohocken, PA, 1993, p. 3–16

49. E.A. Repetto and M. Ortiz, A Micromechanical Model of Cyclic Deformation and Fatigue Crack Nucleation in F.C.C. Single Crystals, *Acta Mater.* Vol 45, No. 6, 1997, p. 2577–2595.

50. J.M. Ambrico and M.R. Begley, Plasticity in Fretting Contact, *J. Mech. Phys. Solids*, Vol 48, No. 11, 2000, p. 2391–2417

51. C.-H. Goh, R.W. Neu, and D.L. McDowell, Influence of Nonhomogeneous Material in Fretting Fatigue, ASTM STP 1425, *Fretting Fatigue: Advances in Basic Understanding and Applications*, 2002, p. 183–205

52. D.R. Swalla, R.W. Neu, and D.L. McDowell, Microstructural Characterization of Ti-6Al-4V Subjected to Fretting, *ASME J. Tribology*, Vol 126, 2004, p. 809–816

53. A. Manonukul and F.P.E. Dunne, High- and Low-Cycle Fatigue Crack Initiation Using Polycrystal Plasticity, *Proc. Royal Soc. Lond. A,,* Vol 460, Issue 2047, 2004, p. 1881–1903

54. F.P.E. Dunne, D. Rugg, and A. Walker, Length Scale-Dependent, Elastically Anisotropic, Physically-Based HCP Crystal Plasticity: Application to Cold-Dwell Fatigue in Ti Alloys, *Int. J. Plasticity*, Vol 23, 2007, p. 1061–1083

55. K. Le Biavant, S. Pommier, and C. Prioul, Local Texture and Fatigue Initiation in a Ti-6Al-4V Titanium Alloy, *Fatigue Fract. Eng. Mater. Struct.*, Vol 25, 2002, p. 527–545

56. T. Neeraj and M.J. Mills, Short-Range Order (SRO) and its Effect on the Primary Creep Behavior of a Ti-6wt.% Al Alloy, *Mat. Sci. Eng. A*, Vol 319–321, 2001, p. 415–419

57. D.J. Morrison and J.C. Moosbrugger, Effects of Grain Size on Cyclic Plasticity and Fatigue Crack Initiation in Nickel, *Int. J. Fatigue*, Vol 19, 1997, p. S51–S59

58. J. Fan, D.L. McDowell, M.F. Horstemeyer, and K. Gall, Computational Micromechanics Analysis of Cyclic Crack-Tip Behavior for Microstructurally Small Cracks in Dual-Phase Al-Si Alloys, *Engng. Fract. Mech.*, Vol 68, 2001, p. 1687–1706

59. K. Gall, N. Yang, M.F. Horstemeyer, D.L. McDowell, and J. Fan, The Influence of Modified Intermetallics and Si Particles on Fatigue Crack Paths in a Cast A356 Al Alloy, *Fatigue Fract. Engng. Mater. Struct.*, Vol 23, No. 2, 2000, p. 159–172

60. K. Gall, M.F. Horstemeyer, D.L. McDowell, and J. Fan, Finite Element Analysis of the Stress Distributions Near Damaged Si Particle Clusters in Cast Al-Si Alloys, *Mech. Materials*, Vol 32, No. 5, 2000, p. 277–301

61. M. Shenoy, R.S. Kumar, and D.L. McDowell, Modeling Effects of Nonmetallic Inclusions on LCF in DS Nickel-Base Superalloys, *Int. J. Fatigue*, Vol 27, 2005, p. 113–127

62. R. Prasannavenkatesan, J. Zhang, D.L. McDowell, G.B. Olson, and H.-J. Jou, 3D Modeling of Subsurface Fatigue Crack Nucleation Potency of Primary Inclusions in Heat Treated and Shot Peened Martensitic Gear Steels, *Int. J. Fatigue*, Vol 31 No. 7, 2009, p 1176–1189

63. D.L. McDowell and G.B. Olson, Concurrent Design of Hierarchical Materials and Structures, *Scientific Modeling and Simulation* (CMNS), Vol 15 (No. 1), 2008, p 207

64. D.L. McDowell and J.-Y. Berard, A ΔJ-Based Approach to Biaxial Fatigue, *Fatigue Fract. Eng. Mat. Struct.*, Vol 15, 1992, p. 719–741

65. D.L. McDowell and V.P. Bennett, Micromechanical Aspects of Small Multiaxial Fatigue Cracks, Proc. 5th Int. Conf. On Biaxial/Multiaxial Fatigue & Fracture, Cracow, Poland, 1997, p. 325–348

66. F. Bridier, P. Villechaise, and J. Mendez, Slip and Fatigue Crack Formation Processes in an α-β Titanium Alloy in Relation to Crystallographic Texture on Different Scales, *Acta Mater.*, doi:10.1016/j.actamat.2008.04.036, 2008

67. K. Chan, J. Lankford, and D. Davidson, A Comparison of Crack-Tip Field Parameters for Large and Small Fatigue Cracks, *ASME J. Engng. Mater. Techn.*, Vol 108, 1986, p. 206–213

68. K. Hussain, E.R. de los Rios, and A. Navarro, A Two-Stage Micromechanics Model for Short Fatigue Cracks,. *Engng Fract Mech,* Vol 44, No. 3, 1993, p. 425–436

69. C. Li, Vector CTD Analysis for Crystallographic Crack Growth, *Acta Met Mater.*, Vol 38, No. 11, 1990, p. 2129–2134

70. M. Zhang, F. Bridier, P. Villechaise, J. Mendez, and D.L. McDowell, Simulation of Slip Band Evolution in Duplex Ti-6Al-4V, *Mater. Sci. Eng. A,* submitted July 2008

71. V.P. Bennett and D.L. McDowell, Polycrystal Orientation Distribution Effects on Microslip in High Cycle Fatigue, *Int. J. Fatigue*, Vol 25, 2003, p. 27–39

72. K.J. Miller, The Short Crack Problem, *Fatigue Fract. Eng. Mat. Struct.*, Vol 5, No. 3, 1982, p. 223–232

73. E. Chen, S. Sauer, M. Meshii, and W.T. Tucker, Fatigue Microcrack Distribution and Reliability of a Ni-Base Superalloy, *Int. J. Fatigue*, Vol 19, No. 1, 1997, p. S75–S82

74. Y.S. Choi, D.M. Dimiduk, M.D. Uchic, and T.A. Parthasarathy, A Crystallographic Constitutive Model for Ni_3Al (L_{12}) Intermetallics, *Mater. Sci. Eng. A*, Vol 400–401, 2005, p. 256–259

75. Y.S. Choi, T.A. Parthasarathy, D.M. Dimiduk, and M.D. Uchic, Microstructural Effects in Modeling the Flow Behavior of Single-Crystal Superalloys, *Met. Mater. Trans. A*, Vol 37A, 2006, p. 545–550

76. T. Hoshide and D.F. Socie, Mechanics of Mixed Mode Small Fatigue Crack Growth, *Engng Fract. Mech*, Vol 26, No. 6, 1987, p. 842–850

77. K.O. Findley and A. Saxena, Low Cycle Fatigue in Rene 88DT at 650°C: Crack Nucleation Mechanisms and Modeling, *Met. Mater. Trans.,* Vol 37A, No. 5, 2006, p. 1469–1475

78. Y. Gao, J.S. Stolken, M. Kumar, and R.O. Ritchie, High-Cycle Fatigue of Nickel-Base Superalloy Rene' 104 (ME3): Interaction of Microstructurally Small Cracks with Grain Boundaries of Known Character, *Acta Mater.,* Vol 55, 2007, p. 3155–3167

Modeling Creep Fatigue

Jeffrey L. Evans, University of Alabama in Huntsville
Ashok Saxena, University of Arkansas

THERE ARE NUMEROUS APPLICATIONS in the modern engineering world that involve the use of metals under conditions of cyclic loading in operating conditions that can cause creep and/or environmental interactions with time-independent, mechanical fatigue processes. The interrelationships between the various damage mechanisms that occur under such severe service conditions are complex. Thus, the development of physics-based models for predicting design or remaining life must be guided by experimental studies that are specifically aimed at fundamental understanding of these mechanisms. These experiments are challenging and demand precise environmental-control capability and extremely high resolution in the measurement of displacements in cracked bodies associated with creep strains and in the measurement of crack increments at high temperatures. Such experimental capabilities and the concepts of nonlinear fracture mechanics that can account for crack-tip creep strains have only recently become available (Ref 1). Also, simulation tools are now available that enable the numerical modeling of the aforementioned complex interactive phenomena that can aid in extracting the most information out of these carefully planned experiments.

Recent advancements in analytical, experimental, and computational techniques can be used in concert to enable very significant advances in the understanding of these life-limiting considerations that are currently not well understood. Much of the recent progress in understanding these interactions has been in the study of materials used in gas turbine machines. Turbine disks are subjected to temperatures in the range of 650 to 700 °C for sustained periods during normal operation. The typical duty cycle for a disk is modeled as a fatigue stress cycle with dwell or hold time for the purposes of predicting crack initiation and propagation life. These operating temperatures are in the range where several time-dependent damage mechanisms can exist during the cyclic loading and during the dwell period and are therefore a design and operational concern. Stress relaxation occurs during cruise due to creep deformation, while simultaneously, oxidation in the crack-tip region and oxygen

ingress along grain boundaries occur that, in addition to causing material damage, also lead to microstructural changes. Each of the distinct time-dependent damage mechanisms is expected to have its unique kinetics that are based on the characteristics of the rate-controlling step. Thus, to accurately predict the dwell-time effects on the fatigue crack growth behavior in turbine disk materials at elevated temperatures, accurate descriptions of the kinetics of each damage mechanism are needed.

Modeling Methodology

Over the past two decades, several experimental studies have been conducted to characterize the elevated-temperature crack growth behavior in nickel-base superalloys, focusing on the effects of loading frequency, temperature, microstructure, and dwell time, but few have attempted to systematically quantify the effects of environment on the kinetics of crack growth at elevated temperatures. Even fewer have attempted to systematically study the kinetics of the various synergistic and competing time-dependent mechanisms. Also, almost all the data have been generated within the framework of linear elastic fracture mechanics (LEFM), which is unable to explicitly account for the effects of creep deformation that may play a significant role in determining the crack growth behavior and its interaction with microstructure and environment.

Figure 1 shows a schematic of a comprehensive methodology for ensuring structural integrity of elevated-temperature components. The importance of the role of crack growth models that apply to creep-fatigue-environment conditions while also accounting for microstructural changes is apparent. If more accurate crack growth models were available, Fig. 1 illustrates how they can be interfaced with constitutive and crack initiation models to significantly impact the ability to predict component behavior during service and in the design of better materials.

Analytical Models. The overarching goal of current life-prediction research is to develop models for the various types of time

dependencies in the crack-tip damage accumulation that occur in materials subjected to elevated temperatures. These consist of:

- Evolution of crack-tip stress fields due to creep
- Oxidation kinetics
- Oxygen ingress
- Change in microstructure

In the subsequent discussion, each of these is considered.

Time-Dependent Damage Evolution

In the past, some limited elevated-temperature hold-time studies have been performed in vacuum to investigate the relative contribution of the environment and creep deformation to the crack growth rate. The results of these studies are briefly described to illustrate how useful some of these experiments can be if performed now with access to far superior experimental and simulation capabilities as well as much better access to time-dependent fracture mechanics concepts.

Figure 2 schematically shows the deformation/damage zones ahead of a crack tip at elevated temperature (Ref 3). The three zones consist of the plastic zone, a creep zone, and an environmentally affected zone. The plastic zone remains fixed with time for a stationary crack, while the environmentally affected zone and the creep zone grow with time. The relative kinetics of the two processes must determine the one that dominates. If one assumes that environmental damage dominates and determines the time-dependent crack growth rates, with the rates being controlled by some diffusion-related phenomena, the following equation has been formulated to represent the dwell-time effects (Ref 3):

$$\frac{da}{dN} = c_1 (\Delta K)^{n_1} + c_2 (\Delta K)^{n_2} \left(\sqrt{t_h} \right) \qquad \text{(Eq 1)}$$

where a is crack size; N is fatigue cycles; ΔK is the cyclic stress-intensity parameter; n_1, n_2, c_1, c_2 are regression constants obtained from crack
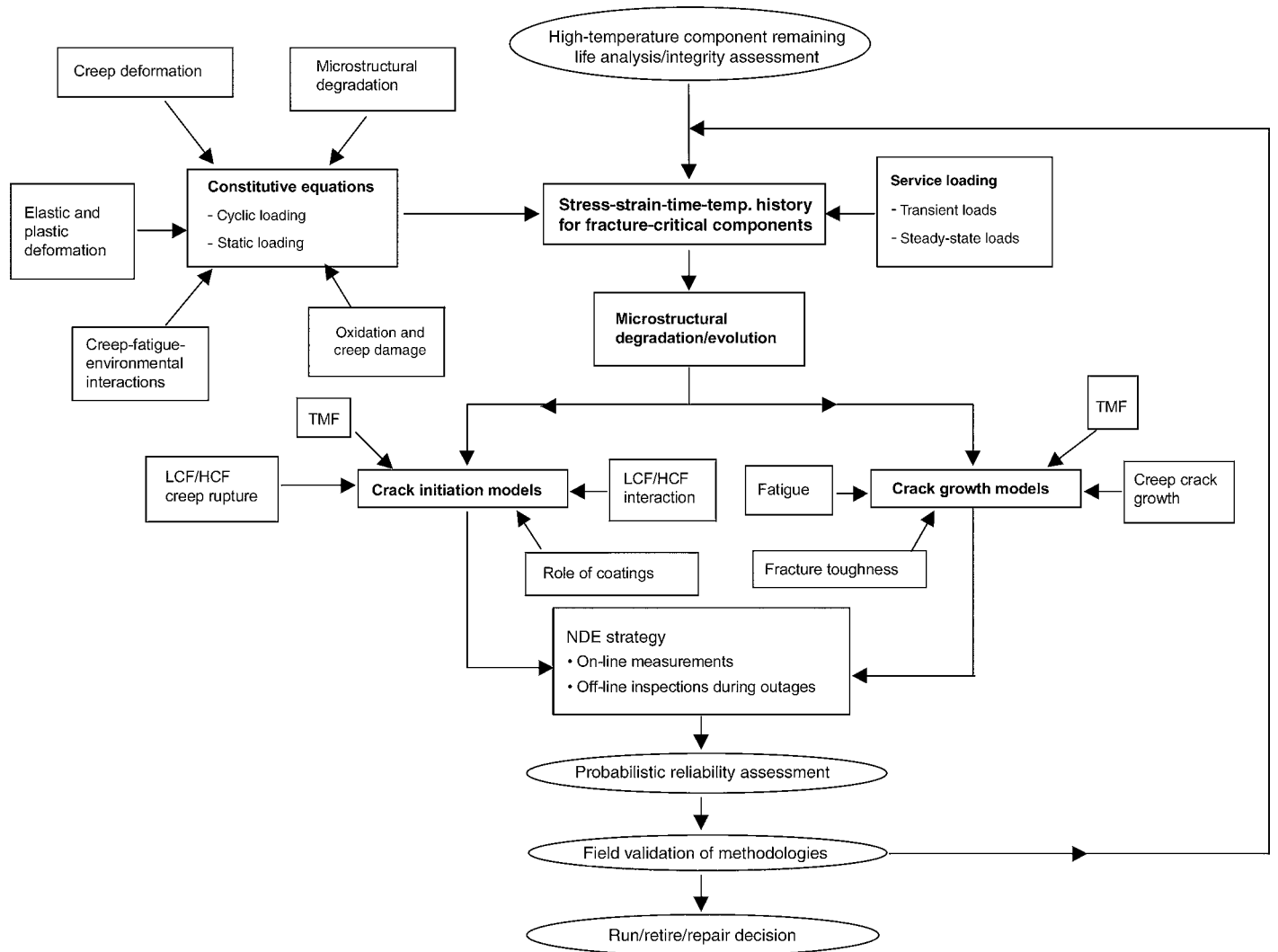
Fig. 1 Methodology for assessing integrity of structural components that operate at high temperatures. TMF, thermomechanical fatigue; NDE, nondestructive evaluation; LCF, low-cycle fatigue; HCF, high-cycle fatigue. Source: Ref 2
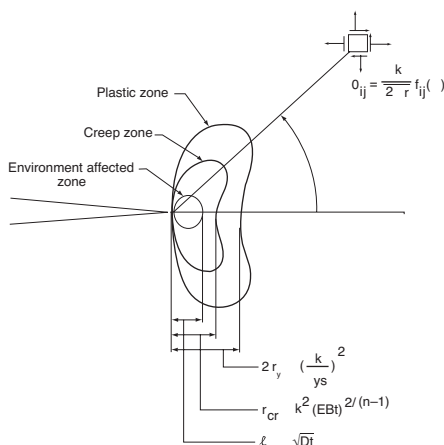
Fig. 2 Schematic of the various zones in the region of a stationary crack showing the various deformation and damage zones ahead of the crack tip. Source: Ref 3

growth data obtained using a fast, continuous fatigue cycle; and t_h is the dwell (or hold) time. Figure 3 shows the predictions of dwell-time effects using this simple model on Astroloy (UNS N13017) at 655 °C (1211 °F), which compare very well with the data. In this figure, the data from 0 and 2 min of dwell time are used to generate the constants in Eq 1, and the rates for 15 min hold time are predicted and compared with the experimental data.

In the aforementioned example, the dominating time-dependent damage mechanism was the environment and the resulting embrittlement, but without due consideration to the separate roles of oxidation and oxygen ingress. In creep-ductile materials, where creep is assumed as the dominating mechanism, the average time rate of crack growth, $(da/dt)_{avg}$, during the dwell period, t_h, is correlated with the average value of the C_t parameter, $(C_t)_{avg}$, defined in the following way (Ref 1):

$$\left(\frac{da}{dt}\right)_{avg} = \frac{1}{t_h} \int_0^{t_h} \left(\frac{da}{dt}\right) dt \qquad \text{(Eq 2)}$$

$$(C_t)_{avg} = \frac{1}{t_h} \int_0^{t_h} C_t dt \qquad \text{(Eq 3)}$$

where C_t is a crack-tip parameter related to the rate of expansion of the creep-zone size under the conditions of small-scale creep and is estimated by Eq 4 for a cracked body of thickness B, width W, and crack size a (Ref 1, 5):

$$C_t = \frac{P\dot{V}_c}{BW}(F/F') \qquad \text{(Eq 4)}$$

where $\dot{V}_c$ is the instantaneous load-line displacement rate in response to the applied load, P, and F and F' are defined as follows:

$$F = \frac{K}{P} B W^{1/2} \qquad \text{(Eq 5)}$$

$$F' = \frac{dF}{d(a/W)} \qquad \text{(Eq 6)}$$

Under the conditions of extensive creep, C_t is identical to a path-independent integral, C^*, defined by the following:

$$C^* = \int_{\Gamma} W^* dy - T_i \left(\frac{\partial \dot{u}_i}{\partial x} \right) ds \qquad \text{(Eq 7)}$$

In Eq 7, W^* is the rate of strain-energy density, T_i are components of the traction vector, $\dot{u}_i$ are components of the displacement vector, and Γ is an arbitrary counterclockwise contour that begins on the lower crack surface and ends on the upper crack surface while enclosing the crack tip. The C^*-Integral uniquely characterizes the crack-tip stress field for a cracked body subjected to extensive creep conditions. The model used to estimate creep-fatigue crack growth rates for loading cycles with hold times is given by Eq 8 (Ref 6):

$$\frac{da}{dN} = c_1 (\Delta K)^{n_1} + H \left[(C_t)_{avg} \right]^q t_h \qquad \text{(Eq 8)}$$

Chromium-Molybdenum Steel. An example of the correlation for 1.25Cr-0.5Mo steel at 538 °C (1000 °F) is shown in Fig. 4 (Ref 6). In Fig. 4(a), the data for various dwell times ranging from 0 to 24 h are plotted as a function of ΔK. It is apparent that time-dependent creep deformation is more widespread, and there is a lack of correlation between crack growth rate and ΔK. The same data when plotted with $(C_t)_{avg}$ consolidates into a single trend, shown in Fig. 4(b). The $(C_t)_{avg}$ can be measured at the loading pins, provided accurate displacement changes during hold time can be measured.

Superalloys. It is not uncommon for time-dependent crack propagation in commercial nickel-base superalloys to be environmentally controlled and/or dominated. Experiments on five commercial nickel-base superalloys under static crack growth conditions are shown in Fig. 5 for air and inert environments (Ref 7).

The data show marked differences in these five alloys in air, and a single scatter band for inert environment.

Evolution of Crack-Tip Stress Fields Due to Creep

During the past decade, significant analytical advances have occurred in nonlinear fracture mechanics concepts to account for creep deformation at the crack tip. A new field, referred to as time-dependent fracture mechanics (TDFM), has emerged that allows the incorporation of creep deformation effects ranging from small-scale creep to extensive creep, including effects of cyclic loading in characterizing the crack growth behavior at elevated temperature (Ref 1). These concepts have been developed and primarily applied to elevated-temperature behavior of ferritic and austenitic stainless steels that are fundamentally different from nickel-base alloys. The ductility of these steels under creep conditions (10% or higher) is substantially higher than in the case of nickel-base alloys (1 to 5%). Consequently, crack growth under elevated-temperature conditions in these materials is always accompanied by significant creep deformation. Crack-tip parameters such as C^* and C_t based on the assumption of stationary crack or very slow crack growth (in comparison to the kinetics of creep deformation) are successful in predicting the crack growth rate. This is not the case for nickel-base alloys that are much more creep resistant, and their ductility is further reduced by the presence of an aggressive environment at high temperatures. Thus, crack growth in these materials is accompanied by creep strains, but the
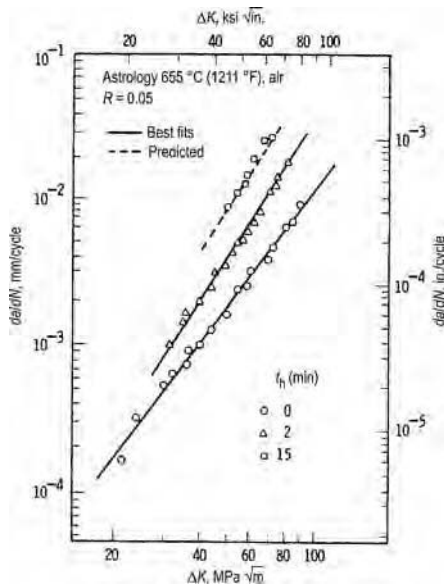
Fig. 3 da/dN versus ΔK behavior in Astroloy at 655 °C (1211 °F) for hold times of 0, 2, and 15 min. Source: Ref 3, 4

LIVE GRAPH
Click here to view

(a)
LIVE GRAPH
Click here to view
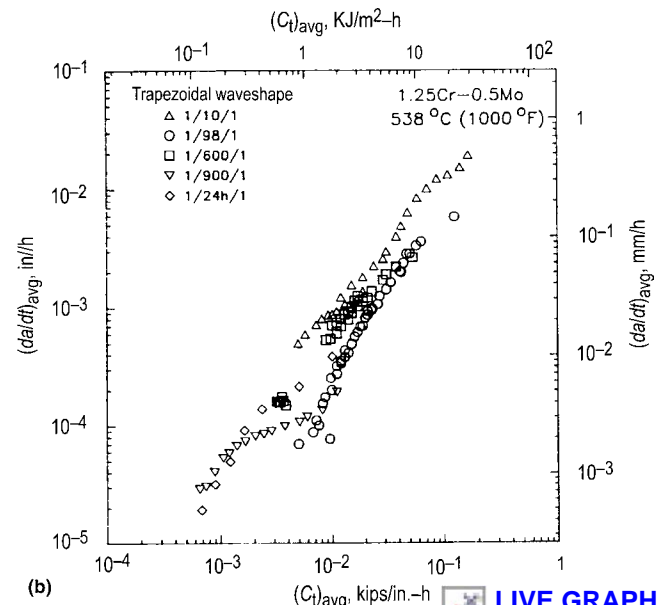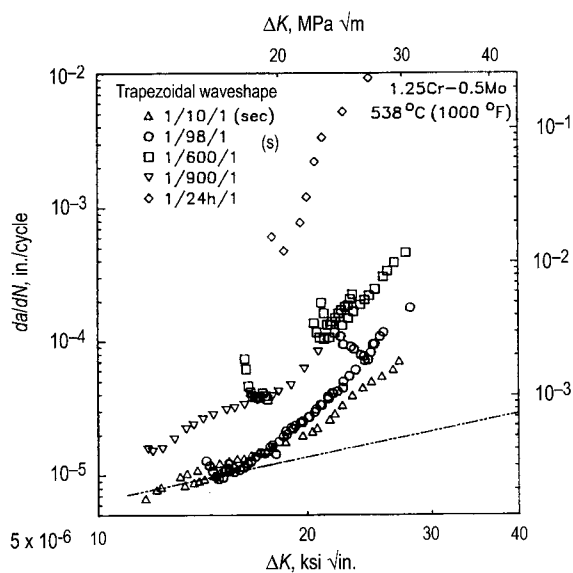
(b)
LIVE GRAPH
Click here to view

Fig. 4 (a) Dwell-time fatigue crack growth data expressed as a function of ΔK for various hold times ranging from 0 to 24 h for 1.25Cr-0.5Mo steels. (b) The same data plotted in the form of $(da/dt)_{avg}$ versus $(C_t)_{avg}$. Source: Ref 6
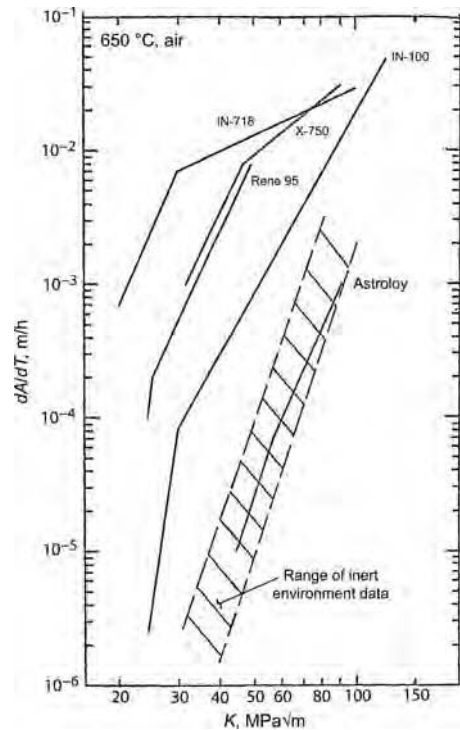
Fig. 5 Static crack growth rates for five nickel-base superalloys at 650 °C, showing that the differences in their behavior in air are dominated by environmental interaction. Source: Ref 7

LIVE GRAPH
Click here to view

Time = 74.5 h; Crack length = 2.26 cm

Time = 97.8 h; Crack length = 2.39 cm

Time = 104.2 h; Crack length = 2.52 cm

Effective creep strain   **Effective stress**

Fig. 6 Evolution of crack-tip creep zone in a 2519 aluminium compact specimen at various times. The measured crack-size-versus-time profile was input into the analysis. Note the change in creep-zone shape as a function of time. Source: Ref 8

magnitude of these strains is not sufficient to completely dominate over the elastic strains. Further, the presence of creep strains may be limited to the immediate vicinity of the crack tip, making the stress-intensity parameter, K, a valid far-field parameter. This explains the limited success in the use of K for characterizing crack growth, but it also is clear that the use of LEFM alone does not provide the range of capability needed.

The applicability of TDFM parameters to nickel-base alloys using continuum mechanics models that clearly account for crack growth in the evolution of crack-tip stress fields in addition to creep still need to be completely examined. The TDFM parameters such as C_t or $C*$ may not be directly applicable to these much more complex phenomena that include creep-brittle behavior exacerbated by the effects of environment, which are not a significant consideration in ferritic and austenitic steels. Combined creep and environmental effects can drive the crack growth at rates that keep pace with the development of creep strains ahead of the crack tip. This condition can lead to non-self-similar creep-zone evolution for an extended period, thus further complicating the establishment of a single crack-tip parameter for characterizing crack growth. This brings to question whether a single parameter even exists that will uniquely characterize the crack-tip conditions and therefore the crack growth rates.

Sufficient understanding now exists from earlier work on aluminum alloys (Ref 8) that behave in a creep-brittle fashion to guide future research to address the aforementioned concerns and bring closure to the questions. Figure 6 shows the finite-element simulation of the development of a crack-tip creep zone in a compact-type specimen of 2519 aluminum alloy (UNS A92519). In this simulation, the crack tip is advanced by releasing nodes at a rate measured in an actual test conducted under loading and temperature conditions that are identical to the conditions used in the experiment. The creep zone in this case is always

small in comparison to the pertinent length dimensions of the specimen, such as the crack size and the uncracked ligament, making the applied value of K a relevant crack-tip parameter. However, it is also clear that in addition to size, the shape of the crack-tip creep zone is constantly evolving in this experiment, questioning the uniqueness of K alone as being the crack-tip parameter. Figure 7 shows the evolution of the creep zone along the x- and y-axes, clearly showing the lack of self-similarity in the growth pattern of the creep zone. Figure 8(a) shows the correlation between creep crack growth rate and K for several specimens of
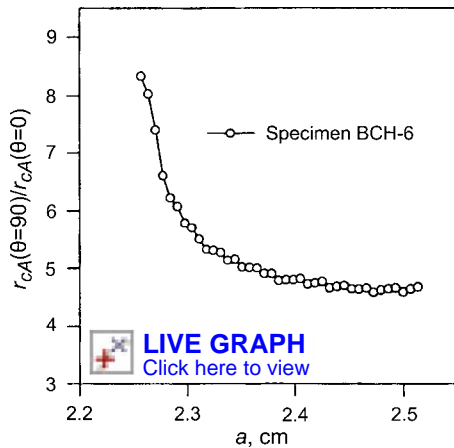
Fig. 7 Ratio of creep-zone size along the *y*-axis (θ = 90) with that along the *x*-axis (θ = 0) in the example shown in Fig 6. Self-similar creep-zone expansion occurs only after approximately 0.15 cm crack extension. BCH-6 is a specimen identification. Source: Ref 8

Fig. 8 Creep crack growth rate, *da/dt*, for aluminium AA2519 alloy. (a) As a function of *K*, the specimen BCH-1 was 6.25 mm thick, while the others were 25.4 mm thick. (b) The same data as in (a) except correlated to K/r_{cA}^q, where r_{cA} is the instantaneous creep zone size, and q is the fitting constant. Source: Ref 8

2519 aluminum, showing that a unique correlation does not exist during the initial portion of the test during which the creep zone evolves in size and shape, but the results do converge after substantial amounts of crack growth have occurred. Figure 8(b) shows the same data plotted against a parameter based on *K* but modified to include the size of the creep zone, K/r_{cA}^q, where r_{cA} is the creep zone size for θ = 90°, and q is a parameter that depends on the creep deformation exponent for the material. This parameter is equivalent to using C_t (Ref 5) but defined for a growing crack (Ref 8).

Time-Dependent Environmental Degradation

The primary time-dependent degradation mechanisms in nickel-base alloys at elevated temperature are microstructural evolution, creep, and environment. This article focuses on describing the models based on creep and oxidation kinetics. However, some aspects of microstructural evolution are intimately linked to oxidation and therefore must be considered as part of oxidation-damage kinetics.

The nucleation and growth of cavities at the grain boundaries due to creep result in almost purely intergranular fracture. When creep deformation is localized near the crack tip, damage in the form of grain-boundary cavitation evolves ahead of the crack tip. The cavities can grow both in numbers and size and can eventually coalesce to advance the crack. Such crack growth is dominantly intergranular. The kinetics of cavity growth have been modeled as either diffusive growth or as growth constrained by the deformation of the surrounding grains. These have led to models for rationalizing the correlation between creep crack growth rates and the time-dependent fracture mechanics parameters (Ref 9–11). The usefulness of

these models lies in understanding the microstructural parameters that may be pertinent in determining the creep crack growth behavior and in understanding the kinetics of this form of crack-tip damage. The rate of grain-boundary cavity growth can be influenced by oxygen diffusion along the grain boundaries, which must be addressed to gain full understanding of this damage-evolution mechanism. The interpretation of these cavities is further complicated by the fact that such cavities can also be formed in the absence of externally applied stress. In studies on commercially pure nickel, grain-boundary cavities have been induced by static oxidation (Ref 12). These authors also found that these cavities were not formed when samples were statically exposed in vacuum or if the material was statically exposed in air after annealing in hydrogen to decarburize the samples.

Environmental degradation potentially plays a major role in crack growth during dwell periods. Oxygen undergoes both short- and long-range diffusion processes in nickel-base alloys at elevated temperature (Ref 13). An oxide layer forms at the crack tip in the short-range diffusion process, resulting in even higher stresses at the crack tip. The rupture of wedge-shaped regions due to oxygen ingress from the short-range diffusion process can accelerate the intergranular crack growth rates. The long-range diffusion process involves the oxygen traveling along paths such as slip bands and grain boundaries to form internal oxide sites, cavities, or solute segregations. Figure 9 shows how the time-dependent crack growth rate in a nickel-base alloy N18 increases significantly with oxygen partial pressure (Ref 13). In Figure 9(a), the crack growth rate is found to be a function of the applied *K* level, but the threshold partial pressure at which oxidation plays a significant role is independent of the applied *K* (at least over the range of *K* tested here). Figure 9(b) shows how the temperature affects the time-dependent crack growth at various oxygen partial pressures. The temperature influences

both the crack growth rate and the threshold oxygen partial pressure, at which instance time-dependent crack growth becomes significant.

For cyclic crack growth, varying the amount of time allowed for environmental interaction at the crack tip gives an analogous transitional effect to varying the oxygen partial pressure at a fixed waveform. Figure 10 shows data from cyclic crack propagation tests on a nickel-base superalloy of 40 μm grain size at 650 °C in air (Ref 14). The crack propagation data at a fixed cyclic stress-intensity factor are plotted in the left-hand graph as crack advance per unit cycle and in the right-hand graph as crack advance per unit time for total cyclic periods ranging from 0.1 to 3200 s. The data shown are characterized by a bilinear fit on the log-log plots. The data are time-dependent at all the periods tested; that is, the data cannot be described by a horizontal line on the left-hand graph. Examination of the right-hand graph shows that for cyclic periods beyond 30 s, the data can be described by a horizontal line. This regime thus represents full time dependence; that is, the crack growth rate per unit time is constant. A critical test of future physics-based models will be their ability to predict the transition from time-independent behavior (not seen in this particular data set) to partially time-dependent behavior (from 0.1 to 30 s in this particular data set) to fully time-dependent behavior (beyond 30 s in this particular data set). The kinetics of these transitions can be used as a metric for the ability of an alloy to resist time-dependent crack propagation.

Oxidation Mechanisms

Nickel-base superalloys exhibit oxidation resistance by forming a protective oxide layer. These alloys are classified into those that form chromia and those that form alumina protective scales. The alumina-forming alloys have higher-temperature oxidation resistance. Ni₃Al
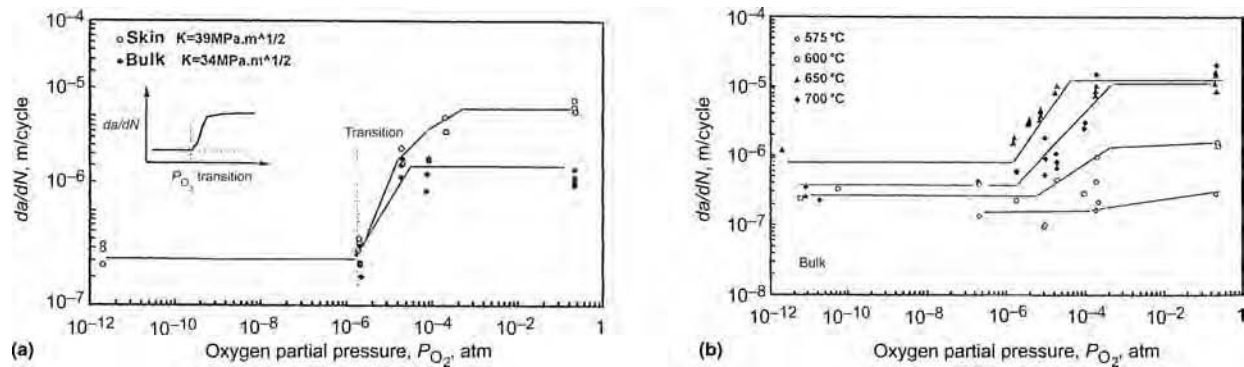
Fig. 9 Effect of oxidation at the crack tip in alloy N18 as a function of (a) applied K value and (b) temperature. Source: Ref 13
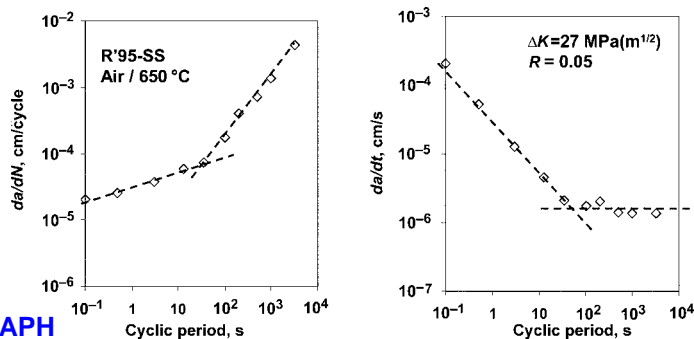
Fig. 10 Cyclic crack growth as a function of total period, showing transition to fully time-dependent crack growth. Source: Ref 14

precipitates form in these types of alloys, providing sufficient high-temperature strength. The chromia-base alloys are limited in their high-temperature use because the protective oxide decomposes into a volatile CrO_3 above 900 to 950 °C, leaving the remaining oxide layer nonprotective (Ref 15). However, another source states that the high-temperature oxidation resistance of nickel-base superalloys depends mostly on the presence of chromium (Ref 16).

The oxidation mechanisms of nickel-base alloys have been shown to contribute to crack initiation and propagation at elevated temperature (Ref 17). Andrieu et al. (Ref 17) state that some early oxidation research indicated that oxygen is most detrimental to an existing crack by penetrating to the crack tip. Oxide formation at the crack tip can produce wedge-shaped intrusions along the crack front, increasing the stress in that area. These wedges can then rupture at grain-boundary intersections, causing an increase in intergranular crack growth. The oxygen can also diffuse along slip planes or grain boundaries, allowing internal oxides or cavities to form, or may take part in other chemical reactions, producing agents that embrittle the grain boundary (Ref 17).

One mechanism for grain-boundary oxidation has been proposed and experimentally studied by Andrieu et al. (Ref 17). That study suggests that the oxidation occurs in a two-stage process. The first step, controlled by the oxygen partial pressure in the atmosphere, results in the formation of NiO and FeO and their spinels. These oxides are porous and allow oxygen diffusion to occur. This diffusion significantly determines the reaction rate over time. Once these oxides form on the surface, the oxygen partial pressure at the oxide-metal interface is reduced. This would provide the required environment for the formation of the protective Cr_2O_3. This can contribute to passivation at the crack front. Fatigue experiments of alloy 718 were conducted at 650 °C in vacuum and in air. Two oxide layers were observed using a transmission electron microscope. The inner oxide was Cr_2O_3 and the outer oxide was the spinel oxide $Ni(FeCr)_2O_4$.

Molins et al. (Ref 18) propose that a nickel-rich oxide forming in the early stage of oxidation is the cause of alloy embrittlement at high temperatures. Three possible mechanisms of oxidation crack propagation are proposed that involve the local mechanical behavior at the crack tip and consequences of the nickel oxide nucleation and growth. One mechanism is the short-range intergranular oxygen diffusion. From the experimental results reported, chromia nuclei act as oxygen getters, slowing the diffusion, and nickel oxide nuclei allow the oxygen to diffuse to the grain boundaries, embrittling the alloy. Another mechanism is vacancy injection into the grain boundaries.

Vacancies are generated during mechanical deformation and as a result of the NiO growth. The nickel oxide grows epitaxially, preventing vacancies from being annihilated at the oxide-metal interface. The vacancies then are forced along the grain boundaries in front of the crack tip, accumulating and contributing to the embrittlement. The third mechanism is that of chromium enrichment at the alloy subscale surface.

An alternative controlling oxidation mechanism has been proposed by Gao et al. (Ref 19). X-ray photoelectron spectroscopy analysis of fracture surfaces of alloy 718, which had been tested in fatigue at 975 K (702 °C) in air, showed a high niobium concentration, most likely in the form of Nb_2O_5-type oxides. These are nonprotective brittle oxides and could be a contributing factor to enhanced crack growth rate. They therefore concluded that niobium is the primary crack growth rate controlling species in elevated-temperature fatigue of alloy 718. They supported this conclusion by testing a niobium-free Ni-18Cr-18Fe alloy under the same conditions as alloy 718 and found that oxygen had no influence on the crack growth rates. Gao et al. then proposed an environmentally enhanced crack growth process of alloy 718 that included, diffusion of oxygen to the crack tip. The oxygen reacted with niobium oxides and other exposed particles, resulting in the formation of niobium oxides and a niobium-enriched fracture surface. The oxide film grew and mechanically ruptured, causing an increase in crack growth rate. Thin, brittle niobium oxide films found at the grain boundaries of the highly stressed crack-tip region were also found by Kang et al. (Ref 20) to be the cause of intergranular embrittlement of alloy 718 when tested in air at 650 °C.

One study evaluated the microstructural characteristics of high-temperature oxidation of nickel-base superalloy Hastelloy C-4 (UNS No 6455) (Ref 21). The authors of that study found large amounts of manganese and chromium in the oxide layer. This indicated that the predominant oxide was Cr_2O_3 and some form of complex spinels. Greater concentrations of chromium and manganese were found in the external oxide as opposed to the internal

oxide. Voids were also present in the oxide and in the matrix material. The external oxide was formed during air exposure at 1200 °C for 6 h, and the layer was observed to have cracked and spalled during the oxidation process. It was obvious that the internal oxidation had internal voids that had formed along grain and twin boundaries. This could be explained by rapid diffusion of oxygen along these boundaries.

Garat et al. (Ref 22) found two modifications at the surface of the material induced by oxidation of alloy 718 in air at 1000 °C for 1 h. Both an external scale of approximately 1 μm thick and an intergranular oxide penetration that had a depth of approximately 7 μm after oxidation were observed. Using secondary ion mass spectrometry, this study found no significant amounts of oxygen ahead of the tip of the intergranular oxide penetration region. The most prevalent internal oxide was observed to be alumina.

Evaluating Rene N4 (single crystal in air at 870 °C during fatigue testing), Khobaib et al. (Ref 23) concluded that the oxidation kinetics are fast enough to produce an oxide layer of reasonable size on the freshly exposed crack surface at frequencies as high as 1.0 Hz. Yuen et al. (Ref 24) also found that the oxidation kinetics significantly affected the oxide growth on the exposed fatigue fracture surfaces in air at 427 to 649 °C for alloy 718.

Fracture Mechanisms

Elevated-temperature creep and fatigue tests in air or in an oxygen partial pressure environment of nickel-base superalloys have been shown to increase the crack growth rate (Ref 17, 25, 26), however, in some cases, only at high strain ranges where plastic deformation is greatest (Ref 27). The actual mechanism for such an enhancement has been studied and proposed by several researchers. Oxidation at the crack tip and at the grain boundaries has been proposed as the most likely reason for this reduction in fatigue and creep strength; however, universal agreement does not exist.

As was previously discussed, oxygen can diffuse rapidly to the crack tip and along the internal grain boundaries. The oxygen then forms an oxide with niobium, chromium, nickel, and/or iron. Researchers disagree as to which is the dominant oxide controlling the embrittlement mechanism. The oxide can form a wedge at the crack tip, creating high stresses that could be transmitted to the substrate material (Ref 17). Long-range diffusion of oxygen into the grain boundaries has been shown to produce an embrittling effect by weakening the grain boundaries through oxidation. Woodford and Bricknell (Ref 28) stated that reactions along the grain boundaries with oxygen could release embrittling agents. Other researchers propose that the brittle nature of the oxide that forms at the grain boundary is the actual weakening mechanism (Ref 26).

Fractographic analysis has been conducted, finding that the fracture mechanism of embrittled specimens is intergranular fracture (Ref 27). Therefore, the grain size, which controls available grain boundary, is also a factor to be considered (Ref 29). An oxygen partial pressure threshold has been observed for the transition from transgranular to intergranular fracture (Ref 17). This variation in fracture mode with a change in partial pressure of oxygen is shown in Fig. 11 for alloy 718.

The effects of oxidation have been observed to decrease the crack growth rate. Yuen et al. (Ref 24) reported experimental results of the threshold ΔK increasing with increasing temperature for alloy 718. In this case, oxidation is described as a decelerating mechanism. The oxide forms at the crack tip and can create a wedge, causing oxide-induced crack closure or crack blunting. Gabrielli (Ref 30) conducted high-temperature fatigue tests of an oxygen-dispersion-strengthened nickel-base alloy (ODS MA 6000) in air and in vacuum. He found that the crack growth rate was substantially higher for the specimens tested in vacuum in all the test conditions. Thick fragments of oxide were observed on the fracture surface, and he concluded that these oxides induced blocking at the crack tip. He also stated that at high temperature, creep may promote ductility of the oxide, thereby enhancing a blocking mechanism. Kang et al. (Ref 20) also found that crack-tip blunting can reduce the embrittlement caused by stress-assisted grain-boundary oxidation.

Fatigue crack growth rate studies of Ni-18Cr-18Fe conducted by Chen and Wei (Ref 31) show that crack growth rate doubled in the presence of oxygen at temperatures of 873, 923, and 973 K. As part of this study, creep tests were conducted at elevated temperature, and they found that grain-boundary cavitation, not oxidation, was the controlling failure mechanism. One conclusion made by Chen and Wei is that, given limited sensitivity to oxygen, the oxidation of nickel, chromium, and iron had little effect on the crack growth rate.

Hancock (Ref 32) provided some interesting insight into the failure mechanisms of nickel-base superalloys at high temperature under cyclic loading. He concluded that for the environment to have an effect, it must have access to the crack tip. He also stated that the environmental effect is only pronounced for the fatigue crack initiation stage and does not increase the fatigue crack growth rate at elevated temperature because, as he states, "at high temperature, when oxidation occurs at the tip of the crack, the environment is effectively excluded when the small strain amplitudes that would be experienced in service conditions are considered." Therefore, he postulates that the environmental influence is dependent on whether or not the structure has existing cracks. Pre-existing surface cracks are not expected on turbine blades; therefore, environment will be the controlling failure mechanism for such an application.

The state of stress of a component has also been shown to play a role in whether or not oxidation controls crack extension in fatigue (Ref 33). Crack-tip oxidation, which promotes intergranular fracture, is accelerated under plain-strain conditions, while a transgranular cyclic plasticity crack growth mechanism is predominant in plane stress. Therefore, the oxidation mechanism occurs in the plane-strain interior, and the cyclic plasticity mechanism was observed near the surface. However, Pandey (Ref 26) concluded that the effect of oxidation on the creep properties of alloy X-750 was more pronounced for thin-sized testpieces, which would increase the likelihood of plane-stress conditions. The section size effect, however, vanished when the depth of oxygen diffusion was normalized by specimen radius. From tests conducted by Kang et al. (Ref 20) of alloy 718, the conclusion was that K_I can be used as the governing parameter of the stress-strain field ahead of the crack tip during creep crack growth in air.

Microstructure also has an effect on the high-temperature creep properties. Pandey (Ref 26) found that the γ' phase was depleted during air exposure of alloy X-750. This led to a degradation of the creep properties of this alloy.

Grain-boundary cavitation has been shown to be enhanced by the formation of gaseous bubbles in high-temperature creep conditions. Osgerby and Dyson (Ref 34) saw a large amount of cavitation below the surface of embrittled Inconel X-750. Pandey (Ref 26) also provided evidence of diffusing oxygen forming titanium oxide and carbon monoxide bubbles during high-temperature creep testing.
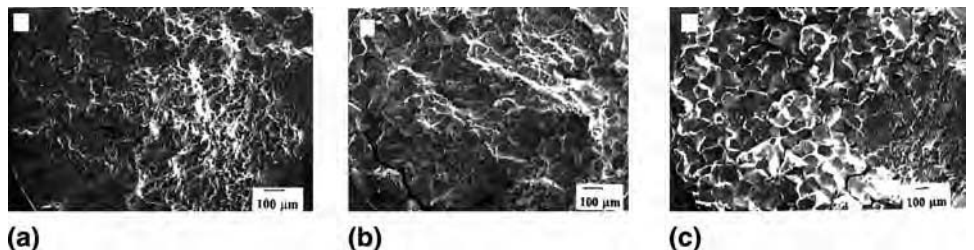
Fig. 11 Fracture surfaces corresponding to different oxygen partial pressures. (a) 10^{-4} Torr. (b) 1 Torr. (c) 4 Torr. Source: Ref 17

Summary of Creep-Fatigue Modeling Approaches

Various approaches for modeling creep-fatigue crack growth behavior have been proposed to capture the various types of damage that is occurring. These include mechanistic modeling, crack growth modeling using the method of superposition to incorporate time-dependent and cycle-dependent components, and numerical modeling.

A number of phenomena have been investigated to obtain a better understanding of the fundamental creep-fatigue crack growth process. As previously mentioned, during the dwell period in creep-fatigue loading, several time-dependent processes are operational at elevated temperature, which include surface oxidation, oxygen diffusion along grain boundaries, and creep deformation/stress relaxation at the crack tip. Each of these can be characterized by evaluating the kinetics of the process. The entire creep-fatigue crack growth process can then be modeled in terms of a rate-controlling step. This requires knowledge of the oxygen diffusion rate along grain boundaries, the oxide growth kinetics, and the rate of the crack-tip creep/stress relaxation. Also, at elevated temperature, dislocation motion is enhanced during fatigue cycling. Several of these processes can be coupled, such as the diffusion of oxygen being affected by the stress-relaxation rate (Ref 17, 35). This makes the kinetic modeling approach a very difficult endeavor.

Predicting the crack growth rate using the method of superposition has also been used as an effective way to model the process. This method was first proposed for modeling a corrosion-fatigue process (Ref 36). It was also later applied to the creep-fatigue process in the following form (Ref 37):

$$\frac{da}{dN} = \left(\frac{da}{dN}\right)_{cyc} + \left(\frac{da}{dN}\right)_{time} \qquad (Eq\ 9)$$

where da/dN is the total crack growth rate, $(da/dN)_{cyc}$ is the cycle-dependent component of the crack growth rate, and $(da/dN)_{time}$ is the time-dependent component of the crack growth rate. This type of approach allows for a generalized crack growth rate model to be developed that can potentially incorporate the cycle-dependent and time-dependent mechanisms in the crack growth rate process. Currently, the superposition model requires a significant amount of detailed test data in order to use it to predict creep-fatigue crack growth for a variety of alloy systems.

The need to reduce cost during the design of new materials and components has led to a substantial increase in the use of computational techniques to model mechanical behavior, including creep-fatigue crack growth. This type of modeling approach has the potential of reducing the time and cost associated with extensive laboratory testing, and it can also provide information not readily obtainable using traditional test methods. Finite-element modeling (FEM), as previously mentioned, has been used to investigate the evolution of crack-tip stress fields during creep fatigue (Ref 8). Also, FEM has been shown to be an effective tool for failure analysis and fatigue-life prediction of a gas turbine disk when subjected to creep-fatigue loading (Ref 38).

In addition to FEM, a new computational technique to evaluate materials behavior on an atomistic level has been developed, known as molecular dynamics (MD). The MD models can be coupled with finite-element solutions to provide a multiscale modeling capability. The MD and FEM are seamlessly coupled using a concurrent length-scale approach suggested by Broughton et al. (Ref 39). By this process instead of considering a number of atoms in the range of 10^8 to 10^9 only 10^3 to 10^4 will suffice. The multiscale approach uses the appropriate theories in the proper places and saves computer time. Preliminary work on modeling crack propagation (Ref 40) using MD is illustrated in Fig. 12. This simulation grows the crack when the stress exceeds the cohesive interatomic forces, providing a means to predict fracture from first principles. However, no such studies have been reported in the literature that model crack growth under creep-fatigue conditions; however, the potential for such modeling is apparent.

Recommendations for Future Work

Although progress has been made in understanding and predicting creep-fatigue behavior, a number of significant challenges remain. The constant desire for improved efficiency of various energy-conversion systems will require components to withstand more aggressive thermal and environmental conditions. The need to accurately predict component life at elevated temperatures in a creep-fatigue-environment operating regime will continually motivate future work in this critical research area. To push the analytical models for predicting creep-fatigue-environment interactions to the next level of sophistication will require systematic experimental and computational work that is carried out side by side. Several recent advancements in computing and experimental capabilities provide a new set of tools that can be used to develop new insights into this convoluted problem.

As described in this article, most crack growth rate models are empirical or semiempirical in nature. These models require expensive and time-consuming tests in order to develop the needed parameters and, at best, provide interpolative capability for predicting time and temperature and loading effects. Future creep-fatigue crack growth rate models will need to be more physics-based, with a good understanding of their limitations but with considerable capability for extrapolation. It will also allow the integration of the various models such that between them a wide temperature and time range is addressed accurately. This type of analytical tool would allow the direct calculation of the crack growth rate, making life prediction more tractable and reliable. However, in order to develop such a model, the unique kinetics of the various time-dependent mechanisms must be quantitatively described, so a rate-limiting step can be determined for wide time and temperature regimes. Models based on these various crack growth mechanisms can then be integrated to provide a more comprehensive approach to life prediction. The kinetics of these mechanisms can be investigated through carefully controlled experiments and accurately developed computational models.

Experimental studies must be designed and conducted to address the interactions between creep, fatigue, and the environment. More extensive creep-fatigue crack growth tests, with

Number of atoms: 10,000
fy = 0.50
Total time - 70, initial time = 10
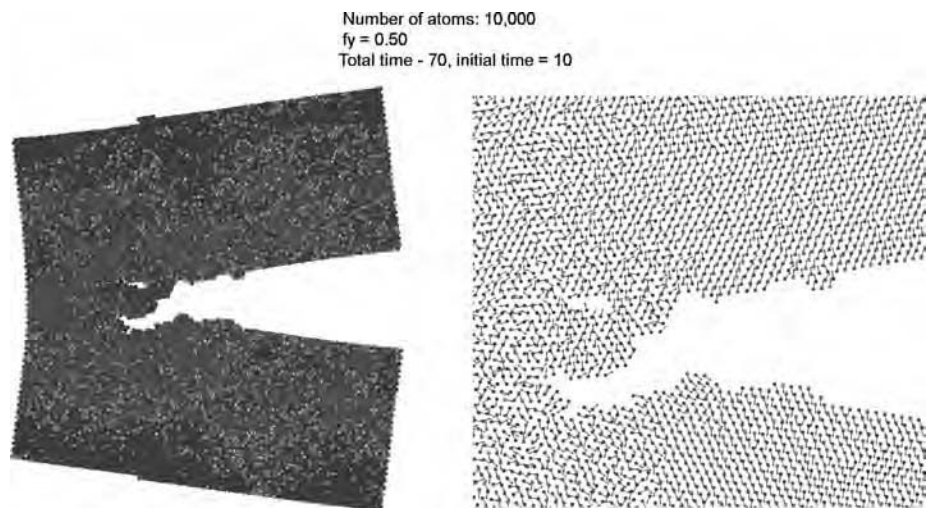
Fig. 12 Studies of crack propagation using molecular dynamics

and without hold times, conducted in air, vacuum, and controlled environments such as varying partial pressures of oxygen would provide necessary experimental evidence to understand the environmental effect on the crack growth rate, as well as potentially shed light on the embrittling mechanism(s) that lead to the creep-brittle behavior in some materials. Also, experiments that characterize creep at the crack tip during long hold times at elevated temperature for creep-brittle materials would be of substantial value in developing a more comprehensive understanding of the crack driving force during the hold time of a creep-fatigue test. Development of a creep-fatigue testing database would also be enormously valuable to the design and prediction of materials behavior. A database could foster the use of more probabilistic design and structural reliability analyses.

Finite-element analysis and molecular dynamics simulations are two computational techniques that can be employed to help provide greater understanding of the fundamental crack growth process at elevated temperature. These tools can also be used to supply the necessary understanding of the fundamental behavior of the materials. This can then be fed back into the design and development of new materials, which would have enhanced creep-fatigue properties. One way this may be achieved is by using molecular dynamics simulations to investigate the grain-boundary character and interaction with some environmental species. The cohesive strength of the grain boundaries could be determined in the presence of some diffusing atomic species. This could lead to new materials with engineered grain boundaries to prevent intergranular fracture, as well as offer more understanding of the damaging mechanisms. The finite-element analysis technique can help in determining local crack-tip stresses and the evolution of the crack-tip plastic zone during elevated-temperature cyclic loading under growing crack conditions. The analytical present solutions are limited to stationary crack assumptions that are simply not valid for creep-brittle materials. Thus, this type of analysis could be used to help in the understanding of creep-brittle behavior.

The experimental and computational future work described previously could immeasurably assist the manufacturers in designing and developing new materials as well as in predicting the life of in-service components.

ACKNOWLEDGMENT

REFERENCES

1. A. Saxena, *Nonlinear Fracture Mechanics for Engineers*, CRC Press, 1998
2. A. Saxena, How Far Have We Come in Predicting High Temperature Crack Growth and Challenges That Remain Ahead, *9th International Meeting — Temperature-Fatigue Interaction,* Paris, France, pp 205–216, 2001.
3. A. Saxena, A Model for Predicting the Environment Enhanced Fatigue Crack Growth Behavior at High Temperature, *Thermal and Environmental Effects in Fatigue-Research Design Interface- PVP-71,* ASME, New York, pp. 171–184, 1984.
4. R.M. Pelloux and J.S. Huang, Creep-Fatigue-Environment Interactions in Astroloy, *Creep-Fatigue—Environment Interactions*, R.M. Pelloux and N.S. Stoloff, Ed., TMS-AIME, pp. 151–164, 1980.
5. A. Saxena, Creep Crack Growth under Non Steady-State Conditions, *Fracture Mechanics: Seventeenth Conference*, ASTM STP 905, pp. 185–201, 1986.
6. K.B. Yoon, A. Saxena, and P.K. Liaw, Characterization of Creep-Fatigue Crack Growth Behavior Under Trapezoidal Waveform Using C_t-Parameter, *International Journal of Fracture,* Vol 59, pp. 95–114, 1993.
7. K.R. Bain, "Creep Crack Growth Rates of Rene-95 in Air and Pure Argon," M.S. Thesis, Massachusetts Institute of Technology, February 1982.
8. D.E. Hall, D.L. McDowell, and A. Saxena, Crack Tip Parameters for Creep-Brittle Crack Growth, *Fatigue & Fracture of Engineering Materials & Structures,* Vol 21, p 387–401, 1998.
9. D.L. Wilkinson and V. Vitek, Propagation of Cracks by Cavitation—A General Theory, *Acta Metallurgica,* Vol 30, pp. 1723–1732, 1982.
10. J.L. Bassani and V. Vitek, Propagation of Cracks Under Creep Conditions, *Proceedings of the 9th National Congress of Applied Mechanics—Symposium on Nonlinear Fracture Mechanics*, pp. 127–133, 1982.
11. S.C. Jani and A. Saxena, Influence of Thermal Aging on the Creep Crack Growth Behavior of a Cr-Mo Steel, *Effects of Thermal and Load Histories,* TMS, pp. 201–220, 1987.
12. R.H. Bricknell and D.A. Woodford, The Mechanism of Cavity Formation During High Temperature Oxidation of Nickel, *Acta Metallurgica,* Vol 30, p 257–264, 1982.
13. A. Pineau, Fatigue and Creep-Fatigue Behavior of Ni-Base Superalloys: Microstructural and Environmental Effects, *Mechanical Behavior of Materials at High Temperature*, Kluwer Publishers, Holland, pp. 135–154, 1996.
14. M.F. Henry and R. L. Casey, "Grain Boundary Serrations in Nickel-Base Superalloys: Mechanism and Impact on Resistance to Time Dependent Fatigue Crack Propagation," MRS Fall Meeting, Boston, Massachusetts, 1996.
15. A.S. Khanna, *Introduction to High Temperature Oxidation and Corrosion,* ASM International, 2002.
16. R.E. Reed-Hill and R. Abbaschian, *Physical Metallurgy Principles,* 3rd ed., PWS Publishing Company, Boston, MA, 1994.
17. E. Andrieu, R. Molins, H. Ghonem, and A. Pineau, Intergranular Crack Tip Oxidation Mechanism in a Nickel-Based Superalloy, *Materials Science and Engineering,* Vol A 154, p 21–28, 1992.
18. R. Molins, J.C. Chassaigne, and E. Andrieu, Nickel Based Alloy Intergranular Embrittlement in Relation with Oxidation Mechanisms, *Materials Science Forum*, Vol 251–254, p 445–452, 1997.
19. M. Gao, S.F. Chen, G.S. Chen, and R.P. Wei, Environmentally Enhanced Crack Growth in Nickel-Based Alloys at Elevated Temperatures, *Elevated Temperature Effects on Fatigue and Fracture,* ASTM STP 1297, p 74–84, 1997.
20. B.S.-J. Kang, G. Zhang, P. Liu, and M. Ellathur, Stress Accelerated Grain Boundary Oxygen Embrittlement on Creep Crack Growth of Nickel-Base Superalloys, *Fatigue and Fracture at Elevated Temperatures,* AD-Vol 50, ASME, 1995.
21. F.A. Khalid, N. Hussain, and K.A. Shahid, Microstructure Characteristics of High-Temperature Oxidation in Nickel-Base Superalloy, *Journal of Materials Engineering and Performance,* Vol 6 (No. 2), p 219–222, 1997.
22. V. Garat, O. Brucelle, J.-M. Bloue, V. Rebeyrolle, D. Monceau, B. Viguier, and E. Andrieu, Comparing Different Methods to Determine the Intergranular Oxidation Damage on a Nickel Based Superalloy, *Materials Science Forum,* Vol 461–464, p 537–544, 2004.
23. M. Khobaib, T. Nicholas, and S.V. Ram, Role of Environment in Elevated Temperature Crack Growth Behavior of Rene N4 Single Crystal, *Environmentally Assisted Cracking: Science and Engineering,* ASTM STP 1049, p 319–333, 1990.
24. J.L. Yuen, P. Roy, and W.D. Nix, Effect of Oxidation Kinetics on the Near Threshold Fatigue Crack Growth Behavior of a Nickel Base Superalloy, *Metallurgical Transactions A*, Vol 15A, p 1769–1775, 1984.
25. J. Tong, S. Dalby, J. Burne, M.B. Henderson, and M.C. Hardy, Creep, Fatigue, and Oxidation in Crack Growth in Advanced Nickel Base Superalloys, *International Journal of Fatigue,* Vol 23, p 897–902, 2001.

26. M.C. Pandey, Oxidation-Creep Interaction in a Nickel Base Superalloy: Experiments and Mechanisms, *Transactions of the Indian Institute of Metallurgy,* Vol 49 (No. 4), p 342–348, 1996.

27. M. Marchionni, G.A. Osinkolu, and G. Onofrio, High Temperature Low Cycle Fatigue Behavior of UDIMET 720 Li Superalloy, *International Journal of Fatigue*, Vol 24, p 1261–1267, 2002.

28. D.A. Woodford and R.H. Bricknell, Environmental Embrittlement of High Temperature Alloy by Oxygen, *Treatise on Materials Science and Technology*, Vol 25, p 157–199, 1983.

29. H.T. Pang and P.A.S. Reed, Fatigue Crack Initiation and Short Crack Growth in Nickel-Base Turbine Disc Alloys — The Effects of Microstructure and Operating Parameters, *International Journal of Fatigue,* Vol 25, p 1089–1099, 2003.

30. F. Gabrielli, High Temperature Fatigue and Creep Crack Growth in Nickel Base Superalloys, *High Temperature Alloys for Gas Turbines and Other Applications,* p 1639–1648, 1986.

31. S.F. Chen and R.P. Wei, Environmentally Assisted Crack Growth in a Ni-18Cr-18Fe Ternary Alloy at Elevated Temperatures, *Materials Science and Engineering,* Vol A256, p 197–207, 1998.

32. P. Hancock, The Influence of Environment on the Fatigue Behavior of Superalloy Components at High Temperatures, *Corrosion Science,* Vol 29 (No. 6), p 657–671, 1989.

33. F.V. Antunes, J.M. Ferreira, C.M. Branco, and J. Byrne, Influence of Stress State on High Temperature Fatigue Crack Growth in Inconel 718, *Fatigue and Fracture of Engineering Materials and Structures,* Vol 24, p 127–135, 2001.

34. S. Osgerby and B.F. Dyson, Metallographic Evidence for the Formation of Gaseous Internal Oxides in Nickel-Base Superalloys, *Materials at High Temperatures*, Vol 9 (No. 1), p 13–16, 1991.

35. J. Telesman, T.P. Gabb, A. Garg, P. Bonacuse, and J. Gayda, Effect of Microstructure on Time Dependent Fatigue Crack Growth Behavior in a P/M Turbine Disk Alloy, *Superalloys 2008,* TMS, Warrendale, PA, p 807–816, 2008.

36. R.P. Wei and J.D. Landes, *Materials Research & Standards,* Vol 9 (No. 7), p 25–46, 1969.

37. R.P. Wei and Z. Huang, Influence of Dwell Time on Fatigue Crack Growth in Nickel-Base Superalloys, *Materials Science and Engineering*, Vol A336, p 209–214, 2002.

38. R.A. Claudio, C.M. Branco, E.C. Gomes, J. Byrne, G.F. Harrison, and M.R. Winstone, Fatigue Life Prediction and Failure Analysis of a Gas Turbine Disc Using the Finite-Element Method, *Fatigue and Fracture of Engineering Materials and Structures,* Vol 27, p 849–860, 2004

39. J.Q. Broughton F.F. Abraham, N. Bernstein, and E. Kaxiras, Concurrent Coupling of Length Scales: Methodology and Application, *Physical Review B,* Vol 60, p 239, 1999.

40. P.C. Millett and R.P. Selvam, "Molecular Dynamics Simulation of Crack Propagation, Report, Computational Mechanics Laboratory," BELL 4190, University of Arkansas, Fayetteville, AR, 2004.

Modeling Fatigue Crack Growth

Andrew H. Rosenberger, Air Force Research Laboratory

UNDERSTANDING FATIGUE CRACK GROWTH is critical for the safe operation of many structural components and has seen a widespread implementation, especially in the aerospace community where failure of a structural component is often catastrophic. Typically, structures are designed for the cyclic loading that they will see in service, and this constitutes the safe life fatigue design philosophy. This assumes that all components are machined to a minimum tolerance and are manufactured from materials meeting a minimum fatigue allowable. The United States Air Force, for example, goes beyond this design approach by additionally requiring damage tolerance, wherein the structure is designed such that a small flaw, due to a rogue material defect (e.g., forge induced porosity, casting defect, or remnant unrecrystallized microstructure) or machining error, will not grow to a catastrophic size before a safety inspection or retirement (Ref 1). This requires the application of crack-growth modeling on the specific component geometry subject to the anticipated loading of the structure. Many different aspects of the loading environment and geometry must be taken into account to accurately model the crack propagation in the structure. The loading environment would include the type of cyclic load applied to a component. A simple rotating shaft having a slight misalignment will see a simple, fully reversed, constant-amplitude load cycle each revolution of the shaft. The wing on an aircraft will see a simple load cycle each flight due to a change in loading, wherein the fuselage supports the wing when the aircraft is on the ground, but the wing will support the fuselage when the aircraft is in the air. In addition to this simple cycle, there will be a highly variable loading due to the wind gust loading and maneuver loading. In this case then, the loading that must be taken into account in the crack-growth analysis is very complicated. The geometry of the crack or the component that contains the crack will have a significant influence on the stress-intensity factor (Eq 1) of the crack, affecting the speed that the crack will grow. Simple closed-form approximations exist (Ref 2) for many of the crack shape geometries and structural configurations, provided that the crack is subject to a uniform applied stress. Cracks that are geometrically simple but in a nonuniform stress field, e.g., notch locations or surfaces having residual stress, can be treated using a weight function K solution (Ref 3). Advanced numerical methods exist to model the crack stress intensity factor in more complicated conditions involving complex geometry, stress conditions, and loading (Ref 4).

This article reviews standard fracture mechanics and the methods to determine the crack-growth rate for a material and loading condition experimentally. Then, the two most important aspects of crack-growth modeling are addressed: the loading environment and the crack geometry. Finally, future directions of crack-growth modeling development are discussed.

Basic Crack-Growth Considerations

This chapter focuses on linear elastic fracture mechanics where the stress field at the crack tip can be described by the stress-intensity factor, K (Ref 5):

$$K = Y\sigma\sqrt{\pi a} \qquad \text{(Eq 1)}$$

where Y is a correction factor for the geometry and loading, σ is the applied stress, and a is the crack length. Simply put, K is a single factor that relates stress and crack size to the growth of a crack in a material. For more depth on this subject, see a fracture mechanics text, for example Ref 5 and 6, for a more detailed description of Y and the stress-intensity factor. Testing has shown that under controlled constant-amplitude fatigue loading, the crack-growth rate, da/dN, depends on the range of the stress-intensity factor, $\Delta K = K_{max} - K_{min}$, and is independent of geometry, provided the correct geometry correction, Y, is used (N is the number of load cycles). ΔK is calculated as:

$$\Delta K = Y\Delta\sigma\sqrt{\pi a} \qquad \text{(Eq 2)}$$

where the terms are as before, and $\Delta\sigma$ is the applied stress range ($\Delta\sigma = \sigma_{max} - \sigma_{min}$).

The crack-growth rate for a material and loading condition is measured using a standardized method, such as ASTM standard E 647 (Ref 7), and can be represented as a crack-growth rate versus stress-intensity factor range curve, as shown in Fig. 1. Here, the three regions of crack-growth behavior can be observed. The first, region I, is characterized by slow growth with a decreasing slope of the crack-growth rate curve. The threshold stress-intensity factor range, ΔK_{th}, bounds this region on the left and is operationally defined as the ΔK that produces a crack-growth rate of 1×10^{-10} m/cycle (Ref 6). The linear region (in log-log space), region II, is called the Paris law or power law region, because Paul Paris first noted that a power law would fit the crack-growth rate data in this region (Ref 8). That is:

$$\frac{da}{dN} = C\Delta K^n \qquad \text{(Eq 3)}$$

where C and n are constants that depend on material and loading conditions. Region III is the accelerating portion of the crack-growth curve that approaches the fracture toughness,
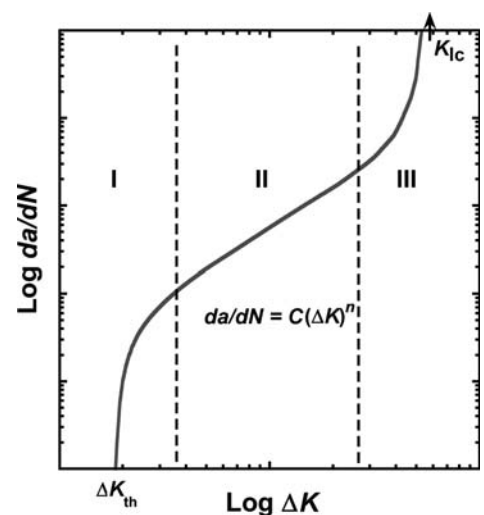
Fig. 1 Schematic of log crack-growth rate (da/dN) versus log applied stress-intensity factor range (ΔK) showing the decreasing crack-growth rate region (I), the linear (or Paris law) region (II), and the accelerating region (III)

K_{Ic}, for the material. From a damage-tolerance point of view, this portion of the curve is of little value, because there is usually precious little life left once the crack-growth rate starts to rapidly accelerate.

There are other, more elegant, ways to model the fatigue crack-growth curve that capture the nonlinearity near threshold and fracture toughness, that is, have a sigmoidal shape. One that does a good job in an empirical sense uses the hyperbolic sine function (Ref 9):

$$\log \frac{da}{dN} = C_1 \sin \mathrm{h}[C_2 \log(\Delta K) + C_3] + C_4 \qquad \text{(Eq 4)}$$

where C_1 is a material constant, and C_2, C_3, and C_4 are functions of load ratio, loading frequency, and temperature. A popular sigmoidally shaped crack growth rate equation is the NASGRO equation from Forman and Mettu (Ref 10):

$$\frac{da}{dN} = C \left[\left(\frac{1-f}{1-R} \right) \Delta K \right]^n \frac{\left(1 - \frac{\Delta K_{th}}{\Delta K} \right)^p}{\left(1 - \frac{K_{max}}{K_{crit}} \right)^q} \qquad \text{(Eq 5)}$$

where C, n, P, and Q are empirical constants and f is a function of the crack closure (crack closure is discussed later). ΔK_{th} is the crack growth threshold that is a function of crack closure and the stress ratio,

$$R = \frac{\sigma_{min}}{\sigma_{max}}$$

and K_{crit} is a thickness modified fracture toughness, K_{IC}. This formulation for crack growth is available in many commercially available crack growth codes.
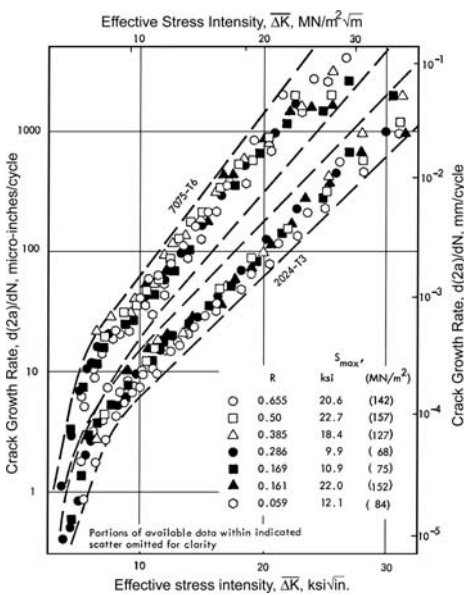
Load Interactions—Empirical Models

The previous models work nicely on simple constant-amplitude fatigue crack growth, which is seldom the case in service. The first additional loading term that must be included is the stress ratio, $R = \sigma_{min}/\sigma_{max}$, of the loading. Stress ratio has a pronounced effect on the crack-growth rate, and any variable-amplitude-load crack-growth model first must address this behavior. Walker (Ref 11) introduced an effective $\overline{\Delta K}$ such that:

$$\frac{da}{dN} = C\overline{\Delta K}^n \qquad \text{(Eq 6)}$$

where

$$\overline{\Delta K} = Y\sigma_{max}{}^{1-m}\Delta\sigma^m\sqrt{\pi a} = Y\sigma_{max}(1-R)^m\sqrt{\pi a}$$

and m is known as the Walker exponent and is experimentally determined. This formulation effectively collapsed the crack-growth curves in 2024-T3 and 7075-T6 aluminum for stress ratios ranging from 0.655 to 0.059 (Fig. 2). It should be noted that different values of the Walker exponent may be required to collapse positive and negative stress ratios to a single curve.

Forman et al. (Ref 12) had a similar formulation but argued that the crack-growth rate should become infinite as the crack reaches a critical size, that is, K_{max} approaches K_{Ic}. They found:

$$\frac{da}{dN} = \frac{C\Delta K^n}{(1-R)K_{Ic} - K_{max}} = \frac{C\Delta K^n}{(1-R)(K_{Ic} - K_{max})} \qquad \text{(Eq 7)}$$

where C and n are material- and frequency-dependent constants.

These two model forms are not significantly different and do not have general applicability but are useful in specific, limited regions of analyses. Generally, they effectively capture the stress-ratio effects and the term $(1 - R)^w$ is found in many "Walkerized" fatigue models. Effects of temperature, frequency, and corrosive environments are introduced into the models by conducting the experiments under those controlled conditions and fitting the model constants to the data. These models are only applicable to simple, constant-amplitude loading.

Real-world loads are incorporated in fatigue crack-growth life analyses using models that account for variable-amplitude or random loading. To illustrate how variable-amplitude loading affects the fatigue crack growth, consider constant-amplitude loading with only a single overload or a single overload/underload (fully reversed cycle) (Fig. 3) (Ref 13).

After the application of a single overload (line C), there is a profound retardation in the crack-growth rate. After a number of cycles and/or crack extension, the crack-growth rate is the same as that for a sample exposed to continuous cycling only (line A). Also note that overloads at longer crack lengths have a greater effect in both the number of cycles and the amount of crack extension (the ordinate is nonlinear) that the overload affects. The application of an overload/underload cycle (line B) resulted in a significantly smaller retardation, but the general trends are the same. The main drivers for retardation are the change in the crack tip plastic zone and the resultant change in crack wake plasticity that affects the crack closure. The overload substantially increases the monotonic plastic zone size (Fig. 4a). This stretched material must fit within the surrounding elastic material, so it results in a compressive zone at the crack tip (Fig. 4b). The crack-growth rate is reduced while the crack tip is within some fraction of the stretch zone. It then makes sense that the retardation is more pronounced when the crack length is longer—the overload plastic zone will be larger. Also, an overload/underload will partially reverse the plasticity at the crack tip so there is less retardation.

Fig. 2 Crack growth in 2024-T3 and 7075-T6 aluminum alloys at various stress ratios. Source: Ref 11
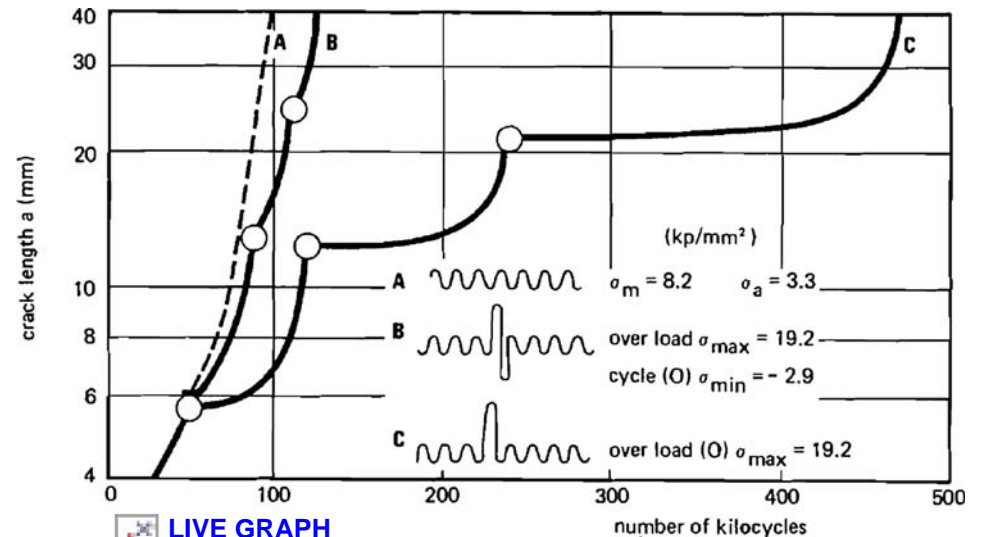
Fig. 3 Retardation as a result of overloads in 2024-T3 aluminum alloy. Source: Ref 13

The two most common load-interaction models deal with the size of the crack tip plastic zone and its relationship to the crack tip. The Wheeler retardation model (Ref 14) modifies the current crack-growth rate using a knock-down factor, γ, based on the ratio of the current yield zone to the difference of the effective crack length of an overload condition and the current crack-growth length. That is:

$$\frac{da}{dN} = \gamma \left(\frac{da}{dN} \right) \qquad \text{(Eq 8)}$$

where

$$\gamma = \left(\frac{R_y}{a_{\text{eff(ol)}} - a} \right)^m$$

$$R_y = \left(\frac{K_{\max}}{\sigma_{\text{YS}}} \right)^2 \left(\frac{1}{\pi PSX} \right)$$

and a is the current crack length, a_{eff} is the current crack length plus the remaining monotonic plastic zone size, σ_{YS} is the yield stress, PSX is the stress-state factor ($PSX = 2$ for plane stress; $PSX = 6$ for plane strain). The subscript (ol) indicates the overload condition that is updated each time a new maximum load is reached or the current yield zone, R_y, grows beyond the overload plastic zone size. Figure 5 shows the relationship between the overload plastic zone and the current plastic zone to help clarify the model approach. The Wheeler exponent, m, is determined empirically from test data. As the crack grows, the knock-down factor increases to 1 when the cyclic crack tip plastic zone reaches the edge of the overload plastic zone.

The generalized Willenborg model (Ref 15, 16) is more popular and is used more often than the Wheeler model. It uses an effective stress-intensity factor based on the size of the yield zone at the crack tip and can be viewed as affecting the maximum K in the cycle due to the overload compressive residual. That is:

$$K_{\max(\text{eff})} = K_{\max} - K_{\text{red}}$$

$$K_{\min(\text{eff})} = K_{\min} - K_{\text{red}}$$

where

$$K_{\text{red}} = \varphi \left(K_{\max(\text{ol})} \sqrt{1 - \frac{(a - a_{\text{(ol)}})}{R_{\text{y(ol)}}}} - K_{\max} \right)$$

$$\varphi = \frac{(1 - \Delta K_{\text{th}}/K_{\max})}{(\text{SOLR} - 1)}$$

$$R_{\text{y(ol)}} = \left(\frac{K_{\max(\text{ol})}}{\sigma_{\text{YS}}} \right)^2 \left(\frac{1}{\pi \, PSX} \right)$$

and a is the current crack length, $a_{\text{(ol)}}$ is the crack length at overload, ΔK_{th} is the threshold value of ΔK at $R = 0$, SOLR (shutoff overload ratio) is the overload required to arrest further crack growth at the nominal loading condition, σ_{YS} is the yield stress, and PSX is the stress-state factor ($PSX = 2$ for plane stress; $PSX = 6$ for plane strain). Again, the subscript (ol) indicates the overload condition that is updated each time a new maximum load is reached or the current yield zone, R_y, grows beyond the overload plastic zone size, $R_{\text{y(ol)}}$.

These two models do predict the number of cycles of retarded crack growth and the length of the affected growth reasonably well. There are limitations to these methods in that they require experimental calibration and cannot account for all levels of load interaction (Ref 17). They also do not address compressive loading or underload effects since under compressive loading there is no change in the crack tip plasticity!

Crack Closure

An alternative approach to account for load interaction uses the concept of crack closure, wherein the crack flanks come into contact before reaching the minimum load of the fatigue cycle, thereby reducing the crack-growth driving force. The overload and underload modifies the crack wake plasticity and affects the crack growth rate. Elber (Ref 18) was the first to discover crack closure in 1968 during his dissertation research. A once-controversial theory has now been generally accepted, and much literature exists to describe its mechanisms, measurements, analyses, and applications (Ref 19). Crack closure affects the total crack driving force, ΔK by affecting the minimum K of the cycle. The reduced driving force is the effective stress intensity factor range, ΔK_{eff}:

$$\Delta K_{\text{eff}} = K_{\max} - K_{\text{op}}$$

where K_{op} is the measured K that must be applied to the sample or structure to first open the crack. Typical crack-opening behavior is shown in Fig. 6. It can be observed that the crack is always open at high stress ratios and approaches a

Fig. 5 Relationship between the overload plastic zone and the current plastic zone

Fig. 6 Typical stress-intensity factor to open a crack, K_{op}, versus the stress-ratio relationship, R, for a fatigue crack
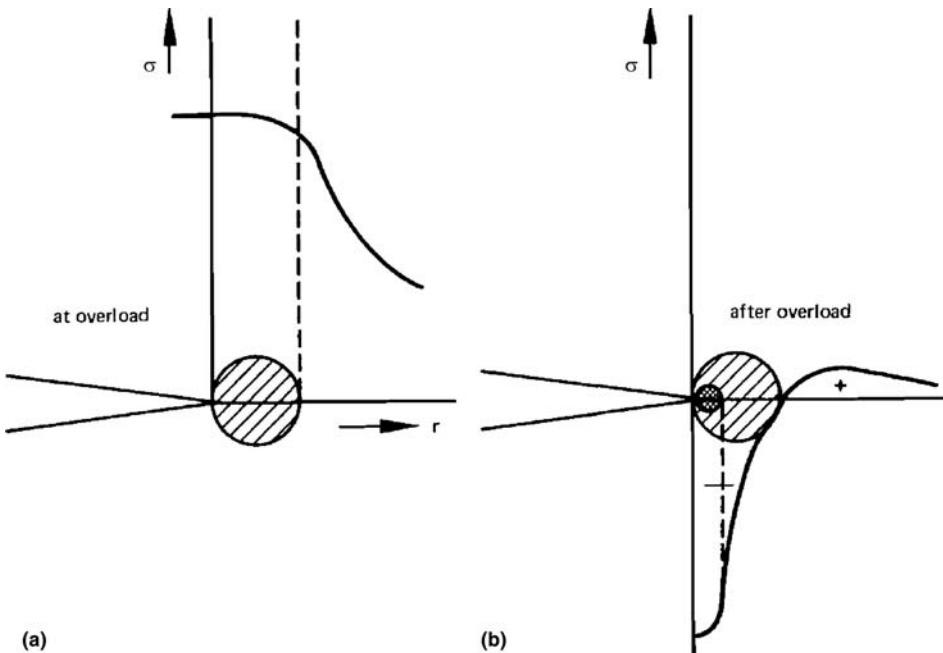
Fig. 4 Residual compressive stresses at a crack tip as a result of an overload. (a) Monotonic plastic zone. (b) Cyclic plastic zone after overload. Source: Ref 5

constant opening level at some value of negative stress ratio.

A popular approach to predict the crack opening load is to use a strip yield model to assess the plasticity-induced crack closure, and it can also predict the closure under variable loading cycles. The strip yield closure model is a Dugdale (Ref 20) model that is modified to leave plastically stretched material in the wake of the crack. The crack opening load is then calculated as the loading moment when the stress at the crack tip ligament crosses zero (Fig. 7). Newman (Ref 21) derived a simple set of equations to calculate the crack opening load:

$$\frac{K_{op}}{K_{max}} = A_0 + A_1 R + A_2 R + A_2 R^2 + A_3 R^3$$
$$\text{for} \quad R \geq 0$$

$$\frac{K_{op}}{K_{max}} = A_0 + A_1 R \text{ for} \quad R < 0$$

where A_0-A_3 are constants that depend on the constraint (α, of the crack, $\alpha = 1$ for plane stress, $\alpha = 3$ for plane strain). This model approach has been implemented in a computer code, FASTRAN-II (Ref 22), that can predict the crack growth under a range of spectrum loading conditions. Additions to this model approach (FASTRAN V5) are shown in the work of Newman et al. (Ref 23) that appears to improve its spectrum crack-growth prediction ability. These additions include refining the elements at the crack tip such that they are 5% or less of the plastic zone and improved crack opening stress calculations under spectrum loading. Crack growth in thin sheet materials is influenced by the loss in constraint at the crack tip as the

stress intensity factor increases. This results in a transition from flat to slant crack growth as ΔK increases. This is not physically modeled but incorporated by linearly transitioning from the plane-strain to plain-stress growth modes dependent on the crack-growth rates. Figure 8 shows crack growth in aluminum alloy 7075-T7351 subject to a wing gust and maneuver load spectrum (Ref 23). A comparison of predictions between the experimental results and the FASTRAN model shows that incorporating the constraint loss is important in this case.

Life-Prediction Codes. The models for crack growth life-prediction have been implemented in several, popular computer codes that are widely used to perform damage tolerance analyses on structures. NASGRO (Ref 24) was originally developed by NASA and includes many of the model approaches discussed above; various crack growth rate equations, load interaction models, crack closure and multiple K solutions. NASGRO also has an extensive database that can be used to perform analyses.

AFGROW is another computer code that is well formulated to incorporate load interaction models, closure models, and many crack geometries (Ref 25). This model can handle two independent cracks (under limited conditions and geometry) and empirically address their interaction. Cracks in plate structures can be modeled to include nonsymmetric corner cracks in holes under tension, bending, and bearing loading. As such, it is ideally suited to assess the structural integrity of built-up airframe structures.

These are just two examples of the damage tolerance life-prediction codes that are available for use. They are commercially supported

and continue to add new capabilities to improve their speed, accuracy and analytical suitability.

Geometric Considerations

The prior sections discuss the model approaches that are used to predict crack growth under simple and complex load spectra for known geometries that have a known stress-intensity factor, K. These geometries include fracture mechanics specimens as well as crack in notches, tubes, bolt holes, and so on subject to axial and bending loading. Cracking in real-world structures seldom have simple K solutions that can be found in a reference book (Ref 2). Weight function K solutions can be used for cracks having a stress gradient on the crack plane. 1-D stress gradients are commonly encountered at bolt holes or notches and were treated by Glinka and Reinhardt (Ref 3) and are available in some standard software, e.g. AFGROW (Ref 25). Now 2-D stress gradients can be addressed by weight functions (Ref 26) and will likely be coded soon. Several finite-element- or boundary-element-based tools exist that can be used to compute the K's and predict the shape of cracks in complex geometric structures subject to complex, multiaxial loading. In the early 2000's these model approaches were two-dimensional, but now fully three-dimensional methods are available, for example, FRANC3D (Ref 4) and Zencrack (Ref 27). These advanced tools can read meshes from commercial finite-element software, such

LIVE GRAPH
Click here to view

Fig. 8 Crack growth in 7075-T7351 aluminum alloy subject to spectrum loading. (a) Crack-growth comparison between model and experiment. LCD is a linear-cumulative damage model, and $\alpha = 1.8$ is a constant constraint model. (b) Wing gust and maneuver loading spectrum. Source: Ref 23

Fig. 7 Schematic of analytical crack-closure model under cyclic loading. (a) Maximum stress. (b) Minimum stress. Source: Ref 22

as ANSYS or ABAQUS, and allow the engineer to insert a crack into the geometry. The shape, size, and orientation of the crack are chosen, and the software automatically remeshes around the crack geometry. The cracked mesh is then sent back to the finite-element software for analysis. A nice example is shown in Fig. 9 for a crack in a spiral-bevel gear. This demonstrates the complexity of geometry that can be incorporated in this type of model.

These types of analyses incorporates mixed-mode crack growth wherein the opening, sliding, and tearing modes (K_I, K_{II}, and K_{III}, respectively) are important. These can be combined in terms of an equivalent energy-release rate, such that:

$$G_{equivalent} = \frac{1-v^2}{E}(K_I^2 + K_{II}^2) + (1+v)K_{III}^2 \text{ (Plane strain)}$$

$$G_{equivalent} = \frac{1}{E}(K_I^2 + K_{II}^2) + (1+v)K_{III}^2 \text{ (Plane stress)}$$

where E is the Young's modulus, and v is the Poisson's ratio. A clear example of mixed-mode crack growth is shown in Fig. 10 (Ref 28). Here, all three modes of cracking appear to be present.

These models accurately portray the cracked geometry and offer the possibility of obtaining the correct crack tip stress fields and driving forces by using a linear elastic fracture mechanics or energy-based continuum approach. Because of this, they can offer some insight into the forces that drive the crack, but they are still dependent upon the experimental crack-growth data that incorporates the stress-ratio effects, load-interaction testing, and so on. Often, they use empirical-based Wheeler or Willenborg (Ref 14, 16) retardation models, because they do not fully capture the deformation history at the crack tip. The material is treated as a continuum (either isotropic or anisotropic) but cannot capture the complex stress-strain history at the crack tip any better
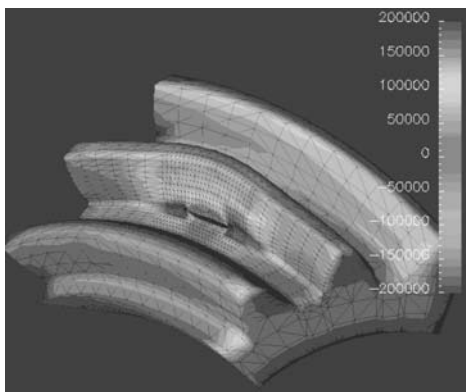
than the constitutive model that is used in the finite-element model.

Recommendations for Future Work

Crack-growth models have been incorporated in a large number of commercial applications where the cost of failure is high. Damage-tolerant structural analysis is here to stay and appears to be growing in the commercial aviation, power generation, and transportation sectors. It will, however, require improvements in the analysis tools and models to reduce the cost of a damage-tolerance analysis. The most advanced crack-growth models still fail to seamlessly track load interaction and do not account for the actual mechanisms of crack growth.

Crack growth is a complex phenomenon depending on the loading and load interaction as well as the geometry of the cracking structure and its multiaxial loading. Currently, tools such as FRANC3D (Ref 3) and Zencrack (Ref 27) do an excellent job of capturing the driving force from a continuum point of view but operate outside of, or on the fringe of, the design finite-element environment. Integrating this capability within the finite-element software could make crack-growth assessments easier to use and therefore more often used in design. They do lack a true predictive capability because they depend on experimental data and other representations of the crack-growth behavior for their prediction of growth. The more-analytic models to predict the crack-growth behavior, that is, Wheeler or Willenborg (Ref 14, 16), do capture the trends in crack growth but lack a strong physical basis in their
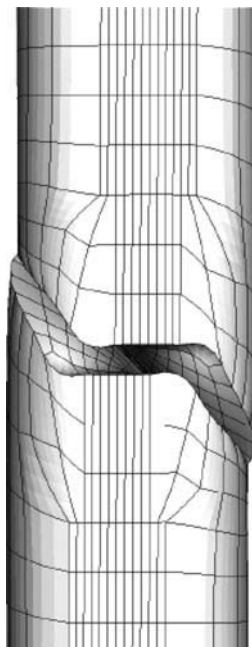
prediction. It is critical that new, advanced models are developed that take into account the deformation history at the crack tip, the mechanisms of cracking, and the slip character at the crack tip.

Future advances will be seen by a continued linking of experiment and analysis. There may be benefit from applying molecular dynamics tools to look at the mechanisms of crack advance, although this appears to be just beginning. The application of cohesive zone models for fatigue crack growth that results in intergranular failure due to environmental or chemical attack may lead to better predictive model approaches that depend less on experimental crack-growth calibration. All modeling efforts should have the goal of developments that will require less experimental calibration and take into account the mechanisms of crack growth. These are high goals, but the advance in computational tools is still very bright.

REFERENCES

1. "General Guidelines for Aircraft Structural Integrity Program," Report MIL-HDBK-1530B, ASC/ENOI, Wright-Patterson AFB, OH, 2002.
2. H. Tada, P.C. Paris, and G.R. Irwin, *The Stress Analysis of Cracks Handbook,* ASME, New York, 2000.
3. G. Glinka and W. Reinhardt, Calculation of Stress Intensity Factors for Cracks of Complex Geometry and Subjected to Arbitrary Nonlinear Stress Field, *Fatigue Fract. Mech.,* Vol 31, ASTM STP 1389, ASTM International, West Conshohocken, PA, 2000, p 348–370.
4. FRANC3D software is described at http://www.cfg.cornell.edu/index.htm.
5. D. Broek, *Elementary Engineering Fracture Mechanics,* Martinus Nijhoff Publishers, Dordrecht, Netherlands, 1986.
6. S.D. Antolovich and B.F. Antolovich, An Introduction to Fracture Mechanics, *Fatigue and Fracture,* Vol 19, *ASM Handbook,* ASM International, 1996, pp. 371–380.
7. "Standard Test Method for Measurement of Fatigue Crack Growth Rates," E 647-05, *Annual Book of Standards,* Section 3, Vol 3.01, ASTM International, West Conshohocken, PA, 2007.
8. P.C. Paris, "The Growth of Cracks due to Variations in Load," Ph.D. thesis, Lehigh University, 1962.
9. R.M. Wallace, C.G. Annis, Jr., and D. Sims, Report AFML-TR-76-176, part 2, 1976.
10. R.G. Forman and S.R. Mettu, Behavior of Surface and Corner Cracks Subjected to Tensile and Bending Loads in Ti-6Al-4V Alloy, *Fracture Mechanics* 22nd *Symposium,* Vol 1, ASTM STP 1131, H.A. Ernst, A. Saxena, and D.L. McDowell, Ed., ASTM International, West Conshohocken, PA, 1992.

Fig. 9 Simulated crack growth in spiral-bevel gears. The model provides color images. The largest values are at the ends of the crack. The smallest values (compressive) are in the concave area above the crack. Source: Ref 4

Fig. 10 Simulated crack growth in a shaft subject to axial and torsional loading. Source: Ref 28

11. K. Walker, *The Effect of Stress Ratio during Crack Propagation and Fatigue for 2024-T3 and 7075-T6 Aluminum,* STP 462, ASTM, 1970, pp. 1–14.

12. R.G. Forman, V.E. Kearney, and R.M. Engle, *Trans. ASME, Series D, J. Basic Eng.,* Vol 89 (3), 1967, pp. 459–464.

13. J. Schive and D. Broek, Crack-Propagation Tests Based on a Gust Spectrum with Variable Amplitude Loading, *Aircr. Eng.,* Vol 34, 1962, pp. 314–316.

14. O.E. Wheeler, Spectrum Loading and Crack Growth, *Trans. ASME, J. Basic Eng.,* 1972, pp.181–186.

15. J. Willenborg, R.M. Engle, and H.A. Wood, "A Crack Growth Model Using an Effective Stress Concept," Report AFFDL-R71-1, Air Force Flight Dynamics Laboratory, Wright-Patterson AFB, OH, 1971.

16. J.P. Gallagher, "A Generalized Development of Yield-Zone Models," Report AFFDL-TM-74-28, Air Force Flight Dynamics Laboratory, Wright-Patterson AFB, OH, 1974.

17. M.V.S Pereira, F.A.I. Darwish, A.F. Camarão, and S.H. Motta, On the Prediction of Fatigue Crack Retardation Using Wheeler and Willenborg Models, *Mater. Res.,* Vol 10 (2), 2007, pp. 101–107.

18. W. Elber, Fatigue Crack Closure under Cyclic Tension, *Eng. Fract. Mech.,* Vol 2, 1970, pp. 37–45.

19. J.C. Newman, Jr. and W. Elber, Ed., *Mechanics of Fatigue Crack Closure,* STP 982, ASTM, Philadelphia, PA, 1988.

20. D.S. Dugdale, Yielding of Steel Sheets Containing Slits, *J. Mech. Phys. Solids,* Vol 8 (2), 1960, pp. 100–104.

21. J.C. Newman, Jr., Prediction of Fatigue Crack Growth under Amplitude and Spectrum Loading Using a Closure Model, *Design of Fatigue and Fracture Resistant Structures*, ASTM STP 761, American Society for Testing and Materials, Philadelphia, PA, 1982, p 255–277.

22. J.C. Newman, Jr., "FASTRAN II—A Fatigue Crack Growth Structural Analysis Program," NASA TM-104159, 1992.

23. J.C. Newman, Jr., A. Brot, and C. Matias, Crack-Growth Calculations in 7075-T7315 Aluminum Alloy under Various Load Spectra Using an Improved Crack-Closure Model, *Eng. Fract. Mech.*, Vol 17, 2004, pp. 2347–2363.

24. NASGRO Software described at http://www.swri.org/4org/d18/mateng/matint/NASGRO/default.htm

25. J.A. Harter, AFGROW Users Guide and Technical Manual, AFGROW for Windows XP/VISTA, Version 4.0012.15, AFRL-VA-WP-TR-2008-XXXX, Air Vehicles Directorate, Air Force Research Laboratory, Wright-Patterson AFB, OH, 2008.

26. A. Jankowiak, H. Jakubczak, and G. Glinka, Fatigue Crack Growth Analysis Using 2-D Weight Function, Int. J. Fatigue, Vol 31 (Issues 11-12), November–December 2009, p 1921–1927

27. Zencrack software is described at http://www.zentech.co.uk/zencrack_overview_what.htm.

28. R. Chandwani and C. Timbrell, "Simulation of 3-D Non-Planar Crack Propagation," NAFEMS World Congress 2007, May 22–25, 2007 (Vancouver, Canada).

Neural-Network Modeling

H.K.D.H. Bhadeshia and H.J. Stone, University of Cambridge

EMPIRICAL METHODS are regarded as less desirable than those based on physical principles, primarily because the former are perceived to be limited to the knowledge base on which they are created and hence may not generalize well. The interpolation or extrapolation of regression equations is associated with unspecified uncertainty, even when the fit with the known domain is good. Yet, when dealing with complicated mechanical properties such as ductility, fatigue, or creep, physical models are less than useful in dealing with the complexities of technology and often are limited to making qualitative and simplistic inferences.

Empirical modeling has, however, taken a turn for the better with the advent of neural networks, which permit the discovery of fundamental relationships and quantitative structure within vast arrays of ill-understood data. The significant factor in this success has been in the understanding of the rules for creating robust models and in the treatment of noise and uncertainties. It is now well established that the method is not only capable of representing known data but can also lead to the discovery of novel concepts (Ref 1 to 3). To begin, the method is introduced.

The Method

In regression analysis, the data are best–fitted to a specified relationship that is usually linear. Each of the inputs, x_j, is multiplied by a weight, w_j; the sum of all such products and a constant, θ, then gives an estimate of the output $y = \sum_j w_j x_j + \theta$. Equations such as these are used widely, for example, in estimating the yield strength (YS) of certain weld metals as a function of the chemical composition:

$$YS \ (MPa) = 326 + 80 \times x_{Mn}$$
$$\text{for } 0.6 x_{Mn} 1.85 \text{ with standard}$$
$$\text{error } \pm 18$$

$$(Eq \ 1)$$

where x_{Mn} is the concentration of manganese in weight percent. The standard error is a reflection of the noise in the data, that is, changes in the strength as a result of uncontrolled variables. It is well understood that there is risk in

using the relationship beyond the range of the fitted data, but the risk is not quantified.

With neural networks, the input data x_j are again multiplied by weights, but the sum of all these products forms the argument of a flexible mathematical function, often a hyperbolic tangent. The output y is therefore a nonlinear function of x_j. The shape of the hyperbolic tangent can be changed by altering the weights (Fig. 1a). Additional flexibility in the nature of the mathematical function used to represent the data can be introduced by combining several of these hyperbolic tangents (Fig. 1b). In this way, the network is able to capture almost arbitrarily nonlinear relationships.

Figure 2 illustrates the complexity of the three-dimensional surface that can be produced when representing the output (vertical axis) as a function of two inputs using just five hyperbolic tangents.

Overfitting

Given that the function representing the neural network can be engineered to comply with any set of data, including a random set, it is possible to overfit. One may be modeling noise rather than an underlying relationship. It is not clear from Fig. 3, in the absence of any guiding physical principles relating x to y, whether it is

the straight line or the polynomial curve that is the correct function to extrapolate in order to make a prediction.

This particular problem is relatively easy to overcome. The model can be tested to avoid this difficulty. The experimental data can be divided into two sets: a training dataset and a test dataset. The model is produced using only the training data. The test data are then used to check that the model behaves itself when presented with previously unseen data. This is illustrated in Fig. 4, which shows three attempts at modeling noisy data for a case where y should vary with x^4. The solid data points represent the arbitrary training data, and the open circles are the test data. A linear model (Fig. 4a) is too simple and does not capture the real complexity in the data. An overcomplex function, such as Fig. 4(c), accurately models the training data but generalizes badly. The optimum model is illustrated in Fig. 4(b). The training and test errors are shown schematically in Fig. 4(d). As expected, the training error tends to decrease continuously as the model complexity increases. The minimum in test error represents the optimum choice of model complexity; having established this, the entire data can be used to retrain the model without changing its complexity from the optimum configuration. There are other techniques to supplement this procedure (Ref 4, 5).
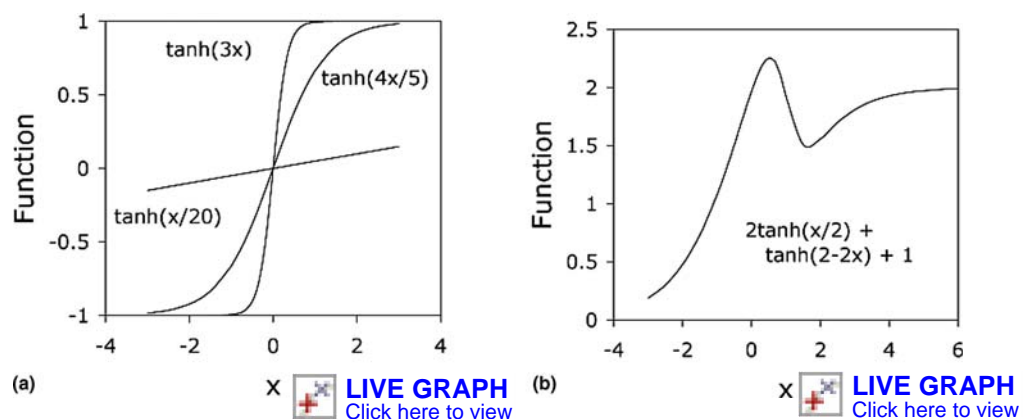
Fig. 1 Fitting hyperbolic tangent functions. (a) Three hyperbolic tangent functions, generated by varying the weights, illustrate the flexibility of the function. (b) A combination of two hyperbolic tangents produces a more complex model.
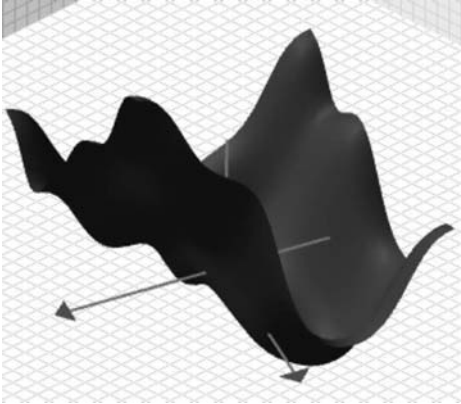
Fig. 2 Variation in the output (vertical axis) as a function of two input variables (horizontal axes). The complex surface is generated using just five hyperbolic tangent functions.
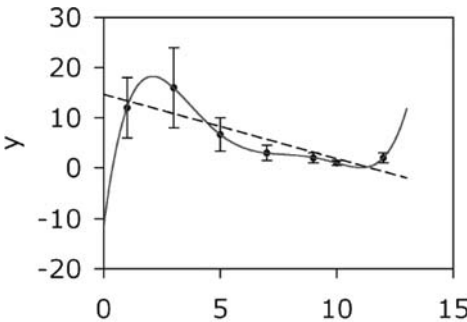
LIVE GRAPH
Click here to view

Fig. 3 The points represent noisy data. There are two fitted functions: a straight line and a nonlinear polynomial curve.

Noise and Uncertainties

There are two kinds of errors to consider when conducting experiments. Noise results in a different output for the same set of inputs when an experiment is repeated. This is because there are variables that are not controlled, so their influence is not included in the analysis. The second kind deals with the uncertainty of modeling. There may exist many mathematical functions that adequately represent the same set of empirical data but behave differently in extrapolation. The noise in the output can be assessed by comparing the predicted values (y_j) of the output against those measured (t_j), for example:

$$E_D \propto \sum_j (t_j - y_j)^2 \text{ or } \sigma = \sqrt{\sum_{j=1}^{n} \frac{(t_j - y_j)^2}{n^2}} \quad \text{(Eq 2)}$$

where E_D and σ_{SE} are expected to increase if important input variables have been excluded from the analysis. Figure. 5(a) shows this kind of noise, where data have been fitted to a straight line with a standard error of $\pm 2\sigma$ in the estimation of the output y. E_D and σ_{SE} have
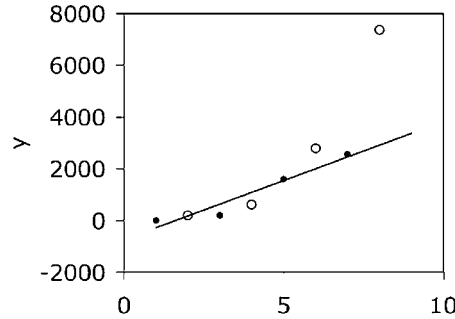
(a) LIVE GRAPH × Click here to view
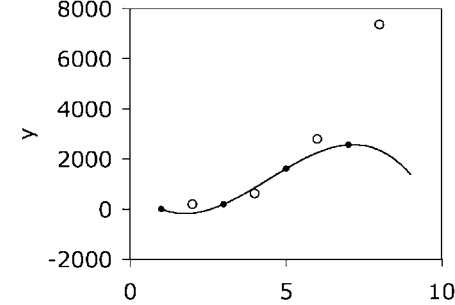
(b) LIVE GRAPH × Click here to view   LIVE GRAPH Click here to view
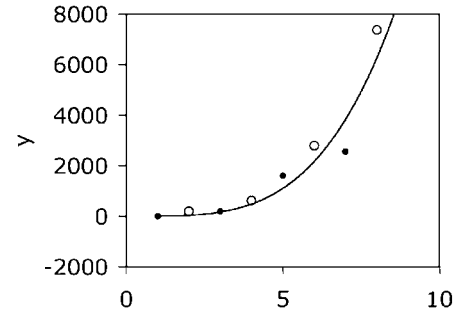
(c) LIVE GRAPH × Click here to view

(d)
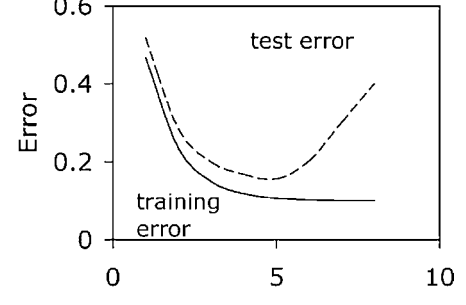
Fig. 4 Test and training errors as a function of model complexity for noisy data in a case where y should vary with x^4. The solid data points were used to create the models (i.e., they represent training data), and the open circles constitute the test data. (a) Linear function that is too simple. (b) Fourth-order polynomial with optimum representation of both the training and test data. (c) Fifth-order polynomial that generalizes poorly. (d) Test and training errors as a function of the model complexity

(a)

(b)
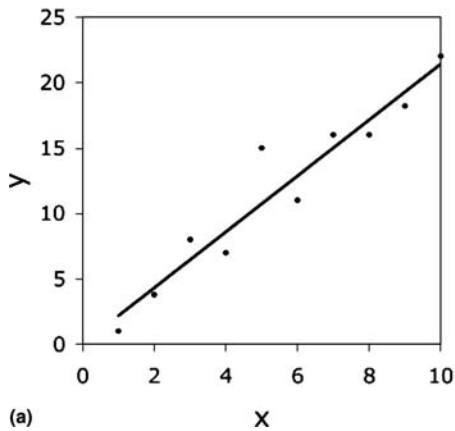
Fig. 5 Types of error. (a) Noise. (b) Uncertainty   LIVE GRAPH Click here to view   LIVE GRAPH Click here to view
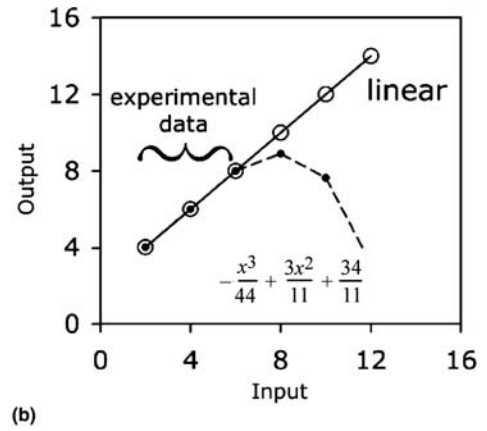
fixed values and indicate noise but are not satisfying descriptions of the uncertainties of prediction. In contrast, the uncertainty of modeling is illustrated in Fig. 5(b), where a set of precise data ($x = 2, 4, 6$) is fitted to two functions, one linear and the other nonlinear (Ref 6):

$$y = -x^3/44 + 3x^2/11 + 34/11 \quad \text{(Eq 3)}$$

Both of the functions illustrated reproduce these experimental data precisely but behave quite differently when extrapolated (or indeed, interpolated, for $x = 3$, $y = 4.931$, not $y = 5$ according to the linear function $y = x +2$). The

difference in the predictions of the two functions in domains where data do not exist is a measure of the uncertainty of modeling, since both functions correctly represent the data $x = 2, 4, 6$ used in creating the models. The magnitude of the modeling uncertainty is not constant but varies as a function of the position in the input space. It becomes larger in domains of the input space where knowledge is sparse or nonexistent (Fig. 5b). One way of representing the uncertainty is to create a large variety of models, all of which reasonably represent the experimental data. The predictions made by these models will not be identical; the standard
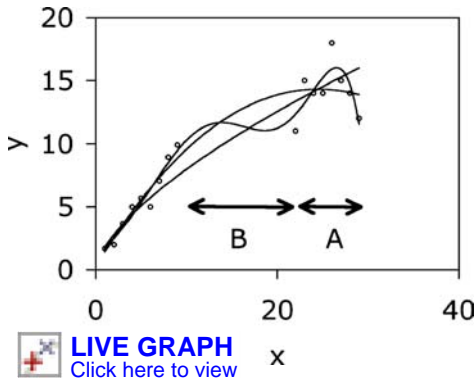
Fig. 6 Illustration of the uncertainty in defining a fitting function in regions where data are sparse (B) or where they are noisy (A). Three possible functions are shown.

Fig. 7 Comparison of the predictions of a neural-network model (central curve) with experimental data that were not included in the creation of the model. The modeling uncertainty is indicated by the upper and lower error bounds (broken lines). Source: Ref 6, 8

deviation in the predicted values then is a quantitative measure of the modeling uncertainty.

Figure 6 illustrates the problem of modeling uncertainty. The practice of using the best–fit function (i.e., the most probable values of the weights) does not adequately describe the uncertainties in regions of the input space where data are sparse (region B) or where the data are particularly noisy (region A).

In the Bayesian method (Ref 4, 5), the modeling uncertainty is expressed by not having a unique set of weights but rather a probability distribution of sets of weights. This recognizes the existence of many functions that can be fitted or extrapolated into uncertain regions of the input space, without unduly compromising the fit in adjacent regions that are rich in accurate data. The error bars depicting the modeling uncertainty then become large when data are sparse or locally noisy, as illustrated in Fig. 6.

Suppose there exist two models $\mathcal{H}_1$ and $\mathcal{H}_2$. Bayes' theorem can then be used to find the probability ratio between the two theories as follows (Ref 7):

$$\frac{P(\mathcal{H}_1|D)}{P(\mathcal{H}_2|D)} = \frac{P(\mathcal{H}_1)}{P(\mathcal{H}_2)} \times \frac{P(D|\mathcal{H}_1)}{P(D|\mathcal{H}_2)} \qquad \text{(Eq 4)}$$

where $P(\mathcal{H}_i|D)$ is the plausibility of $\mathcal{H}_i$ given the data D and $P(\mathcal{H}_i)$ is the prior plausibility of $\mathcal{H}_i$. The first ratio on the right measures how much our initial beliefs favor $\mathcal{H}_1$ over $\mathcal{H}_2$, and the second ratio measures how well the observed data were predicted by $\mathcal{H}_1$ compared to $\mathcal{H}_2$ (Ref 7). The second ratio also incorporates Occam's razor because a more complex model will be capable of making a greater variety of predictions and hence its predictive probability $P(D|\mathcal{H})$ is more thinly spread out over the data, so if both a simple and complex model are capable of representing the data, then the simple model is more probable.

A neural network is capable of producing a vast range of models depending, for example, on different architectures, starting values of weights, etc. In the Bayesian framework, the
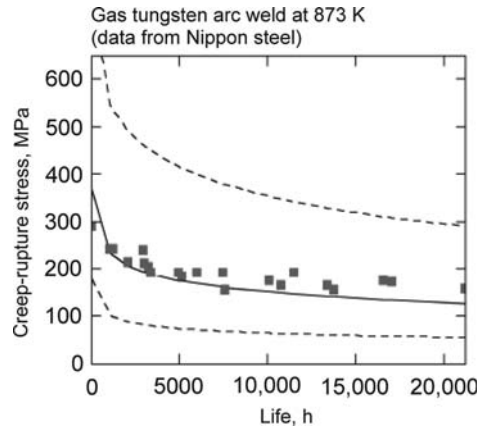
posterior probability of the weight parameters $\mathbf{w}$, $P(\mathbf{w}|D, \mathcal{H})$, for a particular model is assessed using the Bayes' theorem and in this way a probability distribution of $\mathbf{w}$ is obtained, the width of which is related to what we have described as the modeling uncertainty.

This methodology has proved to be extremely useful in materials science (Ref 1, 2), where properties must be estimated as a function of a vast array of inputs. It is then most unlikely that the inputs are uniformly distributed in the input space.

It is important to note that a large modeling uncertainty implies a lack of knowledge (sparse or noisy data). There are advantages to making uncertain predictions in the present context. A large uncertainty indicates a domain where it is useful to conduct new research. A small uncertainty, by contrast, could be interpreted to mean that new experiments are not needed. The estimate of uncertainty can therefore be used to design experimental programs.

A large modeling uncertainty does not necessarily imply that the estimate is incorrect. Figure 7 shows that, in spite of the calculated uncertainties (the upper and lower bounds indicated by the two broken lines), novel data are accurately predicted because of the robust design of the neural network, based on meaningful inputs (Ref 8).

Transparency

Neural networks are sometimes referred to as black boxes, presumably because those who use them cannot access the mathematical functions that describe the relationship between the inputs and outputs. Their mysterious nature is compounded by making inappropriate comparisons to biological systems and synapses. A complete mathematical description of the network is actually straightforward.

It is a fact, however, that a network is a simple combination of transfer functions (e.g., hyperbolic tangents) and weights. The number of hyperbolic tangents used is said to be the number of hidden units. The function for a network with i hidden units, connecting the inputs x_j to the output y, is given by:

$$y = \sum_i w_i^{(2)} h_i + \theta^{(2)}$$

$$\text{with } h_i = \tanh\left(\sum_j w_{ij}^{(1)} x_j + \theta_i^{(1)}\right) \qquad \text{(Eq 5)}$$

where w represents weights and θ the constants, as described in the context of linear regression. The influence of the inputs on the output variable is, together with the transfer functions, implicit in the values of the weights. The weights may not always be easy to interpret, given that there may be high–order interactions between the variables. For example, there may exist more than just pairwise interactions, in which case the problem becomes difficult to visualize from an examination of the weights. This visualization problem is a feature of all nonlinear methods but is not a limitation, because it is simple to use the trained network to make predictions, plot them, and to see how these depend on various combinations of inputs.

Examples

Examples of the application of neural-network modeling to the behavior of metals include the following.

Example 1: Fatigue Crack Growth. It is understood that fatigue crack growth is a consequence of the accumulation of damage by deformation in the plastic zone at the crack tip. At low loads the deformation is governed by the cyclic variation in the stress–intensity range ΔK. The crack extension per cycle (da/dN) becomes measurable at a threshold ΔK_{th}, followed by the slower extension rate in the Paris Law regime (Ref 9–12) described by the proportionality

$$\log \Delta K \propto \log\left\{\frac{da}{dN}\right\}^m \qquad \text{(Eq 6)}$$

where da/dN is the average crack advance per cycle, and m is known as the Paris exponent. The equation can be interpreted in terms of a variety of physical mechanisms (Ref 13, 14) in which case the proportionality constant (C) becomes a function of the Young's Modulus E, the Poisson's ratio v, and the yield and ultimate tensile strengths σ_Y and σ_U respectively. Based on the possible mechanisms consistent with the Paris Law, attempts have been made to generally interpret fatigue crack growth data on the basis of just the mechanical properties of the material (Ref 13, 14), Duggan (Ref 15) for example, expressed the crack growth rate in terms of the elastic modulus, toughness, and ductility:

$$\frac{da}{dN} = \left(\frac{\pi}{32}\right)^{\frac{1}{2\eta}}\frac{1}{\eta}\left[\frac{2}{\varepsilon_f E(K_{I_c} - K_{max})}\left(1 - \frac{K}{K_{I_c}}\right)\right]^{\frac{1}{\eta}} K^{\frac{2}{\eta}}$$

(Eq 7)

where η and ε are the fatigue ductility exponent and coefficient respectively, E is the elastic modulus, K_{I_c} is the critical stress intensity for fracture. Since then, there have been a number of other similar models in which fatigue is expressed as a function of mechanical properties alone, without making explicit reference to the material type. Dimitriu followed this principle to design a neural network model based on the elongation, tensile strength, proof strength, specimen dimensions, stress ratio and frequency of loading (Ref 16). The network was created on the basis of a very large quantity of published data on steels. Since the model did not contain material-specific inputs, it succeeded in making predictions on nickel, titanium and aluminium alloys without modification. In other words, the model generalised well to materials not included in the training data; the accuracy of this generalization is illustrated in Fig. 8. The essential point here is that a well-designed neural network model, one which makes use of known physics, can make predictions well beyond the range of the training information.

Example 2: Creep Rupture and Hot-Strength. Creep rupture tests are expensive to conduct because to obtain results representative of service, the tests may require many years before rupture occurs. It was discovered using neural network analysis (Ref 17) that the temperature dependence of the strength of creep-resistant steel cannot be explained by the variation in the intrinsic strength of pure iron or of solid-solution strengthening as a function of temperature (Fig. 9a). The hot-strength can be categorized into two temperature regimes, one where there is only a minor variation, followed by a more dramatic decline beyond about 780 K. This latter regime is explained by the increased ability of dislocations to overcome obstacles with the help of thermal activation. It is therefore not surprising that the temperature sensitivity to hot-strength at high temperatures is replicated by that of creep rupture data (Fig. 9b). This suggests that hot-strength tests could be used as indicators of the temperature sensitivity of creep rupture data, in which case the number of rupture tests required can be dramatically reduced. This second example demonstrates how a well-designed network is able to capture physical phenomena which can then be exploited more widely than the original intention in creating the model.

For more on the training of neural networks and their uses, see the article "Application of Neural-Network Models" in this Volume.

REFERENCES

Fig. 8 Predictions represented by the uncertainty range, and experimental data presented as points. The model responsible for the predictions was trained only on steel data. (a) A bearing steel not included in the data used for training the model. (b) A nickel based alloy Udimet 700. (b) Inconel 718 with data dues to (Ref 18). (c) A titanium alloy. (d) An aluminium alloy.
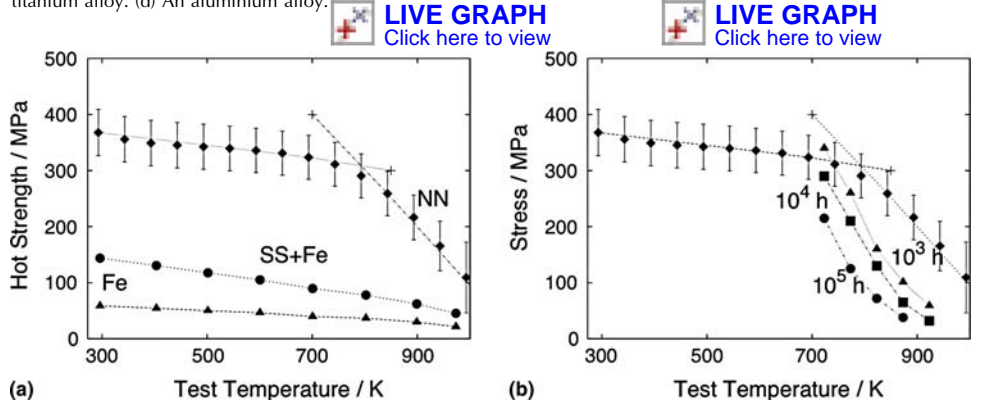
Fig. 9 (a) The hot–strength of 2¹/₄Cr steel. Fe represents the strength of pure, annealed iron, SS the contribution of solid–solution strengthening, and NN is the total strength estimated using a neural network model. (b) Comparison of temperature–sensitivity of creep–rupture (lower three curves) and proof strength. The times indicated are for failure by rupture.

1. H.K.D.H. Bhadeshia, Neural Networks in Materials Science, *ISIJ Int.*, Vol 39, 1999, p 966–979

2. H.K.D.H. Bhadeshia, Neural Networks and Information in Materials Science, *Stat. Anal. Data Mining*, Vol 1, 2009, p 296–305

3. H.K.D.H. Bhadeshia, R.C. Dimitriu, S. Forsik, J.H. Pak, and J.H. Ryu, On the Performance of Neural Networks in Materials Science, *Mater. Sci. Technol.*, Vol 25, 2009, p 504–510

4. D.J.C. MacKay, Practical Bayesian Framework of Back Propagation Networks, *Neural Comp.*, Vol 4, 1992, p 448–472

5. D.J.C. MacKay, Bayesian Non-Linear Modeling for the Energy Prediction Competition, *ASHRAE Trans.*, Vol 100, 1994, p 1053–1062

6. D.J.C. MacKay, Probability Theory and Occam's Razor, *Darwin College Magazine*, March 1993, p 81–85

7. D.J.C. MacKay, Probable Networks and Plausible Predictions – A Review of Practical Bayesian Methods for Supervised Neural Networks, *Comput. Neural Sys.*, Vol 6, 1995, p 469–505

8. D. Cole, C. Martin-Moran, A. Sheard, H.K.D.H. Bhadeshia, and D.J.C. MacKay, Modelling the Creep Rupture Strength of

Ferritic Steel Welds, *Sci. Technol. Weld. Join.*, Vol 5, 2000, p 81–89

9. P.C. Paris, M.P. Gomez, and W.E. Anderson, A Rational Analytic Theory of Fatigue Life, *Trend. Eng.*, Vol 13, 1961, p 9–14

10. P.C. Paris and F. Erdogan, A Critical Analysis of Crack Propagation Laws, *J. Basic Eng. (Trans. ASME)*, Vol 85, 1963, p 528–534

11. P.C. Paris, The Fracture Mechanics Approach to Fatigue, *Proc. 10th Sagamore Conf*, 1965, p 107–132

12. R.O. Ritchie, Near Threshold Fatigue-Crack Propagation in Steels, *Int. Metal. Rev.*, Vol 4, 1979, p 205–230

13. L.N. McCartney and P.M. Cooper, A New Method of Analysing Fatigue Crack Propagation Data, *Eng. Fracture Mech.*, Vol 9, 1974, p 273–290

14. K.H. Schwalbe, Comparison of Several Fatigue Crack Propagation Laws with Experimental Results, *Eng. Fracture Mech.*, Vol 6, 1974, p 325–341

15. T.V. Duggan, A Theory for Fatigue Crack Propagation. *Eng. Fracture Mech.*, Vol 9, 1977, p 735–747

16. R.C. Dimitriu, "Complex Mechanical Properties of Steels," Ph.D thesis, University of Cambridge, Cambridge, U.K., 2009

17. R. Dimitriu and H.K.D.H. Bhadeshia, Hot-Strength of Creep–Rupture Data, *Mater. Sci. Technol.*, Vol 23, 2007, p 1127–1131

18. R.L. Tobler, Low Temperature Effects on the Fracture Behaviour of a Nickel Base Super-alloy. *Cryogenics*, Vol 1, 1976, p 669–674

Material Fundamentals

Phase Equilibria and Phase Diagram Modeling

Y.A. Chang and H.-B. Cao, University of Wisconsin
S.-L. Chen, F. Zhang, Y. Yang, W. Cao, and K. Wu, CompuTherm, LLC

A PHASE DIAGRAM is a graphical representation of the phase equilibria of materials in terms of temperature, composition, and pressure. However, since most, if not all, of the phase diagrams in the field of materials science and engineering are presented at a constant pressure of 1 bar, all diagrams presented in this article are at 1 bar unless noted otherwise. The most familiar phase diagram of a binary is the temperature-composition diagram, and that of a ternary is an isothermal section displaying the compositions of phases in equilibrium with each other using a Gibbs triangular paper, that is, an equal-angle triangular paper. Two other common diagrams of a ternary are an isopleth, that is, a temperature-composition section, with one constraint such as keeping the composition of one component constant, and a projection of the liquidus or solidus surface onto a two-dimensional Gibbs triangular paper. However, an isopleth presents only the stability of phase fields; the tie lines that represent phase equilibria between two phases normally do not lie in the plane of the temperature-composition section. A pseudobinary is a special case of an isopleth when the two end phases exhibit a small range of homogeneity and melt congruently. In such a diagram, the tie lines do lie in the temperature-composition plane, just like a binary phase diagram. The most frequently used phase diagram of a multicomponent (*n*) system beyond binaries is an isopleth, that is, a temperature-composition section with (*n* – 2) constraints—one for a ternary, two for a quaternary, three for a quinary, and so on.

This article provides an overview on the background of phase diagram calculation software. Significant advancements in calculating a variety of multicomponent phase diagrams have been possible with the development of reliable thermodynamic databases and the robustness of the computer software. This has been greatly enhanced with the availability of second-generation phase diagram calculation software. It is noteworthy to point out that many of the first-generation software also have since attempted to calculate the stable diagrams. After the overview, the following topics are discussed in this article:

- An algorithm to calculate binary stable phase equilibria
- A rapid method to obtain a thermodynamic description of a multicomponent system
- Thermodynamically calculated phase diagrams
- Concluding remarks

Overview and Background

Even though phase diagrams and thermodynamics are two aspects of the same subject, historically there are two groups of researchers in studying phase equilibria and thermodynamics. One group of researchers focuses on establishing phase equilibria experimentally using techniques such as cooling/heating curves, x-ray diffraction (XRD), electron probe microanalysis (EPMA), and microstructure characterization using optical, scanning electron, and transmission electron microscopy, denoted as OM, SEM, and TEM, respectively. The other group measures the thermodynamic properties of materials, such as the chemical potentials of the component elements in solutions, using techniques of vapor pressure, galvanic cells, equilibration, and so on and the enthalpies of formation and the specific heats of single phase using a variety of calorimetric methods. Kubaschewski and Alcock (Ref 1) gave a summary of the many techniques used in practice. More recently, many researchers obtained the enthalpies of formation of phases at 0 K using first-principle calculations. In many cases, the calculated values approach the accuracy of those measured experimentally.

On the one hand, the goal of these experimentally measured and theoretically calculated thermodynamic quantities is to attain a basic understanding of the stabilities of the phases and, on the other, to obtain the Gibbs energies of single-phase materials as a function of composition and temperature, and occasionally also of pressure primarily in the geological community, for practical applications. In principle, when the Gibbs energies of all the phases in

an alloy system are known as functions of temperature and composition, it becomes possible to calculate the phase diagrams of the system in question, as shown in Fig. 1 (Ref 2). It is worth noting that more than a century ago, Van Laar (Ref 3, 4) calculated a large number of prototype binary phase diagrams using the one-parameter regular solution model. Approximately half a century later, Meijering extended these calculations from binaries to ternaries (Ref 5–9). Although these calculated phase diagrams in many cases cannot be used for practical applications, they had made a paramount contribution to understanding the relationships between the topological features of phase equilibria in terms of the relative stabilities of the phases involved (Ref 10).

Even though the phase diagram of an alloy system can, in principle, be calculated from the Gibbs energies of the constituent phases, it is, in practice, very challenging to achieve this goal, because a small uncertainty in the Gibbs energy values leads to a different type of phase equilibrium. The experimental phase equilibrium data are essential to calculate a correct phase diagram in many cases. In this article, the phenomenological or Calphad approach is used to calculate phase diagrams. The essence of this approach is to obtain thermodynamic descriptions of the lower-order systems, binaries and ternaries, in terms of both the
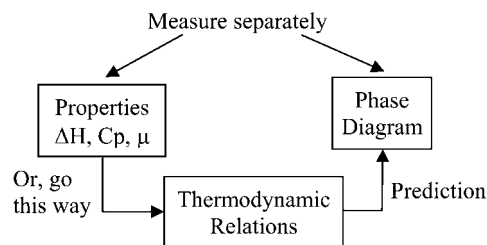
Fig. 1 Schematic diagram showing that a phase diagram can be calculated using thermodynamic relations, knowing the Gibbs energies of the phase involved. These values can be obtained from the enthalpies, specific heats, and the chemical potentials of the components obtained experimentally as well as from first-principles calculations.

444 / Material Fundamentals

experimentally measured/theoretically calculated thermodynamic data and the measured phase equilibria data. The term *thermodynamic description* denotes that the values of a set of thermodynamic parameters for all phases in a system are known. On the basis of the known descriptions of the constituent lower-order systems, a reliable description of a higher-order system can be obtained via an extrapolation method (Ref 11). This description enables one to calculate reliable phase diagrams of the higher-order system in many instances. Experimental investigation is then required for confirmatory purposes and not for the determination of the entire phase diagrams.

Figure 2 shows the methodology of multicomponent phase diagram calculation using the Calphad or phenomenological thermodynamic approach. It is worth noting that the term *thermodynamic database* or simply *database* is occasionally used in this article instead of *thermodynamic description*, particularly for multicomponent systems. Chang et al. (Ref 12) have given a summary of development in this field that needs not be repeated here. However, it should be stressed that significant progress could not have been made in this field without the availability of the first-generation phase diagram calculation software developed in the 1980s and 1990s, such as the Lukas program (Ref 13), ThermoCalc (Ref 14), ChemSage (Ref 15), MTDATA (Ref 16), and FACT (Ref 17).

With the development of the first generation of phase diagram calculation software packages, it became possible to introduce more realistic thermodynamic models for ordered phases such as the beta-brass B2 intermetallic phases in copper-zinc and nickel-aluminum and the ordered phases based on the face-centered cubic (fcc) structures in copper-gold. Both copper and gold exhibit the fcc structure and form a series of continuous solutions at high temperatures. With decreasing temperature, the disordered fcc phase transforms to ordered intermetallic phases at the compositions of $Cu_{0.25}Au_{0.75}$ and $Cu_{0.75}Au_{0.25}$ (with an ordered cubic structure designated as $L1_2$) and that of $Cu_{0.5}Au_{0.5}$ (with a tetragonal structure designated as $L1_0$). The degree of ordering in these phases increases with further decreasing temperature and becomes completely ordered at 0 K, assuming there is no kinetic barrier. For the $L1_2$ phases, three Cu(Au) atoms occupy the face-centered positions of the unit cell, and one Au(Cu) atom occupies the corner position of the unit cell when perfect order prevails. For the $L1_0$ phase, the copper atoms occupy one layer in the *c*-direction and the gold atoms the next layer.

It is obvious that the regular solution-type model, even with additional parameters used for disordered solution phases, is no longer suitable for describing these types of intermetallic phases mentioned previously, as well as many other similar intermetallic phases. Numerous thermodynamic formalisms have been formulated in the literature (Ref 18–23) for these phases. All of them are based on the Bragg-Williams or point approximation and are mathematically equivalent; it is possible to transform the model parameters in a thermodynamic formalism with more constraints to one with fewer, but not the reverse (Ref 23). Because of the importance in developing thermodynamic databases for practical applications and the need for model compatibility, nearly all the software programs adopted the compound energy formalism (Ref 18) first used in ThermoCalc. Unfortunately, the point approximation has its limitation, since it neglects the existence of short-range ordering in the lattice. As a result, this model is unable to describe the correct topological features of the copper-gold phase diagrams, demonstrated as early as 1938 by Shockley (Ref 24). Even with this shortcoming, tremendous advancement of the Calphad approach in phase diagram calculation has been made. In addition to this model for ordered phases, there was also a need to improve the first-generation phase diagram calculation software, since the algorithms used were based on local function minimization routines. This means that (unless appropriate initial values are provided by a user) metastable phase equilibria may be calculated instead of the stable ones.

Occurrence of the metastable equilibria is often a result of multiple minima in the Gibbs energy of a phase involved in the calculation, such as a phase with miscibility gap. Indeed, Chen et al. have found that some of the calculated phase diagrams in the literature are metastable with respect to the authors' model parameters (Ref 25), and Chang et al. (Ref 12) had presented a number of such examples. Realizing the challenge in using the first-generation software for phase diagram calculations, Chen et al. (Ref 26, 27) attempted to develop an algorithm to automatically calculate stable phase diagrams without the user's input of any initial values (Ref 25, 28, 29). Their initial ideas were subsequently extended and implemented to multicomponent systems, with Pandat being the product in 2000 (Ref 28). A new version of this software was released in 2007, including an optimizer and additional features such as superior graphical capability and significantly improved computational speed (Ref 30). Even though Pandat automatically calculates the stable phase diagram with a given thermodynamic database without initial values, the user can always calculate a variety of metastable phase diagrams at the user's choice by suspending some stable phases. All calculations presented in this article have been carried out using Pandat. It is noteworthy to point out that many of the first-generation software have since attempted to calculate the stable diagrams.

An Algorithm to Calculate Stable Phase Equilibria

This section first illustrates that the common tangent approach presented in nearly all the thermodynamic textbooks to calculate phase equilibria is only a necessary but not sufficient condition to obtain the stable equilibrium. The stable equilibrium of a mixture of phases corresponds to the condition that the Gibbs energy of this mixture attains the minimum. The common tangent approach could, in some cases, yield a metastable equilibrium. Figure 3 shows the Gibbs energies of the fcc and hexagonal close-packed (hcp) phases at constant temperature (*T*) and pressure (*p*), with the former exhibiting double minima and the later one minimum. It is evidently clear from this figure that there exist three common tangents between these two

```
┌─────────────────────────────────────────────┐
│  Measured thermodynamic data and "first-     │
│  principles" calculated data at 0 K and      │
│  phase equilibrium data                      │
└─────────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────────┐
│  Select thermodynamic model for each phase   │
└─────────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────────┐
│  Optimize model parameters                   │
└─────────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────────┐
│  Obtain Gibbs energy for each phase in a     │
│  binary system (thermodynamic description)   │
└─────────────────────────────────────────────┘
          ↓                         ↓
┌──────────────────────┐  ┌──────────────────────────────┐
│ Reproduce            │  │ More importantly, extrapolate │
│ experimental         │  │ the Gibbs energies of lower-  │
│ diagrams of the      │  │ order phases to those of      │
│ binaries and         │  │ higher-order phases (predict  │
│ ternaries (describe  │  │ unknown higher-order          │
│ known diagrams)      │  │ multicomponent diagrams)      │
└──────────────────────┘  └──────────────────────────────┘
```

Fig. 2 The Calphad or phenomenological thermodynamic approach to obtain a thermodynamic description or database of a multicomponent system

www.iran‑mavad.com

مرجع مهندسی مواد و متالورژی
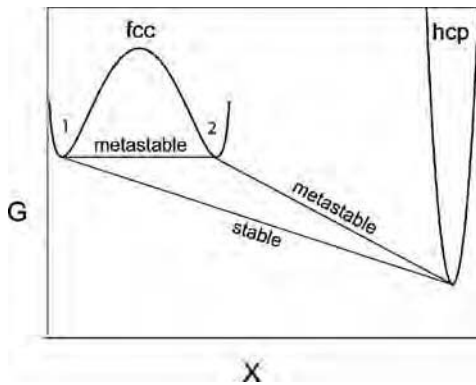
Fig. 3 Given the Gibbs energies of the face-centered cubic (fcc) and hexagonal close-packed (hcp) phases in a binary, shown in this figure at constant temperature and pressure, it becomes possible to calculate a metastable two-phase equilibrium, unless the software used is capable of automatically seeking the global minimum of the Gibbs energy of the stable phases.

Fig. 4 (a) Given the Gibbs energies of two solution phases, α and β, and two stoichiometric intermetallic phases, γ and δ, at constant temperature and pressure, the heavy curve at the bottom of the Gibbs energies from pure A to B is the convex hull or the lowest Gibbs energies of the phases in the binary A-B. The stable phases are α, γ, and β, with two two-phase equilibria of α + γ and γ + β. (b) The convex hull is obtained when connecting the lowest Gibbs energy values of the stoichiometric phases γ and δ as well as the apparent stoichiometric phases portioned from the solutions phase of α and β. When the number of partitions for these solution phases increases, the convex hull becomes the same as that shown in (a). (c) Illustration of the modified Graham scan algorithm to drop point 6, since the angle 5→6→7 is a nonleft turn, but point 6 is a left turn. (d) Illustration of the modified Graham scan algorithm to drop point 5, since the angle 4→5→7 is a nonleft turn. For the same reason, point 4 is dropped until 2→4→7 is a left turn.

phases, but only one of the three is the stable one. It is a mixture of fcc_1/hcp. The other two (fcc_1/fcc_2 and fcc_2/hcp) are metastable, because the Gibbs energy for either of these two mixtures is higher than that of the fcc_1/hcp mixture. A software package to calculate phase equilibria should automatically seek the stable equilibrium of the fcc_1/hcp mixture for the case shown in Fig. 3 and not depend on the user to instruct the software ahead of time that a specific fcc_1/hcp mixture is the stable equilibrium; the only demand on this software is to calculate the compositions of these phases in equilibrium with each other. The user often does not have any prior knowledge of the stable equilibrium of the phases under consideration, even in some binaries not mentioning the higher-order systems. In addition to solution phases exhibiting miscibility gaps, an ordered intermediate phase with multi-sublattices can also exhibit more than one minimum in the Gibbs functions.

Now consider using the Graham scan algorithm to automatically calculate the stable phase equilibria of a binary without a user's input of any initial values, as discussed in the work of Chen et al. (Ref 31). As shown in Fig. 4(a), the Gibbs energy of the solution α has only one minimum but that of β has two; the Gibbs energies of γ and δ exhibit singular values at specified mole fractions of B. It is clear from Fig. 4(a) that the convex hull represented by the thick line (shown at the lower part of the figure) defines the Gibbs energies of the stable phase equilibria. The stable phases are α, β, and γ, and the phase δ is unstable. The phase composition of α in equilibrium with γ can be obtained from a common tangent between α and γ, denoted as $x^\alpha(\alpha + \gamma)$. In a similar manner, the phase composition of β is obtained in equilibrium with γ, that is, $x^\beta(\beta + \gamma)$.

The challenge here is to develop a reliable and efficient algorithm to calculate the stable equilibria. Using a generalized scan algorithm to locate a two-dimensional convex hull

developed by Graham in 1972 at Bell Labs (Ref 32–35), Chen et al. (Ref 31) calculated the stable phase equilibria of binaries. The computational cost of this algorithm is $O(n\log n)$ instead of $O(n^2)$, the cost for a simple comparison method. This represents a significant decrease in the computational cost. Since this program was designed for locating a general convex hull, Chen et al. made a slight modification of the Graham scan algorithm for locating the convex hull of the Gibbs energies of phases for phase diagram calculation. The computational complexity of the modified Graham scan algorithm is $O(n)$, which provides a further decrease in the computational cost.

The modified Graham scan algorithm by Chen et al. is described here to calculate the stable binary phase equilibria, although readers are encouraged to read the paper of Chen et al. (Ref 31) and the reference cited (Ref 32–35) for

details of the original Graham scan algorithm. With the modified Graham scan algorithm, refer to Fig. 4(a) again. The Gibbs energies of the two solutions α and β are plotted as a function of x and those of the two stoichiometric intermetallic phases of γ and δ as two solid circles at specific compositions. The Gibbs energy curve for α and β is first partitioned into n intervals along the compositional axis. These $n + 1$ intersection points are denoted as open circles in Fig. 4(b). In this figure, $n = 10$ for purpose of illustration, counting the pure A and B. As shown in Fig. 4(c), only the lowest intersection points at each composition are plotted. For instance, at $x = 0.1$, there are two points, one from the Gibbs energy of α and the other from the Gibbs energy of β, with the Gibbs energy of α being the lower one. This lower point, denoted as "2" in Fig. 4(c), is so selected. In addition to the lowest intersection points from

each solution phase, the Gibbs energy of the stoichiometric phase also needs to be included for comparison. Therefore, a total of 13 Gibbs energy points are included in Fig. 4(c); pure A is chosen as "1", B as "13," and so on.

According to the modified Graham scan algorithm, the convex hull can be sorted out in the following manner. Start from the line segment 1→2 given in Fig. 4(c) and keep adding points, one by one, from "3" through "6." Each of these added points forms a left turn with two previous points, such as 1→2→3, 2→3→4, and so on until 5→6→7. Point "7" makes a nonleft turn, as shown in Fig. 4(c). Under such a condition, drop point "6" and go back to evaluate 4→5→7, shown as a dashed line in Fig. 4(d). Since this turn is also nonleft, point "5" is thus dropped. For the same reason, point "4" is also dropped. Then, examine 2→3→7 and again find a left turn, as shown in Fig. 4(d). This procedure continues until the last point, "13," is reached. The final lower convex hull vertex list is 1→2→3→7→12→13, which is the same as that shown in Fig. 4(b). On the other hand, when a software package uses a local algorithm to calculate the compositions of two phases at equilibrium using the common tangent approach, the calculated compositions depend on the initial values used. For instance, when calculating the γ/β phase equilibrium, the correct solution is obtained when a value of x^β close to 1 is chosen as the initial value. On the other hand, when the initial chosen value of x^β is small, a metastable γ/β with x^β close to 0.2 is obtained.

The present example using a modified Graham immediately leads to the convex hull of the Gibbs energies of all phases existing in a binary system. As noted earlier, one can use a simple comparison algorithm to calculate the convex hull, but the computational cost in that case would be $O(n^2)$. The computational cost of the modified Graham scan method used here is only $O(n)$. The Pandat software package was developed using a more general global optimization algorithm (Ref 25, 36). It is able to automatically calculate the stable phase diagram for a given thermodynamic description as well as calculate a variety of metastable phase diagrams at the demand of the user by deselecting the stable phases. Chang et al. (Ref 12) had presented a number of examples showing that the Pandat-calculated phase equilibria, using the published thermodynamic descriptions, differ from those calculated from the first-generation software.

In all these cases, Pandat has found the stable phase equilibrium, while others calculated the metastable one without the authors being aware of them. Since the thermodynamic descriptions were obtained by comparing the calculated phase equilibria with those determined by experiments, the consequence of comparing the mistakenly calculated metastable phase equilibria with those experimentally stable ones is the development of incorrect thermodynamic descriptions. After this was pointed out by Chen and Chang (Ref 12, 28),

many of the first-generation software packages have incorporated improvements to minimize some of the pitfalls, yet Pandat is the only software that uses the global minimization algorithm in search of the stable phase equilibrium for zero-, one-, and two-dimensional calculations. More importantly, the calculation speed of Pandat has been greatly enhanced in the past several years and is currently comparable to those using local minimization algorithm (Ref 37).

Rapid Method for Obtaining a Thermodynamic Description of a Multicomponent System

As described in Fig. 2, the development of a thermodynamic database of a multicomponent system, that is, higher than ternaries, depends on thermodynamic descriptions of the constituent lower-order systems, that is, binaries and ternaries. The descriptions of the binaries are obtained based on the first-principles calculated and experimentally measured thermodynamic values of the phases in question as well as the experimentally determined phase equilibria. The authors believe that in the foreseeable future, the phase equilibria data will always be needed, since the uncertainties of the thermodynamic data in many instances are too large to define the stable phase equilibria. Very often, obtaining a thermodynamic description of a ternary is less demanding when compared to that of a binary when thermodynamic descriptions for all three binaries are well developed.

As an example, the combined computational/experimental approach is described here to rapidly obtain a description of a ternary Mg-Al-Sr system. This system was chosen since recent interest in reducing fuel consumption has motivated global effort in developing/commercializing cost-affordable low-density magnesium alloys for heavy powertrain components at high temperatures, such as engine blocks and transmission cases (Ref 38, 39). The descriptions of this and other ternaries are needed to obtain a description of a higher-order system such as Mg-Al-Ca-Sr via extrapolation. Experience has shown that this approach works quite well in many cases (Ref 12, 40). Even though numerous extrapolation methods are available (Ref 11), the one most frequently used is that due to Muggianu et al. (Ref 41). However, the calculated multicomponent phase diagrams using the descriptions obtained by extrapolation from the lower-order systems need to be validated only with selected key experiments. The amount of experimental effort involved is insignificant when compared to pure experimentation over all compositions and temperatures, as was done in the past. For the rare cases when quaternary phases do exist in the systems, such as some of the quaternary aluminum alloys (Ref 42–44), optimization of model parameters for the quaternary phases in question is necessary, but the amount of effort involved is minimal.

The other case is when a phase extends its range of stability into the quaternary compositional space; for such cases, limited additional experimental effort is needed, such as in the Mo-Si-B-Ti system (Ref 45).

Developing a Thermodynamic Description of Mg-Al-Sr

The efficient method to obtain a description of a ternary, using Mg-Al-Sr as an example, is to:

1. Calculate a tentative phase diagram using an approximate thermodynamic description of Mg-Al-Sr obtained via extrapolation from those of the constituent binaries
2. Identify key samples for experimentation
3. Characterize each of the key samples in the as-cast and annealed states using XRD, SEM, EPMA, and TEM when necessary
4. Use these results to improve the approximate thermodynamic description
5. Have additional experimentation to validate this improved description.

For this example, the results obtained from three samples identified from a tentative phase diagram were used to develop a reliable description for the magnesium-rich Mg-Al-Sr ternary, as given subsequently.

Figure 5 is a tentative 400 °C isotherm of Mg-Al-Sr calculated from an approximate thermodynamic description by Cao et al. (Ref 46). This description was obtained from those of the three binaries: magnesium-aluminum (Ref 47), magnesium-strontium (Ref 48), and aluminum-strontium (Ref 49). Since only the phase equilibria in the magnesium-aluminum-rich region is of interest, a total of five samples with strontium concentrations less than 20 mol% were identified for investigation. The composition of sample 1 lies in the calculated three-phase equilibrium of (Mg) + γ-$Mg_{17}Al_{12}$ + Al_4Sr; that of sample 2 in the two-phase equilibrium of (Mg) + Al_2Sr, perhaps in the three-phase equilibrium of (Mg) + Al_2Sr + $Mg_{17}Sr_2$; that of sample 4 in the three-phase equilibrium of (Mg) + Al_4Sr + Al_2Sr; and those of samples 3 and 5 in the three-phase equilibrium of (Mg) + Al_2Sr + $Mg_{17}Sr_2$. These samples prepared in the as-cast and the 400 °C annealed states were subsequently characterized. The experimental results obtained were then used to obtain a thermodynamic description of Mg-Al-Sr, as given subsequently. Since a detailed account of the experimental studies has been published (Ref 46), only a brief summary is given. Alloys were prepared with raw materials of magnesium ingot (99.9%), aluminum lump (99.5%), and strontium ingot (99%) by induction melting in a mild steel crucible with a flowing gas mixture of 99% Ar and 1% SF_6. The solidified samples were then remelted with stirring and next cast into a BN-coated steel bowl. Each of the cast samples was cut into two pieces; one piece was characterized by SEM, EPMA, and XRD

Table 1 **Electron probe microanalysis-measured compositions of phases in the equilibrated samples at 673 K (400 °C) and the primary phase of solidification for these samples**

Alloy No.	mol%	Heat treatment	Phases	Mg, mol%	Al, mol%	Sr, mol%	Primary phase of solidification
1	Mg-3.4Sr-32.5Al	673 K, 770 h	Al$_4$Sr	6.08	72.85	18.78	Al$_4$Sr
			γ-Mg$_{17}$Al$_{12}$	57.13	37.28	3.91	
			(Mg)	92.60	6.36	0.01	
2	Mg-8.5Sr-16.5Al	673 K, 835 h	Al$_4$Sr	13.42	63.93	18.59	Mg$_{17}$Sr$_2$
			Mg$_{17}$Sr$_2$	75.13	13.75	9.26	
			(Mg)	98.18	0.57	0.38	
3	Mg-16Sr-26Al	673 K, 770 h	Al$_4$Sr	15.60	63.46	19.12	Al$_4$Sr
			Mg$_{38}$Sr$_9$	61.56	20.43	15.87	
			Mg$_{17}$Sr$_2$	76.53	11.15	9.71	

Note: The sum of the measured compositions in each of the single phases does not add up to 100%. The difference is the oxygen contents, also measured experimentally. The measured compositions of (Mg) had large uncertainties, since they were obtained from energy-dispersive spectroscopy. Using the measured contents of the metals, the ranges of homogeneity of the (Mg) phase and of the intermetallic phases are as shown in Fig. 6.
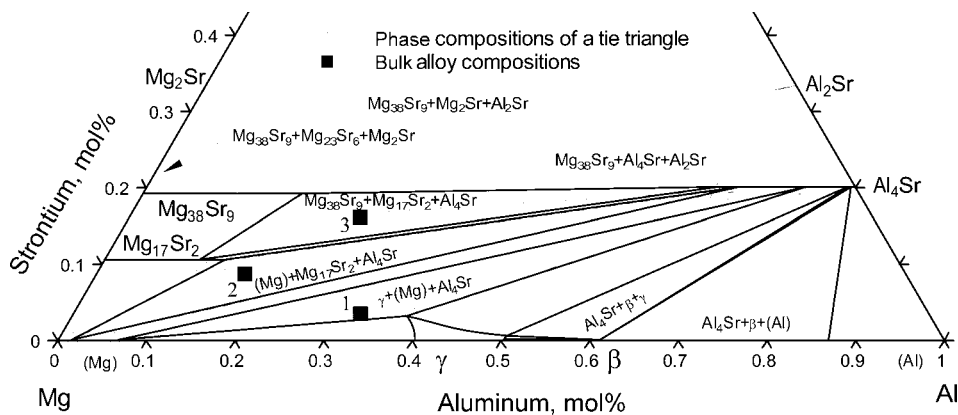
Fig. 6 Comparisons of the calculated isotherm of Mg-Al-Sr at 400 °C (673 K) from the thermodynamic description, obtained in terms of the results of alloy samples 1, 2, and 3, with the experimental data

for the primary phase of solidification and the phases formed, including the compositions and crystal structures of these phases. The other half was annealed at 400 °C for 770/835 h in an ultrahigh-purity argon atmosphere (99.998% Ar). Each of the annealed samples was also characterized using SEM, EPMA, and XRD. Since samples 4 and 5 were overly oxidized,

the results obtained for these samples were not used in developing a thermodynamic description of this ternary.

The thermodynamic models used for the phases in Mg-Al-Sr are now presented. There are three types of solution phases: the liquid phase, the solid metallic phase, and the pseudo-binary phases of stoichiometric binary

intermetallics such as (Mg,Al)$_{17}$Sr$_2$ and one intermetallic phase, γ-Mg$_{17}$Al$_{12}$, with its homogeneity extending to the ternary compositional space. The Gibbs energy of a phase normally consists of three terms: the Gibbs energy of the components in the reference state, the ideal Gibbs energy of mixing, and the excess Gibbs energy. For the liquid and the solid metallic phases, the excess term is represented by the following equation:

$$^{ex}G_m^{\varphi} = X_{Mg}X_{Al}L_{Mg,Al}^{\varphi} + X_{Mg}X_{Sr}L_{Mg,Sr}^{\varphi} \\ + X_{Al}X_{Sr}L_{Al,Sr}^{\varphi} + X_{Mg}X_{Al}X_{Sr}L_{Mg,Al,Sr}^{\varphi}$$

(Eq 1)

The term G is the Gibbs energy, with the superscript "ex" indicting that this G is an excess quantity; x_i is the mole fraction of the component i; and L_s^{φ} is the interaction parameter of the binary and ternary phases, respectively. Values of the binary parameters were obtained in terms of experimental data for the phases in question, and those of the ternary parameters are likewise obtained. However, a preliminary value of the ternary parameter was set to be zero via extrapolation. These parameters were optimized using experimental data of the ternary phase equilibria.

For the second type of phases, a two-sublattice model is represented by (Mg,Al)$_{1-x}$Sr$_x$, with magnesium and aluminum occupying one sublattice and only strontium the other. A familiar example of this type of phase is a solution of NaCl and KCl, normally represented by (Na, K)Cl. It is like a pseudobinary that can be treated as an interaction between magnesium and aluminum on one sublattice in a way similar to binary (Mg,Al), while the other sublattice is occupied exclusively by strontium. The γ-(Mg$_{17}$Al$_{12}$) is described using a simpler three-sublattice model, (**Mg**,Sr)$_{10}$(**Mg**,Al)$_{24}$(**Al**, Mg)$_{24}$ (Ref 46, 47, 50). This is a simplified model, since a four-sublattice model is more consistent with its crystal structure. The element in bold denotes that the element is predominant on that particular sublattice. A detailed description of this model was given by Cao et al. (Ref 46) and is not repeated here.

The experimental results obtained for samples 1 to 3 annealed at 400 °C and the primary phases of solidification for these samples are given in Table 1. While the approximate description was obtained assuming no solubilities in any of the binary intermetallic phases, the experimental results show that this is not the case. An improved thermodynamic description was obtained in terms of the experimental data derived from these three samples. The calculated phase equilibria from this description as well as the experimental data are shown in Fig. 6.

Focus first on sample 1. The overall composition, represented by a solid square, lies in the three-phase equilibrium of (Mg) + (γ-Mg$_{17}$Al$_{12}$) + (Al$_4$Sr). The parentheses indicate that the phase in question has a range of

homogeneity, such as (Mg,Al), (γ-$Mg_{17}Al_{12}$) and (Al_4Sr). The open triangles represent the compositions of the three co-existing phases, (Mg,Al), (γ-$Mg_{17}Al_{12}$), and (Al_4Sr), obtained from EPMA. Figures 7(a) and (b) show the backscattered electron images of sample 1 in the as-cast state and that annealed at 400 °C. The annealed sample shows the existence of three phases, (Mg) + (γ-$Mg_{17}Al_{12}$) + (Al_4Sr), while the as-cast sample shows the presence of (Mg) and (Al_4Sr), with the latter being the primary phase of solidification. The calculated liquidus projection is compared later with the observed primary phases of solidification of the as-cast alloys. Since the calculated compositions of the co-existing phases agree with the experimental data for magnesium-aluminum-rich alloys, the calculated phase boundaries in the magnesium-aluminum-rich region (i.e., less than 20 mol% Sr) are shown as solid lines. On the other hand, since the thermodynamic description was obtained only for the magnesium-aluminum-rich region of Mg-Al-Sr, the calculated phase boundaries in the strontium-rich region are not expected to be correct. The

calculated boundaries are thus shown as dashed lines. In view of the appreciable solubilities observed in the magnesium-strontium and aluminum-strontium intermetallics in the magnesium-aluminum-rich region, it is likely that appreciable solubilities also exist in the strontium-rich intermetallics.

However, when focusing on the phase equilibria in the magnesium-aluminum-rich region, there is agreement between the model-calculated and experimentally measured phase equilibria within the uncertainties of the experimental data, estimated to be approximately 1 to 2 mol%. One noticeable exception is the calculated composition of ($Mg_{38}Sr_9$) in the three-phase equilibrium of ($Mg_{38}Sr_9$) + ($Mg_{17}Sr_2$) + (Al_4Sr). This is due to the use of a simplified two-sublattice model for the (Mg, Al)$_{38}Sr_9$ phase, maintaining constant strontium composition. Had a more realistic model been used, such as a three-sublattice model for γ-$Mg_{17}Al_{12}$ (or a four-sublattice model), it would have been possible to properly describe the range of homogeneity of this phase in terms of the three component elements. However,

because the magnesium concentration in most of the commercial alloys is higher than 90 mol%, this phase is unlikely to form during solidification. Therefore, it is not necessary to use a highly complicated thermodynamic model.

As shown in Fig. 8, the observed primary phases of solidification for samples 1 to 3 are in accord with the calculations, as expected, since these data were considered in obtaining a description of this ternary. The observed primary phase of solidification for sample 4 is (Al_4Sr), which was not used in the modeling, and it is also consistent with the calculation. For sample 5, first calculate the path of solidification of this sample using the Scheil model before examining the observed microstructure of sample 5 with the calculated phase equilibria. The calculated results given in Fig. 9 show the primary phase of solidification is ($Mg_{38}Sr_9$) when the temperature reaches the liquidus surface at 592.6 °C. With an additional decrease of only 0.4 °C, the second solid phase ($Mg_{17}Sr_2$) forms; the liquid now co-exists simultaneously with ($Mg_{17}Sr_2$) and ($Mg_{38}Sr_9$).
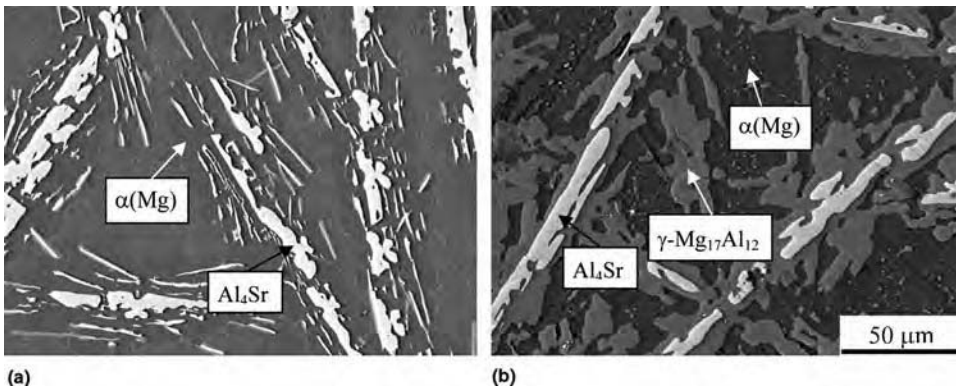
Fig. 7 Backscattered electron images of sample 1 (Mg-32.5Al-3.4Sr) in the (a) as-cast and (b) 400 °C annealed states
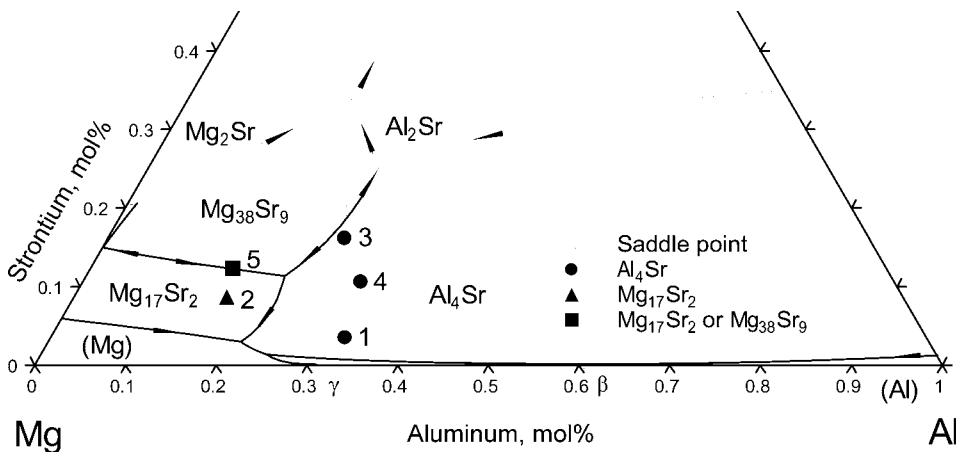
Fig. 8 Comparisons of the calculated liquidus projection of Mg-Al-Sr from the thermodynamic description, obtained in terms of the results of samples 1, 2, and 3, with the experimental data. The results of samples 4 and 5 were not used in obtaining the thermodynamic description.
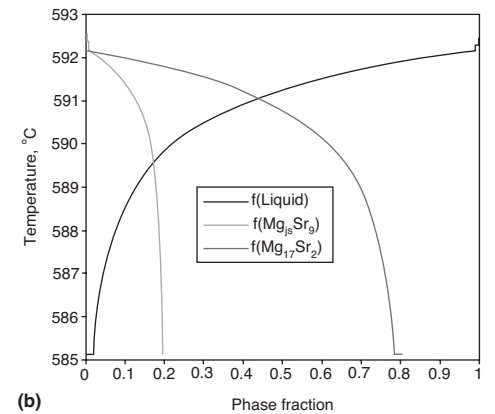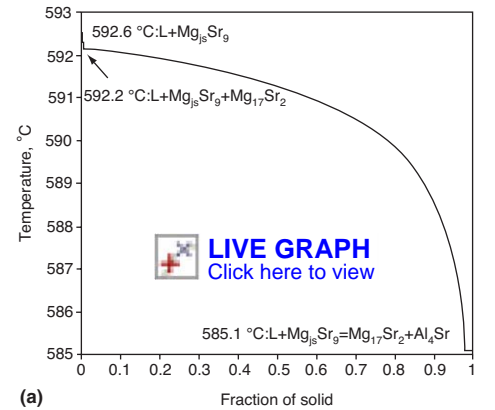
Fig. 9 (a) Calculated temperatures as a function of the fraction of solids of sample 5 (Mg-15.6Sr-12.2Al, at.% composition point in Fig. 5) using the Scheil model and the thermodynamic description obtained in terms of the results of samples 1, 2, and 3. (b) Calculated temperatures as a function of the phase fractions formed. The calculated fraction of the liquid at the invariant reaction is approximately 2%, and that of (Al_4Sr) formed from the invariant reaction is less than 0.05%.

Figure 10 shows a calculated isotherm at ~585.6 °C, 0.5 °C higher than 585.1 °C, the type II invariant temperature. When the temperature decreases to the invariant temperature, the liquid (L) co-exists with $(Mg_{38}Sr_9)$, $(Mg_{17}Sr_2)$, and (Al_4Sr), as shown schematically as an inset in Fig. 10. At this temperature, L can react with $(Mg_{38}Sr_9)$ to form $(Mg_{17}Sr_2)$ and (Al_4Sr). This four-phase equilibrium shows qualitatively that the amount of $(Mg_{17}Sr_2)$ solidified is the greatest among the three solids, that of $(Mg_{38}Sr_9)$ is next, and that of (Al_4Sr) is least. Even though $(Mg_{38}Sr_9)$ is the primary phase of solidification, its amount as shown in Fig. 10 should be less than that of $(Mg_{17}Sr_2)$, which forms subsequently. The microstructure observed, as shown in Fig. 11, is indeed consistent with this calculation, using the thermodynamic description developed coupled with the Scheil model.

One may raise the question why the existence of (Al_4Sr) is not observed in this microstructure. The four-phase equilibrium in the inset of Fig. 10 provides the answer. The amount of (Al_4Sr) formed is anticipated to be small. Unless one is extremely lucky, it is unlikely that the presence of this phase will be seen in the solidified structure. Returning to Fig. 9(b), the calculated fractions of the solids formed at the invariant temperature are approximately 20% for $(Mg_{38}Sr_9)$, 80% for $(Mg_{17}Sr_2)$, and only 0.4% for (Al_4Sr). The amounts of (Al_4Sr) formed are so small that they are not plotted in this figure. The results of sample 5 further support the thermodynamic description obtained for the magnesium-aluminum-rich Mg-Al-Sr ternary.

Subsequent to obtaining the thermodynamic description of Mg-Al-Sr (Ref 46), Parvez et al. published a paper on the phase equilibria of Mg-Al-Sr (Ref 51). They studied a total of 22 alloys prepared by induction melting under an atmosphere of argon with 1% SF_6, followed by slow cooling. These slowly cooled samples were characterized by XRD, metallography, and differential scanning calorimetry (DSC), with a slow heating and cooling rate (2 to 5 °C/min). They reported the existence of several new ternary phases in this system. However, Cao et al. concluded that these phases are not new phases but solid solutions of the binary intermetallic phases. Their alloys were not annealed at a specified temperature but cooled slowly to ambient temperature; the microstructure taken from different locations in the as-cast and post-DSC samples were observed to be similar.

It is likely that the phases obtained by Cao et al. may not differ significantly from those at 400 °C. Indeed, as shown in Fig. 12(a), their experimental results are not inconsistent with the authors' calculated isotherm at 400 °C. The symbols ■, ▲, •, and so on represent the gross compositions of the alloy samples. For example, samples 10 and 11 (solid triangles with one corner pointed upward) were reported by Parvez et al. (Ref 51) to be in the (Mg) + Al_4Sr + τ_2 three-phase field, likely due to their misidentification of the XRD patterns, as a result of the existence of $(Mg_{0.2}Al_{0.8})_4Sr$ instead of Al_4Sr. As shown in Fig. 12(b), the primary phases of solidification for 19 out of their 22 alloy samples are consistent with the calculated liquidus projection. (The primary

phase information for samples 12 to 14 was not clearly given by Parvez et al., Ref 51).

Thus, it can be concluded that the thermodynamic description obtained by Cao et al. (Ref 48) provides a reasonable understanding of the phase equilibria in the (Mg, Al)-rich region of the Mg-Al-Sr system. However, as noted earlier, the calculated phase equilibria using this description for alloys in the compositional vicinity of the $Mg_{38}Sr_9$ phase and for compositions higher than 20 at.% Sr must be used with caution. Additional experimental data are needed.

Obtaining a Thermodynamic Description of Mg-Al-Ca-Sr

To obtain a thermodynamic description of this quaternary, one must first have descriptions of the constituent lower-order systems. Descriptions for two of the four constituent ternaries are known; one is Mg-Al-Sr, as given previously, and the other is Mg-Al-Ca (Ref 52, 53). Thermodynamic descriptions for Mg-Ca-Sr and Al-Ca-Sr are not available due to lack of phase equilibrium data. Thermodynamic descriptions of these two ternaries were obtained via extrapolation from those of the constituent binaries (Ref 54). On the basis of the descriptions of these ternaries, a description of Mg-Al-Ca-Sr was obtained via extrapolation (Ref 54). As shown in Fig. 13, the model-calculated liquidus and solidus for the isopleth from Mg-5Al-3Ca to Mg-5Al-3Sr (in wt %) are in good accord with the measured data (Ref 55). However, there is no phase equilibrium data at lower temperatures to validate the calculated phase equilibria. Suzuki et al. found that in seven as-cast alloys, the primary phase of solidification is α(Mg), which is consistent with the calculated isopleth. When the concentration of strontium is 0 to 1.26%, the second phase formed is C36-$(Mg,Al)_2Ca$, and when the strontium concentration is 1.16 to 1.70 wt%, the second phase formed is $Mg_{17}Sr_2$. The calculated phase equilibria shown in Fig. 13 again account for the
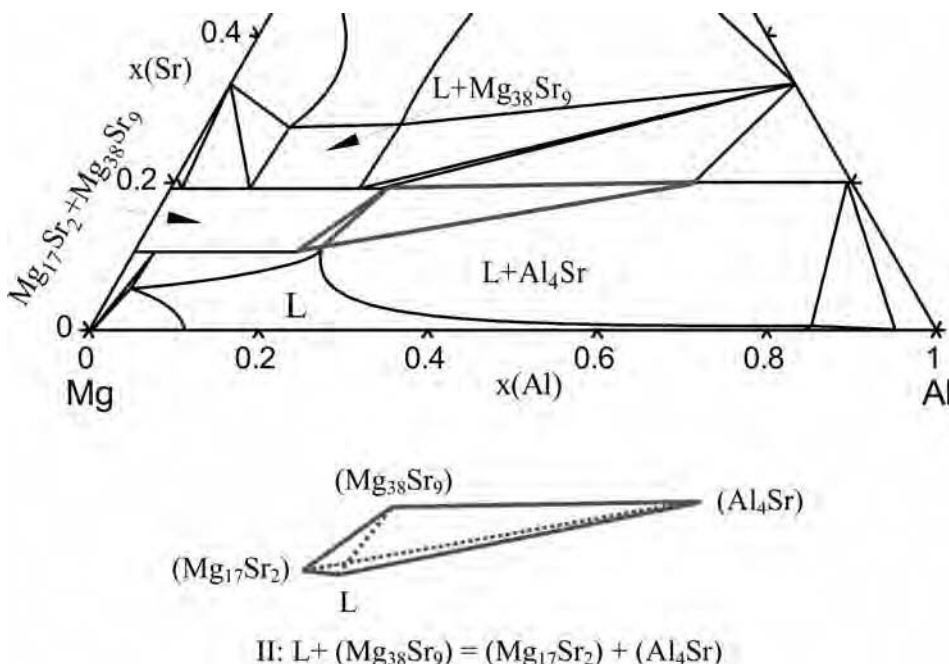
Fig. 10 Calculated isotherm of Mg-Al-Sr at 585.1 °C showing the existence of a type II_1 four-phase equilibrium of L + $(Mg_{38}Sr_9)$ = $(Mg_{17}Sr_2)$ + (Al_4Sr). The II_1 four-phase equilibrium, with some exaggeration in compositions, is shown as an inset.
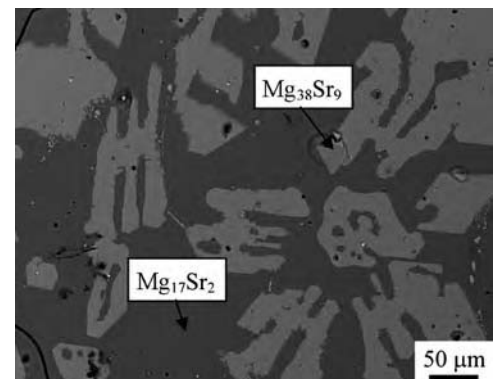
Fig. 11 Backscattered electron image of sample 5 (Mg-15.6Sr-12.2Al) in the as-cast state showing the primary phase of solidification of $(Mg_{38}Sr_9)$, with that of the next phase formed being $(Mg_{17}Sr_2)$
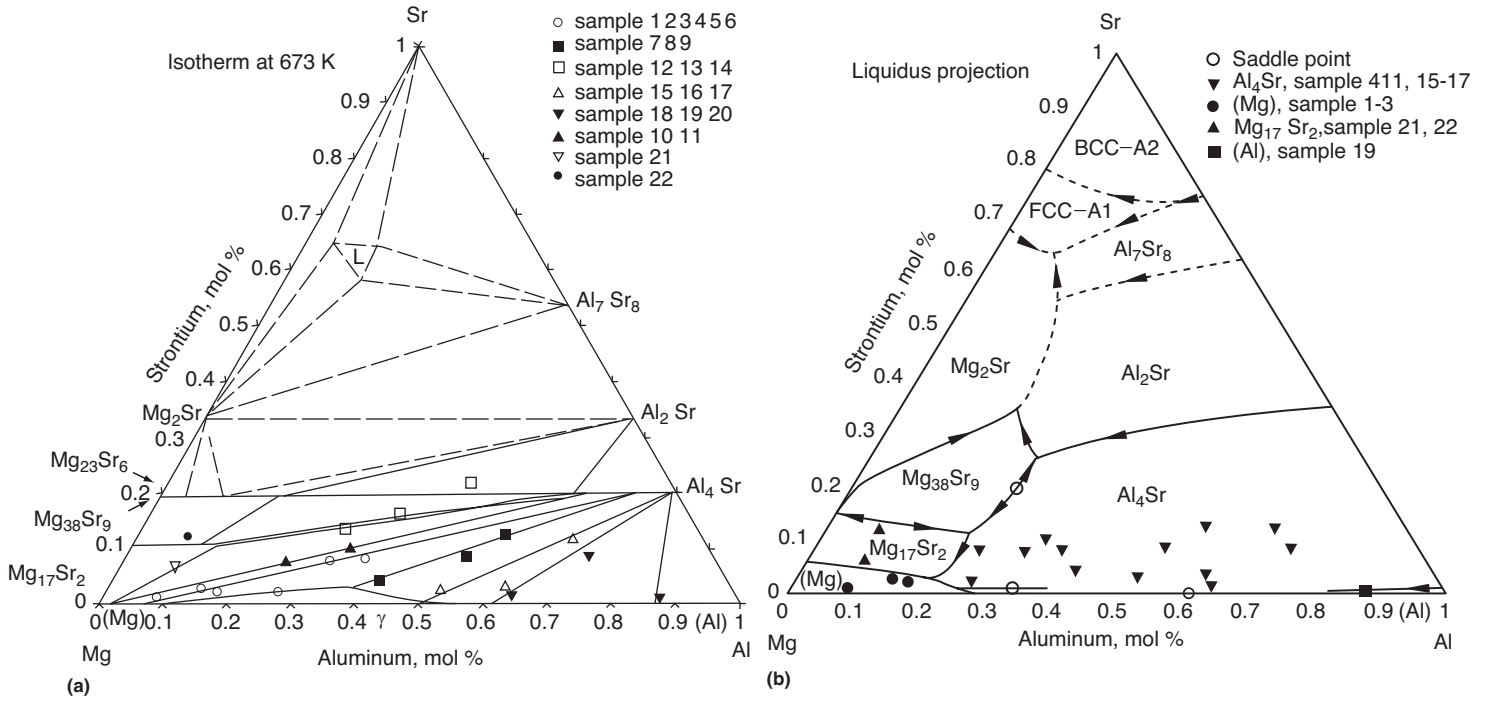
Fig. 12 Comparisons of the calculation with the data of Parvez et al. (Source: Ref 51) between (a) the 400 °C isotherm and (b) the liquidus projection
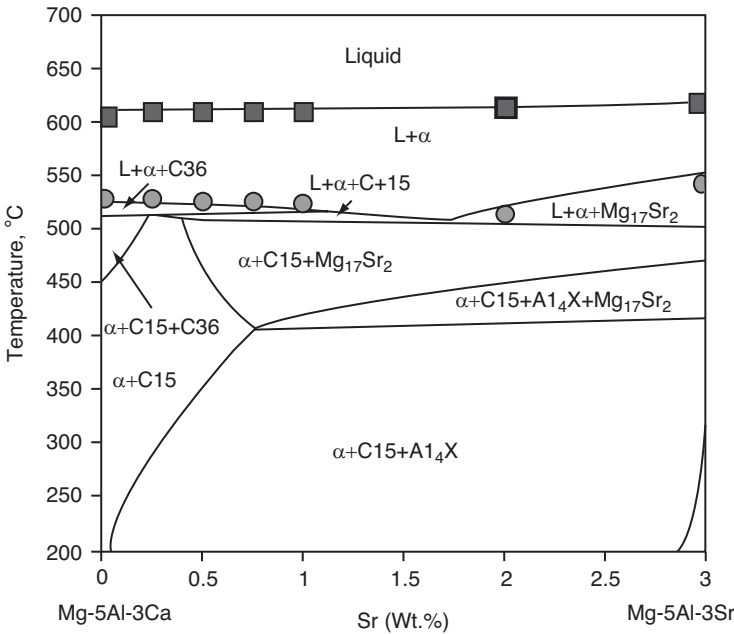
Fig. 13 Comparison between the model-calculated isopleth of Mg-Al-Ca-Sr, from Mg-5Al-3Ca (in wt%) to Mg-5Al-3Sr (in wt%), and the measured liquidus and solidus data. Source: Ref 55

LIVE GRAPH
Click here to view

microstructures observed. In short, the experimental data of Suzuki et al. support the calculated isopleth. A fair question to raise is why the quaternary description is satisfactory when only two of the constituent ternaries were modeled. A simple answer is that the quaternary alloys contain a small amount of strontium and calcium. As shown in Fig. 6, the thermodynamic description obtained is valid up to 20 at.% Sr (approximately 46 wt% Sr), and that for Mg-Al-Ca is valid to even a higher calcium concentration. It is evidently clear that when the concentrations of calcium and strontium increase to much higher concentrations, the calculated phase equilibria are unlikely to be quantitatively reliable.

Thermodynamically Calculated Phase Diagrams

Binary Phase Diagrams

This section presents thermodynamically calculated phase diagrams that do not differ from the experimentally determined ones. However, more information can be obtained from a thermodynamic database of a system in question than simply the temperature-composition phase diagrams given in classical reference books, such as Ref 56. Not only can information be obtained on the compositions of phases in equilibrium with each other as a function of temperature at constant p, but thermodynamic information is also obtainable. Unless noted otherwise, all diagrams presented are at $p = 1$ bar. For instance, the calculated binary phase diagrams in terms of temperature either as a function of composition or chemical potential of a selected element are presented at $p = 1$ bar. In the past, the phase diagrams of a ternary system were usually determined at a limited number of temperatures, often one or two. Educational estimates of the phase equilibria in between the measured temperatures often had to be made. Today (2009), the calculation can

LIVE GRAPH
Click here to view

LIVE GRAPH
Click here to view

LIVE GRAPH
Click here to view
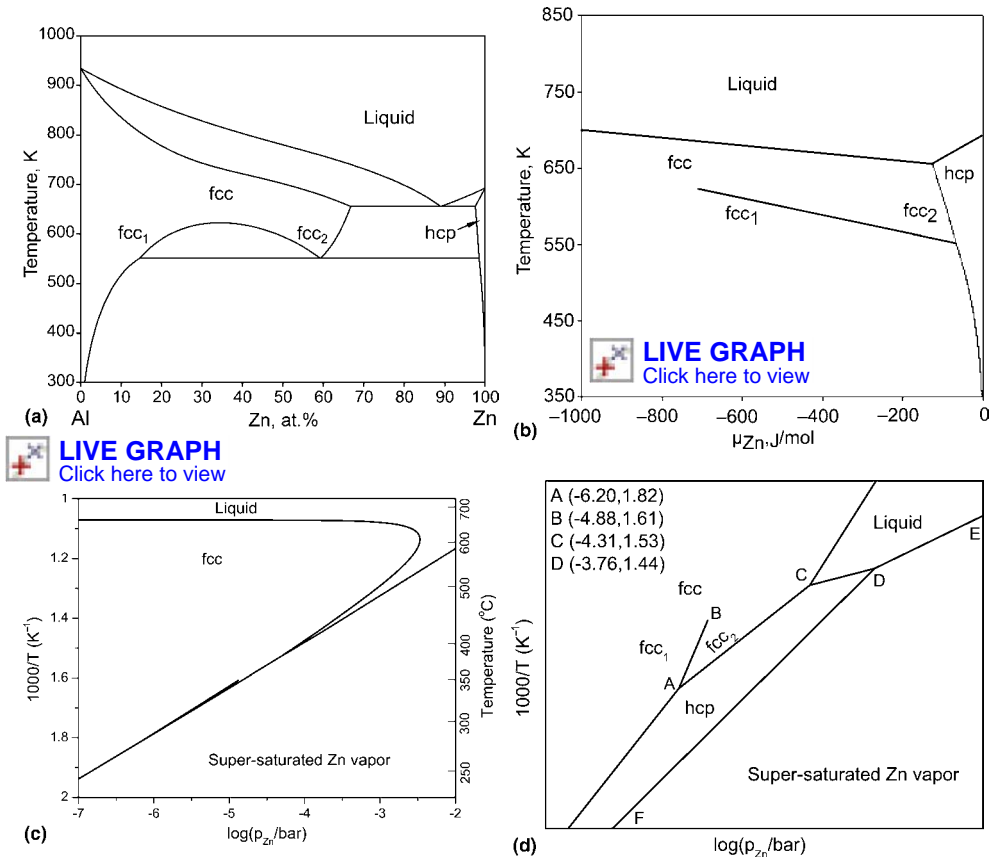
Fig. 14 Calculated phase diagram of aluminum-zinc. (a) Temperature vs. composition. (b) Temperature vs. chemical potential of zinc. (c) Reciprocal of temperature vs. $\log_{10}(p_{Zn})$. (d) Enlarged schematic diagram depicting the invariant equilibria shown in (c). fcc, face-centered cubic; hcp, hexagonal close-packed

simply be carried out using a commercial software package and the thermodynamic database of the system to obtain the phase equilibria at any temperature of interest. Examples are presented to illustrate this point, first for a binary and then a high-order alloy system. The phase diagrams presented in this article are calculated using Pandat software (Ref 31).

Figure 14(a) shows a calculated binary phase diagram of aluminum-zinc in terms of T versus x_{Zn}, using the description of Chen and Chang (Ref 57). As shown in this figure, there exist three phases in this binary, designated as L (the liquid phase), fcc (the aluminum solution phase), and hcp (the zinc solution phase). While the solubility of zinc in the fcc phase is large, that of aluminum in the hcp phase is small. Moreover, there exists a miscibility gap in the fcc phase. Figure 14(b) shows another form of the aluminum-zinc phase diagram: the relationship between the temperature and chemical potential of zinc, μ_{Zn}, with the reference state being pure zinc (hcp). This diagram provides the complementary information to that shown in Fig. 14(a). While the former diagram presents the compositions of the phases in equilibrium with each other at different temperatures, the latter shows the chemical potentials of zinc, μ_{Zn}, for the co-existing phases as a function of temperature. The former is referred to as a

field-density phase diagram and the latter as a field-field phase diagram (Ref 2). There are two invariant equilibria, that is, the eutectic, L = fcc$_2$ + hcp, and the monotectoid, fcc$_2$ = fcc$_1$ + hcp, in this binary. In addition, there is a critical point when the compositions of fcc$_1$ and fcc$_2$ merge to become fcc at this temperature, with a unique value of μ_{Zn}. This diagram is reminiscent of the p versus T diagram for a pure substance, such as water, given in nearly all the thermodynamics and phase diagram textbooks. In addition to Fig. 14(b), a similar diagram can also be plotted in terms of T as a function of μ_{Al}. The topological features would be identical. Instead of Fig. 14(b), a more practical phase diagram is a T versus p_{Zn} diagram, another field-field phase diagram. Such a diagram is given in Fig. 14(c) in terms of $1/T$ versus $\log_{10}(p_{Zn}/bar)$. Because the details involved in the invariant equilibria cannot be resolved as given in Fig. 14(c), an enlarged schematic diagram is presented in Fig. 14(d). Point "A" in this diagram represents the monotectoid, "B" the critical point for the miscibility gap of the fcc phase, and "C" the eutectic. Point "D" is the triple point of zinc, that is, the co-existence of Zn(hcp), Zn(L), and Zn(g) at a total pressure of 1 bar. At any temperature greater than that at the triple point, that is, point "D", when $\log_{10}p_{Zn}$ is higher than that represented by the

line *DE* in Fig. 14(d), pure Zn(L) may form. Conversely, at the temperature lower than that at the triple point, when $\log_{10}p_{Zn}$ is higher than that represented by line *DF*, Zn(hcp) may form. It can thus be concluded that the development of the Calphad approach to obtain a thermodynamic database makes a significant improvement in generating valuable data beyond the temperature-composition phase diagrams for practical utility.

Higher-Order Phase Diagrams

As presented earlier, visualization of binary phase diagrams at constant p is straightforward since they are all two-dimensional diagrams, such as T versus composition and T versus μ_i at constant p. This is also true for a ternary phase diagram at constant T and p. In such an isothermal phase diagram of a ternary system, only the compositions of the phases are in equilibrium with each other at a given T. For information on how the phase compositions change at different temperatures, a three-dimensional diagram is necessary. Visualization of such a diagram then becomes challenging. Nevertheless, three-dimensional representations can be analyzed when the information of the three-dimensional space is projected onto a two-dimensional plane, such as the liquidus projection of a ternary system. Such diagrams can be readily used to obtain the phase formation sequences during solidification of an alloy in a ternary with relative simple phase equilibria. On the other hand, the task becomes much more challenging when carrying out such an analysis for a quaternary, not to mention an even higher-order system. Yet, most real alloys consist of at least four component elements. In the following, an example illustrates the use of a thermodynamically calculated liquidus projection of Mg-Al-Ca-Sr in the magnesium-rich corner to visualize the solidification path of a quaternary magnesium-rich alloy.

Figure 15 shows calculated liquidus projections of three magnesium constituent ternaries—Mg-Al-Sr, Mg-Sr-Ca, and Mg-Ca-Al—of Mg-Al-Ca-Sr, with the compositions given in weight fraction, denoted as w(Mg), w(Al), w(Ca), and w(Sr), respectively. The symbols α, γ, Al$_4$Sr, Mg$_{17}$Sr$_2$, C14, C15, and C36 denote the hcp-(Mg) phase, the γ-(**Mg**,Sr)$_{0.172}$(**Mg**, Al)$_{0.414}$(**Al**,Mg)$_{0.414}$ intermetallic phase, the tetragonal-(Al$_4$Sr) phase, the hexagonal C14-(**Mg**,Al)$_{0.667}$(Ca)$_{0.333}$ Laves phase, the cubic C15-(**Mg**,Al)$_{0.895}$(Sr)$_{0.105}$ Laves phase (Ref 48), and the hexagonal C36-(Al)$_{0.500}$(**Mg**, Al)$_{0.167}$(**Ca**,Mg)$_{0.333}$ Laves phase (Ref 55), respectively. The element in bold denotes the predominant element on the sublattice in question, as noted earlier in this article. The three magnesium-rich binary eutectics are denoted as e$_1$: L = α + Mg$_{17}$Sr$_2$ (589 °C); e$_2$: L = α + C14 (517 °C); and e$_3$: L = α + γ (436 °C). The two ternary eutectics are denoted as I$_1$: L = α + C14 + C36 (Mg-Al-Ca, 516 °C) and

I_2: L = α + C14 + $Mg_{17}Sr_2$ (Mg-Ca-Sr, 508.6 °C). The three type II ternary invariants are II_1: L + $Mg_{17}Sr_2$ = α + Al_4Sr (Mg-Al-Sr, 509 °C); II_2: L + Al_4Sr = α + γ (Mg-Al-Sr, 489 °C); and II_3: L + C36 = α + γ (Mg-Al-Ca, 458 °C). The subscript "1" for the type I invariant indicates it is the highest-temperature type I invariant, and the "2" indicates the next highest and, in this case, also the lowest; the same holds true for the type II invariants, with "1" being the highest temperature and "3" the lowest. In addition to these invariant equilibria, there are two saddle points in the Mg-rich corner of Mg-Al-Ca. There are S_1: L = α + C36 (527 °C) and S_2: L= α + C14 (519 °C). As shown in Fig. 15, both saddle points are temperature maxima. In this section, the notations of Rhines (Ref 58) are adopted for the invariant equilibria, and the format used by Chang and co-workers (Ref 12, 59–62) is followed.

The lines in Fig. 15 are plotted in a three-dimensional diagram in Fig. 16(a). The z-axis (or calcium axis) starts from pure magnesium to $w_{Ca} \approx 0.24$, the x-axis (or aluminum axis) from magnesium to $w_{Al} \approx 0.7$, and the y-axis (or strontium axis) from magnesium to $w_{Sr} \approx$ 0.24. The compositions of the liquid in equilibrium with two solid phases are given in the two-dimensional planes of z-x for Mg-Al-Ca, z-y for Mg-Ca-Sr, and x-y for Mg-Al-Sr, respectively. Next, the compositions of the liquidus emanating from the ternary invariants, I_i and II_i, to the three-dimensional x-y-z space are extended. When four such monovariants intercept each other, a five-phase invariant forms in the quaternary. There are a total of four types of invariant reactions in a quaternary system, represented as I(q), II(q), III(q), and IV(q) (Ref 60). The symbol (q) indicates it is a quaternary invariant, differentiating it from a ternary invariant. In the magnesium-rich Mg-Al-Ca-Sr quaternary shown in Fig. 16(b), there is one type I quaternary invariant, $I_1(q)$, that is, a quaternary eutectic; four type II invariants, from $II_1(q)$ to $II_4(q)$; and one type III quaternary invariant, $III_1(q)$. As shown in this figure, the quaternary monovariant liquidi are shown in red to differentiate them from the ternary monovariant liquidi. Also shown in this figure are four quaternary saddle points, from $S_1(q)$ to $S_4(q)$. The calculated compositions of the co-existing phases at the invariants are given

in Table 2. In addition to the phase equilibria given in Fig. 16(b) and the invariant equilibria given in Table 2, the reaction sequences starting from the liquid to the solid phases at lower temperatures are presented in Fig. 16(c). All data given in these figures aid in better comprehending the complex heterogeneous equilibria involving the liquid and solid phases in the Mg-Al-Ca-Sr quaternary.

Next, the calculated solidification path of the commercial alloy AXJ530 (5 wt% Al, 3 wt% Ca, and 0.15 wt% Sr) is plotted in Fig. 16(d) as open blue circles. When the temperature reaches the liquidus surface, P0 (609 °C), α forms as the primary phase of solidification; L + α exists. With further decrease in temperature to P1 (526 °C), another phase, C36, solidifies, forming a three-phase equilibrium of L + α + C36. When the liquid reaches P2 (510.2 °C), a specific composition on the monovariant liquidus emanating from $S_2(q)$, C15, appears, forming a four-phase equilibrium of L + α + C36 + C15. A four-phase equilibrium of L + α + C36 + C15 now exists. Note that, subsequently, the liquid composition moves from one monovariant liquidus at P2 (510.2 °C) to another monovariant liquidus at P3 (507 °C) of L + α + C15 + C14. In between these two temperatures, the liquid is in equilibrium with α + C15. The liquid composition eventually reaches the quaternary eutectic, $I_1(q)$, that is, L + α + C14 + C15 + $Mg_{17}Sr_2$. The eutectic liquid would next solidify to form a four-phase mixture of α + C14 + C15 + $Mg_{17}Sr_2$. One may raise a question why the liquid does not follow the liquidus of the monovariant reaction to $II_1(q)$ instead of moving to the liquidus of another monovariant reaction, eventually to $I_1(q)$. The answer lies in the geometrical relationship between the compositions of the solid phases with respect to that of the liquid. The reader is referred to a paper by Yang et al. (Ref 63), which gives a clear explanation. Even though it is possible to follow the sequence of the path of solidification of such as quaternary alloy qualitatively after mastering the liquidus projection given in Fig. 16(d) and the reaction sequence given in Fig. 16(c), it is extremely challenging, if not nearly impossible, to comprehend the phases formed sequentially, not to mention the temperatures. This example demonstrates the difficulty in fully understanding the liquidus projection of a quaternary when compared with that for a ternary. The subsequent text demonstrates how a robust software such as Pandat (Ref 31), integrating phase diagram calculation with the basic Scheil solidification model for substitutional alloys, can readily calculate and plot the solidification path of an alloy such as AXJ530. The plots show not only the fractions of solids formed as a function of temperature but also the amounts of the phases formed. As shown in Fig. 17(a), the sequence of phase formation is P0, L $\rightarrow$ L + α (P0, 609 °C) $\rightarrow$ P1, L + α + C36 (P1, 526 °C) $\rightarrow$ P2, L + α + C36 + C15 (P2, 510.2 °C) $\rightarrow$ P2–P3, L + α + C15 (510.2–507 °C) $\rightarrow$ P3, L + α
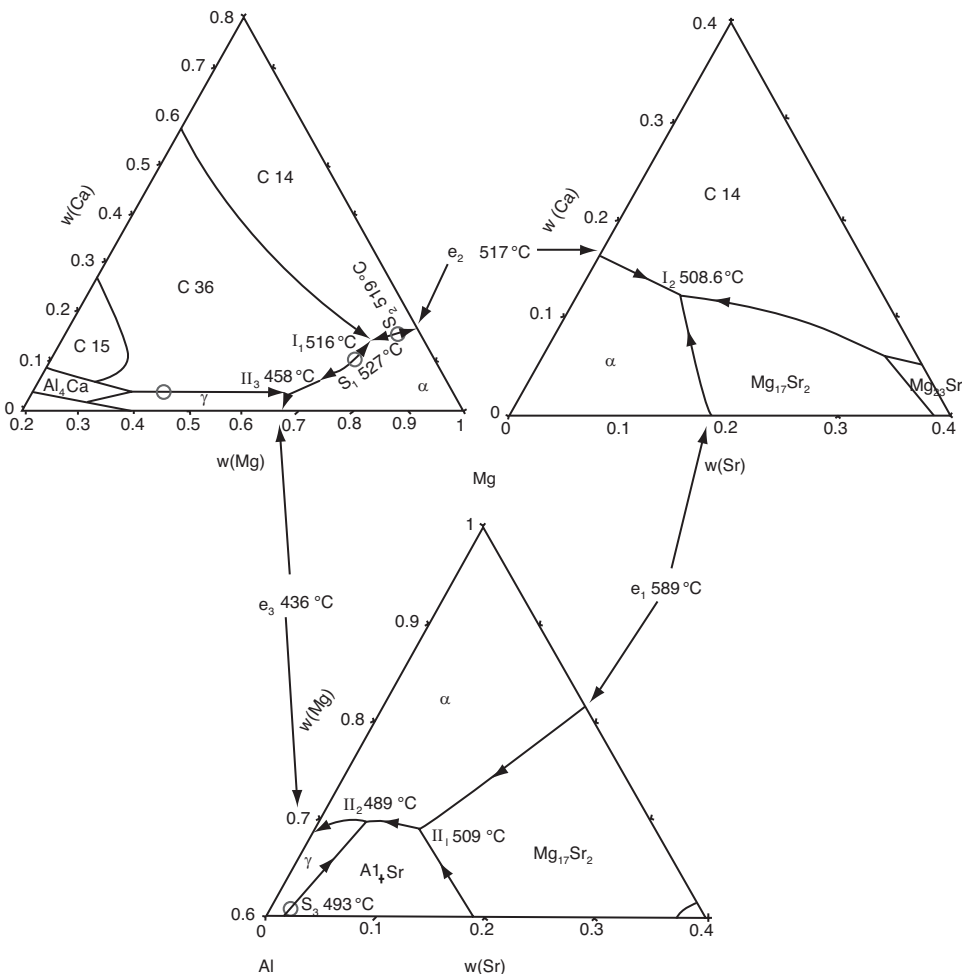
Fig. 15 Calculated liquidus projections of three constituent magnesium ternaries of Mg-Al-Ca-Sr in the magnesium-rich corner, using Pandat. Source: Ref 30
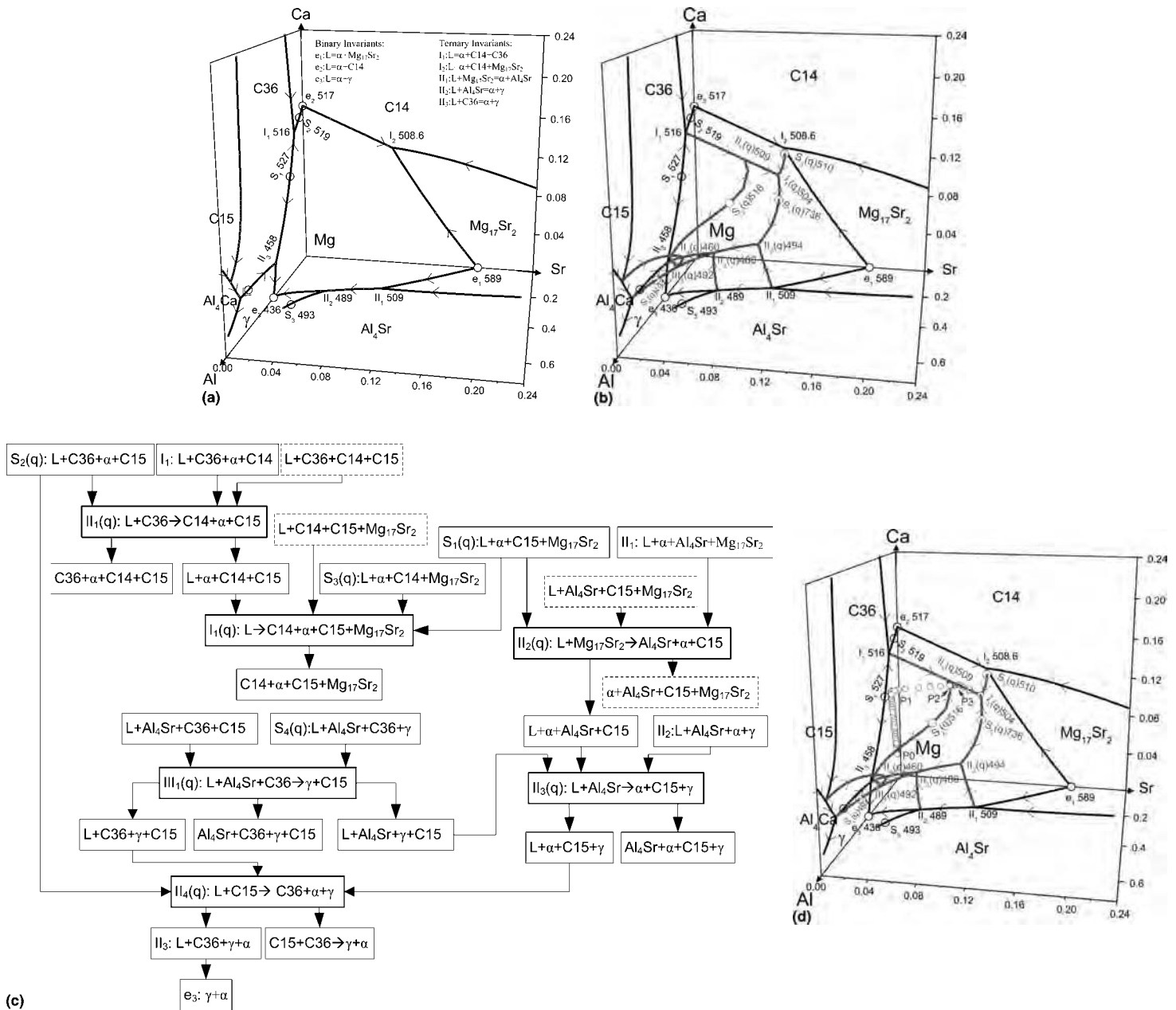
Fig. 16 (a) Calculated liquidus projections of Mg-Al-Ca-Sr in the magnesium corner showing the liquid compositions of the binary eutectics, e_1, e_2, and e_3, as well as those of the monovariant equilibria emanating from the binary eutectics to the three ternary regions, including those at the four-phase invariant equilibria. The primary phases of solidification in the respective ternaries are also shown. The compositions are given in wt%. (b) Calculated liquidus projections of Mg-Al-Ca-Sr in the magnesium corner showing the liquid compositions of the binary invariant reactions, e_1, e_2, and e_3; the ternary invariant reactions, I_1 and I_2, as well as II_1, II_2, and II_3; and those of the monovariant four-phase equilibria emanating from the ternaries to the quaternary space, including those at the five-phase invariants. All the liquid composition curves and the primary phase of solidification are presented in red. The symbols $I_1(q)$, $II_1(q)$ to $II_4(q)$, $III_1(q)$, and so on refer to the types of invariants for the quaternary Mg-Al-Ca-Sr system. The compositions are given in wt%. (c) Reaction sequences for the invariant equilibria in the liquidus projection of Mg-Al-Ca-Sr (see b). The four-phase invariant equilibria enclosed in dashed-line boxes are not shown in (b). (d) Calculated solidification path of alloy AXJ530 (Mg-5Al-3Ca-0.15Sr, wt%) according to the Scheil solidification condition shown on the liquidus surface of the magnesium-rich corner of Mg-Al-Ca-Sr. At P0, 609 °C: L + α; at P1, 526 °C: L + α + C36; at P2, 510.2 °C: L + α + C36 + C15; at P2–P3, 510.2–507 °C: L + α + C15; at P3, 507 °C: L + α + C15 + C14; at 504 °C: ($I_1(q)$) L + α + C15 + C14 + $Mg_{17}Sr_2$

+ C15 + C14 (P3, 507 °C) → $I_1(q)$, L = α + C15 + C14 + $Mg_{17}Sr_2$ ($I_1(q)$, 504 °C) → α + C15 + C14 + $Mg_{17}Sr_2$ (<504 °C). As shown in Fig. 17(a), the quaternary eutectic formed of α, C15, C14, and $Mg_{17}Sr_2$ is approximately 1%.

Next, it is shown that the results obtained from a directionally solidified alloy AXJ530 are consistent with the calculated path of solidification (Ref 54). It is clear from Fig. 17 (a) that the primary phase of solidification is the α(Mg) phase; the second phase formed is C36, followed by C15, C14, and lastly the quaternary eutectic ($I_1(q)$). The fraction of the eutectic is ~1%, as shown in this figure. The calculated temperature as a function of the phase fractions presented in Fig. 17(b) shows that the primary solidified α(Mg) phase has the largest amount, followed by C36; the amounts of the other phases are rather small, 1% or less. Figure 17(c) shows an optical micrograph of a directionally solidified alloy AXJ530 at the solid/liquid interface. This micrograph clearly shows the dendrite structure of the α(Mg) phase and the quenched-in liquid.

Table 2 Invariant equilibria in the quaternary Mg-Al-Ca-Sr system

Type of invariants	Reaction	Temperature, °C	Phase	Composition, wt%			
				Al	Ca	Mg	Sr
$I_1(q)$	L → C14 + C15 + α + $Mg_{17}Sr_2$	504	L	0.105	0.105	0.696	0.094
			C14	0.213	0.346	0.303	0.137
			C15	0.468	0.395	0.086	0.051
			α	0.013	0.002	0.985	0
			$Mg_{17}Sr_2$	0.074	0	0.63	0.296
$II_1(q)$	L + C36 → C14 + C15 + α	509	L	0.102	0.119	0.72	0.060
			C36	0.486	0.424	0.09	0
			C14	0.217	0.384	0.316	0.083
			C15	0.471	0.411	0.089	0.029
			α	0.013	0.002	0.985	0
$II_2(q)$	L + $Mg_{17}Sr_2$ → C15 + α + Al_4Sr	494	L	0.217	0.045	0.656	0.082
			$Mg_{17}Sr_2$	0.18	0	0.528	0.293
			C15	0.518	0.370	0.029	0.083
			α	0.048	0	0.952	0
			Al_4Sr	0.466	0	0.082	0.452
$II_3(q)$	L + Al_4Sr → C15 + α + γ	480	L	0.257	0.038	0.663	0.041
			Al_4Sr	0.487	0.001	0.063	0.449
			C15	0.539	0.394	0.019	0.048
			α	0.067	0	0.933	0
			γ	0.307	0.028	0.555	0.110
$II_4(q)$	L + C15 → γ + C36 + α	460	L	0.3	0.032	0.667	0.001
			C15	0.561	0.425	0.012	0.002
			γ	0.381	0.067	0.541	0.011
			C36	0.535	0.374	0.091	0
			α	0.092	0	0.908	0
$III_1(q)$	L + Al_4Sr + C36 → C15 + γ	492	L	0.372	0.037	0.573	0.017
			Al_4Sr	0.533	0.015	0.028	0.425
			C36	0.542	0.366	0.092	0
			C15	0.557	0.412	0.01	0.021
			γ	0.402	0.037	0.499	0.062

Figure 17(d) shows an SEM/backscattered electron cross-sectional micrograph of the directionally solidified alloy in the mushy zone. This micrograph consists of primary α(Mg) and the eutectic-like microstructure. The TEM/selected-area diffraction shows that the second phase is C36, which is consistent with the calculated sequence of phase formation. The micrograph in the steady-state zone in Fig. 17(e) shows the presence of α(Mg) as the primary phase of solidification, with a subsequent eutectic-like microstructure of α(Mg)/C36. The phase in the lower-right corner is a strontium-rich phase, likely to be $Mg_{17}Sr_2$. In view of the small amount of other phases according to the Scheil calculation, they were not found in the selected microstructure. It can thus be concluded that the results of the solidification experiments support the reliability of the database and the solidification simulation.

Partition Coefficients

Prior to the age of computational thermodynamics, the normal approach in calculating microsegregation in multicomponent alloys was to assume the partitioning coefficients to be constant (Ref 64, 65). A partitioning coefficient is the ratio of the compositions of the solid and liquid in equilibrium with each other at a specified temperature. While it is reasonable to make this assumption for binary alloys for back-of-envelop calculations of microsegregation during solidification, this assumption is not acceptable for multicomponent alloys.

Chang et al. (Ref 62) had demonstrated that while the partition coefficients of copper in binary Al-3.28wt%Cu alloy change slightly with temperature, those of copper in higher-order aluminum alloys, such as Al-3.28wt% Cu-5.9wt%Si, Al-3.28wt%Cu-5.9wt%Si-0.42wt%Mg, and the seven-component B-319 alloys, change appreciably with temperature. Figure 18 shows that while the partition coefficients of aluminum in Mg-5wt%Al change slightly with temperature, those of aluminum in a ternary Mg-5wt%Al-3wt%Ca alloy and a quaternary Mg-5wt%Al-3wt%Ca-0.15Sr alloy change noticeably with temperature. With the availability of phase diagram calculation software such as Pandat and thermodynamic databases such as PanAl, PanNi, PanMg, PanCu, and so on, it becomes straightforward to carry out microsegregation calculations without making this assumption. This clearly is another advancement made by computational thermodynamics.

Concluding Remarks

It is well-recognized that the Calphad approach has made significant advancements in calculating a variety of multicomponent phase diagrams for materials/processing development and/or improvements in a timely manner. This success, which depends on the reliability of the thermodynamic database and the robustness of the computer software, has been greatly enhanced with the availability of second-generation software such as Pandat.

This software was initially available in 2000 (Ref 12) to automatically calculate the stable phase diagrams given by a thermodynamic database of an alloy system in question. The current version of Pandat (Ref 30, 37) makes such complex calculations effortless, so that the materials scientists/engineers can focus their efforts on materials development instead of first becoming experts in such calculations. Nevertheless, there remains a challenge to continuously improve the thermodynamic models used for describing alloys with order/disorder phase transformation, as noted in the opening section of this article. The currently used formulism based on the Bragg-Williams or point approximation (Ref 66–68) is unable to properly account for the topological features of the fcc-based phase diagrams, due to the neglect of the existence of short-range ordering (SRO) in the lattice (Ref 12, 24, 69, 70). On the other hand, the cluster variation method (CVM) considering the existence of SRO does describe the topological features of the fcc-based phase diagrams but is computationally demanding, particularly for multicomponent alloy systems (Ref 71, 72). Oates et al. showed that the cluster/site approximation (CSA) retaining the existence of SRO is computationally efficient (Ref 72). The CSA has since been used successfully in describing the fcc-based phase diagrams in binaries, ternaries, and quaternaries, such as the technologically important nickel-base superalloys (Ref 29, 73–79). Recently Kusoffsky et al. (Ref 80) and Abe and Sundman (Ref 81) made a modification of the compound energy formalism (CEF) by incorporating additional reciprocal parameters. The modified CEF is able to account for the topological features of the calculated fcc-based phase diagrams but still does not consider the existence of SRO in the lattice, as shown in Fig. 19(b) and (c). It is evident from Fig. 19(c) that the CVM- and CSA-calculated entropy values are nearly the same, while those of the modified CEF-calculated values are considerably higher. These differences are due to the fact that SRO is accounted for by CVM and CSA but not by the modified CEF. In other words, the entropy of mixing in the modified CEF remains random. The larger entropy values obtained from the modified CEF are compensated for by the more positive enthalpy values, as shown in Fig. 19(a). Thus, the modified CEF-calculated Gibbs energy values are able to account for the topological features, but this model does not include SRO in the lattice. In view of space limitation, the details of these models are not discussed, and the calculated Gibbs energy, enthalpy, and entropy values are simply presented as a function of temperature using CVM, CSA, CEF, and modified CEF for a binary fcc solid solution at equal atomic composition. As shown in Fig. 19(b) and (c), the calculated enthalpy and entropy values using CSA are essentially the same as those calculated using CVM as a function of temperature but not the modified CEF-calculated values. The modified

P0, 609 °C:L+α

P1, 526 °C:L+α+C36
P2, 510.2 °C:L+α+C36+C15
P2-P3, 510.2-507 °C:L+α+C15
P3, 507 °C:L+α+C15+C14
I₁(q), 504 °C:L+α+C15+C14+Mg₁₇Sr₂

(a)

(b)

α(Mg)

Liquid

200μm

(c)

(d)

(e)
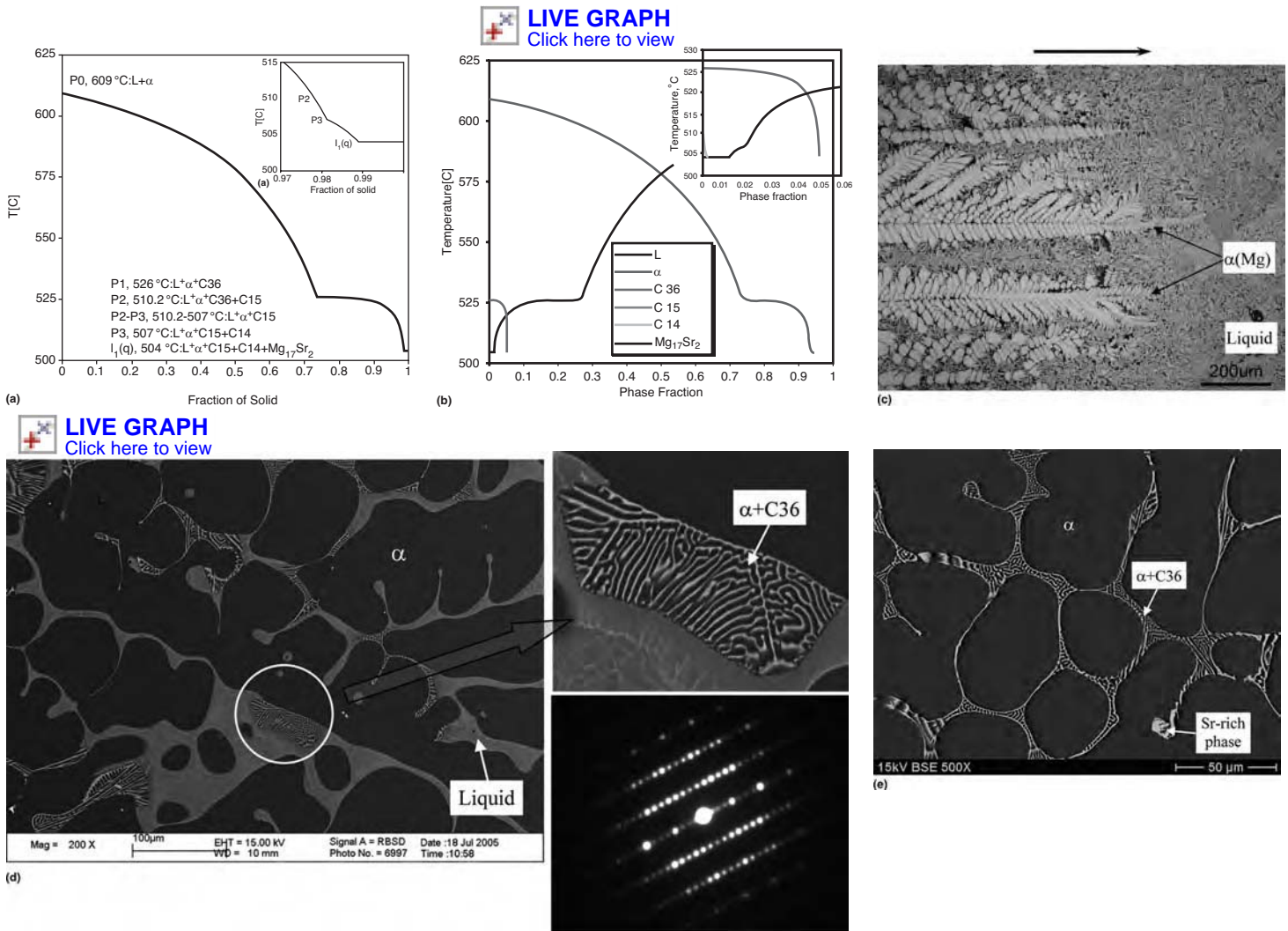
Fig. 17 (a) Calculated temperatures as a function of the fraction of solids of alloy AXJ530 (Mg-5Al-3Ca-0.15Sr, wt%), with an enlarged inset showing the same relationship at low temperatures according to the Scheil solidification condition. (b) Calculated temperatures as a function of the phase fractions formed under the Scheil solidification condition, with an enlarged inset showing the same data at low temperatures. (c) Optical micrograph of a directionally solidified sample at the solid/liquid interface showing the primary solidified α(Mg) phase exhibiting a dendritic structure and quench-in liquid. (d) Scanning electron microscope/backscattered electron (SEM/BE) micrograph of directionally solidified structure in the mushy zone showing primary α(Mg), quenched-in liquid, and the eutectic-like α(Mg) + C36 structure. The transmission electron microscope/selected-area diffraction image shows that one of the phases in the eutectic-like structure is C36. (e) SEM/BE micrograph of directionally solidified structure in the steady-state zone showing primary α(Mg), eutectic-like α(Mg) + C36 structure, and a strontium-rich phase formed in the later stage of the solidification
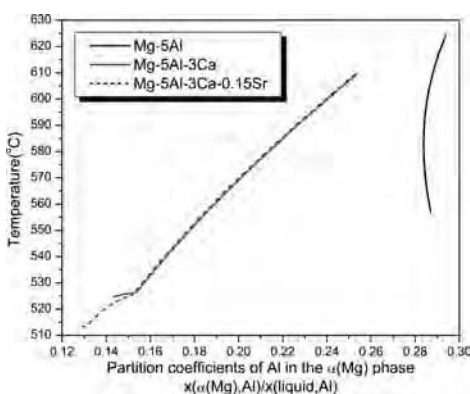
Fig. 18 Calculated partition coefficients of aluminum in the α(Mg) phase for a binary Mg-5Al alloy, a ternary Mg-5Al-3Ca alloy, and a quaternary AXJ530 (Mg-5Al-3Ca-0.15Sr) alloy. All the compositions of these alloys are given in wt%.

CEF-calculated entropy and enthalpy values are better than the CEF-calculated values but still deviate appreciably from those calculated by CVM and CSA. The readers are referred to the literature (Ref 82) for a discussion of the suitability of CSA in describing the thermodynamics of fcc-based phases. The thermodynamically calculated phase diagrams of several alloy systems presented here clearly demonstrate the advancement made in this field for practical applications. There is hope that CSA may be used in database development, particularly for the technologically important nickel-base superalloys in the not-too-distant future.

ACKNOWLEDGMENT

REFERENCES

1. O. Kubaschewski and C.B. Alcock, *Metallurgical Thermochemistry,* 5th Edition, 1979, Pergamon Press, Oxford, U.K.
2. Y.A. Chang and W.A. Oates, Materials Thermodynamics, Wiley, in press, 2009
3. J.J. Van Laar, *Z. Phys. Chem.,* 1908, 63, 216.
4. J.J. Van Laar, *Z. Phys. Chem.,* 1908, 64, 257.
5. J.L. Meijering, *Philips Res. Rep.,* 1950, 5, 333.
6. J.L. Meijering, *Philips Res. Rep.,* 1951, 6, 183.
7. J.L. Meijering and H.K. Hardy, *Acta Metall.,* 1956, 4, 249.
8. J.L. Meijering, *Acta Metall.,* 1957, 5, 257.
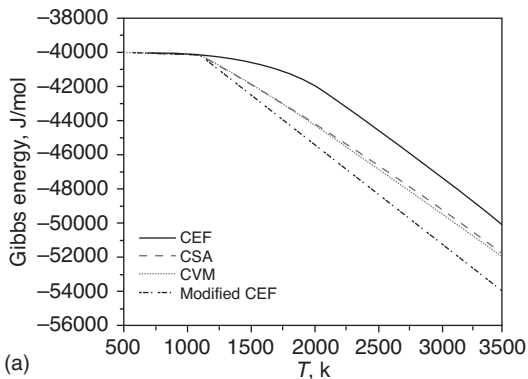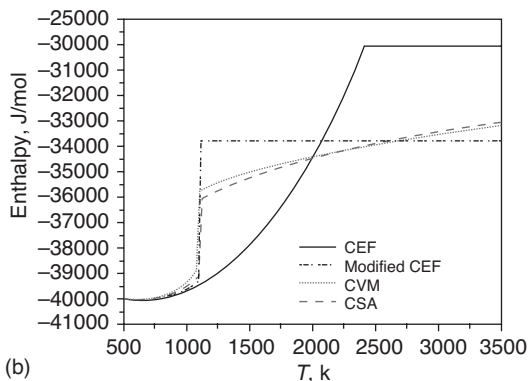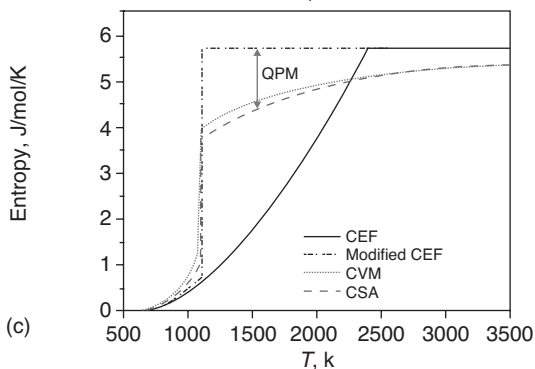
Fig. 19 Cluster variation method (CVM), cluster/site approximation (CSA), compound energy formalism (CEF), and modified CEF used as a function of temperature to calculate (a) Gibbs energies, (b) enthalpy, and (c) entropy for a face-centered cubic alloy with equal mole of the component element A and B. SRO, short-range ordering

9. J.L. Meijering, *Physical Chemistry of Metallic Solutions and Intermetallic Compounds,* Chemical Publishing Company, Inc., New York, NY, 1960, 124–142.

10. L. Kaufman and H. Bernstein, *Computer Calculation of Phase Diagrams,* Academic Press, New York, NY, 1970.

11. K.-C. Chou and Y.A. Chang, *Ber. Bunsenges. Phys. Chem.,* 1989, 93, 735.

12. Y.A. Chang, S.-L. Chen, F. Zhang, X.-Y. Yan, F.-Y. Xie, R. Schmid-Fetzer, and W.A. Oates, *Prog. Mater. Science,* 2004, 49, 313–345.

13. H.L. Lukas, J. Weiss, and E.-Th. Henig, *CALPHAD,* 1982, 6, 229.

14. B. Sundman, B. Jansson, and J.-O. Anderson, *CALPHAD,* 1985, 9, 153.

15. G. Eriksson and K. Hack, *Metall. Trans. B,* 1990, 21B, 1013.

16. A.T. Dinsdale, S.M. Hodson, T.I. Barry, and J.R. Taylor, Computations Using MTDATA of Metal-Matte-Slag-Gas Equilibria, *Proc. 27th Annual Conference of Metallurgists,* CIM, Montreal, 1988, 11, 59.

17. W.T. Thompson, G. Eriksson, C.W. Bale, and A.D. Pelton, *High Temperature Materials Chemistry,* Electrochemical Society, Inc., 1997, 16–30.

18. J.-O. Andersson, A.F. Guillermet, M. Hillert, B. Jasson, and B. Sundman, *Acta Metall.,* 1986, 34, 437.

19. H.L. Lukas, J. Weiss, U. Kattner, and E.-Th. Henig, Manual of the Computer Programs BINGSS, TERGSS, QUAGSS, BINFKT, QUAFKT, AND PMLFKT, Version 11, Jan. 1988.

20. W.A. Oates and H. Wenzl, *CALPHAD,* 1992, 16, 73.

21. W.A. Oates and H. Wenzl, *CALPHAD,* 1993, 17, 35.

22. S.-L. Chen, C.R. Kao, and Y.A. Chang, *Intermetallics,* 1995, 3, 233.

23. F. Zhang, W. Huang, and Y.A. Chang, *CALPHAD,* 1997, 21, 337.

24. W. Shockley, *J. Chem. Phys.,* 1938, 6, 130–144.

25. S.-L. Chen, S. Daniel, F. Zhang, Y.A. Chang, W.A. Oates, and R. Schmid-Fetzer, *J. Phase Equilibria,* 2001, 22, 373–378.

26. S.-L. Chen, K.-C. Chou, and Y.A. Chang, On a New Strategy of Phase Diagram Calculation, 1, Basic Principles, *CALPHAD,* 1993, 17, 237.

27. S.-L. Chen, K.-C. Chou, and Y.A. Chang, *CALPHAD,* 1993, 17, 287.

28. S.-L. Chen, S. Daniel, F. Zhang, Y.A. Chang, X.-Y. Yan, F.-Y. Xie, R. Schmid-Fetzer, and W.A. Oates, *CALPHAD,* 2002, 26, 175–188.

29. S.-L. Chen, F. Zhang, S. Daniel. F.-X. Xie, X.-Y. Yan, Y.A. Chang, R. Schmid-Fetzer, and W.A. Oates, *JOM,* 2003, 55, 48–52.

30. Pandat 7.0—Phase Diagram Calculation Software for Multiplecomponent Systems, Computherm LLC, Madison, WI, 2007.

31. S.-L. Chen, J.-Y. Zhang, X.-G. Lu, K.-C. Chou, and Y.A. Chang, *J. Phase Diagram and Diffusion,* 2006, 27, 121–125.

32. R.L. Graham, *Information Processing Letters,* 1972, 1, 132–133.

33. F.P. Preparata and M.I. Shamos, *Computational Geometry: An Introduction,* Springer-Verlag, New York, 1988.

34. J. O'Rourke, *Computational Geometry in C,* Cambridge University Press, 1993.

35. M.J. Laszlo, *Computational Geometry and Computer Graphics in C++,* Prentice Hall, NJ, 1996.

36. PANDAT Software for Multicomponent Phase Diagram Calculation, CompuTherm, LLC, Madison, WI, 2002.

37. W. Cao, S.-L. Chen, F. Zhang, K. Wu, Y. Yang, Y.A. Chang, R. Schmid-Fetzer, and W.A. Oates, PANDAT Software with PanEmgine, PanOptimizer and PanPrecipitation of Multicomponent Phase Diagram Calculation and Materials Property Simulation, *CALPHAD,* 2009, Vol 33, p 328–342.

38. R.L. Edge, Global Overview on Demand for Magnesium Alloys, *Magnesium Alloys and Their Application,* K.U. Kainer, Ed., Wiley-VCH, New York, 2000, 3–8.

39. A. Luo, Materials Comparison and Potential Applications of Magnesium in Automobiles, *Magnesium 2000,* H.I. Kaplan, J. Hryn, and B. Clow, Ed., TME, Warrendale, PA, 2000, 89–98.

40. U.R. Kattner, *JOM,* 1997, 49(12), 14–19.

41. Y.M. Muggianu, M. Gambino, and J.P. Bros, *J. Chimie Physique,* 1975, 72, 83.

42. H. Liang, Ph.D. Thesis, University of Wisconsin, Madison, WI, 1998.

43. S. Daniel, Ph.D. Thesis, University of Wisconsin, Madison, WI, 2001.

44. X.-Y. Yan, Y.A. Chang, F.-Y. Xie, S.-L. Chen, F. Zhang, and S. Daniel, *J. Alloy Compounds,* 2001, 320, 151–160.

45. Y. Yang, Y.A. Chang, L. Tan, and W. Cao, *Acta Mater.,* 2005, 53, 1711–1720.

46. H. Cao, J. Zhu, C. Zhang, K. Wu, N.D. Saddock, J.W. Jones, T.M. Pollock, R. Schmid-Fetzer, and Y.A. Chang, *Z. Metallk.,* 2006, 97, 422–428.

47. P. Liang, H.-L. Su, P. Donnadieu, M.G. Harmelin, A. Quivy, P. Ochin, G. Effenberg, H.J. Seifert, H.L. Lukas, and F. Aldinger, *Z. Metallkd.,* 1998, 89, 536.

48. H. Li and Z.-K. Liu, Poster CALPHAD, 2000.

49. C. Wang, Z. Jin, and Y. Du, *J. Alloys Comp.,* 2003, 358, 288.

50. I. Ansara, T.G. Chart, A. Fernández Guillermet, F.H. Hayes, U.R. Kattner, D.G. Pettifor, N. Saunders, and K. Zeng, *CALPHAD,* 21, 1997, 171.

51. M.A. Parvez, M. Medraj, E. Essadiqi, A. Muntasar, and G. Dénès, *J. Alloys Comp.,* 2005, 402, 170.

52. H. Cao, C. Zhang, J. Zhu, G. Cao, S. Kou, R. Schmid-Fetzer, and Y.A. Chang, *Scri. Mater.,* 2008, 58, 397–400.

53. H. Cao, C. Zhang, J. Zhu, G. Cao, S. Kou, R. Schmid-Fetzer, and Y.A. Chang, *Acta Mater.,* accepted for publication, 7/15/2007.

54. H. Cao, J. Zhu, C. Zhang, and Y.A. Chang, Thermodynamic Database of Mg-Al-Ca-Sr: A Resource for Alloy Development and Improvement, *Magnesium Technology 2008,* O. Mihriban, N.R. Pekguleryuz, R.S. Neelameggham, and E.A.N. Beals, Ed., TMS, Warrendale, PA, 2008, 101–104.

55. A. Suzuki, N.D. Saddock, L. Riester, E. Lara-Curzio, J.W. Jones, and T.M. Pollock, *Metall. Mater. Trans.,* 2007, A38, 420–427.

56. M. Hansen, Ed., *Constitution of Binary Alloys,* 2nd Ed., McGraw Hill, New York, 1958.

57. S.-L. Chen and Y.A. Chang, *CALPHAD,* 1993, 17, 113.

58. F. Rhines, *Phase Diagrams in Metallurgy, Their Development and Applications,* McGraw-Hill, New York, 1951.

59. Y.A. Chang, J.P. Neumann, A. Mikula, and D. Goldberg, INCRA Monograph VI, *Phase Diagrams and Thermodynamic Properties of Ternary Copper-Metal Systems,* The International Copper Research Association, Inc., New York, 1979.

60. Y.A. Chang, J.P. Neumann, and U.V. Choudary, INCRA Monograph VII, *Phase Diagrams and Thermodynamic Properties of Ternary Copper-Sulfur-Metal Systems,* The International Copper Research Association, Inc., New York, 1979.

61. Y.A. Chang and K.-C. Hsieh, *Phase Diagrams of Ternary Copper-Oxygen-Metal Systems,* ASM International, Metals Park, Ohio, 1989.

62. S.-L. Chen, Y. Yang, S.-W. Chen, X.-G. Lu, Y.A. Chang, Solidification Simulation Using Scheil Model in Multicomponent Systems, *J. Phase Equilib. Diffus.,* 2009, in press

63. Y. Yang and S.-L. Chen, CompuTherm, LLC, Madison, WI, 2008, unpublished results.

64. R. Mehrabian and M.C. Flemings, *Metall. Trans.,* 1970, 1, 455–464.

65. A. Roosz and H.E. Exner, *Acta Metall. Mater.,* 1990, 38, 1990, 375.

66. W.L. Bragg and E.J. Williams, *Proc. Roy. Soc.,* 1934, A145, 699–730.

67. W.L. Bragg and E.J. Williams, *Proc. Roy. Soc.,* 1934, A151, 540–566.

68. W.L. Bragg and E.J. Williams, *Proc. Roy. Soc.,* 1935, A152, 231.

69. D. de Fontaine, Configurational Thermodynamics of Solid Solutions, *Solid State Physics,* 1979, 34, 73.

70. D. de Fontaine, *Solid State Physics,* 1994, 47, 33–176.

71. W.A. Oates and H. Wenzl, *Scripta Mater.,* 1996, 35, 623.

72. W.A. Oates, F. Zhang, S.-L. Chen, and Y.A. Chang, *Phys. Rev.,* 1999, 59B, 11221–11225.

73. J. Zhang, W.A. Oates, F. Zhang, S.-L. Chen, K.-C. Chou, and Y.A. Chang, *Intermetallics,* 2001, 9, 5–8.

74. F. Zhang, W.A. Oates, S.-L. Chen, and Y.A. Chang, in *High Temperature Corrosion and Materials Chemistry III,* M. McNallan and E. Opila, Ed., The Electrochemical Society, Pennington, NJ, 2001, 12, 241–252.

75. W. Cao, Y.A. Chang, J. Zhu, S.-L. Chen, and W.A. Oates, *Acta Mater.,* 2005, 53, 331–335.

76. F. Zhang, Y. Du, W.A. Oates, S.-L. Chen, and Y.A. Chang, *Acta Mater.,* 2003, 51, 207–216.

77. W. Cao, J. Zhu, Y. Yang, F. Zhang, S.-L. Chen, W.A. Oates, and Y.A. Chang, *Acta Mater.,* 2005, 53, 4189–4197.

78. J. Zhu, W. Cao, Y. Yang, F. Zhang, S. Chen, W.A. Oates, and Y.A. Chang, *Acta Mater.,* 2007, 55, 4545–4551.

79. C. Zhang, J. Zhu, A. Bengtson, D. Morgan, F. Zhang, W.-S. Cao, and Y.A. Chang, *Acta Mater.,* accepted for publication, 01/25/2008.

80. A. Kusoffsky, N. Dupin, and B. Sundman, *CALPHAD,* 2001, 25, 549–565.

81. T. Abe and B. Sundman, *CALPHAD,* 2003, 27, 403–408.

82. Y.A. Chang, *Metall. Mater. Trans.,* 2006, 37A, 273–305; also *Metall. Mater. Trans.,* 2006, 37B, 7–39.

Internal-State Variable Modeling of Plastic Flow

H. Mecking, Material Science and Technology, T U Hamburg-Harburg, Germany
A. Beaudoin, University of Illinois at Urbana-Champaign

IF AN ABSOLUTELY PERFECT SINGLE CRYSTAL would be subjected to a tensile test, it would deform elastically up to roughly 10% strain. This corresponds to a stress level of 10% of the elastic modulus, at which point the externally applied force on the atoms would overcome their binding force, and the crystal would fracture. This theoretical limit of strength represents an upper limit, never reached in practice, since a perfect lattice is a theoretical abstraction. In reality, every crystal contains all kinds of lattice imperfections, collectively termed microstructure, that reduce the fracture stress. Even this lower stress level, however, can be established only under the supposition that the crystal does not yield first.

The yield stress is the stress level where dislocations are forced to move and to produce a plastic shear strain by slip of lattice planes in certain lattice directions. The yield strength of a material depends on microstructure and can be varied greatly by thermomechanical processing. There is, however, a lower limit of the yield strength due to the intrinsic lattice resistance to dislocation motion, the so-called Peierls force (Ref 1, 2). It depends on crystal structure and can be quite high, in particular when covalent bonding prevails, as in the case of many ceramics and semiconductors. These materials suffer from inherent brittleness, because the Peierls stress exceeds the fracture stress, and stand in contrast to those materials with close-packed structures, where metallic bonding dominates.

In pure metals, the Peierls stress can be extremely low, down to more than 6 orders of magnitude lower than the shear modulus (Ref 3). In reality, however, additional obstacles to dislocation motion, such as vacancies, interstitials, impurity atoms, and other dislocations, are always present. These obstacles must be passed or be cut by the traveling dislocations. Therefore, even in pure face-centered cubic (fcc) and hexagonal crystals (of approximately 99.99% purity), the yield stress (critical resolved shear stress) is generally much higher

than the Peierls stress; in well-annealed crystals, the room-temperature yield stress is typically found between 0.001 and 0.01% of the shear modulus.

If plastic flow of a crystal is by dislocation motion on only one crystallographic slip system, this mode of deformation is called single slip. It typically occurs when one slip system is much weaker than all others; in some hexagonal metals, for example, cadmium, zinc, or magnesium, slip on the basal plane is much easier than on prismatic and pyramidal planes (Ref 4). In single slip, deformation is carried by only one kind of dislocation, which travels over large distances. By way of continuous reproduction of the moving dislocations, avalanches of dislocations of the same species are created, which form narrow (a few nanometers wide) slip lines at the sample surface along the trace of a slip plane. An individual slip line is active only for a short time and must be replaced by activation of a new one at a different site (Ref 5). Under these conditions, the stress required for deformation to continue rises very slowly with strain; hence, this stage of deformation is called easy glide. Under suitable circumstances, the whole stress-strain curve of a hexagonal crystal consists of the easy glide region.

This is not so in fcc crystals. The fcc lattice offers 12 crystallographically identical slip systems of type $\{111\}\langle110\rangle$ (24 if positive and negative slip directions are counted separately). Generally, deformation starts with easy glide and a corresponding low hardening rate by activation of the so-called primary system, where the applied stress has the highest resolved shear stress compared to all other systems. This regime of easy glide is deemed stage I. As seen in Fig. 1, it persists only for a small range of strain, until the hardening rate increases rapidly and the stress-strain curve turns upward into a steep, almost linear branch, called stage II. The slope in this stage is athermal: It depends on crystal orientation but not on temperature, except for the temperature dependence of the shear modulus. Stage III follows as the slope
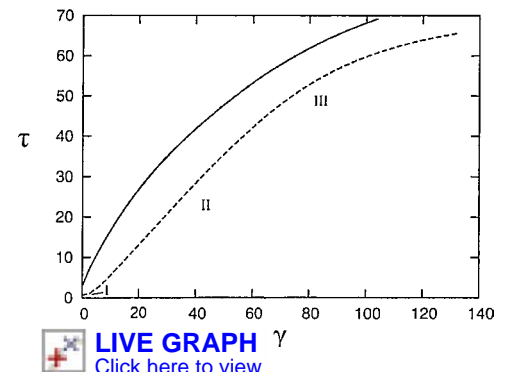
Fig. 1 Resolved shear stress, τ, versus shear strain, γ, of the primary slip system of a silver single crystal, oriented for single slip and deformed in tension at room temperature. The stages I, II, and III can be clearly distinguished, since silver exhibits a distinct stage II, due to its low stacking fault energy (Ref 6). The corresponding curve of a fine-grained polycrystal (Ref 7) exhibits stage III behavior from the beginning. Here, the normal stress, σ, and the strain, ε, are converted to shear stress and shear strain, respectively, with the help of the Taylor factor $\tau = \sigma/3$ and $\gamma = 3\varepsilon$. See the section "Basic Relationships for Work Hardening in Stage III" in this article.

of the stress-strain curve decreases gradually, tending toward a horizontal course with zero slope. Stage III is strongly affected by temperature and reduces the length of stage II continuously with increasing temperature and decreasing rate of deformation.

The discovery of the different stages has certainly been a milestone in the understanding of hardening mechanisms. It has triggered a large amount of experimental and theoretical research activities about the dislocation processes that govern the various stages of work hardening. For a comprehensive update of the actual state of the art, the reader is referred to the literature (Ref 8, 9). – see also the "Dislocations in Solids" series. In spite of still-existing controversies about prevailing mechanisms, the general view is undisputed: Easy glide will end when the applied stress, in combination

with internal stresses, triggers slip on secondary systems. The contribution to the strain by secondary dislocations is negligibly small; however, since they will entangle with primary dislocations, they have a strong catalytic effect on the storage rate of dislocations and cause the high hardening coefﬁcient in stage II. As deformation continues, dislocation arrangements that have been stored at lower strain can become unstable and partly be eliminated under action of the higher stress at larger strain, resulting in a loss of dislocations. These recovery processes are thermally activated; they superimpose on the athermal hardening processes and cause the temperature and rate dependence of stage III. It is worth emphasizing that throughout stages I, II, and III, deformation is carried by slip in the primary system, accompanied by only a small (catalytic) amount of secondary slip.

In symmetrically oriented crystals, with the sample axis parallel to directions $\langle 100 \rangle$ or $\langle 111 \rangle$, under axial loading, 8 or 6 slip systems, respectively, are equally stressed, and deformation is by polyslip. The stress-strain curves resemble those of polycrystalline material, where the grains are also enforced to deform by polyslip. Naturally, stage I is absent under these conditions, but stage II also does not emerge very clearly, as seen from the stress-strain curve of a polycrysal in Fig. 1. The curve starts after yielding, with a high slope of the same magnitude as in stage II of the single crystal (within a factor of 2 or so), but ﬂattens continuously from the beginning, just like stage III of the single-crystal curve. This seems to suggest that the whole curve consists of stage III, and stage II has degenerated into a low-strain asymptote. Considering this behavior, and instead of introducing stages, it seems to be more appropriate to refer to the different mechanisms themselves: athermal storage, thermally activated recovery, and possible large strain mechanisms of stage IV and beyond. While one of them may dominate in a certain regime, they certainly all occur simultaneously.

This short exposition is meant to explain how the various regimes of stress-strain curves can be distinguished and how they are linked to the key processes of strain hardening.

Up to now, ab initio (from the beginning) theories of work hardening exist only in rudimentary form, since the interaction between all dislocation segments in a network can be dealt with only by complex computer codes (Ref 9). There exists, however, a huge body of experimental observations on strain-hardening behavior in relation to the evolution of the dislocation structure that allows modeling, based on phenomenological connections. This article is mainly concerned with analyzing and modeling of stress-strain behavior of polycrystals of pure fcc metals in the range of temperatures and strain rates where diffusion is not important. Rules are presented for parameter identiﬁcation and for classiﬁcation of the various fcc metals according to their stacking fault energy.

The backbone of the analysis is the physical concept elaborated by Kocks and Mecking (Ref 8), which offers a complete description of stage II and III work hardening of pure fcc metals, that is, in the strain regime that is usually covered in stress-strain tests.

The whole approach is based on the assumption that dislocation accumulation is governed by only one state variable, namely, the total dislocation density. This leads to a differential equation, a so-called state function, for dislocation accumulation in which history variables such as strain and time only appear differentially. The integral along a speciﬁc path (constant temperature and strain rate) gives the stress-strain curve. The state function contains only two parameters that must be determined experimentally: The athermal storage rate is a constant, and the recovery rate varies with deformation temperature and strain rate. The values of the parameters contain the complete information about the hardening behavior of a given fcc metal at any temperature and strain rate. The physical concept can be extended to other material classes, body-centered cubic (bcc) and (in part also) hexagonal metals, and can be used for the classiﬁcation of alloying systems.

The emphasis of this article is on the phenomenological description of stress-strain behavior. It also provides information about physical background, alternative interpretations, and directions of research.

Dislocation Movement in a Field of Point Obstacles

The physical basis for the model of work hardening rests upon an understanding of the motion of dislocations through a ﬁeld of obstacles. An outline of the geometrical mechanism underlying work hardening follows, with emphasis placed on relating an evolving physical quantity, the forest dislocation density, to the ﬂow stress.

Dislocation Percolation

Yielding has been simulated by determining the conﬁguration of dislocations in a ﬁeld of point obstacles under the action of a continuously increasing stress (Ref 10).

If the obstacles are weak, the dislocations maintain a somewhat straight form and move stepwise by the unzipping of kinks. In the present context, however, strong obstacles are to be considered, and this case is introduced in Fig. 2 for illustrating the crucial elements of motion and storage of dislocations. As the applied stress increases, the individual dislocation segments bow out between the obstacles and form expanding, isolated loops. These loops build the periphery of the blank (slipped) areas in Fig. 2 (a) and increase in size until the stress reaches a level where some of them become unstable and produce long-range slip along certain (soft) channels. Although the area swept by the dislocations is then inﬁnite, they do not move to every place in the slip plane; rather, they avoid dense (hard) regions, with loops forming the periphery of the shaded (hard) areas in Fig. 2(b).

The stress level at which the slipped area is very nearly contiguous (Fig. 2b) is then considered as the yield stress of the material. It is given by the relationship:

$$\frac{\tau}{\mu} \propto \frac{b}{l} \qquad \text{(Eq 1)}$$

where τ is the resolved shear stress, l is the average obstacle distance, μ is the shear modulus, and b is the length of Burgers vector. The proportionality constant depends on the geometrical arrangement and the proﬁle of the obstacles and, if these are overcome with the help of thermal activation, also depends on temperature and strain rate. The ﬂow stress here is determined by the situation in a few very special areas, the critical gates that just barely connect neighboring soft regions, and may be envisioned as a saddle point between hard regions.
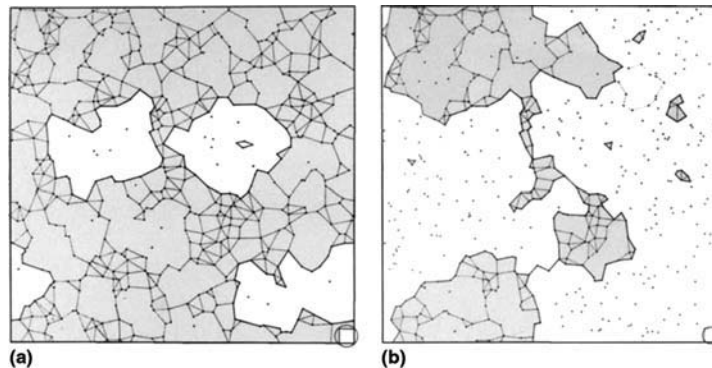
Fig. 2 Dislocation percolation in a field of strong point obstacles. 550 randomly spaced obstacles, connected by lines if they cannot be bypassed by a dislocation under the stress characterized by the circle in the lower-right corner. Source: Ref 11

Dislocation Storage

The hard regions in Fig. 2(b) are surrounded with concave loops (loops that tend to shrink under a positive stress). The fact that a part of the continually ßexing and gliding dislocation was left behind (as a kind of debris), while other parts continued, is considered here as the main hardening process, provided that the debris is not unstable under further loading.

If the incremental number of concave loops per unit volume produced is dN, then the stored dislocation length is $d\rho = L\, dN$, whereby L is the length of the average periphery of a loop. The resulting shear increment is $d\gamma = b\lambda^2\, dN$, where λ^2 is the area swept in the average circuit around each loop and is the square of their average spacing, λ. Therefore, the storage rate of dislocations becomes $d\rho/d\gamma = L/b\lambda^2$. Since L and also λ certainly scale with the obstacle spacing, l, the right side becomes $L/\lambda^2 = 1/\beta l$, with a yet unknown value of the constant β giving:

$$\frac{d\rho}{d\gamma} = \frac{1}{b\beta l} \qquad \text{(Eq 2)}$$

The geometrical mechanism of dislocation storage reviewed here has some very attractive features although a number of problems remain. It naturally leads to a heterogeneous dislocation substructure not by the decomposition of an assumed initially homogeneous structure, but by increasing any existing ßuctuations. It is just the dense areas that become denser through the storage mechanism. One dif culty is that the concave loops stored would be expected to collapse when the next dislocation traverses the slip plane and attempts to add another loop to the same area. Thus, a stabilizing mechanism must be found. It was proposed by Kocks (Ref 11) that secondary dislocations in the neighborhood (i.e. belonging to a slip system not active at this instant) can rearrange locally, and would be likely to react by entangling with the primary loop; this would serve as a potent stabilizing mechanism. Kubin (Ref 10) proposed another possibility based on observations in a dislocation simulation code: that bits of the concave loop itself engage in a local cross-slip process. Both processes have in common that non-coplanar slip systems are required, which can explain rst, why these hardening processes occur preferentially in materials that more easily activate slip systems on different planes and, secondly, how the primarily stored mobile dislocations are converted into the forest dislocations that control the ßow stress.

This leads to one of the major problems in strain-hardening models (this one as well as others): how to explain the generation of a three-dimensional dislocation network (as a consequence of two-dimensional glide). The model of concave loops, stabilized by interactions with other, non-coplanar slip systems (or by cross slip), naturally calls for the formation of a more spherical than circular affected region. Thus, natural mechanisms link neighboring storage centers in the slip plane (as slip stops on that plane) and also propagate storage centers perpendicular to the slip plane. Madec et al. (Ref 12) have shown that cross slip and short-range interactions serve to promote organization of the microstructure. Elevated local stress in dislocation walls promotes cross slip, which in turn provides stabilization through relaxing local internal stresses.

In summary, the two-dimensional structures that form as an immediate consequence of the mobile dislocations having left some parts of themselves behind are associated with an internal stress pattern that tends to propagate these structures into the third dimension. This plastic relaxation of the internal stresses is not an event that actually follows the initial storage in a time sequence but ought to occur simultaneously. Note that the relaxation processes will always be less than perfectly ef cient, so that a polarized internal stress pattern, albeit lowered, remains. The impact of relaxation is threefold: The dislocation structure is stabilized, a more three-dimensional cell structure results, and the effect of primary dislocations is converted into that of forest dislocations.

Certainly, the dislocations of locally activated secondary slip systems will entangle with the primary dislocation stored at a hard region and make the hard spots even harder, but little is known about the level of extra strength. Most likely, under continuing straining the increasing ßow stress enforces rearrangements of the local dislocation tangles, which, in the extreme, may even consist of a partial collapse of a hard spot, accompanied by a loss of stored dislocations. Instabilities of this nature are assumed to counteract the storage processes described at the beginning, with increasing frequency as the ßow stress rises. These, in the authors view, are the mechanisms underlying dynamic recovery that causes a continuous decrease of the net storage rate of dislocations as deformation proceeds. While the storage part, by its very nature, is athermal (not dependent on temperature or strain rate), the loss of dislocations by rearrangement and collapsing is enhanced by thermal activation, and so, dynamic recovery is strongly affected by temperature and strain rate.

Basic Equations for Flow Stress and Strain Hardening

Flow Stress

Although percolation of dislocations in a eld of stationary network dislocations is more complex than the simple situation sketched in Fig. 2, the yield stress can be considered as the transient point where there is a change in the topology of moving dislocations. The critical obstacle distance is then the separation of the obstacle dislocations (the trees) in the critical gates, where the soft areas merge at the yield stress. This distance is assumed to scale with the average dislocation density, ρ, so that l in Eq 1 can be replaced by the average dislocation separation $1/\sqrt{\rho}$. Indeed, one of the most surprising observations in the dislocation theory of plasticity is that the relationship between ßow stress, τ, and dislocation density, ρ:

$$\tau = \alpha \mu b \sqrt{\rho} \qquad \text{(Eq 3)}$$

holds with little regard to the arrangement of dislocations.

The ßow stress discussed here relates to dislocation interactions only. In general, there are other contributions to the yield strength, and the superposition law is not trivial (Ref 13). To simplify the present discussion of strain hardening, τ is identi ed with the dislocation part to the ßow stress only.

Because of thermal activation, α in Eq 3 depends on temperature and strain rate. It decreases with increasing temperature and decreasing strain rate for each material in a characteristic way. As dislocations bow out under stress, interactions lead to changes in the value of α with evolution of the dislocation density (Ref 14). However, this kind of variation is neglected in the following since they are small, much below the experimental and theoretical accuracy for the actual value of α. Thus, α is considered as only depending on temperature and strain rate for a given material:

$$\alpha = \alpha(\dot{\gamma}, T) \qquad \text{(Eq 4)}$$

implying that the Cottrell-Stokes relationship (Ref 15) is obeyed. According to this relationship, the relative rate and temperature sensitivity of the ßow stress, $1/m$ and $1/m^*$:

$$\frac{1}{m} = \left(\frac{d\ln\tau}{d\ln\dot{\gamma}}\right)_{\rho,T} \quad \left(\frac{d\ln\alpha}{d\ln\dot{\gamma}}\right)_T \quad \text{and}$$

$$\frac{1}{m^{\geq}} = -\left(\frac{d\ln\tau}{d\ln T}\right)_{\rho,\dot{\gamma}} \quad -\left(\frac{d\ln\alpha}{d\ln T}\right)_{\dot{\gamma}} \qquad \text{(Eq 5)}$$

are constant along a stress-strain curve. The values for m and m^* can be determined experimentally during a stress-strain test by back-and-forth changes, $\Delta\ln\dot{\gamma}$, of the strain rate (or $\Delta\ln T$ of temperature) in small strain intervals and by monitoring the magnitude of the relative stress jumps, $\Delta\tau/\tau$. According to Cottrell-Stokes, $\Delta\tau/\tau$ stays constant along a stress-strain curve for a given jump, $\Delta\ln\dot{\gamma}$, and so does the rate sensitivity, $1/m = \Delta\ln\tau/\Delta\ln\dot{\gamma}$ (and $1/m^*$ is determined from temperature changes).

As a consequence, the activation enthalpy, $Q = kTm/m^*$, is constant as well during a stress-strain test, where k is the Boltzmann factor. The usual interpretation of these interdependencies is that the thermal and the athermal parts of the ßow stress stay in a xed relation, a rule favoring forest cutting as the controlling process.

The effect of thermal activation on the ßow stress caused by dislocation interactions is not very large, in contrast to strain hardening, which

is also thermally activated and to a much more sensitive extent than the ßow stress. For the reason of simplicity, the temperature dependence of the ßow stress is ignored in the following in particular because this neglect has been found to have practically no effect on the outcome and conclusions of the analysis.

Athermal Hardening

Because the (dislocation-caused) ßow stress is determined by Eq 3, its increase with strain is due to dislocation storage. If in Eq 2 the obstacle separation l is replaced by $1/\sqrt{\rho}$, the rate of accumulation of dislocations can be written as:

$$\frac{d\rho}{d\gamma} = \frac{1}{b\beta}\sqrt{\rho} \qquad \text{(Eq 6)}$$

This dislocation accumulation rate is a fundamental quantity in the dislocation theory of strain hardening (Ref 16). It is accessible from experiment; replacing the dislocation density, ρ, by the ßow stress, τ, with the help of Eq 3, gives $\theta = d\tau/d\gamma$:

$$\frac{\theta_0}{\mu} = \frac{\alpha}{2\beta} \qquad \text{(Eq 7)}$$

There, the subscript points at the athermal nature of θ_0 (except for the negligibly small effect of rate and temperature on α). The experimental values of α depend on how the dislocation density has been measured. For copper at room temperature, the following values have been found (Ref 17): $\alpha = 1$, counting the etch pit density of dislocations piercing the slip plane; and $\alpha = 0.5$, evaluating the total dislocation length in transmission electron microscopy (TEM) thin foils. The value of β can be determined from the slope of the athermal region of stress-strain curves and turns out to be rather sensitive to the orientation of single crystals. Generally, single-slip orientations, after an initial ßat stage I, called easy glide, develop a rather steep, often linear stage II (Ref 18) with a slope of approximately $\mu/300$ not dependent on temperature corresponding to $\beta = 75$ (for $\alpha = 0.5$). Multiple-slip orientations (with straining in directions such as $\langle 100 \rangle$ or $\langle 111 \rangle$) develop no stage I and only a rudimentary stage II, which often degenerates into a temperature-insensitive initial slope of generally up to a factor of 2 steeper than the stage II of single-slip orientations. This behavior is quite similar to that observed with polycrystals, for example, the curves in Fig. 3 for polycrystalline copper obtained at various temperatures between room temperature and 400 •C for two strain rates. Immediately after yielding, the curves all start out with approximately the same slope, $\Theta_0 \leq \mu/200$, given by the dashed line, which can be interpreted as the counterpart of the work-hardening coef cient in stage II for single crystals in polyslip and corresponds to

$\Theta_0 \leq \mu/200$ (since $\Theta = M^2 \theta$, with the Taylor factor $M \leq 3$, see the text following Eq 9).

Dynamic Recovery

As was pointed out in the preceding section, dislocations stored at a certain ßow stress may, at a later stage, rearrange and partly be annihilated under the action of a higher stress. This, in the authors view, is the mechanism underlying dynamic recovery that causes a continuous decrease of the net storage rate of dislocations as deformation proceeds. Since rearrangement and collapsing is enhanced by thermal activation, dynamic recovery is strongly affected by temperature and strain rate, in contrast to the storage part, which is athermal. This effect can be accounted for in Eq 6 by subtracting a recovery term from the storage term in the form:

$$\frac{d\rho}{d\gamma} = \frac{1}{b\beta}\sqrt{\rho} - R(\dot{\gamma}, T)\rho \qquad \text{(Eq 8)}$$

The function $R(\dot{\gamma}, T)$ accounts for the effect of temperature and strain rate on dynamic recovery by thermal activation. This equation takes up the general observation that athermal storage dominates at low strains, where ρ is small, while dynamic recovery prevails as the dislocation density increases with strain. Its simple form allows determining the involved parameters easily from the evaluation of macroscopic behavior, which then can be subjected to physical interpretation.

The physical nature of recovery processes is still a controversial issue in literature. To illustrate the line of discussion some current models will be represented in the following.

In the traditional Seeger model (Ref 21), dynamic recovery is governed by massive cross slip of large groups of screw dislocations, which have been stored previously in front of extended Lomer-Cottrell barriers and are reactivated at higher stresses at the beginning of stage III in single crystals.

In the concept, elaborated in detail by Kocks and Mecking (Ref 8), it is assumed, however, that, in accordance with Eq 8, dislocation accumulation is already counteracted by recovery processes from the beginning of deformation, with continuously increasing frequency as deformation proceeds. Hardening is due to clustering of dislocations in local tangles, which are spatially arranged in a loose cell wall structure and, for example, could be envisaged as the incidental dislocation boundaries, identi ed in TEM studies by Hansen and coworkers (Ref 22), as a dominant microstructural feature in deformed material A local tangle, created at a certain level of the ßow stress, is stabilized by various types of attractive junctions of involved dislocation segments; it will become unstable, however, at some higher stress level. Thus, each tangle site can be characterized by a stress, which breaks the junctions and forces

rearrangement or even collapsing at that tangle site, corresponding with a loss of dislocations. The value of the breaking strength varies from tangle to tangle spanning a wide distribution. Hence, as the ßow stress increases in the course of straining, weak tangles are successively eliminated from the structure until, eventually, only one species survives. In this picture, both hardening and recovery processes are jerky by nature, since they are made up of discrete events that involve rather large groups of dislocations.

Argon takes an alternative approach. On the basis of earlier work with Haasen, dislocation storage is treated as a continuous process (Ref 23). The breaking of Lomer-Cottrell junctions in cell walls is introduced as the controlling process for dynamic recovery, where the dislocation ßux gradually weakens the cell walls by eliminating the strong Lomer-Cottrell junctions, in step with a continuous re nement of the cell size. Assuming an interconnection between several quantities, such as cell size, volume fraction of cell boundaries, and the partition of the total dislocation content in cell and boundary dislocations, the model arrives at a relationship for the hardening coef cient in stage III of quite the same analytical form as follows from Eq. 8.

In a recent publication by Zaiser and Seeger (Ref 24), the storage rate of dislocations has been found to have the same form as Eq 8 if, along the stress-strain curve, the loss of dislocations is estimated by statistical treatment of the annihilation probability of screw segments by cross slip. This view is supported by a computer study of discrete dislocation interactions by Madec et al. (Ref 12), where cross slip has been found to play an important role for stabilization of dislocation clusters and also for dislocation annihilation.

In spite of the many uncertainties involved in modeling the complex dislocation interactions developed during plastic ßow, Eq 8 seems to form a quite robust physical base for the overall description of dislocation accumulation. It contains only one state variable, the total dislocation density, in a simple and straightforward form that allows for assigning phenomenological expressions to modeling parameters and is used in the following for the evaluation of observed stress-strain behavior.

Quantitative Description of Strain Hardening of fcc Polycrystals

Basic Relationships for Work Hardening in Stage III

A word on nomenclature before beginning the evaluation of deformation behavior of polycrystals: Just as in single crystals, also in polycrystals, the shear ßow stress, τ, is the resolved shear stress that causes long-range slip in a crystallographic slip system at a corresponding

shear rate. Material compatibility requires every grain of a polycrystalline aggregate to perform polyslip at a resolved shear stress, τ, in at least five geometrically independent slip systems. The strain rate $\dot{\gamma}_i$ is taken as the sum of the shear rates of all systems activated in grain-number i. This sum varies from grain to grain and is $\dot{\gamma}_i = M_i \dot{\varepsilon}$, where M_i is the individual Taylor factor as dictated by the crystal orientation, and $\dot{\varepsilon}$ is the strain rate. In the Taylor model, the strain rate is assumed to be the same everywhere in the aggregate in order to guarantee material compatibility. Under the assumption that the shear ßow stress is the same in all active slip systems of all grains, the internal work is $\tau\dot{\gamma}$, with $\dot{\gamma}$ being the cumulative shear rate, and the external work is $\sigma\dot{\varepsilon}$, with σ being the applied normal stress in a tension or compression test. Equality of internal and external work rate ($\tau\dot{\gamma} = \sigma\dot{\varepsilon}$) demands:

$$\dot{\gamma} = M\dot{\varepsilon} \quad \text{and} \quad \sigma = M\tau \quad \text{and} \quad \Theta = M^2\theta \quad \text{(Eq 9)}$$

where M is the average Taylor factor, $\dot{\varepsilon}$ is the strain rate (the relative elongation rate), σ is the corresponding normal stress, τ is the shear ßow stress, and $\theta = d\tau/d\gamma$ and $\Theta = d\sigma/d\varepsilon$. It is in this sense that σ, $\dot{\varepsilon}$, and Θ are used in the following discussion.

If in Eq 8 the dislocation density is replaced by the ßow stress, σ, according to Eq. 6 and 9, the evolution of the dislocation content can be written in terms of the ßow stress:

$$\frac{d(\sigma/\mu)}{d\varepsilon} = \frac{M^2\alpha}{2\beta} - \frac{MR(\dot{\varepsilon}, T)}{2}(\sigma/\mu) \quad \text{(Eq 10)}$$

This equation provides the basis of the following discussion, whereby stress and work-hardening rate are always normalized by the shear modulus. In this way, temperature dependences caused by dynamic recovery, expressed by the function R, are separated from the intrinsic temperature dependence of the shear modulus (to which σ and Θ are always proportional). Equation 10 is often written in the abbreviated form:

$$\Theta = \Theta_0\left(1 - \frac{\sigma}{\sigma_v}\right) \quad \text{(Eq 11a)}$$

Here, $\Theta_0/\mu = M^2\alpha/2\beta \leq \frac{1}{20}$ (with $M \leq 3$) and $\sigma_v/\mu = 2\Theta_0/MR(\dot{\varepsilon}, T)$. Integration of Eq 11 (a) gives an equation where the saturation level of the ßow stress, σ_v, is approached exponentially with strain:

$$\frac{\sigma - \sigma_0}{\sigma_v - \sigma_0} = 1 - \exp -\frac{\Theta_0}{\sigma_v}\varepsilon \quad \text{(Eq 11b)}$$

The previous relationship was proposed by Voce in 1948 as an empirical description of stress-strain curves (Ref 25). The integration constant, σ_0, is the initial ßow stress at zero strain, a measure of the initial dislocation density in pure fcc material; intrinsic nondislocation contributions are discussed later in connection with practical applications. The differential form was introduced in 1976 by U.F.

Kocks for the superposition of dislocation storage and recovery processes (Ref 26). It has found widespread attention in literature and is also used in the present evaluation. The first question is about the extent to which this relationship describes measured stress-strain curves with sufficient accuracy and about the values of the parameters Θ_0 and σ_v at a given temperature and strain rate.

Scaling Relationships for Flow Stress and Work Hardening

Figure 3(a) displays stress-strain curves measured in compression between room temperature and 400 °C at two different strain rates between 10^{-4} and 1 s^{-1} (Ref 19, 20). In Fig. 3(b) the corresponding curves for the normalized strain-hardening coefficient are plotted: Θ/μ is given as a function of the normalized ßow stress σ/μ. The curves are not straight; rather, they exhibit tails with an upward curvature at the low- and the high-stress end. The low-stress regimes are not particularly important, since they correspond to only a very small portion of the stress-strain curve, at the beginning at approximately yielding up to strains less than 2%, and can be ascribed to the effect of grain boundaries on dislocation storage (Ref 27, 28). The tails at high stress also span a rather low-stress range, but corresponding strains are rather large.

The curves in Fig. 3(b) look like a single-parameter set, for the most part; thus, much could be gained by scaling them. This examination uses the (straightish) middle part of the curves in Fig. 3(b) which are assumed to be typical of stage III behavior. It turns out that it is possible to scale the Θ-axis by a single value, Θ_0, independent of the strain rate, and dependent on temperature only through the shear modulus, μ. This choice simplifies the derivation of a scaling factor for the stress axis. In principle, this scaling stress could be chosen as the stress level for any Θ/μ-value of this regime, and a convenient value would be just in the middle, that is, at $\Theta_0/2$ (Ref 29). An alternative procedure, which has repeatedly been employed in the literature, is to approximate each Θ/μ-σ/μ curve by a tangent to the rather straight middle part, with a fixed value for the intercept on the ordinate (or at the yield stress), at Θ_0/μ. The intercept on the σ-axis is then used as the scaling stress. Inasmuch as this straight line if it held true for the entire regime would correspond to the Voce law (Eq 11b), this particular scaling stress would be identified as σ_v. (Please note: The subscript in σ_v is not to be confused with the roman numeral for 5.)

The curves of Fig. 3(b) are replotted in Fig. 4, but now in a scaled form as Θ/μ versus σ/σ_v. It is seen that all curves fall into a very narrow band, where the width of the band is practically identical with the experimental uncertainties of an individual curve. Obviously, the set of curves for all temperatures and strain rates forms a single master curve for the dependence of the hardening rate on the ßow stress.

To re-iterate, the Voce law does not fit the stress/strain curves over the entire regime; it does give a reasonable fit over a significant range of Θ/Θ_0 and, more importantly, provides a convenient method for arriving at a scaling stress.

Phenomenological approximations to this function, which differ from the Voce law, have been proposed. As an example, a particular one that has worked well in some cases is (Ref 30):

$$\frac{\Theta}{\Theta_0} = \left(1 - \frac{\sigma}{\kappa\sigma_v}\right)^\kappa \quad \text{(Eq 12)}$$

with the adjustable parameter $\kappa \approx 1.3$.

In summary, there is no doubt that the Θ-σ curves for different temperatures and strain rates form a one-parameter set of curves. Thus, the hardening behavior can be predicted from a knowledge of only two parameters of the master curve, where one is the characteristic strain-hardening value Θ_0 and the other the scaling stress, chosen here as σ_v. θ_0/μ is a true constant, while the scaling stress $\sigma_v = \sigma_v(\dot{\varepsilon}, T)$, which is also proportional to $\mu(T)$, depends in addition on
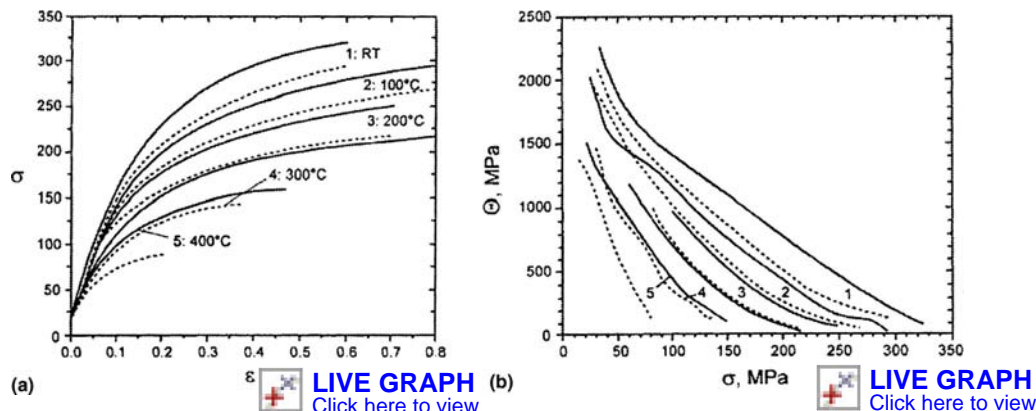
Fig. 3 (a) Stress-strain curves of copper polycrystals for a wide range of temperatures and two strain rates: 1 and 10^{-4} s^{-1}. Source: Ref 19, 20. (b) The slope $\Theta = d\sigma/d\varepsilon$ as a function of σ for the curves in (a)

strain rate and temperature through thermal activation. Details of the nature of the recovery mechanisms can be deduced from the temperature and rate dependence of σ_v, which is the focus of the next section.

Scaling of Deformation Rate and Temperature

In the 1960s and 1970s, in many investigations on single crystals, detailed evaluations of the effect of temperature and strain rate on work hardening were carried out using the so-called τ_{III} analysis introduced by Haasen, Seeger, and others (Ref 31 33). Here, τ_{III} marks the stress level at the beginning of stage III, and the analysis focuses on the effect of thermal activation on τ_{III}.

Mecking and Lucke (Ref 6, 29, 34) have shown that the formalism used in the τ_{III} analysis may be extended to the whole hardening curve, since the relative rate and temperature sensitivity of the stress for any xed value of θ/θ_{II} was found to be a constant for a given rate and temperature, that is, not to vary along a hardening curve. This rule was found to hold not only for single crystals but for polycrystals as well, and it can be written as:

$$\left. \frac{d\ln\sigma}{d\ln\dot\varepsilon} \right|_{\Theta/\mu,\,\mathbf{T}} = \frac{1}{n} \quad\text{and}\quad -\frac{d\ln\sigma}{d\ln T}\Big|_{\Theta/\mu,\dot\varepsilon} = \frac{1}{n^{\ge}} \tag{Eq 13}$$

since θ_{II}, introduced at the beginning, is equivalent to Θ_0. The fact that the derivatives n and n^* do not vary with deformation, points to a striking similarity with the so-called Cottrell-Stokes law, Eq 5, although the n- and m-values generally differ by approximately 1 order of magnitude, with $n \le m$.

In the conventional τ_{III} analysis, the rate and temperature dependence of τ_{III} is evaluated on the basis of an Arrhenius equation with a logarithmically stress-dependent activation energy, which, in terms of σ_v, can be written as:

$$\Delta G = A \ge \ln(\sigma_{v0}/\sigma_v) = -kT\ln(\dot\varepsilon/\dot\varepsilon_0) \tag{Eq 14}$$

where A is assumed to be a material constant (Ref 35), which depends on the relative stacking fault energy. If Eq 14 were true, experimental data would fall on a straight line in a diagram of $\ln\sigma_v$ versus $\Delta G = -kT\ln(\dot\varepsilon_0/\dot\varepsilon)$.

Figure 4(b) displays a set of data on copper polycrystals in this way, with $\dot\varepsilon_0 = 1\,\mathrm{s}^{-1}$, that is, just with $\dot\varepsilon$ measured in units of s^{-1}. There are two things wrong with such a plot. First, the value of $\dot\varepsilon_0$ in any physically based Arrhenius equation is never 1 s^{-1}; it is, in fact, not very well known, except for a general order of magnitude. A good way to choose it is such as to unify the experimental T and $\dot\varepsilon$ dependence, provided the order of magnitude comes out as expected. The second shortcoming, already inherent in Eq 14, is that all ßow stresses are expected to be proportional to the shear modulus, which itself depends on temperature in a nontrivial way; the same effect leads to a proportionality of ΔG to μb^3 (Ref 21, 36, 37). Finally, since it is already evident in Fig. 4(b) that the logarithmic dependence of the activation energy on stress is not exact over the whole stress range, in particular at low stresses, the lines are not straight. A more general form is employed in the following, with an unspeci ed functional dependence of the activation energy on stress:

$$\frac{\Delta G}{\mu b^3} \quad \frac{kT}{\mu b^3}\ln\frac{\dot\varepsilon_0}{\dot\varepsilon} = g\left(\frac{\sigma_v/\mu}{\sigma_{v0}/\mu_0}\right) \tag{Eq 15}$$

If, as required by Eq 15, $\Delta G/\mu b^3$ is a sole function of σ/μ, then a value $\dot\varepsilon_0$ must exist that brings all the data of Fig. 4(b) on a unique curve in a plot of σ/μ versus $(kT/\mu b^3)\ln(\dot\varepsilon_0/\dot\varepsilon)$. The optimum value for $\dot\varepsilon_0$ that ful lls this requirement within the lowest scatter is $\dot\varepsilon_0 = 10^7\,\mathrm{s}^{-1}$, as determined by trial and error. This can be seen from Fig. 5(a), where the data of Fig. 4(b) are replotted in normalized coordinates.

A systematic search for other algebraic approximations to the data (for many materials) has led to the conclusion that the phenomenological equation for the reduced activation energy $\Delta G/\mu b^3$ g:

$$\left(\frac{\Delta G}{\mu b^3}\right)^{1/2} \quad g^{1/2} \propto 1 - \left(\frac{\sigma/\mu}{\sigma_{v0}/\mu_0}\right)^{1/2} \tag{Eq 16}$$

where σ_{v0} is the maximum Voce stress at zero temperature and ts these data best, to the extent that one aims at a straight line as a good representation of these data. The relationship that includes the proper normalizations:

$$\left(\frac{\sigma_v/\mu}{\sigma_{v0}/\mu_0}\right)^{1/2} = 1 - \left(\frac{1}{g_0}\frac{kT}{\mu b^3}\ln\frac{\dot\varepsilon_0}{\dot\varepsilon}\right)^{1/2} \tag{Eq 17}$$

thus can be used as a more accurate (although less convenient) analytical description than Eq 14 of the function $\sigma_v(T,\dot\varepsilon)$. However, neither one is expected on the basis of any rst-principles theory; only the parameter combinations embodied in the coordinates are. Since an accurate and general form for these dependencies is never likely to be found, the best method for
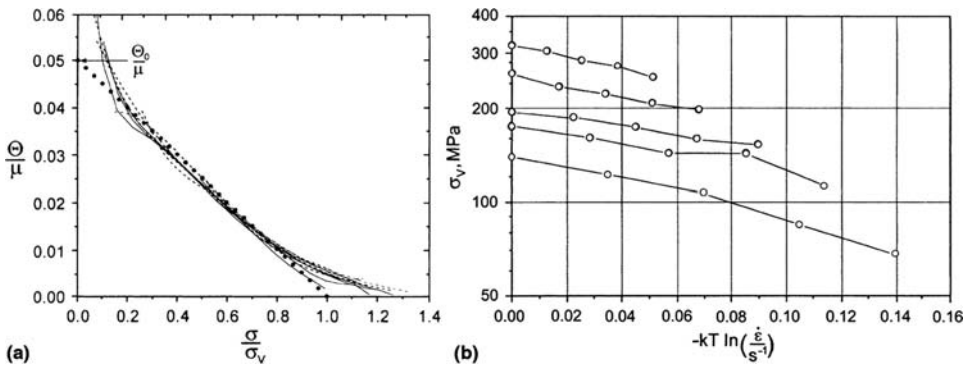
Fig. 4 (a) Θ - σ curves of Fig. 3(b) plotted in a normalized form as Θ/μ vs. σ/σ_v. The dotted line is the Voce approximation, with $\Theta_0 = \mu/20$. (b) Logarithm of σ_v versus $kT\ln(\dot\varepsilon/s^{-1})$ for five strain rates in decades from 1 down to $10^{-4}\,\mathrm{s}^{-1}$
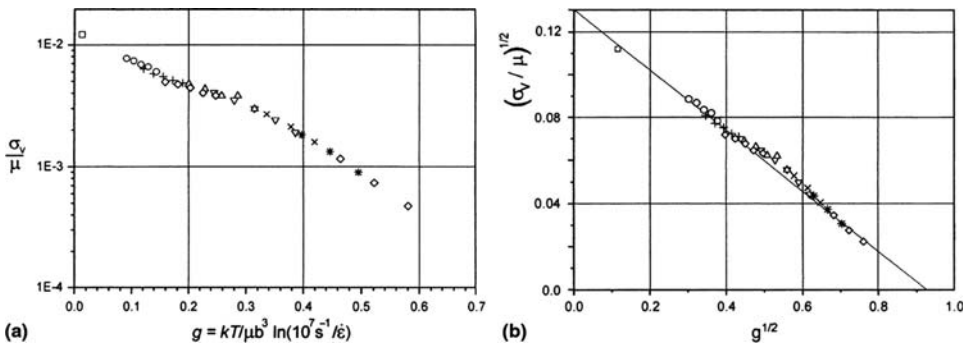
Fig. 5 (a) A plot of $\log(\sigma_v/\mu)$ vs. $kT/\mu b^3 \ln(10^7\,\mathrm{s}^{-1}/\dot\varepsilon)$, verifying the validity of Eq 15. (Compared to Fig. 4b, additional data for high temperature are included). (b) Same as (a) but with coordinates according to Eq 16, square root of σ_v/μ versus square root of g. Note: It is conceivable that the slight bump in the middle of the plots is due to dynamic strain aging in this oxygen-free electronic copper. See the section "Class 2, Dislocation Interaction" in this article.

constitutive descriptions is to employ two master curves for each material, namely one for the dependence of the hardening coef cient on the ßow stress and the other for the dependence of the scaling stress on temperature and strain rate.

The prime advantage of a well- tting equation such as Eq 17 is that the parameters in it (which depend on material) can be determined in a reproducible way, for example, by extrapolating to the axes in Fig. 5(b). One must bear in mind, however, that the extrapolated values themselves, such as g_0, depend on the particular exponents chosen for the plot and really should not be interpreted as having a physical meaning beyond the fact that they describe the measured range well. The authors would also warn against searching for a best t by computer to Eq 29, letting the three parameters ßoat (or even ve, if the two exponents ½ were to be let go, always assuming that measured values are used for $\mu(T)$, or a separate best t to modulus data). The value of $\dot{\varepsilon}_0$ can be measured to suf cient accuracy from overlapping data on the T and $\dot{\varepsilon}$ dependence. The value of g_0 is subject to the constraint that it should be the same for any particular material. Only the value of σ_{v0} is adjustable to a particular set of test samples.

Material Scaling by Stacking Fault Energy (SFE)

It has been found that appropriate scaling uni es the stress-temperature dependence of all investigated materials into a single master curve. A scheme follows where Voce stresses have been corrected for nondislocation contributions, σ_i, to the ßow stress. To this end, the σ_v data have been determined from plots of θ versus $\sigma-\sigma_i$, where the σ_i values were found

by an optimization procedure: Each Θ-σ curve was shifted along the stress axis until the tangent of the (straight) middle part intersected the abscissa at θ_0/μ, a common value for a whole set of curves for one material. The tangent then intersects the new stress axis at σ_v. Generally, the corrections were quite small within the size of the symbols of the graphs for σ_v and thus, they are of no particular relevance within the present context for the pure materials but can be very important for materials of technical purity and, to a much greater extent, for alloys where σ_i can be quite high, as is discussed in later sections.

Figure 6(a) is a replot of Fig. 5(b) for copper, where data for silver, nickel, and aluminum polycrystals are added, obtained at different temperatures for various strain rates (data for silver from Ref 7, 38, 39). Using the same value for $\dot{\varepsilon}_0$ as for copper makes the data gather on separate curves, one for each material. It has already been presupposed in Fig. 6(a) that $\dot{\varepsilon}_0$ is a universal constant for all the considered materials, noting that the database for silver, nickel, and aluminum is not suf ciently complete to prove the accuracy of this assumption. If, however, $\dot{\varepsilon}_0$ would be material dependent, this would not inßuence the general result of this investigation.

It is seen that, as well as for copper, the curves for the other metals are straight over a considerable range, and the intercepts with the axes determine the values for g_0 and σ_{v0} in Eq 17, which are characteristic for each material. It appears that a correlation exists between the two intercepts, so that this is actually a one-parameter set of curves, as seen from Fig. 6(b). They both have been interpreted in terms of the normalized SFE (Ref 42), and the values obtained here for g_0 and σ_{v0} are listed together with the reduced stacking fault energy $\chi/\mu b$.

Range of SFE. The materials of the present analysis cover a wide range of SFEs (Ref 45, 46), and thus, the results may be used to predict the master curve for any other fcc material with known stacking fault by determining the corresponding values for σ_{v0}/μ and of g_0, with the help of Fig. 6(b).

According to this evaluation, the SFE is the material parameter that controls (besides shear modulus and Burgers vector) the level of stress to which a material can be work hardened at a given temperature and strain rate. Thereby, the differences between various materials can differ by up to 1 order of magnitude, as seen from the ratio of the boundary values of the Voce stresses at zero temperature, $[\sigma_{v0}(0)/\mu_0]/[\sigma_{v0}(\infty)/\mu_0] = 13$. The $\chi/\mu b$ values of many pure fcc metals are close to or even larger than that of aluminum (Ref 47), and so, the corresponding master curves will gather in the vicinity of the aluminum curve, with some of them close to the lower bound in Fig. 6(a). In many solid solutions, however, $\chi/\mu b$ is reduced considerably, and the master curves will be shifted up into the vicinity of the silver curve (see the section Single-Phase Alloys in this article).

At the end of this analysis it seems it would be appropriate to address the issue of reliability. Uncertainties of the values of the SFE are certainly a crucial point for the predictive capacity of the interdependencies of the parameters given in Fig. 6(b), since SFE-values found in literature diverge considerably from method to method. Also, stress strain curves and thus Voce stresses, reported by different groups, often show differences even for chemically identical material. These, most probably, are caused by differences of microstructures according to the speci cs of employed fabrication and preparation techniques.

	g_0	σ_{v0}/μ	$\chi/\mu b$
	1.01	.04	.0
● Ag	.95	.026	.0018
▼ Cu	.87	.017	.0047
◇ Ni	.74	.011	.013
○ Al	.65	.007	.023
	.41	.003	∞
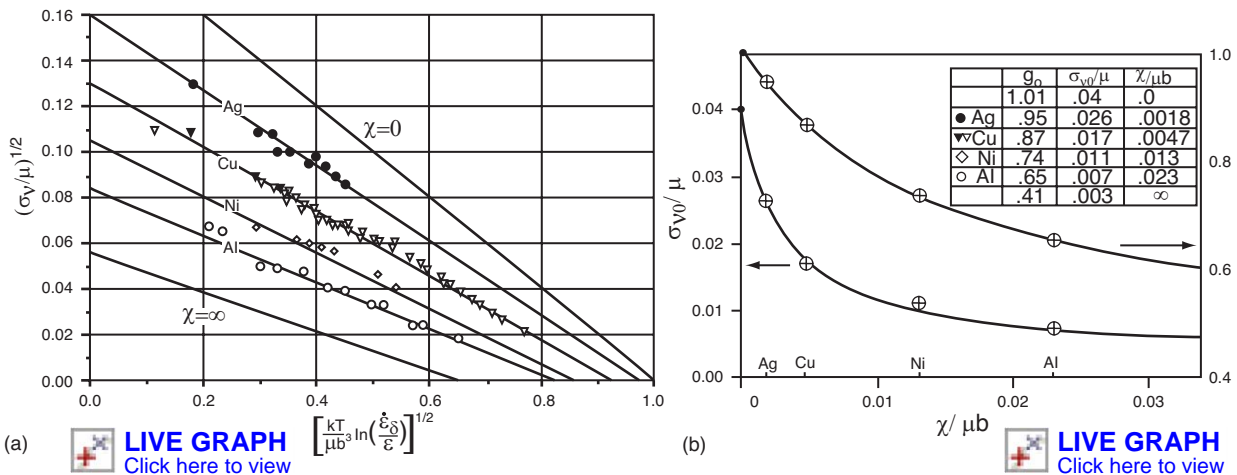
Fig. 6 (a) Same as Fig. 5(b) square root of σ_v/μ vs. square root of g, where data for aluminum and nickel (Ref 19, 26, 40) and silver (Ref 7, 38, 39) have been added, and those for copper have been supplemented (solid triangles) by data from torsion tests (Ref 41). Applied shear stress has been converted to normal stress with the help of the von Mises relationship. (Here, as in preceding figures at intermediate g-values, the data exhibit a small hump that is probably due to strain aging caused by impurities.) The two border lines for the values zero and infinity of the stacking fault energy (SFE) are obtained by the extrapolation method displayed in (b). (b) The intercepts with the axes in (a), σ_{v0}/μ_0 and g_0 versus the SFE $\chi/\mu b$. The empirical expressions $\sigma_{v0}/\mu_0 = 0.04 - 0.037/(1 + 0.003\ \mu b/\chi)$ and $g_0 = 1.01 - 0.6/(1 + 0.016\ \mu b/\chi)$, with the numerical values determined by a least-squares routine, make an optimum fit of the experimental points (Ref 8). The extrapolated coordinates (σ_{v0}/μ, g_0), for zero SFE (0.04, 1.01) and for infinite SFE (0.003, 0.4), determine the two demarcation lines in (a).

In particular, texture is an important microstructural feature in this context, since the real physical quantity to be considered is the resolved shear stress, τ_v, not the normal stress, σ_v, which is inßuenced by the Taylor factor, *M*. This is not known exactly, since the textures have not been measured for all materials considered in Fig. 5 and 6. So far, possible differences of *M* have been ignored in the preceding evaluation, implicitly assuming that ßuctuations were small and did not vary systematically from one material to the other. In view of the conventional thermomechanical treatment of the investigated samples, it is assumed that existing textures were rather weak. As has been discussed (Ref 43), by straining in different deformation modes, the texture evolves in such a manner that the corresponding Taylor factors differ by less than 20%. This is also evident in Fig. 6, where the data for copper obtained in compression and in torsion tests are virtually the same. Therefore, it seems safe just to use the Taylor factor $M \geq 3$ of a random orientation distribution for converting σ_v into $\tau_v = \sigma_v/3$.

Occasionally, however, larger differences have been reported, such as in Ref 44, where the Voce stresses obtained in tension were approximately 50% higher than those for copper in Fig. 5 and 6, but they extrapolate to the same value of g_0.

From these results, it would appear safer to relegate σ_{v0} to a quantity that needs to be determined in each case. The value of g_0, on the other hand, may well be a constant for each material. This would seem reasonable in view of the fact that it relates to a local activation process; it is expected to depend on physical properties such as the SFE but, unlike σ_{v0}/μ, not on microstructure, texture, and so on.

Other Lattice Structures

The bcc Metals

Like fcc, the bcc lattice offers an excessive number of identical crystallographic slip systems. In bcc metals, slip planes and slip directions are interchanged from the fcc case: The dominant slip mode is $\{110\} \langle 111 \rangle$. Slip along $\langle 111 \rangle$ directions in planes other than $\{110\}$ can also be activated, depending on temperature and material. The consequence of these additional slip planes leads to a mode of slip called pencil glide. The details of the slip geometry are of great interest for the evolution of texture and Taylor factors, but, most probably, they are of minor importance for the present discussion. It is to be expected that the principles of dislocation storage and dynamic will follow, qualitatively at least, the same rules as deduced previously from fcc behavior. To make this point, a set of stress-strain curves of low-carbon steel are analyzed, with the help of the Voce equation, and the characteristic parameters are compared with the fcc data. At rst, however,

some speci cs of bcc behavior must be considered:

· At temperatures below (roughly) room temperature, the ßow stress for most bcc metals rises very steeply with decreasing temperature due to the growing inßuence of the Peierls barriers; at low enough temperatures, these eventually control the dislocation mobility. Within the evaluated framework, there would be two effects: (1) The Peierls stress superimposes on the contribution of the dislocation interaction to the ßow stress, and (2) the dislocations are straighter, since the edge parts move out rapidly, and therefore, the areal-glide mechanism that forms the basis of the statistical dislocation-storage processes will be less effective; it may possible to describe this simply by a lower value of Θ_0 (Ref 48). On the other hand, Tang et al. (Ref 49) have postulated, on the basis of dislocation-dynamics simulations, that strain hardening in bcc metals follows an entirely different mechanism.

· Surprisingly low levels of interstitial impurities can strongly affect the strength of bcc metals, since interstitial atoms usually exert very strong interaction forces on dislocations, generally an order of magnitude higher than substitutional atoms. In addition, interstitials typically become highly mobile at a temperature slightly above room temperature and thus easily segregate at dislocations in almost any heat treatment. They cause upper and lower yield points, since dislocations must be unlocked from impurity clouds at the beginning of straining.

· The dependence of the ßow stress on the grain size is much stronger in bcc than in fcc metals, where it is so small for technically pure material that it could be neglected in the preceding evaluation.

In a compact way, the various aspects of deformation behavior of bcc material are reßected in the stress-strain curves for low-carbon steel presented in Fig. 7. The rather high strength level must be ascribed partly to the grain boundaries, but it is mainly due to the interstitial content mainly carbon. Interstitial carbon is also the cause for the occurrence of upper and lower yield points, even though the concentration is below 0.1% C. If the temperature is lowered beyond -75 °C, the behavior is governed by the increasing inßuence of the Peierls-Nabarro force, evident from the steep rise of the ßow stress with decreasing temperature in parallel with a continuous ßattening of the hardening curves (beyond the lower yield point). Finally, at approximately -100 °C, the ductile-to-brittle transition occurs as a consequence of the high strength level achieved at this point.

Between -75 °C and room temperature (RT), the yield stress is much less sensitive to temperature than in the low-temperature regime, and, once the Luders region has been passed, the hardening curves exhibit a shape similar to those
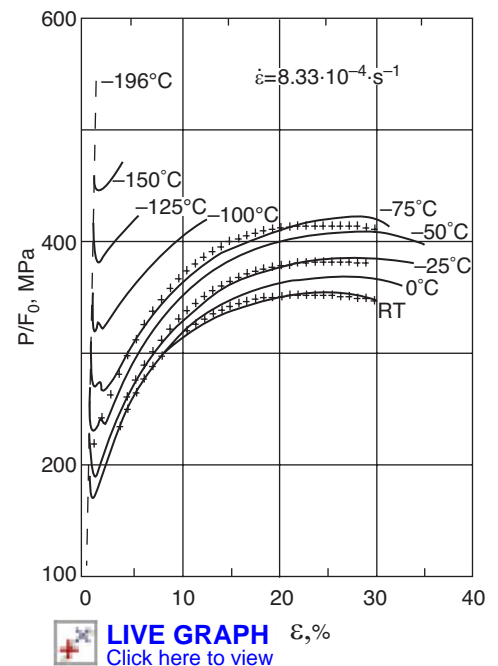
Fig. 7 Stress-strain curves (engineering stress vs. engineering strain) of a low-carbon steel at various temperatures below room temperature at a moderate strain rate. Source: Ref 50

for copper displayed in Fig. 3. From these parts, the characteristic parameters of the Voce equation are determined and compared with the results for fcc material. To this effect, the Voce equation (Eq 11b) is used in a modi ed form where the intrinsic, nondislocation contribution, σ_i, caused by grain boundaries, interstitials, and so on, is subtracted from the ßow stress to obtain the dislocation part $\sigma - \sigma_i$:

$$\sigma - \sigma_0 - \sigma_i = (\sigma_s - \sigma_0 - \sigma_i)$$
$$(1 - \exp -\varepsilon \cdot \Theta_0/[\sigma_s - \sigma_i]) \quad \text{(Eq 18a)}$$

Here, the intrinsic part of the ßow stress, σ_i, is assumed not to change with deformation beyond the Luders strain; σ_0 is the contribution of the initial dislocation density to the ßow stress (assumed to be small compared to σ_i and therefore neglected in the following); σ_s is the saturation level of the stress at large strain, and, in this terminology, the corresponding saturation value of the dislocation density determines the Voce stress, $\sigma_v = \sigma_s - \sigma_i$. It can be interpreted here as the Voce stress of pure iron under the assumption that the interstitials do not affect the hardening behavior, since, at these temperatures, they are not suf ciently mobile for segregation at and trapping of the network dislocations and thus do not impede the recovery processes (see the section Stage IV in this article).

To provide a direct comparison to the data of Fig. 7, as well as provide a general illustration of the approach, engineering quantities are adopted. Using $1 + \varepsilon_{eng} = \exp \varepsilon$ and $P/F_0 = \sigma_{eng} = \sigma \exp - \varepsilon$ (Eq 18a) reads:

$$\frac{P}{F_0} = \{1 + \varepsilon_{eng}\}^{-1} \left\{ \sigma_i + \sigma_0 + [\sigma_s - \sigma_i - \sigma_0 \geq \right.$$

$$\left. \left[1 - (1 + \varepsilon_{eng})^{-\Theta_0/[\sigma_s - \sigma_i \geq} \right] \right\} \qquad \text{(Eq 18b)}$$

In this equation, the values of Θ_0, σ_i, and $\sigma_v = \sigma_s - \sigma_i$ for an optimum t of the stress-strain curves of Fig. 7 were found in the following way. Since the yield points prevent reliable back extrapolation to zero strain for determining both σ_i and the athermal-hardening coef cient Θ_0, the latter was chosen as $\Theta_0 = 0.05\ \mu$, that is, the same as for copper according to Fig. 4(a). With this starting value, σ_i/σ_v pairs were determined for the different temperatures between $-75\ ^\bullet$C and RT by a least-squares routine, and the results in MPa are 162/336 for RT; 159/368 for $0\ ^\bullet$C; 167/392 for $-25\ ^\bullet$C; 193/406 for $-50\ ^\bullet$C; and 216/396 for $-75\ ^\bullet$C.

The curves computed with these values for θ_0, σ_i, and σ_v are inserted as dashed lines in Fig. 7. They match the experimental curves perfectly and thus allow a reliable back extrapolation across the Luders region to σ_i at zero strain. For a quantitative classi cation, the normalized Voce stresses of the steel are plotted in Fig. 8 in the eld of master curves of the various fcc metals. They are located between the aluminum and nickel curves, meaning that iron behaves similar to a fcc material with high SFE.

The general conclusion is that the rules developed in the preceding sections for fcc material hold for bcc metals as well and that the processes of dislocation storage and recovery are physically quite similar in the two material classes. The authors even believe that this is true for any lattice structure that offers a suf - cient number of slip systems of approximately the same strength, provided slip is areal due to an approximately equal slip distance for screw and edge dislocations.

Hexagonal Metals

Hexagonal metals represent a special case. Due to the low symmetry, the hexagonal lattice does not provide a suf cient number of crystallograhically identical slip systems to ful ll the von Mises criterion for polycrystal deformation. To generate a set of ve geometrically independent slip systems, several different crystallographic slip modes with often quite different strength must be activated, and twinning as an additional deformation mechanism comes into play in hexagonal metals much more frequently than in cubic metals (Ref 51, 52). While the strength of polycrystals is generally controlled by the hard modes, the situation is opposite in unconstrained deformation of single crystals. There, predominantly soft modes are activated, except for speci c orientations where the externally applied stress does not create a suf - ciently high shear stress in soft systems. In polycrystals, high interaction stresses build up between soft and hard grains to enforce the required co-deformation of soft and hard grains. These intergranular interaction stresses are strongly inßuenced by the crystallographic texture and depend on the externally imposed straining mode. To illustrate, for a number of magnesium polycrystals with different texture, Fig. 9 displays stress-strain curves obtained in different testing modes, namely, uniaxial tension and compression and also channel-die compression. The curves differ widely; in particular, profuse twinning leads to anomalies in the form of a signi cant reduction of the work-hardening rate at low strains.

Although dislocation mechanisms in hexagonal metals are basically the same as in cubic metals, the hardening behavior is quite different. The individual slip modes can respond, quantitatively as well as qualitatively, quite differently to changes of external or internal parameters, and the corresponding hardening is not so easy to analyze as for cubic material. Every hexagonal metal has its individual characteristics, and a general scheme such as the one developed in the preceding sections for fcc and bcc metals does not exist for hexagonal metals.

Stage IV

The deviation from a linear variation of hardening rate with ßow stress (Fig. 3b) and, as noted previously, the tendency toward a less negative slope at the tail of the curves corresponds to a low rate of work hardening over a relatively large range of strain. Stated another way, a mechanical state with no work hardening, and occurring at the scaling stress σ_v, is never developed. The linear change in hardening rate follows from the postulate of a steady-state process, embodied in Eq 11(a). Modi cations to Eq 11(a) have been proposed to better describe the transition in hardening rate, as given previously in Eq 12.

The deviation from linearity then suggests a change in the underlying mechanism of work hardening. It is common for a state of an approximately constant rate of work hardening to be developed (Fig. 10); such transition represents a change from stage III to stage IV work hardening. While the mechanisms underlying stage III work hardening and the allied mathematical description are fairly well established, models for stage IV are a topic of current research interest and debate. The authors objective then is to highlight how a transition to linear hardening may be effected in the context of the framework for work hardening given previously, with an eye toward practical analysis of data. After an abbreviated review, a framework is presented for the modeling of stage IV in the context of the standard model given previously.

One approach to modeling is to impose a transition from the linear change in work-hardening rate (Eq 11a) to a constant hardening rate for stage IV, Θ_{IV}, as:

$$\frac{d\sigma}{d\varepsilon} = \Theta_{IV} \qquad \text{(Eq 19)}$$

The rate dependence in the stage IV hardening rate may be introduced through a scaling of the saturation stress:
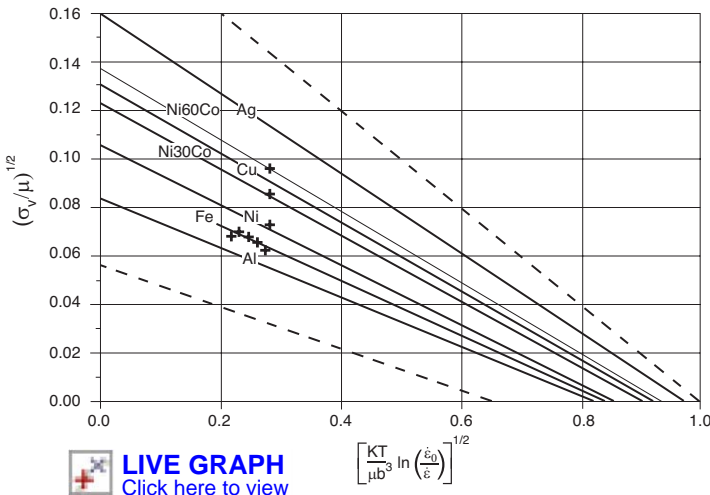
Fig. 8 Comparison with the field of face-centered cubic data of Fig. 6(a)
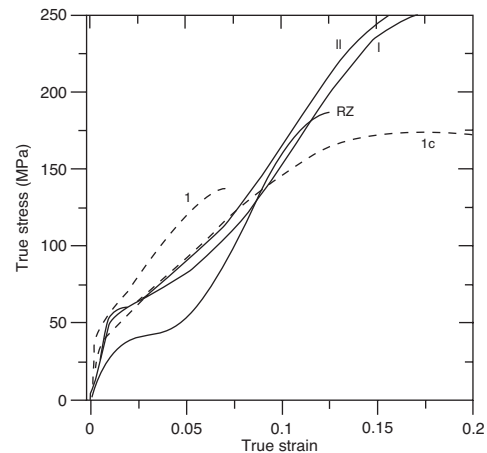
Fig. 9 Flow curves of random magnesium polycrystals, 1 and 1c, tested in tension and compression, respectively, and of textured polycrystals with profuse twinning, I and II, tested in compression. RZ, channel-die compression. Source: Ref 53
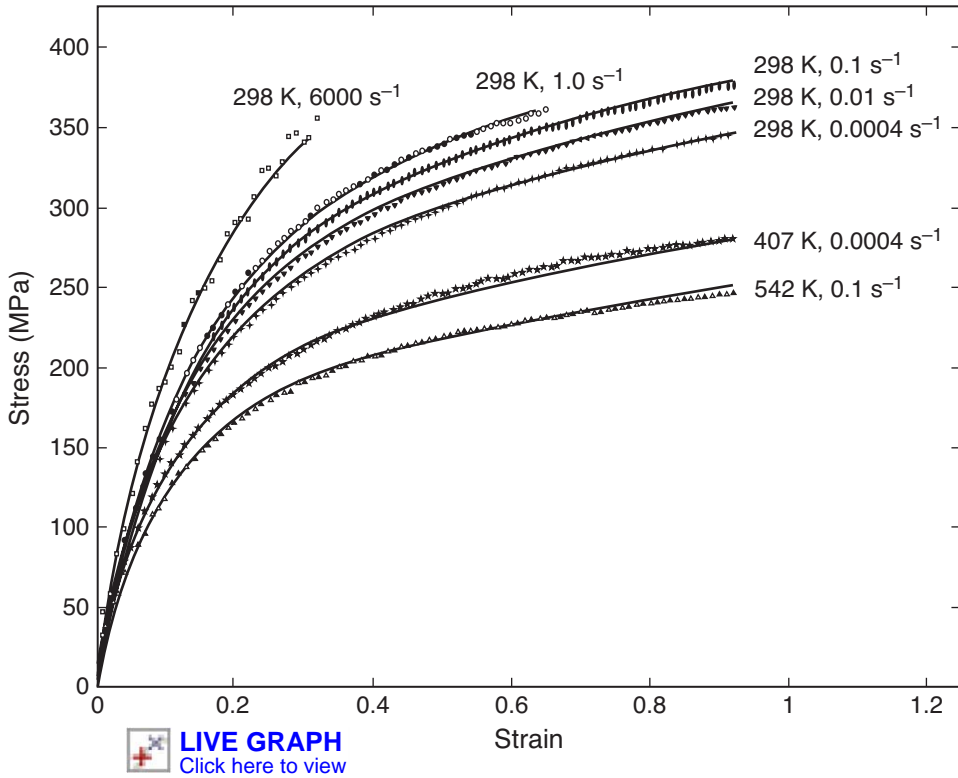
Fig. 10 Temperature and rate dependence of copper showing development of linear hardening at large strain. Symbols indicate experimental data (Ref 63), while solid lines indicate the authors' numerical simulations. Source: Adapted from Ref 62

$$\Theta_{IV} = a\sigma_v(\dot{\varepsilon}, T) \qquad \text{(Eq 20)}$$

where a is a constant between 0.05 and 0.1. Rollett *et al.* (Ref 54) set forth the mechanism of continuous accumulation of dislocation debris as a physical basis for the aforementioned scaling. Goto et al. (Ref 55) combine Eq 11(a) and 19 through:

$$\frac{d\sigma}{d\varepsilon} = \Theta_0\left[1 - \tanh\left(\frac{\kappa\sigma}{\sigma_v}\right)\right] + \Theta_{IV}\tanh\left(\frac{\kappa\sigma}{\sigma_v}\right) \qquad \text{(Eq 21)}$$

The storage of dislocation density, such that an estimation of cumulative Burgers vector leads to a nonzero result, brings about the concept of a net, excess, or polar dislocation density. Because the development of a net Burgers vector is tied to nonuniform plastic straining, or (inverse) elastic distortion (dominated by lattice rotation), these dislocations are often called geometrically-necessary dislocations (GNDs). The development of such a polar dislocation density has been considered as a basis for the development of stage IV by several authors. Argon and Haasen (Ref 56) associate the lattice misorientation between cells with GNDs and suggest evolution proportional to the square root of the plastic strain. Data for the spacing between geometrically-necessary boundaries (GNBs) and the misorientation angle across these GNBs in cold-rolled nickel may be tied to stress-strain response by linking the evolution of these features to the strength (Ref 57). In an in situ study of [001] copper single crystals using synchrotron radiation, Pantleon (Ref 58) assumes that there is an imbalance in the ßux of dislocations, giving a polar density that may be associated with the evolution of GNBs. Schaßer et al. (Ref 59) measure long-range stresses in stages III and IV. These authors infer a microstructural evolution wherein cell walls transition from dipoles bounding thicker cell walls in stage III to a thin monopole structure in stage IV, termed polarized tilt walls (PTWs). Mobile dislocations entering the monopolar cell walls tend to advance lattice misorientation, and the PTW structure provides for effective screening of the large dislocation densities developed in stage IV. The key conclusion to be drawn from the above proposals is that stage IV involves the progressive evolution of lattice misorientation, a geometric feature.

Estrin and Mecking (Ref 60) proposed a modiÞcation to the work-hardening rate to include a contribution to work hardening associated with geometrical obstacles. A second rate of storage is added to Eq 8:

$$\frac{d\rho}{d\gamma} = \frac{1}{b\beta_g l_g} + \frac{1}{b\beta}\sqrt{\rho} - R(\dot{\gamma}, T)\rho \qquad \text{(Eq 22)}$$

where β_g plays a role analogous to β for obstacles with geometrical spacing, l_g. This geometric length scale is taken in proportion to the grain size in Ref 60 and provides for an acceleration of the hardening rate with reduction in grain size. Alternative estimates of this geometric length scale may be posed. Considering incompatibilities following rotation of the lattice, it is possible to develop the density of polar (geometric or excess) dislocations threading the slip plane (Ref 61). This in turn provides an internal variable necessary to model evolution of the geometric features associated with stage IV.

Stage IV Work Hardening Model. In the present setting, it sufÞces to adopt a linear evolution of the inverse of the geometrical spacing, $1/l_g$, leading to the relationships:

$$\frac{d\rho}{d\gamma} = \frac{1}{b\beta_g l_g} + \frac{1}{b\beta}\sqrt{\rho} - R(\dot{\gamma}, T)\rho$$
$$\frac{d}{d\gamma}\left(\frac{1}{l_g}\right) = c \qquad \text{(Eq 29)}$$

The model for work hardening thus involves two internal variables, l_g and ρ, that obey evolution Eq 29 and require initial values. The approach of Estrin and Mecking (Ref 60) is recovered by taking $c = 0$ and setting l_g to the grain size. A generalization of Eq 29 and the association with evolution of various features of microstructure is given in Ref 62.

Following in large part the development leading to Eq 11(a):

$$\frac{d\sigma}{d\varepsilon} = \Theta_0\left(1 - \frac{\sigma}{\sigma_v} + \frac{\sigma_\lambda}{\sigma}\right)$$
$$\frac{d\sigma_\lambda}{d\varepsilon} = c \qquad \text{(Eq 30)}$$

The strength:

$$\sigma_\lambda = M\frac{\beta}{\beta_g}\frac{\alpha\mu_r b}{l_g} \qquad \text{(Eq 31)}$$

with shear modulus also taken at the reference temperature, μ_r, represents a second state variable associated with the evolution of substructure with polar dislocation content. Examples of the use of this model in modeling the development of stage IV in copper are shown in Fig. 10. Because of the presence of σ in the denominator of Eq 30, one must resort to numerical methods for solution using this model.

As an alternative, one may consider the role of polar dislocation content in recovery processes (as opposed to the storage rate). The rate of storage for travelling, mobile dislocations is reduced as polar dislocations become progressively arranged in cell walls or GNBs. Such a mechanism was envisioned by Pantleon (Ref 58), where a difference in signed ßux led to a reduction in annihilation rate relative to strictly statistical interactions. Assuming a linear evolution of this polar density with strain, the recovery term of Eq 10 is modiÞed as:

$$\Theta/\mu = \frac{M^2\alpha}{2\beta} - \frac{MR(\dot{\varepsilon}, T)(\sigma - \Theta_{IV}\varepsilon)}{2} \qquad \text{(Eq 32)}$$

thereby decreasing the recovery rate in proportion to strain through a parameter Θ_{IV}. Defining:

$$\kappa = 1 + \frac{\Theta_{IV}}{\Theta_0} \qquad \text{(Eq 33)}$$

and taking $\kappa\Theta_0 / \mu = M^2\alpha /2\beta$ and $\sigma_v/\mu = 2\kappa\Theta_0/MR(\dot\varepsilon, T)$ gives:

$$\Theta = \kappa\Theta_0\left(1 - \frac{[\sigma - \Theta_{IV}\varepsilon]}{\kappa\sigma_v}\right) \qquad \text{(Eq 34)}$$

Note that the standard model is recovered by taking $\Theta_{IV} = 0$. The evolution of various terms in the previous expression is shown in Fig. 11. This relationship for hardening rate may be integrated, giving:

$$\sigma = \sigma_0 + \Theta_{IV}\varepsilon + (\sigma_v - \sigma_0)[1 - \exp(-\Theta_0\varepsilon/\sigma_v)] \qquad \text{(Eq 35)}$$

Equations 34 and 35 prove effective in the analysis of stress-strain data exhibiting a stage IV response, as is shown in the next section.

Single-Phase Alloys

Solutes can affect flow behavior in two physically quite different ways: either globally, by changing bulk parameters, or locally, by acting as point obstacles to dislocation motion. Although these two contributions generally superimpose on each other, in many cases one of them dominates, so that the alloys fall roughly into two classes, of which representatives are introduced in Fig. 12 and 13.

Class 1, Crystal Properties

Alloying generally comes along with changes of elastic constants, lattice parameter, and SFE. While in most cases the shear modulus and the Burgers vector are changed only moderately, variations of the SFE can be quite large. This effect is made use of in many technical alloys, such as copper alloys (particularly α-brass and copper-aluminum bronze), nickel-cobalt alloys, and austenitic steels (stainless,
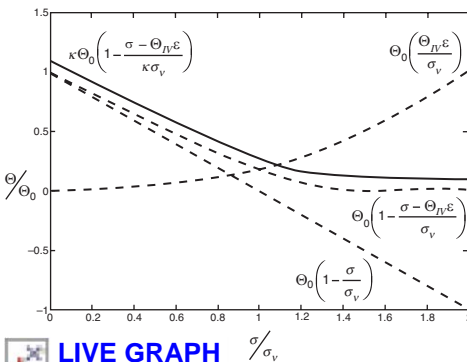
in particular, in combination with strain-induced martensitic transformation); they all have very low SFEs. What matters in these systems is the indirect effect of the alloy additions; they lead to an upward shift of the master curves in Fig 6(a), producing an increase of the work-hardening capacity and, in this way, generally improve the strengthening as well as the forming behavior, such as in deep drawing.

The classical example is the nickel-cobalt system, which serves here to demonstrate the usefulness and capacity of the foregoing analysis for evaluating hardening behavior. Since nickel and cobalt atoms have practically the same size and the same electron structure, cobalt substitutes in the nickel lattice are not noticed as point obstacles by moving dislocation and thus have little effect on the yield stress. They drastically change the work hardening, however, due to a lowering of the SFE (down to zero at the transition from the fcc to the hexagonal phase at approximately 75% Co).

Figure 12 shows stress-strain curves for nickel and nickel-cobalt polycrystals from torsion tests over a rather large strain range. All three curves approach quite a linear stage IV at large strain and thus qualify for determination of the various parameters in the modified Voce equation (Eq 34) developed in the previous section. The actual values for the parameters were found in the following way. The initial slope, $\Theta_{0torsion}/\mu$, and the slope in stage IV were directly taken from the curves in Fig. 12, and the obtained values are a single value for all three curves $\Theta_{0torsion}/\mu = 0.012$ and the individual values for $\Theta_{IVtorsion}/\mu$, 0.001, 0.0006, and 0.0005 for Ni-60%Co, Ni-30%Co, and nickel, respectively. With these starting values, the following $(\tau_{0torsion}/\mu)/(\tau_{vtorsion}/\mu)$ pairs were found by a least-squares routine: 0.00014/0.00429 for Ni-60%Co; 0.00023/0.00368 for Ni-30%Co; and 0.000189/0.00272
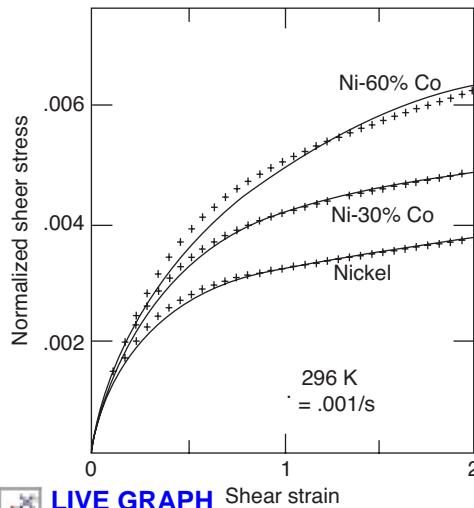
for nickel. These parameters, inserted in the modified Voce equation (Eq 35), lead to the dashed lines in Fig. 12; they match the experimental curves closely.

An interesting question is how these results compare quantitatively with the scheme developed for pure material in the section "Quantitative Description of Strain Hardening of fcc Polycrystals" in this article. For such comparison, applied shear stress and shear strain for torsion must be converted to normal stress and normal strain, which can be done with the help of Taylor factors. The ratio of Taylor factors of randomly textured polycrystals is $M_{uniaxial}/M_{torsion} = 3.06/1.8 = 1.7$; correspondingly, $\sigma/\tau_{torsion} = \gamma_{torsion}/\varepsilon = 1.7$ and $\Theta/\Theta_{torsion} = 1.7^2 = 2.89$. $\Theta_{0torsion} = 0.012$ is practically the same as found for copper in torsion (Ref 41). The corresponding value for uniaxial deformation, $\Theta_0/\mu = 2.89\,\Theta_{0torsion} = 0.035$, is 30% less than $\Theta_0/\mu = 0.05$ of uniaxially deformed copper (Fig 4a). Possibly, this points at an influence of the deformation mode on athermal hardening but also could be a texture effect.

The Voce stresses do not show such a tendency; as seen from Fig. 8, they fit well into the field of Voce stresses for pure fcc material. It must be noted, however, that all data in Fig. 8 stem from fitting the low-strain part of the stress-strain curve to the original Voce form, with $\Theta_{IV} = 0$. Considering only the strain regime $\gamma_{torsion} < 1$, that is, $\varepsilon < 0.6$, in Fig. 12, the values of σ_v/μ for nickel, Ni-30%Co, and Ni-60%Co are 0.0053, 0.0073, and 0.0089, respectively. These are in the sequence of inverse SFEs, with SFE/$\mu b = 0.013$ for nickel, 0.008 for Ni-30%Co, and 0.001 for Ni-60%Co (Ref 29, 64).

The value $\sigma_v/\mu = 0.0089$ for Ni-60%Co appears to be somewhat low in relationship to the reported SFE. The reason for this deviation from the developed scheme could not be

Fig. 11 Evolution of individual terms in Eq 34

Fig. 12 Stress-strain curves of nickel and two nickel-cobalt alloys (30 and 60% Co) measured in torsion at room temperature at a rate of 0.001/s (Ref 64). The stress is normalized by the shear modulus.
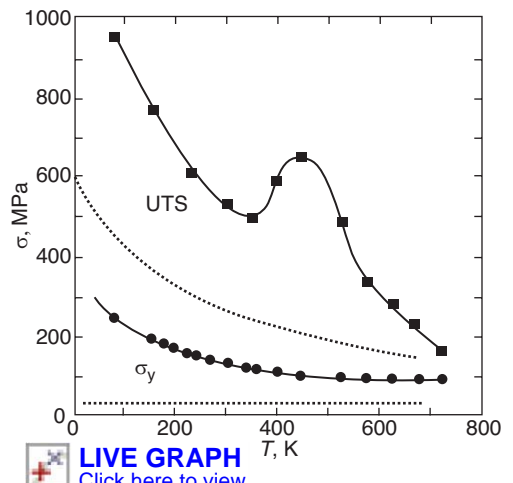
Fig. 13 Yield stress (σ_y) and ultimate tensile stress (UTS) as a function of temperature for a copper-tin alloy (experimental points from Vöhringer 1976) in comparison to pure copper (dashed curves). Source: Ref 76.

cleared up so far. It may be due to uncertainties in the determination of the SFE but could also be caused by mechanical twinning; lowering of the SFE increases the work-hardening capacity and, at the same time, reduces the stress to activate twinning as an additional slip mode. Although the general course of stress-strain curves may be affected by twinning only in a limited strain interval (due to exhaustion of twinning), the inßuence on the microstructure is generally quite large, by way of grain segmentation combined with an enhanced tendency to inhomogeneous deformation such as shear banding, which has a strong inßuence on texture evolution (Ref 65, 66).

So, the nal conclusion is that, in any case, the SFE is the ruling material parameter that accounts for the quantitative differences of the stress-strain behavior in Fig. 12, directly by the effect on hardening parameters and indirectly by the inßuence of twinning on texture and microstructure.

Class 2, Dislocation Interaction

There are, however, many alloy systems where the prime effect is by the local interaction of the solutes as point obstacles with dislocations, rather than by changes of SFE and corresponding modi cations of the dislocation core. Prominent examples for this class of materials are alloys of copper and tin, which have been in service since ancient times. A speci c case is discussed in more detail as follows (Fig. 13).

If a solute atom is of quite different size and of quite different chemical nature, as compared with the host atoms, it carries a local stress eld, which is called parelastic if it exists independently (size effect) and dielastic if it is induced by locally acting stresses (modulus effect) due to local variations of the elastic constants in the vicinity of a solute atom. Then, the alloying atoms can exert strong interaction forces on the dislocations, causing a considerable contribution to the glide resistance of dislocations and resulting in a rather large increase of the initial yield strength. This inßuence can be accounted for by an additive term in Eq 3 for the critical resolved shear stress:

$$\tau = \tau_i(c, \dot{\varepsilon}, T) + \alpha\mu b\sqrt{\rho} \qquad \text{(Eq 36)}$$

where τ_i, the intrinsic contribution of the solute atoms, depends on their concentration, c, and also on temperature and strain rate due to thermal activation. If this were the only change of mechanical properties by solutes, then the developed scheme for work hardening of the pure materials could also be applied in a straightforward fashion for these alloys, just by replacing σ and τ in the developed equations by $\sigma - \sigma_i = M [\tau - \tau_i]$, anticipated already in the modi ed Voce equations (Eq 18a, b). However, the situation is often more complex, as illustrated in Fig. 13, where the behavior of pure copper and a copper-tin alloy is compared for a wide range of deformation temperatures.

The values for the yield stress and the ultimate tensile stress (UTS) are presented; the UTS, the stress at the beginning of necking, is used here as a substitute for the Voce stress, which was not available for the alloy. The UTS and the Voce stress have quite similar values and depend on temperature in quite the same way. For example, if Eq 3 is applicable, they are related by $\sigma_{UTS} = \sigma_v/(1 + \sigma_v/\Theta_0)$ (Ref 8). The ratio σ_v/Θ_0 decreases in the copper case from approximately 0.3 at zero temperature toward negligibly small values at elevated temperature, meaning that σ_v exceeds σ_{UTS} by approximately 30% at very low temperature but attains at medium temperatures practically the same values as the UTS.

As seen in Fig. 13, the difference in the σ_{UTS}-σ_y for the alloy is generally almost twice as large as for the pure material, which means that the tin addition not only increases the yield stress but also the hardening rate (possibly due to a reduction of the SFE). At medium temperature, however, a hump appears in the dependence of the UTS on temperature, and the UTS versus T passes a maximum. This is the regime of strain aging, where the solute atoms develop suf cient mobility for segregating on the dislocations during deformation. Then, the temperature dependence of the UTS as well as the, not shown there, strain-rate dependence become inverse, in the sense that the ßow stress and the work-hardening coef cient increase with temperature and decreasing strain rate. Often, these phenomena come along with plastic instabilities and serrated ßow (Ref 67).

The main conclusion in the present context is that merely a modi cation of Eq 3 for the ßow stress, such as in Eq 19, will not be suf cient for adequately describing the observed behavior. A modi cation of Eq 8 is required as well, where the solute concentration enters the recovery term R in a complex way. Generally, the corresponding inßuence of alloying on the hardening is even stronger than on the yield strength. It has been found that, in many alloys, solute concentrations inßuence the entire stress-strain curve in a multiplicative rather than an additive way (Ref 68). Thereby, the combined effect of the direct and indirect contributions may play a speci c role. The speci c effect of solutes in the temperature regime of strain aging is not only due to their segregation on the moving but, even to a higher extent, it is due to segregation on the network dislocations, where they have a stabilizing inßuence and can slow down the rate of dynamic recovery considerably.

At the end of this section, one aspect seems worth mentioning that cannot be treated in more detail within the scope of this article, although it is important in many technical materials. Alloy additions usually cause an increasing inßuence of the grain size on the yield strength, possibly due to coarsening of slip bands and also to the segregation of solutes on grain boundaries.

Assessment

Modeling Base for Parameter Identification

Work hardening is due to the continuous increase of the dislocation content and can be understood on the basis of an athermal term of dislocation storage counteracted by a thermally activated recovery term. They both depend on only one state variable, the average dislocation density, ρ, although by a different power. In the ideal case, the storage term increases with power ½, the recovery term with power of unity, and superposition of the two terms leads to a dependence of the work-hardening coef - cient, Θ, on the ßow stress, σ. This dependence can be condensed into a single master curve by a simple scaling procedure, namely by normalization of Θ with the shear modulus, μ, and σ by a characteristic stress level, σ_v. The dependence of σ_v on temperature and strain rate can also be condensed in another master curve, by plotting normalized quantities, namely σ_v/μ versus $g = (kT/\mu b^3)\ln(\dot{\varepsilon}_0/\dot{\varepsilon})$. These two master curves contain the complete information about the hardening behavior of a fcc metal at temperature and strain rate outside the diffusion regime. Differences between various materials are governed by their relative SFEs. Therefore, the use of these normalizations in any physics-based modeling exercise of plastic ßow of metals is recommended.

Previously, a physically motivated framework for parameter determination was set forth. The plotting of data with appropriate scaling allows for direct contrast with equations for constitutive description that have a physical basis and are supported by the analysis of a variety of metals and alloys. The predictive capability of the modeling framework in a particular application follows from consistency with the entire database embodied in the master curves. Both presence and absence of such consistency were related in the previous development. Voce stresses of the steel plotted in Fig. 8 suggested that an analogy may be drawn between iron and a fcc material with high SFE, and forest interactions are the source of work hardening; the anomalous value of σ_v/μ for Ni-60%Co suggests that another deformation mechanism is likely at play in this alloy, not considered in this development. Such insight follows from consideration of how the parameters identi ed fall within the overarching constitutive model framed in the master curves. The application of the resulting equations by direct tting of parameters through a regression analysis is not, in general, recommended.

The utility of the one-parameter approach extends to the identi cation of dynamic recrystallization in austenite, through identi cation of a critical strain by double differentiation of the $\Theta - \sigma$ curve (Ref 69). Prior to the onset of dynamic recrystallization, the analysis of stress-strain data for several steels was shown

to follow the hardening relationship Eq 22, without the term for statistical storage (Ref 70). The one-parameter approach can be extended to multiple-phase systems through the introduction of an additional length scale. Examples are the treatment of grain size (Ref 60) and precipitation strengthening (Ref 71). The rule of mixtures may be exploited for the co-deformation of interpenetrating phases through the superposition of separate Voce equations for the distinct phases. Generally speaking, the approach is valid for modeling plastic ßow at constant rate and temperature. Sudden changes of rate and temperature lead to transients that require the introduction of a second state variable (Ref 8).

Outstanding Issues

Given the complexity inherent in the phenomenon, the study of work hardening will continue as an active area of research. The study of length scales in plasticity is an active area of present study (for a review, see Ref 72 and other references in the same issue). The length scale inherent in the development of lattice misorientation was used in the previous presentation of stage IV. Predictive modeling of stage IV (and also V and VI) hardening requires attention to the identi cation of appropriate state variables.

The issue of transients has been mentioned previously. The treatment of long-range internal stresses presents considerable challenge; the state-of-the-art for so-called kinematic work hardening lies far behind that of the description of isotropic work hardening contained herein. This is by virtue of the fact that concern only be lent to the nearest strong obstacle. Anelastic processes, wherein a (possibly recoverable) microplastic strain is developed, pose additional complexity (Ref 73 75).

ACKNOWLEDGMENTS

REFERENCES

1. R. Peierls, Size of Dislocation, *Phys. Soc.—Proc.,* 52(289), 1940, p 34 37
2. F.R.N. Nabarro, Dislocations in a Simple Cubic Lattice, *Proceedings of the Physical Society,* 59, 1947, p 256 272
3. F.R.N. Nabarro, Fifty-Year Study of the Peierls-Nabarro Stress, *Mater. Sci. Eng. A,* Vol 234 236, 1997, p 67 76
4. P. Lukac and Z. Trojanova, Hardening and Softening in Selected Magnesium Alloys, *Mater. Sci. Eng. A,* 462(1 2), 2007, p 23 28
5. H. Neuhauser, *Dislocations in Solids,* F.R.N. Nabarro, Ed., North Holland, 6, 1983, p 319
6. H. Mecking and K. Lucke, Die Verfestigung von Silber-Einkristallen Zwischen 77 und 1200 K, *Z. Metallk.,* 60, 1969
7. M. Biermann, Doctoral Thesis, RWTH Aachen, 1975
8. U.F. Kocks and H. Mecking, Physics and Phenomenology of Strain Hardening: The FCC Case, *Prog. Mater. Sci.,* 48(3), 2003, p 102
9. B. Devincre, T. Hoc, et al., Dislocation Mean Free Paths and Strain Hardening of Crystals, *Science,* 320(5884), 2008, p 1745 1748
10. L.P. Kubin, Dislocation Patterning during Multiple Slip of FCC Crystals. A Simulation Approach, *Phys. Status Solidi (a),* 135(2), 1993, p 433 443
11. U.F. Kocks, A Statistical Theory of Flow Stress and Work-Hardening, *Philos. Mag.,* 13, 1966, p 541 566
12. R. Madec, B. Devincre, et al., Simulation of Dislocation Patterns in Multislip, *Scr. Mater.,* 47(10), 2002, p 689 695
13. U.F. Kocks, *Fifth International Conference on the Strength of Metals and Alloys,* Pergamon Press, 1979
14. B. Devincre, L. Kubin, et al., Physical Analyses of Crystal Plasticity by DD Simulations, *Scr. Mater.,* 54(5), 2006, p 741 746
15. A.H. Cottrell and R.J. Stokes, Effects of Temperature on the Plastic Properties of Aluminium Crystals, *Proc. R. Soc. (London) A,* (Mathematical and Physical Sciences), 233, 1955, p 17 34
16. F.R.N. Nabarro, *International Conference on the Strength of Metals and Alloys,* Pergamon Press, 1986
17. H. Mecking and U.F. Kocks, Kinetics of Flow and Strain-Hardening, *Acta Metall.,* 29(11), 1981, p 1865 1875
18. J. Diehl, Tensile Tests on Copper Single Crystals, Zugverformung von Kupfer-Einkristallen, *Z. Metallk.,* 47(5), 1956, p 331 343
19. S.R. Chen and U.F. Kocks, High Temperature Plasticity in Copper Polycrystals, *High Temperature Constitutive Modelling,* A.D. Freed and K.P. Walker., Ed., ASME, 1991
20. S.R. Chen and U.F. Kocks, *High-Temperature Plasticity in Copper Polycrystals,* ASME, New York, NY, USA, 1991
21. A. Seeger, Kristallplastizitat, *Handbuch der Physik, VII,* F.S. Berlin, Ed., Springer Verlag, 1958
22. N. Hansen, New Discoveries in Deformed Metals, *Metall. Mater. Trans. A:Phys. Metall. Mater. Sci.,* 32A(12), 2001, p 2917 2935
23. A.S. Argon, *Strengthening Mechanisms in Crystal Plasticity,* New York, Oxford University Press, 2008
24. M. Zaiser and A. Seeger, Long-Range Internal Stresses, Dislocation Patterning and Work Hardening in Crystal Plasticity, *Dislocations in Solids,* M.S. Duesbery and J.P. Hirth, Ed., Amsterdam, North-Holland, 11, 2002, p 1 99
25. E. Voce, The Relationship between Stress and Strain for Homogeneous Deformation, *J. Inst. Met.,* 74, 1948, p 536 562
26. U.F. Kocks, Laws for Work-Hardening and Low-Temperature Creep, *J. Eng. Mater. Technol., Trans. ASME,* 98 Ser H(1), 1976, p 7685
27. J.P. Hirth, Inßuence of Grain Boundaries on Mechanical Properties, *Metall. Trans.,* 3(12), 1972, p 3047 3067
28. H. Mecking and Y. Estrin, Microstructure Related Constitutive Modeling of Plastic Deformation, *Constitutive Relations and Their Physical Basis,* S.I. Anderson, et al., Ed., Riso National Laboratory, 1987, p 123
29. K. Lucke and H. Mecking, Dynamic Recovery, *The Inhomogeneity of Plastic Deformation,* R.E. Reed-Hill, Ed., ASM, 1973, p 223
30. C.A. Bronkhorst, S.R. Kalidindi, et al., *Textures Microstruct.,* 14 18, 1991, p 1031
31. P. Haasen, Plastic Deformation of Nickel Single Crystals at Low Temperatures, *Philos. Mag.,* 3, 1958, p 384 418
32. A. Seeger, R. Berner, et al., Experimental Determination of Stacking-Fault Energies in Face-Centered Cubic Metals, *Z. Phys.,* 155(2), 1959, p 247 262
33. S.E. Buhler, K. Lucke, and F.W. Rosenbaurn, Stacking Fault Energy of Silver in the Temperature Range $172{\geq}$ to $873{\geq}$K Determined by Plasticity Measurements, *Phys. Stat. Solid.,* 3(5), 1963, p 886 900
34. H. Mecking and K. Lucke, *Second International Conference on the Strength of Metals and Alloys,* ASM, 1970
35. H. Wolf, The Activation Energy for Cross Slip of a Dissociated Screw-Dislocation, *Z. Naturforsch. A (Astrophysik, Physik und Physikalische Chemie),* 15a(3), 1960, p 180 193
36. U. Kocks, A. Argon, et al., Thermodynamics and Kinetics of Slip, *Prog. Mater. Sci.,* 19, 1975
37. B. Nicklas and H. Mecking, *Fifth International Conference on the Strength of Metals and Alloys,* Pergamon Press, 1979
38. J.E. Bailey and P.B. Hirsch, The Dislocation Distribution, Flow Stress and Stored Energy in Cold-Worked Polycrystalline Silver, *Philos. Mag.,* 5, 1960, p 485 497
39. A.J. Beaudoin, A. Acharya, et al., Consideration of Grain-Size Effect and Kinetics in the Plastic Deformation of Metal Polycrystals, *Acta Mater.,* 48(13), 2000, p 3409 3423
40. U.F. Kocks and S.R. Chen, On the Two Distinct Effects of Thermal Activation on Plasticity: Application to Nickel, *Phys. Status Solidi (a),* 131(2), 1992, p 403 413
41. J.M. Alberdi Garitaonandia, University of Navarra, 1984
42. H. Mecking, B. Nicklas, et al., A Universal Temperature Scale for Plastic Flow, *Acta Metall.,* 34(3), 1986, p 527 535

43. U.F. Kocks, M.G. Stout, et al., *Eighth Int. Conf. Strength of Metals and Alloys,* Pergamon Press, 1988

44. F.P. Bullen and M.M. Hutchinson, Dynamic Recovery from Strain-Hardening in Polycrystalline Copper and Aluminium, *Philos. Mag.,* 7, 1962, p 557 572

45. P.C.J. Gallagher and Y.C. Liu, Diversity of Stacking Fault Energy Determinations and Its Signi cance, 17(2), 1969, p 127 137

46. D.A. Hughes, J.C. Gibeling, et al., *Seventh International Conference on the Strength of Metals and Alloys,* Pergamon, 1986

47. N.M. Rosengaard and H.L. Skriver, Calculated Stacking-Fault Energies of Elemental Metals, *Phys. Rev. B (Condensed Matter),* 47(19), 1993, p 12865 12873

48. U.F. Kocks and S.R. Chen, unpublished work

49. M. Tang, B. Devincre, et al., Simulation and Modelling of Forest Hardening in Body Centre Cubic Crystals at Low Temperature, 7, 1999, p 893 908

50. W. Dahl and H. Rees, The Grain Size Dependence of the Yield Stress of Structural Steels in the Range of Homogeneous Deformation, *Archiv fur das Eisenhuettenwesen,* 50(8), 1979, p 355 360

51. C. Tome and S. Kok

52. C. Tome and U.F. Kocks, The Yield Surface of H.C.P. Crystals, *Acta Metall.,* 33 (4), 1985, p 603 621

53. C.H. Caceres, P. Lukac, et al., Strain Hardening Due to {10(1)2} Twinning in Pure Magnesium, *Philos. Mag.,* 88(7), 2008, p 991 1003

54. A.D. Rollett, U.F. Kocks, et al., Stage IV Work Hardening in Cubic Metals, *Formability and Metallurgical Structure,* A.K. Sachdev and J.D. Embur, The Metals Society, Warrendale, PA, 1987, p 211

55. D.M. Goto, J.F. Bingert et al., Mechanical Threshold Stress Constitutive-Strength Model Description of HY-100 Steel, *Metall. Mater. Trans. A: Phys. Metall. Mater. Sci.,* 31(8), 2000, p 1985 1996

56. A.S. Argon and P. Haasen, New Mechanism of Work Hardening in the Late Stages of Large Strain Plastic Flow in F.C.C. and Diamond Cubic Crystals, *Acta Metall. Mater.,* 41(11), 1993, p 3289 3306

57. D.A. Hughes and N. Hansen, Microstructure and Strength of Nickel at Large Strains, *Acta Mater.,* 48, 2000, p 2985 3004

58. W. Pantleon, Stage IV Work-Hardening Related to Disorientations in Dislocation Structures, *Mater. Sci. Eng. A, Struct. Mater.: Prop. Microstruct. Proc.,* 387 389, 2004, p 257 261

59. E. Schaßer, K. Simon, et al., A Second-Order Phase Transformation of the Dislocation Structure during Plastic Deformation Determined by In-Situ Synchrotron X-Ray Diffraction, *Acta Mater.,* 53, 2005, p 315 322

60. Y. Estrin and H. Mecking, A Uni ed Phenomenological Description of Work Hardening and Creep Based on One-Parameter Models, *Acta Metall.,* 32(1), 1984, p 57 70

61. A. Acharya and J.L. Bassani, Lattice Incompatibility and a Gradient Theory of Crystal Plasticity, *J. Mech. Phys. Solids,* 48(8), 2000, p 1565 1595

62. S. Kok, A.J. Beaudoin, et al., On the Development of Stage IV Hardening Using a Model Based on the Mechanical Threshold, *Acta Mater.,* 50(7), 2002, p 1653 1667

63. A.B. Tanner and D.L. McDowell, *Int. J. Plast.,* 15, 1999, p 375

64. D.A. Hughes and W.D. Nix, Strain Hardening and Substructural Evolution in Ni-Co Solid Solutions at Large Strains, *Mater. Sci. Eng. A: Struct. Mater.: Prop. Microstruct. Proc.,* A122(2), 1989, p 153 172

65. A.S. Malin and M. Hatherly, Microstructure of Cold-Rolled Copper, *Met. Sci.,* 13 (8), 1979, p 463 472

66. K. Morii, H. Mecking, et al., Development of Shear Bands in F. C. C. Single Crystals, *Acta Metall.,* 33(3), 1985, p 379 386

67. U.F. Kocks, Kinetics of Nonuniform Deformation, *Prog. Mater. Sci.,* Chalmers Anniversary Volume, 1981, p 186 241

68. U.F. Kocks, ASM materials seminar Deformation, Processing and Structure, ASM, 89, 1982

69. E.I. Poliak and J.J. Jonas, A One-Parameter Approach to Determining the Critical Conditions for the Initiation of Dynamic Recrystallization, *Acta Mater.,* 44(1), 1996, p 127 136

70. J.J. Jonas, X. Quelennec, et al., The Avrami Kinetics of Dynamic Recrystallization, 2009

71. Y. Estrin, Dislocation-Density-Related Constitutive Modeling, *Unified Constitutive Laws of Plastic Deformation,* A.S. Krausz and K.K. San Diego, Ed., Academic Press, 1996, p 69 106

72. A. Needleman and J.G. Sevillano, Preface to the Viewpoint Set on: Geometrically Necessary Dislocations and Size Dependent Plasticity, *Scr. Mater.,* 48(2), 2003, p 109 111

73. H. Garmestani, M.R. Vaghar, et al., A Uni ed Model for Inelastic Deformation of Poly Crystalline Materials Application to Transient Behavior in Cyclic Loading and Relaxation, *Int. J. Plast.,* 17(10), 2001, p 1367 1391

74. L. Zhu, A.J. Beaudoin, et al., An Assessment of In-Service Stress Relaxation of a Work-Hardened Aluminum Magnesium Alloy, *J. Eng. Mater. Technol., Trans. ASME,* 126(2), 2004, p 157 163

75. L. Zhu, A.J. Beaudoin, et al., *On the Time-Dependent Inelastic Deformation of Metals,* USA, AIP, 2004

76. O. Vohringer 1976, Verformungsverhalten von vielkristallinen Kupferlegierungen *Metall.,* 30, 1150 1160

Constitutive Models for Superplastic Flow

Indrajit Charit, University of Idaho
Rajiv S. Mishra, Missouri University of Science and Technology

SUPERPLASTICITY, in a widely accepted definition, is "the ability of a polycrystalline material to exhibit, in a generally isotropic manner, very high tensile elongations prior to failure" (Ref 1). Note that this definition does not involve any microstructural or mechanistic description. As a rule of thumb, elongation in excess of 200% is termed as superplasticity. Over the years, a number of superplasticity models have been proposed. A majority of experimental evidences suggests that grain-boundary sliding (GBS) is the dominant deformation mechanism. Generally, superplasticity is obtained at homologous temperatures of above ∼0.4 because diffusional processes become rapid enough to impart superplastic property. Superplasticity as a field of study has matured from a mere academic interest to applications in commercial superplastic forming (SPF) operations. Currently, SPF is considered an important near-net shape technique for niche applications. For a detailed historical account of the advent of superplasticity, readers are encouraged to refer to the article references and other sources in the literature (Ref 2–9).

Superplasticity observed in fine-grained microstructures is called structural superplasticity; it is quite universal and has been observed in a wide range of material classes. Superplasticity can be obtained by other ways too, such as transformation superplasticity and thermal cycling superplasticity. However, the latter mechanisms are not as technologically important and universal. Hence, this article is limited to the discussion of constitutive equations applicable to structural superplasticity. Henceforth, reference to the word *superplasticity* will only mean structural superplasticity and not any other types of superplasticity.

At present, SPF is used in a number of manufacturing operations, most notably in aircraft industries. Diffusion bonding/SPF is an important technology that is used to make integrally stiffened structures. This enhances design flexibility and the ability to fabricate complex structures in a single operation, which helps in reducing the number of parts in an assembly.

The materials that are routinely used for SPF are titanium and aluminum alloys. Generally, the flow stresses associated with SPF are less (<10 MPa). Various SPF techniques, such as simple female forming, reverse bulging, snap-back forming, and so on, have been summarized by Pilling and Ridley (Ref 9). Superplasticity is strongly microstructure dependent and requires fine grain size (<15 μm) and high-angle grain boundaries. Therefore, grain refining helps in increasing the optimum strain rate during SPF and reduces the superplastic temperature (Ref 6). It has led to useful concepts of high-strain-rate superplasticity (forming rates $\geq 10^{-2}$ s^{-1}) and low-temperature superplasticity.

As superplasticity continues to gain acceptance in industrial applications, the importance of simulation and prediction of superplastic flow during SPF has increased. Thus, constitutive models have a special significance in the description of superplastic flow. According to the Webster's dictionary, the meaning of *constitutive* is "of or relating to constitution" of the material. So, the constitutive models would present relationships that relate macroscopic properties of the material representative of the inherent behavior of the material. Often, these models attempt to relate material state with the external parameters (such as stress and temperature). Physicists and materials scientists have worked on phenomenological and physics-based equations. On the other hand, engineering mechanics researchers have come up with continuum-based constitutive equations. For more details on constitutive equations, see the article "Constitutive Equations" in *Metalworking: Bulk Forming,* Volume 14A, *ASM Handbook,* 2005, page 563.

For simulating SPF processes, it is important to have a set of reliable constitutive equations. From the general rules of constitutive laws, the ideal constitutive equations for superplasticity should have the following characteristics (Ref 5):

- For useful constitutive equations, the deformation state of the superplastic material should be derivable from the stress state and vice versa. For that, the constitutive

equations should have a closed form; that is, the number of equations should be equal to the number of unknowns.
- Where possible, the constitutive equations should be written in tensor form to be of interest to complicated and industrial SPF equations.
- The equations should indicate what type of experiments must be conducted in order to evaluate the material constants, and those should be feasible in practice.
- The constitutive equations should have material constants that are stable against minor variation in experimental conditions, thus making the constitutive models more stable and useful.
- The range of applicability of the constitutive equations should be experimentally determined and supported theoretically.

Mechanical Description of Superplasticity

The relationship between the strain rate and stress is given by the following simple relation:

$$\sigma = K \dot{\varepsilon}^m \qquad \text{(Eq 1)}$$

where σ is the flow stress, $\dot{\varepsilon}$ is the strain rate, and K is a material constant that also depends on microstructure, temperature, defect structure of the material, and so on. The term m is called the strain-rate sensitivity of the material and is defined by the following equation:

$$m = \frac{d \ln \sigma}{d \ln \dot{\varepsilon}} \qquad \text{(Eq 2)}$$

That is, the strain-rate sensitivity of deformation is determined from the slope of the double logarithmic plot of flow stress against strain rate.

Although several test modes, such as torsion tests, compression tests, load relaxation tests, and so on, have been used to evaluate superplastic materials, uniaxial tensile tests are the

main test mode for superplasticity. Constant cross-head velocity and constant strain-rate tensile tests are common. In uniaxial tensile tests, strain-dependent axial force and elongation are obtained at a constant temperature and microstructure. Instantaneous strain rates are obtained by dividing the cross-head velocity by the instantaneous gage length of the tensile specimen. Flow stress is obtained from the instantaneous axial force data compensated against specimen cross-sectional area using constant volume assumption.

Strain-rate sensitivity (m) determines the ability of a material to resist necking during tensile straining. In general, the higher the strain-rate sensitivity, the higher the resistance to necking and the larger the tensile elongations. This concept is clear from the analysis of Avery and Stuart (Ref 10). Based on the assumption that m is constant over a limited range of strain rates associated with a small diffuse neck, one can write from Eq 1 (by expressing stress and strain rate in terms of load and area):

$$\dot{A} = -\left(\frac{P}{K}\right)^{1/m}(A)^{(m-1)/m} \qquad \text{(Eq 3)}$$

where $\dot{A}$ is the change in the cross-sectional area (A) of the specimen with time, P is the tensile axial force, and the other terms have been defined in Eq 1. Equation 3 implies that if the m-value increases, the rate of cross-sectional change becomes more and more independent of the current cross section. When the m-value approaches 1, there should be no effective necking because of Newtonian viscous flow. However, when the m-value hovers at approximately 0.5 or more, the elongations are large (i.e., superplasticity).

The conditions for conventional superplastic deformation can be summarized as following (Ref 5):

- The average grain size (d) should be smaller than some critical value, that is, <15 μm. However, several other microstructural factors are important, such as grain shape, presence of second phases, and nature and mobility of grain boundaries.
- The temperature during superplastic deformation should be at least 0.4 to 0.5 T_m (where T_m is the melting point in K).
- The strain rates at which superplasticity is observed are within a certain range (10^{-5} to 10^{-1} s^{-1}).
- Superplasticity is observed when the strain-rate sensitivity value is more than 0.3 and thus is only present in region II, as shown in Fig. 1. On the other hand, regions I and III with strain-rate sensitivity values less than 0.3 fail to show any significant superplasticity.

Phenomenological Constitutive Models

Phenomenological constitutive equations do not have a micromechanistic basis to start with.

These equations, along with the material constants, are developed by observing phenomena via conducting several sets of macroexperiments. These equations are generally applicable in the range of experimental conditions under which they are developed. Note that Fick's laws of diffusion are phenomenological in nature. Even though these constitutive equations are developed by purely phenomenological means, most of them embody a general mathematical form that can alternatively be derived from a physical origin as well. Here, only a few of the phenomenological constitutive equations for superplasticity are noted because the phenomenological equations are abundant in superplasticity literature. There are relations ranging from the standard power-law types, various polynomial models, mechanical models (such as generalized Maxwell and Bingham body problems; mechanical threshold models such as Zehr and Backofen, Ref 11; etc.), Smirnov's model based on elasto-viscoplastic analysis (Ref 12), and anelasticity-based equations. However, for a complete treatment of phenomenological constitutive equations for superplastic deformation, readers may consider referring to an excellent text by Padmanabhan et al. (Ref 5).

Bird-Mukherjee-Dorn Equation. This phenomenological equation proved to be an extremely important constitutive relation for describing superplastic deformation and for performing relevant analyses. This was first developed by Bird, Mukherjee, and Dorn (Ref 13) and hence the name. This was an extension of a general form of equation developed earlier by Mukherjee et al. for describing high-temperature deformation behavior (creep) in crystalline materials (Ref 14). Superplasticity is closely related to creep (both are thermally activated processes). This equation presents a general correlation between strain rate, stress, temperature, and grain size and is given by the following general form:

$$\dot{\varepsilon} = \frac{AD_oGb}{kT}\left(\frac{\sigma}{G}\right)^n\left(\frac{b}{d}\right)^p\exp\left(-\frac{Q}{RT}\right) \qquad \text{(Eq 4)}$$

where $\dot{\varepsilon}$ is strain rate, σ is stress, D_o is the frequency factor for diffusivity, Q is the appropriate activation energy for diffusion, R is the universal gas constant, k is Boltzmann's constant, T is temperature (in K), G is shear modulus, b is Burgers vector, d is grain size, n is the stress dependence of strain rate (or stress exponent), p is the inverse grain-size exponent, and A is a microstructural- and mechanism-dependent constant. The constant A can involve microstructural factors other than grain size, such as stacking-fault energy, dislocation density, grain-boundary character distribution, and so on, but their exact description still remains poorly understood. It is important to recognize that n is just the reciprocal of m. Higher m-values mean greater resistance to external neck formation and hence larger ductility (i.e., superplasticity). Generally, an m-value of $\sim$0.5 (or $n = 2$) and a p-value of 2 or 3 imply the operation of GBS.
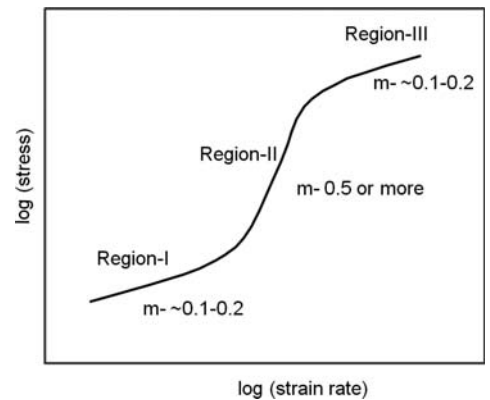
Fig. 1 Schematic illustration of strain-rate dependence on flow stress in a typical superplastic material. m, strain-rate sensitivity factor

Other Phenomenological Equations. Diulgerov et al. (Ref 15) developed the following equation while working on the superplastic behavior of zinc-manganese alloys:

$$\sigma = K\dot{\varepsilon}^m\varepsilon^\theta\left(\frac{T_m - T}{TT_m}\right)^q \qquad \text{(Eq 5)}$$

where T_m is the melting point, as shown before; $\dot{\varepsilon}$ is the associated strain; and θ is the strain-hardening exponent. The values of K, m, and θ could be found out from regression analysis of the data. However, in that particular study, much scatter in the parameter was observed. Later, a similar analysis was applied to a lead-tin eutectic alloy and a Bi-Pb-Sn-Cd alloy (Ref 16). However, there was still considerable scatter in the parameters evaluated.

Ghosh and Hamilton (Ref 17) proposed a polynomial model:

$$\sigma = \sigma_o + B\varepsilon \qquad \text{(Eq 6)}$$

where $\log\sigma_o = a_o + a_1\log\dot{\varepsilon} + a_2\log^2\dot{\varepsilon} + a_3\log^3\dot{\varepsilon} + a_4\log^4\dot{\varepsilon}$.

The polynomial models developed through linear regression methods tend to provide stable and unambiguous results. However, the range of applicability of these models is limited and has not been investigated well.

Smirnov (Ref 12) proposed a model that considers superplastic materials as elasto-viscoplastic medium. The constitutive equation is written as:

$$\sigma = \sigma_s\frac{\sigma_o + K_v\dot{\varepsilon}^{m_v}}{\sigma_s + K_v\dot{\varepsilon}^{m_v}}, \dot{\varepsilon} \neq 0 \qquad \text{(Eq 7a)}$$

where σ_s, σ_o, K_v, and m_v are material constants. Equation 7(a) can also be expressed in the following way:

$$\lim_{\dot{\varepsilon}\to 0}\sigma = \sigma_o \qquad \text{(Eq 7b)}$$

$$\lim_{\dot{\varepsilon}\to\infty}\sigma = \sigma_s \qquad \text{(Eq 7c)}$$

$$\dot{\varepsilon} = 0 \text{ for } \sigma \le \sigma_o \text{ and } \dot{\varepsilon} = \left\{ \frac{\sigma_s(\sigma - \sigma_o)}{K_v(\sigma_s - \sigma)} \right\}^{1/m_v}$$

$$\text{(Eq 7d)}$$

Smirnov (Ref 12) has given the following method to calculate the material constants. First, the point of inflection on the sigmoidal curve (as shown in Fig. 1) is found, and the maximum slope ($m_{\max}$) is calculated. The value of K_v is determined from the relation $K_v = \sigma_{\text{opt}}/\dot{\varepsilon}^{m_{\max}}$. Next, the values of σ_o and σ_s are found using Eq 7(b) and (c) by extrapolating. Finally, the value of m_v is calculated from the following equation:

$$m_v = m_{\max}\left(\frac{\sqrt{\sigma_s} + \sqrt{\sigma_o}}{\sqrt{\sigma_s} - \sqrt{\sigma_o}}\right) \qquad \text{(Eq 7e)}$$

It has been shown that the model proposed by Smirnov can accurately describe the universal superplasticity curve. Further modifications to this equation have been made to take into account the transient region of loading and the influence of grain size. This model can be generalized to describe cases of nonuniform stress-strain rate and is used in Russia in practical calculations. However, the drawback of this model is the need for determination of a large number of unknown empirical parameters.

Physically Based Constitutive Equations

Physical constitutive equations of superplasticity have been developed on the basis of the physics of the process. These equations include microstructural parameters and have a strong micromechanistic basis. Two contiguous grains in a polycrystalline material cannot slide indefinitely because they impinge on other grains. Various processes, such as strain accommodation, grain rotation, grain-boundary migration, grain-switching events, and so on, take place to relax the stress concentrations generated and thus continue GBS. The common school of thought assumes that the process that accommodates GBS is the rate-controlling mechanism for superplasticity. Here, some classical physical constitutive equations are discussed. The accommodation mechanisms could be divided into two major groups: diffusional accommodation and accommodation by dislocations.

Diffusional Accommodation Models

During the later part of 1960s, there were attempts to explain region II in Fig. 1 (the superplastic regime) using diffusional creep models (Ref 18, 19). Both Nabarro-Herring (Ref 20, 21) and Coble creep (Ref 22) were also considered. The features of diffusional creep models, such as high strain-rate sensitivity (=1) and strong grain-size dependence of strain rates, are attractive. However, it was realized that diffusional creep must be accommodated

by some sort of GBS (Lifshitz type) to retain cohesiveness between grains. Raj and Ashby (Ref 23) also argued that the process can be described either as diffusional creep accommodation by GBS or vice versa. However, the following are serious drawbacks of diffusional creep in explaining superplasticity (Ref 4):

- Very little grain elongation is expected (much less than the total specimen elongation that is the hallmark of superplasticity).
- Strain rates expected of diffusional creep models are less than what is observed in superplastic deformation.
- Diffusional creep models do not involve grain rotations that are topologically observed in superplasticity.

Ashby and Verrall (Ref 24) proposed a model that explains superplasticity as a transition region between a diffusion-accommodated flow at low strain rates and a diffusion-controlled dislocation climb at higher strain rates. At low strain rates, where diffusion-accommodated flow accounts for most of the total strain rate, the specimen elongation is obtained by grain rearrangement through GBS (and grain switching). To retain compatibility across the grain boundaries, the grains must have transient yet complex shape changes that are achieved via diffusional transport. The final configuration retains the equiaxed grain shape. Due to the transient increase in grain-boundary surface area during the grain-rearrangement process, a threshold stress arises at very low strain rates. This model is depicted in Fig. 2. However, at higher strain rates, where dislocation creep accounts for most of the total strain rate, specimen elongation is accompanied by the change of shapes of individual grains. At intermediate strain rates, both processes (diffusion-accommodated flow and dislocation creep) would contribute. Because these two mechanisms are independent processes, the total strain rate is given by the sum of the strain rates contributed by each process:

$$\dot{\varepsilon}_{\text{total}} = \dot{\varepsilon}_{\text{diff}} + \dot{\varepsilon}_{\text{dislo}} \qquad \text{(Eq 8a)}$$

$$\dot{\varepsilon}_{\text{diff}} = \frac{100\,\Omega}{kTd^2}\left(\sigma - \frac{0.72\Gamma}{d}\right)D_L\left(1 + \frac{3.3\delta}{d}\frac{D_{\text{GB}}}{D_L}\right)$$

$$\text{(Eq 8b)}$$
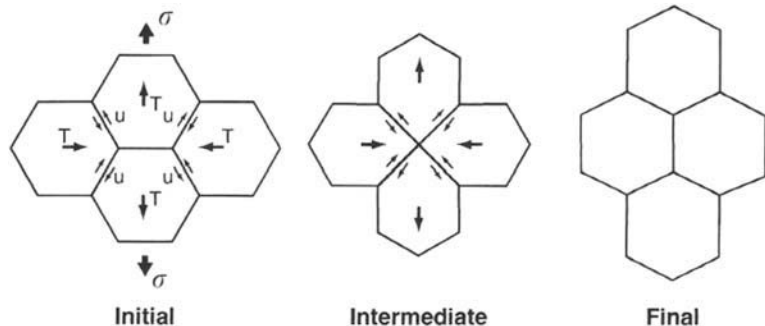
$$\dot{\varepsilon}_{\text{dislo}} = \frac{AD_L\,Gb}{kT}\left(\frac{\sigma}{G}\right)^n \qquad \text{(Eq 8c)}$$

where Γ is the grain-boundary free energy, Ω is the atomic volume, δ is the grain-boundary width (high diffusivity path), and D_L and D_{GB} are the volume and grain-boundary diffusion coefficients, respectively. This equation was thought to provide accurate description within a factor of 2.

When combined, the model predicts strain-rate sensitivity (m) as a strong function of strain rate with a maximum approaching unity. The grain-size dependence may vary from 0 to 3, depending on the strain rate and temperature. The activation energy predicted by this model will be intermediate between the activation energies of lattice diffusion and grain-boundary diffusion.

Even though the Ashby-Verrall model has attractive features and explains topological characteristics well, there are a few shortcomings in this model:

- Differences between the predicted and experimentally observed strain rates are quite large.
- Although the model predicts grain-size exponent (p) values of 0 to 3, experimental results reveal that p lies between 2 and 3.
- The prediction of the model that activation energy increases with temperature has no sound experimental support.

Work by Mohamed and Langdon (Ref 25), Smith et al. (Ref 26), and Spingarn and Nix (Ref 27) has shown that the Ashby-Verrall model has significant drawbacks besides the aforementioned ones.

Dislocation Pileup Accommodation

There are a number of constitutive equations proposed by various researchers over a long span of time on dislocation-accommodation-based concepts. Dislocation-accommodation models can again be classified into three types, depending on the precise nature of the role of dislocations in the accommodation process (pileups within grains, pileups in the grain

Fig. 2 Grain switching in Ashby-Verrall model. Source: Ref 24

Initial   Intermediate   Final

boundaries, and individual dislocation accommodation).

Pileups within Grains. Ball and Hutchison (Ref 28) studied superplasticity in a fine-grained Zn-22.9wt%Al alloy and proposed a physical constitutive model. The physical basis of this model is illustrated in Fig. 3. The model assumes that the easy relative movement of groups of grains is resisted by a few unfavorably oriented grains. Dislocations are generated at such grains and pile up at the opposite grain boundaries until the back stress obstructs further GBS. The stress concentration at the head of the pileup is relieved via accelerated diffusion, and dislocations escape via a climb process, going into and along the grain boundaries and subsequently becoming annihilated. This escape of dislocations from the pileup allows for more GBS in the obstructed groups of grains. This model suggested that each group of grains consists of approximately four grains. The model developed can be expressed by the following constitutive equation:

$$\dot{\varepsilon} = 200 \frac{D_{\mathrm{GB}} G b}{kT} \left(\frac{\sigma}{G}\right)^2 \left(\frac{b}{d}\right)^2 \qquad \text{(Eq 9)}$$

Mukherjee (Ref 29) proposed a modification of the Ball-Hutchison model. In this model, it was assumed that grains slide individually instead of sliding as groups. Dislocations are generated at the ledges and other discontinuities at the grain boundaries; they traverse through the grain and pile up at the grain boundaries. Similar to the Ball-Hutchison model (Ref 28),
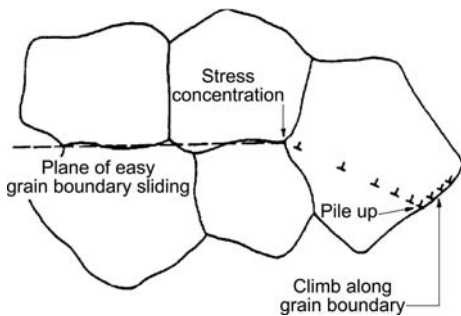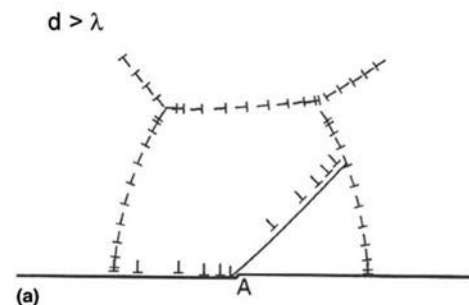
the rate of sliding is governed by the climb rate of the dislocations at the lead of the pileup and subsequent annihilation. The model leads to an equation that is essentially similar to the Ball-Hutchison model (Eq 9) except for the value of the dimensionless constant (A), which is approximately 2 in Mukherjee's model as calculated from theoretical considerations, whereas the A-value of 200 in Ball-Hutchison's model was derived from fitting of the experimental data. However, the lack of dislocation pileups in transmission electron microscopy investigation of superplastically deformed alloys raised questions about the validity of the models. However, these models are still highly regarded by the superplasticity community. Also, additional questions were raised about the presence of multiple dislocation sources in a grain; that is, the dislocation emission rate may be more important than the dislocation annihilation rate. In such a case, the form of the equation would change.

Fukuyo et al. (Ref 30) have proposed a model that is similar to the Ball-Hutchison model of GBS accommodated by slip. The slip-accommodation process involves two sequential steps of glide and climb. When the climb is the rate-controlling step, the strain-rate sensitivity value is 0.5. On the other hand, when the glide is the rate-controlling step, the strain-rate sensitivity is equal to 1. Because the glide and climb processes are sequential, the slowest of the two processes is rate-controlling. So, this model predicts that at lower strain rates, the GBS process is accommodated by climb, whereas it is accommodated by glide at higher strain rates.

Langdon (Ref 31) proposed a unified model for Rachinger GBS (Ref 32), that is, GBS with grain rearrangement but no individual grain elongation, both in creep and superplastic conditions where the rate of sliding is controlled by the rate of accommodation through intragranular slip. The basic model is pictorially shown in Fig. 4. Ball and Hutchison identified in the concluding section of their seminal paper, "...the grain size is stable and smaller than the dislocation cell structure that would normally form under the conditions (stress and temperature) of deformation." Some of the premises of the Ball-Hutchison model were used in

Langdon's model. There are two distinct cases in Langdon's model of Rachinger sliding. For d (grain size) < λ (dislocation self-trapping distance), the relevant constitutive equation is:

$$\dot{\varepsilon}_{\mathrm{gbs}(d<\lambda)} = \frac{A_{\mathrm{gbs}(d<\lambda)} D_{\mathrm{GB}} G b}{kT} \left(\frac{\sigma}{G}\right)^2 \left(\frac{b}{d}\right)^2 \qquad \text{(Eq 10a)}$$

in which $A_{\mathrm{gbs}(d<\lambda)}$ is a dimensionless constant with a theoretical value of ∼10. When $d > \lambda$, the constitutive equation is:

$$\dot{\varepsilon}_{\mathrm{gbs}(d<\lambda)} = \frac{A_{\mathrm{gbs}(d>\lambda)} D_{L} G b}{kT} \left(\frac{\sigma}{G}\right)^3 \left(\frac{b}{d}\right)^1 \qquad \text{(Eq 10b)}$$

where $A_{\mathrm{gbs}(d>\lambda)}$ is a dimensionless constant with a value of 1000, which was determined by fitting relevant experimental data.

Pileups in the Grain Boundaries/ Interfaces. In a different study, Mukherjee (Ref 33) proposed a modification to his earlier model. Here, GBS is considered rate-limited by the motion of dislocations inside the grain boundary by a combined glide-climb process, but dislocation motion across the grains is not involved here. The compatibility between the grains is maintained by diffusion-controlled climb of lattice dislocations along the grain boundaries, and thus, repeated accommodation is achieved. The rate equation is similar to that presented in Eq 9 but differs only in the value of the constant A (=100).

Gifkins (Ref 34) refined the earlier version of his model in a more quantitative way and developed a mathematical description of his model, known as the core-mantle model. Here, the mechanism is described in terms of the motion of grain-boundary dislocations (Fig. 5). In this model, grain-boundary dislocations pile up at the triple junctions. The resulting stress concentration is relieved by the dissociation of lead grain-boundary dislocations into the other two adjoining boundaries of the triple junction or into lattice dislocations that accommodate sliding. These newly formed dislocations will climb into the two grain boundaries and annihilate or may form new grain-boundary dislocations. However, all these unit steps occur near the vicinity of the grain-boundary or mantle. Gifkins predicted that the width of the

Fig. 3 Schematic illustration of the Ball-Hutchison mechanism of superplasticity. Source: Ref 28

Fig. 4 Schematic illustration of grain-boundary sliding in two types of microstructural conditions: (a) for $d > \lambda$ and (b) for $d < \lambda$. Source: Ref 31
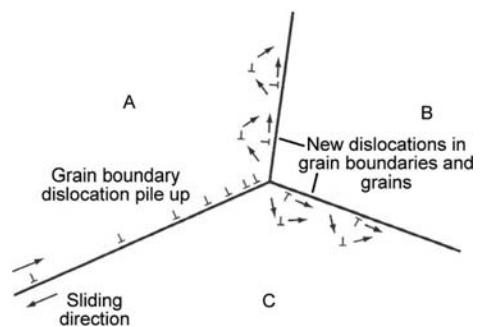
Fig. 5 Core-mantle theory with accommodation by dislocation motion in and near grain boundaries. Source: Ref 34

mantle will be ∼0.07 times the grain diameter. Thus, in a typical superplastic material, the grain-boundary width would be only a few tens of nanometers wide. The core of the grain remains almost free of dislocations. This model leads to the grain rotation and grain rearrangement that are geometrical necessities for superplastic deformation. The constitutive equation appropriate for this model is:

$$\dot{\epsilon} = 64 \frac{D_{GB} G b}{kT} \left(\frac{\sigma}{G}\right)^2 \left(\frac{b}{d}\right)^2 \qquad \text{(Eq 11)}$$

Gifkins (Ref 34) validated the model with experimental observations in tin-lead eutectic, Ni-Cr-Fe, Cu-40wt%Zn, and other alloys. Further work on the core-mantle concept was done by Mayo and Nix (Ref 35) and much later by Ghosh (Ref 36). However, it is currently not known whether these newer models can be applicable to a wider range of experimental conditions or only to a limited range.

Accommodation by Individual Dislocation Motion. Hayden et al. (Ref 37) developed a model where GBS is rate-controlled by intragranular slip (dislocation creep). Dislocations are generated at grain-boundary ledges and triple junctions, move inside the grain via glide and climb processes, and finally climb individually into the opposite grain boundaries (without forming pileups) before being annihilated. The model assumes that at a critical temperature (T_c), there will be a transition in the diffusion mechanism from pipe (at $T < T_c$) to lattice diffusion (at $T > T_c$). Moreover, they suggested that the rate of sliding is related to the rate of intragranular dislocation creep by a geometric constant that is independent of material and temperature, and the ratio of the two rates would vary inversely with grain size. In their model, it was considered that vacancy creep (Coble creep), GBS, and dislocation creep mechanisms are independent, and the total strain rate should be given by the sum of all the strain rates contributed by the mechanisms. They gave a set of complex equations that can be further simplified. For the superplastic regime, the model leads to the following form of equations:

$$\dot{\epsilon} \propto \frac{D_P G b}{kT} \left(\frac{\sigma}{G}\right)^2 \left(\frac{b}{d}\right)^3 \text{ for } T < T_c \qquad \text{(Eq 12a)}$$

where D_P is the dislocation pipe diffusivity, and:

$$\dot{\epsilon} \propto \frac{D_L G b}{kT} \left(\frac{\sigma}{G}\right)^2 \left(\frac{b}{d}\right)^2 \text{ for } T > T_c \qquad \text{(Eq 12b)}$$

For $T > T_c$, the proportionality constant (so-called A-constant) is given by:

$$\frac{\epsilon \pi (1 - \nu)}{10} K_1$$

where ν is the Poisson's ratio, and K_1 is an empirical constant of ∼0.015 cm that is not

expected to change much among various superplastic materials.

Arieli and Mukherjee (Ref 38) proposed that individual lattice dislocations climb into and along the interfaces in a narrow region near the interfaces. During the climb process, dislocations multiply via the Bardeen-Herring (Ref 39) mechanism. Due to the nearness of interfaces, this climb process is controlled by grain-boundary diffusion. The model predicts the rate equation as:

$$\dot{\epsilon} = \left(\frac{4\pi\omega}{hZ^2 \tan\left(\frac{\theta}{2}\right)}\right) \frac{D_{GB} G b}{kT} \left(\frac{\sigma}{G}\right)^2 \left(\frac{b}{d}\right)^2 \qquad \text{(Eq 13)}$$

The significance of this model is that the constant term is not considered a geometric constant; rather, it would vary with interface structure (through h and θ, where h is the proportionality constant between the sliding distance and Burgers vector, and θ is the mismatch angle between adjacent grains) and the structure of the narrow zone near the grain boundary or interface (through ω and Z, which are substructure-related terms).

Other Superplasticity Models. Perevezentsev et al. (Ref 40) came up with an interesting physics-based constitutive model. They analyzed the mechanism of GBS at grain triple points and discontinuities as being possible by a cooperative action of local grain-boundary migration, mass transport via diffusion, and emission of lattice dislocation into the grain interior. The kinetics of GBS is influenced by these processes. They derived the following constitutive relationship for superplasticity:

$$\dot{\epsilon} = 100 \frac{D_{GB} G b}{kT} \left(\frac{\sigma}{G}\right)^2 \left(\frac{b}{d}\right)^2 \qquad \text{(Eq 14)}$$

Their model provided good matching with the experimental data obtained from superplastic alloys, such as Sn-1%Bi, Zn-22%Al, and so on.

Another interesting mechanism for superplasticity could be the solute-drag dislocation-glide mechanism, similar to class I solid-

solution behavior in creep. In this mechanism, glide is the rate-controlling process of the glide-climb dislocation creep. Solute atoms tend to form solute atmospheres around the dislocations in certain temperature/strain-rate regimes, thus imposing a dragging force on the gliding dislocations. This model predicts a strain-rate sensitivity exponent value of 0.33 ($n = 3$). Here, the rate of superplastic deformation does not depend on grain size (i.e., $p = 0$). The constitutive equation (Ref 41) can be written as:

$$\dot{\epsilon} = A' D_{chem} \left(\frac{\sigma}{G}\right)^3 \qquad \text{(Eq 15)}$$

where D_{chem} is the chemical diffusivity of solute atoms in the solvent matrix, and A' is an approximate constant that depends on the solute-solvent atoms misfit-size parameter and solute concentration. If the formalism of the equation is kept the same as the Bird-Mukherjee-Dorn equation, one may obtain a constant of approximately 6. Although the strain-rate sensitivity value is <0.5, it is still high enough (∼0.33) to exhibit large elongations. That is why many coarse-grained, class I solid-solution alloys (such as aluminum-magnesium alloys) exhibit high ductility (>200%) at elevated temperatures.

Applicability of Superplastic Constitutive Equations

Figure 6(a) shows normalized stress and strain-rate data from a superplastic nanocrystalline 1420 alloy (Ref 42). The constitutive equations proposed by Ball and Hutchison (Ref 28, Eq 9) and Mukherjee (Ref 29) are also plotted. The constitutive equation given by Mishra et al. (Ref 43) is included because it represents microcrystalline alloys quite well. For the high-pressure torsioned nanocrystalline samples, the experimental data suggest that the kinetics of deformation is much slower than the rates predicted by constitutive equations used for microcrystalline materials. This is
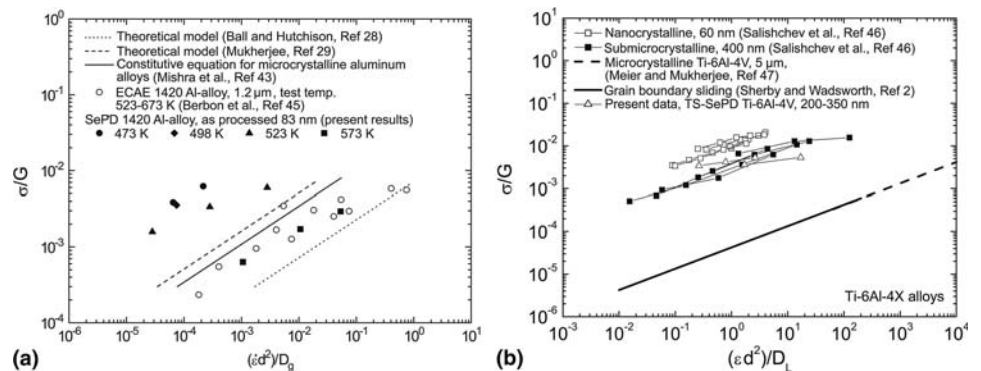
Fig. 6 (a) Normalized stress versus normalized strain-rate plot for SePD (high-pressure torsion, as-processed grain size 83 nm) 1420 aluminum alloy. Source: Ref 42. Other constitutive equations have been plotted along with experimental data of the ECAE 1420 alloy for comparison. (b) Variation of normalized flow stress with grain size and diffusivity-compensated strain rate for superplasticity data across the nanocrystalline-to-microcrystalline state. Source: Ref 44

due to the difficulty of dislocation generation and thus lack of the accommodation mechanism of GBS. Such discrepancy is not limited to aluminum alloys. Figure 6(b) shows a normalized plot for Ti-6Al-4V alloy. Again, the nanocrystalline and ultrafine-grained data do not match with the microcrystalline data. These observations imply that the constitutive equations discussed in prior sections should be used with caution when applied to the microstructural state that is significantly different from the length scale used in original theoretical formulations.

REFERENCES

1. S. Hori, M. Tokizane, and N. Furushiro, Ed., *Preface to Superplasticity in Advanced Materials (ICSAM-91),* JSRS, Osaka, Japan, 1991

2. O.D. Sherby and J. Wadsworth, Superplasticity—Recent Advances and Future Directions, *Prog. Mater. Sci.,* Vol 33, 1989, p 169

3. A.H. Chokshi, A.K. Mukherjee, and T.G. Langdon, Superplasticity in Advanced Materials, *Mater. Sci. Eng.,* Vol R10, 1993, p 237

4. J.W. Edington, K.N. Melton, and C.P. Cutler, Superplasticity, *Prog. Mater. Sci.,* Vol 21, 1976, p 1

5. K.A. Padmanabhan, R.A. Vasin, and F.U. Enikeev, *Superplastic Flow: Phenomenology and Mechanics,* Springer, Berlin, NY, 2001

6. T.G. Nieh, J. Wadsworth, and O.D. Sherby, *Superplasticity in Metals and Ceramics,* Cambridge University Press, Cambridge, U.K., 1997

7. O.A. Kaibyshev, *Superplasticity of Alloys, Intermetallides, and Ceramics,* Springer-Verlag, Berlin, NY, 1992

8. A.A. Presnyakov, *Superplasticity of Metals and Alloys,* Boston Spa, 1976

9. J. Pilling and N. Ridley, *Superplasticity in Crystalline Solids,* The Institute of Metals, London, 1989, p 159

10. D.H. Avery and J.M. Stuart, 14th Sagamore Army Materials Research Conference (Raquette Lake, NY), 1967

11. S.W. Zehr and W.A. Backofen, Superplasticity in Lead-Tin Alloys, *Trans. Am. Soc. Met.,* Vol 61, 1968, p 300–313

12. O.M. Smirnov, *Metal Working in a Superplastic State,* Metallurgy Publishing House, Moscow, 1979 (in Russian)

13. J.E. Bird, A.K. Mukherjee, and J.E. Dorn, *Quantitative Relation between Properties and Microstructure,* D.G. Brandon and A. Rosen, Ed., Jerusalem University Press, 1969, p 255–342

14. A.K. Mukherjee, J.E. Bird, and J.E. Dorn, *Exp. Correlat. High Temp. Creep,* Vol 62, 1969, p 155–179

15. N. Diulgerov, A. Istatkov, N. Mitev, and I. Spirov, Superplastic Low Manganese Zinc-Manganese Alloy, *Superplasticity and Superplastic Forming,* C.H. Hamilton and N.E. Paton, Ed., TMS, Warrendale, PA, 1988, p 419–427

16. R.A. Vasin and F.U. Enikeev, *Introduction to the Mechanics of Superplasticity. Part I,* Gilem, Ufa, Russia (in Russian)

17. A.K. Ghosh and C.H. Hamilton, Influence of Material Parameters and Microstructure on Superplastic Forming, *Metall. Trans. A,* Vol 13, 1982, p 733–743

18. D.H. Avery and W.A. Backofen, Structural Basis for Superplasticity, *Trans. Q. ASM,* Vol 58, 1965, p 551–562

19. M.J. Stowell, J.L. Robertson, and B.M. Watts, Structural Changes during Superplastic Deformation of Al-Cu Eutectic Alloy, *Met. Sci. J.,* Vol 3, 1969, p 41–45

20. F.R.N. Nabarro, *Report on Conference on Strength of Solids,* Physical Society, London, 1948, p 75

21. C. Herring, Polycrystalline Viscosity of Polycrystalline Solids, *J. Appl. Phys.,* Vol 21, 1950, p 437–445

22. R.L. Coble, A Model for Boundary Diffusion Controlled Creep in Polycrystalline Materials, *J. Appl. Phys.,* Vol 34, 1963, p 1679–1682

23. R. Raj and M.F. Ashby, On Grain Boundary Sliding and Diffusional Creep, *Metall. Trans.,* Vol 2, 1971, p 1113–1127

24. M.F. Ashby and R.A. Verrall, Diffusion-Accommodated Flow and Superplasticity, *Acta Metall.,* Vol 21, 1973, p 149

25. F.A. Mohamed and T.G. Langdon, *Acta Metall.,* Vol 32, 1975, p 697

26. C.I. Smith, B. Norgate, and N. Ridley, *Met. Sci.,* May 1976, p 182

27. J.R. Spingarn and W.D. Nix, *Acta Metall.,* Vol 26, 1978, p 1389

28. A. Ball and M. Hutchison, Superplasticity in the Aluminum-Zinc Eutectoid, *J. Met. Sci.,* Vol 3, 1969, p 1

29. A.K. Mukherjee, The Rate-Controlling Mechanism in Superplasticity, *Mater. Sci. Eng.,* Vol 8, 1971, p 83

30. H. Fukuyo, H.C. Tsai, T. Oyama, and O.D. Sherby, Superplasticity and Newtonian-Viscous Flow in Fine Grained Class-I Solid Solution Alloys, *ISIJ Int.,* Vol 31, 1991, p 76

31. T.G. Langdon, A Unified Approach to Grain Boundary Sliding in Creep and Superplasticity, *Acta Metall. Mater.,* Vol 42, 1994, p 2437

32. W.A. Rachinger, Relative Grain Translations in Plastic Flow of Aluminium, *J. Inst. Met.,* Vol 81, 1952–1953, p 33

33. A.K. Mukherjee, Role of Grain Boundaries in Superplastic Deformation, *Grain Boundaries in Engineering Materials,* J.L. Walter et al., Ed., Claitor Publishing, Baton Rouge, LA, 1975, p 93–105

34. R.C. Gifkins, Grain Boundary Sliding and Its Accommodation during Creep and Superplasticity, *Metall. Trans. A,* Vol 7, 1976, p 1225

35. M.J. Mayo and W.D. Nix, *Acta Met.,* Vol 37, 1989, p 1121

36. A.K. Ghosh, Glide-Climb Based GBS Model Applied to Submicrocrystalline Superplasticity, *Mater. Sci. Forum,* Vol 357–359, 2001, p 237–248

37. H.W. Hayden, S. Floreen, and P.D. Goodwell, Deformation Mechanisms of Superplasticity, *Metall. Trans. A,* Vol 3, 1972, p 833

38. A. Arieli and A.K. Mukherjee, A Model for the Rate-Controlling Mechanism in Superplasticity, *Mater. Sci. Eng. A,* Vol 45, 1980, p 61

39. J. Bardeen and C. Herring, *Imperfections in Nearly Perfect Crystals,* W. Shockley, Ed., John Wiley, New York, 1952, p 279

40. V.N. Perevezentsev, V.V. Rybin, and V.N. Chuvil'Deev, The Theory of Structural Superplasticity – II, *Acta Metall. Mater.,* Vol 40, 1992, p 895–905

41. E.M. Taleff, P.J. Nevland, and S.J. Yoon, The Effect of Ternary Alloying Additions on Solute-Drag Creep in Aluminum-Magnesium Alloys, *Deformation, Processing and Properties of Structural Materials,* E.M. Taleff et al., Ed., TMS, 2000, p 373

42. R.S. Mishra, R.Z. Valiev, S.X. McFadden, R.K. Islamgaliev, and A.K. Mukherjee, High-Strain-Rate Superplasticity from Nanocrystalline Al Alloy 1420 at Low Temperatures, *Philos. Mag.,* Vol 81, 2001, p 37–48

43. R.S. Mishra, T.R. Bieler and A.K. Mukherjee, Superplasticity in Powder Metallurgy Aluminum Alloys and Composites, *Acta Metall. Mater.,* Vol 43, 1995, p 877–891

44. R.S. Mishra, V.V. Stolyarov, C. Echer, R.Z. Valiev, A.K. Mukherjee, Mechanical Behavior and Superplasticity of a Severe Plastic Deformation Processed Nanocrystalline Ti-6Al-4V Alloy, *Mater. Sci. Eng. A,* Vol 298, 2001, p 44

45. P.B. Berbon, N.K. Tsenev, R.Z. Valiev, M. Furukawa, Z. Horita, M. Nemoto, and T.G. Langdon, Fabrication of Bulk Ultrafine-Grained Materials Through Intense Plastic Straining, *Metall. Mater. Trans. A,* Vol 29, 1998, p 2237–2243

46. G.A. Salischev, R.M. Galeyev, S.P. Malisheva, and O.R. Valiakhmetov, Low Temperature Superplasticity of Submicrocrystalline Titanium Alloys, *Mater. Sci. Forum,* Vol 243–245, 1997, p 585

47. M.L. Meier, D.R. Leuser, and A.K. Mukherjee, The Effects of the α/β Phase Proportion on the Superplasticity of Ti-6Al-4V and Iron-Modified Ti-6Al-4V, *Mater. Sci. Eng. A,* Vol 154, 1992, p 165

Electronic Structure Methods Based on Density Functional Theory

Christopher Woodward, Air Force Research Laboratory

ELECTRONIC STRUCTURE METHODS based on density functional theory (DFT) have emerged over the last two decades as a powerful tool for assessing the mechanical thermodynamic, and defect properties of metal alloys. These first-principles methods are very appealing because they are based on the culmination of our understanding of quantum mechanics and the electron-ion many-body problem. While the starting point for such calculations requires only the most basic knowledge of chemistry, physics, and crystalline and defect structures, the calculations can quickly become very computationally challenging, with increasing system size and complexity. Practical application of electronic structure methods invariably includes chemical, spatial, or temporal approximations that can curtail a faithful representation of the actual materials problem. However, over the last decade, there have also been significant advances in methods for calculation-free energies (entropy) (Ref 1), activated states (i.e., kinetics) (Ref 2), flexible boundary conditions (Ref 3), lattice dynamics (Ref 4), and reaction-rate theory (Ref 5). Taken with the rapid improvements in computer processor speeds and the maturation of "easy-to-use" DFT methods, there has been an explosive growth in the use of DFT methods in materials science.

This article is a guide to understanding the origins of these methods and their strengths and limitations. It provides the basic procedures for calculating essential structural properties in metal alloys. Before delving into the details for DFT, it is important to place the method in the context of the larger community of scientists interested in the nature of the electronic state in materials.

History

Modern electronic structure emerged from method development in the chemistry and physics communities and falls roughly into two groups: Hartree Fock and its extensions (HF+E) and density functional theory. Historically, HF+E has been considered to be more precise and has been preferred by physical chemists, because systematic improvements to the original approximation are well defined. For several reasons, HF+E is not well suited for metallic systems, as discussed in detail in the next section. Corrections to HF are extremely computationally intensive, scaling with the fourth to sixth power of the number of electrons. Density functional theory has been widely used in metallic systems since its inception in 1962. With improvements in efficiency (speed) and refinements in the underlying approximations, DFT is increasingly being used in quantum chemistry applications. Recently, researchers have begun to blur the line between these two approaches by constructing novel potentials that blend fundamental aspects of the two theories (Ref 6). The resulting approximations show great promise for calculations over a broad range of problems ranging from atoms and molecules to chemically complex metal-oxide interfaces.

For scientists and engineers considering using electronic structure (ES) methods, navigating the sea of DFT acronyms can be challenging (Table 1). In general, the acronyms refer to the numerical scheme, or basis, used to represent the electrons. More recently, as methods have matured, codes have been named after the groups that developed or support the method. All electronic structure methods must deal with the large changes in the electron distribution observed in atoms, molecules, and solids. Some of the electrons are strongly bound to the nuclear sites (core states) and are very similar to that found in isolated atoms. Other electrons are more weakly bound (valence states), producing interatomic and molecular bonding, and are responsible for most of the electronic, optical, thermodynamic, and chemical properties. Electronic structure methods deal with this disparity in a variety of ways, depending on what class of materials problem is under consideration. For example, with isolated atoms and molecules, it is natural to work in real space with methods based on a linear combination of atomic orbitals (LCAO), represented numerically or as a sum of analytic functions (e.g., Gaussians). For crystalline systems, it is more natural to use periodic boundary conditions, and electrons are represented using a linear combination, or basis, of plane waves. Over time, several methods were developed to avoid the large number of plane waves needed to represent the rapidly varying electron core densities. One approach, employed in augmented plane wave (APW) and muffin tin orbital (MTO) methods, is to use a set of local functions centered around each atom and to match that solution on a sphere to a plane wave basis everywhere else. Another technique, the pseudopotential method, maps the strongly bound electron states into a potential that is then used to calculate the valence electrons. These pseudopotential plane wave (PPW) methods are relatively easy to use, and the simplicity of the basis has allowed significant progress in computational efficiency

Table 1   Selected acronyms appearing in this article

Acronym	Meaning
APW	Augmented plane wave
CALPHAD	CALculation of PHAse Diagram methods
CP	Car and Parrinello methods
DFT	Density functional theory
E(ρ)	Functional (of electron density function ρ)
ES	Electronic structure
FLAPW	Full-potential linearized augmented plane wave
FP-LMTO	Full-potential linearized muffin tin orbital
GGA	Generalized gradient approximation
HF	Hartree-Fock
HF+E	Hartree-Fock and its extensions
KS	Kohn and Sham
LAPW	Linearized augmented plane wave
LCAO	Linear combination of atomic orbitals
LDA	Local density approximation
LSDA	Local spin density approximation
MES	Murnaghan's equation of state
MTO	Muffin tin orbital
OPW	Orthogonalized plane wave
PAW	Projector augmented wave
PPW	Pseudopotential plane wave
PW91	Perdew and Wang 1991 (Ref 7, 8)
POSCAR	Input file for Vienna ab initio simulation package
SGGA	Spin-generalized gradient approximation
USPP	Ultrasoft pseudopotential
VASP	Vienna ab initio simulation package

(i.e., parallel processing) and analytic solutions for properties such as atomic forces and stress.

Historically, mixed-basis methods (APW, MTO) have been considered as the benchmark for accuracy in most applications. However, the mixed basis makes these methods more challenging to use and adds significant complications to deriving basic quantities such as atomic forces or stresses. With advances in pseudopotential theory since the mid-1980s, PPW methods routinely reproduce the results of mixed-basis methods. Also, because of the added benefits of larger simulation sizes and automated atomic and cell optimization, it can be argued that PPW methods are producing more accurate results in a wider range of applications.

Mixed-basis methods are still the preferred technique for systems where the pseudopotential approximation breaks down. This happens when changes in valence electrons change the structure of the core electrons. For example, in the actinides, where the f-electrons are coupled to the core states, APW and MTO methods are preferred (Ref 9). Also, simulations of photo absorption and emission are probably best modeled using techniques where the core states are optimized along with the valence states. The quality and availability of pseudopotentials in some software packages can be quite limited. Researchers new to the field should carefully assess the available options before investing time or resources in any particular method.

Scale. One additional criterion to consider is the scale of the problem that needs to be solved. Inevitably, this dictates the method and the required computational platform. Both LCAO and PPW methods scale well on current parallel supercomputers and are typically applied to molecular and crystalline problems, respectively.

Density functional theory is being applied throughout the scientific community to a staggering range of problems. With current multiprocessor supercomputers, PPW methods can simulate system sizes up to approximately 1000 atoms (Ref 10). This varies with the system symmetry and the choice of atomic species, with the transition metals being the most challenging. Researchers are also running ab initio molecular dynamics simulations for cells ranging from 100 to 500 atoms for simulation times up to tens of picoseconds. The future of DFT will be driven by improvements to the underlying approximations, the introduction of new hybrid potentials, and advances in supplementary methods that use DFT results. Research into new novel basis functions (e.g., wavelets) or the introduction of new computer hardware (e.g., field-programmable gate arrays) could revolutionize the field. Finally, while still in its infancy, there is a significant effort underway to directly calculate the electronic state by quantum Monte Carlo methods. If properly coupled to next-generation supercomputers, this eventually could overtake all other developments (Ref 11).

The rest of this article is divided into three independent sections. The first reviews the general underlying theory of electronic structure methods, DFT specifically, and the taxonomy of DFT methods that have emerged over the last 30 years. The second section reviews the approximations and computational details of the most popular method used in metal systems, the PPW methods. The last section reviews a subset of the applications of DFT methods found in metal alloy systems. This includes calculations of a variety of structural, thermodynamic, and defect properties, with particular emphasis on structural metal alloys and their derivatives.

Fundamentals of Density Functional Theory

One would like to model a chunk of matter using only what is known about coulomb interactions between electrons and ions and the underlying principles of quantum mechanics. The approach taken over the last 50 years has been to systematically apply approximations, making the many-body problem more manageable while retaining the essential physics. Part of this evolving approach is to reduce the systems of equations to that subset which captures the problem of interest. We are not interested in solving systems with Avogadro's number of particles; not only would solving such a problem be unfeasible, analyzing the results of such a calculation would be a Herculean task. Therefore, for practical reasons both conceptual and computational, it is considered best practice to minimize the scale (spatial and temporal) of the electronic structure calculation. There are many good reviews of DFT methods. For a general overview of the fundamentals, see Ref 12. The general theory of PPW methods and many practical issues on applying these methods are found in Ref 13; many practical details of PPW and APW methods are reviewed in Ref 14.

Classical to Quantum Mechanics. Beginning from classical mechanics, the many-body Hamiltonian of an ensemble of interacting atoms takes the form:

$$H_{\text{Total}} = \sum_I^M \frac{P_I^2}{2M_I} + \frac{1}{2}\sum_I^M \sum_{I\neq J}^M \frac{Z_I Z_J e^2}{|R_I - R_J|} + \sum_j^m \frac{p_j^2}{2m_e}$$
$$+ \frac{1}{2}\sum_i^m \sum_{j\neq i}^m \frac{e^2}{|r_i - r_j|} - \sum_I^M \sum_i^m \frac{Z_I e^2}{|r_i - R_I|}$$

(Eq 1)

where M_I, P_I, Z_I, and R_I are the mass, momentum, charge, and position of the M possible ions, and m_e, p_e, and r_i are the mass, momentum, and position of the m possible electrons. The Hamiltonian is then separated into two parts: a purely ionic part (the first two terms on the right side of Eq 1) and an ion-electron part:

$$H_{\text{Total}} = H_{\text{Ion}}(R_I) + H_{e-\text{Ion}}(R_I, r_i)$$

(Eq 2)

From this point, a materials scientist can choose to represent $H_{e-\text{Ion}}$ by an effective potential that leads to the field of atomistic modeling or choose to invoke quantum mechanics and solve the many-body Schrödinger equation that leads to the field of ES methods. The latter is generally considered a more faithful representation of the many-body problem, because the electrons are treated explicitly. Standard practice now is to apply the Born-Oppenheimer approximation. Given that the electron cloud responds much faster to an applied electric or magnetic field than the ions ($m_e/M_I \ll 1$), we can decouple the nuclear and electronic motion and solve for the electron degrees of freedom with the ionic positions held fixed. Using separation of variables, the Schrödinger equation corresponding to Eq 1 can be divided into two parts:

$$H\Psi(\{R_I, r_i\}) = (H_e + H_{\text{Ion}})\Psi(\{R_I, r_i\})$$
$$= E\Psi(\{R_I, r_i\})$$

(Eq 3)

using $\Psi(\{R_I, r_i\}) = \Psi_e(\{R_I, r_i\})\Psi_{\text{Ion}}(\{R_I\})$ produces two equations:

$$(T_e + V_{e-e} + V_{e-I})\Psi_e(\{R_I, r_i\}) = E_e\Psi_e(\{R_I, r_i\})$$
and $(T_I + V_{I-I} + E_e)\Psi_{\text{Ion}}(\{R_I\}) = E_I\Psi_{\text{Ion}}(\{R_I\})$

where T and V refer to the kinetic and coulomb potential terms for the electrons and ions, and the eigenvalues (E_e coming from the separation of variables) are incorporated as an effective potential for the ionic problem. The many-body Schrödinger equation for the m electrons is:

$$H_e\Psi_e(\{R_I, r_i\}) = \sum_{i=1}^m \left(\frac{-\hbar^2}{2m}\nabla_i^2 - Ze^2 \sum_I^M \frac{1}{|r_i - R_I|_e} \right.$$
$$\left. + \frac{1}{2}\sum_{i\neq j}^m \frac{e^2}{|r_i - r_j|} \right)\Psi_e(\{R_I, r_i\})$$
$$= E_e\Psi_e(\{R_I, r_i\})$$

Unfortunately, this equation cannot be solved directly. Two approaches are taken to solve this system of equations: the first, Hartree-Fock and its extensions (HF+E), solve for the electron wave-functions; and the second, DFT, solves for the charge density. (Drs. W. Kohn and J.A. Pople split the Nobel Prize in Chemistry in 1998 for aspects of these contributions.) The Hartree-Fock approach is attractive because the derivation allows for well-defined systematic (although costly) improvements to the initial approximation for the third term in Eq 2. These are sometimes called post-Hartree Fock methods.

Limitations for Metallic Systems. The HF+E methods are used extensively in nonmetallic systems; however, they are poorly suited for metal systems for several reasons. First, significant corrections to the initial HF approximation are required to properly represent metal systems. Second, in free-electron metals, HF produces an intrinsic instability in the electron velocity (a logarithmic divergence in $d\varepsilon/dk$), where $\varepsilon(k)$ is the energy dependence of the electron as a function of wave vector (k) at the Fermi surface (Ref 15).

Density functional theory is based on two insights provided by Hohenberg, Kohn, and Sham

in the early 1960's (Ref 16, 17). First, Hohenberg and Kohn proved that for an interacting electron gas in an external potential the ground state is given exactly as a functional of the electronic density: $E = E(\rho(r))$. Here $\rho(r)$ is the electron density and for our many body problem the external potential is the coulomb potential produced by the atomic nuclei. Also, a functional is defined as a function of a function—in this case E is a function of the electron density. While the Hohenberg-Kohn theorem shows that $E(\rho)$ is a unique functional, it does not provide a prescription on how to form the functional, so the usefulness of the theorem is dependent on finding sufficiently accurate approximations (Ref 16).

To this end, Kohn and Sham suggested writing E as:

$$E[\rho] = T_s[\rho] + E_{ei}[\rho] + E_{II}[\rho] + E_H[\rho] + E_{xc}[\rho]$$
(Eq 4)

The functionals on the right represent the kinetic energy of a system of noninteracting electrons (T_s), the electron-ion interactions (E_{ei}), the Hartree potential of electron-electron interactions (E_H), the ion-ion interactions (E_{II}), and the exchange correlation functional (E_{xc}), respectively (Ref 17). The functionals are now integrals over space; that is, the last two terms explicitly are:

$$E_H[\rho] = \frac{e^2}{2} \int\int dr\, dr' \frac{\rho(r)\rho(r')}{|r - r'|}$$
(Eq 5)

$$E_{xc}[\rho] = \int dr\, \rho(r)\varepsilon_{xc}(\rho(r))$$
(Eq 6)

In the last term, Kohn and Sham identified $\varepsilon_{xc}(\rho(r))$ as the exchange and correlation energy/electron of a uniform electron gas of density ρ. This is the local density approximation (LDA), which assumes that given a sufficiently slowly varying density, a function (ε_{xc}), can be defined that represents the effective potential of an electron surrounded by its own "mutual exclusion zone" consistent with the requirements of quantum mechanics.

Using the fact that the functional $E[\rho]$ is an energy minimum with respect to variations in ρ (the Hohenberg-Kohn theorem), they then derived single-particle Schrodinger-like equations that are sometimes referred to as the Kohn-Sham (KS) equations:

$$\left\{ -\frac{\hbar^2}{2m}\nabla^2 - Ze^2 \sum_{I=1}^{N} \frac{1}{|r - R_I|} + e^2 \int \frac{\rho(r')}{|r - r'|} dr' + \mu_{xc}(\rho(r)) \right\} \phi_i(r) = \varepsilon_i \phi_i(r)$$
(Eq 7)

where:

$$\rho(r) = \sum_{i=1}^{m} |\phi_i(r)|^2$$

with m equal to the number of occupied states (the number of electrons in the system), and

$\mu_{xc}(\rho) = d(\rho\varepsilon_{xc}(\rho))/d\rho$, is identified as the exchange-correlation contribution to the chemical potential of a uniform gas of density ρ. The system of equations is solved self-consistently by assuming a $\rho(r)$, constructing the last two terms in the KS equations, and then solving for $\rho(r)$ using $\phi_i(r)$. The total energy is given by:

$$E[\rho(r)] = \sum_{i=1}^{m} \varepsilon_i - \frac{1}{2}\int\int \frac{\rho(r)\rho(r')}{|r - r'|} dr dr' + \int \rho(r)[\varepsilon_{xc}(\rho(r)) - \mu_{xc}(\rho(r))] dr$$
(Eq 8)

The KS equation maps the interacting many-electron system to a set of noninteracting electrons moving in an effective potential of all the other electrons. The utility of the KS equations rests in the ability to find reasonable approximations for the functional $E_{xc}[\rho]$. Fortunately, this function has been studied in detail for the case of a uniform electron gas (Ref 18), derived using Monte Carlo techniques (Ref 19), and parameterized for electronic structure calculations (Ref 20).

The local density approximation has been surprisingly successful in predicting a variety of properties in metals. Lattice parameters are usually accurate to within ~1% and cohesive energies and elastic constants to within ~10%. The method is well suited for studying solids and perfect and defected crystals and is easily extended to include spin dependence, the local spin density approximation (LSDA), which has been widely applied to ferromagnetic and antiferromagnetic systems (Ref 21). The LDA is also the starting point for a variety of improvements based on the local change in the electron density produced by the electron (the exchange-correlation hole). These generalized gradient approximations (GGA) have systematically improved the accuracy of DFT for problems in molecular systems and broadened the application base significantly (these are reviewed in the next section). However, currently the method is still not well suited for systems with large van der Waals interactions or systems sampling infinitesimally small electron densities, such as structures bounded by a vacuum.

Where DFT methods seem to diverge is in the spatial representation of the one-electron wave functions. The wide variety of methods reflects the fact that an accurate representation of the charge density has traditionally required specialized basis functions. The fundamental problem is that as the atomic number increases, additional atomic wave functions are required by the Pauli exclusion principle to be orthogonal to existing lower-lying wave functions. To accomplish this, as the principle quantum number increases, the wave functions take on a rapidly varying radial form near the atomic center. Therefore, for a set of basis functions to accurately describe the electrons, it must be able to both represent the rapidly varying function near atomic centers and the relatively smooth functional form outside that region.

Several strategies have been used to solve this problem. Techniques such as the augmented plane wave, muffin tin orbital, and Korringa-Kohn-Rostoker methods use an efficient and compact basis to describe the wave functions near the atomic centers (Ref 22–27). An additional basis (typically plane waves) is used to describe the wave functions outside this region, and various schemes are used to ensure a proper match of the wave functions at the boundary between these two regions. These methods give a very accurate representation of the core region, while allowing some flexibility in the basis outside this region. Refinements to these techniques, such as the full-potential linearized augmented plane wave (FLAPW) (Ref 28, 29) and full-potential linearized muffin tin orbital (FP-LMTO) (Ref 23, 30), are currently considered to be the most accurate DFT methods, although the implementations are limited to relatively small cell sizes.

The seminal work on orthogonal plane-wave methods led to the development of an alternative technique where the low-lying (core) states are effectively removed from the calculation (Ref 31). In this case, an effective electron-ion potential, or pseudopotential, is derived from an atomic electronic structure calculation. Pseudopotentials incorporate the tightly bound wave functions and ionic charge so that the potential produces the same electronic interactions as the original atomic calculation. In this way, the core electrons, which do not normally influence materials properties, can be removed from the simulation. By construction, the pseudopotentials produce the same interaction with the valence electrons as the original all-electron calculation (as measured through the electron-scattering properties) (Ref 32). Recent pseudopotential schemes have relaxed this philosophy during the construction of the potential, only to reimpose the requirement when constructing the KS orbitals in the system of interest (Ref 33, Ref 34–35). Pseudopotentials also incorporate the effective potential produced by the Pauli exclusion principle, such that the valence wave functions are smooth functions in all space. Therefore, it is natural to combine a plane-wave basis with the pseudopotential representation in what is now called pseudopotential plane-wave (PPW) methods.

While the plane-wave basis is not as compact as that used in the linearized augmented plane-wave methods, PPW methods have been easier to implement because of the simplicity of the plane-wave basis. It is relatively straightforward to calculate atomic forces (through the Hellmann-Feynman theorem), the stress tensor, and phonon properties (Ref 36–38). However, until the mid-1980s, application of PPW methods was somewhat limited because of the size of the required plane-wave basis set. In 1985, Car and Parrinello showed how to simultaneously optimize both the electronic and ionic degrees of freedom by taking advantage of fast Fourier transforms and the plane-wave basis (Ref 39). Iterative diagonalization methods that

have grown out of these insights have shown that by proper preconditioning of the KS wave functions during the optimization procedure, it is possible to directly minimize the KS energy functional (Ref 40). This innovation stabilizes the constrained optimization of Eq 1 and has made it possible to run very large simulations relatively efficiently (Ref 13).

The schematic in Fig. 1 shows the methods developed by the chemistry and physics communities to solve the materials many-body problem. In general, the DFT methods are the method of choice for calculations on metallic systems.

Pertinent Approximations and Computational Details for Calculations in Metal Alloys

To calculate the electronic structure of an alloy, the researcher needs more than just an underlying theory and a working description (basis) of the electrons. First, the methods need efficient and accurate techniques for integrating the quantities described in the last section over the volumes (and k-space) of interest. Second, the approximations and basis must be well matched to the problem of interest. Finally, explicit knowledge of the crystal structure, location, and species of every atom on the simulation volume is required. Fortunately, for most of the alloys of engineering interest, this information is available in tabulations of the *International Tables for Crystallography* in the form of space groups, describing the symmetry of the lattice, and Wyckoff positions, describing the atomic sites (position and chemistry) (Ref 41).

Integration of Cell Quantities. One advantage to working on metallic systems is that the underlying crystalline structure is almost invariably periodic. This allows the researcher to employ simulation cells with periodic boundary conditions (supercells) to represent the material of interest. Periodic boundary conditions also make it possible to represent almost all quantities of interest in terms of real and Fourier space (k-space) components. This is particularly useful when summing up terms numerically in the KS equations or Eq 8 over the Brillouin zone (energy states for the free electrons in a metal, as described by the use of the band, or zone, theory of electron structure).

The most straightforward way to integrate quantities over the supercell is to divide the Brillouin zone using a tetrahedron grid of points (k-points). However, in the early 1970s, numerical schemes were developed to predict the smallest set of k-points, "special k-point", that would yield the most accurate cell integrations (Ref 42, 43). In most modern application codes, the selection of k-points is sufficiently automated that the investigator needs only to input the required density of such points. However, one characteristic of a metal is that valence states, or bands, are partially occupied, and this can produce numerical instabilities. To avoid this and to improve the efficiency of the cell integration, researchers introduced a numerical smearing of the highest-lying bands, specifically those that are near the Fermi surface. The "broadening" methods fall into two classes: those employing ad hoc functional forms such as Gaussians and finite-temperature schemes based on the Fermi-Dirac or Gausslike functions that mimic thermal broadening (Ref 44–46). Both methods are effective, and the latter technique has the advantage of an associated, if ad hoc, temperature.

Computational time for metal simulations scales with the number of k-points. However, as the simulation sizes get larger and reciprocal space gets smaller, the number of required k-points is reduced. For large simulations, say greater than 100 atoms, often only a single k-point is needed. Also, if only the gamma point (k = (0,0,0)) is required, then calculations can gain another factor of 2 in efficiency, because of the symmetry imposed on the complex parts of the wave functions.

While there are rules of thumb for the use of special k-points and broadening methods, it pays to carefully test the convergence of such
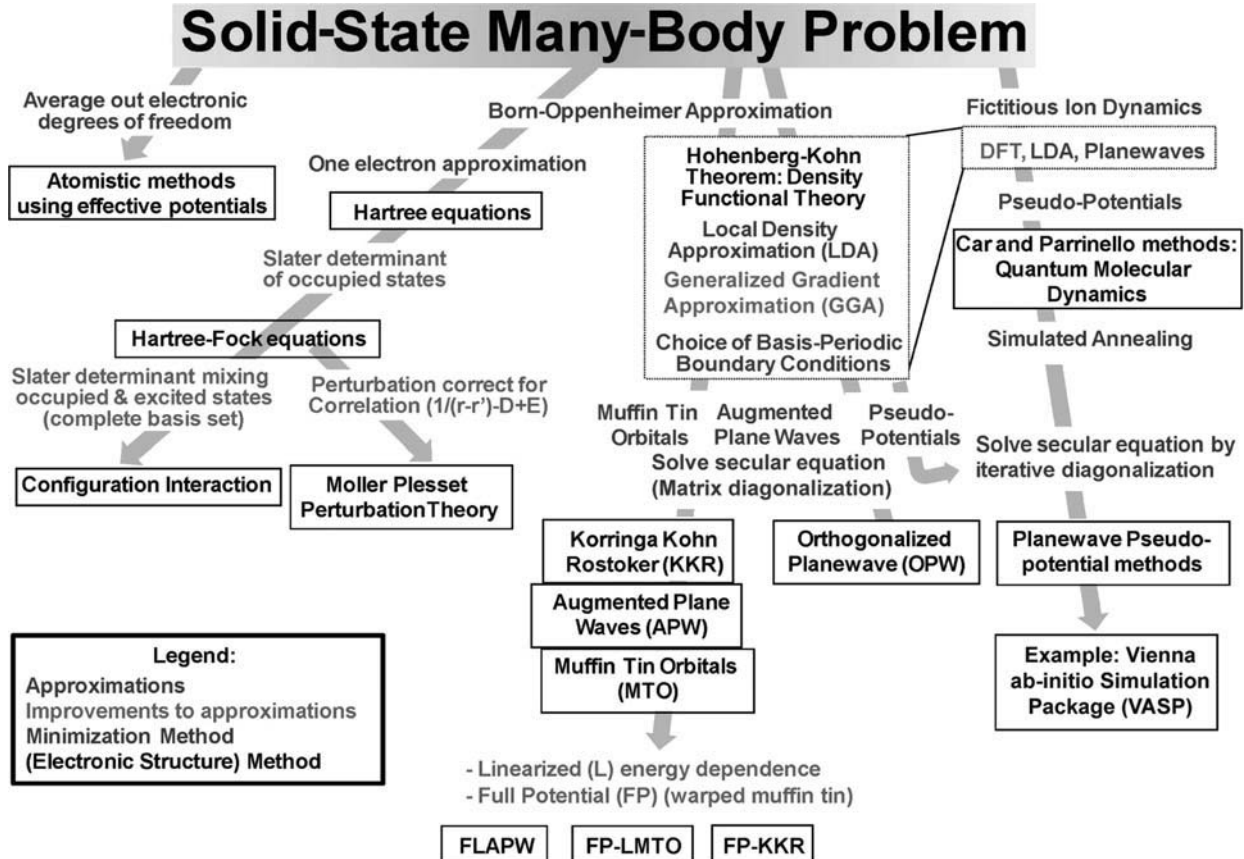
Fig. 1 Schematic of the methods developed by the chemistry and physics communities to solve the materials many-body problem. In general, the DFT methods are the method of choice for calculations on metallic systems.

methods using, for example, the cell energy or other quantity of interest. In general, the density of k-points should be approximately that of the size in dispersion in the bands near the Fermi surface (Ref 14).

Understanding and Choosing a Pseudopotential. Early in the development of electronic structure methods, researchers realized that the electrons contained in full atomic shells (s, p, and d) do not have a strong influence on chemical and mechanical properties. These effects are controlled mostly by the interaction of valence electrons, which have the largest principle quantum number and thus the most rapidly varying radial function in the region around the atom nucleus. Pseudopotentials replace the core-valence electron interactions, the second term of Eq 4 and 7, with an effective potential, producing a realistic pseudovalence wave function that has a smooth and slowly varying radial form. Typically, pseudopotentials are derived from an atomic reference calculation and then used in a crystalline or other environment, so transferability is a serious concern when developing such a scheme. Early local pseudopotentials, and more recent implementations of the same, are severely limited and can only reliably be used in simple free-electron metals where the core electrons have very weak interactions with the valence states (Ref 47, 48).

Modern pseudopotential theory is based on a "norm-conservation" approach that enforces a strict criterion for mapping real- to pseudo-wave functions and includes nonlocal angular momentum-dependent (l) interactions that accurately model the valence-core electron interactions (Ref 32). Transferability is maintained by imposing identical logarithmic derivatives, and thus scattering phase shifts, outside a certain (core) radius about each atomic site. More recent advances in pseudopotential theory, such as Vanderbilt's ultrasoft pseudopotentials, make use of additional functions about the atomic core, which allow for smoother pseudo-wave functions and more efficient PPW calculations (Ref 33). This and later refinements of pseudopotential theory parallel the original strategies used in orthogonalized plane wave (OPW) methods developed by Herring, Callaway, and others from 1940 through the 50's (Ref 49, 50). The most recent advances in pseudopotential theory, the projector augmented wave (PAW) method (Ref 34, 35), retains all the information of the core states and is thus analogous to the most accurate all-electron methods (e.g., OPW, APW and MTO). Implementing the PAW methods in PPW codes required additional development, and using the potentials incurs additional computational overhead.

Using modern numerical methods, current commercial PPW implementations of the PAW method are as efficient as the original Car and Parrinello methods and as accurate as many full-potential methods. They have the added benefit of ease of calculation of atomic forces,

stress tensor, and convergence of basis. A wide variety of ultrasoft and PAW pseudopotentials are available in the user community as well as source code for developing such potentials. More importantly, there are readily available, well-documented suites of pseudopotentials that have been tested by a broad user base.

Exchange-Correlation Potentials, Local Density Approximation, and the Generalized Gradient Approximation. Density functional theory is one of the most successful electronic structure methods precisely because of the simplicity of the underlying exchange-correlation functional. Practical application of the Hohenberg-Kohn theorem through the KS equations requires both an assessment of the exchange correlation functional and a numerically efficient scheme for interpolating the energy for a range of charge densities. Since the original formulation of the KS equations, the nature of the exchange correlation potential has been studied in some detail. The LDA is the foundation of all these approximations. Within the LDA, only knowledge of the exchange correlation energy of the homogeneous electron gas is required. This is approximated as the sum of exchange and correlation potentials, the first given by a basic analytical form and the second calculated using Monte Carlo methods (Ref 19). These data were then fit to functional forms to improve computational efficiency and parameterized for electronic structure calculations (Ref 20).

The LDA has been found to be surprisingly accurate in a wide variety of systems. The initial formulation was expected only to be valid for volumes with slowly varying electron densities, a condition that is not well satisfied in many crystals. It is generally believed that the LDA approximation underestimates the exchange energy (by ~10%) and overestimates the correlations energy (2%) and that these errors partially cancel each other out (Ref 51). However, LDA is particularly unsatisfactory for low-electron densities, such as near a surface, and that has made the approximation problematic for calculations of atoms and molecules. Still, LDA produces reasonably accurate bond lengths and geometries for some molecules.

The efficacy of the KS equations and the need for highly accurate simulations has resulted in systemic improvements to the LDA. The most successful approaches, based on Generalized Gradient Approximation (GGA), include information on the effects of inhomogeneities in the electron gas on the exchange correlation potential. The gradient corrections are constructed to satisfy intrinsic sum rules and are designed to maintain the accuracy of LDA while correcting the errors introduced by large gradients. Using FP-LMTO, Ozolins and Korling calculated the changes in lattice constants and bulk modulus produced by using the GGA proposed by Perdew and Zhang (sometimes referred to as PW91) (Ref 7, 8). They found a systematic improvement in equilibrium volumes and bulk modulus for 3d, 4d, and 5d transition metals, with the

mean error decreasing on average by 50% for both quantities across the series (Ref 52). Other researchers also found that early GGA methods and PW91 correctly produce the correct body-centered cubic ground state for crystalline iron where the LSDA erroneously predicts a face-centered cubic (fcc) ground state (Ref 53, 54).

Recently, other GGA methods have been validated that produce better energetics and better represented low-density regions. Hybrid schemes based on a weighted mixing of HF exchange and DFT correlation have gained favor in the quantum chemistry community (Ref 55, 56). Also, another class of GGA functional has been self-consistently matched to high- and low-electron densities, making it efficient and well suited for metallic systems with internal or external surfaces (Ref 57).

Practical Application of DFT in Metals and Alloys

As illustrated in Fig. 1, DFT methods come in a variety of forms; they also vary widely in their level of maturity and efficacy. Until recently, most of the mixed-basis methods (FP-LMTO and FLAPW) were closely held academic codes; now, there are several freeware and commercial options (Table 2). While highly accurate, these methods have a more complex set of adjustable basis parameters than PPW methods. There are a variety of PPW methods, some available as freeware and some commercially, and some have well-developed pseudopotential libraries. New users should verify that a given code base has the necessary features for running their application before investing resources into a method. All of the methods shown in Table 2 should produce accurate results for the simple applications outlined in this section.

From Crystal Structure to Input File—Examples of VASP Input Files. Setting up the simulation cell for electronic structure calculations requires the lattice vectors, the atomic positions, and the chemical species at each site. The input file required for the PPW method Vienna ab initio simulation package (VASP) is used to illustrate this process (Ref 59, 60). Typically, the researcher starts with a phase and then refers to tables and textbooks to find the required quantities. For example, in nickel-base superalloys, Ni_3Al is an important precipitate that significantly strengthens the nickel matrix phase. Using the tabulated data on alloy phases in W.B. Pearson's *A Handbook of Lattice Spacings and Structures of Metals* (Ref 61, 62), the structure type is listed as having a structure name (a representative material) $AuCu_3$, the Strukturbericht designation $L1_2$, a lattice parameter of 3.567 Å, with a space group of Pm3m. The Pearson classification for $L1_2$, $AuCu_3$ is given as cP4 in the tables leading up to the "Table of Classification of Structures of Metals and Alloys." A shortened version of

Table 2 Partial list of currently available density functional theory programs

Method	Acronym/name	Fee	Point of contact
PPW	ABINIT	No	http://www.abinit.org/ Prof. X. Gonze, Université catholique de Louvain, Physico-Chemistry and Physics of Materials, Louvain-la-Neuve, Belgium
	CASTEP	Yes	Accelrys, Inc., San Diego, CA, 92121 http://accelrys.com/products/materials-studio
	Quantum Expresso	No	P. Giannozzi, Universit'a di Udine and Democritos National Simulation Center, Italy http://www.quantum-espresso.org
	VASP: Vienna Ab initio Simulation Package	Yes	Prof. J. Hafner, Institute of Materialphysik Wien University Austria, http://cms.mpi.univie.ac.at/vasp/
LCAO	DMOL3	Yes	Accelrys, Inc., San Diego, CA, 92121 http://accelrys.com/products/materials-studio
	SIESTA	Yes	Prof. J.A. Torres, Universidad Autonoma de Madrid, Spain http://www.icmab.es/siesta or http://www.nanotec.es/
FP-LMTO	LmtART	No	http://www.fkf.mpg.de/andersen/ S.Y. Savrasov (Ref 57).
	RSPt : Relativistic Spin Polarised (test)	No	http://www.rspt.net/index.php
FP-LAPW	WIEN2K	Yes	http://www.wien2k.at/index.html Prof. Karlheinz Schwarz, Inst. f. Materials Chemistry, TU Vienna
	FLAIR		Prof. M. Wienert, Univ. Wisconsin Milwaukee, weinert@uwm.edu
	QMD-FLAPW	Yes	Prof. A.J. Freeman, Northwestern Univ. Quantum Materials Design, Inc. http://flapw.com/news.html

Some are available for a license fee and others are available at no cost. Many of the methods have associated user groups, and some have graphical user interfaces. Materials science trade organizations are beginning to track the status of software (see, for example, http://iweb.tms.org/forum) and may provide useful updates to this table.

Table 3 Crystallographic information and atomic basis for L1₂, Ni₃Al

Classification symbol	Structure name	Strukturbericht type	Space group
cP4	AuCu₃	L1₂	Pm3m

Origin at center (m3m)
Equivalent positions:
Au: 1 a m3m 0,0,0
Cu: 3 c 4/mmm. 0,½,½; ½,0,½; ½,½,0

```
Example: Ni3Al                Title
3.57640                       lattice Constant (Ang.)
  1.00000   0.00000   0.00000 Lattice vector
  0.00000   1.00000   0.00000 Lattice vector
  0.00000   0.00000   1.00000 Lattice vector
3 1                           Number of each atomic species
Direct                        Format of the atomic positions
  0.000   0.500   0.500       Internal coordinates of atoms
  0.500   0.000   0.500
  0.500   0.500   0.000
  0.000   0.000   0.000
```

Fig. 2 Screen shot of the input file describing crystal system and atomic basis for γ'-Ni₃Al

the entry under cP4 is given in Table 3. After the first line describing the structure designations, the atomic basis and the Wyckoff positions are listed in terms of the atom type, the number of atoms at each symmetry distinct point, and the internal coordinate. All the required information is now determined. The *International Tables for X-Ray Crystallography* has more information for this cubic space group (number 221) and lists much more complicated crystal structures with this space group (Ref 41, 63).

The VASP input file for the lattice vectors and atomic basis is called POSCAR, and a screen shot of the POSCAR file for Ni₃Al is shown in Fig. 2. In POSCAR, the title line is followed by the lattice constant from Pearson and the three lattice vectors for a cubic lattice in Cartesian coordinates. This is followed by the number of each chemical species, in this case, three nickel atoms and one aluminum atom, and a keyword describing the format of the atomic basis. (Note that the order of the chemical species is important and must be consistent with the ordering of the pseudopotential input file.) The atomic basis can be entered in either Cartesian (Keyword: CARTESIAN) or in terms of the three lattice vectors (Keyword: DIRECT). For the "Direct" mode, the atomic positions correspond to $\vec{R} = x_1\vec{a_1} + x_2\vec{a_2} + x_3\vec{a_3}$, where $\vec{a_i}$ are the lattice vectors scaled by the lattice parameter, and x_i are the values entered into POSCAR. Input using the CARTESIAN keyword is scaled only by the lattice constant, $\vec{R} = x_1\hat{i} + x_2\hat{j} + x_3\hat{k}$, where i, j, and k are unit vectors [100], [010], and [001], respectively.

Another, more complex example of an atomic basis is the δ-MoNi phase, which is also important in the nickel-base superalloys. In Pearson, this topologically close-packed phase is listed as orthorhombic with lattice constants $a = 9.108$, $b = 9.108$, and $c = 8.852$ Å, containing 56 atoms with the space group $P2_12_12_1$. Table 4 gives the representative atomic positions listed as 14 roman numbers with corresponding coordinates. The Wyckoff positions, listed in the international tables for $P2_12_12_1$, are shown in Table 5. The atomic coordinates are generated by using the last column of Table 5 with each of the 14 atomic parameters, producing the expected 56 atoms.

Note, however, the atomic species for each site is still unknown. Going back to the original reference for this crystallographic assessment, we can find the chemical assignments for most but not all the atomic sites (Ref 64). At finite temperatures, all alloys show deviations from perfect ordering; however, the composition of the studied δ-MoNi phase was Mo49.2-Ni, and x-ray analysis could not unambiguously determine the chemistry on at least one of the sites. The electronic structure calculations can proceed by assuming a Mo50-Ni composition and the chemistry at the sites designated by XI as being occupied by nickel atoms. However, this is just one possibility, and in principle, a free-energy model would include sampling the formation energy of other atomic arrangements at this composition.

Using Tables 4 and 5, the initial cell was constructed as shown in the screen shot of the POSCAR file in Fig. 3. The figure also shows a screen shot of the final cell configuration. Using VASP, the lattice vectors and atomic positions were optimized within a spin-polarized (ferromagnetic) ultrasoft pseudopotential approximation. Note the slight change in the length of the orthorhombic lattice vectors and the change in atomic positions. At the start of the calculation, the pressure was ~16 kB, and the largest force/atom was ~0.5 eV/10⁻¹⁰ m. After optimization, the pressure was less than 0.5 kB, the atomic forces were less than 2×10^{-3} eV/10⁻¹⁰ m, and the total energy change from the initial configuration was ~0.5 eV.

This and several other configurations were used to develop a simple free-energy model of the nickel-molybdenum system by approximating the configuration entropy (Ref 65).

Lattice Parameters. While current PPW codes can optimize supercell geometries by minimizing in the diagonal components of the stress tensor, it is still useful to know how to calculate the lattice parameters using equations of state. Perhaps the most cited equation of

Table 4   Crystallographic information and atomic parameters for δ-MoNi

Phase	System	Strukturbericht type	Space group					Estimated %Mo
δ-MoNi	Orthorhombic	None	P2₁2₁2₁	Atoms	Atomic parameters			
				IV	0.4519	0.1153	0.5322	0
				VI	0.4424	0.3662	0.5972	0
				VIII	0.3882	0.0523	0.2748	0
				IX	0.1337	0.0707	0.2157	0
				X	0.3768	0.4358	0.8567	0
				XII	0.0680	0.1442	0.9529	0
				I	0.1763	0.4832	0.6425	80
				XIII	0.0338	0.3398	0.1807	80
				II	0.2289	0.2865	0.4098	91
				V	0.2648	0.1993	0.7486	91
				XI	0.3136	0.2464	0.0740	58
				VII	0.0029	0.1969	0.6767	100
				XIV	0.1885	0.0157	0.4960	100
				III	0.1031	0.4192	0.9133	100

Source: Ref 61; %Mo estimate, Ref 64

Table 5   Wyckoff positions for space group P2₁2₁2₁

Multiplicity	Wyckoff letter	Site symmetry	Coordinates
4	a	1	$(x,y,z), (-x+\frac{1}{2},-y,z+\frac{1}{2})$, $(-x,y+\frac{1}{2},-z+\frac{1}{2})$, $(x+\frac{1}{2},-y+\frac{1}{2},-z)$

state used for this purpose was developed by F. D. Murnaghan in 1944 (Ref 66). Assuming that the bulk modulus (K) is a linear function of the pressure:

$$K = -V\frac{dP}{dV} = C(1 + kP) \tag{Eq 9}$$

such that the bulk modulus (K) and its derivative at zero pressure are identified, respectively, as:

$$K(V_0) = -\left(V\frac{dP}{dV}\right)_{p=0}$$
$$= C \text{ and } \left(\frac{dK}{dV}\right)_{p=0} = -\frac{d}{dP}\left(V\frac{dP}{dV}\right)_{p=0} = Ck$$

Integrating Eq 1 from zero pressure yields:

$$P(V) = \frac{C}{Ck}\left(\left(\frac{V_0}{V}\right)^{Ck} - 1\right)$$

Using $P = -\frac{dE}{dV}$, we can identify $K = V\frac{d^2E}{dV^2}$ and finally:

$$V(d^\uparrow 2 E(V))/[dV]^\uparrow 2 = C(1 + kP)$$
$$= C(1 + kC/Ck((V_\downarrow 0/V)^\uparrow Ck - 1))$$
$$= C(V_\downarrow 0/V)^\uparrow Ck \tag{Eq 10}$$

Integrating two times and identifying $C = K_0$ and $Ck = K'_0$ yields Murnaghan's equation of state (MES):

$$E(V) = E(V_0) + \frac{VK_0}{K'_0(K'_0 - 1)}\left[\left(\frac{V_0}{V}\right)^{K'_0} + K'_0 - 1\right] - \frac{V_0K_0}{K'_0 - 1} \tag{Eq 11}$$

This somewhat complicated form can then be used to fit the energy as a function of volume calculated from electronic structure methods. For example, Fig. 4 shows the energy versus volume for fcc nickel using PAW pseudopotentials, and Table 6 gives the constants that produced the fitted curves.

In general, DFT will predict low-temperature lattice constants to within a percent. The LDA will typically underestimate lattice parameters, while the later, improved, gradient-corrected approximations do not follow this trend (Ref 68). The MES also produces an estimate of the bulk modulus and its derivative with respect to volume. There are reasonable alternatives to

Initial Configuration

```
*  Mo-Ni (Ni)24 (Ni4 Mo16) (Mo)12 **
 9.108
   1.00000       0.00000000   0.00000000
   0.00000000    1.00000      0.00000000
   0.00000000    0.00000000   0.97190000
 28 28
Direct
   .45190001     .11530000    .53219998
   .04809999     .88470000    .03219998
   .95190001     .38470000    .46780002
   .54809999     .61530000    .96780002
   .44240001     .36620000    .59719998
   .05759999     .63380000    .09719998
   .94240001     .13380000    .40280002
   .55759999     .86620000    .90280002
   .38820001     .05230000    .27480000
   .11179999     .94770000    .77480000
   .88820001     .44770000    .72520000
   .61179999     .55230000    .22520000
   .13370000     .07070000    .21570000
   .36630000     .92930000    .71570000
   .63370000     .42930000    .78430000
   .86630000     .57070000    .28430000
   .37680000     .43579999    .85670000
   .12320000     .56420001    .35670000
   .87680000     .06420001    .14330000
   .62320000     .93579999    .64330000
   .06800000     .14420000    .95289999
   .43200000     .85580000    .45289999
   .56800000     .35580000    .04710001
   .93200000     .64420000    .54710001
   .10310000     .41920000    .91329998
   .39690000     .58080000    .41329998
   .60310000     .08080000    .08670002
   .89690000     .91920000    .58670002
   .26480001     .19930000    .74860001
   .23513999     .80069999    .24860001
   .76480001     .30069999    .25139999
   .73519999     .69930001    .75139999
   .31360000     .24640000    .07400000
   .18640000     .75360000    .57400000
   .81360000     .25360000    .92600000
   .68640000     .74640000    .42600000
   .00290000     .19690000    .67670000
   .49710000     .80310000    .17670000
   .50290000     .30310000    .32330000
   .99710000     .69690000    .82330000
   .18850000     .01570000    .49599999
   .31150000     .98430000    .99599999
   .68850000     .48430000    .50400001
   .81150000     .51570000    .00400001
   .17630000     .48320001    .64249998
   .32370000     .51679999    .14249998
   .67630000     .01679999    .35750002
   .82370000     .98320001    .85750002
   .03380000     .33980000    .18070000
   .46620000     .66020000    .68070000
   .53380000     .16020000    .81930000
   .96620000     .83980000    .31930000
   .22890000     .28650001    .40979999
   .27110000     .71349999    .90979999
   .72890000     .21349999    .59020001
   .77110000     .78650001    .09020001
```

Final Configuration

```
*  Mo-Ni (Ni)24 (Ni4 Mo16) (Mo)12 **
 9.1080000000000054
   .9996694028707155    .0000000000000000    .0000000000000000
   .0000000000000000    .9997687828277412    .0000000000000000
   .0000000000000000    .0000000000000000    .9767010090926945
 28  28
Direct
   .4506643768099344    .1193068679416401    .5325479584880161
   .0493356231900656    .8806931320583600    .0325479584880161
   .9506643768099343    .3806931320583599    .4674520415119839
   .5493356231900657    .6193068679416400    .9674520415119839
   .4390191950353636315 .3770221180583645    .5972503663884575
   .0609800496463684    .6229778819416355    .0972503663884576
   .9390199503536316    .1229778819416355    .4027496336115424
   .5609800496463684    .8770221180583645    .9027496336115425
   .3880875526506487    .0559391388777245    .2720464605775263
   .1119124473493513    .9440608611222755    .7720464605775261
   .8880875526506486    .4440608611222755    .7279535394224739
   .6119124473493514    .5559391388777245    .2279535394224737
   .1305762636276348    .0701291674759629    .2171473356559112
   .3694237363723653    .9298708325240371    .7171473356559112
   .6305762636276347    .4298708325240371    .7828526643440898
   .8694237363723653    .5701291674759629    .2828526643440889
   .3708537893684525    .4412869934178103    .8579881136158205
   .1291462106315474    .5587130065821898    .3579881136158205
   .8708537893684526    .0587130065821897    .1420118863841867
   .6291462106315474    .9412869934178102    .6420118863841795
   .0672333579610334    .1441663236300990    .9548821936676196
   .4327666420389666    .8558336763699008    .4548821936676197
   .5672333579610335    .3558336763699006    .0451178063323803
   .9327666420389665    .6441663236300992    .5451178063323804
   .0979147884406859    .4168979366954007    .9114546709612028
   .4020852115593141    .5831020633045993    .4114546709612028
   .5979147884406859    .0831020633046064    .0885453290387972
   .9020852115593141    .9168979366954007    .5885453290387972
   .2625803246569123    .2039912353823866    .7483085557329274
   .2374196753430877    .7960087646176135    .2483085557329273
   .7625803246569123    .2960087646176135    .2516914442670727
   .7374196753430877    .7039912353823865    .7516914442670726
   .3069190985937987    .2528840709561812    .0728398268099706
   .1930809014062013    .7471159290438190    .5728398268099706
   .8069190985937985    .2471159290438189    .9271601731900295
   .6930809014062015    .7528840709561810    .4271601731900294
   .0014759749891245    .1925737429781877    .6752655257190752
   .4985240250108754    .8074262570218123    .1752655257190751
   .5014759749891246    .3074262570218124    .3247344742809249
   .9985240250108754    .6925737429781877    .8247344742670726
   .1860527493132290    .0124308173658511    .4945098491188892
   .3139472506867710    .9875691826341489    .9945098491188892
   .6860527493132291    .4875691826341489    .5054901508811108
   .8139472506867709    .5124308173658511    .0054901508811108
   .1761110489577723    .4854782446104694    .6408034101715743
   .3238889510422276    .5145217553895307    .1408034101715943
   .6761110489577723    .0145217553895306    .3591965898284058
   .8238889510422277    .9854782446104693    .8591965898284059
   .0287512823268599    .3372034257831562    .1809562826145831
   .4712487176731401    .6627965742168439    .6809562826145832
   .5287512823268599    .1627965742168438    .8190437173854168
   .9712487176731401    .8372034257831561    .3190437173854170
   .2265797861926567    .2846050956865158    .4105987224600145
   .2734202138073433    .7153949043134844    .9105987224600145
   .7265797861926566    .2153949043134842    .5894012775399855
   .7734202138073434    .7846050956865156    .0894012775399856
```

Fig. 3 Screen shot of initial and final cell configurations for δ-Mo50-Ni

Fig. 4 Murnaghan's equation of state (MES): Energy as a function of volume for face-centered cubic nickel, calculated using VASP

Fig. 5 Calculation of C_{ij} from volume-conserving strains applied to face-centered cubic nickel calculated using the Vienna ab initio simulation package (VASP)

MES, such as the Birch form (Ref 69) that is favored in some applications (Ref 57, 70).

Recent improvements to the gradient corrections that are designed to alleviate problems in low-charge density regions (i.e., internal voids, surfaces, and surface interactions) have produced mean errors in lattice parameters of approximately 0.1% over a wide range of materials (Ref 57, 70).

Elastic Constants. One of the primary uses of DFT in crystalline metals has been to predict lattice parameters and elastic constants. Initially, the calculations were used to assess the validity of the LDA and the computational methods. However, as the LDA became more established and the exchange-correlation functionals became more refined, it became routine for groups to predict the Cij of simple metals. In the early 1990s, DFT was used extensively to predict the elastic constants of a variety of high-temperature intermetallics. Mehl and coworkers at the Naval Research Laboratory were one of the first groups to apply these methods. They developed a robust strategy for assessing Cij for cubic and tetragonal crystal structures (Ref 70).

Table 6 Results of Murnaghan's equation of state (MES) fit to Vienna ab initio simulation package (VASP) total energies for face-centeral cubic nickel

	Ni-SA	Ni.-SP	Experimental(a)
$E(V_0)$, eV	−21.67	−21.867	. . .
K_0, Mbar	1.9681	1.9420	1.876
K_0', eV/Å^6	4.7944	4.7628	. . .
V_0, Å^3	43.429	43.684	. . .
a_0, Å	3.5150	3.5218	3.5238

SA, spin averaged; SP, spin polarized. (a) Source: Ref 61, 62, 67

The general approach is to express the free energy of the system as a function of the strain tensor acting on a small simulation cell volume:

$$dF = -SdT + PdV + dW$$

where dW is the infinitesimal work done by elastically distorting the crystal. Specifically,

$$dW = \sigma_{ij}d\varepsilon_{ij} = C_{ijkl}\varepsilon_{kl}d\varepsilon_{ij},$$

where we have used the definition of the elastic constants relating the applied stress to the resulting strain: $\sigma_{ij} = C_{ijkl}\varepsilon_{kl}$. Assuming reversible and isothermal loading at zero pressure, $dF = dW = C_{ijkl}\varepsilon_{kl}d\varepsilon_{ij}$, and we can write: $d^2F/d\varepsilon_{ij}d\varepsilon_{kl} = C_{ijkl}$. Changing notations from the fourth-rank tensor to the reduced second-rank tensor (Ref 71), we express the energy of the system around equilibrium using a Taylor series expansion in the strain:

$$E(\varepsilon) = E_0 - P(V_0)dV + \frac{V_0}{2}\sum_{i=1}^{6}\sum_{j=1}^{6}C_{ij}\varepsilon_i\varepsilon_j + \mathcal{O}(\varepsilon^2)$$

(Eq 12)

where V and $P(V)$ are the volume and pressure of the undistorted lattice, and dV is the change in volume produced by the strain ε_i. It is natural to apply strains to the simulation cell by transforming the primitive lattice vectors of the cell using the strain tensor, ε:

$$\begin{bmatrix} a_1' \\ a_2' \\ a_3' \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}(\underline{1} + \varepsilon)$$

$$\text{with } \varepsilon = \begin{bmatrix} \varepsilon_1 & \varepsilon_{6/2} & \varepsilon_{5/2} \\ \varepsilon_{6/2} & \varepsilon_2 & \varepsilon_{4/2} \\ \varepsilon_{5/2} & \varepsilon_{4/2} & \varepsilon_3 \end{bmatrix}$$

considering only nonrotating strains. Now, for the specific case of cubic crystals where:

$$C_{ij} = \begin{bmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{bmatrix}$$

then the double summation in the equation for the energy as a function of strain becomes:

$$\sum_{i=1}^{6}\sum_{j=1}^{6}C_{ij}\varepsilon_i\varepsilon_j = (\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2)C_{11}$$
$$+ 2(\varepsilon_1\varepsilon_2 + \varepsilon_1\varepsilon_3 + \varepsilon_2\varepsilon_3)C_{12}$$
$$+ (\varepsilon_4^2 + \varepsilon_5^2 + \varepsilon_6^2)C_{44}$$

The effects of some applied strain are now explicitly coupled by elastic constants to changes in energy for a simulation cell with unit vectors a_i'. For example, take the case of a hydrostatic stress of $\varepsilon_1 = \varepsilon_2 = \varepsilon_2 = \delta$. This yields:

$$E(\varepsilon) = E_0 + \frac{V_0}{2}\delta^2(3C_{11} + 6C_{12})$$

Identifying the bulk modulus:

$$K = \frac{C_{11} + 2C_{12}}{3}$$

we find $E(\varepsilon) = E_0 + \frac{9}{2}V_0K\delta^2$. Applying this form to the data in Fig. 4 for the spin-polarized case yields a bulk modulus of 2.01 Mbar, in good agreement with MES.

To find the three independent elastic constants, two other equations are required, and the normal convention is to apply two other, volume-conserving strains. For cubic systems, usually $C_{11} - C_{12}$ is found by applying $\varepsilon_1 = -\varepsilon_2 = \delta$, with ε_3 being set by the constant volume constraint for a cubic cell:

$$\varepsilon_3 = \frac{\delta^2}{1 - \delta^2}$$

Similarly, C_{44} is found by setting $\varepsilon_6 = \frac{\delta}{2}$ and $\varepsilon_2 = \frac{\delta^2}{4 - \delta^2}$. This yields two other equations for the energy as a function of strain:

$$E(\delta) = E_0 + V_0\delta^2(C_{11} - C_{12}) + \mathcal{O}(\delta^4)$$

$$E(\delta) = E_0 + \frac{V_0}{2}\delta^2C_{44} + \mathcal{O}(\delta^4)$$

The response of a unit cubic cell of nickel to such strains is shown in Fig. 5. When the fits to the two curves are combined with the MES results, the elastic constants can be resolved as shown for fcc nickel and L1$_2$ Ni$_3$Al in Tables 7 and 8. The tabulated results are for spin-averaged and spin-polarized (ferromagnetic) systems using LDA and GGA approximations. Tables 7 and 8 show the results from a PPW calculation (VASP) using ultrasoft and projected augmented-wave pseudopotentials, respectively. Note that the LDA and LSDA underestimate the lattice parameter and overestimate the elastic constants, and the GGA results (PW91) show uniform improvement in lattice parameters and elastic constants. Spin-averaged and spin-polarized (ferromagnetic) calculations in the LDA and a GGA (PW91) were used to predict the lattice parameter (angstrom), elastic constants (Mbar), and misfit parameter: $\delta = 2(a_{Ni_3Al} - a_{Ni})/(a_{Ni_3Al} + a_{Ni})$. As expected, the LDA and LSDA underestimated the lattice parameters for nickel and Ni$_3$Al. The GGA and spin-generalized gradient approximation (SGGA) produced significantly more precise lattice parameters and elastic constants, with the SGGA calculations giving the most accurate

Table 7 Structural parameters for face-centered cubic (fcc) nickel and L1$_2$ Ni$_3$Al calculated using ultrosoft pseudopotentials in the PPW method VASP

USPP		Spin averaged				Spin polarized				
Metal	Property	LDA	Error, %	GGA	Error, %	LSDA	Error, %	SGGA	Error, %	Experimental(a)
fcc Ni	a_0, Å	3.4294	−2.7	3.5258	0.1	3.4221	−2.9	3.5337	0.3	3.5238
	K, Mbar	2.515	35	1.985	6.8	2.383	28	1.962	5.5	1.860
	C_{11}	3.154	27	2.506	1.0	3.034	22	2.380	−4.1	2.481
	C_{12}	2.195	42	1.725	11	2.057	33	1.753	13	1.549
	C_{44}	1.358	9	1.057	−15	1.363	10	1.253	0.9	1.242
	\<err\>	. . .	23	. . .	6.8	. . .	19	. . .	4.8	. . .
L1$_2$ Ni$_3$Al	a_0, Å	3.4893	−2.2	3.5769	0.3	3.4928	−2.1	3.5784	0.3	3.5670
	K, Mbar	2.163	24	1.777	2.1	2.159	24	1.787	2.7	1.740
	C_{11}	2.749	21	2.271	0.3	2.778	23	2.397	5.9	2.264
	C_{12}	1.870	26	1.529	3.3	1.849	25	1.481	0.1	1.480
	C_{44}	1.431	11	1.173	−8.7	1.471	15	1.240	−3.4	1.284
	\<err\>	. . .	17	. . .	2.9	. . .	18	. . .	2.5	. . .
	δ	0.0173	42	0.0143	18	0.0204	68	0.0125	3	0.0121

(a) PPW, Pseudopotential plane wave; USPP, ultrasoft pseudopotential; LDA, local density approximation; GGA, generalized gradient approximation; LSDA, local spin density approximation; SGGA, spin-generalized gradient approximation. (a) Source: Ref 61, 62, 67

Table 8 Structural parameters for face-centered cubic (fcc) nickel and L1$_2$ Ni$_3$Al calculated using projected augmented wave pseudopotentials in the PPW method VASP

PAW		Spin averaged				Spin polarized				
Metal	Property	LDA	Error, %	GGA	Error, %	LSDA	Error, %	SGGA	Error, %	Experimental(a)
fcc Ni	a_0, Å	3.4197	−3.0	3.5150	−0.2	3.4258	−2.8	3.5219	−0.1	3.5238
	K, Mbar	1.144	−38	1.968	5.8	1.175	−37	1.942	4.4	1.860
	C_{11}	3.185	28	2.480	0.0	3.476	40	2.704	9.0	2.481
	C_{12}	2.227	44	1.712	11	2.019	30	1.561	0.8	1.549
	C_{44}	1.383	11	1.121	−10	1.618	30	1.294	4.2	1.242
	\<err\>	. . .	25	. . .	5	. . .	28	. . .	4	. . .
L1$_2$ Ni$_3$ Al	a_0, Å	3.4823	−2.4	3.5685	0.04	3.4927	−2.1	3.5699	0.1	3.5670
	K, Mbar	2.183	25	1.779	2.2	2.171	25	1.773	1.9	1.740
	C_{11}	2.774	23	2.264	0.0	2.787	23	2.343	3.5	2.264
	C_{12}	1.888	28	1.537	3.9	1.863	26	1.488	0.5	1.480
	C_{44}	1.441	12	1.189	−7.4	1.488	16	1.248	−2.8	1.284
	\<err\>	. . .	18	. . .	2.7	. . .	18	. . .	1.8	. . .
	δ	0.0181	49	0.0151	24	0.0193	59	0.0135	11	0.0121

PAW, projector augmented wave; LDA, Local density approximation; GGA, generalized gradient approximation; LSDA, Local spin density approximation; SGGA, Spin-generalized gradient approximation.
(a) Source: Ref 61, 62, 67

misfit parameter. Finally, although PAW is assumed to be a better representation of the core states, the ultrasoft pseudopotentials (USPPs) produce a more accurate misfit parameter.

Entropic Contributions to the Free Energy. In the last 10 years, significant progress has been made in calculating the entropic contributions to the free energy of bulk phases and defects. This includes configurational, vibrational, and electronic entropic terms. Examples of applications, including references reviewing the techniques, are given here. Electronic entropy has been shown to be important in calculating defect energies, such as vacancies in body-centered cubic metals (Ref 72, 73). Contribution of thermal vibrations to the free energy as a function of volume (harmonic and anharmonic terms) has been used to estimate the thermal expansion of a variety of metals (Ref 9, 74). Configurational entropy for dilute solute concentrations is treated using the Bragg-Williams approximations in conjunction with either lattice gas models or low-temperature expansion (Ref 75, 76). For solid solutions at high concentrations, cluster expansion methods (Ref 77, 78) are used to approximate the free energy on an Ising model lattice. Recent progress in methods development has automated parts of the construction and use of these techniques (Ref 1, 79). Van de Walle and coworkers have also attempted to include all three entropic contributions in modeling phase stability and to inform CALculation of PHAse Diagram (CALPHAD) methods (Ref 80). These developments have significantly improved the efficiency and accuracy of the cluster expansion approach, particularly in its application to phase diagrams.

REFERENCES

1. A. van de Walle and G. Ceder, The Effect of Lattice Vibrations on Substitutional Alloy Thermodynamics, *Rev. Mod. Phys.,* Vol 74, 2002, p 11
2. A. Janotti, M. Krcmar, C.L. Fu, and R.C. Reed, Solute Diffusion in Metals: Larger Atoms Can Move Faster, *Phys. Rev. Lett.,* Vol 92, 2004, p 085901
3. C. Woodward and S.I. Rao, Flexible Ab Initio Boundary Conditions: Simulating Isolated Dislocations in BCC Mo and Ta, *Phys. Rev. Lett.,* Vol 88, 2001, p 216402
4. B. Grabowski, T. Hickel, and J. Neugebauer, *Phys. Rev. B,* Vol 76, 2007, p 024309
5. D.J. Siegel, C. Wolverton, and V. Ozoliņš, Thermodynamic Guidelines for the Prediction of Hydrogen Storage Reactions and Their Application to Destabilized Hydride Mixtures, *Phys. Rev. B,* Vol 76, 2007, p 134102
6. M. Ernzerhof and G.E. Scuseria, Assessment of the Perdew-Burke-Ernzerhof Exchange-Correlation Functional, *J. Chem. Phys.,* Vol 98, 199, p 5029–5036
7. J.P. Perdew, in *Electronic Structure of Solids'91,* P. Ziesche and H. Eschrig, Ed., Akademie-Verlag, Berlin, 1991
8. J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, and C. Fiolhais, *Phys. Rev. B,* Vol 46, 1992, p 6671
9. P. Söderlind, L. Nordström, L. Yongming, and B. Johansson, Relativistic Effects on the Thermal Expansion of the Actinide Elements, *Phys. Rev. B,* Vol 42, 1990, p 4544–4552
10. C. Woodward, D.R. Trinkle, L.G. Hector, Jr., and D.L. Olmsted, Prediction of Dislocation Cores in Aluminum from Density Functional Theory, *Phys. Rev. Lett.,* Vol 100, 2008, p 045507
11. W.M.C. Foulkes, L. Mitáš, R.J. Needs, and G. Rajagopal, Quantum Monte Carlo

Simulations of Solids, *Rev. Mod. Phys.,* Vol 73, 2001, p 33–83

12. R. Martin, *Electronic Structure, Basic Theory and Practical Methods,* Cambridge University Press, Cambridge, U.K., 2004

13. M.C. Payne, M.P. Teter, D.C. Allen, T.A. Arias, and J.D. Joannopoulos, Interative Minimization Techniques for Ab Initio Total Energy Calculations: Molecular Dynamics and Conjugate Gradients, *Rev. Mod. Phys.,* Vol 64, 1992, p 1045–1097

14. D. Singh, *Planewaves, Pseudopotentials and the LAPW Method,* Kluwer Academic Press, Norwell, MA, 1994

15. N.W. Ashcroft and N.D. Mermin, *Solid State Physics,* Saunders College, Philadelphia, PA, 1976

16. J. Hohenberg and L.J. Kohn, *Phys. Rev. B,* Vol 136, 1962, p 864

17. W. Kohn and L.J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects, *Phys. Rev.,* Vol 140, 1965, p 1133–1138

18. D. Pines, *Elementary Excitations in Solids,* Benjamin Inc., NY, 1963

19. D.M. Ceperley and B.J. Alder, *Phys. Rev. Lett.,* Vol 45, 1980, p 566

20. P. Perdew and A. Zunger, *Phys. Rev. B,* Vol 23, 1981, p 5048

21. L. Moruzzi, J.F. Janak, and A.R. Williams, *Calculated Electronic Properties of Metals,* Pergamon, New York, 1978

22. J.C. Slater, Wavefunctions in a Periodic Potential, *Phys. Rev.,* Vol 51, 1937, p 846

23. J.M. Wills and B.R. Cooper, *Phys. Rev. B,* Vol 36, 1987, p 3809

24. D.L. Price and B.R. Cooper, *Phys. Rev. B,* Vol 39, 1989, p 4945

25. J. Korringa, *Physica* (Amsterdam), Vol 13, 1947, p 392

26. W. Kohn and N. Rostoker, *Phys. Rev.,* Vol 94, 1954, p 1111

27. N. Papanikolaou, R. Zeller, P.H. Dederichs, and N. Stefanou, *Comput. Mater. Sci.,* Vol 8, 1997, p 131

28. M. Weinart, E. Wimmer, and A.J. Freeman, Total-Energy All Electron Density Functional Method for Bulk Solids and Surfaces, *Phys. Rev. B,* Vol 26, 1982, p 4571–4578

29. H.J.F. Jansen and A.J. Freeman, Total Energy Full Potential Linearized Augmented-Planewave Method for Bulk Solids, *Phys. Rev. B,* Vol 30, 1984, p 561–569

30. K.H. Weyrich, Full Potential Linear Muffin-Tin Orbital Method, *Phys. Rev. B,* Vol 37, 1988, p 10269–10282

31. F. Herman, Theoretical Investigation of the Electronic Energy Band Structure of Solids, *Rev. Mod. Phys.,* Vol 30, 1958, p 102

32. N. Troullier and J.L. Martins, *Phys. Rev. B,* Vol 43, 1991, p 1993

33. D. Vanderbilt, *Phys. Rev. B,* Vol 41, 1990, p 7892

34. P.E. Blochl, *Phys. Rev. B,* Vol 50, 1994, p 17953

35. M. Marsman and G. Kresse, *J. Chem. Phys.,* Vol 125, 2006, p 104101

36. R. Feynman, *Phys. Rev.,* Vol 56, 1939, p 340

37. O.H. Nielsen and R.M. Martin, *Phys. Rev. B,* Vol 32, 1985, p 3780

38. S.Y. Savrasov, D.Y. Savrasov, and O.K. Anderson, Linear Response Calculations of Electron-Phonon Interactions, *Phys. Rev. Lett.,* Vol 72, 1994, p 372–375

39. R. Car and M. Parrinello, *Phys. Rev. Lett.,* Vol 55, 1985, p 2471–2475

40. M.P. Teter, M.C. Payne, and D.C. Allen, *Phys. Rev. B,* Vol 40, 1989, p 12255

41. T. Hahn, Ed., *International Tables for Crystallography,* Vol A, *Space-Group Symmetry,* 5th ed., Dordrecht, Kluwer Academic Publishers, 2002

42. D.J. Chadi and M.L. Cohen, *Phys. Rev. B,* Vol 8, 1973, p 5747

43. H.J. Monkhorst and J.D. Pack, *Phys. Rev. B,* Vol 13, 1976, p 5188

44. M. Weinert and J.W. Davenport, *Phys. Rev. B,* Vol 45, 1992, p 13709

45. C.-L. Fu and K.-M. Ho, First Principles Calculations of the Equilibrium Ground State Properties of Transition Metals: Application to Nb and Mo, *Phys. Rev. B,* Vol 28, 1983, p 5480–5486

46. C. Elsasser, M. Fahnle, C.T. Chan, and K.M. Ho, Density-Functional Energies and Forces with Gaussian-Broadened Fractional Occupancies, *Phys. Rev. B,* Vol 49, 1994, p 13975

47. F. Nogueiray, C. Fiolhaisy, J. Hez, J.P. Perdew, and A. Rubio, Transferability of a Local Pseudopotential Based on Solid State Electron Density, *J. Phys., Condens. Matter,* Vol 8, 1996, p 287–302

48. Y.A. Wang and E.A. Carter, in *Theoretical Methods in Condensed Phase Chemistry,* S.D. Schwartz, Ed., within the series *Progress in Theoretical Chemistry and Physics,* Kluwer, Dordrecht, 2000, p 117–184

49. W.C. Herring, A New Method for Calculating Wave Functions in Crystals, *Phys. Rev.,* 57, 1940, p 1169

50. F. Herman and J. Callaway, Electronic Strcuture of the Germanium Crystal, *Phys. Rev.,* 1953, p 518

51. W. Kohn, *Rev. Mod. Phys.,* Vol 71, 1999, p 1253

52. V. Ozoliņš and M. Körling, Full Potential Calculations Using the Generalized Gradient Approximation: Structural Properties of Transition Metals, *Phys. Rev. B,* Vol 48, 1993, p 18304–18307

53. T.C. Leung, C.T. Chan, and B.N. Harmon, *Phys. Rev. B,* Vol 44, 1991, p 2923

54. J.-H. Cho and M. Scheffler, *Phys. Rev. B,* Vol 53, 1996, p 10685

55. J.P. Stevens, F.J. Devlin, C.F. Chablowski, and M.J. Frisch, *J. Phys. Chem.,* Vol 80, 1994, p 11623

56. A.D. Becke, *J. Chem. Phys.,* Vol 98, 1993, p 5648

57. A.E. Mattsson, R. Armiento, J. Paier, G. Kresse, J.M. Wills, and T.R. Mattsson, *J. Chem. Phys.,* Vol 128, 2008, p 084714

58. S.Y. Savrasov, *Phys. Rev. B,* Vol 54, 1996, p 16470

59. G. Kresse and J. Furthmüller, Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set, *Comput. Mater. Sci.,* Vol 6, 1996, p 15–50

60. G. Kresse and J. Furthmüller, Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set, *Phys. Rev. B,* Vol 54, 1996, p 11169

61. W.B. Pearson, *A Handbook of Lattice Spacings and Structures of Metals and Alloys,* Pergamon Press, NY, 1967

62. "Crystal Lattice Structures," Naval Research Laboratory, http://cst-www.nrl.navy.mil/lattice/index.html, Oct 29, 2008

63. "Bilbao Crystallographic Server," Condensed Matter Physics Department, University of the Basque Country, http://www.cryst.ehu.es/

64. C.B. Shoemaker and D.P. Shoemaker, Crystal Structure of the d Phase, Mo-Ni, *Acta Crystallogr.,* Vol 16, 1963, p 997

65. Y. Wang, C. Woodward, S.H. Zhou, Z.-K. Liu, and L.-Q. Chen, Structural Stability of Ni-Mo Compounds from First-Principles Calculations, *Scr. Metall.,* Vol 52, 2005, p 17–20

66. F.D. Murnaghan, *Proc. Nat. Acad. Sci.,* USA, Vol 30, 1944, p 244

67. G. Simmons and H. Wang, *Single Crystal Elastic Constants,* MIT Press, Cambridge, MA, 1971

68. P. Söderlind, O. Eriksson, J.M. Wills, and A.M. Boring, Theory of Elastic Constants of Cubic Transition Metals and Alloys, *Phys. Rev. B,* Vol 48, 1993, p 5844–5851

69. F. Birch, *J. Geophys. Res.,* Vol 83, 1978, p 1257

70. M.J. Mehl, J.E. Osborn, D.A. Papaconstantopoulos, and B.M. Klein, *Phys. Rev. B,* Vol 41, 1990, p 10311–10323

71. J.F. Nye, *Physical Properties of Crystals,* Clarendon Press, Oxford, 1987

72. A. Satta, F. Willaime, and S. de Gironcoli, First Principles Study of Vacancy Formation and Migration Energies in Tantalum, *Phys. Rev. B,* Vol 60, 1999, p 7001–7005

73. A. Satta, F. Willaime, and S. de Gironcoli, Vacancy Self-Diffusion in Tungsten: Finite Electron-Temperature LDA Calculations, *Phys. Rev. B,* Vol 57, 1998, p 11184–11192

74. V.L. Moruzzi, J.F. Janak, and K. Schwarz, *Phys. Rev. B,* Vol 37, 1988, p 790

75. J. Mayer, C. Elsasser, and M. Fahnle, *Phys. Stat. Solidi B,* Vol 191, 1995, p 283

76. C. Woodward, M. Asta, G. Kresse, and J. Hafner, Density of Constitutional and Thermal Point Defects in $L1_2$ Al_3Sc, *Phys. Rev. B,* Vol 63, 2001, p 094103

77. J.M. Sanchez, F. Ducastelle, and D. Gratis, *Physica A,* Vol 128, 1984, p 334

78. D. de Fontaine, Configurational Thermodynamics of Solid Solutions, *Solid State Physics,* Vol 34, Academic Press, New York, 1979, p 73

79. A. van de Walle, M. Asta, and G. Ceder, The Alloy Theoretic Automated Toolkit: A User Guide, *Calphad,* Vol 26, 2002, p 539–553

80. A. van de Walle, G. Ghosh, and M. Asta, Ab Initio Modeling of Alloy Phase Stability, *Applied Computational Materials Modeling: Theory, Simulation and Experiment,* G. Bozzolo, R.D. Noebe, and P. Abel, Ed., Springer, New York, 2007, p 1–34

SELECTED REFERENCES

- R. Martin, *Electronic Structure, Basic Theory and Practical Methods,* Cambridge University Press, Cambridge, U.K., 2004
- M.C. Payne, M.P. Teter, D.C. Allen, T.A. Arias, and J.D. Joannopoulos, Interative Minimization Techniques for Ab Initio Total Energy Calculations: Molecular Dynamics and Conjugate Gradients, *Rev. Mod. Phys.,* Vol 64, 1992, p 1045–1097
- "Pearson's Crystal Structure Database for Inorganic Compounds," http://asmcommunity.asminternational.org/portal/site/www/MatInformation/Pearsons/
- D. Singh, *Planewaves, Pseudopotentials and the LAPW Method,* Kluwer Academic Press, Norwell, MA, 1994
- J.C. Slater, *Quantum Theory of Matter,* McGraw-Hill, New York, NY, 1968
- J.C. Slater, *Quantum Theory of Molecules and Solids,* Vol 1–3, McGraw-Hill, New York, NY, 1974

Modeling of Microstructures

Simulation of Microstructural Evolution in Steels

P.M. Pauskar, The Timken Company
R. Shivpuri, The Ohio State University

STEEL is certainly one of the most widely used materials. It has a broad range of uses, primarily as raw material in the manufacture of automobiles, machinery, appliances, and also in the construction of buildings, bridges, roads, and railways. In its basic form, steel is a combination of iron and carbon. Many steels also contain other alloying elements in various combinations and quantities. During high-temperature thermomechanical processing, steel ordinarily consists of single-phase austenite face-centered cubic. The condition of the austenite (grain size, degree of recrystallization, etc.) affects the phase transformations that occur during cooling following thermomechanical processing. During service at ambient temperature, steel ordinarily consists of two phases: ferrite and cementite. Different morphological arrangements of these phases are termed microconstituents (pearlite, bainite, tempered martensite, allotriomorphic ferrite, etc.). The aggregate arrangement of microconstituents in the steel represents the steel microstructure. The microstructure in steel depends greatly upon the steel chemistry and its thermomechanical history. A wide range of mechanical properties can be obtained by engineering the steel chemistry and microstructure. This ability to engineer desired mechanical properties in steel by modifying chemical composition and altering microstructure via heat treatment or thermomechanical processing makes steel an extremely versatile material.

Hot working involves deforming steel under conditions above its recrystallization temperature. Common hot working processes include hot rolling, hot forging, and hot extrusion. Computer simulation of microstructural evolution during hot rolling of steels has been a major topic of research and development in academia and industry for over two decades. A good portion of this article is therefore focused on simulation of microstructural evolution during hot rolling. However, it is important to recognize that the principles of simulation discussed in this article are also applicable to other hot working and heat treatment processes. The primary intent of this article is to provide a basic description of the methodology and procedures commonly employed to develop microstructural evolution models and using them to simulate microstructural evolution in steels. Readers are encouraged to refer to the article "Models of Recrystallization" in this Volume for additional information.

Microstructural Evolution during Hot Working

Microstructural evolution during hot working depends primarily on the processing conditions, such as the hot working temperature, amount of deformation (strain), deformation rate (strain rate), and cooling rate. Figure 1 illustrates a typical heating, deformation, and cooling sequence in a hot rolling process. In rolling mills, billets, blooms, or slabs of steel are reheated to a high temperature, typically approximately 1200 °C, before being hot rolled into flat products such as strips and plates or long products such as bars, rails, and structural shapes. Reheating of the steel is followed by successive reductions in rough and finish rolling mills. Rolling is generally finished at specific temperatures, followed in many cases by controlled cooling. In some cases, such as hot strip rolling, accelerated water cooling is used to attain the desired microstructure and properties in the product. By means of these controlled processing steps during hot working, a broad range of microstructure and mechanical properties can be achieved in steel alloys (Ref 1). Microstructural changes occurring at different stages in the rolling process are described as follows.

Microstructural Changes during Reheating. A billet or a slab is initially heated to a specific temperature and soaked for a period of time
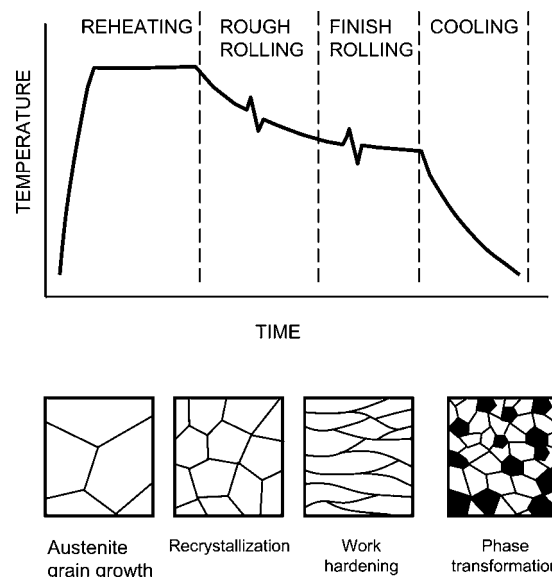
Fig. 1 Typical thermomechanical schedule and associated microstructural changes during hot rolling

before rolling. This is referred to as reheating. The heating temperature and soaking time are dependent upon steel chemistry. For microalloyed or alloyed steels, in particular, the reheating temperature is so chosen as to dissolve the right amount of carbide or nitride precipitates. In general terms, the reasons for reheating are:

- To homogenize and soften the steel for rolling
- To ensure that the rolling process is completed with steel generally in the austenitic state
- To ensure that carbides or nitrides, where applicable, are in solution before hot rolling

During the initial stages of reheating, grain growth of ferrite takes place, followed by the ferrite-to-austenite $(\alpha - \gamma)$ transformation. Finally, grain growth of austenite occurs. The size of austenite grains prior to hot rolling is considered one of the important variables that affect recovery and recrystallization phenomena that occur during hot working.

Microstructural Changes during Hot Rolling. After reheating, the material undergoes deformation in the rolls. Deformation leads to an increase in dislocation density, causing the material to work harden. The stored energy due to accumulated dislocations is generally lowered by three processes: recovery, recrystallization, and grain growth. Recovery is a process by which annihilation and rearrangement of the dislocations occurs. Recrystallization, on the other hand, is a process in which new dislocation-free grains are formed in place of old deformed grains, resulting in a new structure with lower dislocation density. Grain growth is the process in which grains coarsen and lower the grain-boundary area per unit volume. Recrystallization that occurs while the metal is being deformed is called dynamic recrystallization, whereas recrystallization that occurs after deformation is completed is classified as either static or metadynamic recrystallization. Recrystallization, in general, results in finer grains compared to those existing before.

Microstructural Changes in the Interstand Region. The dislocation structures introduced during a rolling pass influence the microstructural changes in the time interval before the next deformation pass. Static microstructural changes typically consist of static recovery, static recrystallization, and subsequent grain growth. Under certain conditions, dynamic recrystallization that nucleates during deformation undergoes completion in the interstand region. This is referred to as metadynamic recrystallization. The extent of these microstructural changes in the interstand region depends on their kinetics in relation to the time interval available before the next pass. Complete recrystallization is usually followed by grain growth, if the existing conditions favor it.

Phase Transformation after Rolling. Hot rolling is mostly carried out with steel in the austenite phase. After hot rolling, depending on the steel chemistry and cooling conditions, austenite decomposes into a variety of transformation products, such as ferrite, pearlite, bainite, or martensite. The resulting mix of transformation products determines the mechanical properties of the rolled part.

All of these microstructural changes taking place in the material affect the properties of the final product. These also affect the forming process, because the flow stress and deformation behavior of the material are affected by changes in the microstructure. For accurate mathematical modeling of the hot rolling process, it is therefore important to take into account the physical and metallurgical changes taking place in the material during both forming and postforming cooling periods.

Hot rolling of steels has been the primary focus of many studies with regards to computer modeling of microstructural evolution in steels. Pioneering work in this area conducted by Sellars and Whiteman (Ref 2) in the 1970s has been advanced by significant contributions on this subject by numerous researchers from all over the world (Ref 3–11). Early work emphasized modeling of austenite grain structure evolution during multipass hot rolling—particularly modeling of recrystallization and grain growth. Numerous studies have been published on the subject of recrystallization and grain growth in carbon-manganese and high-strength low-alloy HSLA steels. Most of the early studies did not attempt to model postrolling austenite decomposition, primarily due to the complexities involved in modeling phase transformations. In recent years, there has been a growing emphasis on modeling of austenite decomposition. Recent advances in finite-element modeling of metal deformation processes, coupled with advances in materials characterization, have resulted in the development of integrated models for computer simulation of metal flow and microstructural evolution in a number of metalworking processes. Such integrated models have been used successfully for simulation of metal flow and microstructural evolution in hot working processes (Ref 4, 7, 9–11).

Development of Models for Austenite Evolution and Decomposition

To be able to simulate microstructural evolution in a hot deformation process, it is first necessary to develop the following models:

- Models for grain growth kinetics
- Models for recrystallization kinetics
- Models for recrystallized grain size
- Models for austenite decomposition

While attempts have been made by researchers to develop physically based models for recrystallization and grain growth (Ref 12–15), most of the models being used today (2009)

are considered as semiempirical or phenomenological models. These models are obtained by fitting experimental data to an equation whose general form is derived from theoretical analysis. Models developed in the early work conducted by Sellars and Whiteman (Ref 2) on carbon-manganese steels are considered phenomenological in nature. However, the general forms of these models are also applicable to most other steels, with minor changes. To develop such semiempirical or phenomenological models for microstructural evolution, experimental techniques, such as hot compression tests and hot torsion tests, have been extensively used by materials scientists worldwide. The following sections describe the general methodology used in developing these models.

Austenite Grain Growth

During hot rolling, austenite grain size affects the final transformed microstructure and hence the resulting mechanical properties. Austenite grain size during thermomechanical processing also affects recrystallization kinetics and flow stress. Therefore, modeling grain growth at various stages in the process (e.g., grain growth during reheating, grain growth after recrystallization) is considered important. To develop models for grain growth, experiments are commonly performed with austenitization temperature and hold time at that temperature as the control variables. Figure 2 illustrates the experimental procedure. The experiment consists of heating the specimen at a relatively fast rate (e.g., 30 °C/s) to a selected reheat temperature (T,°C) and holding the specimen at that temperature for a given amount of time (t, s). The rapid heating rate is generally used to limit the grain coarsening to the temperature of interest and to avoid vanadium carbonitride, V(CN), and aluminum nitride, AlN, reprecipitation (Ref 16). A variety of techniques can be used to metallographically measure the prior-austenite grain size after each test. A common approach with hypoeutectic steels it to treat the sample so as to form ferrite at the prior-austenite grain boundaries. Here, after
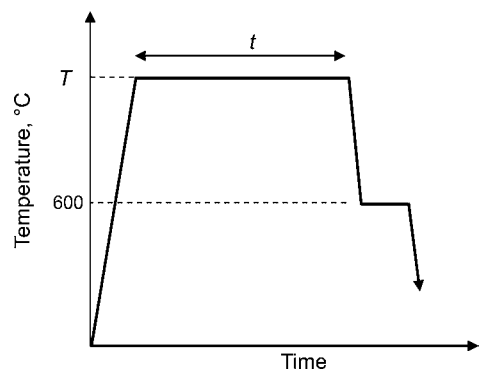
Fig. 2 Experimental procedure for grain-growth studies. T, reheat temperature

the grain-coarsening thermal cycle, the specimen is cooled rapidly to approximately 600 °C. The specimen is held at this temperature for a few seconds to allow formation of proeutectoid ferrite at the prior-austenite grain boundaries. The specimen is then quenched to room temperature. Prior-austenite grain size is then measured in these samples. This method of measuring grain size on a large number of specimens is time-consuming and expensive. New techniques using laser ultrasonics (Ref 17) have been shown to be very effective in real-time measurement of austenite grain size during hot deformation of steels.

Austenite grain growth is generally modeled using the empirical power law:

$$d^m = d_0^m + kt \exp\left[\frac{-Q}{RT}\right] \qquad \text{(Eq 1)}$$

Here, d is the grain size (in microns) at time t (seconds), d_0 is the initial grain size (in microns), Q is the apparent activation energy (J/mol), T is the absolute temperature (K), t is time in seconds, m and k are material constants, and R is the universal gas constant (8.31 J/mol.K). This model clearly illustrates that the higher the temperature and longer the hold time, the larger the grain size will be. The constants in the grain-growth equation, namely m, Q, and k, are obtained by means of regression analysis. A few published grain-growth models are listed in Table 1. Note that the best-fit exponent m is usually found to be higher than the theoretically expected value of 2 (parabolic growth law). Higher values of m are attributed to various factors, such as pinning forces due to precipitates and solute drag forces on grain boundaries. Published studies (Ref 2, 19, 20, and others) show values of m ranging from 2 to 13 for a variety of steels. Figure 3 shows the experimental measurements and the best-fit curve representing

the grain-growth equation for two different steels (Ref 16).

While these models show good correlation between measurements and predictions, it is important to note that this form of equation does not explicitly model the pinning effect and drag forces and other phenomena that affect grain growth. In that sense, it is not a true mechanism-based model. The value of activation energy can be misleading in such regression-based models. A higher value of the exponent m generally results in a higher value of Q. Therefore, in materials such as microalloyed steels that exhibit the pinning effect due to precipitates, m and consequently the Q-value obtained by means of regression analysis are generally higher than theoretical estimates.

Applying Grain-Growth Model to Non-isothermal Conditions. Grain-growth models discussed previously are developed under isothermal conditions, whereas the temperature changes constantly in a hot working process such as rolling or forging. Semiatin et al. (Ref 21) have developed closed-form equations for applying the isothermal grain-growth model under constant heating-rate conditions. During cooling, however, the temperature does not usually change at a constant rate. One analytical approach proposed by Ashby and Easterling

(Ref 22) is based on the idea of integrating the grain-growth equation. In this approach, it is assumed that grain growth is driven by the reduction in grain-boundary energy and requires no nucleation. Grain growth is therefore calculated by the following:

$$d_t^m = d_0^m + k \int_0^t \exp\left(\frac{-Q}{RT(t)}\right) \cdot dt \qquad \text{(Eq 2)}$$

where d_t is the average grain size at time t, and $T(t)$ is the absolute temperature expressed as a function of time. Defining temperature as a simple function of time is not always possible. Therefore, in a simplified approach, the temperature history is approximated by dividing it into several small isothermal time steps (Ref 9, 23, 24). During each time step, the temperature is assumed to be constant, as illustrated in Fig. 4. Therefore, the grain growth is calculated by:

$$d_t^m = d_0^m + k \sum_i \delta t_i \cdot \exp\left(\frac{-Q}{RT_i}\right) \qquad \text{(Eq 3)}$$

Table 1   A few published grain-growth models

Basic equation: $d^m = d_0^m + kt \exp\left[\frac{-Q}{RT}\right]$

Steel type(a)	m	k	Q	Reference
Low C-Mn	2	4.27×10^{12}	66,600	18
C-Mn and C-Mn-V	7	1.45×10^{27}	400,000	4
Ti-HSLA	10	2.60×10^{28}	437,000	4
Nb-HSLA	4.5	4.05×10^{23}	435,000	4
V-HSLA	5	1.26×10^{32}	655,800	9
0.2C Cr-Mo (4120)	4	3.00×10^{30}	675,000	16

(a) HSLA, high-strength low-alloy
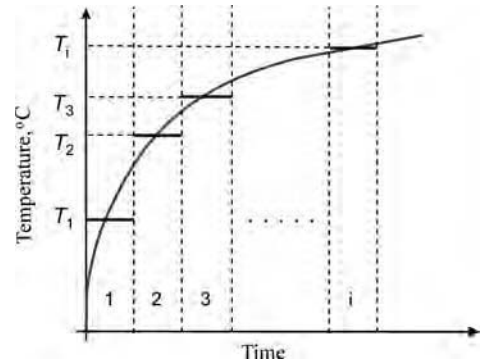
Fig. 4 Approximating the continuous time-temperature curve with discrete isothermal steps
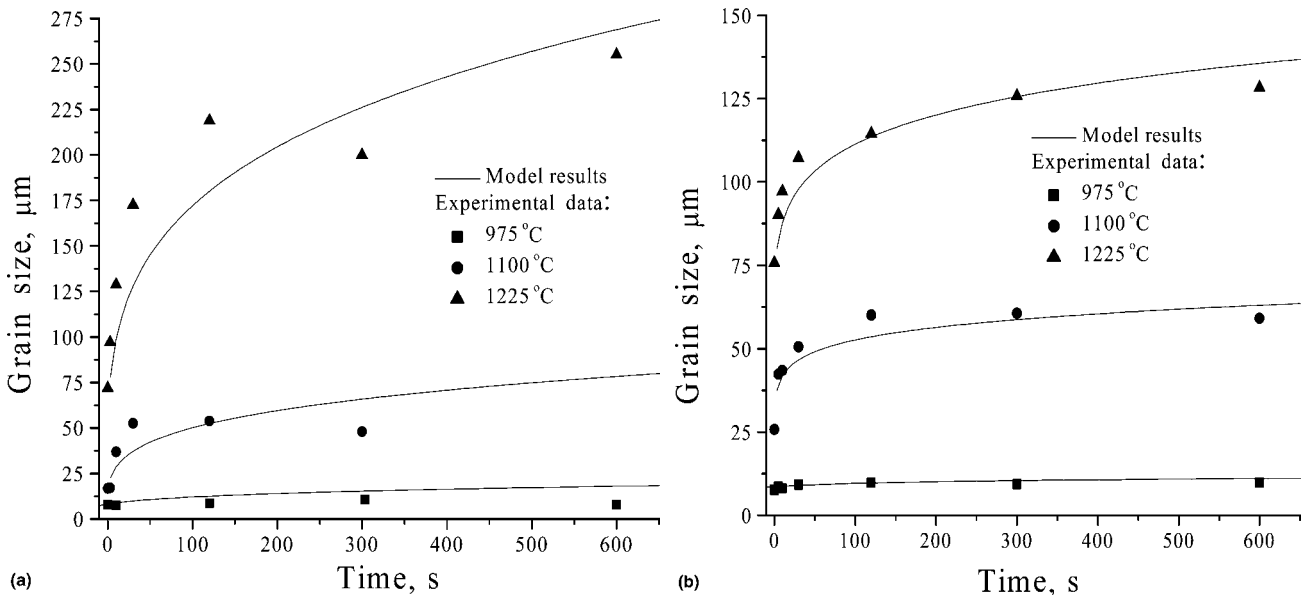
Fig. 3 Measurements from isothermal grain-growth tests (discrete points) and predictions from the grain-growth model for (a) 4120 steel and (b) 1538V steel. Source: Ref 16

(a)

(b)
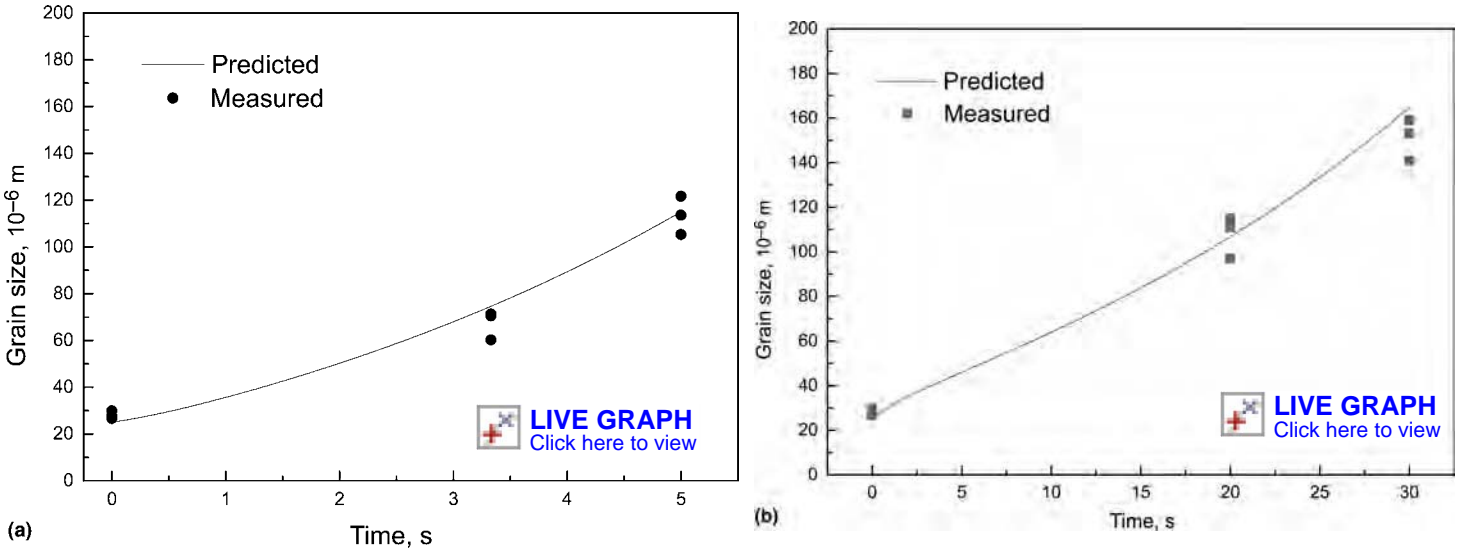
Fig. 5 Measurements vs. predictions in nonisothermal grain-growth tests for heating rates of (a) 30° C/s and (b) 5° C/s. Source: Ref 9

Figure 5 compares the predicted and measured austenite grain size in vanadium-microalloyed steel specimens heated from 1100 to 1250 °C at two different heating rates (5 °C/s and 30 °C/s). In this case, grain-size predictions were made using the time discretization approximation discussed previously. Austenite grain-size measurements were made at 1100, 1200, and 1250 °C. The model predictions match well with the measurements.

In addition to the simplified analytical approaches discussed previously, newer mesoscale modeling techniques, such as the cellular automata method, Monte Carlo simulation, and phase-field approach, are also being used by researchers to model grain growth under nonisothermal conditions. Mishra and DebRoy (Ref 25) present a comprehensive overview of the merits and limitations of these techniques for simulating grain growth under nonisothermal conditions

Recrystallization

Recrystallization is the primary mechanism by which austenite grains become refined during hot working. As discussed previously, recrystallization that starts and finishes after deformation is complete is called static recrystallization, while recrystallization that occurs during deformation is called dynamic recrystallization. A third type of recrystallization that starts during deformation (formation of recrystallization nuclei) and undergoes completion after deformation is referred to as metadynamic recrystallization. The type of recrystallization that occurs during a hot rolling process depends on the deformation conditions, particularly temperature, accumulated strain, strain rates, and interpass times.

A commonly used measure to differentiate static recrystallization from dynamic or metadynamic recrystallization is the strain corresponding to the peak of the flow-stress curve. There is a fair amount of published literature on how flow stress changes under hot deformation conditions (Ref 16, 26, 27). A typical flow-stress curve for steel at an elevated temperature and a constant strain-rate condition is illustrated in Fig. 6. In the initial part of the curve, the increase in dislocation density, along with the pileup of dislocations at the grain boundaries, causes the stress to rise with strain. In other words, the material strain hardens. The strain-hardening rate, $d\sigma/d\varepsilon$, decreases with increasing strain due to dynamic recovery, which results in annihilation of dislocations in the deformed grains. Under certain conditions (high temperatures and low strain rates), the stress-strain curve displays a clear peak at the peak strain (ε_p), after which the flow stress decreases due to dynamic recrystallization and eventually stabilizes to a value (σ_{ss}^{drex}) between the yield stress and the peak stress. The other steady-state stress, σ_{ss}^* shown in Fig. 6, is the estimated steady-state stress if recovery were the only softening mechanism present in the material. The critical strain (ε_c) for the onset of dynamic recrystallization is understood to be somewhat lower than the peak strain (ε_p) (Ref 28) and depends on temperature, strain rate, and the grain size. While it is difficult to accurately pinpoint the strain for the start of dynamic recrystallization, researchers have found that the following relationship between critical strain and peak strain works very well for many steels:

$$\varepsilon_c = A\varepsilon_p \qquad \text{(Eq 4)}$$

where A is a material-specific constant and generally has a value between 0.5 and 0.9. Studies have shown that an A-value of 0.83 works as a good estimate for many steels.

The peak strain is generally modeled as an equation of the following form (Ref 28–30):
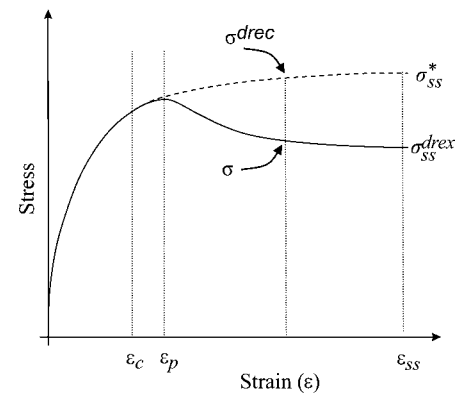
Fig. 6 Typical flow-stress curve under hot working conditions

$$\varepsilon_p = Ad_0^m \left(\dot{\varepsilon} \exp\left(\frac{Q}{RT}\right) \right)^n \qquad \text{(Eq 5)}$$

where d_0 is the as-heated austenite grain size, T is the absolute temperature, $\dot{\varepsilon}$ is the strain rate, and A, Q, m, and n are material-specific constants. A few published models for peak strain and critical strain for various steels are listed in Table 2.

Modeling recrystallization involves developing models for static, dynamic, and metadynamic recrystallization kinetics as functions of the processing conditions and the existing microstructure. Several models for recrystallization kinetics have been developed based on the well-known Johnson-Mehl-Avrami-Kolmogorov (JMAK) equation:

$$X = 1 - \exp\left(-0.693 \left(\frac{t}{t_{0.5}}\right)^n \right) \qquad \text{(Eq 6)}$$

where X is the fraction recrystallized at time t, $t_{0.5}$ is the time for 50% recrystallization, and n is the time exponent and is assumed to be a material-specific constant. Note that the

aforementioned general equation can be applied to static as well as metadynamic recrystallization processes. While static and metadynamic recrystallization are modeled as functions of time, dynamic recrystallization is commonly described using strain as an independent variable instead of time. The general equation describing dynamically recrystallized volume fraction is:

$$X_{DRX} = 1 - \exp\left(C\left(\frac{\varepsilon - \varepsilon_c}{\varepsilon_p}\right)^n\right) \qquad \text{(Eq 7)}$$

where X_{DRX} is the fraction dynamically recrystallized at strain ε, ε_c is the critical strain for the onset of dynamic recrystallization, ε_p is the strain corresponding to the peak of the flow stress-strain curve, n is an exponent and is assumed to be a material-specific constant, and C is a material-specific constant. Similar other equations for dynamic recrystallization have also been used by other researchers (Ref 34). Due to the complexity of modeling dynamic recrystallization, many instead prefer to model dynamic recrystallization as metadynamic recrystallization.

Modeling Recrystallization Kinetics

Recrystallization kinetics (static and metadynamic) are characterized by the time for 50% recrystallization and are a function of the strain, strain rate, temperature, and grain size, as discussed previously. To determine the time for 50% recrystallization, single-hit deformation tests under hot working conditions followed by quenching to lock the microstructure can be performed. However, these single-hit tests must be followed by extensive metallography and measurements to determine the recrystallized fraction in each test specimen. On the other hand, indirect measurements, such as a double-hit deformation test, can give instantaneous estimates of the fraction recrystallized based on measurements on the stress-strain curve. Figure 7 shows a typical stress-strain curve during a double-hit compression test. σ_1 and σ_2 are the yield stresses for the first and second hits. The fraction softening, F_s, occurring in the interval between the two hits is given by:

$$F_s = \frac{\sigma_m - \sigma_2}{\sigma_m - \sigma_1} \qquad \text{(Eq 8)}$$

where σ_m is the stress at the end of the first hit, as shown in Fig. 7. It is generally assumed that an initial 20% softening occurs due to static recovery. The fraction recrystallized is therefore estimated by:

$$F_x = \frac{F_s - 0.2}{0.8} \qquad \text{(Eq 9)}$$

Some researchers believe that examination of the microstructure is a more accurate way of measuring the fraction recrystallized and that the softening due to static recovery can vary between 15 and 30% (Ref 35–36). Therefore,

Table 2 Models for peak strain and critical strain

Steel(a)	Peak strain	Critical strain	Reference
V-HSLA	$\varepsilon_p = 1.20 \times 10^{-3} d_0^{0.27} \dot{\varepsilon}^{0.18} \exp\left(\frac{50,400}{RT}\right)$	$\varepsilon_c = 0.83\varepsilon_p$	9
Cr-Mo steel (4120)	$\varepsilon_p = 3.5 \times 10^{-3} d_0^{0.15} \dot{\varepsilon}^{0.17} \exp\left(\frac{46,326}{RT}\right)$	$\varepsilon_c = 0.83\varepsilon_p$	16
V-HSLA	$\varepsilon_p = 6.05 \times 10^{-4} d_0^{0.32} \dot{\varepsilon}^{0.17} \exp\left(\frac{55,530}{RT}\right)$	$\varepsilon_c = 0.83\varepsilon_p$	16
A36, DQSK	$\varepsilon_p = 1.32 \times 10^{-2} d_0^{0.174} \dot{\varepsilon}^{0.165} \exp\left(\frac{24,360}{RT}\right)$	$\varepsilon_c = 0.83\varepsilon_p$	31
Medium C-Mn	$\varepsilon_p = 6.97 \times 10^{-4} d_0^{0.3} \left(\dot{\varepsilon} \exp\left(\frac{300,000}{RT}\right)\right)^{0.17}$	$\varepsilon_c = 0.81\varepsilon_p$	24
316 stainless steel	$\ldots$	$\varepsilon_c = 1.97 \cdot \left(\dot{\varepsilon} \exp\left(\frac{413,000}{RT}\right)\right)^{0.057}$	32
Low-carbon steel	$\varepsilon_p = 4.83 \times 10^{-3} d_0^{0.09} \left(\dot{\varepsilon} \exp\left(\frac{274,362}{RT}\right)\right)^{0.15}$	$\varepsilon_c = 0.85\varepsilon_p$	33
High-carbon steel	$\varepsilon_p = 1.84 \times 10^{-3} d_0^{0.24} (Z)^{0.15}$ $where$ $Z = \dot{\varepsilon} \exp\left(\frac{Q_{def}}{RT}\right)$	$\varepsilon_c = 0.85\varepsilon_p$	33
Medium-carbon	$\varepsilon_p = 2.73 \times 10^{-3} d_0^{0.20} (Z)^{0.15}$ $where$ $Z = \dot{\varepsilon} \exp\left(\frac{Q_{def}}{RT}\right)$	$\varepsilon_c = 0.85\varepsilon_p$	33

(a) V-HSLA, Vanadium micro alloyed high strength low alloy; DQSK, drawing quality special killed
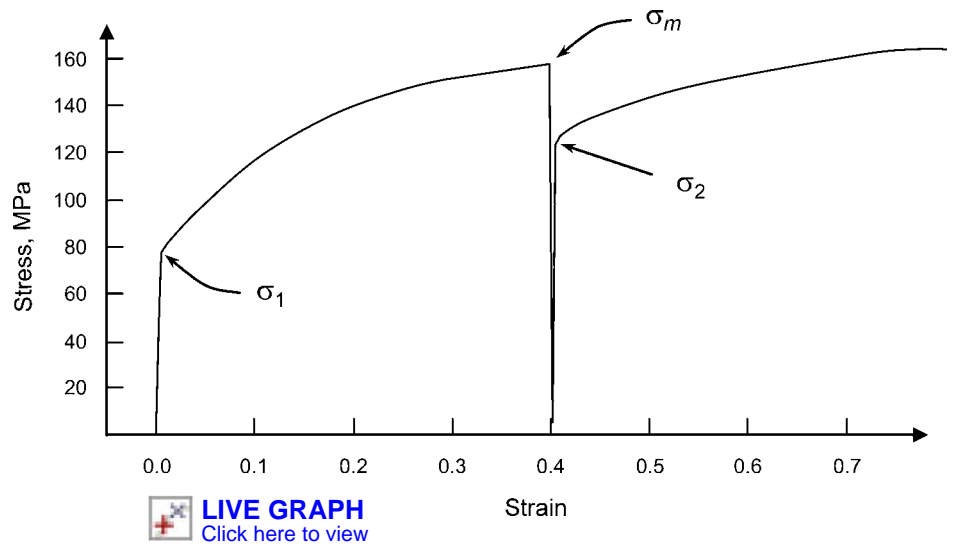
Fig. 7 Typical stress-strain curve in a double-hit compression test

the assumption of 20% initial softening due to recovery may not yield accurate results. However, the efficiency and ease afforded by double-hit tests make it attractive, particularly when a large number of experiments must be conducted. In any case, both techniques are known to yield reasonably good models and are used extensively.

A typical test procedure for the double-hit compression test is shown in Fig. 8. Using the grain-growth model, the reheat temperature and the corresponding hold time necessary to obtain a desired grain size is determined. The specimen is then heated to this temperature and held there until the desired grain size is obtained. The specimen is then cooled to the test temperature (T). After homogenizing at the test temperature, the specimen is axially compressed a predetermined amount and held at the test temperature for a certain period of time (t) before being compressed again. While planning these experiments for static recrystallization, care must be taken to ensure that the strain attained in the first hit is lower than the critical strain, ε_c to ensure that dynamic recrystallization is not initiated during the first hit of this test. Conversely, when
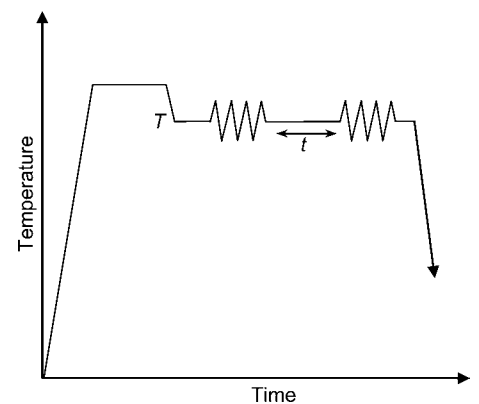
Fig. 8 Depiction of a double-hit compression test

studying metadynamic recrystallization, care is taken to assure that the applied strain of the first hit exceeds the critical strain for the onset of dynamic recrystallization (Eq. 4). The stress-strain data during these tests are recorded using a high-speed data acquisition system. The fraction recrystallized in each test is computed using the stress-strain data via Eq. 8 and 9.
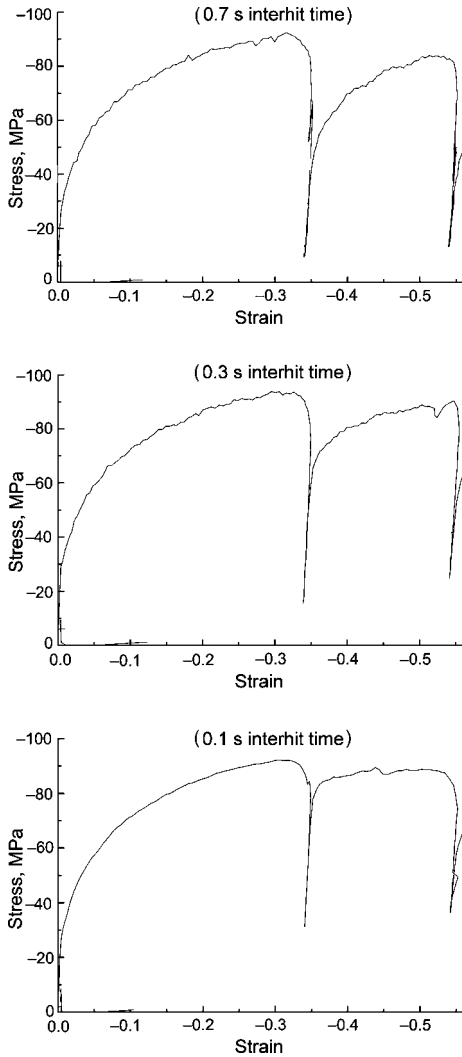
Fig. 9 Stress-strain curves in double-hit compression tests conducted at 1100 °C at a strain rate of 1.0/s with an as-heated grain size of 135 10⁻⁶ m

LIVE GRAPH
Click here to view

Figure 9 shows three stress-strain curves recorded during one such study (Ref 9). These tests were conducted at 1100 °C at a strain rate of 1.0/s, with a compression strain of 0.35 in the first hit and a strain of 0.20 in the second hit. The only difference among these tests was the interhit time. From these tests it is clear that longer interhit time leads to more recrystallization and consequently results in lowering of the yield stress during the second hit.

Model for Static Recrystallization. After collecting data from the previously described double-hit compression tests, regression analysis is used to fit the model. First, the value of time exponent n is determined from a least-squares analysis on the fraction recrystallized measurements. Figure 10 shows a plot of $\ln(\ln(1/(1 - X)))$ versus $\ln(t)$ for different static recrystallization test conditions used in one study (Ref 9). The slope of the lines is the value of n. Under static recrystallization conditions, typical values of n for low-alloy and microalloyed steels range
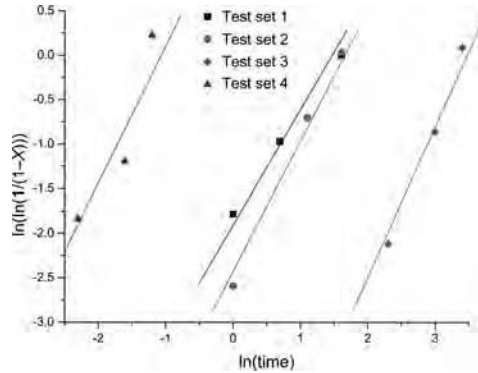
Fig. 10 Determining time exponent, n, in the Avrami-type equation (Eq 10) from the slope of lines. Source: Ref 9

LIVE GRAPH
Click here to view

from 0.7 to 1.50, as reported by Medina and Mancilla (Ref 37). Beynon and Sellars (Ref 38) have reported an n-value of 1 to 2 for carbon-manganese steels. Matja et al. (Ref 39) have also reported that n-values for microalloyed steels are generally between 1.0 and 2.0.

After determining the time exponent, n, the Avrami-type equation is used to determine the time for 50% recrystallization ($t_{0.5}$) for the tests conducted. An equation for $t_{0.5}$ as a function of as-heated grain size, strain, strain rate, and temperature is then developed using a regression analysis:

$$t_{0.5} = A_2 \varepsilon^l \dot{\varepsilon}^m d_0^n \exp\left(\frac{Q}{RT}\right) \qquad \text{(Eq 10)}$$

where ε is the strain, $\dot{\varepsilon}$ is the strain rate (per second), d_0 is the as-heated grain size (microns), T is the absolute temperature (K), and A_2, Q, l, m, and n are material-specific constants.

A few select published static recrystallization models for various steels are listed in Table 3. Fraction recrystallization of 4120 (UNS G41200) steel and 1538V (V-HSLA) steel under various test conditions is plotted on charts shown in Fig. 11 and compared with model predictions (Ref 16).

Model for Metadynamic Recrystallization. To develop models for metadynamic recrystallization, double-hit compression tests can be used, similar to those used for modeling static recrystallization. Care must be taken to ensure that the strain imparted in the first hit is higher than the critical strain for the start of dynamic recrystallization. This ensures that dynamic recrystallization is initiated before the first hit ends and that the softening observed during the period between the two hits is primarily due to metadynamic recrystallization. Table 4 lists a few published models for metadynamic recrystallization.

Comparing the exponents of various terms in Tables 3 and 4, it can be seen that the effect of grain size and strain is significantly less in metadynamic recrystallization as compared to that in static recrystallization. It has generally been noted in published empirical studies that static and metadynamic recrystallization follow

two different kinetic laws. Static recrystallization depends more on strain and only slightly on strain rate, whereas metadynamic recrystallization depends more on strain rate and only slightly on strain (Ref 43).

Recrystallized Grain Size. The size of the recrystallized grains is dependent on the strain, initial grain size prior to deformation, strain rate, and temperature. Experiments are performed with strain, strain rate, grain size, and temperature as the control variables. Figure 12 depicts the experimental procedure used to develop a model for recrystallized grain size under static and metadynamic recrystallization conditions. To develop models for recrystallized grain size, the specimens are first heated to obtain the desired grain size, after which they are cooled to the test temperature (T_t) and deformed to a certain strain at a constant strain rate. After deformation, the specimens are held at T_t to allow complete recrystallization to take place. For practical purposes, 95% recrystallization is considered complete recrystallization. The time for complete recrystallization (95% recrystallization) is determined from the recrystallization kinetics models discussed earlier. The specimens are then cooled rapidly to 600 °C and held there for a short time to allow proeutectoid ferrite to form along the prior-austenite grain boundaries before being quenched to room temperature. The microstructures in the specimens are later examined to measure the grain size.

Several models of recrystallized grain size (d_{rex}) for static and metadynamic recrystallization for various steels are listed in Tables 3 and 4. From these models, it is evident that the higher the temperature of deformation and the lower the strain rate, the larger are the recrystallized grains formed during high-temperature deformation.

Dealing with Partial Recrystallization Conditions. Rolling is a multiple-stage deformation process. Other than dynamic recrystallization, the primary mode of grain refinement is the static and metadynamic recrystallization that occurs during the period between two successive deformations. Often during the rolling process, the time in the interstand is not sufficient for complete recrystallization to occur. Under such circumstances, only partial recrystallization occurs, resulting in a mixed microstructure, one portion of which consists of fine, strain-free recrystallized grains, while the other portion consists of deformed grains that are usually relatively coarser and elongated compared to the newly recrystallized grains. In other words, some amount of strain is retained in the structure when it enters the next deformation pass, while, at the same time, the structure is nonuniform. This presents a critical question about how to handle the two fractions.

To illustrate the effect of partial recrystallization, a physical simulation experiment using isothermal three-hit compression tests has been used. In the three-hit compression tests, the first interhit time (t_1) is deliberately kept short to cause partial recrystallization. The second hit

Table 3 Published models for static recrystallization kinetics and recrystallized grain size

Steel(a)	Fraction recrystallized	Time for 50% recrystallization	Recrystallized grain size	Reference
C-Mn	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^k\right]$	$t_{0.5} = 2.5 \times 10^{-19}\varepsilon^{-4}d_0^2\exp\left(\frac{300,000}{RT}\right)$	$d_{SRX} = 0.743\cdot\varepsilon^{-1.0}d_0^{0.67}$	38
Low C-Mn	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)\right]$	$t_{0.5} = 2.3 \times 10^{-15}\varepsilon^{-2.5}d_0^2\exp\left(\frac{230,000}{RT}\right)$	$d_{SRX} = 343\cdot\varepsilon^{-0.5}d_0^{0.4}\exp\left(\frac{-45,000}{RT}\right)$	23, 40
Medium C-Mn	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)\right]$	$t_{0.5} = 1.14 \times 10^{-13}\varepsilon^{-3.8}\dot{\varepsilon}^{-0.41}\exp\left(\frac{252,000}{RT}\right)$	$d_{SRX} = 0.5\cdot\varepsilon^{-0.67}d_0^{0.67}$	24
316 stainless	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^{0.31}\right]$	$t_{0.5} = 3.72 \times 10^{-8}\varepsilon^{-2.2}\dot{\varepsilon}^{-0.66}\exp\left(\frac{137,000}{RT}\right)$	$d_{SRX} = 0.57\cdot d_0(t_0)^{0.042}$	32
V-HSLA (1538V)	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^{1.46}\right]$	$t_{0.5} = 1.6 \times 10^{-4}\varepsilon^{-1.17}\dot{\varepsilon}^{-0.28}d_0^{0.26}\exp\left(\frac{84,362}{RT}\right)$	$d_{SRX} = 2.14\cdot\varepsilon^{-0.76}d_0^{0.81}\exp\left(\frac{-21,000}{RT}\right)$	16
Cr-Mo steel (4120)	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)\right]$	$t_{0.5} = 1.0 \times 10^{-11}\varepsilon^{-2.36}\dot{\varepsilon}^{-0.24}d_0^{0.20}\exp\left(\frac{241,000}{RT}\right)$	$d_{SRX} = 1.3\cdot\varepsilon^{-1.4}d_0^{0.46}\exp\left(\frac{-7,200}{RT}\right)$	16
High C-Mn	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^{1.3}\right]$	$t_{0.5} = 4.2 \times 10^{-16}\varepsilon^{-2}\dot{\varepsilon}^{-0.34}d_0^2\exp\left(\frac{270,000}{RT}\right)$	$d_{SRX} = 0.3\cdot\varepsilon^{-0.9}d_0^{0.67}$	41
A36	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^2\right]$	$t_{0.5} = 8.31 \times 10^{-15}\varepsilon^{-1.5}\dot{\varepsilon}^{-0.33}d_0^{1.5}\exp\left(\frac{263,000}{RT}\right)$	...	31
DQSK	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^2\right]$	$t_{0.5} = 5.22 \times 10^{-13}\varepsilon^{-0.67}\dot{\varepsilon}^{-0.33}d_0\exp\left(\frac{248,000}{RT}\right)$	...	31
Ti-Nb-HSLA	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^{0.6}\right]$	$t_{0.5} = 1.18 \times 10^{-11}\varepsilon^{-1.31}\dot{\varepsilon}^{-1.07}d_0^{0.353}\exp\left(\frac{240,000}{RT}\right)$ for $T < 1423\ K$ $t_{0.5} = 3.81 \times 10^{-4}\varepsilon^{-1.31}\dot{\varepsilon}^{-1.07}d_0^{0.353}\exp\left(\frac{35,400}{RT}\right)$ for $T > 1423\ K$	$d_{SRX} = 8.49\cdot\varepsilon^{-0.872}d_0^{0.33}\exp\left(\frac{-10,600}{RT}\right)$	42

(a) V-HSLA, Vanadium microalloyed high strength low alloy; DQSK, drawing quality special killed
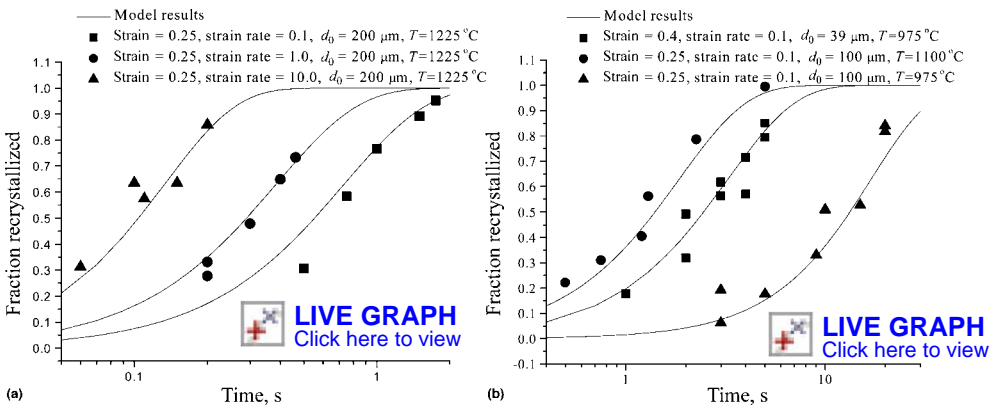
Fig. 11 Recrystallization behavior and model results for (a) 4120 and (b) 1538V steel. Source: Ref 16

is followed by a third hit, with an interhit time (t_2) between the two, as shown in Fig. 13. The amount of recrystallization during the second interhit period is measured using the same procedure as used in the double-hit compression tests. Figures 13(b) and (c) show examples of flow stress-strain data recorded from two such three-hit tests conducted. The deformation in each hit was equivalent to an increment in strain of 0.15. The noise in the data is because of the low strain rate used in these particular tests.

In Fig. 13(b), the low temperature and the low strain in the first hit led to very little softening in the 10 s between the first hit and the second hit. Consequently, the retained strain at the beginning of the second was large and is additive to the strain in the second hit. The net result was that the accumulated strain in the specimen was greater, which led to much faster recrystallization after the second hit. This is reflected in the significant softening that is seen in the stress-strain curve for the third hit. Note

that the second interhit time is only 3 s, while the first interhit time was 10 s.

In Fig. 13(c), the significantly larger duration of the interstand (30 s) led to a larger fraction of recrystallization to occur in the first interstand. As a result, the retained strain at the beginning of the second hit was very small. After, the second hit, the net accumulated strain was much smaller than that in Fig. 13(a). Consequently, the recrystallization kinetics after the second hit were significantly slower. This is reflected in very little softening that is seen in the stress-strain curve for the third hit. It may be noted that the second interhit time in this case is the same as that in the earlier case, that is, 3 s.

Partial recrystallization may be easily encountered in a hot strip mill or hot bar rolling mill. For example, a hot strip mill or a hot bar rolling mill may use a reversing mill for roughing passes and a continuous rolling mill for finishing passes. In the reverse mill, the temperature is generally high, and the interpass times

are long. This generally allows for full recrystallization to occur between the passes. Also, due to high temperature and relatively large strains and low strain rates, there is a possibility for dynamic recrystallization to occur during deformation. On the other hand, in the continuous finishing mill where the temperatures are lower and the interpass times are shorter, partial or no recrystallization may occur at one finishing mill stand. However, accumulated strain from pass to pass could eventually lead to dynamic recrystallization or metadynamic recrystallization in a later pass.

A number of approaches have been proposed to handle partial recrystallization. One of the most commonly used approaches is to treat the microstructure as an aggregate. In this approach, the retained strain and the effective grain sizes are determined using the rule of mixtures, as follows:

$$\varepsilon_{ret} = \varepsilon\cdot(1 - X) \qquad \text{(Eq 11)}$$

$$d_{eff} = \left(X\cdot d_{rex}^3 + (1 - X)\cdot d_0^3\right)^{1/3} \qquad \text{(Eq 12)}$$

where X is the fraction recrystallized, ε_{ret} is the retained strain, d_{eff} is the effective grain size, d_0 is the initial as-heated grain size, and d_{rex} is the recrystallized grain size. This approach has been adopted in many studies (Ref 3, 30, etc.). It is easy to implement but may give rise to inconsistency in the modeling of recrystallization kinetics. This is because the grain size and strain that are used for calculating recrystallization kinetics in the subsequent pass are not representative of either of the fractions (recrystallized and unrecrystallized fractions) of the material. It is an average of fractions that are expected to behave differently in terms of recrystallization kinetics. However, the error is

Table 4 Published models for metadynamic recrystallization kinetics and recrystallized grain size

Steel(a)	Fraction recrystallized	Time for 50% recrystallization	Recrystallized grain size	Reference
Low C-Mn	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^{1.5}\right]$	$t_{0.5} = 1.1\left(\dot{\varepsilon}\exp\left(\frac{300,000}{RT}\right)\right)^{-0.8}\exp\left(\frac{230,000}{RT}\right)$	$d_{MDX} = 2.6 \times 10^4 \cdot \left[\dot{\varepsilon}\exp\left(\frac{300,000}{RT}\right)\right]^{-0.23}$	23, 39
Medium C-Mn	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^{1.5}\right]$	$t_{0.5} = 0.44 \cdot \dot{\varepsilon}^{-0.8}$	$d_{MDX} = 6.8 \times 10^4 \cdot \left[\dot{\varepsilon}\exp\left(\frac{300,000}{RT}\right)\right]^{-0.27}$	24
316 stainless	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^{0.237}\right]$	$t_{0.5} = 2.14 \times 10^{-9}d_0\dot{\varepsilon}^{-0.85}\exp\left(\frac{155,000}{RT}\right)$	$d_{MDX} = 1.57d_0^{0.99} \cdot \left[\dot{\varepsilon}\exp\left(\frac{460,000}{RT}\right)\right]^{-0.0327}$	32
V-HSLA	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^{1.28}\right]$	$t_{0.5} = 5.78 \times 10^{-8}\varepsilon^{-1.17}d_0^{0.26}\dot{\varepsilon}^{-0.28} \cdot \exp\left(\frac{154,614}{RT}\right)$	$d_{MDX} = 1.22 \times 10^4 \cdot \left[\dot{\varepsilon}\exp\left(\frac{330,000}{RT}\right)\right]^{-0.193}$	16
Cr-Mo steel (4120)	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)\right]$	$t_{0.5} = 3.0 \times 10^{-5}\left(\dot{\varepsilon}\exp\left(\frac{320,000}{RT}\right)\right)^{-0.5}\exp\left(\frac{270,000}{RT}\right)$	$d_{MDX} = 2900 \cdot \left[\dot{\varepsilon}\exp\left(\frac{330,000}{RT}\right)\right]^{-0.16}$	16
High C-Mn	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^{1.2}\right]$	$t_{0.5} = 1.5 \times 10^{-6}\dot{\varepsilon}^{-0.61}\exp\left(\frac{146,000}{RT}\right)$	$d_{MDX} = 2.55 \times 10^4 \cdot \left[\dot{\varepsilon}\exp\left(\frac{300,000}{RT}\right)\right]^{-0.22}$	41
Low carbon (A36)	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)\right]$	$t_{0.5} = 2.13 \times 10^{-6}\dot{\varepsilon}^{-0.67} \cdot \exp\left(\frac{133,000}{RT}\right)$	...	31
DQSK	$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)\right]$	$t_{0.5} = 1.54 \times 10^{-6}\dot{\varepsilon}^{-0.67} \cdot \exp\left(\frac{127,000}{RT}\right)$	...	31

(a) V-HSLA, Vanadium microalloyed high strength low alloy; DQSK, drawing quality special killed
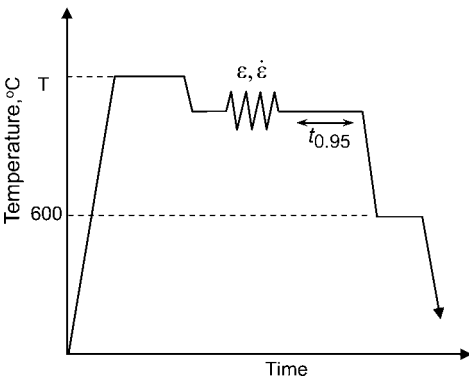
Fig. 12 Test procedure for modeling recrystallized grain size

small if the fraction recrystallized is either close to zero or close to 1.0. The latter is usually the case in hot rolling with reversing mills. A variation of the rule of mixtures was used by Jin et al. (Ref 6) following Hodgson and Gibbs (Ref 4), who reported that the following model predicted the softening kinetics more accurately than the rule of mixtures for low-carbon steels:

$$\varepsilon_{ret} = 0.5 \cdot \varepsilon \cdot (1 - X) \qquad \text{(Eq 13)}$$

Another approach is to treat the recrystallized and unrecrystallized fractions independently (Ref 44). This allows close representation of the true behavior of the material. However, the number of fractions to be handled increases very rapidly. If three are m-number of passes, the total number of fractions can be as high as 2^m. Typically in hot rolling, 10 to 15 passes are fairly common. This approach therefore calls for a tremendous amount of computer memory and time and is generally not widely used. Yanagimoto et al. (Ref 45) proposed a variation of Karhausen's model. In this approach, the number of fractions increases linearly versus exponentially in Karhausen's model and therefore requires considerably less memory. However, as with the rule of mixtures, considerable

approximation is involved, and the true behavior of the system may not be represented.

Recrystallization Kinetics under Changing Temperature Conditions. Recrystallization models, similar to grain-growth models, are generally developed under isothermal conditions. Temperature changes continuously during hot rolling. A common approach to handle continuously changing temperature conditions is based on Scheil's additivity rule (Ref 46). For the Scheil additivity rule to be applicable, the reaction must be isokinetic. In other words, all parameters in the equation must have the same temperature dependence. This is generally true in the case of recrystallization. In this technique, the temperature history curve in the interstand region is divided into several small time steps. The temperature is assumed to be constant in each of these steps. In other words, the cooling curve is approximated by a large number of isothermal steps. Figure 14 illustrates how this technique works. The figure shows two sigmoidal curves representing recrystallization kinetics for two successive iteration steps, i and $i+1$, during which the temperatures are assumed to be constant at T_i and T_{i+1}, respectively. The two curves have different start and finish times for recrystallization because of the change in temperature. At the end of time step i, the recrystallized fraction is X_i. For the next time step $i + 1$, the fraction transformed, X_{i+1}, must be computed. The procedure for computing X_{i+1} is described as follows:

1. First, a fictitious time, $t_{i+1, fict}$, is computed as the time needed at the new temperature, T_{i+1}, for transformation of the fraction, X_i, that was completely transformed at the end of the previous iteration.
2. Then, the new time step is added to $t_{i+1, fict}$, and at this new time, $t_{i+1, fict} + \Delta t$, the new fraction transformed, X_{i+1}, is determined.

This incremental computation procedure is carried out through all the time steps in the interstand region to determine the recrystallized fractions and the corresponding recrystallized grain sizes.

Recrystallization Kinetics under Changing Strain-Rate Conditions. From the models discussed in the previous sections, it is evident that strain rate has a significant effect on recrystallization kinetics. Increasing the strain rate increases the rate of recrystallization. The driving force for recrystallization is the strain energy stored in the material. The higher strain rate gives rise to higher dislocation density, higher stored energy, and consequently, faster recrystallization. Like temperature, strain rate changes continuously in a rolling process from the entry into the rolls to the exit of the roll bite. The recrystallization kinetics models are generally developed under constant strain-rate conditions. To use these models for modeling microstructural evolution in a rolling process, a procedure must be adopted to determine recrystallization kinetics under changing strain-rate conditions. Despite a great amount of research on recrystallization kinetics, the effect of strain-rate variations on recrystallization kinetics remains unclear (Ref 47). Many studies have used the average strain rate during deformation to model recrystallization kinetics. In a recent study, Poliak and Jonas (Ref 47) report that variation in strain rate along the roll bite significantly influences the critical strain for dynamic recrystallization. They also conclude that approximating the variable strain rate in the roll bite with a mean strain rate of the pass gives reasonable estimates for calculating critical strain, using models developed under constant strain-rate conditions.

Modeling Austenite Decomposition

As the hot-worked steel is cooled, austenite decomposes to form transformation products, such as ferrite, pearlite, bainite, and martensite. Modeling of phase transformation is considered very important with regards to prediction of microstructure and mechanical properties in the hot-worked product. Numerous studies have been published on this subject (Ref 12–14, 48–53).
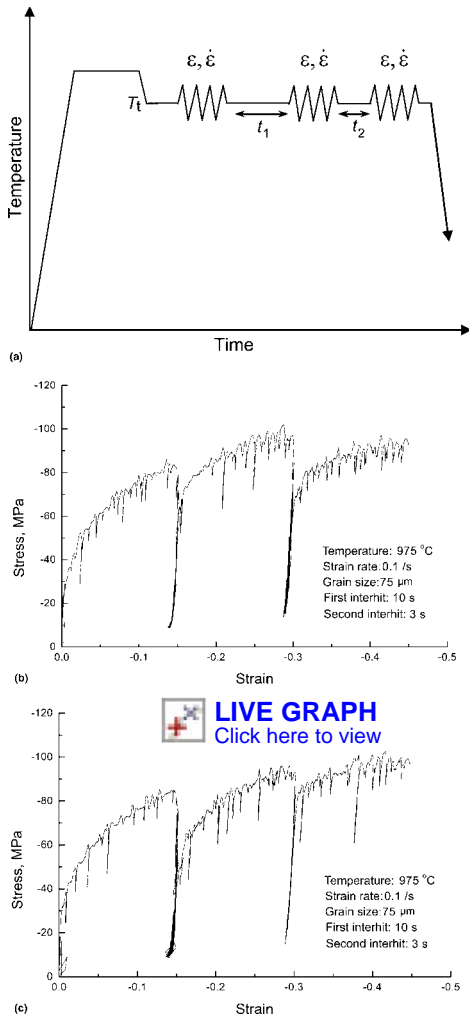
Fig. 13 Three-hit compression test. (a) Test procedure for an isothermal three-hit test. Examples of stress-strain data. (b) First interhit time 10 s and second interhit time 3 s. (c) First interhit time 30 s and second interhit time 3 s

Fig. 14 Adapting JMAK equation to nonisothermal conditions. (a) Iteration step *i*. (b) Iteration step *i* + 1

While simulation of austenite evolution during hot rolling has been the subject of research for the past three decades, only recently has simulation of phase transformations become the focus of research (Ref 53).

In the case of hypoeutectoid steels, the formation of proeutectoid ferrite, pearlite, and bainite occurs by means of diffusional transformation. Diffusion is a time- and temperature-dependent phenomenon. Diffusional transformation proceeds by a nucleation and growth mechanism and can be described by the JMAK equation similar to the one used to describe recrystallization kinetics. The JMAK equation estimates the amount of transformed product, $X(t)$, at time, t, at a constant temperature:

$$X(t) = 1 - \exp\left[-\beta(t)^k\right] \qquad \text{(Eq 14)}$$

where β and k are parameters derived from the consideration of nucleation and growth rates.

β and k are functions of the transformation product, steel chemistry, temperature, and initial austenite microstructure, and their values or functions are generally obtained by means of empirical studies. For ferrite, the value of exponent k is generally found to be approximately 1, while for pearlite and bainite, the exponent is computed from isothermal transformation (IT) data. Tamura et al. (Ref 51) have generalized and proposed that $k = 1$ for ferrite, $k = 2$ for pearlite, and $k = 4$ for bainite. These generalized values for ferrite and pearlite appear to be consistent with those reported in other published studies on this subject. However, for bainite, values of k from 2 to 7 have been reported (Ref 53).

While ferrite, pearlite, and bainite form by means of time-dependent diffusional transformations, the progress of martensitic transformation is independent of time and depends only on the temperature. The transformation of martensite is described by a semiempirical relation proposed by Koistinen and Marburger (Ref 50). Using precise x-ray techniques to estimate the amount of martensite formed during transformation, they have proposed the following equation for the fraction of martensite formed, $X(T)$, at a given temperature, T:

$$X(T) = 1 - \exp\left[-1.10 \times 10^{-2}(M_s - T)\right] \qquad \text{(Eq 15)}$$

where M_s denotes the martensite start temperature. The fraction of martensite therefore depends only on the undercooling below the martensite start temperature. Similar relationships have been developed by other researchers (Ref 34) to model transformation to martensite.

The kinetics of diffusional phase transformations can be described by the IT diagram (also known as the time-temperature-transformation, or TTT, diagram). Dilatometry is one of the techniques most commonly used to determine the start and end of phase transformations in steels. To construct a TTT diagram, a sample of steel is first austenitized at a certain temperature, then cooled rapidly to a predetermined lower temperature and held at that temperature until transformation is completed. Progress of transformation is monitored via dilation strain measurements in the sample. Dilational strain under isothermal conditions is primarily due to volumetric changes associated with phase transformations. From the analysis of the dilatometry data and examination of the resulting microstructure in the specimen, it is possible to identify the starting point and completion of phase transformations under different processing conditions. Generally, a large number of such dilatometry experiments at different temperatures are required to build a complete TTT diagram. Figure 15 shows a typical IT diagram for AISI 4130 steel (UNS G41300). Reference 54 is a good source for IT diagrams of many different steel grades.

For computer simulation of phase transformation, some form of digital representation of the various curves in the IT diagram is required. Researchers have used various techniques for digitizing the IT diagram. Some of the commonly used techniques are spline interpolation, function interpolation, and actual measurement (Ref 55). In the spline interpolation method, the curves in the IT diagram are represented by splines (curves defined mathematically by points called knots). On the other hand, the function interpolation method uses a time-temperature function to represent the IT curves. The actual measurement method is considered the easiest to digitize and is also considered to be the most accurate. It consists of tabulating the transformation start time (5%) and time for complete transformation (95%) for the various phases at different temperatures and using interpolation for intermediate values.

The JMAK equation used for describing transformation kinetics of diffusional transformations, such as recrystallization models, is valid only under isothermal conditions. However, transformation during cooling of the hot-rolled steel product is a nonisothermal process. To adapt the isothermal transformation JMAK models under nonisothermal conditions, a time discretization approach, based on Scheil's additivity principle is generally used. In this approach, the cooling curve is divided into several isothermal time steps. The calculations for fraction transformed are very similar to the calculations described earlier to model recrystallization under nonisothermal conditions.

At any instant during transformation, there can be up to five different phases and microconstituents present in the material, namely austenite, ferrite, pearlite, bainite, and martensite. For modeling austenite decomposition kinetics, two approaches have been proposed by researchers. The simultaneous transformation model (Ref 49) and the sequential transformation model. For simulation purposes, the sequential transformation model has been more commonly used, primarily because of its

simplicity. The primary assumption in this approach is that only one transformation reaction occurs at any instant during the transformation process. The sequence of transformation is ferrite, then pearlite, then bainite, and finally martensite, with distinct transition conditions separating these reactions.

Effect of Microstructure Evolution on Flow Stress

Flow stress of steels is known to be a function of strain, strain rate, temperature, chemistry, and microstructure. Modeling flow stress as a function of microstructure is a complex problem, because the microstructure constantly changes during a multiple-pass hot working process. As a result, traditionally flow stress has been modeled as a function of strain, strain rate, and temperature while ignoring the effect of microstructural evolution. Accurate modeling of flow stress involves taking into consideration different phenomena, such as work hardening, dynamic recovery, and dynamic recrystallization. Significant progress was made by Laasraoui and Jonas (Ref 26) in developing a microstructure-based flow stress model. Pauskar et al. (Ref 27) have extended this model to include the effect of austenite grain size.

A typical stress-strain curve was illustrated and described earlier in Fig. 6. In the initial part of the curve, the increase in dislocation density and pileup of dislocations at the grain boundaries causes the stress to rise with strain. The strain-hardening rate, $d\sigma/d\varepsilon$, decreases with increasing strain due to dynamic recovery, which results in annihilation of dislocations in the deformed grains. Under high temperatures and low strain-rate conditions, the stress-strain curve exhibits a clear peak at the peak strain (ε_p), after which, the flow stress decreases due to dynamic recrystallization. Dynamic recrystallization occurs only after a critical strain is exceeded. Following Laasraoui and Jonas's approach, the stress-strain curve can be divided into two regions:

- The region before the critical strain, where dynamic recovery is the predominant softening mechanism ($\varepsilon \leq \varepsilon_c$)
- The region after the critical strain, where dynamic recovery and dynamic recrystallization act together ($\varepsilon > \varepsilon_c$)

Under conditions that display a clear peak due to dynamic recrystallization, the stress-strain curve before ε_c is extrapolated to a hypothetical saturation stress (σ_{ss}^*), as illustrated in Fig. 6. This represents the saturation stress if recovery were the only softening mechanism during deformation. The difference between the saturation stress and the steady-state stress ($\sigma_{ss}^* - \sigma_{ss}^{drex}$) is a measure of the additional softening taking place due to dynamic recrystallization. The fractional softening due to dynamic recrystallization, X_d, is represented by:

$$X_d = \frac{\sigma^{drec} - \sigma}{\sigma_{ss}^* - \sigma_{ss}^{drex}} = 1 - \exp\left(-0.693\left(\frac{\varepsilon - \varepsilon_c}{\varepsilon_{0.5} - \varepsilon_c}\right)^n\right)$$
(Eq 16)

where n is a constant, $\varepsilon_{0.5}$ is the strain corresponding to 50% dynamic recrystallization, σ is the flow stress corresponding to strain ε, and σ^{drec} is the extrapolated hypothetical flow stress corresponding to strain ε if recovery were the only softening mechanism.

Flow stress versus strain data acquired at different temperature, strain rate, and austenite grain-size conditions can be analyzed using statistical methods to develop flow-stress models. Flow-stress models developed for a vanadium microalloy steel (V-HSLA) and AISI 4140 (UNS G41200) steel are summarized in Table 5. Figures 16(a) and (b) compare stress-strain curves predicted by the model with the experimentally measured flow-stress data. It is seen that the predictions are fairly good over the range of conditions considered in this study. The model is able to accurately model the effect of strain hardening, dynamic recovery, as well as dynamic recrystallization.

Application to a Multiple-Stage Deformation Process. In many processes, deformation progresses in multiple deformation stages. In hot rolling, for example, after deformation in a roll bite, there is a small time gap anywhere from a fraction of a second to a few seconds before the material undergoes deformation again in the next roll bite. During the period

between successive roll passes, metallurgical changes such as static recovery, static recrystallization, and grain growth change the flow characteristics of the material. A good flow-stress model not only predicts the flow stress accurately in a single-hit deformation process but also in a multiple-hit deformation process.

While modeling a multiple-deformation process, one of the important issues that must be addressed is the handling of partial recrystallization. In a reversing bar mill, the time available between two successive roll passes may be sufficient for complete recrystallization to occur. On the other hand, the interstand time in a continuous rolling mill may not be long enough for complete recrystallization. Researchers have proposed different approaches for handling partial recrystallization. One of the approaches uses the law of averages for the effective grain size and the retained strain (Eq 11 and 12). Researchers have used the average grain size and retained strain to determine recrystallization kinetics with reasonable success. However, this method yields relatively poor results when applied to flow-stress modeling (Ref 9). Another approach to handling partial recrystallization was suggested by Karhausen and Kopp (Ref 11, 44). It consists of treating the recrystallized and unrecrystallized fractions separately, unlike averaging as described previously. While this technique is known to be accurate and convenient for modeling a few stages of deformation, difficulties arise when modeling a large number of deformation passes

LIVE GRAPH
Click here to view
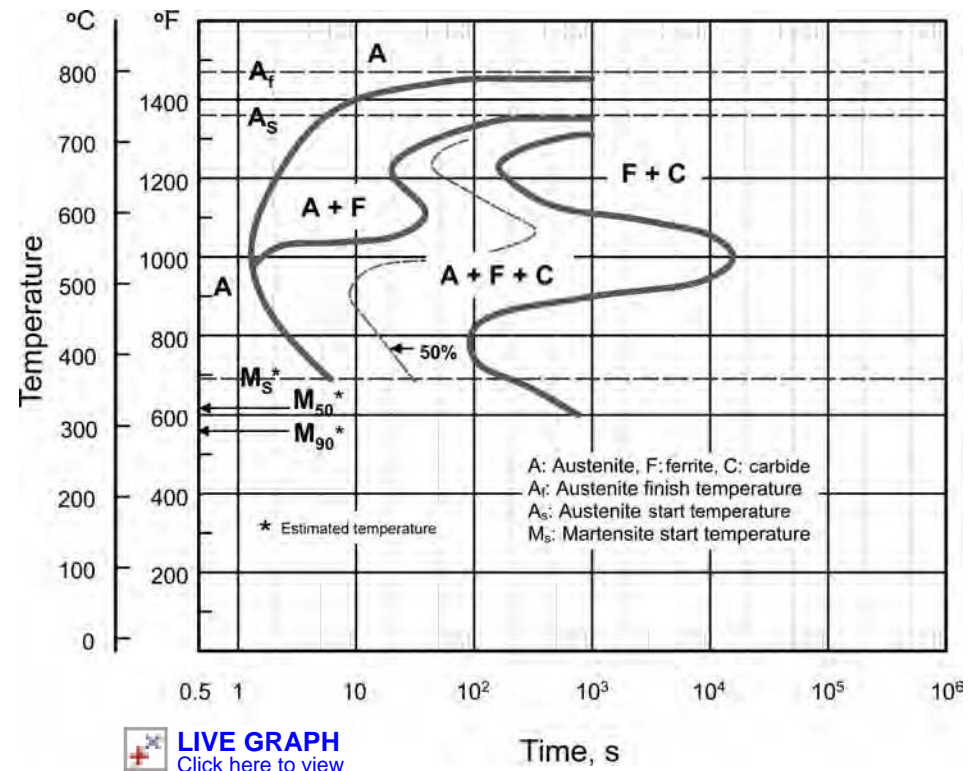
Fig. 15 Isothermal transformation (IT) diagram for AISI 4130 steel austenitized at 843 °C (1550 °F) and austenite grain size 9–10 ASTM. Source: Ref 54

due to rapidly increasing numbers of substructures. Therefore, a further modification is proposed to counter the difficulty in applying this method to multistage deformation. This approximation using the rule of mixtures is illustrated in Fig. 17. Consider the second hit, during which two substructures (1 and 2) exist in the material. After deformation, the grain sizes and the accumulated strains, in the two substructures are averaged to obtain the effective grain size and the average accumulated strain which are then associated with substructure 3. The static recrystallization model is then applied to substructure 3 to determine the fraction and size of the evolving substructure (substructure 4). Consequently, substructures 3 and 4 are carried forward to the third deformation step. This procedure is repeated over the entire deformation sequence. At any stage of deformation except the first, there are only two substructures that must be used in computing the flow stress, thereby making it computationally very efficient. This technique relies on the fact that grain size has a lesser effect on the flow stress at greater strains. Consequently, this technique has been found to work well for incremental strains larger than 0.1 under the conditions of temperature, strain rates, and grain sizes considered in this study. This method also makes use of the notion that the average grain size and average strain can provide reasonable predictions for recrystallization kinetics.

This technique has been tested by comparing predictions and measurements in five-hit compression tests conducted on a Gleeble thermomechanical testing machine. The compression test consisted of heating the test specimen to obtain a desired initial grain size. This was followed by cooling to the test temperature before being compressed to a strain of 1.0 in five equal strain increments. A short hold time was provided between two successive hits. Figures 18(a) and (b) illustrate two such cases studied with differing interhit times. The predicted flow-stress values agreed very well with the measurements, thereby validating not only the flow-stress model but also the austenite evolution models. Overall, the flow-stress model has been found to yield fairly accurate results under a variety of test conditions, including changing strain-rate deformation conditions. Additional details regarding this model can be found in Ref 9.

Physical Simulation in the Laboratory Environment

Physical simulation techniques attempt to replicate real-world processes on the laboratory scale in such a way that the resultant data can be used to explain, optimize, or solve real-world problems. Uniaxial hot compression tests and hot torsion tests are among the most common techniques used to simulate and study thermomechanical processes. These techniques make it possible to simulate off-line various

processing schedules for a particular material and to predict with reasonable accuracy the microstructure and mechanical properties that will result in the production environment. Such simulation techniques have been widely used (e.g., Ref 34, 56) to study and optimize rolling mill schedules.

The following is an example of a physical simulation experiment conducted on the Gleeble 3500 thermo-mechanical test system to validate an austenite evolution model. A five-stage hot uniaxial compression test is depicted in Fig. 18. In this test, a vanadium-microalloyed steel specimen was initially heated to a high temperature, T_0, to obtain an initial grain size of 200 μm (Ref 9). The specimen was then cooled to temperature T_1 and compressed to a strain of ε_1 at a certain strain rate. It was then cooled to a temperature T_2 at a predetermined cooling rate before being hit again. The specimen was thus successively compressed and cooled, as illustrated in Fig. 19. The grain sizes in this processing sequence were measured at different stages. The conditions under which one such physical simulation experiment was conducted are:

	Temperature, °C
T_1	1200
T_2	1175
T_3	1150
T_4	1125
T_5	1100
Incremental strain each hit	0.2
Strain rate each hit, s^{-1}	3
Cooling rate between hits, °C/s	1.67

The conditions of temperature, strains in each pass, and interhit times were chosen to reflect conditions encountered in a bar rolling process (Ref 9). Figure 20 compares the austenite grain size predicted using the microstructural evolution models with measurements at different stages. It is clearly seen that with each deformation step, the grains progressively become finer due to the ensuing recrystallization. It is also seen that the predicted austenite grain sizes match the measured austenite grain sizes quite well. The simplicity and ease of conducting such laboratory simulations allow one to conveniently investigate the effect of changing some of the process parameters (e.g., temperature or cooling rate) and ultimately identify the set of conditions that will yield the desired microstructure and properties in the rolled product. Such physical tests are also commonly used to validate microstructural evolution models, namely models for recrystallization, grain growth, and phase transformation.

Simulation Using Finite-Element Analysis

To model microstructural evolution in a hot working process, one must know the thermomechanical history of the material. Finite-element modeling (FEM) of the process provides a convenient and reasonably accurate way of obtaining the thermomechanical history of various points in the workpiece at different instants in time. The finite-element method has been widely recognized as an effective tool for the analysis of metalforming processes. The finite-

Table 5 Microstructure-dependent flow-stress models

Critical strain: $\varepsilon_c = 0.83 * \varepsilon_p$
For $\varepsilon < \varepsilon_c$:
$\sigma = \sigma^{drec} = \left[\sigma_{ss}^{*2} + \left(\sigma_0^2 - \sigma_{ss}^{*2} \right) e^{-\Omega\varepsilon} \right]^{0.5}$
For $\varepsilon > \varepsilon_c$:
$\sigma = \sigma^{drec} - \left[\sigma_{ss}^{*} - \sigma_{ss}^{drex} \right] \cdot \left[1 - \exp\left(-0.693\left(\frac{\varepsilon-\varepsilon_c}{\varepsilon_{0.5}-\varepsilon_c} \right)^n \right) \right]$
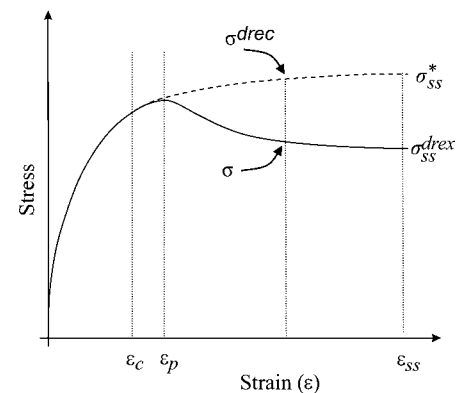
Vanadium-microalloyed steel (V-HSLA) (Ref 27)	AISI 4120 steel (Ref 16)
$n = 1.46$	$n = 1.42$
$\varepsilon_p = 1.20 \cdot 10^{-3} \cdot d_0^{0.263} \cdot \dot{\varepsilon}^{0.216} \cdot \exp\left(\frac{230,474 \times 0.216}{RT}\right)$	$\varepsilon_p = 3.50 \cdot 10^{-3} \cdot d_0^{0.15} \cdot \dot{\varepsilon}^{0.17} \cdot \exp\left(\frac{46,326}{RT}\right)$
$\sigma_0 = 3.25 \cdot \dot{\varepsilon}^{0.073} d_0^{-0.09} \exp\left(\frac{314,223 \times 0.073}{RT}\right)$	$\sigma_0 = 2.96 \cdot \dot{\varepsilon}^{0.11} d_0^{-0.11} \exp\left(\frac{36,300}{RT}\right)$
$\Omega = 39.25 \cdot \dot{\varepsilon}^{-0.076} d_0^{0.08} \exp\left(\frac{-228,215 \times 0.076}{RT}\right)$	$\Omega = 129 \cdot \dot{\varepsilon}^{-0.09} d_0^{-0.20} \exp\left(\frac{-24,346}{RT}\right)$
$\sigma_{ss}^{*} = 90.9 \left(\frac{d_0}{75}\right)^{-0.107} \sinh^{-1}\left[3.31 \times 10^{-3} \left[\dot{\varepsilon} \cdot \exp\left(\frac{354,819}{RT}\right)\right]^{0.192} \right]$	$\sigma_{ss}^{*} = 100 \left(\frac{d_0}{125}\right)^{-0.07} \sinh^{-1}\left[3.01 \times 10^{-3} \left[\dot{\varepsilon} \cdot \exp\left(\frac{396,800}{RT}\right)\right]^{0.175} \right]$
$\sigma_{ss}^{drex} = 90.91 \cdot \sinh^{-1}\left[1.97 \times 10^{-3} \cdot \left[\dot{\varepsilon} \cdot \exp\left(\frac{290,000}{RT}\right)\right]^{0.247} \right]$	$\sigma_{ss}^{drex} = 90.91 \cdot \sinh^{-1}\left[1.18 \times 10^{-3} \cdot \left[\dot{\varepsilon} \cdot \exp\left(\frac{238,000}{RT}\right)\right]^{0.238} \right]$
$\varepsilon_{0.5} = 1.43 \times 10^{-3} \cdot d_0^{0.239} \left[\dot{\varepsilon} \cdot \exp\left(\frac{256,305}{RT}\right)\right]^{0.21}$	$\varepsilon_{0.5} = 9.95 \times 10^{-3} \cdot d_0^{0.1} \left[\dot{\varepsilon} \cdot \exp\left(\frac{278,772}{RT}\right)\right]^{0.145}$

element method has the advantage over other analytical methods, such as slip line, upper bound, and so on, in its ability to generate detailed and accurate information. The main advantage of FEM is that it can simulate a wide range of metalforming problems with little restriction on workpiece geometry and boundary conditions. This makes it a very useful tool for analyzing metalforming processes.

Two basic formulations are primarily adopted in the nonlinear FEM to deal with the large displacements, rotations, and strains that are typical in bulk forming processes such as hot rolling and forging. These are the Lagrangian formulation and the Eulerian formulation (Ref 57, 58). In the Lagrangian formulation, the mesh is fixed to the material that is being deformed, and therefore, the mesh follows the movement of the material throughout the process. Finite-element simulation with the Lagrangian formulation follows the deformation path of the workpiece incrementally from the beginning to the end by dividing the process time into a number of incremental small time steps. In the Eulerian formulation, the mesh is fixed in space, and the material flows across the elements of the mesh. The Eulerian formulation has been used effectively in the simulation of processes such as rolling and extrusion that can be considered as kinematically steady

processes. In such processes, kinematically steady state means that the shape and temperature of the workpiece at any planar section in the process (for example, exit section of a roll bite) remain reasonably steady or constant with time. Many FEMs for multipass shape rolling adopt a steady-state approach (Eulerian approach), mainly because of its simplicity. In the steady-state approach, stream lines and flow-stress distributions are iteratively updated until the analysis converges to the solution. In addition to the simulation being simplified, other benefits of the steady-state approach include shorter computation time and significantly smaller memory and file size requirements.

Early research studies on three-dimensional (3-D) FEMs for simulation of shape rolling were published in the late 1980s (Ref 59–64). Kiuchi and Yanagimoto (Ref 62) introduced a different approach to modeling the rolling. Their method, called the complex-element method, is a combination of the rigid-plastic FEM and the slab method. A variation of this approach was adopted by Kim et al. (Ref 65). Recent advancements in computer technology and FEM have led to the development of more sophisticated 3-D FEMs for rolling, along with thermomechanical coupling, and consequently led to finite-element-based simulation of microstructural evolution. There exists a great deal of published literature on the topic of FEM of hot rolling and other metalforming processes (Ref 57–65). Readers are encouraged to refer

to published literature for details about FEM of metalforming processes.

Simulation of the hot rolling process is a transient analysis that includes, at the very least, deformation and heat-transfer computations of the workpiece in the roll bite, heat-transfer computations of the workpiece in the interstand region, and postrolling cooling on the runout table. Rolling, as mentioned earlier, is generally considered a steady-state process. Therefore, a Eulerian-formulation-based steady-state approach is often used for simulating hot rolling. A Eulerian-formulation-based 3-D FEM program, ROLPAS (Ref 9), is discussed next.

ROLPAS: A 3-D FEM for Hot Rolling

ROLPAS is a Eulerian-formulation-based 3-D FEM program (Ref 9) for simulating hot rolling. In this FEM, the roll pass is broken down into control volumes, as illustrated in Fig. 21. A solution is obtained for each of the control volumes in a sequential manner, starting with deformation in the first roll bite and progressively modeling subsequent interstand regions and roll bites. Both deformation and heat transfer are modeled in the roll bites, whereas only heat transfer is modeled in the interstand regions.

Deformation Analysis. In this approach, the deformation analysis in the first roll bite begins with a mesh that is an initial guess. For a given
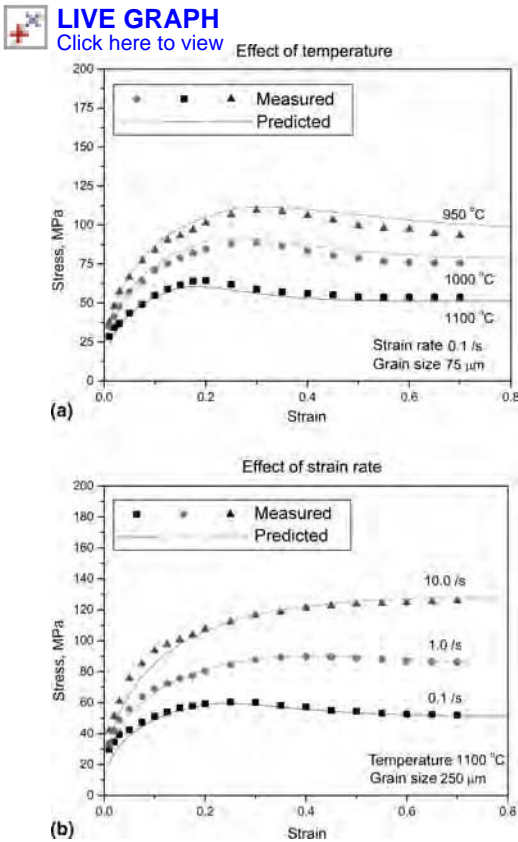
Fig. 16 Flow-stress model: measurements vs. predictions for V-HSLA steel. (a) Effect of temperature $\dot{\varepsilon} = 0.1/s$, $d_0 = 75$ μm. (b) Effect of strain rate (temperature = 1100 °C, $d_0 = 250$ μm)
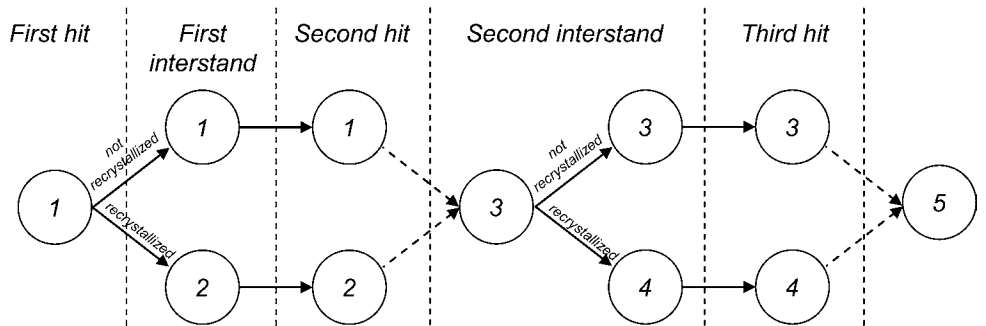
Fig. 17 Representation of substructures in the approximate method, using the rule of mixtures
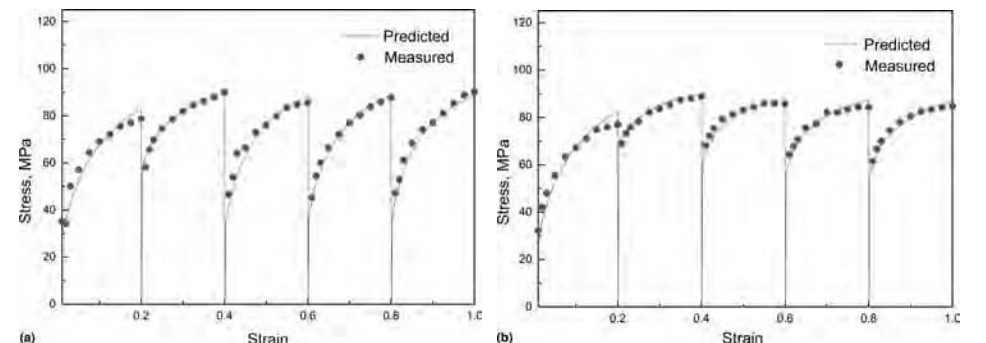
Fig. 18 Five-hit hot compression test: flow-stress measurements vs. model predictions for V-HSLA steel. (a) Test conditions: temperature, 1100 °C; strain rate, 1.0/s; initial austenite grain size, 200 μm; interhit time, 1.0 s. (b) Test conditions: temperature, 1100 °C; strain rate, 1.0/s; initial austenite grain size; 200 μm; interhit time, 0.2 s

workpiece cross-sectional geometry (two-dimensional, or the 2-D shape representing the cross section of the incoming slab), the ROLPAS program generates the 3-D finite-element control-volume mesh in the roll gap using eight-node hexahedral isoparametric elements (Fig. 22). After the initial control-volume mesh is established, traction- and velocity-boundary conditions are assigned to the nodes. The program then solves a set of simultaneous nonlinear equations based on rigid-plastic material behavior. To find the kinematically steady-state geometry, ROLPAS updates the geometry of the workpiece after every iteration in the solution process based on the velocity field obtained. Iterations are continued until a convergence of the solution occurs. In short, the stream lines (representing the flow of material) and the corresponding flow-stress distributions are iteratively solved and updated until convergence is obtained. Convergence is said to be attained when the numerical difference between the solutions of two successive iterations is smaller than a specified tolerance.

Heat-Transfer Analysis. The workpiece undergoes temperature changes continuously during the hot rolling process. The heat loss at the workpiece surface is due to convection, radiation, and conduction. On the other hand, heat is generated in the workpiece due to plastic deformation and friction at the roll-workpiece interface. In steady-state hot rolling, the longitudinal temperature gradients in the workpiece are small compared to transverse thermal gradients. Therefore, following the approach of Kim et al. (Ref 65), the heat flow along the rolling direction is neglected to simplify heat-transfer computations. This assumption makes it possible to model heat-transfer using 2-D FEM instead of 3-D FEM formulation. In ROLPAS, the temperature distribution and heat-transfer rates are determined at each cross section in the control volumes using a 2-D finite-element heat-transfer analysis (Fig. 23). The instantaneous time at each of the sections in the interstand region is estimated using the distance between the sections and the longitudinal velocity of metal flow obtained from deformation analysis (i.e., exit velocity from the previous roll bite).

Integrated Finite-Element System for Simulation of Microstructural Evolution

Earlier sections of this article described the development of the microstructural evolution models and FEM for hot rolling. In this section, an example of integration of FEMs and microstructural evolution models for simulation of metal flow and microstructural evolution in a hot rolling process is presented. Figure 24 shows the overall computational structure used in this integrated system. The main computational unit of the integrated system is the 3-D finite-element program ROLPAS, discussed

previously. ROLPAS computes the thermomechanical history at each node in the control volumes, which is used to model microstructural evolution during hot rolling. A microstructural evolution module, MICON, uses the thermomechanical history computed by the FEM in conjunction with recrystallization and grain-growth models to simulate the evolution of austenite during hot rolling. The flow chart shown in Fig. 25 describes the computation logic for simulating microstructural evolution. For the first pass, the program calculates the initial austenite grain size based on the reheating temperature and soak time. The FEM computes the thermomechanical history (strain, strain rate, and temperature at different instants in time) at every node in the control volume for the first pass and also for the first interstand. After deformation and heat-transfer computations for each roll bite, the microstructural evolution module in conjunction with the heat-transfer analysis module computes the recrystallized fraction and the austenite grain size at each node in the interstand region. If the fraction recrystallized is found to be greater than 95%, complete recrystallization is assumed. Otherwise, the partial recrystallization condition is carried over to the next pass. In the event of complete recrystallization, grain growth after recrystallization is modeled. Partial recrystallization is handled using the rule of mixtures. The last unit in the integrated system is a module called AUSTRANS, for modeling phase transformation. It uses the temperature history after rolling (computed by the FEM), austenite grain size, fraction recrystallized, retained strain, and so on in the last interstand (computed by MICON) and relevant isothermal transformation data to model the transformation of austenite to ferrite, pearlite, bainite, and martensite. This module also uses available structure-property relationships to estimate the mechanical properties of the rolled product. At any instant t during transformation, there can be up to five different phases present in the material, namely austenite, ferrite, pearlite, bainite, and martensite. Initially, all nodes are assumed to be in the austenite state. Also, at any instant t, the sum of all volume fractions must equal 1:

Fig. 19 Thermomechanical history in a physical simulation experiment

$$\sum_{i=1}^{5} V_{i,t} = 1 \quad for \ t > 0 \qquad \text{(Eq 17)}$$

where $V_{i,t}$ is the volume fraction of a phase i at time t. The interstand region is discretized into several small time steps. For each time step, the temperature is assumed to be constant. The volume fractions of the ferrite, pearlite, and bainite at each time step are determined using the JMAK equation and the sequential transformation approach. In the case of martensitic transformation, the Koistinen-Marburger model is used. This computational procedure is repeated over each time step for every node until austenite is completely transformed or until it reaches the room temperature.

Finite-Element Simulation of Microstructural Evolution during Shape Rolling. An example of an eight-pass hot rolling schedule to convert a 250 mm. square billet to a 100 mm round bar is presented. The simulations were conducted for AISI 1050 steel with an initial reheat temperature of 1200 °C and specified interstand and postrolling cooling conditions. Figure 26 shows the shapes at each of the roll bites predicted by the FEM. The FEM ROLPAS in combination with MICON simulate the evolution of austenite by modeling recrystallization and grain-growth phenomena. The results from this analysis are used as input for AUSTRANS, which simulates austenite decomposition. Contours showing the distribution of volume fractions of austenite, pearlite, bainite, and martensite are illustrated in Fig. 27. The center of the rolled bar sees higher temperatures and lower cooling rates, resulting in a high percentage of pearlite. Near the surface, where the temperatures and cooling rates are intermediate, a significant quantity of bainite is formed, while martensite forms at the surface that experiences the highest cooling rates.

One advantage of such a system is that it allows the user to conduct a quick sensitivity analysis to see the effect of varying certain process conditions on the final properties. Figure 28 illustrates the variation in mechanical properties along the radial direction predicted
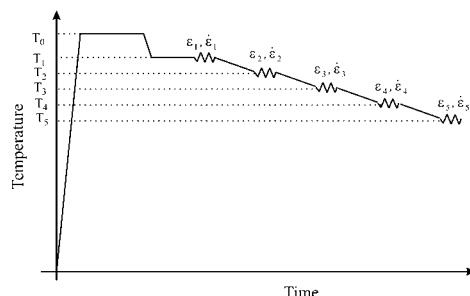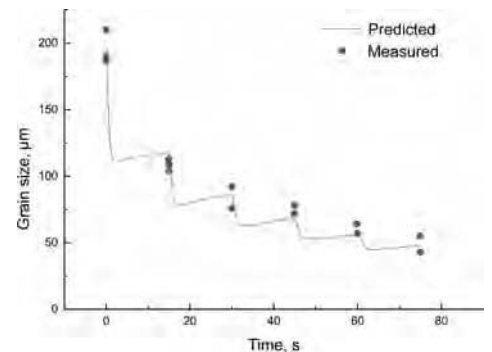
Fig. 20 Physical simulation experiment: comparison between measured and predicted austenite grain size. (Source: Ref 9)

by the simulation for three different reheat temperatures. For the purpose of comparison, similar data can be generated easily by changing other parameters, such as the roll pass sequence, interstand cooling, and postrolling cooling conditions.

Improved Metal Flow Prediction due to Microstructure Evolution Modeling. Accurate prediction of material spread is important while designing roll pass sequences for bar and shape rolling. Accuracy of material spread prediction in FEMs is strongly dependent on the accuracy of flow-stress models. The integrated approach makes it possible to model microstructural evolution and metal flow simultaneously and thereby leads to better accuracy in the prediction of metal flow and forming loads.

Figure 29 shows a three-pass rough rolling schedule being used to convert a 6.625 in. square billet to a 5 in. diameter round billet in a rolling mill. This process has been chosen to illustrate the effect of microstructure modeling on metal flow using the integrated approach described earlier. Figure 30(a) shows the finite-element analysis (FEA)-predicted shape at the exit of the rolls in the second pass without microstructure modeling, while Fig. 30(b) shows the FEA-predicted shape with microstructure modeling. Note that a microstructure-dependent flow-stress model was used in the FEA with microstructure modeling. For comparison, a sketch of the actual shape seen at the end of the second pass is shown in Fig. 30 (c). From these results, it is evident that the FEM without microstructure modeling greatly underpredicts the material spread. It also fails to predict the bulge profile on the sides. On the other hand, predictions of material spread with microstructure modeling are more accurate, and the shape predicted matches the actual shape more closely. It is generally recognized that softening due to recovery and recrystallization during hot rolling has a major effect on the flow stress and metal flow behavior. It is therefore important to account for these metallurgical changes while modeling a hot working process. In addition to improvements in metal flow predictions, the integrated approach to modeling metal flow and microstructural evolution also leads to better predictions of forming loads (Ref 66, 67).

Finite-Element Simulation of Microstructural Evolution during Induction Hardening. The examples presented thus far discuss simulation of microstructural evolution in hot rolling. However, the modeling principles are applicable to other hot working and heat treatment processes. In the following example, simulation of microstructural evolution during induction hardening is discussed.

The induction hardening process is an advanced and efficient method for hardening localized regions in medium- and high-carbon steel parts. One of the primary advantages of the induction hardening process is its ability to harden parts with repeatable results in a very short cycle time. The induction hardening process consists of rapidly heating the part to an austenitic temperature, followed by rapid spray cooling. Simulation of the induction hardening process is generally considered difficult, due to the coupling effects of electromagnetic, thermal, mechanical, and metallurgical field variables that must be considered. Simulation of the induction hardening process involves modeling of electromagnetic, thermal, mechanical, and metallurgical phenomena, as illustrated in Fig. 31. The governing equations can be divided into three models: electromagnetic, thermomechanical, and metallurgical. The simulation is carried out over a number of small time-step increments. For each small time step, the electromagnetic, thermomechanical, and metallurgical models are solved sequentially. First, the electromagnetic model determines ohmic power (or heating power) in each object of the FEM. Next, the thermomechanical model, coupled with the metallurgical model, calculates nodal temperatures, displacements, stresses, and phase transformations.

The aforementioned approach has been used for simulating induction hardening of outer races on a double cone (Ref 68). A double cone is a component in a particular tapered roller-bearing assembly. During induction hardening of this component, first one race is induction hardened, then the part is flipped around to harden the other race. Each race is induction hardened using a setup consisting of a copper coil, flux concentrator, and stainless steel holder. The flux concentrator is used to concentrate the magnetic flux over the region in the workpiece that needs to be heated. The workpiece (double cone) is made of AISI 1080 steel (UNS G10800). In this example, the workpiece is modeled as a mixture of different phases during the simulation. Each phase has its own electrical, magnetic, thermal, and mechanical properties. The properties of the material at any given time during the simulation are estimated using the rule of mixtures. At the start of the simulation, the microstructure in the workpiece is assumed to be 100% pearlite. As the workpiece is heated, transformation to austenite first occurs in the heated areas. Upon rapid cooling, much of this austenite transforms to martensite.

The 2-D axisymmetric finite-element mesh of the double cone, coil, flux concentrator, holder, and air is shown in Fig. 32. Each side of the outer race is heated at a 105 kHz frequency using 150 kW power for 3.8 s. After heating, each race is cooled using a water spray system after a short 1 s dwell time. Figure 33 shows the profile of the induction-hardened layer as predicted by the FEM. The predicted hardened profile matches the measured profile very well. This simulation model has been used successfully to develop the coil design and also to develop induction heating process parameters in order to obtain the desired hardened layer profile underneath the raceways.

Summary and Concluding Remarks

Advances in FEM techniques for metalforming processes, coupled with advances in microstructure modeling, have led to the development of integrated models for simulation of microstructural evolution in steels. Such simulation tools have been shown to be very effective in the analysis and optimization of various hot working and heat treatment processes. Today (2009), many leading steel companies now use such tools routinely for process development and optimization.

The semiempirical and empirical models for microstructural evolution discussed in this
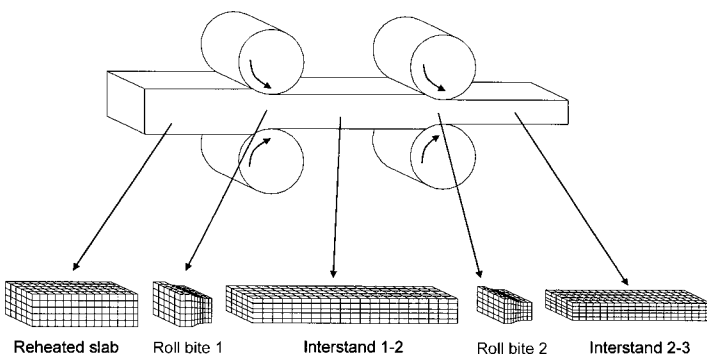
Fig. 21 Control-volume approach for finite-element modeling of hot rolling
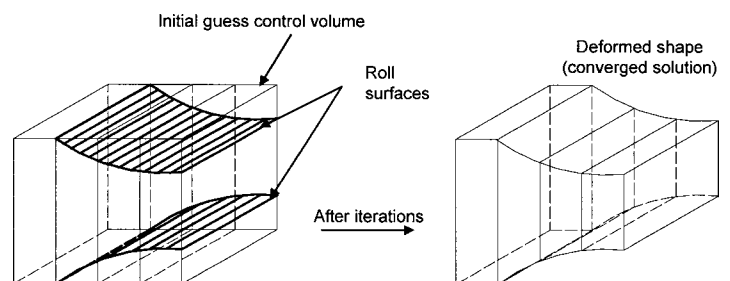
Fig. 22 Initial guess of the control volume and converged solution from the finite-element analysis
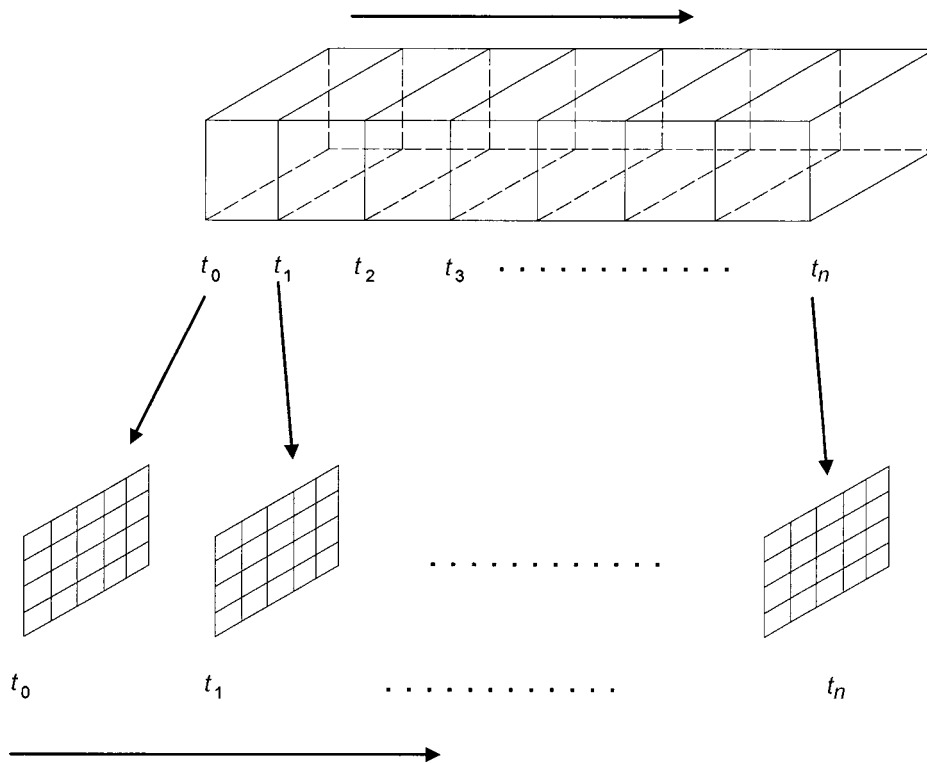
Fig. 23 Heat-transfer analysis in the interstand computes nodal temperatures for the control volumes using a two-dimensional analysis (only transverse heat transfer is analyzed while neglecting longitudinal heat transfer).
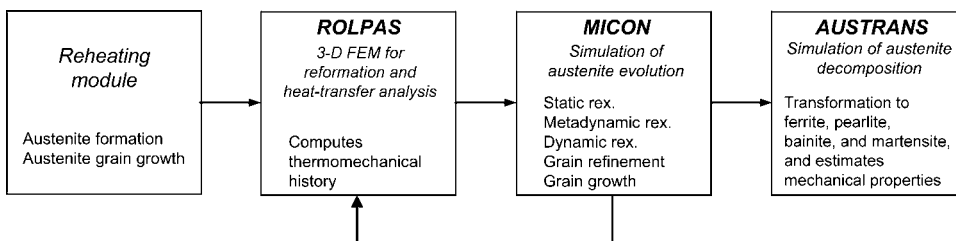
Fig. 24 Framework of the integrated system

article are considered as macroscale models in which the microstructure is described using internal state variables such as average grain size, fraction recrystallized, and fraction transformed. New modeling methods are being developed that allow modeling of microstructural evolution on the mesoscale. In this methodology, simulation is conducted on the length scale of the microstrutural features (e.g., grains). Using the mesoscale approach, microstructures and their distributions can be predicted rather than mean values, which have traditionally been used to characterize microstructures. Mesoscale techniques are broadly classified into the following types: Monte Carlo simulations, cellular automata, and phase-field models (Ref 53) (see the corresponding articles "Monte Carlo Models for Grain Growth and Recrystallization," "Cellular-Automata Models

for Recrys tallization," and "Phase-Field Microstructure Modeling" in this Volume). While significant differences exist among these techniques, each is considered to be an attractive tool for simulation of microstructural evolution at the mesoscale level. However, much research and advancement is needed to make these novel modeling techniques suitable for everyday use in the industry.

REFERENCES

1. B.K. Panigrahi, Processing of Low Carbon Steel Plate and Hot Strip—An Overview, *Bull. Mater. Sci.,* Vol 24 (No. 4), August 2001, p 361–371
2. C.M. Sellars and J.A. Whiteman, Recrystallization and Grain Growth in Hot Rolling, *Met. Sci.,* March-April 1979, p 187–194
3. Y. Saito and C. Shiga, Computer Simulation of Microstructural Evolution in Thermomechanical Processing of Steel Plates, *ISIJ Int.,* 1992, p 414–422
4. P.D. Hodgson and R.K. Gibbs, A Mathematical Model to Predict the Mechanical Properties of Hot Rolled C-Mn and Microalloyed Steels, *ISIJ Int.,* Vol 32 (No. 12), 1992, p 1329–1338
5. P.C. Campbell, E.B. Hawbolt, and J.K. Brimacombe, Microstructural Engineering Applied to the Controlled Cooling of Steel Wire Rod: Parts I, II and III, *Metall. Trans. A,* Vol 22A, November 1991, p 2769–2805
6. D.Q. Jin, I.V. Samarasekhara, and J.K. Brimacombe, Modelling Inhomogeneous Deformation during Hot Rolling of Plain Carbon Steels, *Proceedings of the 38th Mechanical Working and Steel Processing Conference,* 1996, p 329–338
7. A. Yoshie, M. Fujioka, Y. Watanabe, K. Nishioka, and H. Morikawa, Modeling of Microstructural Evolution and Mechanical Properties of Steel Plates Produced by Thermo-Mechanical Control Process, *ISIJ Int.,* Vol 32, 1992, p 395–404
8. W. Roberts, H. Boden, and B. Ahlbom, Dynamic Recrystallization Kinetics, *Met. Sci.,* March-April 1979, p 195–205
9. P.M. Pauskar, "An Integrated System for Analysis of Metal Flow and Microstructural Evolution in Hot Rolling," Ph.D. Dissertation, The Ohio State University, 1998
10. Y. Watanabe et al., Integrated Model for Microstructural Evolution and Properties of Steel Plates Manufactured in Production Line, *ISIJ Int.,* 1992, p 405–413
11. K. Karhausen and R. Kopp, Model for Integrated Process and Microstructure Simulation in Hot Forming, *Proceedings of the International Symposium on Mathematical Modeling of Hot Rolling of Steel,* 1990, p 99–118
12. M. Avrami, Kinetics of Phase Change I—General Theory, *J. Chem. Phys.,* Vol 7, 1940, p 1103–1112
13. M. Avrami, Kinetics of Phase Change II—Transformation-Time Relations for Random Distribution of Nuclei, *J. Chem. Phys.,* Vol 8, 1940, p 212–224
14. M. Avrami, Kinetics of Phase Change III—Granulation, Phase Change and Microstructure, *J. Chem. Phys.,* Vol 9, 1941, p 177–184
15. W.A. Johnson and R.F. Mehl, Reaction Kinetics in Processes of Nucleation and Growth, *Trans. Metall. Soc., AIME,* Vol 135, 1939, p 416–458
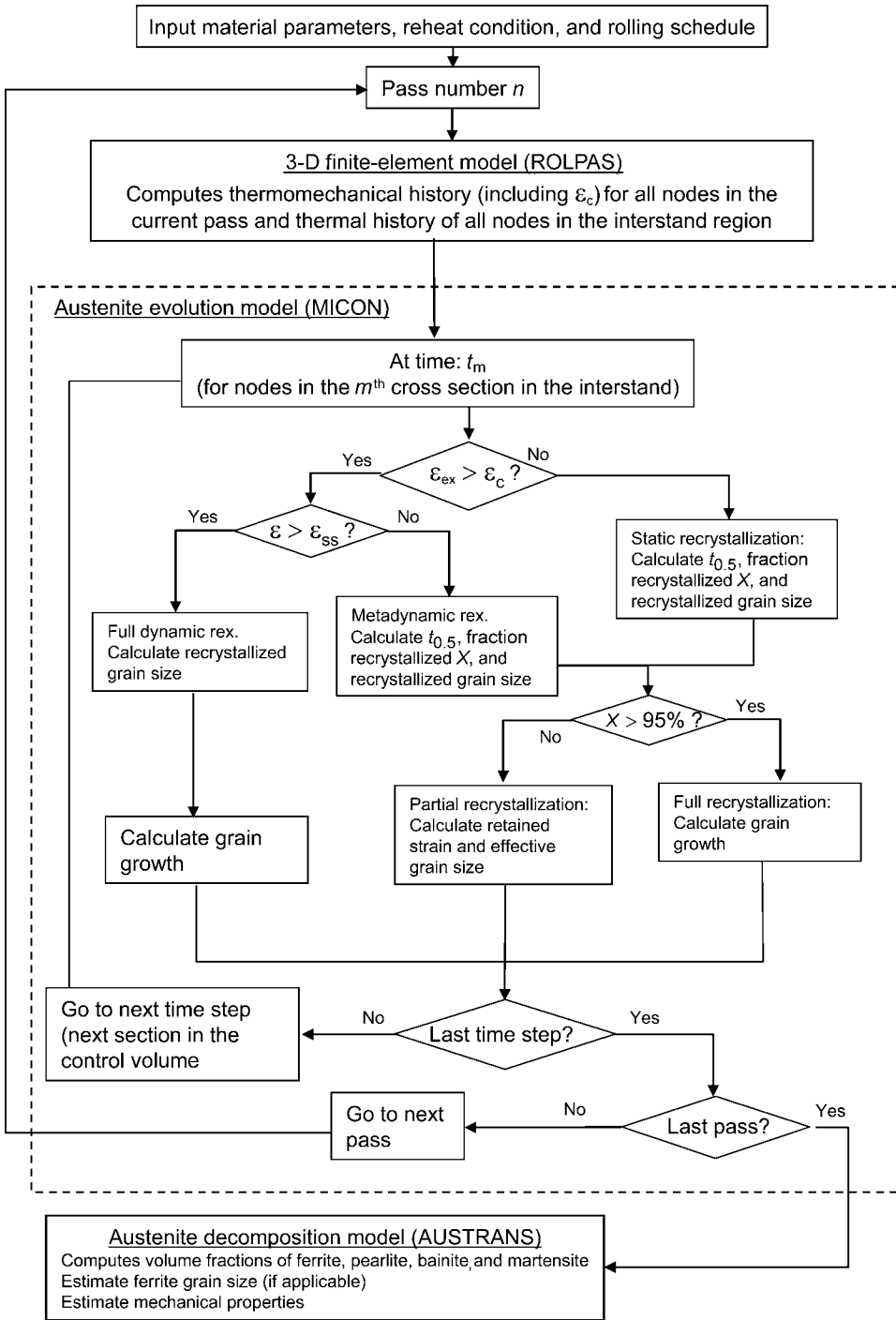
Fig. 25 Flow-chart for modeling microstructural evolution
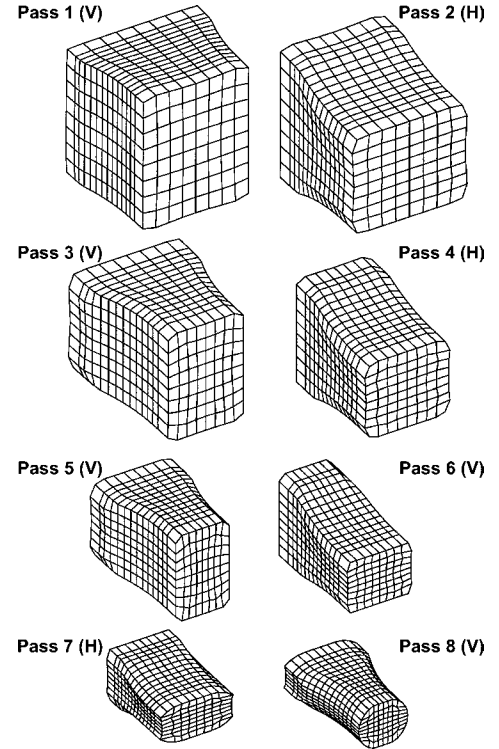
Fig. 26 Finite-element-predicted shapes of control volumes at the roll bites in an eight-pass hot rolling process. V, vertical rolls; H, horizontal rolls

16. E.B. Damm, P.M. Pauskar, J. Ives, K. Sawamiphakdi, and M. Conneely, Modeling Hot-Bar Rolling to Predict Surface Quality, *Proceedings of the 42nd Mechanical Working and Steel Proceeding Conference,* 2000
17. M. Dubois, M. Militzer, A. Moreau, and J. Bussiere, A New Technique for the Quantitative Real-Time of Austenite Grain Growth in Steel, *Scr. Mater.,* Vol 42, 2000, p 867–874
18. S. Nanba et al., Prediction of Microstructure Distribution in the Through Thickness Direction During and After Hot Rolling in Carbon Steels, *ISIJ Int.,* 1992, p 377–386
19. P.A. Manohar, D.P. Dunne, T. Chandra, and C.R. Killmore, Grain Growth Predictions in Microalloyed Steels, *ISIJ Int.,* Vol 36 (No. 2), 1996, p 194–200
20. M. Militzer, A. Giumelli, E.B. Hawbolt, and T.R. Meadowcroft, Austenite Grain Growth Kinetics in Al-Killed Plain Carbon Steels, *Metall. Mater. Trans. A,* Vol 27A, November 1996, p 3399–3409
21. S.L. Semiatin, J.C. Soper, and I.M. Sukonnik, Short-Time Beta Grain Growth Kinetics for a Conventional Titanium Alloy, *Acta Mater.,* Vol 44 (No. 5), 1996, p 1979–1986
22. M.F. Ashby and K.E. Easterling, A First Report on Diagrams for Grain Growth in Welds, *Acta Metall.,* Vol 30, 1982, p 1969–1978
23. A. Zufia and J.M. Llanos, Mathematical Simulation and Controlled Cooling in an EDC Conveyor of a Wire Rod Rolling Mill, *ISIJ Int.,* Vol 41 (No. 10), 2001, p 1282–1288
24. P.A. Manohar, K. Lim, A.D. Rollett, and Y. Lee, Computational Exploration of Microstructural Evolution in a Medium C-Mn Steel and Applications to Rod Mill, *ISIJ Int.,* Vol 43 (No. 9), 2003, p 1421–1430
25. S. Mishra and T. DebRoy, Non Isothermal Grain Growth in Metals and Alloys, *Mater. Sci. Technol.,* Vol 22 (No. 3), March 2006, p 253–278
26. A. Laasraoui and J.J. Jonas, Prediction of Steel Flow Stresses at High Temperatures and Strain Rates, *Metall. Trans. A,* Vol 22A, July 1991, p 1545–1558
27. P.M. Pauskar, S. Phadke, and R. Shivpuri, A Microstructure Dependent Flow Stress Model for Steel Rolling, *Proceedings of the 39th Mechanical Working and Steel Processing Conference, ISS,* Vol 35, 1997

Label	Value
1	0.162
2	0.325
3	0.487
4	0.650
5	0.812

(a)

Label	Value
1	0.098
2	0.197
3	0.296
4	0.394
5	0.492

(b)

Label	Value
1	0.166
2	0.333
3	0.500
4	0.665
5	0.831

(c)
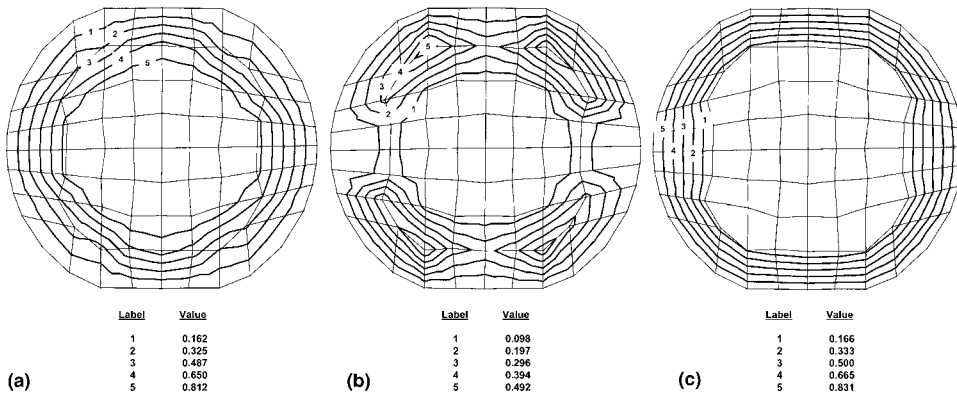
Fig. 27 Contours showing distribution of volumetric fractions of (a) pearlite, (b) bainite, and (c) martensite

LIVE GRAPH
Click here to view
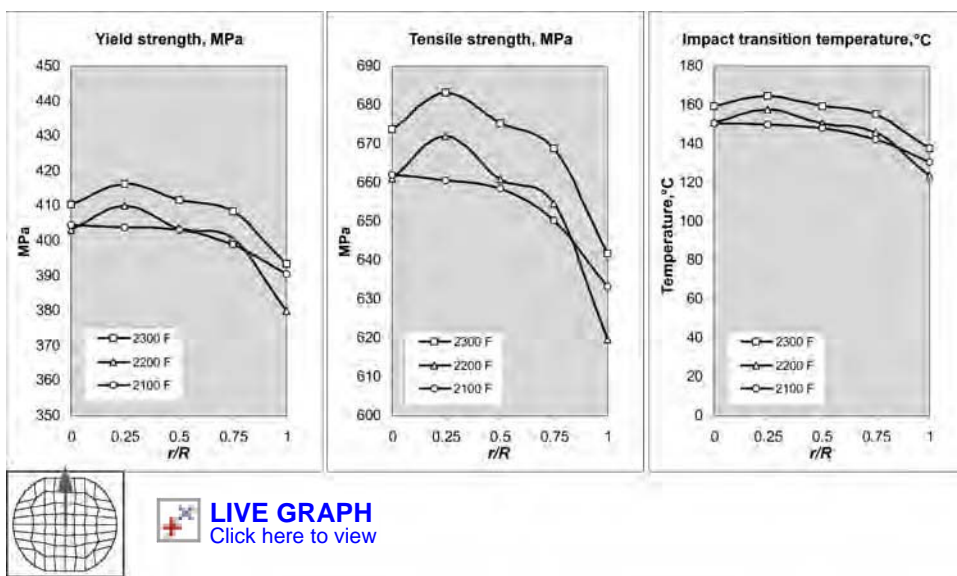
Fig. 28 Variation of mechanical properties along the radial direction for three different initial temperatures

6.625 in. square

First pass

5.25

Second pass

4.44

Third pass
5.0 in. round
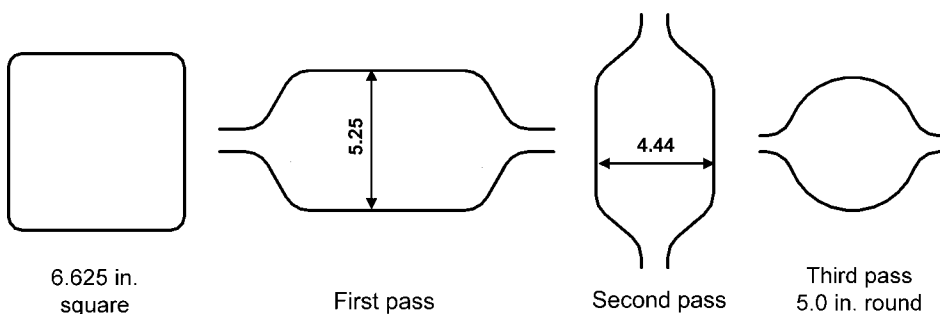
Fig. 29 Three-pass rolling sequence (6.625 in. square to 5 in. round)

28. C.M. Sellars, Modeling—An Interdisciplinary Activity, *Proceedings of the International Symposium on Mathematical Modeling of Hot Rolling of Steel,* 1990, p 1–18

29. C. Rossard, *Proceedings of the International Conference on Strength of Metals and Alloys,* Vol 2, 1973, p 175–203

30. I.V. Samarasekhara, D.Q. Jin, and J.K. Brimacombe, The Application of Microstructural Engineering to the Hot Rolling of Steel, *38th Mechanical Working and Steel Processing Conference Proceedings, ISS,* Vol 34, 1996, p 313–327

31. W.P. Sun and E.B. Hawbolt, Comparison between Static and Metadynamic Recrystallization—An Application to the Hot Rolling of Steels, *ISIJ Int.,* Vol 37 (No. 10), 1997, p 1000–1009

32. S. Kim, Y. Lee, and B. Jang, Modeling of Recrystallization and Austenite Grain Size for AISI 316 Stainless Steel and Its Application to Hot Bar Rolling, *Mater. Sci. and Eng.,* Vol A357, 2003, p 235–239

33. E. Anelli, Application of Mathematical Modelling to Hot Rolling and Controlled Cooling of Wire Rods and Bars, *ISIJ Int.,* Vol 32 (No. 3), 1992, p 440–449

34. Y. Lee, *Rod and Bar Rolling: Theory and Applications,* CRC Press, 2004

35. H.J. McQueen and J.J. Jonas, Recovery and Recrystallization during High Temperature Deformation, *Treatise on Materials Science and Technology,* Vol 6, R.J. Arsenault, Ed., Academic Press, 1975, p 393–493

36. H.J. McQueen and J.J. Jonas, Role of the Dynamic and Static Softening Mechanisms in Multistage Hot Working, *J. Appl. Metalwork.,* Vol 3, 1985, p 410–420

37. S.F. Medina and J.E. Mancilla, Static Recrystallization Modelling of Hot Deformed Steels Containing Several Alloying Elements, *ISIJ Int.,* Vol 36 (No. 8), 1996, p 1070–1076

38. J.H. Beynon and C.M. Sellars, Modelling Microstructure and Its Effects during Multipass Hot Rolling, *ISIJ Int.,* Vol 32 (No. 3), 1992, p 359–367

39. J. Matja, J.G. Lenard, and M. Pietrzyk, Modeling the Evolution of the Microstructure of a Nb Steel, *ISIJ Int.,* Vol 36 (No. 8), 1996, p 1094–1102

40. H. Kwon, Y. Lee, S. Kim, J. Woo, and Y. Im, Numerical Prediction of Austenite Grain Size in Round-Oval Round Bar Rolling, *ISIJ Int.,* Vol 43 (No. 5), 2003, p 676–683

41. A.M. Elwazri, P. Wanjara, and S. Yue, Metadynamic and Static Recrystallization of Hypereutectoid Steel, *ISIJ Int.,* Vol 43 (No. 7), 2003, p 1080–1088

42. N. Nakata and M. Militzer, Modelling of Microstructure Evolution during Hot Rolling of a 780 MPa High Strength Steel, *ISIJ Int.,* Vol 45 (No. 1), 2005, p 82–90

43. C. Roucoules, S. Yue, and J.J. Jonas, Effect of Alloying Elements on Metadynamic Recrystallization in HSLA Steels, *Metall. Mater. Trans. A,* Vol 26A, 1995, p 181–190

44. K. Karhausen, R. Kopp, and R. Schneiders, Application of FEM to the Prediction of Microstructure in Hot Forming of Metals, *Proceedings of the Fourth International Conference on Technology of Plasticity,* 1993, p 1203–1211

45. J. Yanagimoto, K. Karhausen, A.J. Brand, and R. Kopp, Incremental Formulation for the Prediction Flow Stress and Microstructural Change in Multipass Hot Forming, *Proceedings of the 1995 Japanese Spring Conference for the Technology of Plasticity,* 1995, p 329–331

46. E. Scheil, *Arch. Eisenhüttenwes,* Vol 8, 1934, p 565
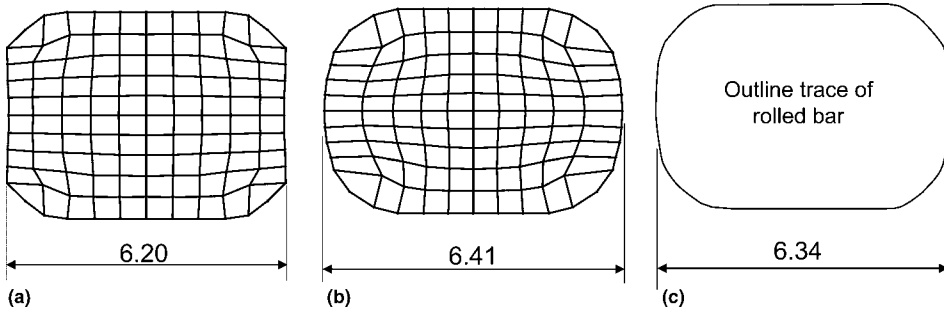
Fig. 30 Comparison of spread and shape predictions in the second pass (units, in.). (a) Finite-element analysis (FEA) without microstructure modeling. (b) FEA with microstructure modeling. (c) Actual rolled shape

Fig. 31 Coupling between electromagnetic, thermal, mechanical, and metellurgical models. (1) Heating power generation due to ohmic loss. (2) Temperature-dependent electric conductivity and magnetic permeability. (3) Phase transformation due to temperature changes. (4) Latent head generation from phase changes. (5) Phase transformation due to stress induced. (6) Transformation stress due to phase changes. (7) Thermal stress due to temperature changes. (8) Heat generation due to mechanical work. Source: Ref 68
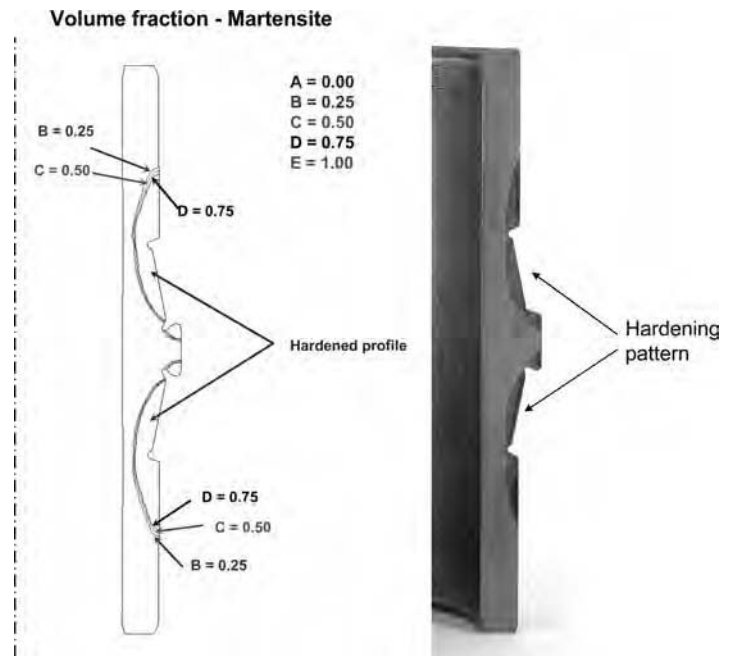
Fig. 33 Comparison between finite-element prediction and actual profile of the induction-hardened layer

Fig. 32 Finite-element model of the induction hardening process

47. E.I. Poliak and J.J. Jonas, Critical Strain for Dynamic Recrystallization in Variable Strain Rate Hot Deformation, *ISIJ Int.,* Vol 43 (No. 5), 2003, p 692–700

48. J.W. Cahn, Transformation Kinetics during Continuous Cooling, *Acta Metall.,* Vol 4, 1956, p 572–575

49. S.J. Jones and H.K.D.H. Bhadeshia, Kinetics of the Simultaneous Decomposition of Austenite into Several Transformation Products, *Acta Mater.,* Vol 45, 1997, p 2911–2920

50. D.P. Koistinen and R.E. Marburger, A General Equation Describing the Extent of the Austenite-Martensite Transformation in Pure Iron-Carbon Alloys and Plain Carbon Steels, *Acta Metall.,* Vol 7 (No. 1), 1959, p 59–60

51. I. Tamura, H. Sekine, T. Tanaka, and C. Ouchi, *Thermomechanical Processing of High-Strength Low-Alloy Steels,* Butterworth and Co., London, U.K., 1988, p 21

52. M. Murugaiyan et al., Phase Transformations in Two C-Mn-Si-Cr Dual Phase Steels, *ISIJ Int.,* Vol 46 (No. 10), 2006, p 1489–1494

53. M. Militzer, Computer Simulation of Micro-structure Evolution in Low Carbon Sheet Steels, *ISIJ Int.,* Vol 47 (No. 1), 2007, p 1–5

54. *Atlas of Isothermal Transformation and Cooling Transformation Diagrams,* American Society for Metals, 1977

55. S. Phadke, P. Pauskar, and R. Shivpuri, Computational Modeling of the Phase Transformations and Mechanical Properties during Cooling of Hot Rolled Rod, *J. Mater. Proc. Technol.,* Vol 150, 2004, p 107–115

56. L.N. Pussegoda and J.J. Jonas, Comparison of Dynamic Recrystallization and Conventional Controlled Rolling Schedules by Laboratory Simulation, *ISIJ Int.,* Vol 31 (No. 3), 1991, p 278–288

57. K.J. Bathe, *Finite Element Procedures,* Prentice-Hall, 1996

58. S. Kobayashi, S.I. Oh, and T. Altan, *Metal Forming and the Finite Element Method,* Oxford University Press, Oxford, 1989

59. J.J. Park and S.I. Oh, Application of Three Dimensional Finite Element Analysis to Shape Rolling Processes, *Trans. ASME,* Vol 112, February 1990, p 36–46

60. K. Mori and K. Osakada, Simulation of Three-Dimensional Deformation in Rolling by the Finite-Element Method, *Int. J. Mech. Sci.,* Vol 26 (No. 9/10), 1984, p 515–525

61. K. Mori and K. Osakada, Finite Element Simulation of Three Dimensional Deformation in Shape Rolling, *Proceedings of Numiform 89,* E.G. Thompson et al., Ed., 1989, p 337–342

62. M. Kiuchi and J. Yanagimoto, Computer Aided Simulation of Universal Rolling Processes, *ISIJ Int.,* Vol 30 (No. 2), 1990, p 142–149

63. C. David, C. Bertrand, and J.L. Chenot, A Transient 3D FEM Analysis of Hot Rolling of Thick Slabs, *Proceedings of Numiform'86,* Balkema, Gothenburg, 1986, p 219–224

64. J.L. Chenot, P. Montmitonnet, P. Buessler, and F. Fau, Finite Element Computation of Spread in Hot Flat and Shape Rolling with a Steady State Approach, *Eng. Comput.,* Vol 8, 1991, p 245

65. N. Kim, S. Kobayashi, and T. Altan, Three Dimensional Analysis and Computer Simulation of Shape Rolling by the Finite and Slab Element Method, *Int. J. Mach. Tools Manuf.,* Vol 31 (No. 4), 1991, p 553

66. P. Pauskar and S. Shivpuri, Microstructure and Mechanics Interaction in the Modeling of Hot Rolling of Rods, *Annals CIRP,* Vol 48, 1999, p 191–194

67. P. Pauskar and R. Shivpuri, Integrated Microstructural-Phenomenological Approach to the Analysis of Roll Pass Design in Bar Rolling, *40th MWSP Conf., Iron and Steel Society,* October 25–28, 1998 (Pittsburgh, PA)

68. K. Sawamiphakdi, J. Yang, and W. Wu, Finite Element Modeling of Induction Hardening Process, *Proceedings of the Third International Conference on Thermal Process Modelling and Simulation,* 2006

Simulation of Microstructure and Texture Evolution in Aluminum Sheet

Olaf Engler, Kai Karhausen, and Jürgen Hirsch, Hydro Aluminium Deutschland GmbH, R&D Center, Bonn, Germany

Introduction

Many metallic sheets exhibit significant plastic anisotropy that can be attributed to the presence of crystallographic texture. For instance, the in-plane anisotropy in textured sheet can cause major problems in subsequent forming operations. A prominent example is the occurrence of earing in the production of beverage cans of Al-Mn-Mg alloys (Ref 1–3). It is generally known that the formation of a suitable texture with minimum earing at final gage is critically dependent on the texture in the hot strip and the amount of subsequent cold rolling. The evolution of microstructure and texture during hot rolling is largely controlled by the recrystallization processes involved, which, in turn, depend on deformation temperature, strain rate, and the amount of recovery and, potentially, recrystallization between the individual hot rolling passes. In industrial production of aluminum sheet materials, models relating to the evolution of microstructure, texture, and the resulting properties have gained increasing interest, because the material producers are obliged to deliver material within strictly defined property-limit specifications (Ref 4). Because of the strong dependency of the evolution of microstructure and resulting properties on the time/temperature history during thermomechanical processing, it is important to take all process steps into account, which requires a coupling of thermomechanical models with appropriate microstructure simulation tools accounting for deformation and recrystallization (Ref 5–9).

This article explores the potential of through-process simulations of the development of microstructure, texture, and resulting properties during the thermomechanical processing of Al-Mn-Mg alloys, starting from the as-cast ingot to final-gage sheet. The next section provides a brief introduction in the thermomechanical production of aluminum sheet and, in particular, highlights the main effects governing the evolution of microstructure and texture. The simulation tools used to model the evolution of microchemistry, microstructure, and texture upon deformation and recrystallization of aluminum alloys are described. The analytical softening model AlSoft is combined with the rolling texture model GIA to simulate texture evolution during deformation and with a model designated ClaNG to treat the evolution in microchemistry, that is, the variation in solute level and precipitation, during thermomechanical processing. The resulting anisotropic properties — here, earing — are simulated with a polycrystal-plasticity model. The possibility of a coupled microstructure and texture simulation is then illustrated by way of the industrial production of Al-Mn-Mg AA 3104 (UNS A93104) can body stock and the final-gage earing properties.

Evolution of Microstructure and Texture during the Thermomechanical Processing of Al-Mn-Mg Sheet

The most common aluminum alloy for the body of beverage cans is AA 3104, which provides an optimal combination of strength and formability. The alloy contains approximately 1 wt% of both manganese and magnesium. Continuous efforts to increase strength for further materials savings by down-gaging led to the addition of small amounts of copper up to approximately 0.2%. For the can lids (ends), the higher-strength alloy AA 5182 (UNS A95182) with approximately 5% Mg is used. Figure 1 illustrates the conventional fabrication route for aluminum can stock. It consists of casting of large ingots, a two-step homogenization, breakdown hot rolling to 20 to 40 mm transfer slab gage, tandem hot rolling to approximately 2.5 to 5 mm hot strip, and final cold rolling, resulting in the finish-gage sheet with a thickness below 0.3 mm in the fully hardened temper (H19).

Production starts with a direct chill (DC) casting operation. Because the overall efficiency of the process is, to a large degree, determined by batch size, it is usually aimed for the largest ingots that can be processed in the hot mill (up to 30 tons). The resulting ingots with dimensions of up to 600 mm in thickness, 2 m in width, and up to 9 m in length are characterized by a coarse dendrite cell structure and a close-to-random texture. The microstructure contains coarse (10 to 50 μm) and inhomogeneously distributed constituent particles of various Al-Mn- and Al-Mn-Fe-Si phases. The sheet ingots are sawn at the ends and scalped to remove surface blemishes and chemistry variations caused by surface and subsurface segregation and inverse segregation that occur during solidification.

In preparation for hot rolling, ingots are preheated to a temperature up to 600 °C. Depending on the temperatures achieved, internal stresses are relieved and elements in supersaturated solid solution (e.g., manganese, iron, silicon, chromium, etc.) are precipitated. Soluble phases in the material are dissolved, and microsegregation is reduced by diffusion processes, while constituent phases are spheroidized and partially redissolved. From homogenization temperature, the slab is usually cooled down
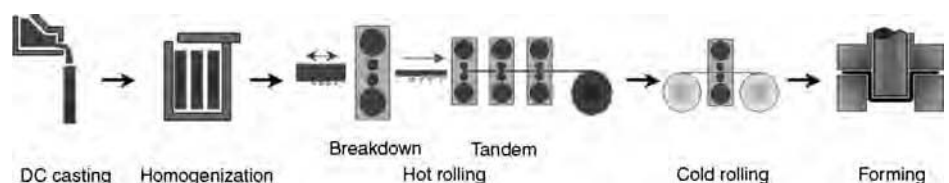
Fig. 1 Typical processing steps involved in the thermomechanical production of aluminum can stock. DC, direct chill

at controlled rates to the hot rolling temperature of approximately 500 °C. This may lead to further precipitation along with growth of existing particles, including some constitutional changes (from β-Al_6(Mn,Fe) to α-Al_{15}(Mn,Fe)$_3$Si$_2$). The microstructure state of the slab before hot rolling in terms of both solutes and precipitates significantly influences the evolution of the subsequent microstructure by affecting recrystallization mechanisms and kinetics and thereby also final-gage properties of aluminum sheets. Therefore, homogenization and the resulting microstructure are critically important.

The hot ingots are then transferred to the rolling line, which is commonly comprised of one or two single-stand reversible breakdown (roughing) mills followed by a multistand finishing mill. Rolling conditions (temperature, velocity, and reductions per pass) must be adjusted in a way that high reductions can be achieved and recrystallization and precipitation will occur under controlled conditions. During hot rolling, the coarse as-cast structure is transformed into a highly fragmented structure. The hard, coarse constituent phases are crushed and distributed more evenly, while decreasing temperatures during rolling together with plastic deformation may lead to further precipitation of phases from the supersaturated aluminum matrix. The alternating cycles of deformation and recrystallization during hot rolling have a strong impact on the overall microstructure and texture evolution. Because of the high deformations and temperatures of approximately 500 °C involved in breakdown rolling, the material readily recrystallizes between rolling passes, which usually results in a fairly mild cube recrystallization texture (001)[100].

In the following tandem mill, hot strip is produced unidirectionally through typically three or four connected rolling stands. The finished hot-rolled strip of approximately 2.5 to 5 mm thickness is edge trimmed and wound into coils. Because of the high strains and strain rates with short interstand times, modern high-speed tandem mills can be run in a process window that diminishes recrystallization between two

consecutive tandem passes — the so-called interstand (or interpass) recrystallization — yet, in turn, ensures recrystallization during cooling of the coiled hot strip (which is referred to as self-annealing).

Figures 2 and 3 show microstructure and texture results of three hot strips of alloy AA 3104 that were produced with different coiling temperatures, T_{coil} (Ref 3). The textures are represented in the form of the $\varphi_2 = 0°$ sections through the three-dimensional orientation space $\{\varphi_1, \Phi, \varphi_2\}$. (For measurement, analysis, and representation of texture data, see Ref 10). It is seen that for this hot rolling configuration, coiling temperatures in excess of 300 °C are required to achieve self-annealing and a fully recrystallized hot strip with a pronounced cube recrystallization texture (Fig. 2c, 3c), which is suitable for production of can stock (see subsequent text). At lower coiling temperatures, the textures comprise significant fractions of rolling texture components (Fig. 3a, b), which indicates that these temperatures are not sufficient to enable complete recrystallization upon self-annealing (Fig. 2a, b). The texture of AA 5182 hot strip is shown in Fig. 4. Here, recrystallization is complete, but the recrystallization texture sharpness is appreciably lower than that typically found in AA 3104 can body stock.

After hot rolling is complete, the coiled hot strip is allowed to cool down to ambient temperature before it is cold rolled to final gage (Fig. 1). During cold rolling, the grain structure is flattened, and the material strain hardens by formation of a typical dislocation and/or subgrain substructure. For aluminum beverage cans, highly cold-rolled states (H19) are used to meet strength requirements (>275 MPa). Thus, the sheets always comprise a strong β-fiber rolling texture, which is accompanied by the formation of ears at the four positions $\pm 45°$ from the rolling direction (RD). These 45° ears are offset by the cube texture that is retained from the recrystallized hot strip with pronounced 0°/90° earing (Fig. 3c; see insets in Fig. 10). Due to the slow rotation rate of the cube orientation, the resultant 0°/90° ears

are able to balance the newly forming 45° ears up to rather high rolling reductions, typically ranging between 80 and 90%. Figure 5(a) shows the texture of final-gage can stock, which
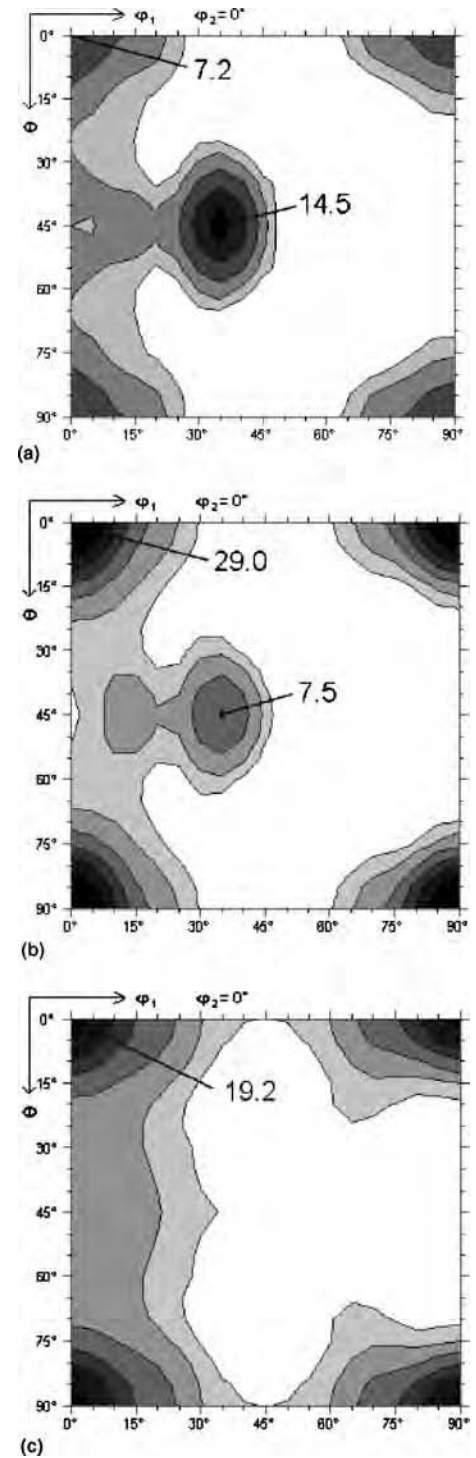
Fig. 3 Hot strip textures of AA 3104 as a function of coiling temperature, T_{coil} (orientation distribution function $\varphi_2 = 0°$ sections, intensity levels 1 - 2 - 4 - 7 - 10 - 15). (a) $T_{coil} = 260$ °C. (b) $T_{coil} = 280$ °C. (c) $T_{coil} = 315$ °C

Fig. 2 Microstructure of AA 3104 (UNS A93104) hot strip as a function of coiling temperature, T_{coil}. The cross section is longitudinal, with rolling direction (RD) and normal direction (ND.) (a) $T_{coil} = 260$ °C. (b) $T_{coil} = 280$ °C. (c) $T_{coil} = 315$ °C
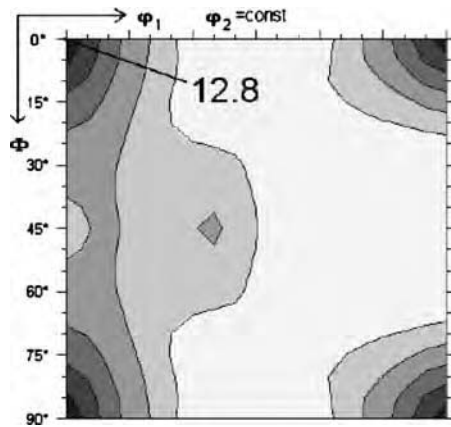
Fig. 4 Hot strip texture of AA 5182 (UNS A95182) can end stock (orientation distribution function $\varphi_2 = 0°$ section, intensity levels 1 - 2 - 4 - 7 - 10 - 15)

consists of a mixture of a retained cube texture and the newly formed rolling texture components. The resulting earing cup reveals a characteristic six-ear profile $h(\alpha)$ with ears under angles (α) of 0°, 45°, 135°, 180°, 225°, and 315° to the RD (see inset in Fig. 5b). However, because of the orthotropic symmetry of sheet rolling, earing profiles are principally symmetric with respect to RD and transverse direction (TD). Accordingly, Fig. 5(b) shows the symmetrized cup height profile of a typical six-ear profile in the range 0°/RD to 90°/TD. Such material with six ears is generally deemed satisfactory for the deep drawing and ironing of beverage cans.

If the cube texture is too weak, by contrast, the rolling texture components with the resultant 45° ears may become too strong. An example is included in Fig. 5(b), showing excessive 45° earing (dashed line). Vice versa, under certain conditions the 0°/180° ears may be higher than the remaining 45° and 90° ears (dotted line in Fig. 5b) (Ref 11). Such specimens are particularly deleterious for beverage can production, since the two main ears under 0°/180° cannot withstand the hold-down pressure supporting the blank during can forming without some circumferential flow and pinching. The resulting pinched ears tend to jam the body maker during subsequent redrawing and ironing, leading to machine downtime and hence reducing productivity.

Simulation Tools

Optimization Strategies in Sheet Processing and Material Quality. Early efforts in sheet process development were of a purely empirical nature. The production methods of flat sheets by rolling were experimentally developed in the plants and — after testing of, at best, a few key material properties — the resulting sheet was then used as is. This method requires numerous plant trials under varying conditions, selecting
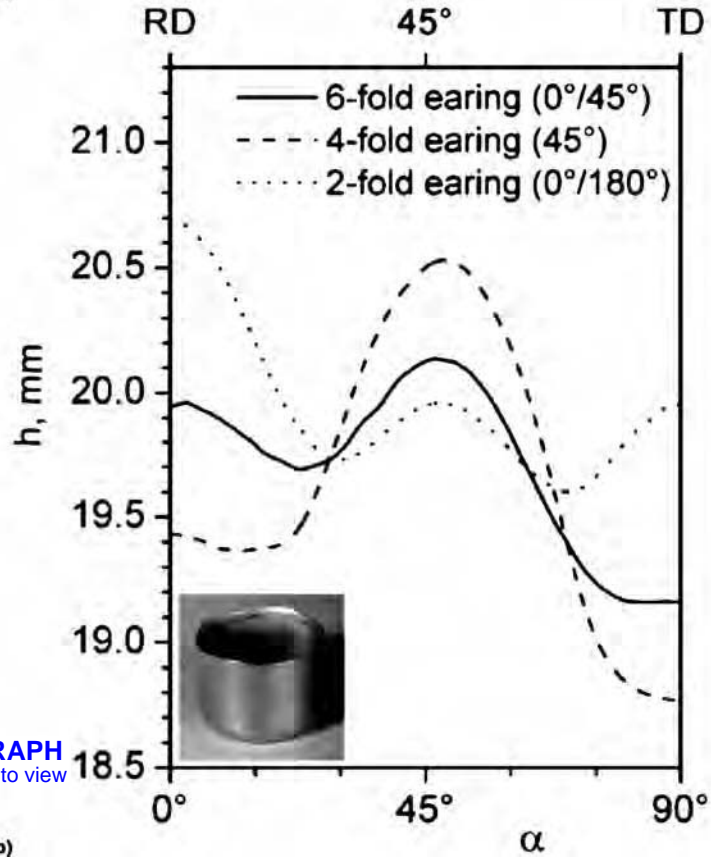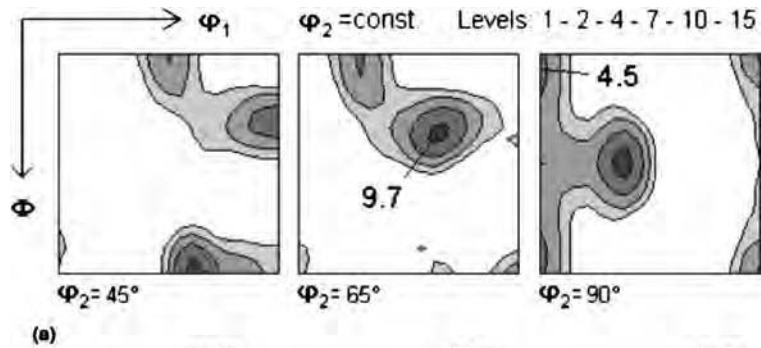
Fig. 5 Cold-rolled AA 3104-H19. (a) Texture. (b) Various earing profiles. h is the cup height (symmetrized). RD, rolling direction; TD, transverse direction

the best combination of process parameters. Thus, any knowledge about processing and material properties was developed empirically.

Obviously, this traditional way of process optimization has a number of major shortcomings:

- Plant trials are expensive and time-consuming. They reduce production capacities, with significant impact on high-volume-production facilities.
- The results tend to be less than adequate due to the limited amount of trials. The amount of data often does not allow a statistically relevant evaluation and sound conclusions.
- Some decisive parameters in the process are not easily accessible by direct measurement.
- The results are limited to the specific equipment used and the specific alloys tested. It is

difficult to extrapolate to new equipment or different alloy compositions.

Thus, a more intelligent way of process optimization is required, based on the sound knowledge and quantitative description of thermomechanical process parameters and metallurgical effects involved. With their integration into full simulation programs, such models are capable of also addressing complex interactions (Ref 12–17). For this purpose, it is necessary to systematically analyze and understand all important material parameters under the different thermomechanical conditions that affect the microstructure and ultimately control the resulting material properties. These effects need to be quantified, which can be done either empirically or — wherever possible — new algorithms

based on physical principles are introduced and translated into suitable material models applied in practice.

Such models can be integrated into industrial process models — even including potential on-line applications — that determine the local process conditions in a material. Thus, fully coupled through-process models can be used to simulate the corresponding process chain on a computer in all possible or even imaginary variations, with the material properties as a result. Once established and verified, only a limited number of validation plant trials need to be performed in order to assess the effectiveness of practices and equipment. Such an intelligent control for modern hot and cold rolling lines enhances performance and flexibility, since it allows a large variety of alloys to be processed individually and almost simultaneously according to their specific behavior under varying process conditions, even in complex interdependent production schedules.

Thermomechanical Simulation of Flow Stress, Static Recovery, and Static Recrystallization during Rolling. Thermomechanical simulation of rolling processes for the automated control of rolling passes is a common feature in modern, high-speed hot and cold rolling operations. An essential feature is the correct description of material flow stresses that determine the stress-strain state of the material in the roll gap. This, in turn, affects all decisive process parameters, such as roll force, torque, adiabatic heating, friction, and so on; for example, strip temperature is a result of the entire thermomechanical processing history. Thus, correct information on material strength under various forming situations is essential, and errors tend to accumulate during the process simulation of several consecutive passes. Since the value of flow stress essentially represents an integral value over a number of intrinsic material properties, it is a key parameter connecting plastomechanical, thermal, and material models, requiring full coupling.

Flow stress is classically determined in basic material tests, such as uniaxial compression. As an example, Fig. 6 shows a set of measured flow stress curves for hot compression testing of AA 3104 at different temperatures and strain rates. This testing method has proven to be adequate since it allows homogeneous deformation if lubrication fully reduces friction between the workpiece and tools. Under such conditions of constant and homogeneous strain rates and temperatures, a constitutive model can be derived by relating flow stress to nominal compression conditions. The homogeneous temperature rise in the sample due to adiabatic heating can be corrected by iterative procedures (Ref 13). Metallurgical analysis of the deformed samples reveals homogeneous microstructures, and thus, recrystallization kinetics can be quantified and empirical constitutive models can be adjusted accurately.

In the absence of dynamic recrystallization, the flow stress of aluminum alloys reaches a steady-state value at elevated temperatures due to concurrent work hardening and dynamic recovery. However, at combinations of high strain rates and low temperatures, flow stress does not saturate within the strain range of upsetting. For this reason, a generalized Voce equation, as proposed by Tomé et al. (Ref 18), is most suitable for aluminum alloys:

$$\sigma = \sigma_0(Z) + (\sigma_1(Z) + \Theta_1(Z) \cdot \varepsilon) \cdot$$
$$\left(1 - \exp\left(\frac{-\Theta_0(Z) \cdot \varepsilon}{\sigma_1(Z)}\right)\right) \quad \text{with}$$
$$Z = \dot{\varepsilon} \cdot \exp\left(\frac{Q}{R \cdot T_D}\right) \qquad \text{(Eq 1)}$$

where Z is the Zener-Hollomon parameter; σ_0 and $(\sigma_0 + \sigma_1)$ symbolize the flow stress at the onset of straining and at the onset of constant $d\sigma/d\varepsilon$, respectively; Θ_0 and Θ_1 are the corresponding strain-hardening coefficients; Q is the activation energy; R is the universal gas constant; and T_D is temperature. Equation 1 can be used over the whole range of industrial hot and cold forming conditions defined by different values of the Zener-Hollomon parameter Z. The generalized Voce law allows describing a saturation of flow stress at hot forming conditions as well as a steady work hardening with strain, especially at lower temperatures, as it occurs for magnesium-containing alloys. Also, a dependence of the Voce parameters on alloy composition can be formulated, which allows interpolating such data for different alloys.

For many materials, close-to-full sliding conditions can be achieved in uniaxial compression (Ref 19) by applying suitable lubrication on so-called Rastegaev specimens (cylindrical specimen with a defined groove milled into the face of the cylinder). However, this is not always the case for aluminum alloys at temperatures above 300 °C. Figure 7(a) shows the microstructure of a Rastegaev sample of AA3104 after inhomogeneous deformation due to severe surface friction. Strong gradients in static recrystallization are clearly visible, indicating an inhomogeneous state of stored energy immediately after deformation.

Sticking of the surface of aluminum alloy samples to the tool cannot always be prevented. However, when friction is known, such inhomogeneous tests can be of some advantage. Without any lubrication, complete sticking prevails that can be simulated accordingly (Fig. 7b). The quality of the prediction can be assessed from the barreling of the compressed specimen and also from the pattern of recrystallization (Fig. 7a); both agree well with the strain distribution determined by finite-element (FEM) simulation of the compression deformation under sticking conditions.

To derive suitable constitutive equations from inhomogeneous compression tests, the development of the local strain distribution during compression is analyzed. This requires a recursive method to provide the data, which again uses the constitutive equation as input. Such inverse computational methods have been successfully applied to this problem (Ref 14). A multidimensional optimization of FEM calculations is
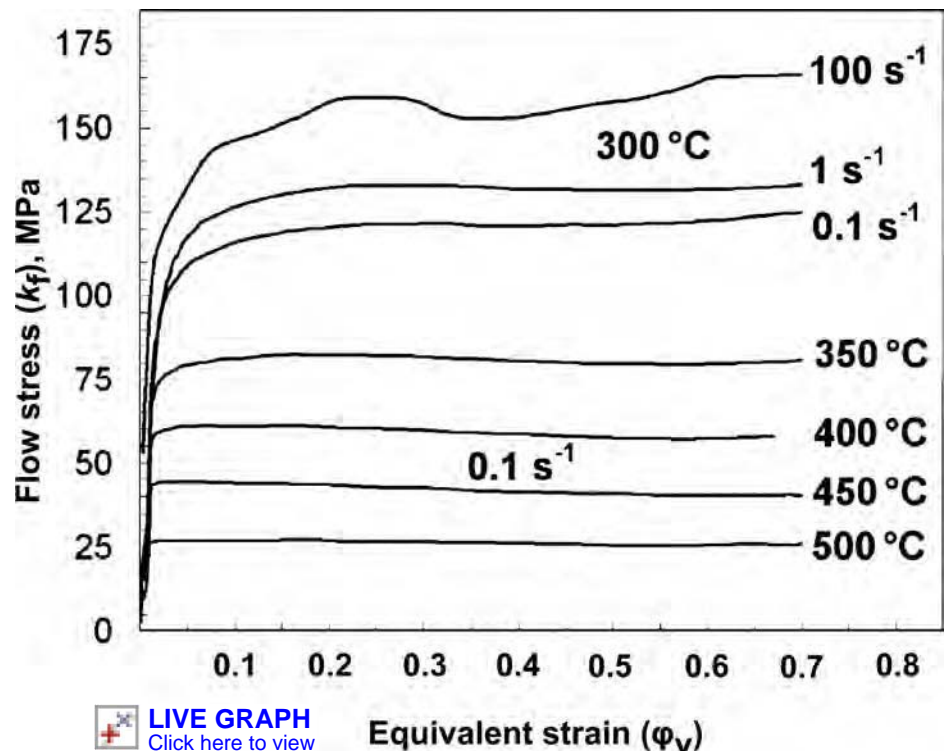
Fig. 6 Variation of flow stress during hot compression testing of AA 3104
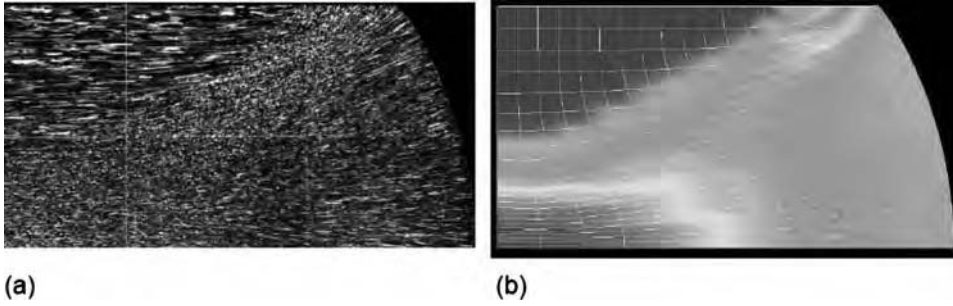
Fig. 7 (a) Microstructure of a cylindrical sample (¼ section) inhomogeneously hot compressed at 1 s⁻¹ at 350 °C, average strain of 0.8, partially recrystallized (after 10 s). (b) Corresponding finite-element method simulation. Local strain variation: blue = 0 (upper left); red = 1.54 (predominantly lower right)

LIVE GRAPH
Click here to view

LIVE GRAPH
Click here to view
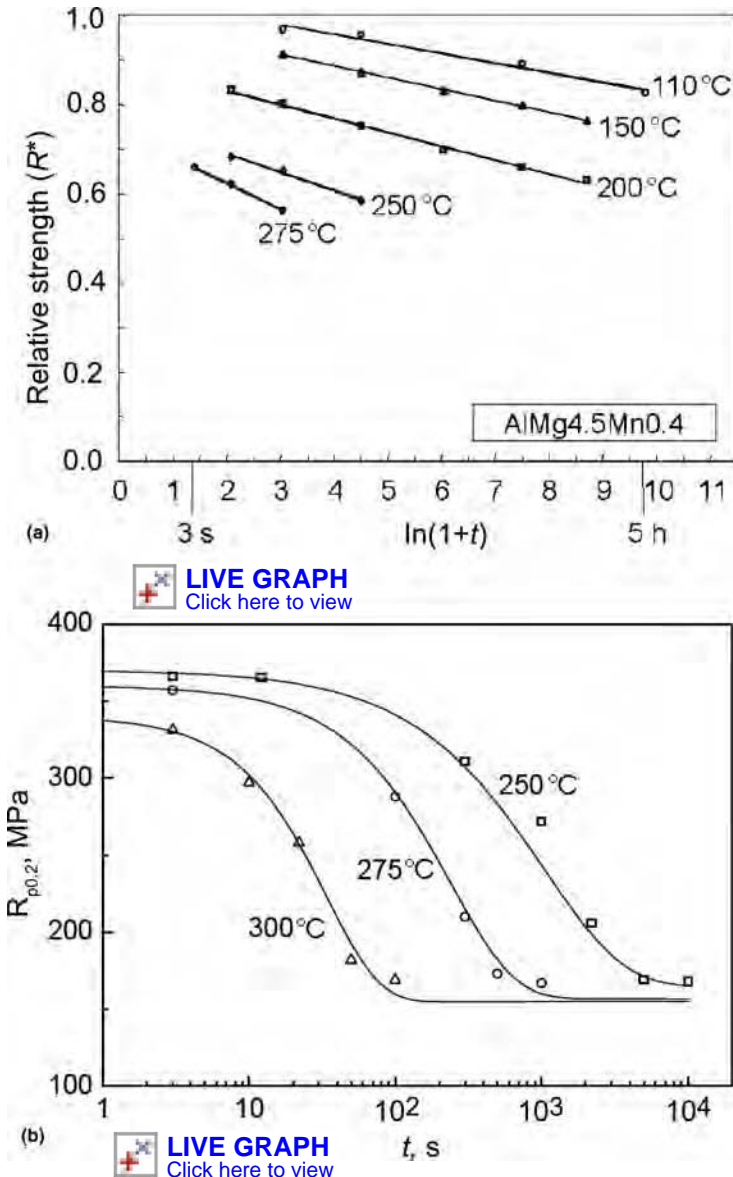
Fig. 8 Progress of (a) recovery and (b) recrystallization in AA 5182 for different annealing temperatures. Softening behavior by recrystallization is observed as a change in the 0.2% proof strength ($R_{p0.2}$).

performed in which the parameters of the generalized Voce equation (Eq 1) are adjusted by combinatory algorithms in an iterative procedure until the FEM code provides accurate force-displacement curves. Since the specimens are axially symmetric and sticking friction prevails, the evolving sample geometry is forced into a predescribed barrel shape. Therefore, the solution space with respect to an optimal constitutive equation is restricted, and reasonable convergence of the solving routine is achieved.

After hot deformation and/or during a subsequent annealing treatment, the material undergoes recovery due to annihilation of dislocations; thus, strength (σ) decreases with time (t). Figure 8(a) shows the effect of recovery on the relative yield strength (R^*) during annealing of a cold-rolled sheet of alloy AA 5182. In a logarithmic plot, a linear behavior can be observed, and recovery can be modeled on an empirical level by the following normalized form:

$$R^* = \frac{\sigma(t,T)}{\sigma_0} = 1 - a \cdot \ln(1 + b \cdot \tau) \quad \text{with}$$
$$\tau = \int_t dt \cdot \exp\left(-\frac{Q_{RV}}{RT}\right) \tag{Eq 2}$$

where R is the universal gas constant, T is temperature, Q_{RV} is the activation energy for recovery, and a and b are alloy-dependent parameters.

For simulation of recrystallization in terms of progress of recrystallized volume X with time t, a standard Avrami-type expression may be used:

$$X(t) = 1 - \exp\left[-0.693 \cdot \left(\frac{t}{t_{0.5}}\right)^n\right] \quad \text{with}$$
$$t_{0.5} = a \cdot Z^b \cdot \exp\left(\frac{Q_{RX}}{RT}\right) \tag{Eq 3}$$

where n, a, and b are alloy-dependent parameters and Q_{RX} is the apparent activation energy for recrystallization. As an example, Fig. 8(b) shows the softening behavior of cold-rolled AA 5182 at three different annealing temperatures. Evidently, Eq 3 gives a reasonable description of the progress of recrystallization and the resultant decrease in yield strength. (See the section "Simulation of Recovery and Recrystallization" that follows.)

Figure 9 shows an example of the integration of materials models into a full process-simulation environment. The temperature profile in a hot rolling line, consisting of a breakdown and three-stand tandem mill, has been simulated (Fig. 9a) using the coupled models RoseRoll, StrucSim, 4IVM, RoseTem, and RoseStat (Ref 17). The results of such integrated simulation models can be used for a quantitative description of the dislocation densities (Fig. 9b) and recrystallization kinetics and the resulting microstructure evolution as a function of process conditions, such as rolling temperature and velocity (Ref 12–17). Figure 10 illustrates the impact of interstand recrystallization on the texture of the hot strip for the high-magnesium alloy AA 5182. Here, the variations of cube
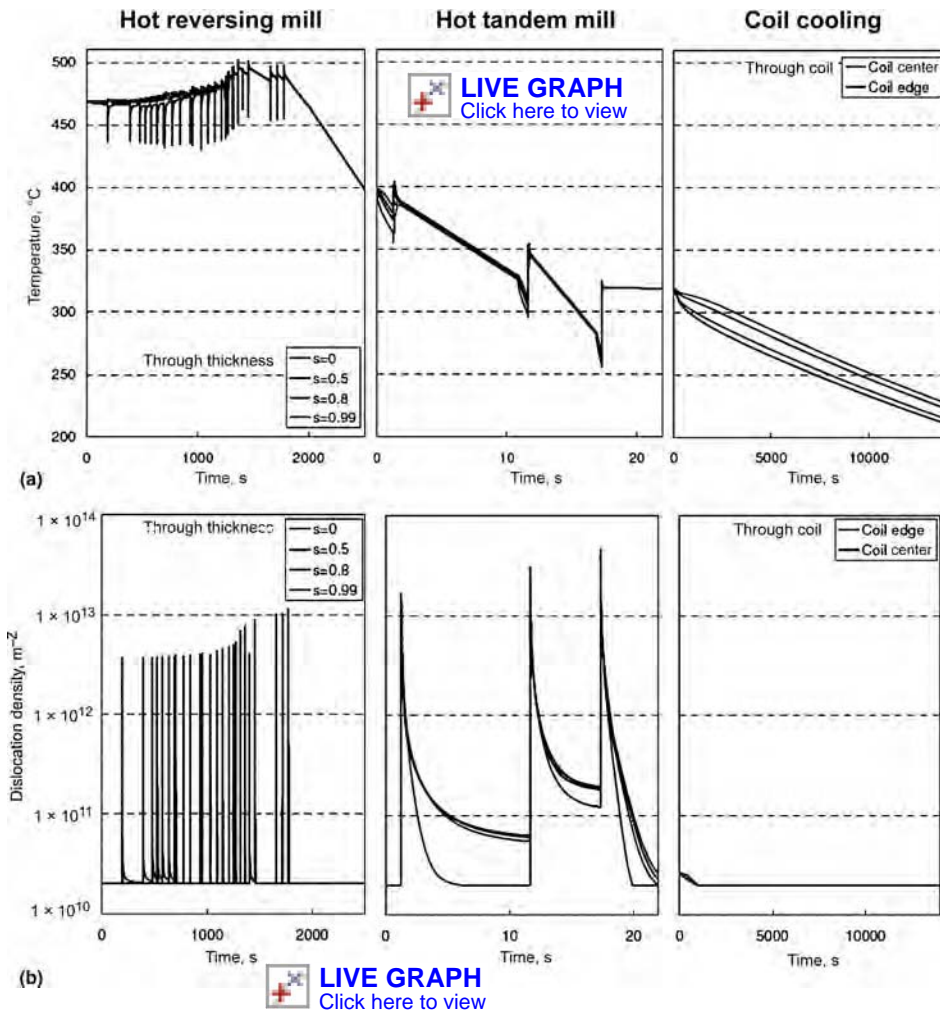
Fig. 9 Simulation of (a) temperature and (b) dislocation density in a breakdown and three-stand tandem mill hot rolling line, using integrated simulation models RoseRoll, StrucSim, 4IVM, RoseTem, and RoseStat (for details see Ref 17)

texture intensity $f(g)^{cube}$ and resulting anisotropy (in terms of the percentage of 0°/90° earing, Z) in the hot strip are plotted versus the (computed) recrystallized volume fraction $X(t)$ at the entry of the second stand (Ref 12). Despite the constant exit temperatures yielding fully recrystallized microstructures with similar grain sizes, systematic variations in texture occur that evidently have a strong impact on the earing behavior. In fast-recrystallizing high-magnesium alloys, a relatively low cube intensity usually develops (Fig. 4). For slow-recrystallizing alloys, tandem lines usually produce a distinct cube recrystallization texture together with a small grain size (Fig. 2c, 3c).

Microchemistry Simulation. Through-process microstructure simulation requires information about the state of the material with regard to its local chemistry — that is, variation in solute level and precipitation — which is subsumed under the term *microchemistry*. Today, a broad thermodynamic database is available to compute phase diagrams based on the well-known CAL-PHAD approach (*Cal*culation of *Pha*se *Dia*grams). For aluminum alloys, many of the

required data have been determined in the European research program COST 507 (Ref 20). One of the most sophisticated databases available — including the 16 most important alloy elements of wrought aluminum alloys (boron, calcium, chromium, copper, iron, lanthanum, magnesium, manganese, nickel, scandium, silicon, strontium, titanium, vanadium, zinc, zirconium) and capable of handling quaternary phases such as α-$Al_{15}(Fe, Mn)_3Si$ — is distributed commercially by ThermoTech Ltd. of the United Kingdom (Ref 21).

In this article the evolution in microchemistry during thermomechanical processing is calculated with a newly developed precipitation model termed ClaNG, which is based on the classical Becker-Döring theory for nucleation and growth of precipitates (Ref 22). A numerical implementation of this theory has the advantage that full size distributions can be tackled. The main improvement of the ClaNG model over former approaches is the introduction of multicomponent thermodynamics into a kinetics model through the programming library ChemApp, provided by GTT-Technologies of Germany. Chemical driving forces and

equilibrium compositions of phases are derived by calls of ChemApp using a subset of the ThermoTech aluminum database. The only parameters that are known only imprecisely are the interfacial energies. Hence, interfacial energies of the relevant phases were determined by inverse modeling of a set of simple model alloys with systematic variation of the composition (AlMn, AlMnFe, AlMnFeSi, ...). Finally, the equations for single nuclei are combined through the continuity equation, and therefore, entire size distributions of multicomponent and multiphase systems can be modeled.

To determine the microchemistry evolution along a given process chain, the resulting time/temperature profile is input into the model, for example, the time/temperature history of homogenization and hot rolling plus subsequent coil cooling and, if necessary, further intermediate or final annealing. If deformation processes are involved (e.g., upon hot or cold rolling), the dislocation density is derived from appropriate evolution equations, as described in Ref 17. Dislocations can both act as nucleation sites for particles enhancing the precipitation process and speed up the diffusion by pipe diffusion along the dislocation cores. A more detailed description of the ClaNG model, including the part describing the transformation of nonstochiometric phases with regard to their composition, can be found in Refs 23 to 25.

Simulation of Rolling Textures. In the literature, there are numerous attempts to model the texture evolution accompanying rolling with the help of Taylor-type deformation models (Ref 26, 27). In such approaches, the individual crystallites are assumed to deform by slip on a number of crystallographic slip systems so as to accommodate the prescribed macroscopic strain rate D_{ij}. With a view to a coupled through-process modeling, the strain states in the various sheet layers can be deduced from plastomechanical simulations that incorporate the effects of temperature and strain gradients on hot rolling results. Alternatively, for tandem hot rolling and especially for cold rolling of thin strip, the deformation is usually approximated by a plane-strain state, where $D_{11} = -D_{33}$, and all other strain-rate components are assumed to be zero. However, although the Taylor-type approaches are capable of simulating the main features of typical rolling textures, they all fail in reproducing the real texture evolution in any detail. The main reason for this poor texture simulation is ascribed to the fact that the classical one-point models do not take grain-to-grain interaction into consideration. Thus, two grains with the same orientation will behave absolutely identically, whereas, in reality, the two grains will deform differently, leading to different reorientations and thereby weaker rolling textures. More advanced multigrain or *N*-point formulations, such as the *LAMEL* model (Ref 28) or the grain interaction (*GIA*) model (Ref 29), indeed yield more realistic rolling textures, in particular with a significantly reduced rate of texture evolution. (For a
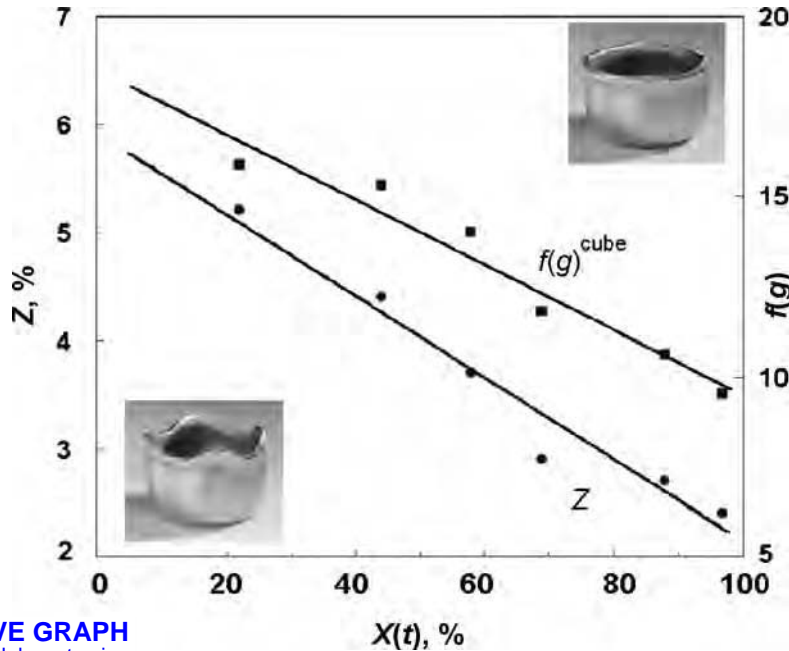
Fig. 10 Intensity of the cube texture, $f(g)^{cube}$, and resultant $0°/90°$ earing percentage, Z, of AA 5182 hot strip as a function of the amount of interstand recrystallization after the first hot rolling pass

quantitative assessment of various texture models, see Ref 30).

Throughout this article, the latter approach for simulating rolling textures is used. The GIA model considers the deformation of an aggregate of eight brick-shaped grains embedded in a homogeneous surrounding. While the deformation of the whole aggregate is fully prescribed, a compatible relaxation of the deformation of the individual grains within the aggregate is possible. The resulting misfits are accommodated by a simplified arrangement of geometrically necessary dislocations (GND). The amount of relaxation of the shears, the slip rates, and, ultimately, the grain reorientation is derived from a minimization of the total plastic work due to slip of the eight grains plus the energy introduced by the GNDs. In combination with two recent refinements — random perturbations of the overall plane-strain state and consideration of material parameters (work hardening and grain size) — the GIA model has proven to be able to simulate the rolling textures of aluminum alloys with high accuracy (Ref 31).

Simulation of Recovery and Recrystallization. To calculate the softening behavior between the various passes of multipass hot rolling and during coil cooling, the analytical softening model AlSoft is used (Ref 32–35). This model is based on a two-parameter description of the microstructure of the as-deformed or recovered state, where the microstructure is composed of subgrains with average size δ and dislocations in the subgrain interior with density ρ_i. After hot deformation, there is a simple phenomenological relationship of subgrain size, δ, with deformation temperature, T_D, and strain rate, $\dot{\varepsilon}$ (Ref 32):

$$\frac{1}{\delta} = \frac{RT_D}{A^*} \ln\left(\frac{Z\delta^2}{B^*}\right)$$

with the Zener-Hollomon parameter (Eq 4)

$$Z = \dot{\varepsilon} \cdot \exp\left(\frac{Q}{RT_D}\right)$$

where R is the universal gas constant, A^* and B^* are alloy-dependent constants, and Q is the activation energy. Subgrain size δ and dislocation density ρ_i are linked through the principle of similitude, $\sqrt{\rho_i} = C/\delta$, where C is a constant of the order of 2.

Recovery kinetics is assumed to be controlled by solute drag, where the rate-controlling mechanism is given by the thermal activation of the alloy elements in solid solution. During recovery, the increase in subgrain size (δ) and decrease in dislocation density (ρ) as a function of time (t) and temperature (T) is described by the following differential equations:

$$\rho_t = -v_D b A_\rho B_\rho \rho_i^{3/2} \cdot \exp\left(-\frac{U_{RV}}{RT}\right) \cdot 2$$
$$\cdot \sinh\left(\frac{A_\rho \mu b^4 \sqrt{\rho_1}}{kT}\right) \quad \text{with} \quad A_\rho = \omega_\rho c_{ss}^{-2/3}$$
(Eq 5)

$$\dot{\delta} = -v_D b A_\delta B_\delta \cdot \exp\left(-\frac{U_{RV}}{RT}\right) \cdot 2 \cdot \sinh\left(\frac{A_\delta \mu b^4}{\delta kT}\right)$$
$$\text{with} \quad A_\delta = \omega_\delta c_{ss}^{-2/3}$$
(Eq 6)

where v_D is the Debye frequency, k is the Boltzmann constant, ω_i are constants of the order of unity, and B_i are alloy-specific constants that must be fitted to the actual experiments. The total concentration of solutes, c_{ss}, is derived from summation of the solute concentration of the individual alloy elements, weighted by their activation energy for diffusion (Ref 35). U_{RV} is an activation energy, which, in the case of solute drag, equals that for diffusion of the solutes. Here, the value of the slowest element in the alloy, that is, manganese, was adopted, so that $U_{RV} = 200$ kJ/mol. The recrystallization model is an extension of the classical Johnson-Mehl-Avrami-Kolmogorov approach, treating recrystallization as a nucleation and growth process:

$$\dot{X} = (1 - X(t)) \cdot N(t) \cdot 4\pi \cdot r(t)^2 \cdot G(t) \quad \text{(Eq 7)}$$

where $r(t)$ is the size and $G(t)$ is the growth rate of the recrystallization nuclei. $N(t)$ is the density of nuclei per unit volume, which critically depends on both the processing parameters (strain, strain rate, deformation temperature) and the microstructure of the material analyzed. The derivation of the density of recrystallization nuclei, $N(t)$, as well as simulation of the resulting recrystallization textures, has already been discussed in detail (Ref 36, 37).

The size of the nuclei, $r(t)$, is linked to their growth rate, $G(t)$, through:

$$\dot{r} = G(t) \quad \text{(Eq 8)}$$

where the latter can be expressed through the well-know relation:

$$G(t) = m(t) \cdot (P_D(t) - p_Z) \quad \text{(Eq 9)}$$

Here, $m(t)$ is the mobility, which depends on temperature with the following Arrhenius relation:

$$m(t) = \frac{m_0}{c_{SS}kT} \cdot \exp\left(-\frac{U_{RX}}{RT}\right) \quad \text{(Eq 10)}$$

where U_{RX} is the activation energy for migration of high-angle grain boundaries, p_z is the back-driving force due to finely dispersed particles (Zener drag), and p_D is the driving force for recrystallization due to dislocations and subgrains (see the aforementioned):

$$P_D(t) = \frac{\alpha_2 \gamma_{SB}}{\delta(t)} + \frac{1}{2}\mu b^2 \cdot \rho_i(t) \quad \text{(Eq 11)}$$

where μ and b denote the shear modulus and Burgers vector of aluminum, respectively; α_2 is a constant of the order of 3; and γ_{SB} denotes the specific energy of subgrain boundaries. Note that the required microchemistry data — including the particle density entering into the Zener drag, p_Z (Eq 9), as well as the concentration of solids, c_{ss} (Eq. 5, 6, 10) — are provided by the microchemistry model ClaNG. In addition to providing the microstructural changes during recovery and recrystallization in terms of changes in dislocation density and subgrain size, the model provides the flow stress, σ, recrystallized volume fraction, $X(t)$, grain size, D_{RX}, and texture changes upon recrystallization.

It is seen that the recrystallization model is governed by four differential equations (Eq. 5–8), which are solved for a given cycle of time (t) and temperature (T) by numerical routines with an adaptive-step size control. For simulation of isothermal annealing between two consecutive hot rolling passes, T is set to the rolling temperature, while t increases up to the total annealing time. For simulation of self-annealing, T decreases from the coiling temperature (T_{coil}) of the hot strip down to ambient temperature according to $T(t) = T_{coil} - \dot{T} \cdot t$, where $\dot{T}$ represents the cooling rate.

Polycrystal-Plasticity Modeling of Earing Profiles. Simulation of texture-induced anisotropy, such as earing during cup deep drawing, is performed with a polycrystal-plasticity model (Ref 11, 38) that is based on the visco-plastic self-consistent (VPSC) scheme advanced by Lebensohn and Tomé (Ref 39). To derive the boundary conditions for the VPSC computations, the evolution of stresses and strains operating during cup deep drawing were studied by FEM (Ref 38). These FEM simulations have shown that plastic flow and therefore the formation of ears and troughs is concentrated in the flange of the blank under the blankholder. In the flange, a plane-stress state prevails in that the material is exposed to a tensile stress in the radial direction (σ_r) and a compressive stress in the tangential direction (σ_t). The stress in the through-thickness direction (σ_z) as well as the off-diagonal shear stresses (σ_{rt}) are almost zero. This stress state generates a positive radial strain rate (D_r) and a negative tangential strain rate (D_t) while the through-thickness strain rate (D_z) is not necessarily zero. Thus, making use of a feature of the VPSC model that permits enforcing mixed displacement/stress-boundary conditions on the polycrystalline aggregate, the following load case is imposed:

$$D_{ij} = \begin{pmatrix} * & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & * \end{pmatrix} \cdot \dot{\varepsilon},$$

$$\sigma_{ij} = \begin{pmatrix} \sigma_r & * & * \\ * & * & * \\ * & * & \sigma_z = 0 \end{pmatrix} \quad \text{(Eq 12)}$$

The asterisks (*) indicate that the corresponding values are not prescribed; $\dot{\varepsilon}$ is a scalar measure of the strain rate (s^{-1}). It is seen that the stress components σ_r and σ_z are prescribed rather than the corresponding strain-rate components D_r and D_z. All simulations in this study were performed with a nonzero radial stress of $\sigma_r = -0.25 \cdot \sigma_t$ (Ref 40). Because the stress values are not yet known in the first time step, this is accomplished iteratively within a few VPSC steps until stable values for σ_t and thereby σ_r are achieved.

Equation 12 is expressed in the frame defined by radial direction (r), tangential direction (t), and through-thickness direction (z). This frame is related to the standard rolling frame {RD, TD, ND} through a rotation by angle α about the normal direction z = ND. Thus, to derive the earing tendency under a given angle α, the strain in radial direction is determined for the texture in this particular position of the blank. For this purpose, the strain and stress tensors, or alternatively, the sheet texture, must be rotated by $\triangle\alpha$ in steps of typically 5° about z. The earing profile along the rim of a deep-drawn cup is then given by the radial strain or, more precisely, by the normalized strain-rate ratio $q = D_r / D_t$ as a function of α.

The absolute height of a deep-drawn cup is controlled by the radii of initial blanks (r_B) and punch (r_P) and, to a lesser extent, by sheet thickness (t_0) and the radius of the punch profile (r_P). As detailed in Ref 41, the exact cup height profile, $h(\alpha)$, can be computed from the strain-rate ratio (q) as a function of the in-plane angle (α) from:

$$h(\alpha) = t_0 + (1 - \pi/4)(2r_{pp} + t_0)$$
$$+ \frac{r_B^{q+1} - (r_P + t_0)^{q+1}}{(q+1)(r_P + t_0/2)^q} \quad \text{(Eq 13)}$$

Coupled Through-Process Simulation of Microstructure and Texture Evolution in AA 3104

Simulation of Recrystallization during Hot Rolling of AA 3104 Hot Strip. In this section, the recrystallization behavior of alloy AA 3104 during the interstand times in between two consecutive hot rolling passes is tracked with the help of the combined microstructure models described earlier. The changes in microchemistry along the thermomechanical process chain — representing an essential input for the softening module AlSoft — were computed with the microchemistry model ClaNG. Figure 11 shows the time/temperature cycle during a typical two-step homogenization practice of can body stock plus subsequent hot rolling and coil cooling. The ingot is heated to 600 °C, held for a few hours, cooled to 500 °C, discharged after a further 2 h, and then hot rolled. The entire hot rolling (i.e., breakdown plus tandem rolling) takes place in a few minutes, such that on the time scale of Fig. 11, hot rolling is represented by an almost vertical decrease in temperature to the exit temperature (in this example, 320 °C). Finally, the coiled hot strip slowly cools down to ambient temperature. Furthermore, Fig. 11(a) provides information on the solute level (c_{ss}) of the five most important alloying elements of alloy AA 3104, that is, magnesium, manganese, copper, silicon, and iron. This information is required for the recovery and growth kinetics equations of the softening model (Eq. 5, 6, 10). Figure 11(b) shows the evolution of precipitated phases, that is, size (radius) and volume fraction of β-Al$_6$(Mn,Fe) (dotted lines) and α-Al$_{15}$(Mn, Fe)$_3$Si$_2$ (solid lines) dispersoids, which retard recrystallization through exerting a back-driving force (Zener drag, p_Z, see Eq 9).

It is seen that during heating of the supersaturated as-cast ingot, beginning at a temperature of approximately 350 °C up to the homogenization temperature of 600 °C, β-Al$_6$(Mn,Fe) dispersoids with sizes below 0.1 μm form. During the soaking period, the β-particles coarsen, while their volume fraction decreases slightly, pointing at coarsening and partial redissolution of these phases. The homogenized ingot is then cooled down to the hot rolling temperature of approximately 500 °C. This is accompanied by further precipitation along with growth of existing β-particles, plus some constitutional changes toward α-Al$_{15}$(Mn,Fe)$_3$Si$_2$ particles (Ref 23). Thus, as already alluded to earlier, the microchemistry of AA 3104 in terms of solutes and dispersoid phases can be fine-tuned by suitable heat treatment conditions so as to control recrystallization of the hot strip and thereby achieve optimal final-gage properties.

The texture evolution during deformation is simulated with the GIA model. Hot rolling at elevated temperature is taken into account by allowing the activation of additional nonoctahedral slip systems of type {011}<011> (Ref 42), as discussed in more detail elsewhere (Ref 43) ($\tau\{111\}/\tau\{110\} = 1/1.15$). The kinetics of recovery and recrystallization, as well as the resulting textural changes, are modeled with the softening tool AlSoft.

For the analysis of interstand recrystallization, the temperature is assumed to stay constant between two consecutive rolling passes (isothermal annealing). As described previously, the driving force is derived from the deformation conditions through the Zener-Hollomon parameter, Z (Eq 4). Tandem hot rolling commences at approximately 450 °C with a strain rate of the order of 10 s^{-1}, which corresponds to a Zener-Hollomon parameter $Z \approx 10^{12}$ s^{-1}. During subsequent passes, the deformation temperature drops, while the strain rate increases significantly. This leads to an increase of the Zener-Hollomon parameter to values of up to $Z \approx 10^{16}$ s^{-1} at the fourth pass.

Figure 12 shows the progress of recrystallization after the various tandem passes in terms of the recrystallized volume fraction (X) as a function of time (t). After the first pass, the temperature is quite high, such that recrystallization would occur in approximately 25 s. However, considering the high rolling velocities in industrial tandem hot rolling, the times that the material actually spends between the various rolling stands are very short, at the most a few seconds. Thus, the interstand time of 5 to 10 s between first and second rolling pass only allows for partial recrystallization, with volume fractions X ranging from ~5 to 40% (shaded in Fig. 12). After the following rolling passes, the temperatures are lower (i.e., higher Z), such that recrystallization kinetics are slowed, although the driving force increases slightly (Eq 11). Furthermore, the interstand times decrease appreciably because of the increase in rolling velocity. Hence, after the later rolling passes, interstand recrystallization is much less likely to occur than after the first pass. This result is in accord with earlier thermomechanical simulations that have

(a)

(b)

Fig. 11 Results of the ClaNG microchemistry simulation of the evolution of (a) concentration of solutes (c_{ss}) and (b) volume (V) and size (r) of precipitating dispersoids during the thermomechanical processing of AA 3104 hot strip

Fig. 12 Simulation of the progress of recrystallization in AA 3104 after the various passes of a four-stand tandem mill

rolling pass. After hot rolling is complete, the hot strip is assumed to recrystallize completely (this point is qualified in a later section).

In a first trial, hot rolling is assumed to take place without recrystallization between the various rolling passes. Thus, the rolling texture evolution was simulated up to a total strain of ~93% (ε = 2.8) prior to complete recrystallization. Starting from the weak cube recrystallization texture of the transfer slab (Fig. 13a), hot rolling without interstand recrystallization gives rise to the formation of a well-defined rolling texture (Fig. 13b). During subsequent recrystallization, a pronounced cube texture with a maximum intensity of ~24 develops (Fig. 13c), which is notably sharper than the experimental hot band texture commonly observed after self-annealing (e.g., Fig. 3c). It may be noted in this context that the cube texture of the partially recrystallized hot strip obtained at a lower coiling temperature (Fig. 3b) is actually sharper than that of the fully recrystallized hot strip shown in Fig. 3(c). Since the lower coiling temperature is controlled by a lower rolling temperature, it is likely that this material has experienced less or no interstand recrystallization. Thus, the increase in texture sharpness of the partially recrystallized hot strip may well be linked to the lower level of interstand recrystallization in this material.

Figure 14(a) shows the simulation results obtained under the assumption of 20% interstand recrystallization after the first tandem rolling pass. The first pass is accompanied by formation of a weak rolling texture, yet the strain (50%) is too low to give rise to a pronounced texture. Correspondingly, there is still a significant amount of cube orientation left after rolling. During (partial) recrystallization, the cube texture sharpens further, which is attributed to its favorable nucleation conditions at this stage. During subsequent rolling passes,

shown that in modern high-speed tandem mills, AA 3104 can stock does not usually recrystallize.

Evidently, prediction of the exact amount of interstand recrystallization is critically dependent on the exact details of the softening model. Hence, in the present investigation, the impact of interstand recrystallization on the texture of the recrystallized hot strip is examined in a parametric study. For that purpose, the rolling texture evolution is simulated in four idealized 50% passes (ε = 0.7) up to a total strain of ~93% (ε = 2.8). The aforementioned softening simulations have shown that in high-speed tandem rolling of AA 3104, only incomplete recrystallization may occur (Fig. 12). Therefore, in the parametric study, partial recrystallization — with recrystallized volume fractions X ranging from 0 to 50% — is included after the first

(a)

(b)

(c)

Fig. 13 Simulation of the texture during hot rolling and subsequent self-annealing without interpass recrystallization (orientation distribution function $\varphi_2 = 0°$ sections, intensity levels 1 - 2 - 4 - 7 - 10 - 15). (a) As-received plate (transfer gage). (b) Four passes, no interstand recrystallization. (c) Four passes, no interstand recrystallization

no further recrystallization is assumed to occur. Nonetheless, driving forces and texture sharpness accumulated upon the later rolling passes

(a)

(b)
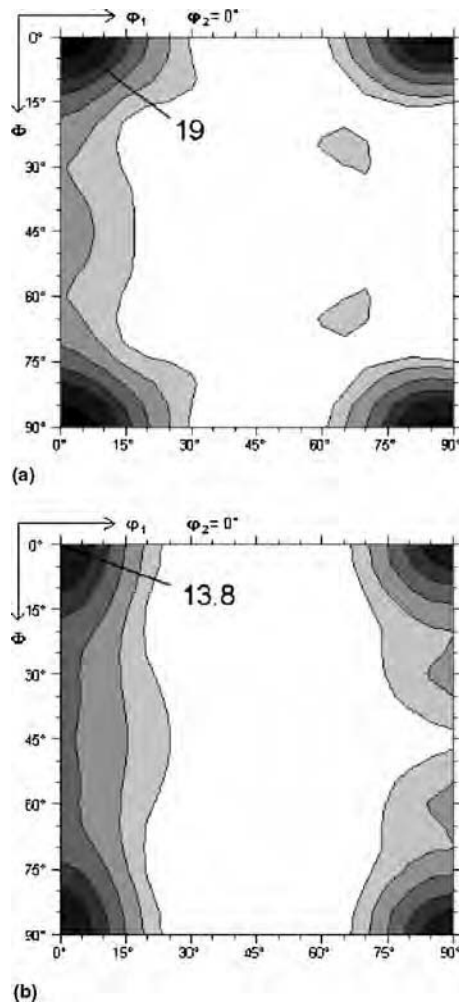
Fig. 14 Simulation of the texture during hot rolling and subsequent self-annealing with (a) 20% interpass recrystallization after the first pass and (b) complete interpass recrystallization after every pass (orientation distribution function $\varphi_2 = 0°$ sections, intensity levels 1 - 2 - 4 - 7 - 10 - 15)

are weaker because of the softening and partial recrystallization after pass 1. Hence, the texture sharpness of the recrystallized hot strip displayed in Fig. 14(a) is weaker than that simulated without interstand recrystallization (Fig. 13c). Note that this texture is in very good agreement with the experimental hot band texture obtained after self-annealing (Fig. 3c).

As an extreme case, complete interstand recrystallization ($X = 100\%$) is assumed to occur after each rolling pass. Here, the alternating cycles of deformation and recrystallization counteract the accumulation of large driving forces and the formation of a well-defined rolling texture. Hence, in comparison to the hot band texture simulated under the assumption of no or limited interstand recrystallization, the resulting hot strip recrystallization texture (Fig. 14b) is notably weaker. Note that this texture compares quite well with the hot strip texture observed in the high-magnesium alloy AA
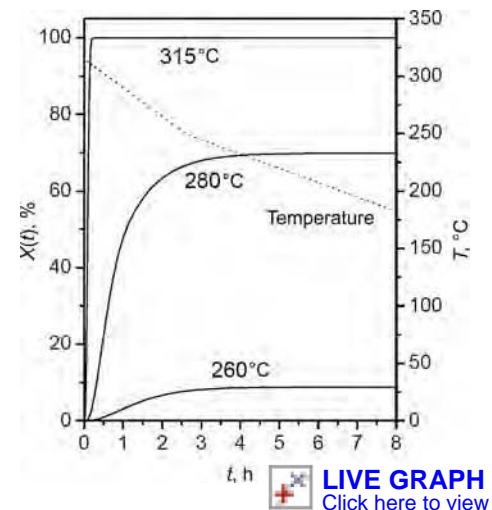
Fig. 15 Simulation of the progress of recrystallization during coil cooling (self-annealing) of AA 3104 hot strip produced with different exit temperatures

5182 (Fig. 4), where significant interstand recrystallization is likely to occur.

Simulation of Recrystallization and Texture during Self-Annealing of AA 3104 Hot Strip. So far, it has been assumed that complete recrystallization takes place during self-annealing, which, obviously, is too simple an assumption. The results shown in Fig. 2 and 3 imply that coiling temperatures below 300 °C are too low to enable full recrystallization of AA 3104. Figure 15 shows simulation results for the progress of recrystallization for the three hot strip exit temperatures of 260, 280, and 315 °C. A typical time/temperature history during coil cooling is given in Fig. 9(a). Obviously, the resulting recrystallization kinetics data agree qualitatively well with the aforementioned findings that coiling temperatures in excess of 300 °C are required to enable full recrystallization during self-annealing of the coiled hot strip (Fig. 2, 3).

The texture of an incompletely recrystallized hot strip can be simulated by superimposing the rolling texture orientations with the newly forming recrystallization texture components, with a ratio according to the recrystallized fraction X. Following the aforementioned recrystallization kinetics simulations (Fig. 15), at hot strip exit temperatures of 260 and 280 °C, recrystallization should proceed by approximately 10 and 70%, respectively; Fig. 16 displays the simulated textures obtained for these two cases. Both textures comprise mixed rolling and recrystallization textures, which are qualitatively similar to the experimental hot strip textures obtained with the corresponding coiling exit temperatures (Fig. 3a, b). However, the ratio of rolling and recrystallization texture components, including the exact texture intensities, is not matched that well. This again points at the high sensitivity of through-process simulations on the precise prediction of recrystallization and especially recrystallization kinetics.

(a)

(b)
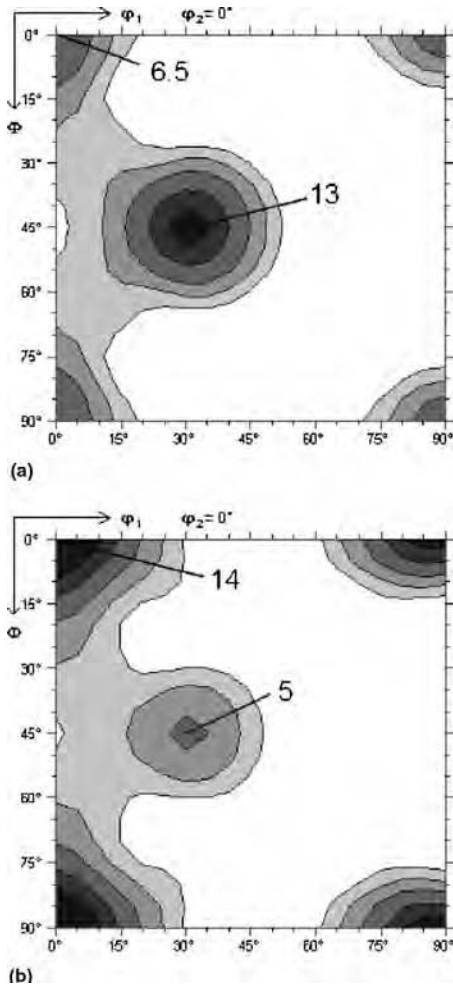
Fig. 16 Texture of partially recrystallized hot strip simulated under the assumption of different amounts of partial recrystallization, $X(t)$, during self-annealing (orientation distribution function φ_2 = 0° sections, intensity levels 1 - 2 - 4 - 7 - 10 - 15). (a) X $(t) = 10\%$. (b) $X(t) = 70\%$

Simulation of Texture and Resulting Earing Profiles during Cold Rolling of AA 3104 Strip.

As mentioned previously, upon cold rolling of a hot strip with a cube recrystallization texture, there will be a transition from 0°/90° toward 45° earing. This behavior is traced by simulating the texture evolution accompanying cold rolling with the GIA model; the simulated rolling textures are then used as an input for the subsequent polycrystal-plasticity earing simulations. Three test cases are selected on the basis of showing distinctively different hot strip textures:

- Fully recrystallized hot strip obtained without interstand recrystallization with a too-sharp cube recrystallization texture (Fig. 13c)
- Fully recrystallized hot strip with 20% interstand recrystallization with a regular-strength cube recrystallization texture (Fig. 14a)
- Incompletely (70%) recrystallized hot strip without interstand recrystallization with a mixed rolling/recrystallization texture (Fig. 16b).
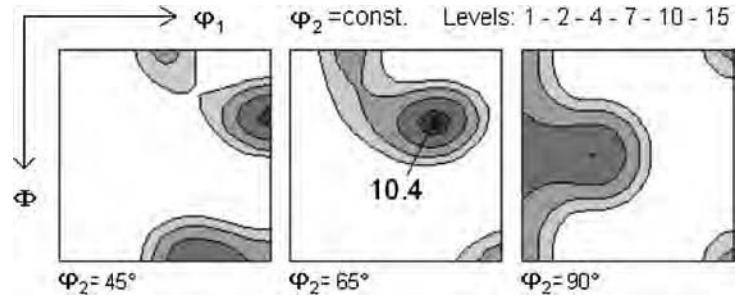
Fig. 17 Simulated cold rolling texture (ε = 1.9) of hot strip with regular-strength cube texture (Fig. 14a). Compare with experimental texture of Fig. 5(a).
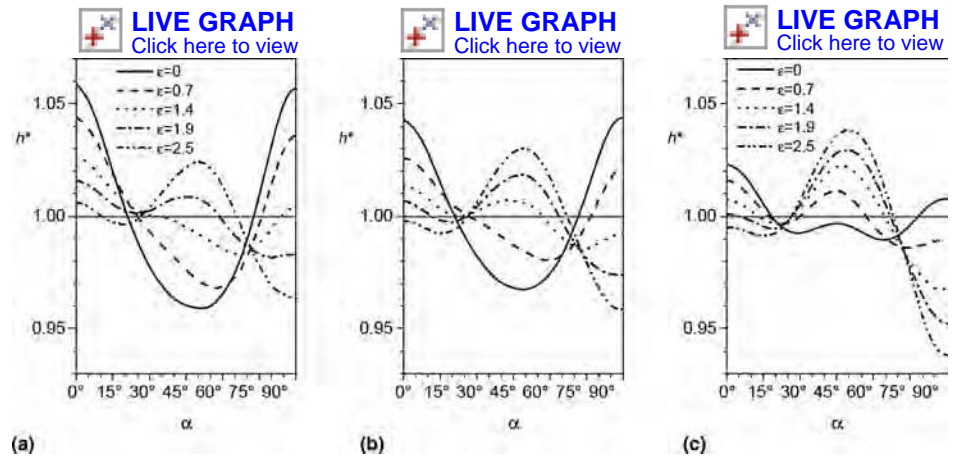
Fig. 18 Evolution of earing profiles as a function of cold rolling strain simulated for three AA 3104 hot strips with various initial (hot strip) textures (see text for details). (a) No interstand recrystallization (0%)/fully recrystallized hot strip. (b) Partial interstand recrystallization (20% after first pass)/fully recrystallized hot strip. (c) No interstand recrystallization/partially recrystallized hot strip (70%)

The simulated hot strip textures are discretized into sets of approximately 1000 individual orientations. Then rolling is simulated with the GIA model for strains (ε) of 0.7, 1.4, 1.9, and 2.5, corresponding to thickness reductions of 50, 75, 85, and 92%, respectively. Figure 17 shows an example of the simulated rolling textures for 85% rolling of the hot strip with regular-strength cube texture (Fig. 14a). Comparison with the experimental texture given in Fig. 5(a) validates the quality of rolling-texture simulations achievable with the refined GIA model (Ref 31).

To simulate earing, the modeled rolling textures are entered into the polycrystal-plasticity earing model. The evolution of the earing profiles modeled for the hot strip with limited interstand recrystallization (20%) very much resembles the behavior of standard can body stock (Ref 3, 11, 15, 38), where, with increasing level of cold rolling, 45° ears form at the expense of the original 0°/90° ears (Fig. 18b). At cold rolling reductions in excess of 75% ($\varepsilon \geq 1.4$), balanced earing profiles with six ears form, which are well suited for the production of beverage cans (compare with Fig. 5).

For the hot strip with no interstand recrystallization (0%), the overall evolution of earing is similar to that obtained for the regular-strength hot strip, yet the transition toward 45° earing is delayed toward higher strains (Fig. 18a). In consequence, at lower strains (up to strains, ε, of almost 2, or 85%), excessive 0/180° ears may prevail (Ref 11). At very high strains, earing is less pronounced, such that higher strains may be tolerable without ending up with too-high 45° earing. In total, the range with balanced earing is narrower than in the regular-strength material.

The non-fully recrystallized hot strip (labeled 70%) shows the opposite behavior. Here, the mixed texture, consisting of cube recrystallization texture plus the remaining rolling texture orientations, results in eightfold earing, with distinct 45° ears already in the hot strip. Upon cold rolling, the 45° ears intensify, leading to pronounced four-fold 45° earing already after moderate strains ($\varepsilon \geq$ 1.4; Fig. 18c). Accordingly, the range with suitable earing profiles is shifted toward smaller reductions, where the material strength is usually not sufficient.

Summary and Conclusions

The development of a comprehensive through-process model of microstructure and texture evolution during the thermomechanical processing of aluminum sheet products necessitates coupling of models for simulating the evolution of microstructure, microchemistry, and crystallographic texture during the various process steps involved.

This article has focused on linking softening models to simulate recovery and recrystallization to deformation and microchemistry models. As an example, the texture evolution during the production of AA 3104 is simulated, starting from the as-cast ingot to final-gage cold strip. The resulting textures were then input into a polycrystal-plasticity model to simulate the earing behavior of the rolled strip as a function of comprehensive consideration of the upstream thermomechanical processing steps.

It appears that the coupled model approaches are capable of reproducing the most important effects regarding texture and related properties. In particular, the impact of interstand recrystallization during tandem hot rolling as well as recrystallization during self-annealing on the texture of the coiled hot strip and, in consequence, on texture and earing properties at final gage are well captured. This illustrates the potential of simulations of texture and resulting texture-related properties by coupling models for rolling and recrystallization texture evolution. However, it must be noted that many aspects of the present through-process simulation are still rather crude. The individual models are not fully integrated, which requires somewhat tedious, manual transfer of the data between the various models. Furthermore, simulation of recovery and recrystallization kinetics largely depends on the exact choice of the various parameters of the softening model, such that the predicted values of the recrystallized volume fraction $X(t)$ may be subject to quite large errors. Despite this uncertainty, the afore mentioned examples point at the prospect of performing parametric studies or, more generally speaking, case or sensitivity studies of certain aspects of the process chain so as to scrutinize the impact of the decisive steps of the thermomechanical processing on the final product properties.

ACKNOWLEDGMENT

REFERENCES

1. W.B. Hutchinson, A. Oscarsson, and Å. Karlsson, *Mater. Sci. Technol.,* 5, 1989, p. 1118.
2. B. Ren, *Aluminum Alloys for Packaging III,* S.K. Das, Ed., TMS (Warrendale, PA), 1998, p. 49.
3. J. Hirsch, *Mater. Sci. Forum,* 495–497, 2005, p. 1565.
4. J. Hirsch, *Virtual Fabrication of Aluminium Products,* J. Hirsch, Ed., Wiley-VCH (Weinheim, Germany), 2006, p. 27.
5. K. Marthinsen, S. Abtahi, K. Sjølstad, B. Holmedal, E. Nes, A. Johansen, J.A. Sæter, T. Furu, O. Engler, Z.J. Lok, J. Talamantes-Silva, C. Allen, and C. Liu, *Aluminium,* 80, 2004, p. 729.
6. M. Goerdeler, M. Crumbach, P. Mukhopadhyay, G. Gottstein, L. Neumann, and R. Kopp, *Aluminium,* 80, 2004, p. 666.
7. M. Crumbach, L. Neumann, M. Goerdeler, H. Aretz, G. Gottstein, and R. Kopp, *Modell. Simul. Mater. Sci. Eng.,* 14, 2006, p. 835.
8. K.F. Karhausen and J. Hirsch, *Virtual Fabrication of Aluminium Products,* J. Hirsch, Ed., Wiley-VCH (Weinheim, Germany), 2006, p. 315.
9. O. Engler, L. Löchte, and J. Hirsch, *Acta Mater.,* 55, 2007, p. 5449.
10. V. Randle and O. Engler, *Introduction to Texture Analysis, Macrotexture, Microtexture and Orientation Mapping,* Gordon & Breach (Amsterdam), 2000.
11. O. Engler and J. Hirsch, *Mater. Sci. Eng. A,* 452–453, 2007, p. 640.
12. J. Hirsch, K. Karhausen, and R. Kopp, *Proc. ICAA 4,* T.H. Sanders and E.A. Starke, Ed., Georgia Inst. of Tech. (Atlanta, GA), 1994, p. 476.
13. K.F. Karhausen, *Integrierte Prozeß- und Gefügesimulation bei der Warmumformung,* Umformtechnische Schriften Bd. 52, Verlag Stahleisen (Düsseldorf), 1995.
14. K.F. Karhausen, M. Tschirnich, R. Luce, and R. Kopp, *Proc. 3rd Conf. on Modeling of Metal Rolling Processes,* J. Beynon et al., Ed., IOM Communications (London), 1999, p. 300.
15. J. Hirsch, K.F. Karhausen, and P. Wagner, *Mater. Sci. Forum,* 331–337, 2000, p. 421.
16. J. Hirsch, K.F. Karhausen, and O. Engler, *Continuum Scale Simulation of Engineering Materials,* D. Raabe, F. Roters, F. Barlat, and L.Q. Chen, Ed., Wiley-VCH (Weinheim, Germany), 2004, p. 705.
17. K.F. Karhausen, L. Neumann, R. Kopp, M. Goerdeler, and G. Gottstein, *Aluminium,* 80, 2004, p. 690.
18. C. Tomé, G.R. Canova, U.F. Kocks, N. Christodoulou, and J.J. Jonas, *Acta Metall.,* 32, 1984, p. 1637.
19. R. Kopp, M. Heußen, F.-D. Philipp, and K.F. Karhausen, *Steel Research,* 64, 1993, p. 377.
20. I. Ansara, A.T. Dinsdale, and M.H. Rand, *Thermochemical Database for Light Metal Alloys* (final report COST 507 V2), European Commission (Brussels), 1998.
21. N. Saunders, *J. Japanese Inst. Light Metals,* 51, 2001, p. 141.
22. R. Becker and W. Döring, *Ann. Phys.,* 32, 1938, p. 128.
23. M. Schneider, G. Gottstein, L. Löchte, and J. Hirsch, *Mater. Sci. Forum,* 396–402, 2002, p. 637.
24. G. Gottstein, M. Schneider, and L. Löchte, *Proc. ICAA 9,* J.F. Nie, A.J. Morton, and B.C. Muddle, Ed., Inst. Mater. Eng. Australasia (Melbourne), 2004, p. 1116.
25. L. Löchte, M. Schneider, and G. Gottstein, *Aluminium,* 80, 2004, p. 685.
26. J. Hirsch and K. Lücke, *Acta Metall.,* 36, 1988, p. 2883.
27. U.F. Kocks, C.N. Tomé, and H.R. Wenk, *Texture and Anisotropy, Preferred Orientations and Their Effect on Materials Properties,* Cambridge Univ. Press (Cambridge), 1998.
28. P. Van Houtte, S. Li, M. Seefeldt, and L. Delannay, *Inter. J. Plast.,* 21, 2005, p. 589.
29. M. Crumbach, G. Pomana, P. Wagner, G. Gottstein, *Proc. 1st Joint Int. Conf. on Recrystallization and Grain Growth,* G. Gottstein and D.A. Molodov, Ed., Springer (Berlin), 2001, p. 1053.
30. P. Van Houtte, S. Li, and O. Engler, *Aluminium,* 80, 2004, p. 702.
31. O. Engler, M. Crumbach, and S. Li, *Acta Mater.,* 53, 2005, p. 2241.
32. H.E. Vatne, T. Furu, R. Ørsund, and E. Nes, *Acta Mater.,* 44, 1996, p. 4463.
33. H.E. Vatne, K. Marthinsen, R. Ørsund, and E. Nes, *Metall. Trans.,* 27A, 1996, p. 4133.
34. J.A. Sæter, B. Forbord, H.E. Vatne, and E. Nes, *Proc. ICAA 6,* T. Sato et al., Ed., The Japan Inst. Light Metals (Tokyo), 1998, Vol. 1, p. 113.
35. H.E. Vatne, *Modeling of Metal Rolling Processes 3,* J.H. Beynon et al., Ed., IOM Communications (London), 1999, p. 247.
36. O. Engler and H.E. Vatne, *JOM,* 50, 1998 (No. 6), p. 23.
37. O. Engler, *Textures and Microstr.,* 32, 1999, p. 197.
38. O. Engler and S. Kalz, *Mater. Sci. Eng.,* A373, 2004, p. 350.
39. R.A. Lebensohn and C.N. Tomé, *Acta Metall. Mater.,* 41, 1993, p. 2611.
40. P. Van Houtte, G. Cauwenberg, and E. Aernoudt, *Mater. Sci. Eng.,* 95, 1987, p. 115.
41. R. Schouwenaars, P. Van Houtte, A. Van Bael, J. Winters, and K. Mols, *Textures and Microstr.,* 26–27, 1996, p. 553.
42. C. Maurice and J. Driver, *Acta Mater.,* 45, 1997, p. 4627.
43. M. Crumbach, G. Gottstein, L. Löchte, D. Piot, J. Driver, C.M. Allen, and J.F. Savoie, *Mater. Sci. Forum,* 396–402, 2002, p. 357.

Modeling of Microstructure Evolution during the Thermomechanical Processing of Titanium Alloys

S.L. Semiatin, Air Force Research Laboratory
D.U. Furrer, Rolls-Royce Corporation

TITANIUM AND TITANIUM ALLOYS are widely used for a variety of aerospace, chemical, marine, and other specialty applications because of their high strength, low density, good ductility, fatigue and corrosion resistance, as well as other properties. Depending on the application, a specific property (or combination of properties) can be obtained through microstructural modification.

Microstructure evolution and control in titanium alloys is heavily dependent on the allotropic transformation from a hexagonal close-packed crystal structure (denoted as alpha phase) found at low temperatures to a body-centered cubic crystal structure (denoted as beta phase) at high temperatures. For pure titanium, this transformation occurs at 882 °C. In many titanium alloys, the beta phase is partially stabilized at lower temperatures, and the equilibrium volume fractions of alpha and beta vary with temperature. The temperature at which a specific alloy becomes entirely beta (i.e., alpha + beta → beta) is called the beta transus. The conditions used for hot working and heating/heat treatment are often selected relative to the beta-transus temperature.

There are a number of important classes of titanium alloys whose designations reflect the relative volume fraction of beta retained at room temperature. These include beta and near-beta alloys and near-alpha and alpha/beta alloys (Ref 1, 2). Beta and near-beta alloys have moderate-to-large amounts of beta-stabilizing elements such as vanadium, molybdenum, tungsten, niobium, chromium, and iron. The beta transi of near-beta alloys are typically in the range of 700 to 850 °C (Ref 1). Near-alpha and alpha/beta alloys have large amounts of alpha-stabilizing elements such as aluminum; the beta transi of near-alpha and alpha/beta alloys are usually of the order of 950 to 1050 °C (Ref 2).

Processing of Titanium Alloys

The commercial processing of titanium alloys is most often based on an ingot-metallurgy approach in which large ingots are melted and refined using vacuum arc or cold hearth techniques, converted to mill products such as billet, plate, and sheet via a series of thermomechanical processing (TMP) steps, and fabricated into finished components via closed-die forging, superplastic forming, and so on (Ref 3, 4). The TMP of ingots of alpha/beta and near-alpha titanium alloys comprises a large amount of hot work in the beta-phase field, a small increment of hot work in the alpha+beta field, recrystallization heat treatment in the beta field, and secondary alpha+beta hot working (Ref 5). The initial steps of the process (beta and initial alpha+beta working and beta recrystallization) are employed to transform the ingot structure of relatively large columnar beta grains into a structure of finer (~0.5 to 1 mm) recrystallized equiaxed beta grains. Secondary alpha+beta hot working and post-hot-working alpha+beta annealing are then used to spheroidize the colony-alpha microstructure developed during cooldown of partially converted billet products following the beta recrystallization heat treatment. The TMP of ingots and billets of near-beta titanium alloys usually comprises hot working and recrystallization heat treatment in the beta-phase field.

Regardless of alloy class, the manufacture of finished components also typically includes a final heat treatment. These heat treatments are conducted in the alpha+beta or beta-phase field and may be used to control the volume fraction and morphology of the equiaxed alpha phase (alpha/beta alloys), to develop a fully transformed, colony-alpha microstructure (alpha/beta alloys), or to control the volume fraction of acicular alpha (near-beta alloys).

The modeling of microstructure evolution during the TMP of titanium alloys has seen significant advances during the last decade. Some of the principal efforts in this area have focused on the following:

- Recrystallization and grain growth phenomena for titanium alloys consisting of single-phase alpha or single-phase beta titanium
- Dynamic and static spheroidization of colony alpha in alpha/beta titanium alloys during TMP in the alpha+beta phase field
- Static and dynamic coarsening of equiaxed alpha or colony alpha in alpha/beta alloys during TMP in the alpha+beta phase field
- Evolution of primary, equiaxed alpha and secondary (lamellar) alpha in alpha/beta titanium alloys during cooling following heat treatment in the alpha+beta phase field
- Precipitation of acicular/lamellar alpha in alpha/beta and near-beta alloys following beta solution treatment

The following discussion focuses on the modeling of microstructure evolution during TMP in the two-phase (alpha+beta) field for alpha/beta and beta titanium alloys. A companion article in this Volume, "Modeling and Simulation of Texture Evolution during the Thermomechanical Processing of Titanium Alloys," treats recrystallization and grain-growth phenomena in the single-phase alpha or beta fields, inasmuch as such processes are strongly influenced by the dependence of nucleation and growth events on the local orientation of individual grains as well as the misorientation across grain boundaries. As such, microstructure and texture evolution are coupled and must be treated simultaneously during modeling. The evolution of deformation and transformation texture in alpha/beta titanium alloys during TMP in the alpha+beta field is also described in this latter article. The effect

Table 1 List of symbols

a	slip vector along close-packed direction in hexagonal close-packed (hcp) crystal structure
b	length of Burgers vector
C	composition (atomic fraction)
C_F	composition factor (Eq 8)
c	lattice parameter for hcp crystal structure
D	diffusivity
d_α	thickness of alpha platelet
d	grain/particle size
f_α	volume fraction of alpha
G	growth rate during phase transformation
G_s	shear modulus
J	mass flux
K	strength coefficient
k	rate constant in Johnson-Mehl-Avrami-Kolmogorov (JMAK) equation (Eq 23)
k_B	Boltzmann's constant
k_{LSW}	Lifshitz, Slyosov, and Wagner (LSW) coarsening-rate constant
k_{MLSW}	LSW coarsening-rate constant modified for volume-fraction effects
L_b	length of lamellar branch
N	nucleation rate during phase transformation
m	strain-rate sensitivity of the flow stress
m_g	slope of alpha-beta interface groove
n	strain-hardening exponent
n_a	JMAK (Avrami) exponent
p	grain-size exponent of the strain rate
Q	activation energy
R	radius of curvature, particle radius
r_α	radius of globular alpha particle
r	activity coefficient
R_g	gas constant
T	absolute temperature
t	time
V_M	molar volume
W	boundary-groove interface shape, $W(x, t)$
w	diameter of alpha platelet
X	fraction transformed
X_s	fraction spheroidized
x, y, z	Cartesian coordinates
α	alpha phase
β	beta phase
Γ	shear strain
$\dot{\Gamma}$	shear strain rate
γ	surface energy (grain boundary, interface phase, etc.)
δ	groove angle
$\varepsilon, \dot{\varepsilon}$	normal/principal strain, strain rate
ϕ	volume fraction
λ	particle-growth parameter (Eq 17 and 18)
σ	normal/principal stress
Ω	atomic volume
Ω_s	supersaturation (Eq 19b)
τ	shear stress
τ_{vd}	time for diffusional spheroidization via termination migration

of microstructure and local texture on cavitation in alpha/beta titanium alloys during alpha +beta hot working is discussed in the article "Modeling and Simulation of Cavitation during Hot Working" in this Volume. Symbols used in this article are listed in Table 1.

Dynamic and Static Spheroidization

The breakdown of the colony-alpha microstructure during alpha+beta TMP plays a pivotal role in obtaining a uniform, fine, equiaxed-alpha microstructure in forging bar stock and billet, plate for hog-out, and other alpha/beta titanium mill products (Fig. 1). In the literature, the mechanism of this microstructure change has been ascribed to various
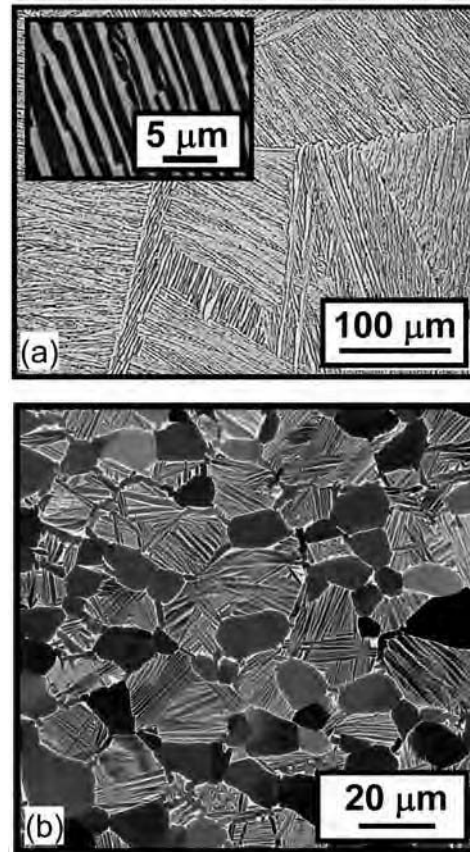
Fig. 1 Microstructures developed in Ti-6Al-4V during primary processing. (a) Optical and (inset) scanning electron microscopy (SEM) backscattered electron (BSE) micrographs of the colony-alpha microstructure. (b) SEM BSE micrograph of the fine, equiaxed-alpha microstructure. In SEM micrographs taken via BSE imaging, the dark phase is alpha, and the lighter phase is beta (or very fine transformed beta).

recrystallization-like and spheroidization phenomena. Most of the research in this area has focused on the workhorse alloy of the aerospace titanium industry, Ti-6Al-4V.

A number of early observations (Ref 6, 7) suggested that recrystallized alpha grains are formed *within* the alpha lamellae during hot working or post-hot-working annealing. Surface tension requirements do not permit 90° angles to exist between alpha/alpha boundaries (developed during recrystallization) and alpha-beta interfaces, however. Thus, a driving force is provided for the penetration of beta phase *along* the alpha/alpha boundaries and subsequent pinch-off/segmentation of alpha lamellae.

The research of Weiss et al. (Ref 8–10) expanded upon the work by Margolin and Cohen (Ref 6, 7) but did not invoke *recrystallized* alpha grains as the source of alpha-plate pinch-off. Rather, the main driving force was hypothesized to be the formation of intense shear bands within the alpha lamellae or the development of high-angle boundaries within alpha lamellae, spaced at periodic intervals along their length, due to dynamic/static recovery (Fig. 2). If the strain is large enough, shear bands may lead to fracture
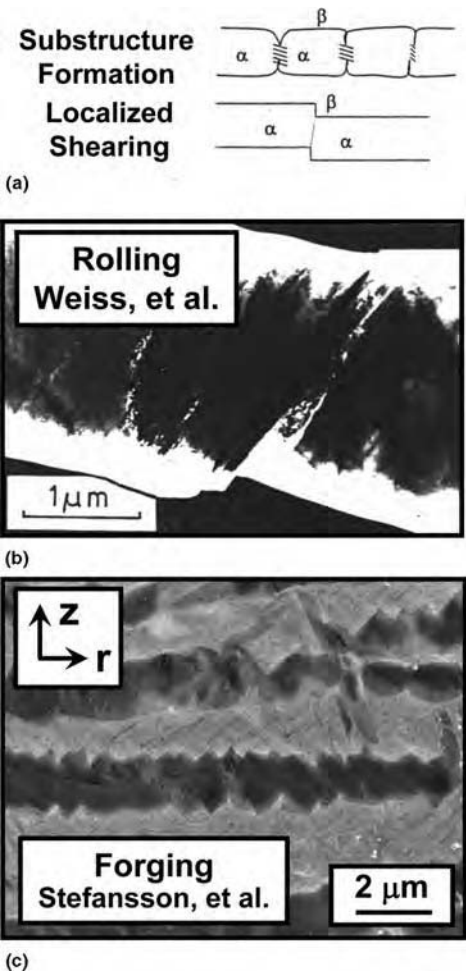
Fig. 2 Mechanisms of the spheroidization of alpha lamellae. (a) Spheroidization driven by the formation of subboundaries or shear bands within alpha lamellae. Source: Ref 9. (b, c) Observation of shear bands developed during hot deformation. Source: Ref 8, 11

and complete segmentation of an alpha lamella. In the absence of fracture, the alpha/alpha interfaces formed by shear bands or internal high-angle boundaries can give rise to surface-tension-driven penetration of the alpha plates by beta phase, as postulated by Margolin and Cohen. The rate of penetration by the beta phase depends on diffusion kinetics and the ratio of the interfacial energies of alpha/alpha boundaries and the alpha/beta interface.

Recent work by Salishchev, Furuhara, Stefansson, and their colleagues (Ref 12–15) has shed further light on some of the early observations. For example, Salishchev *et al.* (Ref 12, 13) and Furuhara *et al.* (Ref 14) used electron backscatter diffraction (EBSD) imaging techniques to show that extended dynamic recovery during hot working (so-called *continuous* dynamic recrystallization) can give rise to subboundaries within the alpha plates. The misorientation across these boundaries increases with increasing hot deformation; at large strains, high-angle boundaries and a structure resembling recrystallized equiaxed-alpha grains

are formed. High temperatures and low strain rates ($\sim 10^{-3}$ s^{-1}) favor such continuous recrystallization behavior. Furthermore, it has been found that the size of the resulting alpha *particles* is of the same order of magnitude as the original platelet thickness, thus suggesting that beta phase may indeed penetrate along the high-angle boundaries developed in the alpha phase and result in complete segmentation of the original lamellae.

Stefansson and Semiatin (Ref 15) investigated static spheroidization behavior following hot working. During relatively short times, spheroidization of alpha platelets was found to be controlled by the pinch-off mechanism suggested by Margolin, Weiss, and their colleagues. At longer times, spheroidization was controlled by mass transport from the ends to the center of residual alpha platelets, thereby resulting in alpha particles whose diameter was substantially greater than the initial alpha platelet thickness.

Dynamic and static spheroidization observations for Ti-6Al-4V have provided the basis for mechanism-based models based on flow localization/shear banding/shear fracture, the pinch-off of alpha plates by beta, and spheroidization controlled by termination migration. Each of these is discussed as follows.

Flow Localization Analysis. Observations by Weiss et al. (Ref 8) and Stefansson and Semiatin (Ref 11) have revealed that the spheroidization of colony alpha can occur by the localization of strain into intense shear bands within alpha lamellae (Fig. 2b, c) that eventually leads to fracture and segmentation. The local shearing also produces relative local rotations of adjacent elements of a given alpha platelet, thereby providing a means of randomizing the orientation of equiaxed alpha particles and preventing microtexture, that is, groups of alpha particles derived from a given prior colony of alpha plates, each of which has a similar orientation.

The development of shear bands appears to be driven by the difficulty of accommodating deformation uniformly within hexagonal close-packed (hcp) titanium due to its limited number of slip systems and the fact that the deformation resistance varies greatly among the possible slip systems. In particular, the critical resolved shear stress is relatively low for basal $<a>$ and prism $<a>$ systems but considerably higher for the pyramidal $<c+a>$ systems that must be activated to accommodate compression or extension along the c-axis of hcp crystal structures at hot working temperatures (Ref 16). Furthermore, the presence of beta phase between adjacent lamellae introduces a further anisotropy in slip behavior, inasmuch as only one of the three $<a>$ directions in the alpha phase is co-linear with a $<111>$ slip direction in the beta phase (Ref 10). Hence, one of alpha-phase $<a>$ slip directions tends to be considerably softer than the other two at both hot and cold working temperatures (Ref 17, 18).

A simple analysis can be used to estimate the effect of nonuniformity in thickness along an alpha plate on the shear localization rate.

Although each alpha plate is surrounded by a layer of considerably softer beta, useful qualitative insight can be obtained by focusing solely on the deformation of alpha. The simplest case comprises simple shear along one slip system in a single alpha lamella. It is assumed that strain localization arises from regions in which the alpha plate is locally thinner than the rest. By denoting the thickness of the nominally uniform and thinner (defect) regions as t_u and t_d, respectively, equilibrium considerations (Ref 19) yield the following expression for a material whose constitutive behavior ($\tau = \tau(\Gamma, \dot{\Gamma})$, in which τ denotes shear stress as a function of shear strain, Γ, and shear strain rate, $\dot{\Gamma}$) is characterized by a power-law strain-hardening and strain-rate hardening:

$$K\Gamma_u^n \dot{\Gamma}_u^m t_u = K\Gamma_d^n \dot{\Gamma}_d^m t_d \qquad \text{(Eq 1)}$$

Here, K denotes the strength coefficient, n and m are the strain-hardening exponent and strain-rate sensitivity, and the subscripts "u" and "d" refer to the uniform and defect regions, respectively. Rewriting $\dot{\Gamma}$ as $d\Gamma/dt$, rearrangement of Eq 1 provides an expression relating the increments of shear strain in the uniform and defect regions, that is, $d\Gamma_u$ and $d\Gamma_d$, respectively:

$$d\Gamma_d = d\Gamma_u (t_u \Gamma_u^n / t_d \Gamma_d^n)^{1/m} \qquad \text{(Eq 2)}$$

Equation 2 can be solved numerically to relate Γ_u and Γ_d as a function of n and m and the magnitude of the thickness defect. Such solutions reveal that large-thickness nonuniformity (small values of t_d/t_u), low values of m, and low (or negative) values of n increase the kinetics of flow localization. An example calculation assuming typical values for the thickness inhomogeneity and material coefficients (i.e., a 10% thickness defect ($t_d/t_u = 0.90$), $m = 0.15$, and $n = 0$) is shown in Fig. 3. The results indicate a rapid, but not catastrophic, rate of flow localization; that is, $d\Gamma_d/d\Gamma_u$ is relatively large but not infinite. This type of trend suggests that
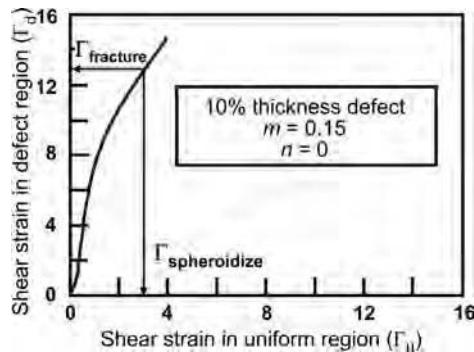
localization is likely terminated once Γ_d reaches a critical fracture strain.

Calculations such as that described in the previous paragraph may be useful for interpreting dynamic spheroidization behavior at high strain rates (of the order of 0.1 s^{-1} or greater), in which flow softening (negative n) is enhanced by both microstructural influences (Ref 16) and deformation heating. In these cases, spheroidization kinetics have been found to depend very weakly on temperature for a given preform microstructure (Ref 20, 21). This behavior may be rationalized qualitatively on the basis of the compensatory effects of increases in both the thickness inhomogeneity (associated with decreasing average platelet thickness) and the strain-rate sensitivity with increasing temperature on localization kinetics. Further research documenting the kinetics of the localization process, including the effect of local colony orientation and the orientation of neighboring colonies on spheroidization (Ref 22), is warranted. Detailed EBSD analysis of the activated slip systems that are associated with shear localization would also be beneficial. In this regard, it appears likely that systems whose slip planes form a large angle with each other (e.g., prism and basal planes) must both undergo localization in order to segment alpha lamellae into individual equiaxed particles.

Platelet Pinch-Off via Boundary Grooving. The penetration of beta phase along the boundaries formed within alpha platelets (or vice versa) can be analyzed using the classic approach for grain-boundary grooving developed originally by Mullins (Ref 23, 24). In the original work, grooving was assumed to be controlled by volume diffusion, surface diffusion, or evaporation/condensation (as for a boundary meeting a free surface). The analysis is similar for all three cases. Thus, attention is focused here on the situation of most interest: the volume-diffusion-controlled grooving/penetration of alpha platelets by beta.

Grooving is initiated by atomic-scale processes near the region of intersection of the alpha-alpha and alpha-beta boundaries. For the semi-infinite, two-dimensional case shown in Fig. 4, these processes establish a groove angle, δ, that satisfies the equilibrium equation for the surface tensions associated with the two different interfaces (Ref 25):

$$2\gamma_{\alpha\beta} \sin \delta = \gamma_{\alpha\alpha} \qquad \text{(Eq 3)}$$

Fig. 3 Predictions of a flow localization calculation for shear banding/shear fracture. The arrows indicate the strain in the nominally uniform region at which a hypothetical fracture strain is reached in the defect region.
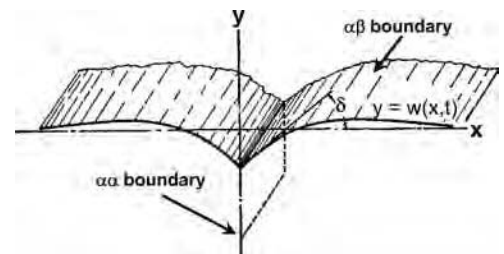
Fig. 4 Schematic illustration of the geometry during boundary grooving. Source: Ref 23, 24

in which $\gamma_{\alpha\beta}$ and $\gamma_{\alpha\alpha}$ denote the surface energy of the alpha-beta interface and the alpha-alpha boundary within the alpha lamella. The development of the initial groove generates sharp edges of high energy adjacent to it. These asperities flatten rapidly, leading to the diffusional development of a broad convex shape for the alpha-beta interface. The continued transport of mass from the curved alpha-beta interface to distant regions of lower curvature upsets the equilibrium associated with the groove angle, δ, and thereby drives the groove deeper. For a platelet of finite thickness, the mass transport continues until the groove has completely segmented the plate.

The Mullins grooving analysis (Ref 23, 24) determines the solute concentration field that satisfies the steady-state diffusion (Laplace's) equation subject to appropriate boundary and initial conditions. For the case in which the platelet and matrix phases are both terminal solid solutions, the latter relations comprise the classical Gibbs-Thomson equation, the equilibrium equation (Eq 3 reformatted in terms of the groove slope), and the initial condition (an initially flat interface). Hence, the complete set of equations is as follows:

$$\nabla^2 C_\beta(x, y, t) = 0 \qquad \text{(Eq 4a)}$$

$$\frac{C_\beta(R)}{C_\beta(R = \infty)} = 1 + \frac{\gamma_{\alpha\beta}\Omega}{k_B T} \frac{1}{R(x,t)} \qquad \text{(Eq 4b)}$$

$$\frac{dW(0,t)}{dx} = m_g = \tan\delta \qquad \text{(Eq 4c)}$$

$$W(x, t = 0) = 0 \qquad \text{(Eq 4d)}$$

In these expressions, $C_\beta(R)$ and $C_\beta(R = \infty)$ denote the concentration of the rate-limiting solute in the matrix adjacent to a curved surface of radius R and a flat surface, respectively; Ω is the atomic volume; k_B is the Boltzmann's constant; T is the absolute temperature; $W(x, t)$ is the interface shape as a function of the x-coordinate and time, t (Fig. 4); and m_g is the slope of the groove at $x = 0$.

The boundary-grooving analysis of Mullins reveals that the interface profile $W(x, t)$ retains a fixed shape in normalized coordinates and that the groove deepens as the cube root of time. The normalization constant is $(At)^{1/3}$, in which A is defined as:

$$A = \frac{C_\beta \gamma_{\alpha\beta} \Omega^2 D_\beta}{k_B T} = \frac{C_\beta \gamma_{\alpha\beta} V_M D_\beta}{R_g T} \qquad \text{(Eq 5)}$$

Here, D_β is the diffusivity of the solute through the matrix (beta) phase, V_M is the molar volume of the matrix, and R_g is the gas constant. The concentration C_β is defined in terms of atoms per unit volume or as an atomic fraction in the first and second expressions, respectively, for A. The time for the groove to move a distance d (relative to the plane $y = 0$, Fig. 4) is then given by (Ref 24):

$$d = 0.86 \; m_g (At)^{1/3} \qquad \text{(Eq 6)}$$

For an alpha platelet of thickness d_α, the time to complete the boundary-grooving process is that corresponding to $d = d_\alpha/2$, inasmuch as beta-phase penetration would proceed inward from both alpha-beta interfaces. The time for pinch-off (t_p) is therefore:

$$t_p = \frac{0.2 d_\alpha^3}{A \; m_g^3} \qquad \text{(Eq 7)}$$

Strictly speaking, the expressions for A (Eq 5) and hence t_p (Eq 7) are valid only when both the matrix phase and the second phase are terminal solid solutions. For Ti-6Al-4V, neither the alpha nor the beta phase is a terminal solid solution. In this case, two corrections are needed; one is required for the Gibbs-Thomson equation (Eq 4b) and the other to account for the actual amount of solute that must be transported to deepen the groove. The two corrections are (1) $\{(1 - C_\beta) /[(C_\alpha - C_\beta)(1 + \partial \ln r / \partial \ln C_\beta)]\}$, in which the second parenthetical term in the denominator is a thermodynamic factor ($r \equiv$ activity coefficient of the solute in the beta phase), and (2) $(C_\alpha - C_\beta)$ (Ref 26). In both terms, C_α denotes the concentration of the rate-limiting solute in the alpha phase. The term C_β in Eq 4(b) and 5 is thus replaced by the quotient of these two terms (Ref 26):

$$\text{Composition factor, } C_F = \frac{C_\beta(1 - C_\beta)}{(C_\alpha - C_\beta)^2 [1 + \partial \ln r / \partial \ln C_\beta]} \qquad \text{(Eq 8)}$$

Corrections such as that given in Eq 8 are particularly important for alloys such as Ti-6Al-4V in which there is only a small composition difference, $C_\beta - C_\alpha$.

The applicability of the Mullins analysis for the pinch-off of alpha platelets can be assessed using the static heat treatment observations of Stefansson and Semiatin (Ref 15). In this work, it was found that boundary splitting was completed in ~14 h at 900 °C and ~1 h at 955 °C. At both temperatures, the slope m_g (tan δ) was measured as ~0.35. Input data at the lower and higher temperatures comprise the values of the composition factor (30.6, 61.3) (Ref 27), diffusivity of the rate-limiting solute (vanadium) (0.025, 0.05 $\mu m^2/s$) (Ref 28), and the alpha-platelet thicknesses (~1, 2 μm) (Ref 15). The molar volume (10, 440 mm^3) and alpha-beta surface energy (0.4 J/m^2) are taken to be the same at both temperatures (Ref 27). Applying Eq 7, the predicted time (1 h) is in excellent agreement with the observation for 955 °C, but the prediction for 900 °C (32 h) is approximately twice that observed. The difference between measured and predicted times for heat treatment at 900 °C may be partially due to dislocation substructure in the beta matrix retained from prior hot working and a

concomitant enhancement in kinetics due to pipe diffusion.

Pipe diffusion may also play a role in the kinetics of the pinch-off of alpha platelets during deformation. In such cases, the required substructure within the alpha lamellae is developed at low-to-moderate strains, and grooving takes place at larger strains. If concurrent hot working enhances the diffusivity of solutes through the beta matrix by a factor of the order of 50, the times required for splitting alpha lamellae would be of the order of 1 min at 955 °C or 20 min at 900 °C. Thus, the breakdown of the colony-alpha microstructure during deformation at low strain rates ($\leq 10^{-2}$ s^{-1}) may indeed involve an element of boundary grooving.

Static Spheroidization via Termination Migration. The complete breakdown of the colony-alpha microstructure via shear localization and/or boundary grooving usually requires strains much in excess of those that can be imposed in conventional production practices. Moreover, the dislocation substructure that drives boundary grooving during static heat treatment tends to be eliminated in relatively short times. At the completion of the pinch-off process, sections of alpha lamellae having a pancakelike shape remain. Subsequent (longer-time) completion of spheroidization occurs via termination migration, that is, the transfer of solute from the periphery to the flat (plan) surfaces of each (assumed noninteracting) alpha lamella, resulting in equiaxed alpha particles whose diameters are substantially greater than the original platelet thickness (Ref 15).

The time to complete static spheroidization via termination migration can be estimated using an approach originally developed by Courtney and Malzahn Kampe (Ref 29) and later extended by Semiatin et al. (Ref 30) for the specific geometry involved here. The analysis comprises a relatively simple application of the one-dimensional form of Fick's law to solute diffusion at the beginning and intermediate points of the process:

$$\text{Flux, } J = -D_\beta \frac{dC_\beta}{dx} \qquad \text{(Eq 9)}$$

in which C_β and D_β are taken to be the concentration and diffusivity of the rate-limiting solute.

As for the boundary-grooving analysis, solute concentration gradients are described with the aid of the Gibbs-Thompson relation, using the two principal radii of curvature of the alpha platelet. For example, in the beginning of the diffusional process, during which the plan surfaces of remnant lamellae are still relatively flat, the concentration difference between the edges and plan surfaces is the following:

$$[C_\beta(R = \infty)](\gamma_{\alpha\beta}\Omega/k_B T)\left(\frac{2}{d_\alpha} + \frac{1}{\frac{w}{2} + \frac{d_\alpha}{4}}\right) \qquad \text{(Eq 10)}$$

Here, d_α and w are the platelet thickness and diameter, respectively, and the other symbols

(a)

(b)

20 μm

(c)

20 μm
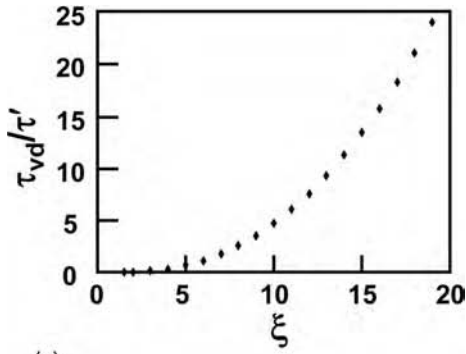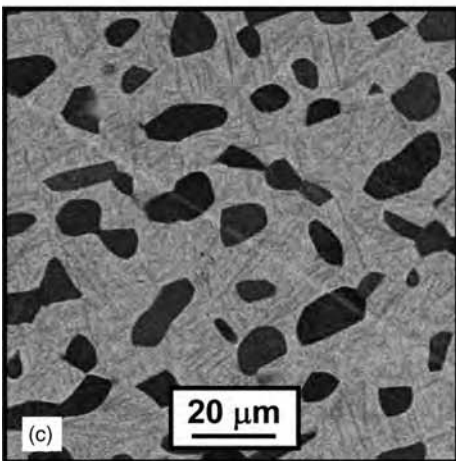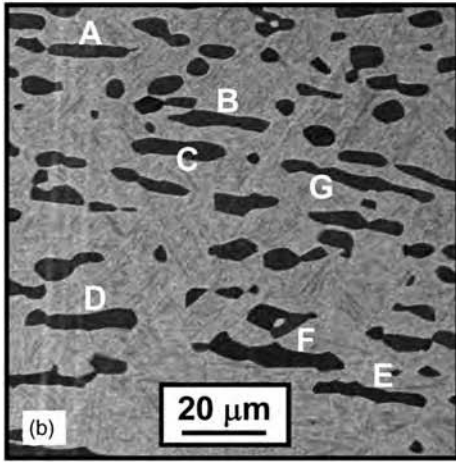
Fig. 5 Static spheroidization via termination migration. (a) Plot of τ_{vd}/τ' as a function of ξ. (b, c) SEM backscattered micrographs of the microstructure developed in Ti-6Al-4V samples deformed at 955 °C to an effective strain of 1.1 and water quenched after holding at temperature for (b) 1 h or (c) 14 h. Source: Ref 30. The letters adjacent to the unspheroidized alpha platelets in (b) correspond to the data in Table 2. (Alpha, darker phase; beta, lighter phase)

are as defined for Eq 4(b). Geometrical considerations provide estimates of the diffusion distance (and hence concentration gradient) and the area through which the solute flux passes.

The final result for the spheroidization time, τ_{vd}, incorporating the composition-factor

Table 2 Geometry of remnant alpha lamellae and model predictions of globularization time at 955 °C

Lamella identification	Lamella diameter (w), μm	Lamella thickness (d_α), μm	η	τ_{vd}/τ'	τ_{vd}, h
A	16.4	2.7	6.5	1.39	6.1
B	17.7	3.2	6.1	1.13	7.8
C	18.2	3.6	5.5	0.84	8.6
D	21.8	3.6	6.5	1.39	14.4
E	24.1	2.3	11.1	6.16	15.6
F	27.3	3.6	8.0	2.52	26.1
G	33.2	2.3	15.1	13.6	34.4

Source: Ref 30

correction (Eq 8) as shown previously is the following:

$$\frac{\tau_{vd}}{\tau'} = \frac{\xi^3 - \left[0.328\xi^{7/3}\left(1 + \sqrt{1 - 0.763\xi^{-4/3}}\right)^2\right]}{4\left[\frac{2(1+\xi)}{3(0.5 - 0.572\xi^{-1/3})} + \frac{0.5\xi^{1/3} + 0.665\xi^{2/3}}{3(0.143 + 0.934\xi^{-1/3})}\right]}$$

(Eq 11a)

in which:

$$\xi \equiv (w/d_\alpha) + 0.5 \qquad \text{(Eq 11b)}$$

$$\tau' \equiv d_\alpha^3 R_g T / D_\beta C_F \gamma_{\alpha\beta} V_M \qquad \text{(Eq 11c)}$$

The dependence τ_{vd}/τ' on ξ is shown in Fig. 5(a).

The validity of Eq 11 to quantify the spheroidization behavior and persistence of remnant lamellae was established by Semiatin et al. (Ref 30) using observations for Ti-6Al-4V heat treated at 955 °C (Fig. 5b, c). At this temperature, the shapes of the lamellae that existed after the completion of boundary grooving (time ≈ 1 h) approximated the idealized pancake geometry assumed in the diffusion model. For Ti-6Al-4V, the diffusion of vanadium through the beta matrix is rate-limiting in comparison to the diffusion of aluminum.

Model predictions of τ_{vd}/τ' and hence τ_{vd} for spheroidization of the remnant lamellae are given in Table 2. The predicted values for the spheroidization time for the pancake-shape geometry ranged from approximately 5 to 15 h for 5 of the 7 platelets; the other two lamellae were predicted to require times of the order of 30 h. These values of spheroidization time compare well with the measured time of 12 to 14 hours to achieve approximately 90 vol% of spheroidized microstructure, as defined by alpha particles with an aspect ratio of less than 2:1 (Ref 15, 30). Thus, it is not surprising that a small volume fraction of modest-aspect-ratio, partially spheroidized lamellae would remain after 14 h (Fig. 5c).

Phenomenological Models. Because of the complexity of the various mechanisms that control the dynamic and static spheroidization of the colony-alpha microstructure, several attempts have been made to fit phenomenological models to measured data for the regime of processing parameters most common in
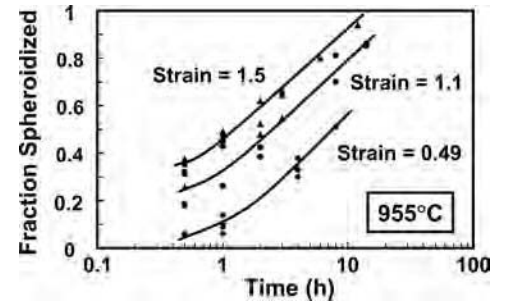
Fig. 6 Statically spheroidized fraction of colony-alpha microstructure as a function of time during heat treatment at 955 °C following predeformation to various strains at 955 °C. Source: Ref 31

production practice (Ref 31–34). For example, the data of Stefansson et al. (Ref 31) for the overall kinetics of static spheroidization following deformation at 900 or 955 °C and a strain rate of 0.1 s^{-1} (Fig. 6) have served as the basis for a phenomenological fit for the fraction spheroidized, X_s, as a function of temperature, T, and prestrain, ε (Ref 32–34):

$$X_s = (a\varepsilon + f) \log t + b\varepsilon + cT + d \qquad \text{(Eq 12)}$$

in which a, b, c, d, and f are constants. Recent additional measurements by Brooks et al. (Ref 34) have shown that static spheroidization kinetics are also dependent on the strain rate during the predeformation.

Static and Dynamic Coarsening

The term *coarsening* refers to the increase in the average size of a dispersion of second-phase particles, lamellae, and so on during TMP. Also known as Ostwald ripening, coarsening in alpha/beta titanium alloys generally refers to the increase in the average size of equiaxed-alpha particles in the beta matrix or the increase in the average thickness of the lamellae within a colony-alpha microstructure. Such phenomena may occur during static preheating/heat treatment or during deformation in the alpha+beta phase field. Several models have been developed and applied to quantify the static and dynamic coarsening of equiaxed alpha and the static coarsening of colony alpha in alpha/beta titanium alloys.

Static Coarsening of Equiaxed Alpha. Static coarsening of a distribution of equiaxed particles in a matrix phase comprises the shrinkage (and elimination) of the smaller particles and the growth of the larger particles. The process is driven by the reduction of the total surface energy of the system. The rate of coarsening is frequently limited by diffusion of solutes through the matrix or by a reaction at matrix-particle interfaces.

Most approaches for describing coarsening are based on the classic work of Lifshitz, Slyosov, and Wagner (LSW) (Ref 26, 35, 36). This early work assumed an infinitesimal

volume fraction of particles and matrix/particle compositions that are both terminal solid solutions. For coarsening controlled by bulk diffusion of a solute through a matrix denoted as beta, the classical LSW theory predicts the following for the variation of the *average* radius of the particles ($\bar{r}_\alpha$) as a function of time, t:

$$\bar{r}_\alpha^3 - \bar{r}_{\alpha o}^3 = \left[\frac{8D_\beta \gamma_{\alpha\beta} C_\beta V_M}{9R_g T}\right](t - t_o) = k_{LSW}(t - t_o)$$

(Eq 13a)

in which D_β denotes the diffusivity in the beta matrix of the rate-limiting solute, $\gamma_{\alpha\beta}$ is the energy associated with particle-matrix (alpha-beta) interfaces (in J/m^2), C_β is the equilibrium concentration in the beta matrix of the rate-limiting solute (expressed as an atomic fraction), V_M is the molar volume of the particles, R_g is the gas constant, T is the absolute temperature, and the subscript "o" signifies quantities at an initial (reference) time. When coarsening is limited by a matrix-particle interface reaction, the coarsening exponent (exponent of $\bar{r}_\alpha$ in Eq 13a) takes a value of 2 instead of 3. The bracketed term in Eq 13(a) is known as the LSW rate constant, k_{LSW}:

$$k_{LSW} \equiv \frac{8D_\beta \gamma_{\alpha\beta} C_\beta V_M}{9R_g T}$$

(Eq 13b)

The form of the rate constant, k_{MLSW}, for a finite volume fraction of particles in a system comprising two concentrated alloy phases is the following (Ref 37, 38):

$$k_{MLSW} = \frac{8f(\phi)D_\beta \gamma_{\alpha\beta} C_\beta (1 - C_\beta)V_M}{9RT(C_\alpha - C_\beta)^2[1 + \partial \ln r / \partial \ln C_\beta]}$$

(Eq 14)

in which $f(\phi)$ describes the functional dependence of the rate constant on volume fraction ϕ, C_α is the equilibrium concentration in the alpha particle of the rate-limiting solute (expressed as an atomic fraction), and r is the activity coefficient of the rate-limiting solute in the beta matrix. A number of analyses have been conducted to estimate $f(\phi)$ and hence the interaction of the solute sources/sinks comprising a field of second-phase particles and its effect on overall coarsening behavior (Ref 39). As summarized by Doherty (Ref 38), the terms in Eq 14 that are related to the volume-fraction function, $f(\phi)$, and the phase concentrations can readily increase the magnitude of predictions of the rate constant by 1 or 2 orders of magnitude.

The applicability of Eq 13 and 14 to describe the static coarsening of Ti-6Al-4V with an equiaxed-alpha microstructure was established by Semiatin, Sargent, and their coworkers for heat treatment temperatures between 775 and 955 °C (Ref 27, 40); typical microstructure observations are shown in Fig. 7. The coarsening exponent was found to be 3 (Fig. 8a), indicating that coarsening was indeed diffusion controlled. The ratio of the measured rate constant and the calculated value of $k_{MLSW}/f(\phi)$
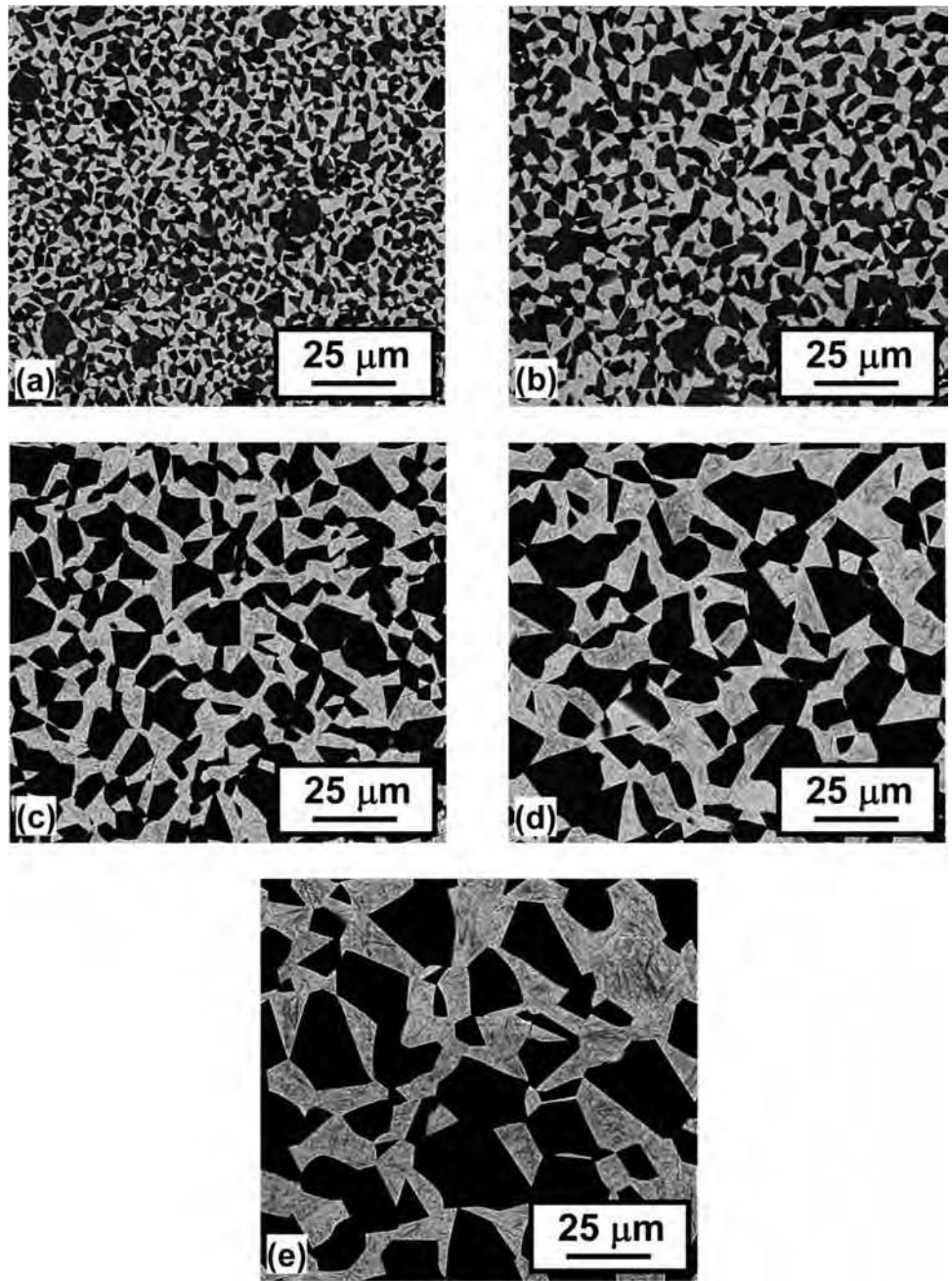
Fig. 7 SEM backscattered micrographs taken at the same magnification, illustrating static coarsening of Ti-6Al-4V (with an equiaxed-alpha microstructure) that was heat treated at 900 °C and water quenched after times of (a) 1, (b) 4, (c) 16, (d) 48, or (e) 144 h. Source: Ref 27. (Alpha, darker phase; beta, lighter phase)

for each temperature (and hence corresponding volume fraction) was determined to estimate $f(\phi)$. A comparison of the values of $f(\phi)$ so determined to various model predictions (Fig. 8b) indicated that the Voorhees and Glicksman model (Ref 39) was most appropriate for quantifying the effect of volume fraction on coarsening kinetics for Ti-6Al-4V.

Dynamic Coarsening of Equiaxed Alpha. Coarsening of equiaxed alpha in alpha/beta titanium alloys may also occur during hot deformation (Fig. 9). A number of investigations (Ref 40–42) have shown that the kinetics of coarsening

in such situations are enhanced by approximately an order of magnitude relative to those for static coarsening. The effect is most noticeable at strain rates that characterize superplastic (or near-superplastic) flow, that is, ~10^{-4} to 10^{-3} s^{-1}. Under these conditions, the time of deformation is typically of the order of 10 to 100 min.

As for static behavior, dynamic coarsening kinetics have been found to be bulk-diffusion controlled with a coarsening exponent equal to 3 (Ref 42). With respect to the coarsening constant, k_{MLSW}, an examination of Eq 14 suggests that the principal factor that would be
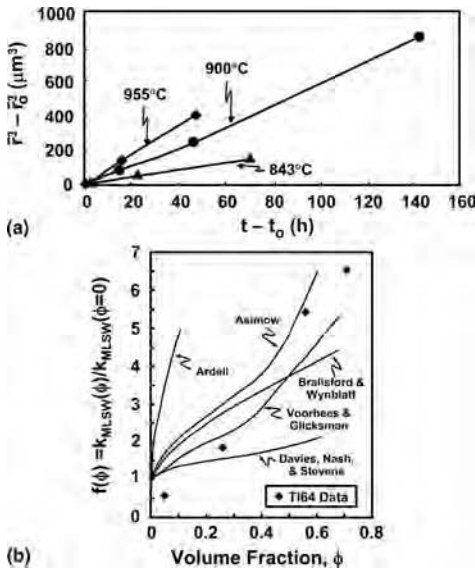
Fig. 8 Static coarsening behavior of Ti-6Al-4V with an equiaxed-alpha microstructure. Source: Ref 27. (a) Coarsening kinetics in terms of average alpha-particle size as a function of time. (b) Comparison of the ratio of the measured coarsening rate at different temperatures (and hence volume fractions, φ) and the calculated value of $k_{MLSW}/f(\phi)$ to various model predictions (Ref 39) to determine $f(\phi)$.

Fig. 9 SEM backscattered micrographs taken at the same magnification, illustrating dynamic coarsening of Ti-6Al-4V samples with an equiaxed-alpha microstructure deformed at 900 °C, 10^{-4} s^{-1} to true strains of (a) 0, (b) 0.5, and (c) 1.4, followed by water quenching. Source: Ref 42. The compression axis is vertical for each micrograph. (Alpha, darker phase; beta, lighter phase)

affected by concurrent deformation is the diffusivity D_β.

Defects such as dislocations and a nonequilibrium concentration of vacancies generated during deformation can contribute significantly to diffusion. Insight into the effect of such pipe diffusion on kinetics can be obtained from work on microalloyed steels (Ref 43). For these materials, precipitation kinetics were found to increase by approximately 2 orders of magnitude during concurrent deformation at conventional strain rates ($\sim$0.1 to 1 s^{-1}); at these strain rates, the micromechanism of plastic flow is principally dislocation glide/climb. (Precipitation rates were found to increase by 1 order of magnitude during static heat treatment following a prestrain of 0.05.) By contrast, superplastic flow occurs in a regime in which deformation is accomplished largely by grain-boundary sliding and secondarily by dislocation-glide processes. Thus, the enhancement of diffusion during superplastic flow would be expected to be less than that during higher-strain-rate conventional forming. Hence, the observed increase in coarsening kinetics during superplastic flow by 1, rather than 2, orders of magnitude appears to be reasonable (Ref 40, 42).

Dynamic coarsening in alpha/beta titanium alloys can have a significant effect on stress-strain behavior at superplastic strain rates (Ref 40–42). Deformation occurs primarily by sliding at the alpha-beta interfaces, and the accommodation of resulting stress concentrations developed at triple points, grain edges, and so on becomes more difficult as the alpha-particle size increases. Hence, dynamic coarsening typically gives rise to flow hardening in stress-strain curves (Fig. 10a).

An evolving microstructure should be taken into account when formulating constitutive models to describe plastic flow. One of the most widely used constitutive approaches to relate flow stress, σ, and strain rate, $\dot{\varepsilon}$, is the single-state-variable model of Bird, Mukherjee, and Dorn (Ref 44), typically applied for single-phase alloys:

$$\dot{\varepsilon} = \left(\frac{A_{BMD} D\, G_s \mathrm{b}}{k_B T} \right) \left(\frac{\sigma}{G_s} \right)^{1/m} \left(\frac{\mathrm{b}}{d} \right)^p \quad \text{(Eq 15)}$$

In Eq 15, A_{BDM} is a constant, D is a diffusivity, k_B is Boltzmann's constant, T is absolute temperature, G_s is the shear modulus, b is the length of the Burgers vector, m is the rate sensitivity ($1/m$ is the stress exponent of the strain rate), d is the grain size, and p is the grain-size exponent of the strain rate. For superplastic flow characterized by grain-boundary sliding accommodated by glide/climb of dislocations, $m \sim 0.5$ and $p \sim 2$. For grain-boundary sliding accommodated by diffusional flow, $m \sim 1$ and $p \sim 2$ or 3, depending on whether bulk (lattice) or boundary diffusion predominates.

It has been shown that superplastic flow in two-phase alpha/beta titanium alloys can also be modeled using Eq 15, provided the instantaneous alpha-particle size (accounting for dynamic coarsening) is used for d, and the values of G_s and b pertain to the softer beta phase (Ref 40). In this respect, the alpha particles act like the core and the beta matrix like the mantle in the classical Gifkins model of superplastic flow (Ref 45). To determine the appropriate diffusivity to be used in Eq 15, Sargent et al. (Ref 40) inverted Eq 15 and plotted log $A_{BMD}D$ as a function of $1/T$ using the measured values of stress, alpha-particle size, m, and p corresponding to the imposed strain rates in the superplastic regime (Fig. 10b). The resulting plot was linear, yielding an activation energy $Q = 160$ kJ/mol. This value of Q is comparable to that for the diffusion of vanadium (in annealed beta titanium), the solute which controls static (and dynamic) coarsening (Fig. 10b). The comparison of the lines for $A_{BMD}D$ versus $1/T$ for superplastic flow and D_β^V suggests that A_{BMD} is of the order of 10 when D_{BMD} is taken to be equivalent to the diffusivity of vanadium in annealed beta titanium. By analogy with the relation between the rates of dynamic and static coarsening, A_{BDM} thus quantifies the enhancement of diffusion due to concurrent superplastic deformation.

Static Coarsening of Lamellar Alpha. The static coarsening/thickening of alpha lamellae in colony- or Widmanstätten- (basketweave-) alpha microstructures represents an important consideration with regard to the control of the final alpha-particle size in mill products of alpha/beta titanium alloys. This is because the sizes of alpha particles are generally no smaller than the thickness of the alpha lamellae from which they originate. Furthermore, the ease of dynamic and/or static spheroidization is greatly dependent on the thickness of the lamellae, regardless of whether the mechanism is boundary grooving or shear localization. Hence, an understanding of the kinetics of the coarsening of lamellar microstructures during static heat treatment is important.

Thickening of alpha lamellae during static heating occurs by mass transport, which results in a reduction in overall surface area and thus surface energy. The transport may occur between a lamella and so-called "branches" that are attached to it as well as between adjacent lamellae due to a classical coarsening-type process (Fig. 11). The former mechanism, that is, branch elimination, is analogous to the phenomenon of fault migration treated previously in the literature (Ref 46, 47) and is relatively easy to quantify, inasmuch as the entity that dissolves (a branch or lamellar fault) forms a well-defined geometric relation with the lamella or lamellae onto which its mass is transferred (Fig. 11a, b). By contrast, classical coarsening of an aggregate of lamellae (Fig. 11c) is a much more difficult problem because of the irregular plan-view shapes of typical alpha platelets, the complex spatial arrangement of the platelets within a colony relative to each other, and the irregular shape of the colonies themselves (Ref 48). Due to these complexities, a method to model the latter coarsening phenomenon has not yet been developed.
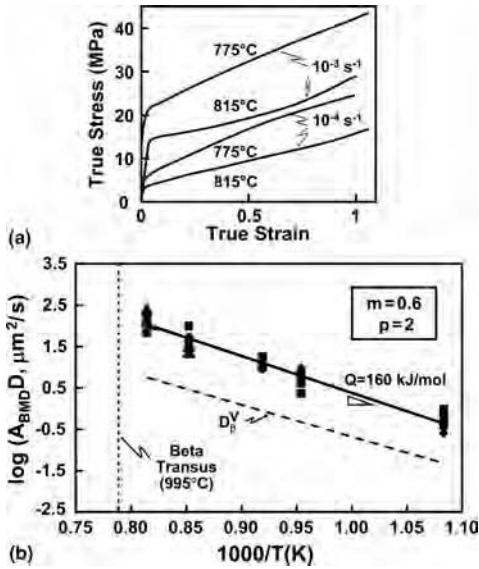
Fig. 10 Effect of dynamic coarsening on plastic flow of Ti-6Al-4V with an equiaxed-alpha microstructure. Source: Ref 40. (a) Selected flow curves. (b) Constitutive analysis to determine the appropriate activation energy and diffusivity to describe superplastic flow

LIVE GRAPH
Click here to view

LIVE GRAPH
Click here to view

An approximate Fick's law treatment of the coarsening of colony alpha in alpha/beta titanium alloys via branch elimination has been performed recently (Ref 49). The analysis is similar to that for static spheroidization via termination migration, summarized earlier in this article. In brief, the concentration gradient and associated flux are determined using (1) the Gibbs-Thompson equation, assuming that the branch has a radius of curvature equal to one-half its thickness, and the lamella onto which its mass is transferred is flat (Fig. 11a, b); and (2) a composition factor (Eq 8) to correct for the fact that the alpha and beta phases are not terminal solid solutions. The branch recession rate is given by the following expression:

$$\frac{dL_b}{dt} = \frac{\pi D_\beta C_F V_M \gamma_{\alpha\beta}}{R_g T \ T_b(Y_b + T_b/2)} \qquad \text{(Eq 16)}$$

In this equation, the geometric factors are defined in Fig. 11(b); and D_β, C_F, V_M, $\gamma_{\alpha\beta}$, R_g, and T denote the diffusivity of the rate-limiting solute, the composition factor (Eq 8), the molar volume of the alpha phase, the alpha-beta interface energy, the gas constant, and the absolute temperature, respectively.

Predictions of the branch recession rate from Eq 16 show reasonably good agreement with observations after correcting for the stereological (section-plane) effect (Ref 49). In particular, for a heat treatment at 955 °C, the recession rate was predicted to be 7.5 μm/h for branches 1.66 μm thick lying 1.86 μm from the lamella to which they are attached. The measured recession rate was ~10 μm/h.
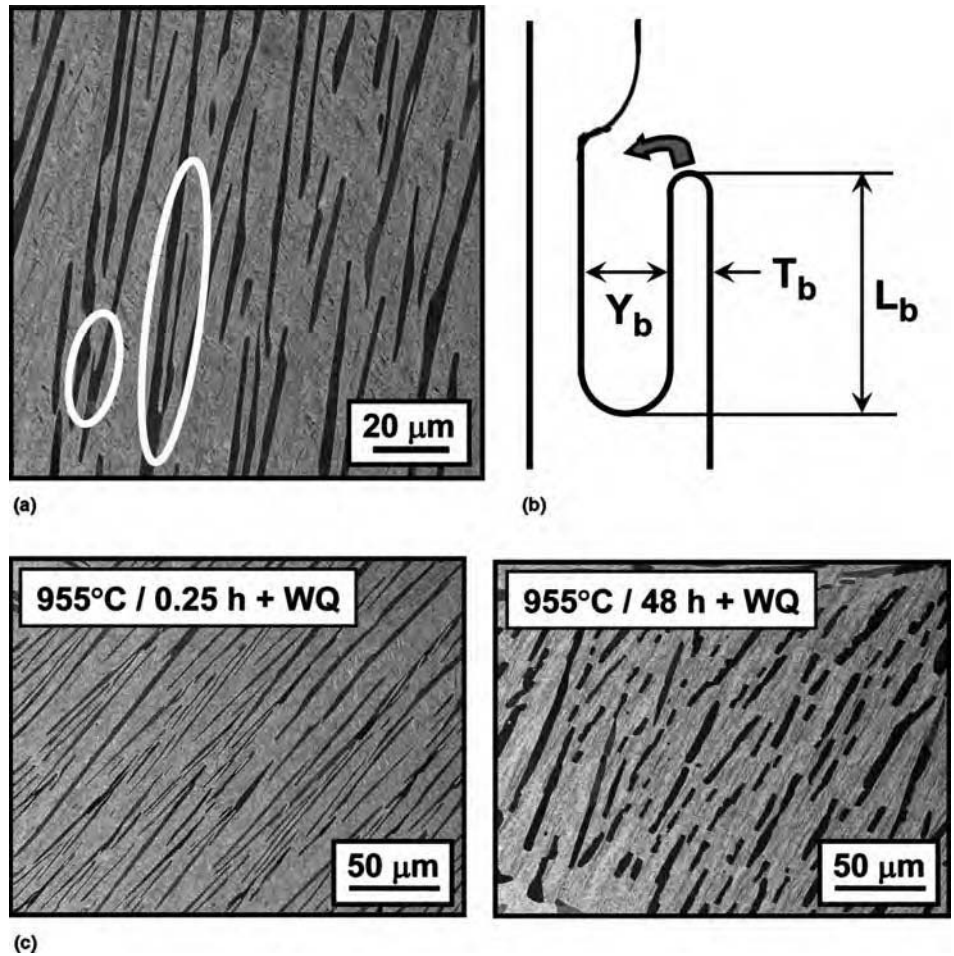
Fig. 11 Static coarsening of a colony-alpha microstructure in Ti-6Al-4V. (a) SEM backscattered micrograph illustrating lamellar branching observed at 955 °C. (b) Idealized geometry used to describe the microstructure. See Eq 16. Source: Ref 49. (c) SEM backscattered micrographs of samples heat treated at 955 °C for times of 0.25 or 48 h, followed by water quenching (WQ)

Final Heat Treatment

Heat treatment is used to control final microstructure and properties and to relieve residual stresses for both alpha/beta and beta titanium alloys (Ref 50, 51). For alpha/beta alloys, these treatments include (1) solution treatment at relatively high temperatures in the alpha+beta field, followed by water quenching or air cooling, (2) so-called recrystallization treatment at slightly lower temperatures, followed by slow cooling, (3) mill annealing at a temperature of the order of 700 °C, and (4) beta annealing followed by water quenching, forced convection, and so on. Each of these treatments may be followed with a final aging/stabilization heat treatment. The principal final heat treatment for beta and near-beta alloys consists of aging of material previously beta annealed and water quenched.

In this section, two broad categories of heat treatment models are discussed dealing with (1) microstructure evolution during cooling following alpha+beta solution treatment of alpha/beta titanium alloys with an equiaxed-alpha microstructure and (2) decomposition of beta following beta annealing of alpha/beta alloys or during the aging of beta titanium alloys.

Alpha+Beta Heat Treatment of Alpha/Beta Alloys. Final alpha+beta heat treatment of alpha/beta titanium alloys with an equiaxed-alpha microstructure is used to control the volume fraction and size of the primary alpha and the nature of the transformed matrix phase. To this end, components are typically solution treated high in the alpha+beta phase field and cooled via water quenching or forced convection. Because of section size variations, the cooling rate and hence microstructure vary within the component, thereby leading to nonuniform properties. Predictive models can be very useful to quantify these variations.

Diffusion Analysis. A diffusion analysis can be applied to estimate the growth of primary alpha particles and the onset of the

decomposition of residual beta during continuous cooling following alpha+beta solution treatment (Ref 52–54). The approach is based on the "exact" one-dimensional solution for the growth of an isolated spherical precipitate as a function of time, t, and diffusivity, D (Ref 55, 56) and consideration of the effect of soft impingement on the supersaturation controlling the rate of growth.

The diffusion solution is the following:

$$R(t) = 2\lambda(Dt)^{1/2} \qquad \text{(Eq 17)}$$

To treat continuous cooling, the differential form of Eq 17 is used:

$$\mathrm{d}R/\mathrm{d}t = 2\lambda^2 D/R \qquad \text{(Eq 18)}$$

In Eq 17 and 18, the parameter λ is given by the relation:

$$\{\lambda^2 \exp(\lambda^2)\}\bullet[(\exp(-\lambda^2)) - (\lambda\pi^{1/2}\mathrm{erfc}(\lambda))]$$
$$= \Omega_s/2 \qquad \text{(Eq 19a)}$$

in which Ω_s denotes the supersaturation:

$$\Omega_s = (C_M - C_I)/(C_P - C_I) \qquad \text{(Eq 19b)}$$

Here, C_M, C_I, and C_P represent the compositions of the matrix far from the matrix-particle interface, the matrix at the matrix-particle interface, and the particle at the matrix-particle interface, respectively. For a diffusion-controlled reaction, C_I and C_P correspond to the equilibrium matrix and particle compositions, respectively.

Equation 19(a) is not readily inverted to obtain λ as a function of Ω_s. However, $\lambda = \lambda(\Omega_s)$ may be obtained from a simple FORTRAN program or using commercial software such as the Solver tool of Microsoft Excel (Fig. 12).

It should be mentioned that the approximate constant-radius diffusion solution (Ref 38, 57) yields a relation analogous to Eq 18 but with a prefactor of Ω_s instead of $2\lambda^2$. Figure 12 shows that $2\lambda^2/\Omega_s$ is greater than unity even for relatively small values of the supersaturation Ω_s. Thus, the constant-radius solution underestimates the growth rate except for values of Ω, approaching zero. Even for $\Omega_s < 0.1$ (the range of applicability often quoted in the literature), substantial errors in the prediction of particle size are inevitable when using the constant-radius solution.

To simulate actual situations involving a collection of alpha particles, the interaction of the diffusion fields (i.e., the soft impingement effect) must be taken into account. In such cases, the far-field matrix composition C_M is usually adjusted using an approximation derived from a mass balance (Ref 57):

$$C_M = (C_o - f_\alpha C_\alpha)/(1 - f_\alpha) \qquad \text{(Eq 20)}$$

in which C_o and f_α denote the overall alloy composition and the volume fraction of the
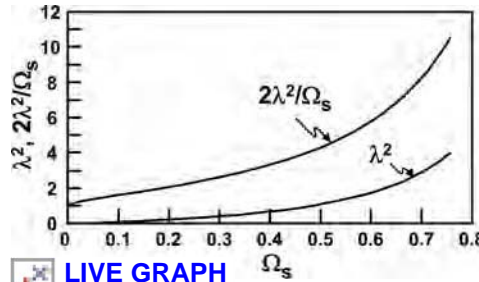
Fig. 12 Dependence of the growth parameter λ^2 and the ratio $2\lambda^2/\Omega_s$ on the supersaturation, Ω_s. Source: Ref 52

precipitate/particle phase (i.e., primary alpha). Because the composition of the alpha phase in alloys such as Ti-6Al-4V often shows relatively little variation with temperature, the particle compositions C_α and C_P are equivalent and constant.

Numerical Solution. The solution of Eq 18 subject to the soft impingement criterion (Eq 20) can readily be accomplished using a spreadsheet method. The required input data include the initial alpha-particle radius, solution temperature/volume fraction of alpha phase, phase compositions, diffusivity as a function of temperature, and cooling rate (Ref 54). The beta approach curve (showing the volume fraction of beta as a function of temperature) and phase compositions for Ti-6Al-4V are given in Fig. 13. The diffusivity of aluminum and vanadium in beta titanium with a composition similar to that of beta in Ti-6Al-4V is as follows (Ref 28):

$$\text{Aluminum}: D_\beta^{Al}(\mu m^2/s)$$
$$= 199{,}200 \exp(-18{,}040/T(K)) \qquad \text{(Eq 21a)}$$

$$\text{Vanadium}: D_\beta^{V}(\mu m^2/s)$$
$$= 77{,}000 \exp(-17{,}460/T(K)) \qquad \text{(Eq 21b)}$$

The instantaneous alpha-particle radius, R, can be converted to volume fraction, f_α, using the following expression:

$$f_\alpha = f_{\alpha o}(R/R_o)^3 \qquad \text{(Eq 22)}$$

Here, $f_{\alpha o}$ and R_o denote the initial volume fraction and particle radius, respectively. The solution procedure then comprises the following steps:

1. Specify initial temperature/volume fraction of alpha, initial alpha-particle radius, diffusing element and its overall content (C_o), cooling rate, and time/temperature decrement.
2. Impose a temperature/time decrement.
3. Calculate the supersaturation, Ω_s (Eq 19b), using measured equilibrium phase compositions for alpha and beta (e.g., Fig. 13) and
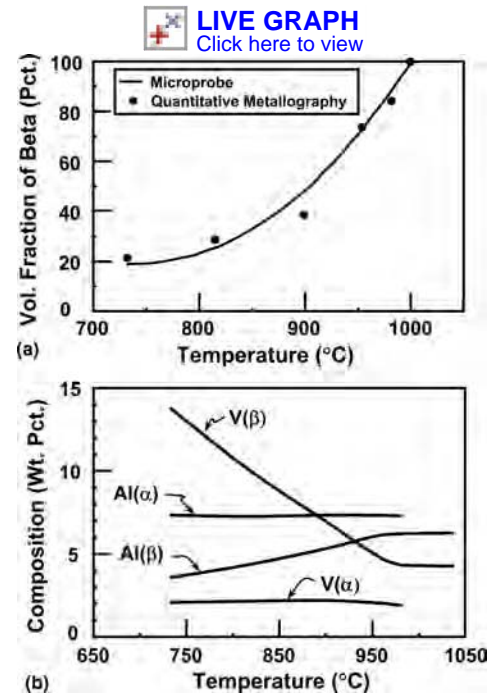
Fig. 13 Phase equilibria data for Ti-6Al-4V with a regular interstitial level. (a) Beta-approach curve. (b) Phase compositions. Source: Ref 52

the matrix composition (C_M) determined from Eq 20).
4. Calculate λ (Eq 19a) and the rate of change of the alpha-particle radius (Eq 18).
5. Impose a temperature/time decrement.
6. Calculate the new alpha-particle radius and the volume fraction of alpha (Eq 22).
7. Repeat steps 3 to 6 until the desired final temperature is reached.

The output of the spreadsheet calculations includes the alpha-particle radius and supersaturation as a function of temperature.

The usefulness of the model for diffusional growth of equiaxed alpha following solution treatment high in the alpha+beta phase field has been verified using Ti-6Al-4V (Ref 52). Experiments and corresponding model calculations were conducted for several different peak temperatures (and hence initial volume fractions of alpha) and cooling rates. The model calculations were done assuming two different initial alpha-particle radii that bounded those measured (4 μm, 5 μm) and diffusion controlled by the supersaturation of aluminum or vanadium. Sample microstructural observations and model results for a peak (solution) temperature of 955 °C and a cooling rate of 42 °C/min are shown in Fig. 14. Model predictions of the volume fraction of alpha as a function of temperature during cooldown tended to lie above the measurements. This behavior can be explained on the basis of the decomposition of the beta matrix during cooling, a factor which would reduce the supersaturation that drives the growth of primary alpha. A parameter known as the secondary alpha factor (which is a
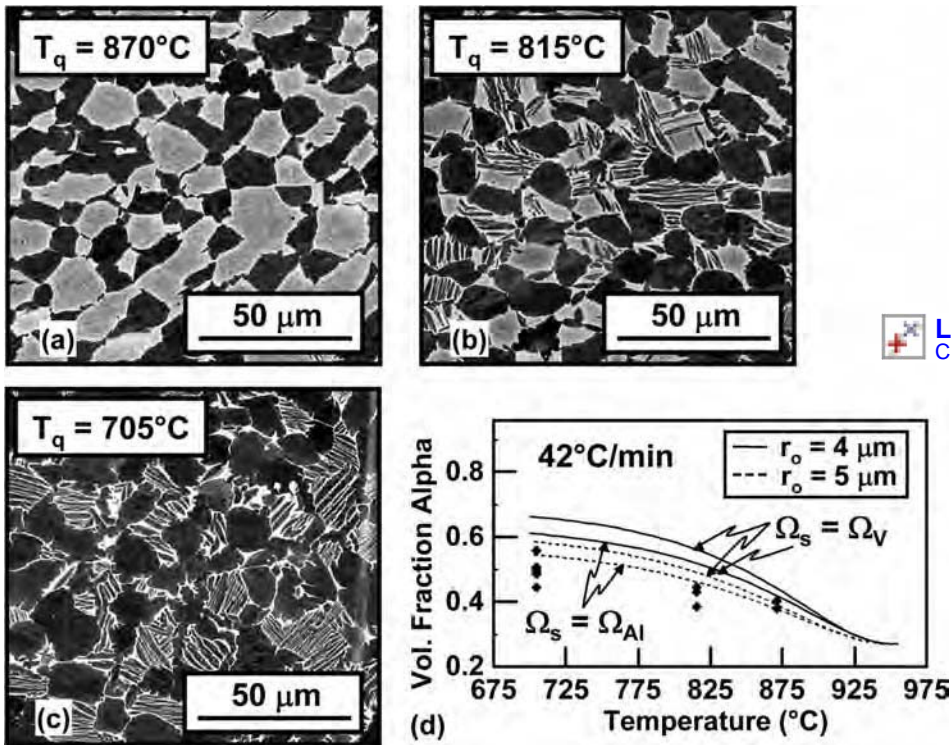
Fig. 14 Alpha/beta heat treatment results for Ti-6Al-4V. (a,b,c) SEM backscattered micrographs of samples solution treated at 955 °C, cooled at 42 °C/min, and water quenched at the temperatures (T_q) indicated. (d) Comparison of corresponding measurements of the volume fraction of primary alpha (data points) and diffusion-model predictions (lines). Source: Ref 52

Fig. 15 Phase-field model predictions of the growth of primary alpha particles in a hypothetical titanium alloy that was solution treated, rapidly cooled to a lower temperature, and soaked isothermally. Simulations were done for an initial alpha-particle distribution that was random or uniform. Source: Ref 59

function of cooling rate, diffusivity, and supersaturation) was therefore derived to predict the onset of beta transformation and the temperature below which diffusional growth of primary alpha would be retarded or cease (Ref 52).

Model Enhancements. A number of enhancements have been made to the diffusion model described in the previous section. These include the ability to treat the growth of primary alpha with (a) a prespecified size distribution (Ref 53) or (b) a mixture of equiaxed and lamellar-platelet morphologies (such as would be present in material that is not fully spheroidized) (Ref 54). The former work has shown that the average alpha-particle size for an arbitrary initial size distribution varies with temperature in a manner identical to that predicted for an initial distribution of mono-size particles, that is, the case summarized previously. The other effort revealed that lamellar alpha particles grow substantially slower than equiaxed particles.

The phase-field (PF) method has also been applied to obtain detailed information on the temporal and spatial evolution of alpha-particle size and shape during final alpha/beta heat treatment of titanium alloys with an equiaxed-alpha microstructure (Ref 58, 59). The PF approach uses the numerical solution of (1) the generalized (Cahn-Hilliard) diffusion equation (Ref 60) in terms of chemical-potential gradients and chemical mobility (rather than

concentration gradients and diffusivities) and (2) the time-dependent Ginzburg-Landau equation that describes structural changes that accompany phase transformations. Because of the use of chemical potentials in the analysis, an important factor in this approach is the estimation of the Gibbs free energy as a function of concentration, order parameter (i.e., phase), and their gradients; methods to calculate the Gibbs free energy for PF simulations are described in the article "Phase-Field Microstructure Modeling" in this Volume. An example of the application of the PF method for the growth of primary alpha particles in a hypothetical titanium alloy that was solution treated, rapidly cooled to a lower temperature, and soaked is shown in Fig. 15 (Ref 59). The results show noticeable differences in growth behavior for the cases of randomly versus uniformly spaced particles.

Beta Annealing and Beta Decomposition. The decomposition of single-phase beta serves as the basis for two types of common heat treatments: the beta annealing of alpha/beta titanium alloys and the aging of beta-annealed beta and near-beta alloys.

Alpha/Beta Titanium Alloys. The microstructure formed in alpha/beta titanium alloys that are beta annealed is heavily dependent on the cooling rate. High rates give rise to a martensitic-type transformation, moderate rates to a basketweave/Widmanstätten microstructure, and
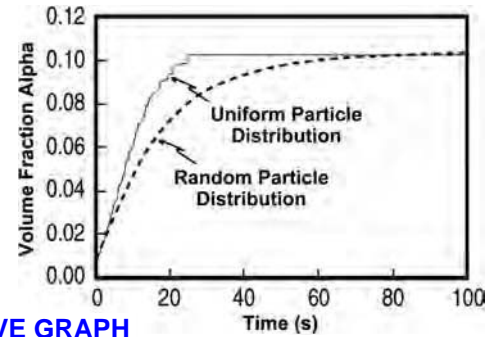
slow rates to colony-alpha structure. The formation of Widmanstätten on colony alpha in alloys such as Ti-6Al-4V involves nucleation and growth of the alpha phase and has been modeled using both phenomenological and phase-field models.

Phenomenological models, which generally provide only temporal averages, have most often relied on the application of the Johnson-Mehl-Avrami-Kolmogorov (JMAK) approach (Ref 61, 62) and modifications of it. For example, a JMAK approach can be used to describe phase decomposition following beta annealing of alpha/beta alloys such as Ti-6Al-4V and Ti-6Al-2Sn-4Zr-2Mo-0.1Si (Ti6242Si) under both isothermal and continuous cooling conditions (Ref 63, 64). In the simpler isothermal instance, the standard phenomenological JMAK relation is used to fit measurements:

$$X = 1 - \exp(-k\ t^{n_a}) \qquad \text{(Eq 23)}$$

in which X denotes the fraction transformed, the rate constant k is a function of temperature, t is time, and n_a denotes the JMAK (Avrami) exponent. Plots of $\log(\ln[1/(1 - X)])$ versus $\log t$ are made to determine the Avrami exponent. In Ref 63, it was found that the values of n_a for both Ti-6Al-4V and Ti6242Si were between 1.15 and 1.6. The lower values ($n_a \sim 1$) pertained to isothermal transformation at temperatures from ~900 to 950 °C, at which alpha nucleated heterogeneously at and grew from the beta grain boundaries; such values are in agreement with classical predictions in the literature for continuous nucleation at grain boundaries (Ref 65). The higher n_a values were found for temperatures below 900 °C, at which alpha nucleated heterogeneously at the grain boundaries as well as homogeneously within the beta grains.

The interpretation of decomposition kinetics under continuous-cooling conditions can be somewhat more complex using the JMAK formalism. The challenge associated with such cases is best described by rewriting Eq 23 in

terms of the explicit nucleation and growth rates, N and G, respectively:

$$X = 1 - \exp(-C_a N \, G^q t^{n_a}) \qquad \text{(Eq 24)}$$

in which C_a and q are constants. The temperature dependence of N and G are both described by Arrhenius-type relations, that is, N, $G \sim \exp(-Q/R_g T)$, in which Q is an activation energy, R_g is the gas constant, and T is absolute temperature. The activation energy for growth is relatively constant. However, Q for nucleation decreases with decreasing temperature (Ref 65), thus leading to a nucleation rate that rapidly increases with decreasing temperature and a complex function for the overall rate constant.

To simplify the modeling of beta decomposition under continuous-cooling conditions, therefore, an alternate approach based on isothermal JMAK data can be applied (Ref 64). Specifically, the fractions transformed under nominally isothermal conditions are summed for a series of closely spaced temperature decrements (Ref 66). The calculation scheme makes use of a series of fictitious times, each of which corresponds to that which had pertain if the entire fraction that had transformed at prior higher temperatures had occurred at the current temperature. Using this basic approach, Malinov et al. (Ref 64) developed a computer code to simulate continuous-cooling transformation behavior of Ti-6Al-4V following beta annealing. It was assumed that n_a was constant ($= 1.13$); the code used an error-minimization routine to derive $k = k(T)$ in the JMAK relation (Fig. 16a). With this information, the computer simulation yielded predictions of transformed fractions as a function of cooling rate and temperature in excellent agreement with the measurements (Fig. 16b).

Beta and Near-Beta Titanium Alloys. The precipitation of alpha from solution-treated beta and near-beta titanium alloys during isothermal or continuous-cooling heat treatments has been analyzed in detail by Da Costa Teixeira, Appolaire, and their coworkers (Ref 67–69). In work similar to that of Malinov et al. (Ref 63) for alpha/beta titanium alloys, a numerical method based on the incremental application of the JMAK relation was applied to quantify the cooling-transformation behavior of the beta-rich alloy Ti-17 (Ref 67). The model treated the precipitation of grain-boundary alpha, Widmanstätten alpha that grew from the grain-boundary alpha, and Widmanstätten alpha that nucleated homogeneously with the beta grains.

More detailed mesoscale models that describe the temporal and some aspects of the spatial evolution of microstructure have also been developed for (1) isothermal transformation of an annealed equiaxed beta-grain microstructure (Ref 67) and (2) isothermal and cooling transformation of alloys with residual substructure from prior-beta hot working (Ref 67).

The mesoscale model for isothermal transformation (Ref 68) focuses on four discrete
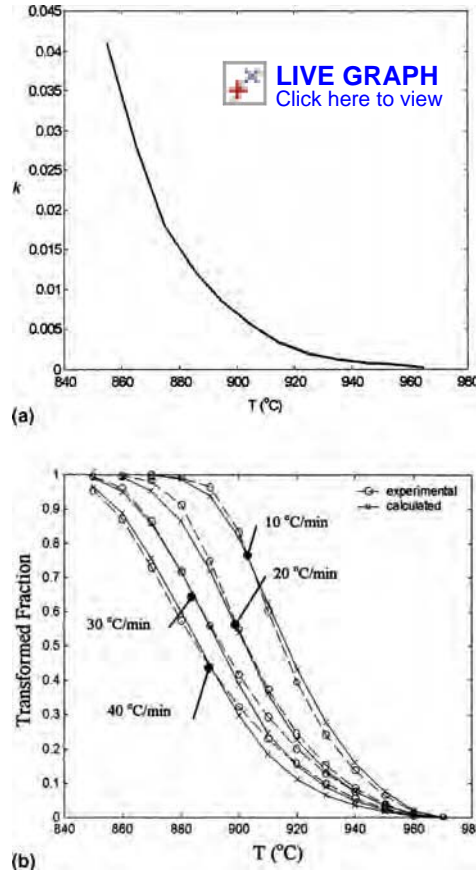
Fig. 16 Application of the Johnson-Mehl-Avrami-Kolmogorov relation to predict cooling-transformation behavior of Ti-6Al-4V following beta annealing. (a) Fitted dependence of the rate constant k on temperature. (b) Comparison of predicted and measured transformed fractions as a function of cooling rate and temperature. Source: Ref 64

LIVE GRAPH
Click here to view

phenomena: the nucleation of alpha at the beta grain boundaries (α_{gb}), the growth of this α_{gb}, the appearance of Widmanstätten-alpha sideplates (α_{wgb}) along the α_{gb} and the growth of the α_{wgb}. Classical nucleation theory and Fick's law are used to quantify the rate of nucleation and the planar growth of α_{gb}, respectively. The appearance of α_{wgb} sideplates is described empirically based on a critical thickness of the α_{gb}, and their growth is modeled using a diffusion approach analogous to that for α_{gb}.

Detailed transformation behavior is analyzed by implementing the various rate equations into a mesoscale simulation comprising 1000 tetrakaidecahedral (14-faced) grains. To introduce a statistical aspect into the simulations, the beta grain-boundary energy (and hence the wetting angle of α_{gb}, nuclei) are varied from one grain boundary to another. Transition from one stage of the transformation to another at a specific grain boundary is based on several rules: (1) α_{gb} is allowed to grow once a grain boundary is covered 90% by nuclei, and (2) α_{wgb} sideplates are allowed to grow once the α_{gb} layer reaches the critical thickness. The growth of a
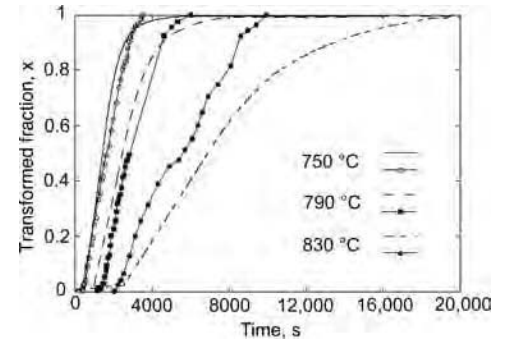
Fig. 17 Comparison of measured (solid lines, data points) and predicted (broken lines) isothermal transformation kinetics for the titanium alloy Beta-Cez. Source: Ref 68

LIVE GRAPH
Click here to view

given colony of α_{wgb} sideplates proceeds until the total transformed volume is equal to the size of the grain in which it is located, or its length is equal to the grain size minus the length of any colony growing from an opposite grain boundary. Predictions of the fraction transformed as a function of time and the average number of colonies per beta grain show reasonably good agreement with measurements for the Beta-Cez alloy (Fig. 17).

The mesoscale model and associated simulations for transformation of beta titanium alloys with residual substructure (subgrains) (Ref 69) are similar. The principal enhancements include: (1) modification of the geometric representation to include subgrain boundaries of various misorientations within the beta grains, (2) nucleation of α_{gb} at both subgrain and grain boundaries at a rate dependent on boundary misorientation, and (3) the appearance of α_{wgb} at a thickness of α_{gb} which depends on boundary misorientation. Simulation results revealed that it must be assumed that α_{wgb} plates are able to cross subgrain boundaries in order to approach the equilibrium volume fraction of alpha during long-time isothermal aging. With this assumption, good predictions of the dependence of transformation kinetics on subgrain size, number of colonies per grain, and so on are obtained for alloys such as Beta-Cez (Fig. 18).

The phase-field technique has also been applied to investigate the decomposition of metastable beta to form colony- and basket-weave-alpha microstructures (Ref 59, 70). In this work, alpha sideplates were assumed to nucleate along grain-boundary alpha layers (by the classic Mullins-Sekerka interface-instability mechanism) (Ref 71) and then grow into a supersaturated beta matrix. To this end, random fluctuations were introduced into the initial interface between the grain-boundary alpha and beta phase. Model results showed that the morphology of the alpha plates (breadth-to-thickness ratio) was sensitive to the assumed anisotropy in interfacial energy and the elastic strain energy associated with semicoherency of the alpha and beta phases. By introducing homogeneous nucleation (within beta grains),
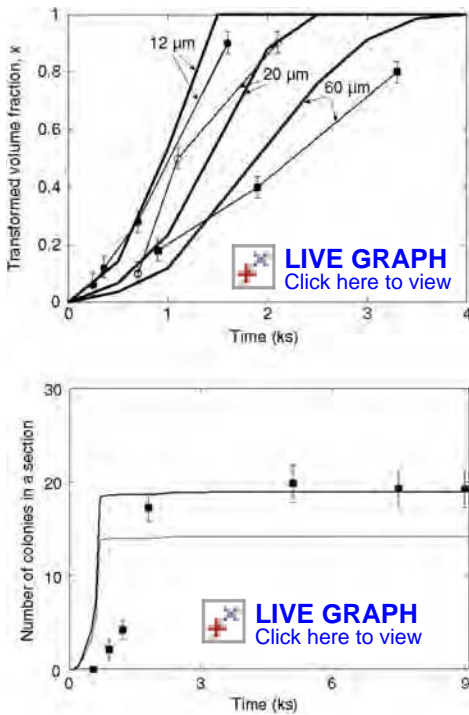
Fig. 18 Comparison of measurements (data points, thin lines) and mesocale-model predictions (thick lines) of microstructure evolution in Beta-Cez. (a) Isothermal transformation kinetics at 800 °C for material prior-beta worked to develop different subgrain sizes. (b) Number of colonies observed in a two-dimensional section as a function of time at 790 °C for material predeformed 25% in the beta-phase field at a strain rate of 0.01 s^{-1}. Source: Ref 69

the evolution of microstructures containing both alpha sideplatelets and Widmanstätten (basketweave) alpha within beta grains was simulated.

Summary and Future Outlook

Models to describe the evolution of microstructure during the thermomechanical processing of titanium alloys have been described. The mechanisms of spheroidization, coarsening, particle growth, and phase decomposition in titanium alloys frequently follow classical nucleation and diffusional-growth behaviors and thus are relatively easy to quantify. In this regard, accurate thermodynamic (phase equilibria) and kinetic data are key to the validation and application of the models for industrially significant alloys. Avenues for future developments in this area are varied and include the following:

- Models for continuous dynamic recrystallization for both alpha titanium and beta titanium, with applications to the dynamic spheroidization of lamellar-colony microstructures as well as for fully wrought, equiaxed microstructures

- Internal-state variable models that describe concurrent plastic flow and the evolution of both microstructure and texture
- Development of general models for thermodynamic, kinetic, and boundary properties needed for simulation of microstructure evolution
- Coupled models of microstructure evolution and the formation/healing of defects such as cavities

REFERENCES

1. I. Weiss and S.L. Semiatin, Thermomechanical Processing of Beta Titanium Alloys—An Overview, *Materials Science and Engineering*, 1998, Vol A243, pp. 44–65.
2. I. Weiss and S.L. Semiatin, Thermomechanical Processing of Alpha Titanium Alloys—An Overview, *Materials Science and Engineering*, 1999, Vol A263, pp. 243–256.
3. A. Choudhury and E. Weingartner, Vacuum Arc Remelting, *ASM Handbook, Volume 15, Casting*, ASM International, Materials Park, OH, 1988, pp. 406–408.
4. G. Luetjering and J.C. Williams, *Titanium*, Springer Verlag, Berlin, 2007.
5. S.L. Semiatin, V. Seetharaman, and I. Weiss, Hot Working of Titanium Alloys—An Overview, *Advances in the Science and Technology of Titanium Alloy Processing*, I. Weiss, et al., Ed., TMS, Warrendale, PA, 1997, pp. 3–73.
6. H. Margolin and P. Cohen, Evolution of the Equiaxed Morphology of Phases in Ti-6Al-4V, *Titanium '80: Science and Technology*, H. Kimura and O. Izumi, Ed., TMS, Warrendale, PA, 1980, pp. 1555–1561.
7. H. Margolin and P. Cohen, Kinetics of Recrystallization of Alpha in Ti-6Al-4V, *Titanium '80: Science and Technology*, H. Kimura and O. Izumi, Ed., TMS, Warrendale, PA, 1980, pp. 2991–2997.
8. I. Weiss, G.E. Welsch, F.H. Froes, and D. Eylon, Mechanisms of Microstructure Refinement in Ti-6Al-4V, *Titanium: Science and Technology*, G. Luetjering,

U. Zwicker, and W. Bunk, Ed., Deutsche Gesellschaft fur Metallkunde e.V., Oberursel, Germany, 1985, pp. 1503–1510.
9. I. Weiss, F.H. Froes, D. Eylon, and G.E. Welsch, Modification of Alpha Morphology in Ti-6Al-4V by Thermomechanical Processing, *Metall. Trans. A*, 1986, Vol 17A, pp. 1935–1947.
10. G. Welsch, I. Weiss, D. Eylon, and F.H. Froes, Lamellar to Equiaxed Grain Structure by Torsional Deformation of Ti-6Al-2Sn-4Zr-2Mo Alloy, *Advances in the Science and Technology of Titanium Alloy Processing*, I. Weiss, et al., Ed., TMS, Warrendale, PA, 1997, pp. 169–183.
11. N. Stefansson and S.L. Semiatin, Unpublished Research, Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, 2001.
12. G.A. Salishchev, S.V. Zherebtsov, S. Yu. Mironov, and S.L. Semiatin, Formation of Grain Boundary Misorientation Spectrum in Alpha-Beta Titanium Alloys with Lamellar Structure Under Warm and Hot Working, *Materials Science Forum*, 2004, vols. 467–470, pp. 501–506.
13. S. Zherebtsov, M. Murzinova, S. Mironov, G. Salishchev, and S.L. Semiatin, Microstructure Refinement in Ti-6Al-4V during Warm Deformation, *Collected Proceedings, 2008 TMS Annual Meeting*, Volume I, *Materials Processing and Properties*, TMS, Warrendale, PA, 2008, pp. 211–217 (pdf only).
14. T. Furuhara, B. Poorganji, H. Abe, and T. Maki, Dynamic Recovery and Recrystallization in Titanium Alloys by Hot Deformation, *JOM*, January 2007, Vol 59(1), pp. 64–67.
15. N. Stefansson and S.L. Semiatin, Mechanisms of Globularization of Ti-6Al-4V during Static Heat Treatment, *Metall. Mater. Trans. A,* 2003, Vol 34A, pp. 691–698.
16. S.L. Semiatin and T.R. Bieler, Effect of Texture and Slip Mode on the Anisotropy of Plastic Flow and Flow Softening during Hot Working of Ti-6Al-4V, *Metall. Mater. Trans. A*, 2001, Vol 32A, pp. 1787–1799.
17. A.A. Salem and S.L. Semiatin, Hot Deformation of Ti-6Al-4V Single-Colony Samples, *Maten. Sci. Eng.* Vol A508, 2009, pp. 114–120.
18. S. Suri, G.B. Viswanathan, T. Neeraj, D.-H. Hou, and M.J. Mills, Room Temperature Deformation and Mechanisms of Slip Transmission in Oriented Single-Colony Crystals of an Alpha/Beta Titanium Alloy, *Acta Mater.*, 1999, Vol 47, pp. 1019–1034.
19. S.L. Semiatin and J.J. Jonas, *Formability and Workability of Metals: Plastic Instability and Flow Localization*, ASM International, Materials Park, OH, 1984.
20. S.L. Semiatin, V. Seetharaman, and I. Weiss, Flow Behavior and Globularization Kinetics during Hot Working of Ti-6Al-4V with a Colony Alpha Microstructure, *Materials Science and Engineering*, 1999, Vol A263, pp. 257–271.

21. E.B. Shell and S.L. Semiatin, Effect of Initial Microstructure on Plastic Flow and Dynamic Globularization during Hot Working of Ti-6Al-4V, *Metall. Mater. Trans. A*, 1999, Vol 30A, pp. 3219–3229.

22. T.R. Bieler and S.L. Semiatin, The Origins of Heterogeneous Deformation during Hot Working of Ti-6Al-4V, *Inter. J. Plasticity*, 2002, Vol 18, pp. 1165–1189.

23. W.W. Mullins, Theory of Thermal Grooving, *J. Applied Physics*, 1957, Vol 28, pp. 333–339.

24. W.W. Mullins, Grain-Boundary Grooving by Volume Diffusion, *Trans. TMS-AIME*, 1960, Vol 218, pp. 354–361.

25. G.L.J. Bailey and H.C. Watkins, Surface Tensions in the System Solid Copper Molten Lead, *Proc. Phys. Society*, 1950, Vol B63, pp. 350–358.

26. J.W. Martin, R.D. Doherty, and B. Cantor, *Stability of Microstructure in Metallic Systems*, Cambridge University Press, Cambridge, UK, 1997.

27. S.L. Semiatin, B.C. Kirby, and G.A. Salishchev, Coarsening Behavior of an Alpha-Beta Titanium Alloy, *Metall. Mater. Trans. A*, 2004, Vol 35A, pp. 2809–2819.

28. S.L. Semiatin, T.M. Brown, T.A. Goff, P.N. Fagin, D.R. Barker, R.E. Turner, J.M. Murry, J.D. Miller, and F. Zhang, Diffusion Coefficients for Modeling the Heat Treatment of Ti-6Al-4V, *Metall. Mater. Trans. A*, 2004, Vol 35A, pp. 3015–3018.

29. T.H. Courtney and J.C. Malzahn Kampe, Shape Instabilities of Plate-Like Structures—II. Analysis, *Acta Metall.*, 1989, Vol 37, pp. 1747–1758.

30. S.L. Semiatin, N. Stefansson, and R.D. Doherty, Prediction of the Kinetics of Static Globularization of Ti-6Al-4V, *Metall. Mater. Trans. A*, 2005, Vol 36A, pp. 1372–1376.

31. N. Stefansson, S.L. Semiatin, and D. Eylon, The Kinetics of Static Globularization of Ti-6Al-4V, *Metall. Mater. Trans. A*, 2002, Vol 33A, pp. 3527–3534.

32. A.F. Wilson, V. Venkatesh, R. Pather, J.W. Brooks, and S.P. Fox, The Prediction of Microstructure Development during Ti-6Al-4V Billet Manufacture, *Ti-2003: Science and Technology*, G. Luetjering and J. Albrecht, Ed., Wiley-VCH Verlag GmbH, 2004, pp. 321–329.

33. J.W. Brooks, M.D. Griffiths, A.J. Storton, and R. Waddingham, Modeling Microstructure Development—Measurement and Validation, *Proceedings, Materials Science and Technology 2006 (MS&T'06)*, TMS, Warrendale, PA, 2006, pp. 595–606.

34. J.W. Brooks, M. Mulyadi, and M. Rist, "Constitutive Based Modeling of the Alpha Lath Width and Spheroidization in Ti-6-4," TMS Annual Meeting, New Orleans, LA, March 2008.

35. I.M. Lifshitz and V.V. Slyozov, The Kinetics of Precipitation from Supersaturated Solid Solutions, *J. Phys. Chem. Solids*, 1961, Vol 19, pp. 35–50.

36. C. Wagner, Theory of Precipitate Change by Redissolution, *Zeit. Elektrochem.*, 1961, Vol 65, pp. 581–591.

37. H.A. Calderon, P.W. Voorhees, J.L. Murray, and G. Kostorz, Ostwald Ripening in Concentrated Alloys, *Acta Metall. et Mater.*, 1994, Vol 42, pp. 991–1000.

38. R.D. Doherty, Diffusive Phase Transformations in the Solid State, *Physical Metallurgy*, R.W. Cahn and P. Haasen, Ed., North-Holland, Amsterdam, 1996, ch. 15.

39. P.W. Voorhees and M.E. Glicksman, Solution to the Multi-Particle Diffusion Problem with Applications to Ostwald Ripening—II. Computer Simulations, *Acta Metall.*, 1984, Vol 32, pp. 2001–2011.

40. G.A. Sargent, A.P. Zane, P.N. Fagin, A.K. Ghosh, and S.L. Semiatin, Low-Temperature Coarsening and Plastic-Flow Behavior of an Alpha/Beta Titanium Billet Material with an Ultrafine Microstructure, *Metall. Mater. Trans. A*, Vol 39, 2008, pp. 2949–2964.

41. A.K. Ghosh and C.H. Hamilton, Mechanical Behavior and Hardening Characteristics of a Superplastic Ti-6Al-4V Alloy, *Metall. Trans. A*, 1979, Vol 10A, pp 699–706.

42. S.L. Semiatin, M.W. Corbett, P.N. Fagin, G.A. Salishchev, A.K. Ghosh, and C.S. Lee, Dynamic-Coarsening Behavior of an Alpha-Beta Titanium Alloy, *Metall. Mater. Trans. A*, 2006, Vol 37A, pp. 1125–1136.

43. I. Weiss and J.J. Jonas, Interaction between Recrystallization and Precipitation during the High Temperature Deformation of HSLA Steels, *Metall. Trans. A*, 1979, Vol 10A, pp. 831–840.

44. J.E. Bird, A.K. Mukherjee, and J.E. Dorn, Correlations between High-Temperature Creep Behavior and Structure, *Quantitative Relation Between Microstructure and Properties*, D.G. Brandon and A. Rosen, Ed., Israel Universities Press, Jerusalem, Israel, 1969, pp. 255–342.

45. R.C. Gifkins, Grain-Boundary Sliding and Its Accommodation during Creep and Superplasticity, *Metall. Trans. A*, 1976, Vol 7A, pp. 1225–1232.

46. L.D. Graham and R.W. Kraft, Coarsening of Eutectic Microstructures at Elevated Temperatures, *Trans. TMS-AIME*, 1966, Vol 236, pp. 94–102.

47. H.E. Cline, Shape Instabilities of Eutectic Composites at Elevated Temperatures, *Acta Metall.*, 1971, Vol 19, pp. 481–490.

48. N. Vanderesse, E. Maire, M. Darrieulat, F. Montheillet, M. Moreaud, and D. Jeulin, Three-Dimensional Microtomographic Study of Widmanstatten Microstructures in an Alpha/Beta Titanium Alloy, *Scripta Mater.*, 2008, Vol 58, pp. 512–515.

49. S.L. Semiatin and P.S. Poteet, Branch Elimination during Heat Treatment of Titanium Alloys with a Colony-Alpha Microstructure, *Metall. Mater. Trans. A*, Vol 39, 2008, pp. 2538–2541.

50. J.C. Williams and E.A. Starke, Jr., The Role of Thermomechanical Processing in Tailoring the Properties of Aluminum and Titanium Alloys, *Deformation, Processing, and Structure*, G. Krauss, Ed., ASM International, Materials Park, OH, 1984, pp. 279–354.

51. G.W. Kuhlman, A Critical Appraisal of Thermomechanical Processing of Structural Titanium Alloys, *Microstructure/Property Relationships in Titanium Aluminides and Alloys*, Y.-W. Kim and R.R. Boyer, Eds., TMS, Warrendale, PA, 1991, pp. 465–491.

52. S.L. Semiatin, S.L. Knisley, P.N. Fagin, F. Zhang, and D.R. Barker, Microstructure Evolution during Alpha-Beta Heat Treatment of Ti-6Al-4V, *Metall. Mater. Trans. A*, 2003, Vol 34A, pp. 2377–2386.

53. J.D. Miller and S.L. Semiatin, Effect of the Size Distribution of Alpha Particles on Microstructure Evolution during Heat Treatment of an Alpha/Beta Titanium Alloy, *Metall. Mater. Trans. A*, 2005, Vol 36A, pp. 259–262.

54. S.L. Semiatin, T.M. Lehner, J.D. Miller, R.D. Doherty, and D.U. Furrer, Alpha/Beta Heat Treatment of a Titanium Alloy with a Non-Uniform Microstructure, *Metall. Mater. Trans. A*, 2007, Vol 38A, pp. 910–921.

55. H.S. Carslaw and J.C. Jaeger, *Conduction of Heat in Solids*, Oxford University Press, London, 1959.

56. H.B. Aaron, D. Fainstein, and G.R. Kotler, Diffusion-Limited Phase Transformations: A Comparison and Critical Evaluation of the Mathematical Approximations, *J. Appl. Physics*, 1970, Vol 41, pp. 4404–4410.

57. O. Grong and H.R. Shercliff, Microstructural Modeling in Metals Processing, *Progress in Materials Science*, 2002, Vol 47, pp. 163–282.

58. Q. Chen, N. Ma, K. Wu, and Y. Wang, Quantitative Phase Field Modeling of Diffusion-Controlled Precipitate Growth and Dissolution in Ti-6Al-4V, *Scripta Mater.*, 2004, Vol 50, pp. 471–476.

59. Y. Wang, N. Ma, Q. Chen, F. Zhang, S.-L. Chen, and Y.A. Chang, Predicting Phase Equilibrium, Phase Transformation, and Microstructure Evolution in Titanium Alloys, *JOM*, September 2005, Vol 57(9), pp. 32–39.

60. J.W. Cahn and J.E. Hilliard, Free Energy of a Nonuniform System: I. Interfacial Energy, *J. Chemical Physics*, 1958, Vol 28, pp. 258–267.

61. M. Avrami, Kinetics of Phase Change: I. General Theory, *J. Chemical Physics*, 1939, Vol 7, pp. 1103–1112.

62. W.A. Johnson and R.F. Mehl, Reaction Kinetics in Processes of Nucleation and Growth, *Trans. AIME*, 1939, Vol 135, pp. 416–430.

63. S. Malinov, P. Markovsky, W. Sha, and Z. Guo, Resistivity Study and Computer Modeling of the Isothermal Transformation

Kinetics of Ti-6Al-4V and Ti-6Al-2Sn-4Zr-2Mo-0.08Si Alloys, *J. Alloy. Compd.*, 2001, Vol 314, pp. 181–192.

64. S. Malinov, Z. Guo, W. Sha, and A. Wilson, Differential Scanning Calorimetry Study and Computer Modeling of Beta-to-Alpha Transformation in a Ti-6Al-4V Alloy, *Metall. Mater. Trans. A*, 2001, Vol 32A, pp. 879–887.

65. J.W. Cahn, The Kinetics of Grain Boundary Nucleated Reactions, *Acta Metall.*, 1956, Vol 4, pp. 449–459.

66. J.W. Christian, *The Theory of Transformation in Metals and Alloys: Equilibrium and General Kinetic Theory*, 2nd ed., Pergamon Press, Oxford, UK, 1975.

67. J. Da Costa Teixeira, B. Appolaire, E. Aeby-Gautier, S. Denis, G. Cailletaud, and N. Spath, Transformation Kinetics and Microstructures of Ti17 Titanium Alloy during Continuous Cooling, *Mater. Sci. Eng., A*, 2007, Vol A448, pp. 135–145.

68. B. Appolaire, L. Hericher, and E. Aeby-Gautier, Modeling of Phase Transformation Kinetics in Ti Alloys—Isothermal Treatments, *Acta Mater.*, 2005, Vol 53, pp. 3001–3011.

69. J. Da Costa Teixeira, B. Appolaire, E. Aeby-Gautier, S. Denis, and F. Bruneseaux, Modeling of the Effect of the Beta-Phase Deformation on the Alpha-Phase Precipitation in Near-Beta Titanium Alloys, *Acta Mater.*, 2006, Vol 54, pp. 4261–4271.

70. N. Ma, F. Yang, C. Shen, G. Wang, G.B. Viswanathan, P.C. Collins, D. Xu, R. Yang, H.L. Fraser, and Y. Wang, Modeling Formation of Alpha Sideplates in Alpha/Beta Ti Alloys—Effect of Interfacial Energy Anisotropy and Coherency Elastic Strain Energy, *Titanium 2007: Science and Technology*, M. Ninomi, S. Akiyama, M. Ikeda, M. Hagiwara, K. Maruyama, Ed., Japan Institute of Metals, Tokyo, 2007, pp. 287–290.

71. W.W. Mullins and R.F. Sekerka, Morphological Stability of a Particle Growing by Diffusion or Heat Flow, *J. Appl. Phys.*, 1963, Vol 34, pp. 323–329.

Modeling and Simulation of Texture Evolution during the Thermomechanical Processing of Titanium Alloys

S.L. Semiatin, Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA
M.G. Glavicic, Rolls-Royce Corporation, Indianapolis, IN, USA
S.V. Shevchenko and O.M. Ivasishin, Kurdyumov Institute for Metal Physics, Kyiv, Ukraine
Y.B. Chun and S.K. Hwang, Inha University, Incheon, Korea

THE DEVELOPMENT of crystallographic texture, the preferred orientation of grains in a polycrystalline aggregate, during thermomechanical processing can play an important role with regard to the secondary forming response (e.g., deep drawing of sheet) and service performance (e.g., strength, elastic modulus, ductility, fracture toughness) of metallic materials. Crystallographic texture, or simply texture for succinctness, may arise as a result of large-strain deformation, dynamic/static recrystallization, grain growth, or phase transformation (Ref 1). A second form of anisotropy, mechanical texturing or mechanical fibering, refers to the alignment of microstructure, inclusions, and so forth during deformation processes and may also affect mechanical properties such as ductility and fracture toughness. This latter form of texture is not discussed in the present article.

Because of its technological importance, a number of modeling and simulation techniques have been formulated to describe the evolution of crystallographic texture (Ref 2, 3). A majority of these techniques have been developed for predicting textures in face-centered cubic (fcc) and body-centered cubic (bcc) metals. Considerably less work has been done in the area of hexagonal close-packed (hcp) materials, let alone two-phase alloys, one of whose phases is hcp, such as is the case for a number of industrially important titanium alloys.

Unalloyed titanium and titanium alloys comprise a class of material for which texture can be unusually strong and therefore is very important with regard to mechanical properties (Ref 4, 5). The tendency to form strong textures results principally from the low symmetry of the hcp crystal structure that characterizes many titanium alloys at low temperatures, the limited number of slip and twinning systems that can be activated to accommodate imposed deformation of hcp

crystals, and the allotropic transformation of titanium from the bcc beta phase (at high temperatures) to the hcp alpha phase (at low temperatures). The first two of these factors play an important role with regard to the formation of deformation texture in both alpha and alpha/beta titanium alloys. The nature of prior deformation and the allotropic transformation affect transformation-texture development in both alpha/beta and near-beta titanium alloys.

In the broad sense, crystallographic texture is but one, albeit an important one, of the many attributes that are needed to fully characterize the metallurgical state of a material. Thus, the modeling and simulation of texture evolution for titanium alloys is often tightly coupled to microstructure evolution. This article focuses on a number of problems for titanium alloys in which such coupling is critical in the development of quantitative models. For some phenomena, such as spheroidization or coarsening, it is not as important to model microstructure and texture evolution simultaneously. Such situations are summarized in the article "Modeling of Microstructure Evolution during the Thermomechanical Processing of Titanium Alloys" in this Volume.

A number of general considerations related to the characterization, modeling, and simulation of texture are treated in other articles in this Volume: "Crystallographic Texture," "Crystal-Plasticity Fundamentals," "Self-Consistent Texture Modeling," and "Crystal-Plasticity FEM." Hence, this article focuses solely on aspects pertaining to titanium and titanium alloys. To this end, the key aspects of phase equilibria, the description of crystallographic orientations, and slip/twinning behavior of importance for titanium alloys are summarized first. Subsequent sections of this article describe the modeling and simulation of recrystallization and grain growth of single-phase beta and single-phase

alpha titanium and deformation- and transformation- texture evolution in two-phase (alpha/beta) titanium alloys.

Fundamental Considerations for Titanium

The description of phase equilibria, crystallography, and deformation behavior of titanium and titanium alloys is of great importance with regard to understanding and modeling texture evolution. Each of these topics is briefly discussed in this section.

Phase Equilibria/Phase Transformations

Microstructure and texture evolution and control in titanium alloys is heavily dependent on the allotropic transformation from the hcp crystal structure (denoted as alpha phase) found at low temperatures to a bcc crystal structure (denoted as beta phase) at high temperatures (Ref 6). For pure titanium, this occurs at 882 °C (1620 °F). In many titanium alloys, the beta phase is partially stabilized at lower temperatures, and the equilibrium volume fractions of alpha and beta vary with temperature. The temperature at which a specific alloy becomes entirely beta (i.e., alpha + beta → beta) is called the beta transus. The conditions used for hot working and heating/heat treatment are often selected relative to the beta-transus temperature.

There are a number of important classes of titanium alloys whose designations reflect the relative volume fraction of beta retained at room temperature. These include beta and near-beta alloys and near-alpha and alpha/beta alloys (Ref 7, 8). Beta and near-beta alloys have moderate-to-large amounts of beta-stabilizing

elements such as vanadium, molybdenum, tungsten, niobium, chromium, and iron. The beta transi of near-beta alloys are typically in the range of 700 to 850 °C (1300 to 1560 °F) (Ref 7). Near-alpha and alpha/beta alloys have large amounts of alpha-stabilizing elements such as aluminum; the beta transi of near-alpha and alpha/beta alloys are usually of the order of 950 to 1050 °C (1740 to 1920 °F) (Ref 8).

The cooling of alpha/beta alloys from a temperature above the beta transus or the aging of beta-annealed and water-quenched near-beta alloys results in the decomposition of metastable beta. The alpha that is formed has a lath or platelet morphology. The orientation of the alpha is related to that of the parent beta by a classical Burgers relation (Ref 9). The single close-packed plane in the hcp alpha phase (i.e., the basal plane) is parallel to one of the six close-packed planes in the bcc beta phase. In addition, one of the three close-packed ("**a**") directions in the basal plane of the alpha is parallel to one of the two close-packed directions lying within the specific close-packed plane in the beta (Fig. 1). Thus, decomposition of a beta grain may give rise to one or more of 12 (= 6 × 2) possible alpha-phase variants, each with its own orientation within the beta grain. It is rare that a specific beta grain would contain a large number of the 12 possible alpha variants. Often, a subset of these may be found due to a variant-selection process. Various factors may affect variant selection. These include stress (which may be applied macroscopically or develop as a result of temperature gradients, for example) (Ref 10) and, if the beta has been hot worked, residual dislocation substructure on the activated slip systems. Possible variant-selection rules are discussed in the following section on the modeling of transformation texture in alpha/beta titanium alloys.

Crystallography and Description of Orientation

Descriptions of crystallography and the orientations of bcc beta and hcp alpha are critical to the modeling and simulation of texture evolution in titanium alloys.

Crystallography is rather simple for the bcc phase whose unit cell comprises atoms at each corner and an atom at the center of a cube (Fig. 1b). Using standard Miller indices for cubic crystals, the three principal directions are parallel to the cube edges, that is, [100], [010], and [001]. Similarly, the planes comprising cube faces are denoted (100), (010), and (001). Each of the six close-packed planes contains four corner atoms and the central atom and are denoted as {110}-type planes, for example, (110) and ($\bar{1}$10). The four close-packed directions are the cube diagonals and are of the type <111>, for example, the [1$\bar{1}$1] and [$\bar{1}$11] directions lying in the (110) plane.

The crystallography of the hcp unit cell shows less symmetry than the bcc unit cell (Fig. 1a). The four Miller indices for hcp crystals denote relative lengths/inverse intercepts along the three close-packed directions and the normal to the close-packed plane. The close-packed (basal) plane is thus (0001) and its normal is [0001], known also as the c-axis. The three close-packed directions lying in the basal plane are denoted a_1, a_2, and a_3, or [2$\bar{1}\bar{1}$0], [$\bar{1}$2$\bar{1}$0], and [$\bar{1}\bar{1}$20], respectively. Because these three directions lie in the same plane at 120° angles to each other, the third index of the Miller indices for hcp crystals is always equal to minus the sum of the first two indices. Other planes of importance comprise the prism planes ({10$\bar{1}$0}type) and various pyramidal planes, the most important of which are the {10$\bar{1}$1} planes, which contain the directions consisting of the vector sum **c** + **a**, that is, the <11$\bar{2}$3> directions.

The description of texture relies on quantifying the orientation of bcc and hcp crystal axes relative to an orthogonal set of material (sample) reference directions, X, Y, Z (or rolling direction, or RD; transverse direction, or TD; and normal direction, or ND, for sheet or plate materials). For the bcc phase, the crystal axes are taken to be the [100], [010], and [001] directions. Several different choices are possible for the orthogonal set for hcp crystals. The most common consists of the [2$\bar{1}\bar{1}$0], [01$\bar{1}$0], and [0001] directions.

The 3 × 3 rotation (orientation) matrix g, consisting of the cosines of the angles between the crystal axes and the material reference directions, is used to quantify the orientation of a crystal relative to the material reference directions. The ij th member of this matrix is the cosine of the angle between the i th crystal axis and the j th sample reference direction (Fig. 2a) (Ref 1, 11):

$$g = \begin{pmatrix} \cos\alpha_1 & \cos\beta_1 & \cos\gamma_1 \\ \cos\alpha_2 & \cos\beta_2 & \cos\gamma_2 \\ \cos\alpha_3 & \cos\beta_3 & \cos\gamma_3 \end{pmatrix}$$
$$= \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix} \quad \text{(Eq 1)}$$

The description of orientation via the nine-parameter rotation matrix is actually an over-specification, inasmuch as the sample reference directions can be brought into coincidence with the crystal axes (or vice versa) by three independent rotations. The three rotation angles are known as Euler angles. Although a number of conventions have been used to specify Euler angles, the most common is that due to Bunge (Ref 1, 11, 12). The three Bunge Euler angles
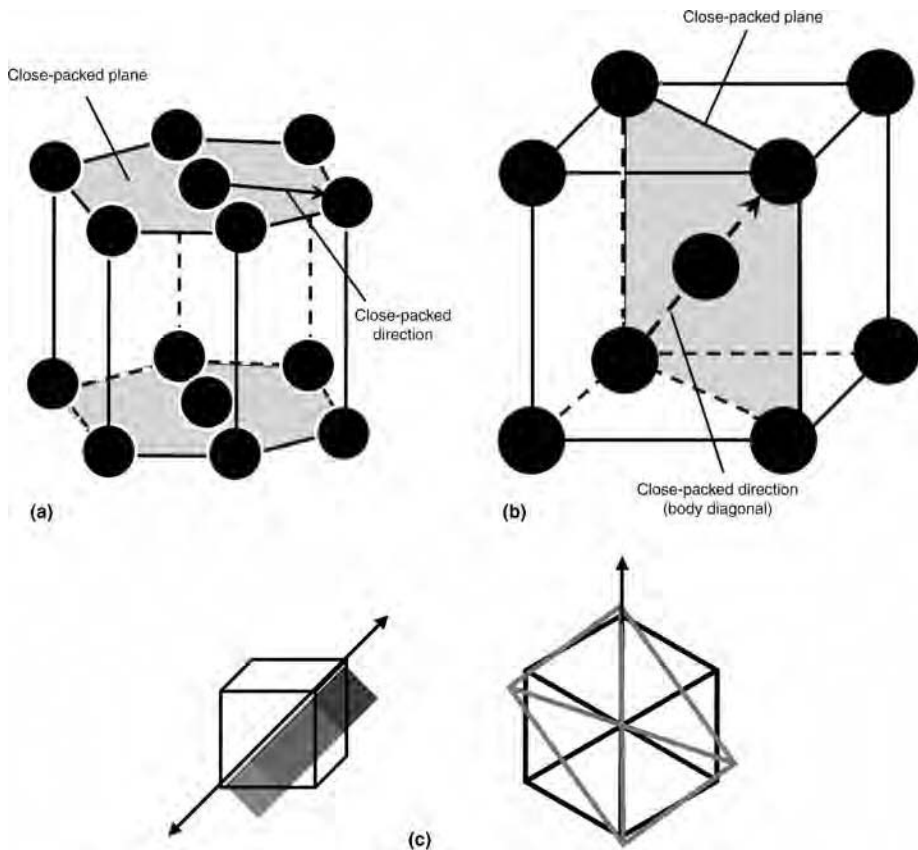
Fig. 1 Close-packed planes/directions and crystallography of (a) hexagonal close-packed alpha titanium and (b) body-centered cubic beta titanium. The close-packed layer of atoms lying between the upper and lower close-packed layers has been removed from (a) for clarity. (c) Burgers orientation relation between the beta and one of the 12 possible alpha variants
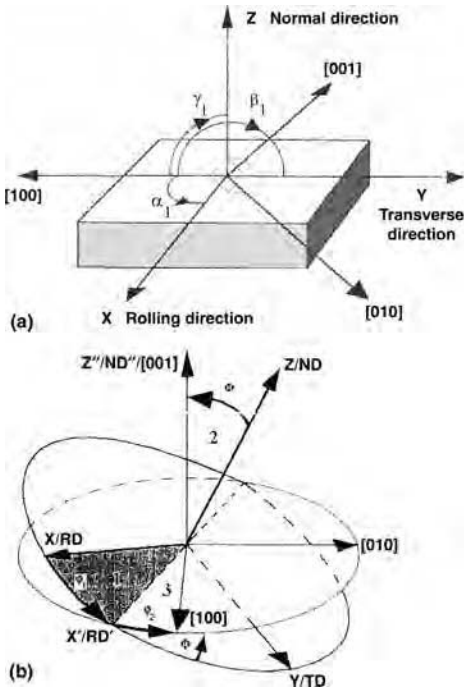
Fig. 2 Orientation relation between the orthogonal sample reference directions (X,Y,Z or rolling direction, or RD; transverse direction, or TD; and normal direction, or ND, for a sheet material) and the orthogonal crystal reference directions [100], [010], and [001] for a cubic crystal. (a) Geometry in terms of angles between the two systems of axes (for clarity, only the angles α_1, β_1, and γ_1 are shown). (b) Euler angles using the Bunge convention. Source: Adapted from Ref 11

Fig. 3 Common slip systems for hexagonal close-packed (alpha) titanium crystals. (a) Basal $<a>$. (b) Prism $<a>$. (c) Pyramidal $<c + a>$

are denoted φ_1, Φ, and φ_2 and comprise the following successive rotations (Fig. 2b):

- Rotation of φ_1 about Z/ND, transforming the Y/TD direction into Y'/TD$'$ and the X/RD direction into X'/RD$'$
- Rotation of Φ about X'/RD$'$, transforming the Y'/TD$'$ direction into Y''/TD$''$ and the Z/ND direction into Z''/ND$''$
- Rotation of φ_2 about Z''/ND$''$

The elements of the rotation matrix (Eq 1) are related to the Euler angles by the following set of relations (Ref 11):

$$g_{11} = \cos \varphi_1 \cos \varphi_2 - \sin \varphi_1 \sin \varphi_2 \cos \Phi$$
$$g_{12} = \sin \varphi_1 \cos \varphi_2 + \cos \varphi_1 \sin \varphi_2 \cos \Phi$$
$$g_{13} = \sin \varphi_2 \sin \Phi$$
$$g_{21} = -\cos \varphi_1 \sin \varphi_2 - \sin \varphi_1 \cos \varphi_2 \cos \Phi$$
$$g_{22} = -\sin \varphi_1 \sin \varphi_2 + \cos \varphi_1 \cos \varphi_2 \cos \Phi$$
$$g_{23} = \cos \varphi_2 \sin \Phi$$
$$g_{31} = \sin \varphi_1 \sin \Phi$$
$$g_{32} = -\cos \varphi_1 \sin \Phi$$
$$g_{33} = \cos \Phi \qquad \text{(Eq 2)}$$

The orientation relation between the reference axes and an arbitrarily oriented crystal or the misorientation between two adjacent crystals can also be represented as a single rotation about a specific axis. For the case of the misorientation between two adjacent crystals 1 and 2, for example, a misorientation matrix Q_{12} is calculated first:

$$Q_{12} = g_1^{\mathrm{T}} g_2 \qquad \text{(Eq 3)}$$

in which g_1^{T} denotes the transpose of the rotation matrix for grain 1, and g_2 is the rotation matrix for grain 2. The misorientation angle θ is then given by the following expression (Ref 11):

$$\cos \theta = (g_{11} + g_{22} + g_{33} - 1)/2 \qquad \text{(Eq 4)}$$

The components r of the axis of rotation (relative to the axes of one of the crystals) are the following (Ref 11):

$$r_1 = g_{23} - g_{32}$$
$$r_2 = g_{31} - g_{13}$$
$$r_3 = g_{12} - g_{21} \qquad \text{(Eq 5)}$$

In Eq 4 and 5, the quantities g_{ij} are the elements of the misorientation matrix Q_{12}.

The overall texture of polycrystalline aggregates is represented by pole figures or orientation distribution functions (ODFs). A pole figure is a two-dimensional stereographic projection of the probability of finding a specified crystallographic direction (pole) at various orientations relative to the sample reference directions. Typical pole figures for beta titanium are 100, 110, and 111. For alpha titanium, 0001, $10\bar{1}0$, $11\bar{2}0$, and $10\bar{1}1$ pole figures are used to quantify texture. Similarly, an ODF summarizes the probability of finding crystals with given Euler angles relative to the sample reference directions.

Slip/Twinning Systems

Alpha Phase. Deformation textures in unalloyed titanium and two-phase titanium alloys tend to be strong because of the low symmetry of hcp crystals and the limited number of modes of deformation due to crystallographic slip or twinning.

The most common slip systems for hcp titanium consist of basal $<a>$ (slip on the basal plane along an $<a>$ direction), prism $<a>$ (slip on a prism plane along an $<a>$ direction), and pyramidal $<c + a>$ (slip on pyramidal planes along a $<c + a>$ direction) (Fig. 3). In the case of an arbitrary imposed deformation, five independent slip systems must be activated within each grain of a polycrystalline aggregate. Because there are only two independent basal $<a>$ systems and two independent prism $<a>$ systems, pyramidal $<c + a>$ slip (or mechanical twinning) must also be activated. Compressive or tensile deformation along the c-axis of an hcp crystal is enabled by $<c + a>$ slip or by twinning.

Deformation of hcp titanium by $<c + a>$ slip involves stresses usually much higher than those required for basal $<a>$ or prism $<a>$ slip. Hence, twinning may provide alternate, softer modes to accommodate deformation along the c-axis, particularly under ambient or cold-working temperatures. The key twinning systems (denoting the twinning plane by the first index and the twinning direction by the second) include $\{11\bar{2}2\} < \bar{1}\bar{1}23 >$, which provides a compressive strain along the c-axis, and $\{10\bar{1}2\} < \bar{1}011 >$ and $\{11\bar{2}1\} < \bar{1}\bar{1}26 >$, each of which gives a tensile strain along the c-axis (Ref 13, 14).

The analysis of deformation texture often involves the application of the strain (or stress) tensor transformation law from a set of reference axes to the hcp crystal axes. To determine the appropriate direction cosines (Eq 1), it is often simplest to apply the dot product formula. In this case, the use of an orthogonal set of reference axes for hexagonal crystals is preferable. As mentioned previously, a typical choice of axes comprises the $[2\bar{1}\bar{1}0]$, $[01\bar{1}0]$, and $[0001]$ directions; that is, $[2\bar{1}\bar{1}0]$ is represented as [100], $[01\bar{1}0]$ as [010], and $[0001]$ as [001]. With this choice, the Miller indices for various slip systems in an equivalent orthogonal system can be calculated (Table 1).

Beta Phase. The bcc metals such as beta titanium are known to deform by slip along $<111>$

Table 1 **Miller index equivalents for hexagonal close-packed (hcp) slip systems in an orthogonal coordinate system**

Slip system	hcp indices	Equivalent indices in cubic coordinates
Basal $\langle a \rangle$	$(0001)\ [2\bar{1}\bar{1}0]$	$(001)\ [1\ 0\ 0]$
Basal $\langle a \rangle$	$(0001)\ [\bar{1}2\bar{1}0]$	$(001)\ [-1/2\ \sqrt{3}/2\ 0]$
Basal $\langle a \rangle$	$(0001)\ [\bar{1}\bar{1}20]$	$(001)\ [-1/2\ -\sqrt{3}/2\ 0]$
Prism $\langle a \rangle$	$(10\bar{1}0)\ [\bar{1}2\bar{1}0]$	$(\sqrt{3}/2\ 1/2\ 0)[-1/2\ \sqrt{3}/2\ 0]$
Prism $\langle a \rangle$	$(1\bar{1}00)\ [\bar{1}\bar{1}20]$	$(\sqrt{3}/2\ -1/2\ 0)[-1/2\ -\sqrt{3}/2\ 0]$
Prism $\langle a \rangle$	$(01\bar{1}0)\ [2\bar{1}\bar{1}0]$	$(0\ 1\ 0)\ [1\ 0\ 0]$
Pyramidal $\langle a \rangle$	$(10\bar{1}1)\ [\bar{1}2\bar{1}0]$	$(1\ 1/\sqrt{3}\ 1/\rho)[-1/2\sqrt{3}/2\ 0]$
Pyramidal $\langle a \rangle$	$(\bar{1}011)\ [\bar{1}2\bar{1}0]$	$(-1\ -1/\sqrt{3}\ 1/\rho)[-1/2\ \sqrt{3}/2\ 0]$
Pyramidal $\langle a \rangle$	$(01\bar{1}1)\ [2\bar{1}\bar{1}0]$	$(0\ 2/\sqrt{3}\ 1/\rho)[1\ 0\ 0]$
Pyramidal $\langle a \rangle$	$(0\bar{1}11)\ [2\bar{1}\bar{1}0]$	$(0\ -2/\sqrt{3}\ 1/\rho)[1\ 0\ 0]$
Pyramidal $\langle a \rangle$	$(1\bar{1}01)\ [\bar{1}\bar{1}20]$	$(1\ -1/\sqrt{3}\ 1/\rho)[-1/2\ -\sqrt{3}/2\ 0]$
Pyramidal $\langle a \rangle$	$(\bar{1}101)\ [\bar{1}\bar{1}20]$	$(-1\ 1/\sqrt{3}\ 1/\rho)[-1/2\ -\sqrt{3}/2\ 0]$
Pyramidal $\langle c+a \rangle$	$(10\bar{1}1)\ [a_1-c],[-a_3-c]$	$(1\ 1/\sqrt{3}\ 1/\rho)[1\ 0-\rho],[1/2\ \sqrt{3}/2-\rho]$
Pyramidal $\langle c+a \rangle$	$(\bar{1}011)\ [-a_1-c],[a_3-c]$	$(-1\ -1/\sqrt{3}\ 1/\rho)[-1\ 0-\rho],[-1/2\ -\sqrt{3}/2-\rho]$
Pyramidal $\langle c+a \rangle$	$(01\bar{1}1)\ [a_2-c],[-a_3-c]$	$(0\ 2/\sqrt{3}\ 1/\rho)[-1/2\ \sqrt{3}/2-\rho],[1/2\ \sqrt{3}/2-\rho]$
Pyramidal $\langle c+a \rangle$	$(0\bar{1}11)\ [-a_2-c],[a_3-c]$	$(0\ -2/\sqrt{3}\ 1/\rho)[1/2\ -\sqrt{3}/2-\rho],[-1/2\ -\sqrt{3}/2-\rho]$
Pyramidal $\langle c+a \rangle$	$(1\bar{1}01)\ [a_1-c],[-a_2-c]$	$(1\ -1/\sqrt{3}\ 1/\rho)[1\ 0-\rho],[1/2\ -\sqrt{3}/2-\rho]$
Pyramidal $\langle c+a \rangle$	$(\bar{1}101)\ [-a_1-c],[a_2-c]$	$(-1\ 1/\sqrt{3}\ 1/\rho)[-1\ 0-\rho],[-1/2\sqrt{3}/2-\rho]$

$\rho \equiv c/a = 1.59$. Source: Ref 15

Fig. 4 Observations of beta grain growth in Ti-6Al-4V sheet with a strong initial texture showing periods of rapid and slow growth. Source: Ref 22

slip directions. However, the slip plane(s) are not well defined. Some have modeled deformation-texture evolution for bcc metals, such as beta titanium assuming pencil glide, that is, slip on any plane containing a <111> direction (Ref 16, 17). Others have assumed a slightly more restrictive approach, assuming slip only on {110}, {112}, and {123} planes (Ref 18, 19). In yet other work, it has been suggested that the active slip planes, particularly at intermediate temperatures, are of the {112} type (Ref 20).

Texture Evolution during Recrystallization and Grain Growth

Modeling of recrystallization and grain growth of single-phase beta or single-phase alpha titanium has a number of important industrial applications. These include the recrystallization and grain growth of alpha/beta and beta titanium alloys during heat treatment in the single-phase beta field (i.e., above the beta-transus temperature) and the recrystallization and grain growth of alpha titanium during heat treatment in the alpha phase field. Grain growth of alloys with no prior hot work during beta annealing is perhaps the simplest problem in texture evolution for titanium alloys and is discussed first. Subsequently, the simulation of coupled recrystallization and grain growth of materials that have been prior hot or cold worked is addressed.

Beta Grain Growth

Beta annealing is often used for alpha/beta titanium alloys to develop a transformed-beta microstructure for fracture-critical aerospace applications. Beta annealing is also applied as a solution treatment for beta and near-beta alloys prior to aging in the alpha/beta phase field.

In both cases, the control of the beta grain size can be very important.

A number of measurements of beta-grain-growth kinetics for alloys such as Ti-6Al-4V and Ti-5Mo-5V-5Al-1Fe-1Cr (VT22) have shown substantial deviations from the parabolic behavior that characterizes classical normal grain growth (that is, $d^n \sim t$, in which d denotes the grain size, t is time, and n is the grain-growth exponent) (Ref 21–25). Specifically, periods of rapid and slow growth have been observed during both isothermal heat treatments and processes involving continuous heating (Fig. 4). Furthermore, grain-growth kinetics have been found to vary quite noticeably in different lots of a given alloy such as Ti-6Al-4V with identical composition and initial microstructure but different initial textures. These observations can therefore be ascribed to initial texture and the evolution of texture during grain growth. Such texture-controlled grain growth is a result of the anisotropy in energy and mobility of beta grain boundaries.

Texture-controlled grain growth during beta annealing of titanium alloys has been simulated using both the Monte Carlo (Potts) and phase-field modeling approaches. In the work of Ivasishin et al. (Ref 24, 25), for example, the Monte Carlo (MC) method was used. A detailed description of the MC technique is contained in the article "Monte Carlo Models for Grain Growth and Recrystallization" in this Volume. The MC formulation of Ivasishin et al. was similar to that in prior efforts with regard to:

- Calculation of energy before and after an elementary orientation "flip" (except that the trial orientation was taken to be that of a randomly chosen neighboring site and not the entire model lattice)
- Quantitative description of the probability for atoms at/near the grain boundaries to change their orientations
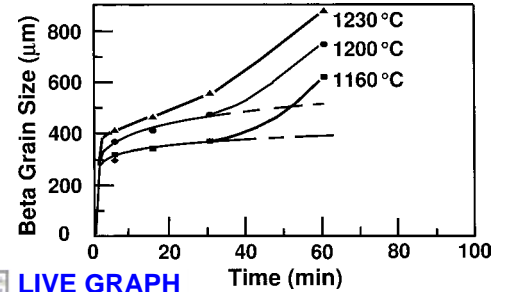
However, a number of improvements were incorporated to obtain quantitative information on the coupling of grain growth and texture evolution. These improvements included:

- Application of a three-dimensional (3-D), rather than a two-dimensional (2-D), analysis
- Use of a large number of possible orientations defined by their respective Euler angles
- Ability to input measured or hypothetical textures using pole figures or ODFs
- Extensive postprocessor software enhancements to quantify intermediate and final grain-size and grain-boundary misorientation distributions, textures/volume fractions of texture components, and so on

To simplify the calculations for large 3-D modeling domains, the separate effects of grain-boundary energy and mobility were combined into a single effective mobility parameter.

To obtain insight into the specific interaction of texture and grain growth, Ivasishin et al. (Ref 24, 25) applied the 3-D code to a number of special cases, including grain growth in (1) an initially textured or untextured material with a misorientation-dependent or independent grain-boundary mobility and (2) a material with an initial two-component texture and misorientation-dependent grain-boundary mobility. Sample MC results for the material with two initial texture components (denoted "A" and "B") are shown in Fig. 5 to 7. The MC-simulated kinetics revealed periods of fast and slow grain growth (Fig. 5a), similar to experimental observations (Fig. 4). As seen in (100) pole figures (Fig. 5b), there were also pronounced variations in texture as grain growth occurred. (Only the (100) poles and not the (010) and (001) poles are shown for the sake of clarity.) The texture changes were characterized by periodic interchanges of the volume fractions of the two components. Time is measured in Monte Carlo steps (MCS). A comparison of Fig. 5(a) and (b) indicates that the periods of rapid grain growth were associated with the times (~15 and ~1200 MCS).at which the volume fractions of the two texture components were approximately equal, and there was thus a large fraction of high-angle boundaries.

(a)

(b)
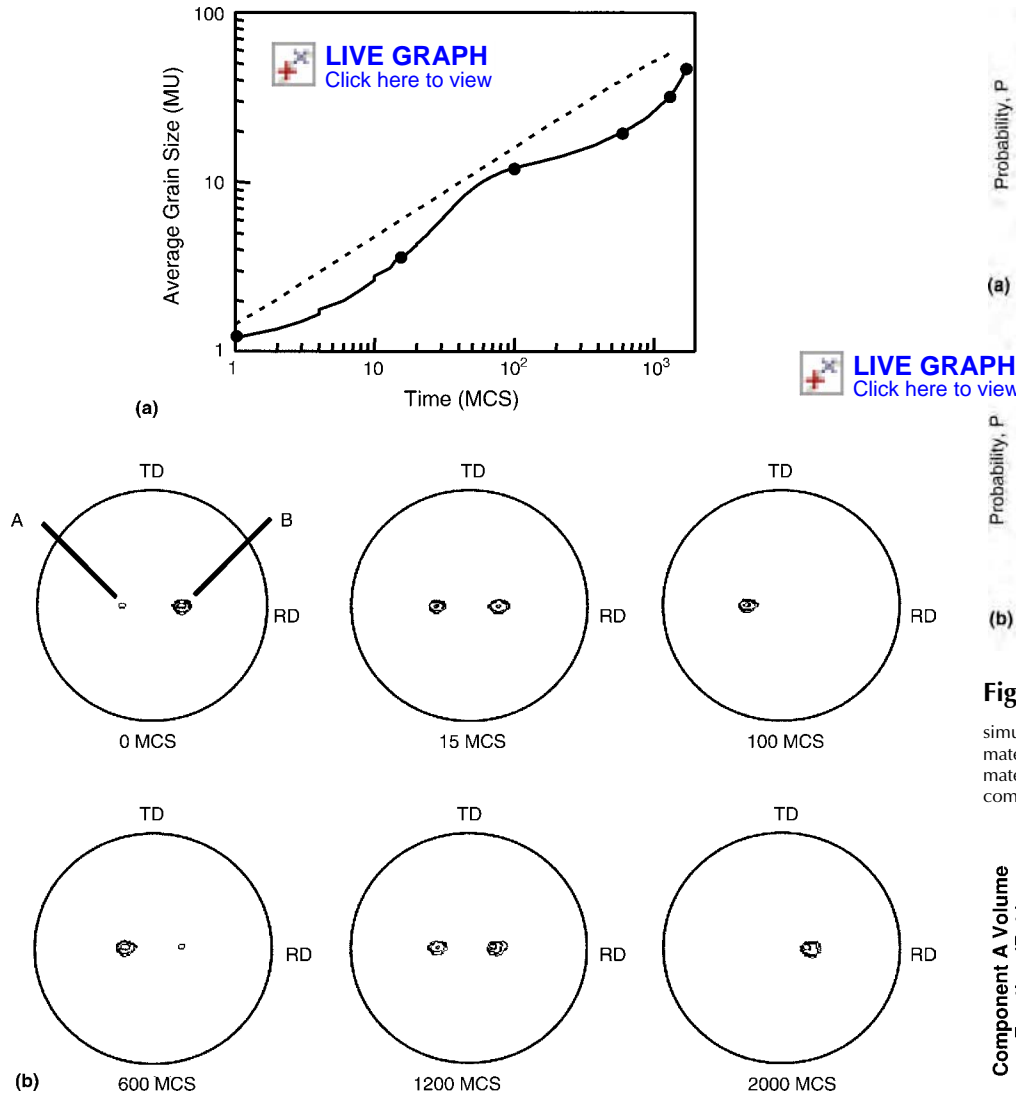
Fig. 5 Monte Carlo model simulation of texture-controlled grain growth for a material with two texture components. (a) Comparison of predicted grain-growth kinetics (solid line) and normal grain-growth kinetics (broken line). MU, model lattice units; MCS, Monte Carlo steps. (b) Simulated (100) pole figures after various times. The iso-intensity lines correspond to 1, 2, 5, 10, and >30 times random. The 010 and 001 poles are not shown for clarity. TD, transverse direction; RD, rolling direction. Source: Ref 25

Fig. 6 Monte Carlo model predictions of the grain-size distributions after 15 Monte Carlo steps for the simulation of texture-controlled grain growth in a material with two texture components. (a) For the entire material. (b) For the grains belonging to texture component "A." MU, model lattice units. Source: Ref 25
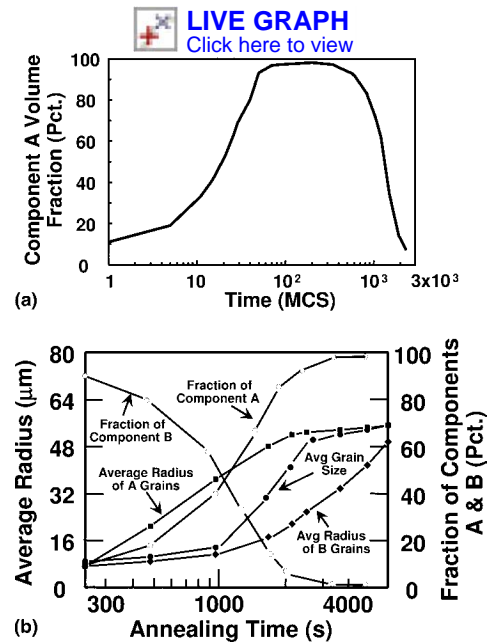
Fig. 7 Predictions of the volume fraction (%) of the texture components and the grain size as a function of annealing time for a two-component initial texture. (a) A Monte Carlo method (Ref 25) is used. MCS, Monte Carlo steps. (b) An analytical approach (Ref 27). The labels for components "A" and "B" in (b) have been switched from those in Ref 27 to coincide with the nomenclature in Ref 25.

The MC-simulated texture evolution for the two-component case can be explained as follows. Because of their lower volume fraction, the "A" grains initially had a much higher probability of having "B" grain neighbors and thus had high-mobility "A-B" boundaries. Hence, small "A" grains were rapidly consumed by large "B" grains, but large "A" grains were able to grow rapidly. Because the "B" grains were in contact preferentially with other "B" grains, they had mainly low-mobility "B-B" boundaries; therefore, most of the "B" grains participated only slightly in the initial stages of grain growth. This led to a very different grain-size distribution for the modeling domain as a whole (Fig. 6a) compared to that for those grains with the "A" texture component (Fig. 6b). After the "A" grains had consumed a majority of the volume, the growth rate of the "A" grains decreased, because there

was now a high probability that each "A" grain was surrounded by other "A" grains. Because "B" had become the minority component, the growth rate of "B" grains exceeded that of the "A" grains. Over long times, such phenomena produced cyclic changes in the relative volume fractions of the "A" and "B" texture components.

The MC predictions of the periodic evolution of a two-component texture (Ref 25) mirror deterministic numerical calculations, such as those of Lücke and his coworkers (Ref 26, 27) (Fig. 7). These earlier calculations were based on the classical expression for the velocity, v, of a grain boundary between grains ν and μ:

$$v_{\nu\mu} = 2M_{\nu\mu}\gamma_{\nu\mu}(R_\nu^{-1} - R_\mu^{-1}) \qquad \text{(Eq 6)}$$

in which $M_{\nu\mu}$, $\gamma_{\nu\mu}$, R_ν, and R_μ denote the grain-boundary mobility, grain-boundary surface

energy, and radii of the adjacent grains, respectively. For the case of a two-component texture (Ref 26), the calculations comprise (1) the

discretization of the grain-size distributions for each of the two texture components into bins and (2) the application of a relation of the form of Eq 6 for the grains in each bin of each texture component. The value of Mγ used in these expressions is an average of those for "AA" and "AB" interfaces (for "A" grains) or for "BB" and "AB" interfaces (for "B" grains) weighted by the volume fractions of "A" and "B" grains.

A comparison of MC predictions of texture-controlled grain growth to experimental observations has shown only qualitative agreement (Ref 24) to date. The differences can be attributed largely to the dearth of quantitative data on grain-boundary energy and mobility as a function of misorientation for titanium alloys. Furthermore, it has been shown using 3-D MC simulations that periods of rapid and slow grain growth may also be due to the anisotropy of boundary properties associated with tilt-twist character/boundary inclination (Ref 28–30).

In similar work, Ma and his coworkers (Ref 31, 32) simulated texture-controlled grain growth during beta annealing of Ti-6Al-4V using the phase-field method. The effect of grain-boundary energy/mobility and specific type of initial texture on grain-growth kinetics and the evolution of the different texture components were evaluated.

Recrystallization of Beta Titanium

The recrystallization of beta titanium plays a key role in breaking down coarse ingot structures to obtain semifinished billet products of essentially all conventional titanium alloys as well as in refining the beta grain size during the thermomechanical processing of plate and sheet product of beta and near-beta titanium alloys. Modeling efforts in this area have begun only recently and have focused primarily on static recrystallization of beta titanium alloys.

Obtaining a uniform (statically) recrystallized microstructure in beta-rich titanium alloys can be very difficult. The challenge arises from the nonuniformity in stored energy within grains (from prior hot deformation in the beta phase field) as well as from grain to grain. Static recrystallization kinetics are thus different in different areas, giving rise to simultaneous propagation of the recrystallization front in some areas and grain growth in other, previously recrystallized areas. Two approaches have been proposed to address problems of this sort. One involves the development of phenomenological processing maps that delineate combinations of forging and heat treatment parameters that produce uniform recrystallization (Ref 33). A second approach, which is just emerging, comprises the application of modeling techniques such as the Monte Carlo (MC) and cellular automata (CA) techniques (Ref 34–36). The CA method (Ref 35, 36) has been applied to treat dynamic recrystallization. The specific method involves tracking the generation/storage and annihilation of dislocations, nucleation events, and migration of a recrystallization front. However, the details of the relationship between the orientations of the deformed matrix and nuclei and the mobility of different types of boundaries have yet to be incorporated, thus limiting the application of CA techniques for quantitative texture predictions. These drawbacks have been remedied using the MC technique, at least for static recrystallization problems, and thus, attention below is focused on this method.

Formulation of Monte Carlo Simulations. An MC routine (Ref 34) has been formulated and applied to provide insight into static recrystallization and grain growth in beta titanium alloys. The calculation procedure in already recrystallized regions is similar to that for standard MC simulations of grain growth, in that the probability, P, for reorientation during an elementary flip trial at site i is taken to be the following:

$$P = U(\theta)W(\theta)\exp(-\Delta E_L/k_b T_s) \text{ for } \Delta E_L > 0 \tag{Eq 7a}$$

and

$$P = U(\theta)W(\theta) \text{ for } \Delta E_L \leq 0 \tag{Eq 7b}$$

In Eq 7(a) and (b), U and W denote the relative (normalized) grain-boundary mobility (M/M_{max}) and energy (γ/γ_{max}) as a function of the scalar misorientation angle θ between the initial and trial orientations; ΔE_L is the change in system energy before and after the flip trial; and k_b is Boltzmann's constant. In the absence of stored energy due to cold or hot working, the system energy E_L is the total grain-boundary energy and is typically given by the following relation for 3-D MC simulations:

$$E_L = \frac{1}{2}\sum_{i=1}^{N}\left(\sum_{j=1}^{26} W(\theta)\left[1 - \delta(q_i, q_j)\right]\right) \tag{Eq 7c}$$

Here, the first summation is taken over the total number, N, of lattice sites in the simulation; the second summation is taken over the 26 first, second, and third nearest neighbors (for a simple cubic lattice) whose orientations/states are taken to be those specifically possible for the trial flip; q_i and q_j denote the orientations of sites i and j; and $\delta(q_i, q_j)$ denotes the Kronecker delta function ($\delta = 1$ for $q_i = q_j$ and $\delta = 0$ for $q_i \neq q_j$). In Eq 7(a), T_s represents the simulation temperature, and $k_b T_s$ is an energy that defines the thermal fluctuations/noise present in MC simulations. When the probability P (Eq 7a and b) is greater than that obtained from a random-number generator that produces values between 0 and 1, the flip is retained; that is, boundary migration occurs. Otherwise, the original orientation/state is restored.

The boundary energy in Eq 7 is often assumed to follow the classic Read-Shockley behavior for low-angle boundaries ($\theta < 15°$) and exhibit a constant (maximum) energy for
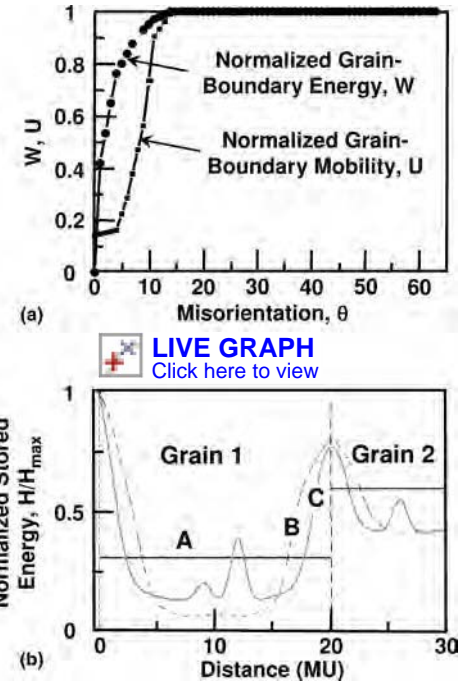
Fig. 8 Baseline dependences. (a) Normalized grain-boundary energy and mobility ($m_a = 1$) dependence on misorientation. (b) Normalized stored energy dependence on location used in Monte Carlo simulations of recrystallization and grain growth. "A," "B," and "C" are energy-level distributions used in Fig. 9. MU, model lattice units. Source: Ref 34

nonspecial, high-angle boundaries ($\theta \geq 15°$) (Fig. 8a):

$$\gamma = \gamma_{max}\left(\frac{\theta}{\theta_{max}}\right)\left[1 - \ln\left(\frac{\theta}{\theta_{max}}\right)\right] \text{ for } \theta < 15° \tag{Eq 8a}$$

$$\gamma = \gamma_{max} \text{ for } \theta \geq 15° \tag{Eq 8b}$$

The mobility dependence on misorientation ($M(\theta)$) can be based on specific measurements or the formulation of Rollett and Holm (Ref 37) (Fig. 8a):

$$M(\theta) = M_{max}\left\{1 - m_a \exp\left[\left(-\frac{\theta}{10°}\right)^3\right]\right\} \tag{Eq 9}$$

In this expression, M_{max} is the maximum mobility (corresponding to nonspecial, high-angle boundaries), and the parameter m_a is set equal to 0 or 1 for the isotropic and the most anisotropic case, respectively.

For flip trials focusing on a site i located at the recrystallization front, Eq 7(a) is replaced by an expression of the following form:

$$P = U(\theta)\frac{H_i}{H_{max}}\exp(-\Delta E_L/k_b T_s) \tag{Eq 10}$$

Here, H_i denotes the stored energy at site i, and H_{max} is the maximum stored energy in the

material; the grain-boundary energy in the pre-multiplier term is neglected because its value is usually several orders of magnitude less than the typical levels of stored work associated with dislocations. An expression similar to Eq 7(c), but which incorporates the stored energy H_i/H_{max}, is used to estimate E_L and hence ΔE_L associated with the flip trial, that is, migration of the recrystallization front. Various spatial dependences of stored energy can be used. These include uniform stored energy within each grain (with each grain's energy scaled by its Taylor factor, M_T) and a stored-work distribution whose magnitude decreases from a maximum at the grain boundary toward the center of the grain (Fig. 8b).

In the MC simulations described in Ref 34, nucleation was assumed to occur at a constant rate (10^{-4} nuclei/ MU^3 per MCS; $MU \equiv$ model lattice unit/ lattice spacing) at grain boundaries and local regions of high stored energy. Oriented nucleation was introduced by assigning various probabilities for the possible misorientations between the deformed matrix and the recrystallization nucleus. Specifically, it was assumed that nuclei were either randomly oriented or formed a $\Sigma 9$, $\Sigma 19$, or $\Sigma 27$ special boundary with the deformed matrix grain, thus giving rise to misorientation angles/rotation axes of $39°<110>$, $26.5°<110>$, or $35°<110>$, respectively; these misorientations are commonly observed in the recrystallization of bcc metals (Ref 38, 39). A special boundary refers to the boundary between two grains (or a matrix grain and a twin within it) for which a certain fraction $1/X$ of the lattice sites of the two grains coincide. The special boundary is then denoted as being of the type ΣX.

Monte Carlo Simulation Results. A number of interesting predictions were obtained by Ivasishin et al. (Ref 34) from 3-D MC simulations using the aforementioned formulation. The simulated materials and initial textures were based on cold-drawn beta titanium alloy Timetal LCB (Ti-4.5Fe-6.8Mo-1.5Al), with partial fibers denoted as β and η (Ref 1) or hot rolled sheet of VT22 (Ti-5Al-5Mo-5V-1Cr-1Fe) with an α fiber (Ref 1). Some of the key findings were:

- Not surprisingly, oriented nucleation assuming a preponderance of $\Sigma 9$ nuclei gives rise to a noticeably different texture from the deformation texture after recrystallization is complete.
- The nature of the stored-energy distribution can have a strong effect on recrystallization kinetics. This behavior can be demonstrated using three cases, each assuming randomly oriented nuclei and the same initial total stored energy but with one of three different initial stored-energy distributions: (A) uniform within each grain, scaled by M_T, (B) energy gradient varying smoothly within each grain, and (C) energy gradient within each grain with a number of local maxima (Fig. 8b). The most rapid recrystallization rate is

found for case C due to the large number of recrystallization nuclei (Fig. 9). The recrystallization rate for case B is initially high due to the high level of stored energy near the grain boundaries but decreases substantially at long times; this behavior is accentuated in Avrami plots of the recrystallization kinetics (Fig. 9b).

- The mobility of special boundaries can have a pronounced effect on microstructure/texture evolution and recrystallization kinetics, as illustrated in Fig. 10. These dependencies can be illustrated using two simulations, both of which assumed oriented nucleation with nuclei probabilities, P, of $P(\Sigma 9) = 0.4$, $P(\Sigma 19) = 0.3$, $P(\Sigma 27) = 0.3$, and $P(random) = 0$. In one case, the special boundaries were given the same mobility (Fig. 10b), and in the other, the mobility was 40% higher than that of nonspecial high-angle boundaries (Fig. 10c). A noticeable difference in microstructure and recrystallized fraction after a time of 100 MCS is evident in simulation predictions for the two cases.
- Oriented nucleation has a strong effect on recrystallization texture, which in turn affects subsequent grain-growth kinetics, as discussed earlier in the section on the modeling of beta grain growth.

Recrystallization of Alpha Titanium

Significant progress has been made in modeling the recrystallization of alpha titanium. The principal focus in this area has been on the static recrystallization of cold- or warm-worked unalloyed (alpha) titanium.

Experimental Observations. Modeling and simulation of the static recrystallization of alpha titanium has been guided by detailed characterization of the deformed microstructure and experimental observations of recrystallization kinetics and so on. For both commercial-purity (CP) and high-purity unalloyed titanium, very inhomogeneous microstructures are developed as a result of the competition between crystallographic slip and twinning at cold and, to a lesser extent, warm working temperatures (Ref 13, 14, 40, 41). At cold working temperatures, for example, twinning is activated at very low strains ($\sim$0.05) and reaches a saturation level at effective strains of $\sim$0.3 to 0.6 in both uniaxial compression and rolling due to a Hall-Petch type of strengthening. Furthermore, as shown by Salem et al. (Ref 13) for a highly textured lot of pure titanium, the amount of twinning that is activated is heavily dependent on the imposed strain state. Imposed strains that require compressive or tensile strains along the c-axis of many of the grains in a polycrystalline aggregate lead to high-volume fractions of twins. Those strain states that require limited extension or compression along the c-axis result in the accommodation of the imposed deformation primarily by slip processes in the majority of the grains.

(a)

LIVE GRAPH
Click here to view

(b)

LIVE GRAPH
Click here to view
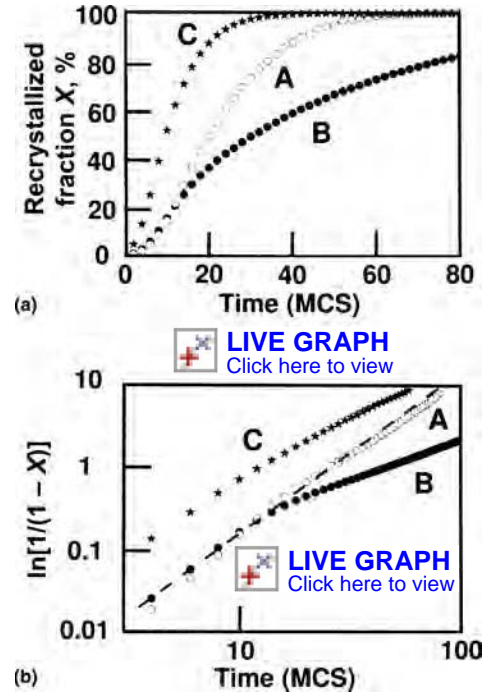
Fig. 9 Monte Carlo simulation predictions of the effect of initial stored-energy distribution on recrystallization, assuming randomly oriented nuclei. (a) Recrystallized fraction X. (b) Corresponding Avrami plots (Ref 34). The letters "A," "B," and "C" refer to the stored-energy distributions shown in Fig. 8(b). MCS, Monte Carlo steps.

Chun et al. (Ref 14, 40) have done extensive characterization of the microstructure developed during the cold rolling and recrystallization annealing of CP titanium (CP Ti) sheet. It was found that cold rolling to a moderate reduction ($\leq$40%) activates slip and mechanical twinning, primarily $\{11\bar{2}2\} < \bar{1}\bar{1}23 >$ compressive twins and $\{10\bar{1}2\} < \bar{1}011 >$ tensile twins (Fig. 11a). The formation of twins results in an inhomogeneous microstructure, in which regions containing twins are refined and the grains that deform primarily by slip remain relatively coarse (Fig. 11b). The occurrence of twinning can weaken the starting texture, leading to a moderate basal texture (Fig. 12). Above 40% reduction, during which deformation by slip predominates, a split-basal texture typical of cold-rolled hcp metals (with a principal ODF component of $\varphi_1 = 0°$, $\Phi = 35$ to $90°$, $\varphi_2 = 30°$) is developed.

During subsequent recrystallization annealing of CP Ti at temperatures of the order of 600 °C (1110 °F), twinned grains, which contain high stored energy and numerous high-angle boundaries, become the preferential sites for nucleation. Despite their impingement, the recrystallized grains grow quickly, consuming the neighboring deformed regions in short times ($\sim$15 min). The coarse, elongated remnant grains with low stored energy hinder the recrystallization process and require considerably longer times (of the order of 100 min) to be consumed. Although the overall recrystallization
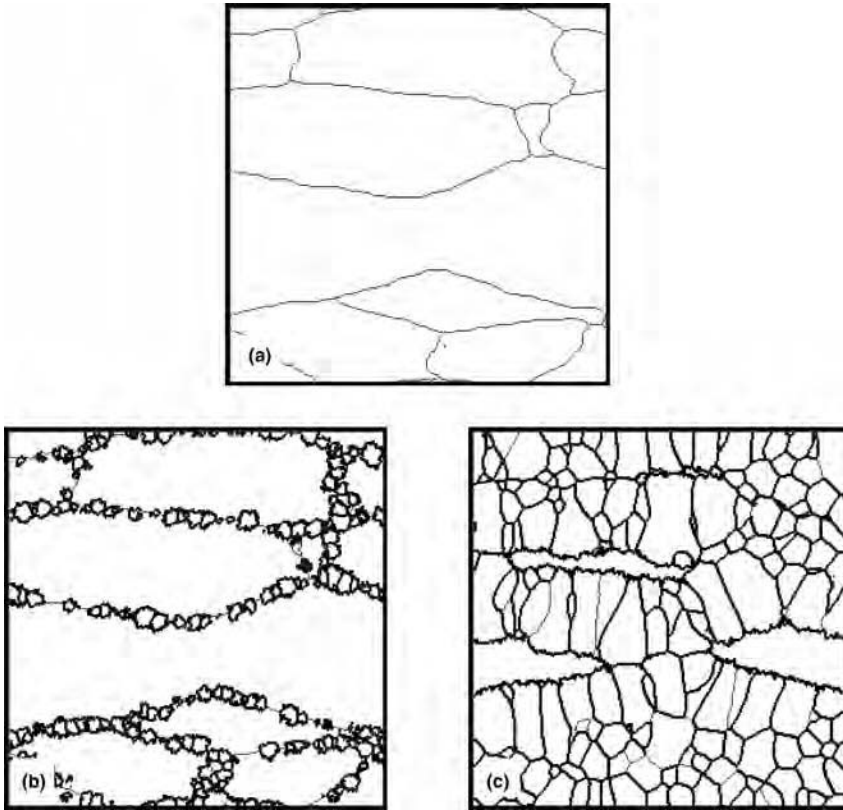
Fig. 10 Monte Carlo-predicted dependence of microstructure evolution. (a) Initially wrought material. (b) After 100 Monte Carlo steps (MCS), assuming identical nuclei orientations and a mobility of the special boundaries which was the same as that for nonspecial boundaries. (c) After 100 MCS with the same orientation and a mobility 40% higher than that of other nonspecial high-angle boundaries. Source: Ref 34

kinetics may exhibit a sigmoidal behavior, the corresponding Avrami plot is bilinear (Fig. 13a, b).

During recrystallization, the cold rolling texture diminishes in intensity, and a recrystallization-texture component ($\varphi_1 = 15°$, $\Phi = 35°$, $\varphi_2 = 35°$) appears. The recrystallization texture component strengthens during the subsequent grain growth stage as grains with this orientation become larger than average.

Formulation for Monte Carlo Simulations. Similar to the work of Ivasishin et al. (Ref 34) dealing with the recrystallization of beta titanium, Chun et al. (Ref 42) developed and applied a 2-D MC code to describe the recrystallization of cold-worked alpha titanium. The key components of this formulation consist of descriptions of (1) the stored energy associated with the deformed (cold-worked) state, (2) the decrease in stored energy during annealing due to static recovery, (3) nucleation behavior, and (4) the flip probability for recrystallization/migration of the recrystallization front.

The spatial distribution of cold work can be quantified by deformation models such as crystal-plasticity techniques (Ref 39, 43) or direct measurements based on image-quality maps derived from electron backscatter diffraction (EBSD) (Ref 44, 45). Because of the complexity of the microstructure developed during the cold deformation of CP Ti, the EBSD image-quality-map approach is preferable. The technique uses

EBSD Kikuchi-pattern bands whose sharpness is dependent on elastic strains that distort the crystal lattice. Chun et al. used the following relation to quantify this effect:

$$H_i = k\left(1 - \frac{IQ_i}{f \cdot IQ_{max}}\right) \qquad \text{(Eq 11)}$$

in which H_i and IQ_i denote the stored energy (in arbitrary units) and the image quality, respectively, at MC lattice site i; IQ_{max} is the maximum value of the image quality; and k and f are constants that specify the upper and lower bounds for the stored-energy distribution. In line with other measurements, Chun et al. chose $k = 70$ and $f = 1.3$, respectively, which resulted in stored-energy values ranging from 16 to 65. An example image-quality map for CP Ti rolled to a thickness reduction of 60% is shown in Fig. 11 (c). Regions with fine microstructure (developed in grains that underwent both twinning and slip) have high stored energy. The coarse remnants of the initial grain structure (which deformed only by slip) exhibit low values of stored energy.

Static recovery is described using the following relation:

$$H_{new} = (1 - b)H_{old} \qquad \text{(Eq 12)}$$

in which H_{old} and H_{new} are the stored energy before and after recovery, and b is a simulation

parameter. This relation quantifies the exponential decay of stored energy with time, which is typical of the static recovery of metals (Ref 46).

In the work of Chun et al. (Ref 42), nucleation was assumed to comprise site saturation; that is, all nuclei were assumed to be formed at the beginning of the annealing process. The nucleation sites were placed either randomly in the lattice or at sites with high stored energy. For the latter high-stored-energy nucleation case, 1000 lattice sites among those belonging to the upper 1% of stored energy were chosen. The stored energies of these nuclei were set to zero, and their orientations were taken to be the same as that for the lattice site in the deformed state.

The probability for recrystallization/migration of the recrystallization front can be treated in a manner analogous to that embodied in Eq 10. In the work of Chun et al. (Ref 42), however, an explicit expression for strain-induced migration of the recrystallization front between sites i and j as a function of the local variation in stored energy and the grain-boundary mobility was used:

$$v_{ij} = a_r \cdot \Delta H_{ij} \cdot M \qquad \text{(Eq 13)}$$

In this equation, v_{ij} denotes a velocity that varies between 0 and 1; a_r is a constant that ensures a realistic rate of recrystallization corresponding to the range of energy differences ΔH_{ij} between sites i and j (e.g., 0 to 65), and M is the mobility (as a function of misorientation), as given in Eq 9. When v_{ij} is greater than the probability obtained from a random-number generator that produces values between 0 and 1, the recrystallized grain grows at the expense of the deformed site j. As a consequence, the stored energy of site j is reset to zero, and the orientation of site j is changed to that of site i.

The calculation procedure in already recrystallized regions is similar to that for standard MC simulations of grain growth.

Monte Carlo Simulation Results. The MC simulations of Chun et al. (Ref 42) illustrate the effects of concurrent recovery, the spatial distribution of nuclei, the heterogeneous distribution of stored energy, and anisotropy of grain-boundary energy and mobility on static recrystallization kinetics and texture evolution and thus aid in the interpretation of observations for CP Ti. For the ideal case of no recovery, site saturation with randomly placed nuclei, uniformly distributed stored energy, and isotropic boundary properties, MC predictions reveal a linear Avrami plot with a slope p of ~2.15 (Fig. 13c), in good agreement with the classical Johnson-Mehl-Avrami-Kolmogorov (JMAK) prediction of 2.0.

A number of useful predictions can be obtained for special cases in which only one parameter (i.e., recovery, nuclei distribution, stored-energy distribution) is varied relative to the ideal case that mirrors JMAK behavior. When static recovery is included in MC

Fig. 11 Microstructures developed during cold rolling of commercially pure titanium to a thickness reduction of (a) 20% or (b) 60%. (c) An electron backscatter diffraction image-quality map for material cold rolled to a 60% reduction. The rolling direction is horizontal, and the sheet normal direction is vertical in all micrographs. Source: Ref 14, 40, 42

of the recrystallized fraction X, using the Cahn-Hagel relation (Ref 47):

$$v_{CH} = \frac{1}{S_v} \cdot \frac{dX}{dt} \qquad \text{(Eq 14)}$$

Here, S_v and t denote the total area of the recrystallization front per unit volume and the time, respectively. For 2-D MC simulations, S_v and t can be replaced by the total length of the recrystallization front per unit area, L_A, and the simulation time in MCS, respectively.

As examples, plots of V_{CH}, L_A, and the recrystallization rate ($\Delta X/\Delta \text{MCS}$) as a function of X are shown in Fig. 14. For the ideal case giving rise to linear JMAK kinetics (Fig. 14a), the recrystallization front moves at a nearly constant speed during the entire recrystallization process; the parameters L_A and $\Delta X/\Delta \text{MCS}$ exhibit maxima at $X \sim 0.42$. By contrast, for the simulation involving recovery, heterogeneous nucleation at nonuniformly distributed sites of high stored energy, and anisotropic grain-boundary properties (Fig. 14b), the front velocity increases rapidly during the very early stage of recrystallization and then decreases; L_A and $\Delta X/\Delta \text{MCS}$ both reach maxima at lower values of the recrystallized fraction (i.e., $X \sim 0.3$) than the classic JMAK case. This situation corresponds to the Avrami plot in Fig. 13(c). Thus, a negative deviation in recrystallization kinetics can be correlated to the onset of decreasing L_A at a lower value of X compared to that for JMAK kinetics or an average velocity of the recrystallization front that decreases with time.

The MC simulations also provide an invaluable tool to investigate the coupling of microstructure and texture evolution during recrystallization of a material such as CP Ti whose as-deformed condition is so complex. For instance, observed and simulated microstructures for recrystallization of CP titanium cold rolled to a thickness reduction of 60% and annealed at 600 °C (1100 °F) show excellent agreement (Fig. 15). These results illustrate the heterogeneity of recrystallization at short times in local regions with high stored energy and the subsequent grain growth in these areas during longer times at which the balance of the microstructure with lower initial stored energy is still undergoing recrystallization. Similar remarks apply to predictions of texture, quantified, for example, by ODFs (Fig. 16). For CP Ti, the as-deformed texture used for MC simulations (Fig. 16b, top) was obtained by EBSD over a limited area of observation (120 × 100 μm) and thus differs slightly from that obtained by x-ray diffraction (Fig. 16a, top); that is, the peak intensity of the x-ray diffraction analysis appeared at $\varphi_1 = 0°$, $\Phi = 35°$, and $\varphi_2 = 30°$, while that of the EBSD analysis was at $\varphi_1 = 0°$, $\Phi = 55°$, and $\varphi_2 = 30°$. However, both the experimental and MC-predicted recrystallization textures (Fig. 16a, bottom and 16b bottom, respectively) show a weakening of the $\varphi_1 = 0°$, $\Phi = 50$ to $\sim 90°$, and $\varphi_2 = 30°$

simulations, for example, Avrami curves show a negative deviation from the linear JMAK behavior. The deviation becomes more dramatic as the parameter b (Eq 12) is increased, accentuating the rate of recovery and thus the loss of the driving force for boundary migration associated with the stored energy. The distribution of the nuclei can also lead to deviations from the linear JMAK trend. When the nuclei are uniformly or heterogeneously distributed rather than at random locations, positive or negative deviations, respectively, are predicted. The range of stored energy (e.g., $15 < H < 65$, $20 < H < 86$, $50 < H < 92$) has little influence on recrystallization kinetics. Increasing

the average level of stored energy reduces the recrystallization time, but the distribution per se does not result in deviations from linear JMAK behavior. When these various factors and anisotropic grain-boundary properties are all incorporated into the MC simulation, Avrami plots that mirror experimental observations are obtained (e.g., Fig. 13b versus Fig. 13c).

The spatial and temporal information from MC simulations provides a useful tool for investigating the source of deviations from linear JMAK behavior. Such interpretations often involve microstructural-path analysis, for example, determination of the average velocity V_{CH} of the recrystallization front as a function

Fig. 12 Pole figures for commercially pure titanium cold rolled to percent thickness reductions of (a) 0%, (b) 40%, and (c) 60%. The maximum x-random texture intensities were (a) 4.4, (b) 3.9, and (c) 3.7. The contour levels for all pole figures are 1.5, 2.0, 2.5, ...7.5. RD, rolling direction; TD, transverse direction. Source: Ref 14

orientations and a strengthening of $\varphi_1 = 0$ to ~30°, $\Phi = 20$ to ~40°, and $\varphi_2 = 30°$ orientations. The strengthened texture component corresponds to the orientation of regions of high stored energy in the as-cold rolled CP Ti. This finding tends to support the MC-simulation assumption of identical orientations for both the deformed material and the nuclei and the importance of nucleation behavior in controlling recrystallization textures of alpha titanium.

Simulation of Deformation Texture Evolution

The simulation of deformation textures in titanium alloys can be performed using a number of different approaches, including the Taylor (isostrain) formulation; viscoplastic self-consistent (VPSC) analysis; and the crystal-plasticity, finite-element method (CPFEM).

Taylor Analysis

The Taylor analysis can be used for simulating the evolution of deformation texture for both single- and two-phase titanium alloys to various degrees of accuracy.

Single-Phase Beta Titanium. Glavicic et al. (Ref 19) and Gey et al. (Ref 48) demonstrated that a rate-sensitive Taylor-type model can be used to simulate beta-texture development during hot working of Ti-6Al-4V in the single-phase beta field. The former effort focused on texture evolution during the initial beta hot working breakdown operations for production-scale ingots. To this end, the local strains imparted to the workpiece were first estimated using a commercial continuum finite-element method (FEM) code, DEFORM (Scientific Forming Technologies Corporation, Columbus, OH). The strain increments so determined were inserted in the Los Alamos polycrystalline plasticity (LApp) code (Ref 49), assuming deformation via slip on {110}<111> and {112}<111> systems. By this means, a <110> fiber texture was predicted for the billet center at which the initial (solidification) texture had been random. Model predictions showed good agreement with the experimental beta texture, whose determination required measurements of the alpha-phase texture due to the limited amount of beta phase for Ti-6Al-4V at room temperature (Ref 50, 51).

Gey et al. (Ref 48) analyzed the rolling of Ti-6Al-4V plate in the beta phase field. They assumed the same slip systems as in the work of Glavicic et al. (Ref 19) but chose a relaxed-constraints approach. That is to say, two of the three shear strains were not forced to be identical in all grains because of the pancake shape of the grains developed during rolling to heavy reductions. The simulated texture for a 75% thickness reduction showed relatively good agreement with measurements (Fig. 17). The major difference lay in the absolute intensity of texture components; the simulations showed stronger maxima compared to the observations. Such differences are common for crystal-plasticity models that typically neglect accommodation of the imposed deformation via dynamic-recovery/dislocation-climb processes or heterogeneous strain in different regions of the same grain.

Two-Phase Titanium Alloys. Taylor-type models can also be used to estimate the deformation textures developed in the individual phases during the hot working of two-phase titanium alloys with an equiaxed alpha microstructure, such as occurs during alpha/beta forging or rolling of alpha/beta titanium alloys. One such approach is that developed by Glavicic et al. (Ref 52) for the prediction of texture evolution during the hot pancake forging of Ti-6Al-4V. In this approach, a continuum FEM code is used first to estimate the macroscopic strains/strain increments and rigid-body rotations at each point of the workpiece. An approximate analysis is then used to estimate the partitioning of strain between the (harder) alpha-phase particles and the (softer)
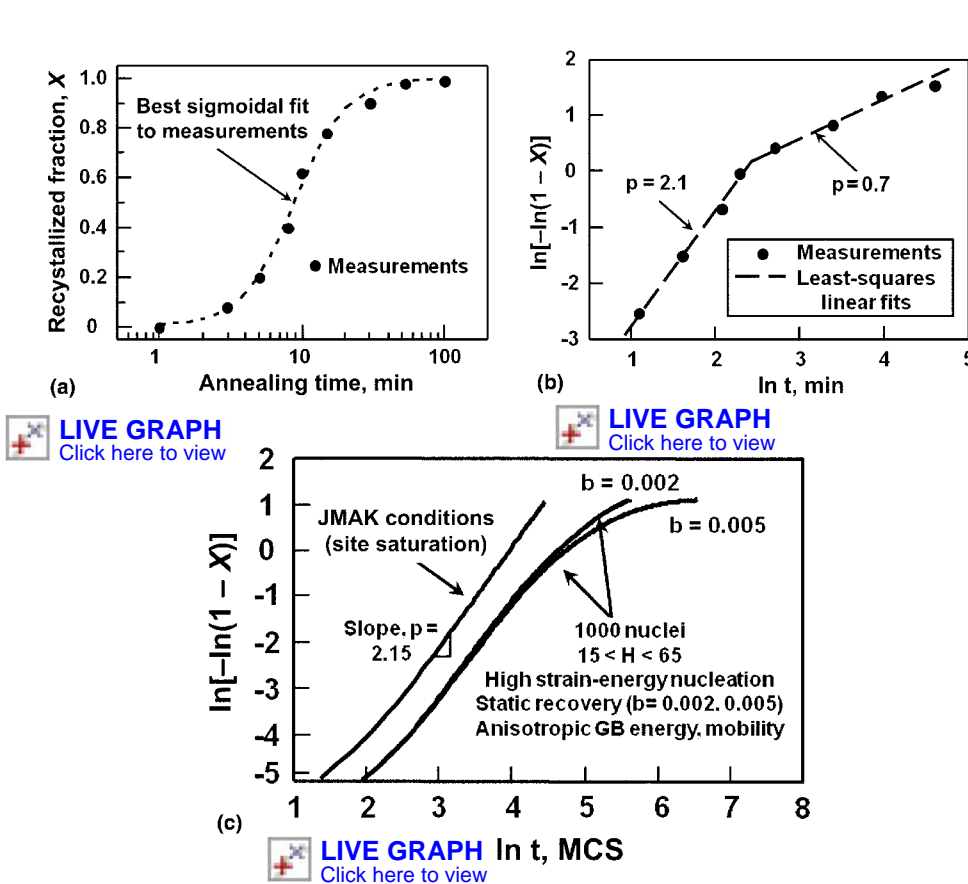
Fig. 13 Static recrystallization behavior of commercially pure titanium rolled to a thickness reduction of 60% and annealed at 600 °C (1100 °F). (a) Measured recrystallization kinetics. (b) Corresponding experimental Avrami plot. (c) Predicted kinetics from a Monte Carlo simulation; see text for details. JMAK, Johnson-Mehl-Avrami-Kolmogorov; GB, grain boundary; MCS, Monte Carlo steps. Source: Ref 42

Fig. 14 Monte Carlo predictions of the dependence on recrystallized fraction X of the average velocity (V_{CH}) and total length per unit area (L_A) of the recrystallization front and the rate of recrystallization ($\Delta X/\Delta$MCS). (a) Classical Johnson-Mehl-Avrami-Kolmogorov (JMAK) condition. (b) A condition involving recovery, heterogeneous nucleation at sites of high stored energy, and anisotropic grain-boundary (GB) properties. MCS, Monte Carlo steps. Source: Ref 42

beta-phase matrix grains at specific points of interest. Next, a rate-sensitive Taylor analysis is applied to estimate the rotations associated with slip within each separate phase. Last, the rotations due to rigid-body motion (metal flow) and crystallographic slip are summed to determine the final texture of the alpha and beta phases.

To partition the strain between the alpha and beta phases, Glavicic et al. (Ref 52) used an approximate self-consistent analysis for two-phase materials, each of whose constituents has the same strain-rate sensitivity, m, and a constitutive response of the form $\sigma_i = k_i \dot{\varepsilon}_i^m$; $\sigma(\dot{\varepsilon})$ and k_i denote the flow stress as a function of strain rate and the strength coefficient for the specific phase, respectively. The model was based on the approach developed by Hill (Ref 53) for linearly elastic solids, which was later extended to the case of rate-sensitive, incompressible materials by Suquet (Ref 54). Subsequently, it was applied to titanium alloys by Briottet et al. (Ref 55) and Semiatin et al. (Ref 56). The analysis leads to the determination of an effective strength coefficient, k, for the two-phase aggregate, whose overall constitutive response is assumed to be $\sigma = k\dot{\varepsilon}^m$. The strength coefficient k depends on the ratio of the strength coefficients of the two phases, k_a/k_β, and the volume fraction of the harder phase (alpha for the case of alpha/beta titanium

alloys), that is, f_α. Results of the analysis for $m = 0.23$ (i.e., the m for the alpha and beta phases of Ti-6Al-4V) are shown in Fig. 18(a). Results for $m = 0.15$ and $m = 0.30$ are similar.

The values of k_α and k_β depend on phase composition and temperature (Ref 56, 57). An analysis of a large collection of flow stress data for various alloys has led to the finding that the ratio k_α/k_β for the alpha and beta phases in Ti-6Al-4V is almost constant at temperatures between 815 and 982 °C (1500 and 1800 °F) and has a value of 3 (Ref 56).

The average strain rates in the alpha and beta phases (and hence the strain partitioning between the phases) are readily calculated from the values of k, k_α, k_β, and f_α by noting that the aggregate flow stress and strain rate are volume averages of the corresponding quantities for the components (Ref 58). This leads to the following expressions for $\dot{\varepsilon}_\alpha/\dot{\varepsilon}_{ov}$ and $\dot{\varepsilon}_\beta/\dot{\varepsilon}_{ov}$, in which $\dot{\varepsilon}_{ov}$ is the overall (aggregate) strain rate:

$$k/k_\alpha = f_\alpha(\dot{\varepsilon}_\alpha/\dot{\varepsilon}_{ov})^m + [(1-f_\alpha)^{(1-m)}(k_\beta/k_\alpha)\{1-f_\alpha(\dot{\varepsilon}_\alpha/\dot{\varepsilon}_{ov})\}^m] \quad \text{(Eq 15a)}$$

$$\dot{\varepsilon}_\beta/\dot{\varepsilon}_{ov} = [1-f_\alpha(\dot{\varepsilon}_\alpha/\dot{\varepsilon}_{ov})]/(1-f_\alpha) \quad \text{(Eq 15b)}$$

Equation 15(a) cannot be solved analytically but is readily evaluated using numerical

techniques, leading to nomographs such as those in Fig. 18(b, c).

Using this approach for strain partitioning, Glavicic et al. (Ref 52) applied the rate-sensitive Taylor code LApp to predict the deformation texture of the primary alpha phase at various locations in a pancake of Ti-6Al-4V forged at 955 °C (1750 °F), which is ~40 °C (~72 °F) below the beta-transus temperature. The predictions showed reasonable qualitative agreement with measurements (Fig. 19). The angular difference in the location of the texture maxima may be ascribed to small errors in the metal-flow part of the texture prediction. In addition, the magnitudes of the predicted texture components were high, a trend similar to that discussed previously with regard to the work of Gey et al. (Ref 48).

VPSC and CPFEM Analyses

Viscoplastic self-consistent (VPSC) and crystal-plasticity FEM (CPFEM) approaches can also be used to model the evolution of deformation texture in two-phase titanium alloys. In the former approach, the strain is assumed to be homogeneous with each grain but may vary from one grain/phase to another. The CPFEM technique enables the investigation of strain variations within grains/phases as well as from grain to grain. Details on these simulation techniques are contained in other

Fig. 15 Comparison of microstructure evolution during recrystallization of commercially pure titanium cold rolled to a 60% thickness reduction and then annealed at 600 °C (1110 °F). (a) Experimental observations. (b) Monte Carlo predictions. MCS, Monte Carlo steps. Source: Ref 42

alpha and beta phases were the same in each case. The evolution of deformation texture was simulated using the VPSC approach of Lebensohn and Tome (Ref 60). Using a fitting procedure to get the best deformation-texture predictions for single-phase alloys, the critical resolved shear stresses (CRSS) for slip were taken to be in the ratio of 1:1:1.5:3:1/3:1/3:1/3 for the four assumed alpha-phase systems (prism $<a>$, pyramidal $<a>$, basal $<a>$, pyramidal $<c + a>$) and the three assumed beta-phase systems ($\{110\}<111>$, $\{112\}<111>$, and $\{123\}<111>$). The strain-rate sensitivity (m) was set at 0.2 for both phases.

Alpha-phase rolling textures predicted by Dunst and Mecking using the VPSC approach showed good agreement with measurements (Fig. 20). The calculated activity of different slip systems and microstructure observations revealed that hard alpha-grain orientations (i.e., those whose c-axes lay in the plate normal direction) tended to remain undeformed as the volume fraction of alpha decreased. On the other hand, soft alpha grains deformed predominantly by single slip (along prism $<a>$ systems) as the alpha volume fraction increased; the strain incompatibility was accommodated by the deformation of the beta phase. Predicted beta-phase textures showed poorer agreement with measurements, an effect ascribed to local interactions between the beta and alpha phases that were not treated in the VPSC model.

Deformation-texture development in alpha/beta titanium alloys with a colony-alpha microstructure can also be described using a VPSC approach (Ref 61). The best alpha-phase predictions are obtained when the correlation between alpha lamellae and the beta matrix (i.e., the Burgers relation between the phases and an alpha-beta interface close to a prism plane in the alpha phase) is taken into consideration. Local interactions and nonuniform local flow of the beta phase make the prediction of beta deformation textures somewhat problematical, as in VPSC simulations for equiaxed microstructures, however.

Crystal-plasticity FEM simulations can also be used to quantify strain partitioning and texture evolution in two-phase titanium alloys. For example, Turner and Semiatin (Ref 62) examined the effect of the degree of microstructure discretization on strain partitioning between a hard and a soft phase. Predictions were found to differ noticeably for so-called meso-scale and micro-scale modeling approaches. For the former, each grain is represented by a single element in the CPFEM simulation, whereas each grain is divided into multiple elements for micro-scale simulations. The higher level of discretization produces predictions of strain partitioning similar to those from the approximate self-consistent model of Suquet, Semiatin, and their co-workers (Ref 54, 56) (Fig. 21a).

The usefulness of CPFEM to quantify local deformation and texture evolution for two-phase alpha/beta titanium alloys has been

articles in this Volume. Hence, only some key results of such analyses are summarized here.

The VPSC analysis has been applied to alpha/beta titanium alloys with an equiaxed alpha or a lamellar colony microstructure. In the work of Dunst and Mecking (Ref 59),

for example, texture evolution during the rolling of plate to a 70% thickness reduction in a single pass at 800 °C (1470 °F) was investigated for a series of equiaxed alpha alloys with different volume fractions of the two phases. Alloying was chosen so that the compositions of the

Fig. 16 Comparison of textures developed during recrystallization of commercially pure titanium cold rolled to a 60% thickness reduction and then annealed at 600 °C (1110 °F). (a) Measured. (b) Monte Carlo prediction. In both (a) and (b), the orientation distribution function on the top is for the as-deformed condition, and that on the bottom is for the fully recrystallized condition. Source: Ref 42

Fig. 17 Comparison of (a) measured and (b) predicted (100) beta-phase pole figures for Ti-6Al-4V plate hot rolled to a 75% thickness reduction in the beta phase field. RD, rolling direction. Source: Ref 48

demonstrated by Barton and Dawson (Ref 63). In this work, the hot rolling of plate in which the volume fractions of the alpha and beta phases were varied was simulated; because of excessive mesh distortions, the thickness reduction was limited to 25%. The values of CRSS were similar to those used in the VPSC simulations of Dunst and Mecking (Ref 59). As for other modeling approaches, the partitioning of strain between the alpha and beta phases is predicted by CPFEM to increase with a decrease in the volume fraction of the harder alpha phase (Fig. 21b). Furthermore, nonnegligible shear strain components can be quantified in CPFEM simulations. In the work of Barton and Dawson (Ref 63), it was shown that the shear strains tend to be higher in the beta phase and increase in magnitude with an increase in the volume fraction of alpha, thus underlining the need to quantify local neighborhood effects on texture evolution in the beta phase. In a similar vein, CPFEM simulations reveal that strain localization tends to occur in materials with roughly equal amounts of alpha and beta or slightly

beta-rich alloys. In titanium alloys with a large volume fraction of beta, on the other hand, the alpha particles tend to undergo little deformation (but may experience noticeable rotations), and thus, they are essentially carried along with the plastic flow of the beta. Textures predicted by CPFEM mirror measured ones, but limitations on the reduction level that can be simulated (due to excessive mesh distortion) preclude quantitative comparisons. Current and future increases in computing power are now overcoming such limitations for CPFEM simulations of texture evolution.

Transformation Texture Evolution

The decomposition of the high-temperature metastable bcc beta phase during cooling often results in the preferential selection of a subset of the 12 possible hcp lamellar-alpha variants. This effect has been investigated most often for alpha/beta titanium alloys such as Ti-6Al-4V, and various approaches have been proposed
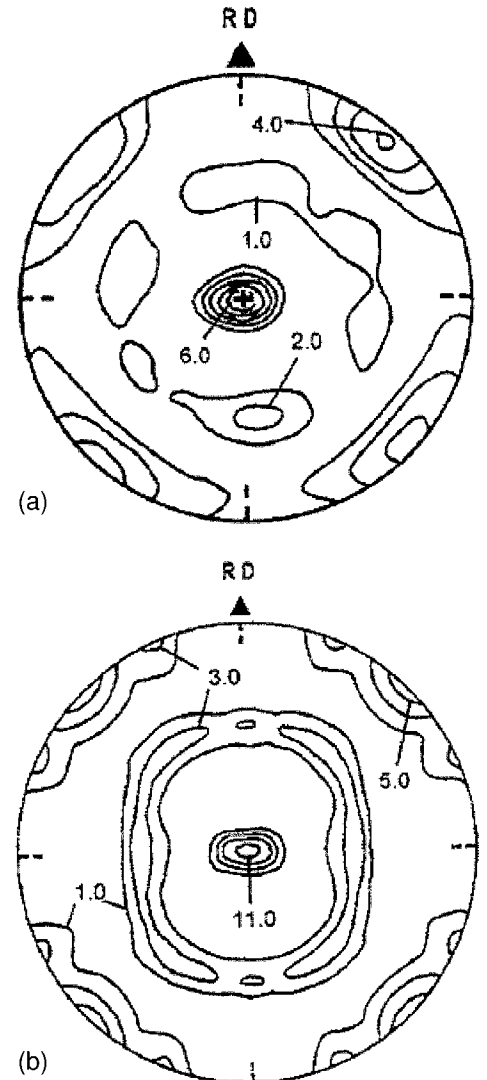
to interpret observations for both heat treatment and deformation processes.

In the area of heat treatment, it has been found that variant selection tends to be weak for cold- or hot-rolled Ti-6Al-4V plate that is beta annealed and slow cooled (Ref 64, 65). By contrast, a strong variant-selection process can occur for this material during rapid cooling from a heat treatment in the alpha + beta phase field (Ref 64). For example, an analysis of the data of Moustahfid et al. (Ref 64) by Divinski et al. (Ref 66) suggests that the orientation of the beta-phase plane relative to the rolling plane controls the variants that are formed. In particular, good agreement between measured and predicted pole figures (Fig. 22) is obtained when only one-half of the possible variants are chosen for each of the two principal beta-texture components ($\{112\} < 1\bar{1}0 >$ and $\{111\} < 11\bar{2} >$, in which the two indices refer to the
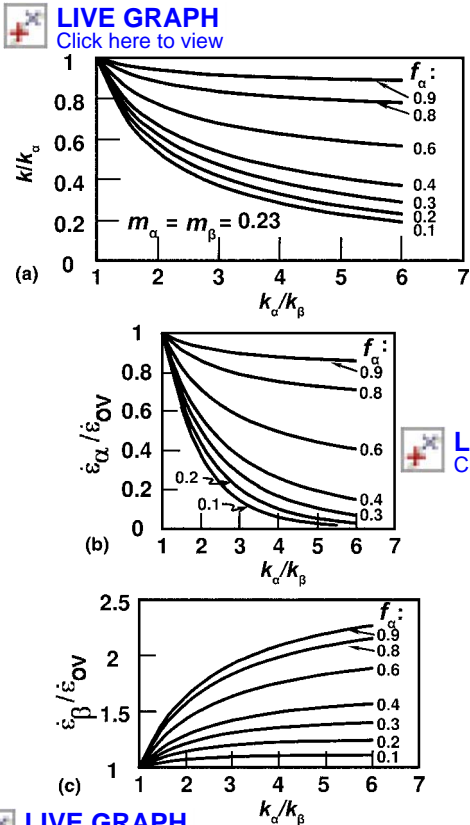
Fig. 18 Predictions from an approximate self-consistent model for strain partitioning in a two-phase material with a rate sensitivity of 0.23. The graphs show the dependence on the ratio k_α/k_β and the volume fraction of the harder phase f_α (a) Aggregate strength coefficient k relative to k_α (b) Strain rate in the harder phase ($\dot{\varepsilon}_\alpha$). (c) Strain rate in the softer phase ($\dot{\varepsilon}_\beta$). Both relative to the overall (aggregate) strain rate ($\dot{\varepsilon}_{ov}$). Source: Ref 56
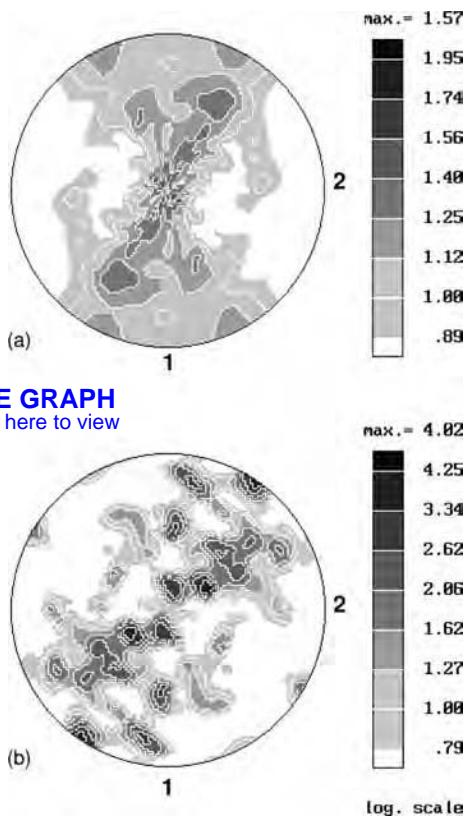
Fig. 19 Comparison of the deformation texture (in terms of (0001) pole figures) developed in the primary alpha phase at the corner location of a Ti-6Al-4V billet that was pancake forged at 955 °C (1550 °F). (a) Measured. (b) Predicted. Source: Ref 52

Fig. 21 Strain partitioning in two-phase alloys. (a) Comparison of crystal-plasticity finite element method (CPFEM) meso-scale and micro-scale model predictions with those from a self-consistent (SC) model. (b) CPFEM predictions of the strain-rate components, *ij*, in the alpha and beta phases during rolling of titanium plate with two different volume fraction ratios. *i* or *j*: 1, normal direction; 2, transverse direction; 3, rolling direction. Source: (a) Ref 62, (b) Ref 63

Fig. 20 Comparison of measured (a-d) and simulated (e-h) alpha-phase (0001) pole figures for heavily rolled titanium alloys having alpha/beta volume fractions (in %) of (a) and (e) 96/4, (b) and (f) 78/22, (c) and (g) 36/64, and (d) and (h) 18/82. Source: Ref 59

crystallographic plane/direction parallel to the rolling plane/rolling direction). It has been hypothesized that local stresses developed in two-phase material during cooling (as a result of differences in coefficients of thermal expansion) tend to favor the formation of alpha variants whose orientation is similar to that of the primary alpha (Ref 66). In later work by Humbert and Gey (Ref 67), such biasing of the texture of the secondary alpha by that of the primary alpha was modeled based on a minimum strain-energy hypothesis. In yet other work, Zeng and Bieler (Ref 68) have postulated that the anisotropic contraction of primary alpha during cooling may activate slip on selected {110} slip planes in the beta phase, thus biasing the secondary-alpha variants that are formed.

When beta-phase decomposition follows hot deformation in the beta field, noticeable preferred variant selection tends to occur (Ref 48). In such cases, the selected alpha variants tend to correlate with the beta slip systems that have been most active during the prior deformation. The plausibility of this approach was demonstrated by Gey et al. (Ref 69) for beta hot working via plate rolling. As described in the section on the modeling of deformation texture, the relative slip activity in the beta phase can be estimated using a Taylor-type model assuming deformation on {110}<111> and {112}<111>

(a)

(b)

(c)
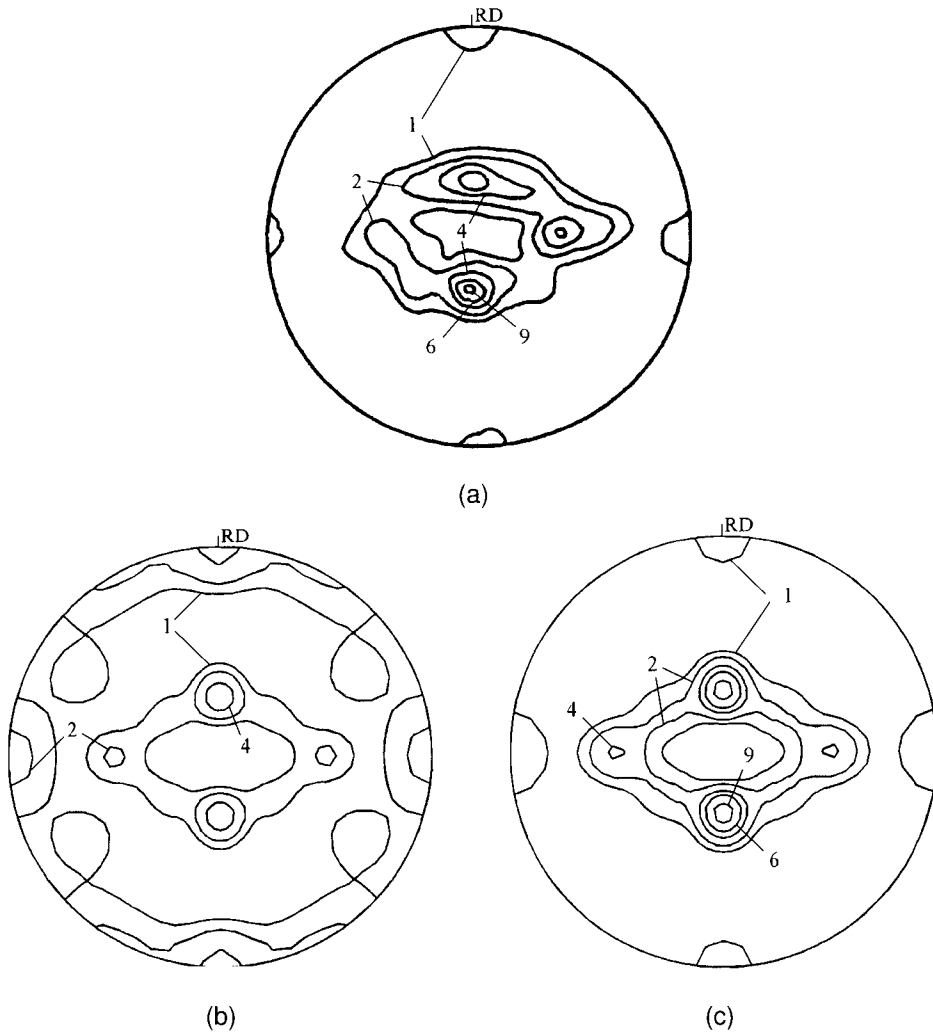
Fig. 22 Comparison of (0001) alpha-phase pole figures for Ti-6Al-4V plate that was heat treated high in the alpha + beta phase field and then rapidly cooled. (a) Measured. (b) Predicted assuming random variant selection. (c) Predicted assuming biased variant selection. RD, rolling direction. Source: (a) Ref 64, (b) and (c) Ref 66

(a)

(b)

Fig. 23 Comparison of (0001) pole figures for colony alpha formed during cooling following hot rolling of Ti-6Al-4V plate to a 75% thickness reduction in the beta-phase field. (a) Measured. (b) Predicted using a Taylor-type crystal-plasticity model and alpha-variant selection based on relative slip-system activity. RD, rolling direction. Source: Ref 69

systems. Alpha variants may be presumed to nucleate on highly active {110}<111> beta-phase systems because of the direct Burgers orientation relation between the close-packed planes and directions in the two phases. In addition, Burgers (Ref 9) suggested that the beta-to-alpha transformation begins with shear movements of atoms on {112} planes in <111> directions. Thus, prior activity of specific {112}<111> systems may also favor the formation of related alpha variants. In the specific work of Gey et al. (Ref 69), therefore, the volume fractions of alpha variants for each beta crystal orientation were assigned in direct proportion to the relative activity on those slip systems that underwent a certain minimum shear strain; the minimum was taken to be 50% of the maximum shear strain for the particular crystal orientation. With a variant-selection rule of this form, texture predictions in good agreement with observation can be made (Fig. 23).

An alternate but related variant-selection rule for the decomposition of hot-worked beta has been proposed and validated by Moustahfid et al. (Ref 70). This alternate rule chooses among the various possible {110}<111> and {112}<111> slip systems for each crystal orientation based on their relative Schmid factors/resolved shear stresses. Those systems whose Schmid factors are within 10% of the maximum Schmid factor are assumed to form an alpha variant.

Future Outlook

The current status of the modeling and simulation of the evolution of crystallographic texture in titanium and titanium alloys due to recrystallization, grain growth, deformation, and phase transformation has been summarized. For these materials, the development of a fundamental understanding of texture formation tends to be challenging from an experimental standpoint because of the high temperatures involved in typical industrial processes and the
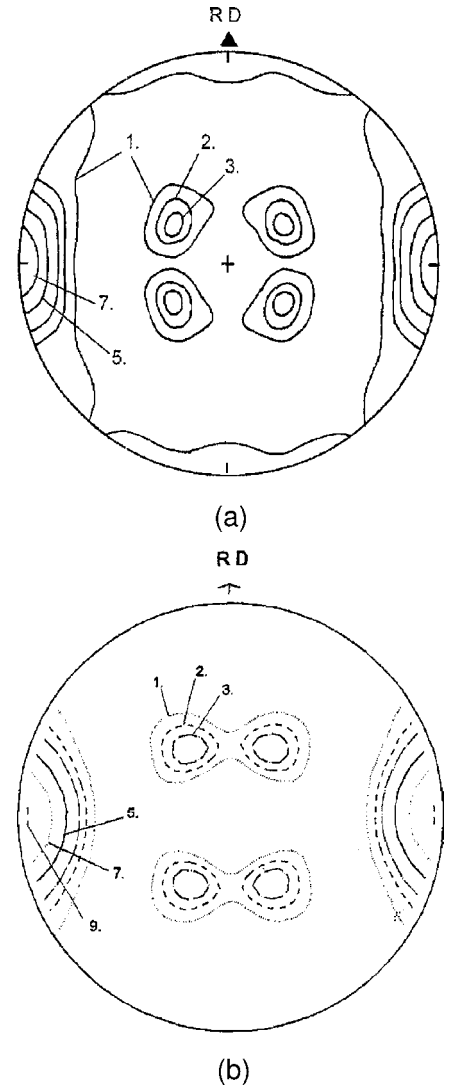
allotropic transformation of many alloys of commercial interest. Hence, modeling and simulation provide useful tools to investigate phenomena that are not easily observed. Recent developments in this area underscore the coupled nature of the evolution of texture and microstructure and the need to treat both aspects in realistic physics-based models. Important areas of future development include the following:

- Development of databases of material properties needed for simulations. These properties include grain-boundary energy and mobility and stress-strain/strain-rate constitutive behavior at large strains for slip in both alpha and beta titanium.

- Development of an understanding of the effect of strain path on large-strain constitutive behavior at the crystal scale
- Incorporation of recovery in deformation-texture models
- Development of variant-selection rules to describe the decomposition of beta during cooling following heat treatment or hot working, especially for alpha/beta titanium alloys processed in the two-phase field

ACKNOWLEDGMENTS

REFERENCES

1. U.F. Kocks, C.N. Tome, and H.R. Wenk, *Texture and Anisotropy*, Cambridge University Press, Cambridge, UK, 1998.
2. C.N. Tome and A.J. Beaudoin, "Polycrystal Modeling, Plastic Forming, and Deformation Textures," *ASM Handbook, Volume 14A, Metalworking: Bulk Forming*, ASM International, Materials Park, OH, 2005, pp. 671–684.
3. L. Kestens and J.J. Jonas, "Transformation and Recrystallization Textures Associated with Steel Processing," *ASM Handbook, Volume 14A, Metalworking: Bulk Forming*, ASM International, Materials Park, OH, 2005, pp. 685–700.
4. F. Larson and A. Zarkades, "Properties of Textured Titanium Alloys," Report MCIC-74-20, Metals and Ceramics Information Center, Battelle's Columbus Laboratories, Columbus, OH, 1974.
5. G. Lütjering and M. Peters, "Mechanical Properties of a Titanium Blading Alloy," Report EPRI-CS-2933, Electric Power Research Institute, Palo Alto, CA, 1983.
6. G. Lütjering and J.C. Williams, *Titanium*, Springer Verlag, Berlin, 2007.
7. I. Weiss and S.L. Semiatin, "Thermomechanical Processing of Beta Titanium Alloys—An Overview," *Mat. Sci. Eng. A*, 1998, Vol A243, pp. 46–65.
8. I. Weiss and S.L. Semiatin, "Thermomechanical Processing of Alpha Titanium Alloys—An Overview," *Mater. Sci. Eng. A*, 1999, Vol A263, pp. 243–256.
9. W.G. Burgers, "On the Process of Transition of the BCC Modification into the HCP Modification of Zirconium," *Physica*, 1934, Vol 1, pp. 561–586.
10. B. Hutchinson and B. Wynne, "Effects of Stress on Structure Transformations in Metals," *Mater. Sci. Forum*, 2007, Vol 550, pp. 149–158.

11. V. Randle and O. Engler, *Introduction to Texture Analysis: Macrotexture, Microtexture, and Mapping*, Taylor & Francis Ltd., London, 2000.
12. H.J. Bunge, *Texture Analysis in Materials Science*, Butterworths, Boston, 1982.
13. A.A. Salem, S.R. Kalidindi, and R.D. Doherty, "Strain Hardening of Titanium: Role of Deformation Twinning," *Acta Mater.*, 2003, Vol 51, pp. 4225–4237.
14. Y.B. Chun, S.H. Yu, S.L. Semiatin, and S.K. Hwang, "Effect of Deformation Twinning on Microstructure and Texture Evolution during Cold Rolling of CP Titanium," *Mater. Sci. Eng. A*, 2005, Vol A398, pp. 209–219.
15. S.L. Semiatin and T.R. Bieler, "Effect of Texture and Slip Mode on the Anisotropy of Plastic Flow and Flow Softening during Hot Working of Ti-6Al-4V," *Metall. Mater. Trans. A*, 2001, Vol 32A, pp. 1787–1799.
16. H.R. Piehler and W.A. Backofen, "A Theoretical Examination of the Plastic Properties of BCC Crystals Deforming by <111> Pencil Glide," *Metall. Trans.*, 1971, Vol 2, pp. 249–255.
17. P.R. Morris and S.L. Semiatin, "The Prediction of Plastic Properties of Polycrystalline Aggregates of BCC Metals Deforming by <111> Pencil Glide," *Texture of Crystalline Solids*, 1979, Vol 3, pp. 113–126.
18. G.Y. Chin and W.L. Mammel, "Computer Solutions of the Taylor Analysis for Axisymmetric Flow," *Trans. AIME*, 1967, Vol 239, pp. 1400–1405
19. M.G. Glavicic, P.A. Kobryn, R.L. Goetz, K.O. Yu, and S.L. Semiatin, "Texture Evolution during Primary Processing of Production-Scale Vacuum-Arc-Remelted Ingots of Ti-6Al-4V," *Ti-2003: Science and Technology*, G. Lütjering and J. Albrecht, Eds., Wiley-VCH GmbH, Weinheim, Germany, 2004, pp. 1299–1306.
20. W. Pichl, "Slip Geometry and Plastic Anisotropy of Body-Centered Cubic Metals," *Physica Status Solidi (a)*, 2002, Vol 189, pp. 5–25.
21. S.P. Fox, "A Study of Grain-Growth Behavior in Titanium Alloys," *Titanium '92: Science and Technology*, F.H. Froes and I.L. Caplan, Eds., TMS, Warrendale, PA, 1992, pp. 769–776.
22. O.M. Ivasishin, S.V. Shevchenko, and S.L. Semiatin, "Effect of Crystallographic Texture on the Isothermal Beta Grain-Growth Kinetics of Ti-6Al-4V," *Mater. Sci. Eng. A*, 2002, Vol A332, pp. 343–350.
23. O.M. Ivasishin, S.L. Semiatin, P.E. Markovsky, S.V. Shevchenko, and S.V. Ulshin, "Grain Growth and Texture Evolution in Ti-6Al-4V during Beta Annealing under Continuous Heating Conditions," *Mater. Sci. Eng. A*, 2002, Vol A337, pp. 88–96.
24. O.M. Ivasishin, S.V. Shevchenko, P.E. Markovsky, and S.L. Semiatin,

"Experimental Investigation and 3D Monte-Carlo Simulation of Texture-Controlled Grain Growth in Titanium Alloys," *Ti-2003: Science and Technology*, G. Lütjering and J. Albrecht, Eds., Wiley-VCH GmbH, Weinheim, 2004, pp. 1307–1314.
25. O.M. Ivasishin, S.V. Shevchenko, N.L. Vasiliev, and S.L. Semiatin, "3D Monte-Carlo Simulation of Texture-Controlled Grain Growth," *Acta Mater.*, 2003, Vol 51, pp. 1019–1034.
26. G. Abbruzzese and K. Lücke, "A Theory of Texture Controlled Grain Growth—Derivation and General Discussion of the Model," *Acta Metall.*, 1986, Vol 34, 1986, pp. 905–914.
27. H. Eichelkraut, G. Abbruzzese, and K. Lücke, "A Theory of Texture-Controlled Grain Growth——II. Numerical and Analytical Treatment of Grain Growth in the Presence of Two Texture Components," *Acta Metall.*, 1988, Vol 36, pp. 55–68.
28. O.M. Ivasishin, S.V. Shevchenko, and S.L. Semiatin, "Implementation of Exact Grain-Boundary Geometry into a 3D Monte-Carlo (Potts) Model for Microstructure," *Acta. Mater.*, Vol 57, 2009, pp. 2834–2844.
29. G. Gottstein and L.S. Shvindlerman, *Grain-Boundary Migration in Metals*, CRC Press, Boca Raton, Florida, 1999.
30. Y. Amouyal, E. Rabkin, and Y. Mishin, "Correlation between Grain-Boundary Energy and Geometry in Ni-Rich NiAl," *Acta Mater.*, 2005, Vol 53, pp. 3795–3805.
31. N. Ma and Y. Wang, "Beta Grain-Growth Kinetics in Ti-6Al-4V," *Numiform 2004*, S. Ghosh, J.M. Castro, and J.K. Lee, Eds., American Institute of Physics, College Park, Maryland, 2004, pp. 1700–1704
32. N. Ma, A. Kazaryan, S.A. Dregia, and Y. Wang, "Computer Simulation of Texture Evolution during Grain Growth: Effect of Boundary Properties and Initial Microstructure," *Acta Mater.*, 2004, Vol 52, pp. 3869–3879.
33. I. Weiss and F.H. Froes, "The Processing Window for the Near Beta Ti-10V-2Fe-3Al Alloy," *Titanium: Science and Technology*, G. Lütjering, U. Zwicker, and W. Bunk, Eds., Deutsche Gesellschaft für Metallkunde e.V., Oberursel, Germany, 1985, pp. 499–506.
34. O.M. Ivasishin, S.V. Shevchenko, N.L. Vasiliev, and S.L. Semiatin, "A 3-D Monte-Carlo (Potts) Model for Recrystallization and Grain Growth in Polycrystalline Materials," *Mater. Sci. Eng. A*, 2006, Vol A433, pp. 216–232.
35. R. Ding and Z.X. Guo, "Microstructural Modelling of Dynamic Recrystallization Using an Extended Cellular Automaton Approach," *Comp. Mater. Sci.*, 2002, Vol 23, pp. 209–218.
36. R. Ding and Z.X. Guo, "Microstructural Evolution of a Ti-6Al-4V Alloy during Beta-Phase Processing: Experimental and

Simulative Investigations," *Mater. Sci. Eng. A*, 2004, Vol A365, pp. 172–179.

37. A.D. Rollett and E.A. Holm, "Abnormal Grain Growth—The Origin of Recrystallization Nuclei?," *ReX'96, Proceedings of the Third International Conference on Recrystallization and Related Phenomena*, T.R. McNelley, Ed., Monterey Institute of Advanced Studies, Monterey, CA, 1997, pp. 31–42.

38. R.D. Doherty, D.A. Hughes, F.J. Humphreys, J.J. Jonas, D. Jul Jensen, M.E. Kassner, W.E. King, T.R. McNelly, H.J. McQueen, and A.D. Rollett, "Current Issues in Recrystallization: A Review," *Mater. Sci. Eng. A*, 1997, Vol A238, pp. 219–274.

39. B. Radhakrishnan, G.B. Sarma, and T. Zacharia, "Modeling the Kinetics and Microstructure Evolution during Static Recrystallization—Monte-Carlo Simulation of Recrystallization," *Acta Mater.*, 1998, Vol 46, pp. 4415–4433.

40. Y.B. Chun, S.L. Semiatin, and S.K. Hwang, "Role of Deformation Twinning in Cold Rolling and Recrystallization of Titanium," *Mater. Sci. Forum*, 2005, Vol 495–497, pp. 651–656.

41. Y.B. Chun and S.K. Hwang, "Static Recrystallization of Warm-Rolled Pure Titanium Influenced by Microstructural Inhomogeneity," *Acta Mater.*, 2008, Vol 56, pp. 369–379.

42. Y.B. Chun, S.L. Semiatin, and S.K. Hwang, "Monte-Carlo Modeling of Microstructure Evolution during the Static Recrystallization of Cold-Rolled, Commercial-Purity Titanium," *Acta Mater.*, 2006, Vol 54, pp. 3673–3689.

43. D.E. Solas, C.N. Tomé, O. Engler, and H.R. Wenk, "Deformation and Recrystallization of Hexagonal Metals: Modeling and Experimental Results for Zinc," *Acta Mater.*, 2001, Vol 49, pp. 3791–3801.

44. F. Caleyo, T. Baudin, and R. Penelle, "Monte-Carlo Simulation of Recrystallization in Fe-50 Pct. Ni Starting from EBSD and Bulk Texture Measurements," *Scripta Mater.*, 2002, Vol 46, pp. 829–835.

45. S.H. Choi and J.H. Cho, "Primary Recrystallization Modeling for Interstitial-Free Steels," *Mater. Sci. Eng. A*, 2005, Vol A405, pp. 86–101.

46. F.J. Humphreys and M. Hatherly, *Recrystallization and Related Annealing Phenomena*, Elsevier Science Ltd., Oxford, UK, 1996.

47. J.W. Cahn and W. Hagel, "Theory of the Pearlite Reaction", *Decomposition of Austenite by Diffusional Processes*, V.F. Zackay and H.I. Aaronson, Eds., Interscience, New York, 1960, pp. 131–196.

48. N. Gey, M. Humbert, M.J. Philippe, and Y. Combres, "Investigation of the Alpha- and Beta- Texture Evolution of Hot Rolled Ti-64 Products," *Mater. Sci. Eng. A*, 1996, Vol A219, pp. 80–88.

49. Los Alamos Polycrystal Plasticity Code, Los Alamos National Laboratory report LA-CC-88-6, Los Alamos, NM, 1988.

50. M. Humbert, H. Moustahfid, F. Wagner, and M.J. Philippe, "Evaluation of the High-Temperature Texture of the Beta Phase of a Ti-6Al-4V Sample from the Individual Orientations of Grains of the Low-Temperature Alpha Phase," *Scripta Metall. et Mater.*, 1994, Vol 30, pp. 377–382.

51. M.G. Glavicic, P.A. Kobryn, T.R. Bieler, and S.L. Semiatin, "A Method to Determine the Orientation of the High-Temperature Beta Phase from Measured EBSD Data for the Low-Temperature Alpha Phase in Ti-6Al-4V," *Mater. Sci. Eng. A*, 2003, Vol A346, pp. 50–59.

52. M.G. Glavicic, R.L. Goetz, D.R. Barker, G. Shen, D. Furrer, A. Woodfield, and S.L. Semiatin, "Modeling of Texture Evolution during Hot Forging of Alpha/Beta Titanium Alloys," *Metall. Mater. Trans. A*, 2008, Vol 39A, pp. 887–896.

53. R. Hill, "A Self-Consistent Mechanics of Composite Materials," *J. Mech. Phys. Solids*, 1965, Vol 13, pp. 213–222.

54. P.M. Suquet, "Overall Potentials and External Surfaces of Power Law or Ideally Plastic Composites," *J. Mech. Phys. Solids*, 1993, Vol 41, pp. 981–1002.

55. L. Briottet, J.J. Jonas, and F. Montheillet, "A Mechanical Interpretation of the Activation Energy of High-Temperature Deformation in Two-Phase Materials," *Acta Mater.*, 1996, Vol 44, pp. 1665–1672.

56. S.L. Semiatin, F. Montheillet, G. Shen, and J.J. Jonas, "Self-Consistent Modeling of the Flow Behavior of Wrought Alpha/Beta Titanium Alloys under Isothermal and Nonisothermal Hot-Working Conditions," *Metall. Mater. Trans. A*, 2002, Vol 33A, pp. 2719–2727

57. P. Vo, M. Jahazi, S. Yue, and P. Bocher, "Flow Stress Prediction during Hot Working of Near-Alpha Titanium Alloys," *Mater. Sci. Eng. A*, 2007, Vol A447, pp. 99–110.

58. R. Hill, "The Essential Structure of Constitutive Laws for Metal Composites and Polycrystals," *J. Mech. Phys. Solids*, 1967, Vol 15, pp. 79–95.

59. D. Dunst and H. Mecking, "Analysis of Experimental and Theoretical Rolling Textures of Two-Phase Titanium Alloys," *Zeitschrift für Metallkunde*, 1996, Vol 87, pp. 498–507.

60. R. Lebensohn and C. Tomé, "A Self-Consistent Anisotropic Approach for the Simulation of Plastic Deformation and Texture Development of Polycrystals: Application to Zirconium Alloys," *Acta Metall. Mater.*, 1993, Vol 41, pp. 2611–2624.

61. R.A. Lebensohn and G.R. Canova, "A Self-Consistent Approach for Modelling Texture Development of Two-Phase Polycrystals: Application to Titanium Alloys," *Acta Mater.*, 1997, Vol 45, pp. 3687–3694.

62. T.J. Turner and S.L. Semiatin, "A Crystal-Plasticity Model for the Flow Behavior of Two-Phase Alloy Systems," *Numiform 2004*, S. Ghosh, J.M. Castro, and J.K. Lee, Eds., American Institute of Physics, College Park, Maryland, 2004, pp. 1792–1797.

63. N.R. Barton and P.R. Dawson, "On the Spatial Arrangement of Lattice Orientation in Hot-Rolled Multiphase Titanium," *Model. Simul. Mater. Sci. Eng.*, 2001, Vol 9, pp. 433–463.

64. H. Moustahfid, N. Gey, M. Humbert, and M.J. Philippe, "Study of the Beta-Alpha Phase Transformations of a Ti-64 Sheet Induced from a High-Temperature Beta State and a High-Temperature Alpha+Beta State," *Metall. Mater. Trans. A*, 1997, vol. 28A, pp. 51–59.

65. M.G. Glavicic, P.A. Kobryn, T.R. Bieler, and S.L. Semiatin, "An Automated Method to Determine the Orientation of the High-Temperature Beta Phase from Measured EBSD Data for the Low-Temperature Alpha-Phase in Ti-6Al-4V," *Mater. Sci. Eng. A*, 2003, Vol A351, pp. 258–264.

66. S.V. Divinski, V.N. Dnieprenko, and O.M. Ivasishin, "Effect of Phase Transformation on Texture Formation Ti-Base Alloys," *Mater. Sci. Eng. A*, 1998, Vol A243, pp. 201–205.

67. M. Humbert, L. Germaine, N. Gey, P. Bocher, and M. Jahazi, "Study of the Variant Selection in Sharp Textured Regions of Bimodal IMI 834 Billet," *Mater. Sci. Eng. A*, 2006, Vol A430, pp. 157–164.

68. L. Zeng and T.R. Bieler, "Effects of Working, Heat Treatment, and Aging on Microstructural Evolution and Crystallographic Texture of α, α′, α″, and β Phases in Ti-6Al-4V Wire," *Mater. Sci. Eng. A*, 2005, Vol A392, pp. 403–414.

69. N. Gey, M. Humbert, M.J. Philippe, and Y. Combres, "Modeling the Transformation Texture of Ti-64 Sheets after Rolling in the Beta Field," *Mater. Sci. Eng. A*, 1997, Vol A230, pp. 68–74.

70. H. Moustahfid, M. Humbert, and M.J. Philippe, "Modeling of the Texture Transformation in a Ti-64 Sheet after Hot Compression," *Acta Mater.*, 1997, Vol 45, pp. 3785–3790.

Application of Neural-Network Models

Wei Sha and Savko Malinov, Queen's University, Belfast, United Kingdom

MANY COMPUTER-BASED MODELS and software have been developed to aid the understanding of metallurgical processes and to help reduce the investment of time and money for experimental work in research or for plant trials in production. Generally, the modeling techniques can be classified into two large groups: physical modeling and statistical modeling. Each has advantages and may be more suitable for certain areas of application. Both are constantly being improved and applied to a variety of processes and correlations in metallurgy and materials science.

Physical models and software are usually based on fundamental theories and laws describing the physical nature of the process. There are major achievements in the field of physical modeling of the correlations between processing parameters, microstructure formation, and product properties of metals and alloys, including thermodynamics and kinetics modeling and many other areas covered in other sections of this Handbook.

Statistical models, on the other hand, have applications in areas where large quantities of data exist and there are no physical models to adequately describe the process. A huge amount of data for various correlations in metals and alloys at different conditions is currently available in the literature. However, these data are sometimes rather confusing for use in the engineering practice because of contradicting data from various sources, for example.

Artificial neural network (ANN) is currently one of the most powerful and advanced modeling techniques based on a statistical approach, with very quick return for the practice. Neural-network (NN) modeling is most suitable for simulations of correlations that are hard to describe or cannot be accurately predicted by physical models. Because ANN modeling is a nonlinear statistical technique, it can be used to solve problems that are not amenable to conventional statistical methods. The ANNs have been applied to model complicated processes in many engineering fields—aerospace, automotive, electronic, manufacturing, robotics, telecommunication, and so on—and the method

is now a standard modeling technique. Since the1990s, there has been increasing interest in ANN modeling in various fields of materials science (Ref 1–4). The ANN models have been developed to model various correlations and phenomena in steels (Ref 5–13), aluminum alloys (Ref 14), nickel-base superalloys, mechanically alloyed materials, and so on. A special feature of the models is the ability to provide upper and lower limits of the predicted value, thanks to the introduction of probability theory into nonlinear data statistical analysis. How to calculate the error associated with the prediction, that is, the confidence level, is described by Ref 15. As a few examples, ANN modeling has been employed to study the mechanical properties of microalloyed steels as functions of alloy composition and rolling process parameters, the effect of carbon content on the hot strength of austenitic steels, continuous cooling transformation diagrams of vanadium-containing steels, and the hardness after surface nitriding of titanium alloys (Ref 16).

This article describes the NN technique, its software development, and its applications. The organization and features of NN models are presented. The effectiveness and applications of the programs are discussed. Examples for use of the models for simulation and optimization of various processes are demonstrated. Ways to improve and upgrade the models are given.

Principles and Procedures of NN Modeling

One direction of titanium research has been dedicated to ANN modeling and software development for simulation of processes, correlations, and phenomena in titanium alloys (Ref 16–23). This section describes the principles of NN modeling, using mostly examples from this program of work for demonstration purposes where needed, while the next section demonstrates NN applications, expanding to models concerning other types of materials.

Examples of the types of correlations that can be modeled are schematically shown in Fig. 1. The input parameters for each particular case of output are chosen based on the physical background of the process; all relevant input parameters must be represented. The graphical user interfaces of the corresponding software products are shown in Fig. 2. The basic principles of NN modeling and the algorithms of software programs are discussed in the following sections. The use of graphical user interfaces is discussed further.

Basic Principles

Artificial neural-network modeling is essentially an operation linking input to output data by using a particular set of nonlinear basis functions. An NN consists of simple synchronous processing elements that are inspired by the biological central nervous systems in living organisms. It comes to a conclusion, given the relevant information, or stimuli, and experience. The basic unit, or building block, in the ANN is the neuron, also referred to as the node. Neurons are connected to each other by links known as synapses. Figure 3 shows the schematic layout of the neurons within a network, with each arrow representing a link, or a synapse, between neurons. Each of these synapses has a weight attached to it that governs the output of the neuron. As the synapses are built up, a network is formed. For metallurgical modeling, hierarchical types of networks are most suited. In all cases in this article, feed-forward hierarchical (also referred to as multilayer perceptron) ANNs are used (Fig. 3). In a feed-forward NN, the input information is processed in a one-way direction—from input to output—and the neurons are ordered in layers, with an input set, hidden set(s), and an output set of neurons (input layer, hidden layer(s), and output layer in Fig. 3). Most of the NNs for metallurgical studies follow this structure, fully connected and feed-forward.

When developing a model, these steps should be followed:
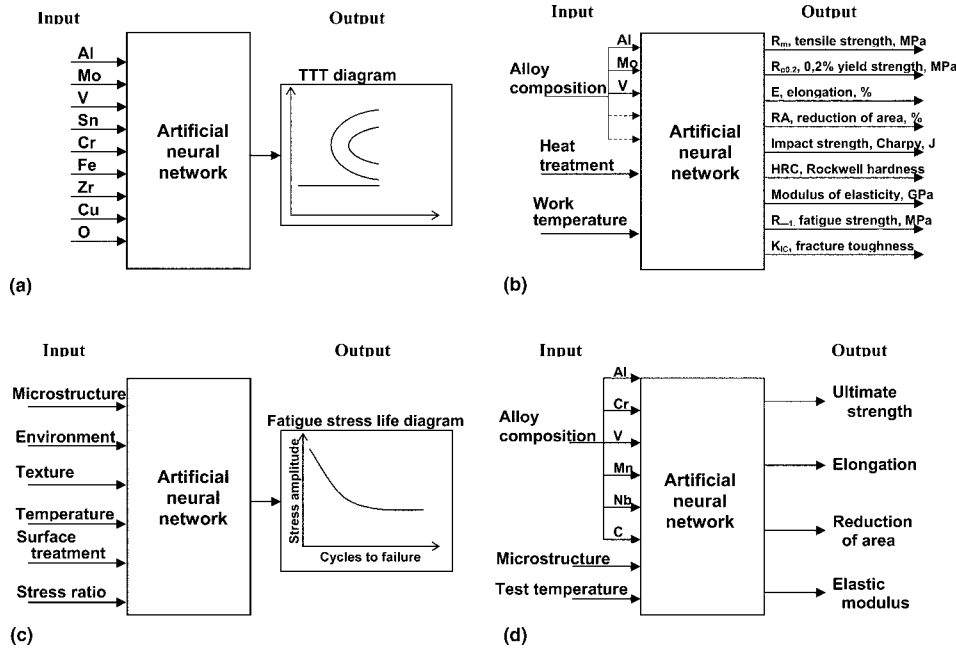
Fig. 1 Schematic models of artificial neural networks for simulation and prediction of various correlations in titanium alloys. (a) Time-temperature transformation (TTT) diagrams. (b) Mechanical properties of conventional titanium alloys. (c) Fatigue stress life diagrams. (d) Mechanical properties of titanium aluminides

Fig. 2 Graphical user interfaces of artificial neural-network software for simulation and prediction of various correlations in titanium alloys. (a) Time-temperature transformation (TTT) diagrams. (b) Mechanical properties of conventional titanium alloys. (c) Fatigue stress life diagrams. (d) Mechanical properties of titanium aluminides
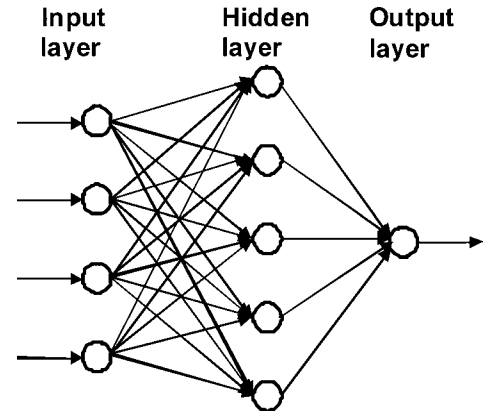
Fig. 3 Schematic of the structure of a feed-forward hierarchical artificial neural network

1. Determination of input/output parameters
2. Database collection
3. Analysis and preprocessing of the data
4. Training of the NN
5. Test of the trained network
6. Use of the trained network for simulation and prediction

Input Parameters. Selection of the property-related parameters, or input parameters, is based on the physical background of how the target property is determined. Omitting the unimportant parameters benefits the development of the model and simplifies further application. The input parameters should ideally be independent variables. However, closely related input parameters, such as processing temperature and processing phase fraction although one is a function of the other (but not solely due to, for example, factor(s) in addition to temperature also contributing to the phase fraction), can be used as separate input parameters. Sometimes, the selection is limited by data availability, but this could lead to an inadequate model if an important input variable is not included.

Example 1: Maraging Steels. For maraging steels (Ref 6), there are two major thermal treatment processes: austenitizing and aging. A cold deformation procedure is sometimes used between the two treatments to increase the achievable strength level. As long as the austenitizing process ensures a full transformation to austenite, the temperature and time only marginally affect the mechanical properties after aging, in that an increase in austenitizing temperature (T_{aus}) or time (t_{aus}) usually leads to slightly better toughness and a slight drop in strength. Therefore, T_{aus} and t_{aus} need not be used as inputs. In practice, the chosen T_{aus} and t_{aus} should be, respectively, sufficiently high and long to ensure a fully austenitic structure and then as low and short as possible to avoid austenite grain growth. The cooling method after austenitizing is usually air cooling or water quenching (occasionally oil quenching). It is chosen to ensure a full martensitic transformation and need not be taken as an
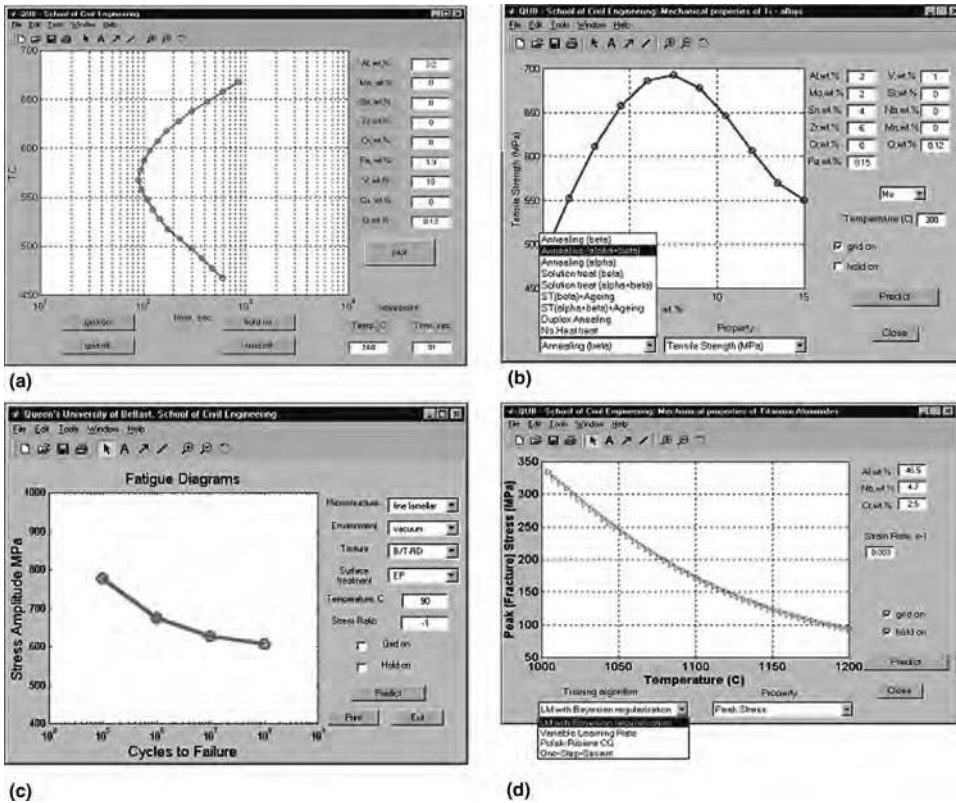
input parameter. The cold deformation and aging treatment are important to the mechanical properties of maraging steels. Therefore, the input parameters include cold deformation degree, aging temperature, and aging time. Maraging steels are frequently used and sometimes essential for many high-temperature applications, so the working temperature is another input parameter to the ANN model.

The chemical composition of the maraging alloys contains 13 elements: carbon, aluminum, cobalt, chromium, copper, manganese, molybdenum, niobium, nickel, silicon, titanium, vanadium, and tungsten. Elements boron and zirconium are not efficient alloying elements, and their uses are rare in the development of maraging steels. Element beryllium, although classified as a strong hardener, has not been widely employed because of toxicity concerns. Influences from residual impurities of calcium, nitrogen, oxygen, phosphorus, and sulfur are ignored.

Contradicting Information from Various Sources. In some cases, there is a contradiction between the data (for example, time-temperature transformation diagrams and mechanical properties) (Ref 23) for the same alloy taken from different sources and authors. If there is no reason to discard one source against the others, all data should be included in the database. However, the contradiction in the experimental data will be a main source of deviation of the network response. Usually, the calculated results are within the range of experimental data published in various sources (Ref 23), showing an averaging effect. The difference between the NN predictions and the experimental data should be comparable with the difference between experimental data published in various sources.

Sometimes, the input parameter is qualitative, for example, microstructural types (Ref 23). There may be overlap in microstructural types due to the different terminology used in various sources. For example, for gamma titanium aluminides, fully transformed microstructure is similar to fully lamellar. These different terminologies should be preserved in modeling, however, to avoid possible distortion from original literature data.

Training. As indicated by step four, the NN must be trained. The training procedure is the most significant part of NN modeling. Training is a procedure whereby a network is adjusted to do a particular job. Usually, NNs are trained using a large amount of data containing input with corresponding output values, called input/output pairs, so that a particular set of known inputs produces, as nearly as possible, a specific set of known target outputs. Training involves adjusting the weight associated with each connection (synapse) between neurons within the network by comparison of the computed outputs and the targets, until the computed outputs for each set of data inputs are as close as possible to the target data outputs. The weight of a synapse, multiplied by the strength of the signal on that synapse, defines the contribution of that synapse to the activation of the neuron for which it is input. The total activation of a neuron is then the sum of the activation of all its inputs plus a bias value, and this defines the value of the output signal for that neuron, via a transfer function. By adjusting the values of these synaptic weights throughout the network, the outputs of the NN for any given set of inputs can be altered. Training is a continuous process, until the network correctly simulates the known behavior of the system to be modeled. The simulation will rarely be exact; training is usually aimed at minimizing the sum of the squares of the differences (the errors) between the predicted and experimentally measured values of the outputs.

Training algorithm (also known as learning rule) refers to the procedure for modifying the weights and biases of a network. The training algorithm is applied to train the network to perform some particular task. It is the mathematical apparatus by which the data are used to fit (train) the network. There are many different training algorithms. To achieve the best result, various training algorithms may be attempted. For feed-forward NN, the most commonly used ones are batch gradient descent, batch gradient descent with momentum, one-step secant, scaled conjugate gradient, resilient backpropagation, Polak-Ribiere conjugate gradient, Fletcher-Powell conjugate gradient, Powell-Beale conjugate gradient, variable learning rate, and Levenberg-Marquardt. These training algorithms are standard. Their mathematics can be found in many NN books. In software, the training algorithms are implemented in the computer code.

The method used in this article is the Levenberg-Marquardt algorithm. Typically, the use of Levenberg-Marquardt leads to a reduction of orders of magnitude in the number of training iterations required compared with backpropagation, and it is highly reliable. From a comparative work (Ref 21), the algorithms based on batch gradient descent require approximately 1000 times longer training time compared to the Levenberg-Marquardt training and do not give better results. The algorithms based on conjugate gradient require 20 to 30 times longer training time compared to the Levenberg-Marquardt training. The Levenberg-Marquardt training algorithm is the fastest, but the results for the test and the whole dataset for this algorithm are not acceptable. The most probable reason for this is the problem with overfitting. When Bayesian regularization in combination with the Levenberg-Marquardt training was used, the training time was increased, but the R-values (Ref 23) for the test and the whole dataset were better. The training with variable learning rate is slower than Levenberg-Marquardt.

The two most popular algorithms are the Levenberg-Marquardt algorithm and Bayesian regularization. The Levenberg-Marquardt algorithm is the fastest training algorithm for networks of moderate size. It has a memory-reduction feature for use when the training set is large. The time required for training can be dramatically reduced using this method. However, it is sometimes difficult to find the best model using this algorithm. Bayesian regularization is a modification of the Levenberg-Marquardt algorithm for obtaining networks that generalize well. It reduces the difficulty in determining the optimum network parameters.

Generally, the features of ANNs can be summarized as follows:

- The NN models are statistical models; that is, they are not based on any physical theory. However, they can be used to model complex processes and correlations.
- The NNs work with numerical characteristics (input/outputs).
- A large amount of data is required. Through the use of such a database, an NN can be trained to perform complex functions.

Algorithm of Computer Program for NN Training

This part of the development encompasses step two, populating the database; step three, analyzing and processing the data; and step four, training the NN.

Pretraining Procedures The block diagram (Fig. 4) starts with the accumulation of the database. The database can only be constructed by collecting available data for the correlation being modeled. The data can be from handbooks, journal papers, and one's own experimental data and usually span back over a long period. The initial organization of the accumulated data can be in a standard spreadsheet file format. Each row in this file represents one input/output data pair. Each column in this file represents one input or output. Some of the inputs or outputs are properties that are not numerical values but rather categories. Examples of this are heat treatment (Ref 20), microstructure (Ref 18, 22), material grade (Ref 18), environment (Ref 18, 22), surface treatment (Ref 22), and cooling method (Ref 9). At this stage of pretraining, these should be digitized by means of attributing different digits to the various categories. In some cases, the output is a graph (Fig. 1a, 1c). In these cases, the graph should be appropriately digitized and presented as a set of numbers (Ref 9, 19, 22). Once the data accumulation is completed, the file containing the database is converted and saved in ASCII format. This file is read by the computer program for creation of the model (Fig. 4) and is put as a matrix **M** for further manipulations. The matrix **M** has the dimension $m \times n$, where m is the number of rows and is equal to the number of data pairs in the database, and n is the number of the columns and is equal to the
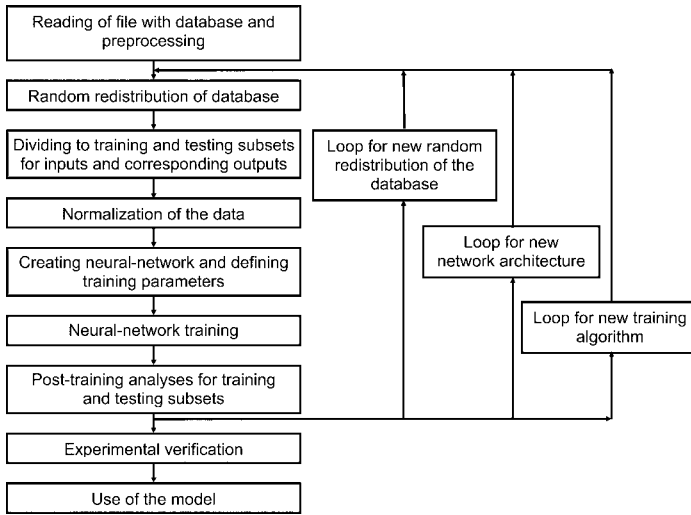
Fig. 4 Algorithm of computer program for creation of neural-network model

number of inputs plus the number of outputs. The spreadsheet and the ASCII files are dynamically linked, so that each change and addition in the spreadsheet file results in an automatic upgrade of the ASCII file and the matrix **M** to be used for the model design.

The next step in the computer program is the random redistribution of the database (Fig. 4). In NN modeling, one part of the data (usually two-thirds) is used for model training, and the remaining part (usually one-third) is not used in the training procedure but is used for testing of the model. Before training, the database is randomly divided into these two parts. In the computer program, this is done by the block for random redistribution of the database. First, a vector with random numbers between 0 and 1 is generated. The size of this vector is equal to the number of the data pairs (m). This vector is thereafter stored as an additional $(n + 1)$th column of the matrix **M**. The next operation is an automatic rearrangement of the rows of matrix **M** in ascending order of the numbers in the $(n + 1)$th column. After implementation of this operation, the $(n + 1)$th column is taken out, and the new matrix **Mr** is produced. **Mr**, with dimension $m \times n$, contains the same data as **M**. The only difference is that the rows (data pairs) in **Mr** are randomly redistributed as compared to the original **M** matrix. Written in this way, the computer program allows new random distribution of the whole database into the sub-datasets each time it is run, because the vector with generated random numbers will be different each time.

The next step in the computer program is to extract the training (**Mr$_{tr}$**) and the testing (**Mr$_{tst}$**) matrices for input and corresponding output (Fig. 4). The first ⅔ rows from **Mr** are extracted as matrix **Mr$_{tr}$** and the last ⅓ rows as matrix **Mr$_{tst}$**. These new matrices have dimensions $m_1 \times n$ and $m_2 \times n$. Obviously,

$m_1 + m_2 = m$; $m_1 = (⅔)m$, and $m_2 = (⅓)m$. **Mr$_{tr}$**(m_1,n) and **Mr$_{tst}$**(m_2,n) contain data pairs (input with corresponding output) that will be used for training and testing of the model, respectively. The next step is to divide these matrices into matrices containing inputs and outputs only—**M$_{trin}$**(m_1,n_1), **M$_{trout}$**(m_1,n_2), **M$_{tstin}$**(m_2,n_1), and **M$_{tstout}$**(m_2,n_2). Here, **M$_{trin}$**(m_1,n_1) is the matrix containing the inputs for training, **M$_{trout}$**(m_1,n_2) is the matrix containing outputs for training, and so on. n_1 is equal to the number of the inputs, n_2 is equal to the number of the outputs, and $n_1 + n_2 = n$. These four matrices are further used for training and testing of the model. It should be restated that each time the program is run, new random redistribution of the database will be executed, and, as a result, these four matrices will contain different data. It should also be mentioned that the various random redistribution of the database results in different NN performance. This is discussed in a following section.

The values in the matrices for training and testing have different dimensions and ranges. To overcome this, the next block in the computer program is for normalization of the data. Depending on the transfer function used, the data are normalized between 0 and 1 or between −1 and 1, applying Eq 1(a) or (b), respectively:

$$x_N = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \qquad \text{(Eq 1a)}$$

$$x_N = 2\frac{x - x_{\min}}{x_{\max} - x_{\min}} - 1 \qquad \text{(Eq 1b)}$$

where x_N is the normalized value of a certain parameter, x is the measured value for this parameter, and $x_{\min}$ and $x_{\max}$ are the smallest and the largest values in the database for this parameter, respectively.

Training Parameters. The next blocks in the computer program are for creating the NN, defining the training parameters, and the actual training of the model. The following matters are important in the design and training of NNs:

- Database and its distribution
- Choice of architecture of the NN
- Training algorithm and parameters
- Transfer function

Other training parameters, such as learning rate, performance goal, and minimum performance gradient, may have minor influence but aim mainly at the optimization of the training time and computer memory use and have only little influence on the final performance of the trained model.

If software is developed on the user level, an option can be incorporated to enable the user to add his own data and retrain the model. This option is discussed in a following section. Here, some recommendations on the selection of the important parameters and their influences on the NN performance are discussed. The reliability of the NN model for any particular combination of database distribution, architecture, training algorithm, and transfer function can be tested by an analysis of the network response in a form of linear regression between network outputs (predictions) and corresponding targets (experimental data) for the training and testing datasets (Fig. 5).

Effect of Creating Random Sets. Dividing **Mr** into training and testing subsets can result in significantly different model performance. This difference can be explained with the ranges of input data variation for the training and testing datasets. In some cases of random dividing, instances may occur where the range of variation of the data for the training dataset is narrow compared to the range of variation of the test dataset. In these cases, most of the data pairs in the test datasets are a new encounter for the trained NN and fall outside the range of variables for which the model has been trained. An appropriately trained NN model can give reliable predictions for new instances within the ranges that it has been trained (interpolation), but appreciable errors are possible for predictions outside these ranges (extrapolation). Hence, it is recommended to run not one but a number of training cycles with different random redistribution to find the case with the most suitable dividing of the data to training and testing sets.

Architecture of the NN refers to the number of layers in the NN and the number of neurons in each layer. The numbers of neurons in the input layer and the output layer are determined by the numbers of input and output parameters, respectively, while the number(s) of neurons in the hidden layer(s) can vary. The number of hidden layers could be one or more. One hidden layer is enough for appropriate model performance. Increasing the layers
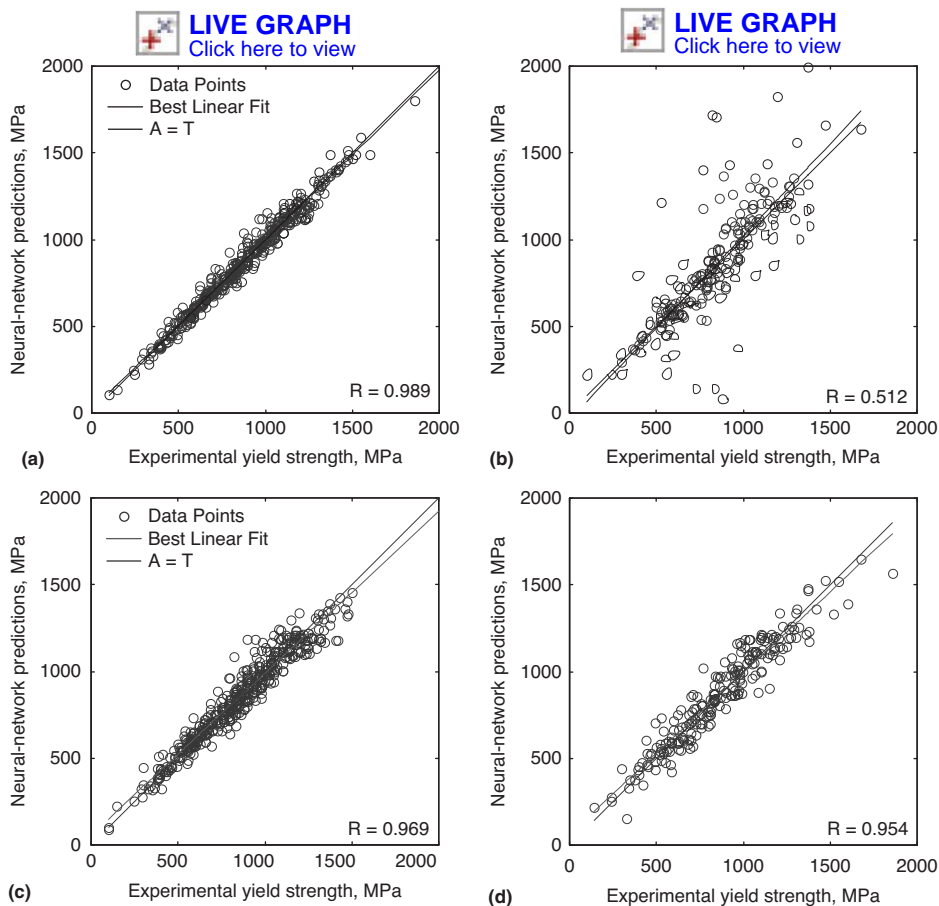
Fig. 5 Posttraining linear regression analysis between experimental data (horizontal axis, T values) and neural-network predictions (vertical axis, A values). The diagonal dashed line is A = T. (a) and (c) Training datasets. (b) and (d) Testing datasets. Levenberg-Marquardt training alone is used for (a) and (b). Levenberg-Marquardt with Bayesian regularization is used for (c) and (d). All cases have 14 hidden neurons.
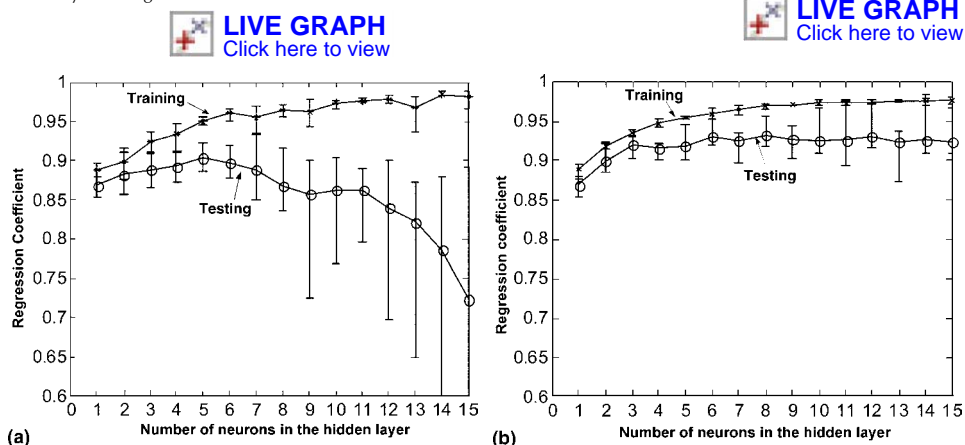
Fig. 6 Regression coefficients between the neural-network predictions and the experimental data for various numbers of neurons in the hidden layer using Levenberg-Marquardt training (a) without and (b) with Bayesian regularization

to two results in a remarkable increase in the unknown parameters (connections) to be fitted, which itself drastically increases the requirements for the amount of data without any noticeable improvement in the final model performance (Ref 20). Hence, the general structure of input, one hidden, and one output layer is preferred. To find the optimal architecture, various numbers of neurons in the hidden layer should be attempted. The number of neurons in the hidden layer (NNHL) considerably influences the model performance. Some practical recommendations for selection of the NNHL are given. Two factors are important for determining the number of neurons:

- First, an increase in the NNHL increases the connections and weights to be fitted. The NNHL cannot be increased without limit. The maximum NNHL is different for each particular case and depends on the number of inputs, number of outputs, and number of available data pairs. The number of connections to be fitted can be worked out by drawing the scheme of the NN (Fig. 3) for the particular case of inputs, outputs, and NNHL and calculating the connections to be fitted.

- Second, the initial increase in the NNHL usually results in improvement of the model performance. However, one problem that can occur when training with a large number of neurons is that the network can overfit on the training set and not generalize well to new data. In other words, the network is too flexible, and the error of the training set is driven to very small values, but when new data are presented to the network, the error is large. The optimal NNHL depends on the database, nature of the problem to be modeled, and the training algorithm and should be determined for each particular case. Overfitting can be prevented by various techniques aimed at better generalization of the model.

Example 2: Yield Strength of Titanium Alloys. Figure 6 demonstrates one example of finding the optimal architecture of the NN in terms of the number of neurons in the hidden layer. The case is for the prediction of yield strength of titanium alloys (Fig. 1b). The number of inputs is 13, the number of outputs is 1 (if just yield strength is considered), and the number of available data pairs is 662 (441 for training and 221 for testing). To find the optimal architecture, various numbers of neurons in the hidden layer were attempted. The Levenberg-Marquardt training algorithm was used for this study. The results for the influence of the number of neurons in the hidden layer on the NN response are given in Fig. 5 and 6. The results in Fig. 5 are for single training with 14 neurons in the hidden layer and show the network response in a form of linear regression analysis between the network outputs (predictions) and the corresponding targets (experimental data) for two different training algorithms. The results in Fig. 6 are presented in the form of correlation of the regression coefficient (R) between the NN predictions and the experimental data for the training and testing datasets for various numbers of neurons in the hidden layer. For each case, the values for the R coefficient are averaged from five separate training runs at the same conditions and using different random division of the dataset. When the number of neurons in the hidden layer is increased from 1 to 5, the R coefficient for both training and test datasets quickly increases (Fig. 6a). Further increase in the number of neurons results in further increase of the

R coefficient for the training dataset that is approaching the value of 1 (Fig. 5, 6). However, the regression coefficient for the test dataset quickly decreases to average values <0.8 with large error bars (Fig. 5, 6). This observation shows that there is an effect of overfitting when the number of neurons is increased to six and above. One way to prevent overfitting is by applying Bayesian regularization in combination with the Levenberg-Marquardt training (Fig. 5, 6). This is one of the techniques aimed at better generalization. The results obtained for the various numbers of neurons in the hidden layer are appreciably stable (Fig. 6b). Values of coefficient *R* for both training and test datasets increase when the number of neurons is increased to eight. Further increase of the neurons does not result in overfitting, and the results for the regression analyses of both training and test datasets are comparable (Fig. 5c, 5d). However, at the same time, an increase in the number of neurons above eight does not give appreciable improvement, while the training time was significantly increased. Considering all of the aforementioned, the optimum for this case of NN architecture is eight neurons in the hidden layer.

The type of training algorithm used is important for both NN response and the computational resources necessary for training. A summary analysis of the influence of various training algorithms on the performance of the NN model is given subsequently. Various training options can be carried out, including training on variations of mean square error for better generalization, training against a validation set, and training until the gradient of the error reaches a minimum. Improvement of the generalization can be achieved by means of regularization and early stopping with validation. The general opinion is that automated regularization based on Bayesian regularization in combination with Levenberg-Marquardt training usually gives the best result in terms of model performance and training speed. However, in some cases, other training algorithms, such as one-step secant, Polak-Ribiere conjugate gradient, and variable learning rate, may give competitive results.

The transfer function transforms the neuron input value into the output value. The most popular functions are hard limit (hardlim), linear (purelin), log sigmoid (logsig), and hyperbolic tangent sigmoid (tansig). For all cases of NN, a linear transfer function is suitable in the output layer. The transfer functions in the hidden layer are generally s-shaped curves, with the output value confined within limits of (0,1) or (−1,1) for log sigmoid and tan sigmoid functions, respectively. Mathematical formulations of the main transfer functions and their general shapes can be found in Ref 19. Usually, there is little influence whether log sigmoid or tangent sigmoid function is used. Hence, this is not considered as a significant parameter for the NN response.

Optimal Model Parameters. The computer program for design of the NN model can be written in a way to find the best combination of database distribution, network architecture, and training algorithm by grid search (Fig. 4). Loops are organized for the previously mentioned parameters. By varying these, all possible combinations of network architecture and training algorithm with different database distribution are attempted and stored. For each combination, the correlation coefficients between the NN predictions and the experimental data for the training and testing datasets are stored and used for further evaluation. The program realization is as follows. First, the number of neurons in the hidden layer is put equal to one, and the first training algorithm, for example, one-step secant, is used. For this combination, at least five attempts at training with different random redistribution of the database are executed. Next, the number of neurons is increased to two, and again, five attempts at training with different random redistribution (but the same as those for the aforementioned case and for all cases) of the database are executed. The same is executed for three, four, five, and so on NNHL. The next loop is to change the training algorithm, and the same procedure is repeated for the new and for all training algorithms given previously. In this way, the NNs that have been trained with all possible combinations of network architecture, training algorithm, and various random redistributions of the datasets are stored. The next step is to find the best case. To do this, the following procedure is applied. First, using the runs with various random redistributions, the average regression coefficients between NN predictions and the experimental data for the training and testing datasets for various network architecture and training algorithms are calculated. For one training algorithm, the regression coefficients after training with various NNHL are compared, and the best case (the winner) in terms of best architecture for this training algorithm is found. This procedure is executed for each training algorithm. The winners for the various training algorithms are thereafter compared, and the best combination of network architecture and training algorithm is found. For this combination, the best case among the five cycles of training with various random redistributions is extracted, and this is the final winner in a sense of best database distribution, network architecture, and training algorithm. This case is then stored, and the NN trained with these parameters is used for further simulations and predictions. Readers should be reminded that the optimal combination of database distribution, network architecture, and training algorithm depends on many factors, such as nature of the problem to be modeled, number of inputs and outputs, and database available, and can differ considerably from case to case. The procedure described here for finding the optimal case of trained NN, and reliable NN modeling, in general, requires a significant amount of

data, but, once trained and created, the model can quickly and easily be used for the prediction of new instances.

Posttraining Procedures. The next block of the work procedure is for further experimental verification of the model by statistical analysis of the model predictions error. The relative error (relative difference between the model simulations and the experimental data) is calculated for each data pair. The results are presented by plotting them in the classical plot of occurrence versus error, also referred to as frequency versus error. This plot is more informative compared to the plots of linear regression analysis (Fig. 5). Using it, specialists in the field can conclude whether the achieved accuracy in terms of error variation is acceptable for the phenomenon being modeled. Here, an example is given of such an analysis for modeling the modulus of elasticity of titanium alloys (Fig. 7). Additional plots and analyses for other correlations are given and discussed in Ref 9.

Further verification of the model simulations can be performed by the user by means of direct comparison of the software predictions and experimental data. The user can calculate the results for new input values and thereafter can compare these predictions with experimental data taken from the literature or with one's own experimental data. Some comparisons of this type are given in Ref 9.

Use of NN Modeling

General Organization of Software and Graphical User Interfaces. Software systems should be designed in a modular fashion, with each module designed in such a way that it can be upgraded by a later version without substantially affecting the operation of the remainder of the model. Various modules can be linked according to the block diagram given in Fig. 8. The modules for databases, the computer program for training, and trained ANN are described in the previous section. Once the NN models are trained and verified, they can be used for further simulations and predictions for an arbitrary, user-defined new (unused in the training) combination of input values. One way of organizing the interactions with the system is by interface modules. Two types of interface are possible: text user interface and graphical user interface (GUI). In the text version, the user is asked interactively for input of the necessary information; once it is completed, the output from the system is generated.

Over the years, MatLab Neural Network Toolbox (Ref 24) has become one of the most popular commercial software that can be used to develop and use NN models. An example of an NN program code is given in Ref 23. Graphical user interfaces (examples shown in Fig. 2) can be developed using the MatLab programming language that allow easy and further use of the models (matching examples shown in
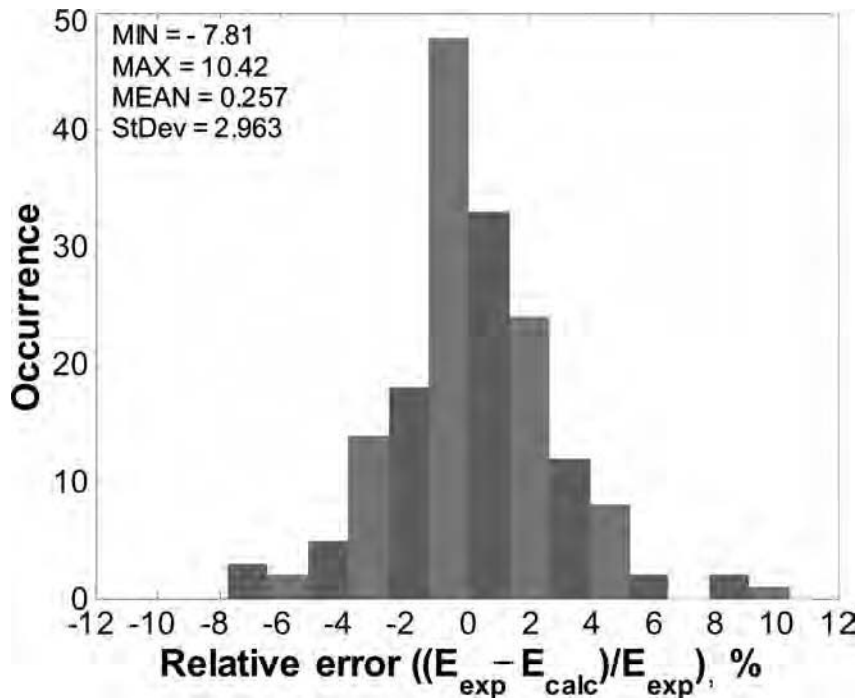
Fig. 7 Posttraining validation of software simulations by statistical analysis of the error of model simulations for modulus of elasticity
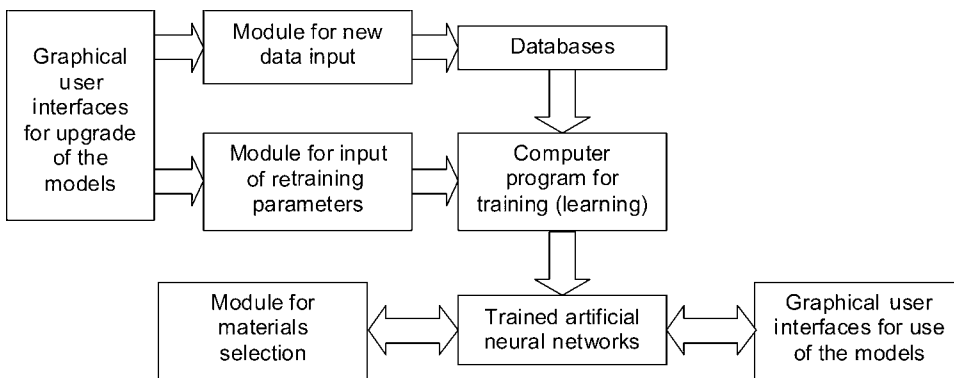
Fig. 8 Block diagram of a software system for modeling based on the artificial neural-network approach

Fig. 1). The user can run GUIs in the way shown in Fig. 2 to input the desired new combination of input values using the textboxes, menus, list boxes, pop-up menus, radio buttons, sliders, and so on. Once the information is entered, the user can click on "predict" or "plot." The entered information is then acquired, organized, and processed (Fig. 8) as a new input combination to the trained NN for the corresponding GUI. The trained ANN calculates output for this combination of inputs. The output is returned to the GUI and is visualized in the form of a graph (Fig. 1a, c, for example) or the dependence of the properties on a desired parameter. In this way, the user can very easily obtain and evaluate different kinds of dependence of any property involved, that is, the dependence of the outputs from the chosen input. Furthermore, the GUI provides the opportunity for easy comparison of output properties corresponding to various sets of inputs by plotting them together. In the cases where the output is a graph, the GUIs also allow smoothing of the predicted profile with a user-defined coefficient.

Training of the various modules is performed using the collected databases for each particular case. However, the system can be made open for further upgrading by the users by adding additional data and training (see the left block in Fig. 8). This feature is discussed in a following section.

In practice, integrating and linking by an introductory GUI window allows easy switching between the various applications. Such organization is convenient for the user. Once any of the GUIs have been used to predict a property (for example, a mechanical property), it is straightforward to switch to others to check if the other properties (for example, a time-temperature transformation diagram) are satisfactory.

Predicting Properties of Existing Materials. New materials with special properties are continuously being developed to meet engineering requirements. The aim is usually to achieve desirable values of one or a combination of properties. However, most often, the other properties of these materials are unknown when they are first designed, and these properties are studied by additional experiments after the materials are introduced in certain applications. The lack of knowledge of these other properties inhibits full exploitation of the new materials. Neural-network models can be used to predict unknown properties of existing materials. For this purpose, the user can input and fix the composition of the alloy and can predict the unknown property (mechanical, kinetics of phase transformation, etc., as desired) for various conditions. To demonstrate this mode of application of the NN technique, three examples are given.

Example 3: Time-Temperature Transformation Diagrams. Neural-network modeling is used for the prediction of time-temperature transformation (TTT) diagrams (Fig. 1a, 2a). Experimental investigation of TTT (and continuous cooling transformation) diagrams is both costly and time-consuming. Hundreds of experiments are necessary to construct such diagrams for a single alloy. For that reason, when a new material is designed, its TTT diagrams are usually unknown. Presently, there is still a lack of such diagrams, even for widely used alloys. Using an NN model, TTT diagrams are predicted (Ref 18, 19) for many existing titanium alloys for which experimentally determined TTT diagrams are not available in the literature. These simulation results can be used to explain the transformation kinetics during cooling of corresponding alloys. The diagrams will also be useful in the design of experimental studies of TTT diagrams for these alloys. In this way, the amount of experimental work could be significantly reduced. In this respect, it should be mentioned that, for some of the alloys, experimental TTT diagrams have become available after simulations and good correspondence between the predictions and the later available experimental diagrams are demonstrated. These agreements provide confidence in the model simulations. Other cases of using the developed model to predict properties of existing titanium alloys include the calculation of fatigue strength S-N diagrams and mechanical properties (Ref 23).

Example 4: Correlation of Composition, Processing, and Properties of Austenitic Stainless Steels. This work (Ref 25) includes:

- Correlation between chemical composition, process variables, and flow stress of austenitic stainless steels under hot compression
- Constitutive flow behavior of type AISI 304L stainless steel during hot torsion
- Microstructural evolution during dynamic recrystallization
- Correlation between chemical composition and tensile properties

Multilayer perceptron-based feed-forward networks have been trained by comprehensive in-house datasets. Very good performance of the NNs is achieved in modeling the following:

- Influence of alloy composition and processing parameters on flow behavior of austenitic stainless steels
- Effect of strain rate on torsional flow behavior of 304L stainless steel
- Combined influence of temperature and strain on dynamic recrystallization behavior

The simulated results are found to be consistent with the expected metallurgical trends. Figure 9 shows an example of model calculation results, demonstrating the capability of using NNs to evaluate the importance of different input parameters.

Example 5: Applications of ANN in Modeling Maraging Steel Properties. Artificial neural-network models can be used to predict the martensitic start temperature, M_s, and mechanical properties of maraging steels with sufficient accuracy within the data range used in model development. In maraging steels, the interactions between cobalt and molybdenum are complicated. Many alloy developments are based on these interactions. They are examined here based on model calculations. Since the models have been designed on statistical analysis and not on physical theories, some of the model outcomes are discussed from the metallurgical point of view, to justify and validate them.

To demonstrate the interactions between cobalt and molybdenum on M_s temperature, the alloy system is chosen to base around the chemical composition of the classical Fe-18Ni maraging steels, Fe-0.01C-0.1Al-18Ni-0.4Ti-9Co-4Mo (referred to as Fe-18Ni-9Co-4Mo). The amounts of cobalt and molybdenum are variables with values as low as zero, so that the influence from these elements can be shown (Fig. 10). Cobalt always raises the M_s temperature when no molybdenum is present. However, with molybdenum present, the influence of cobalt becomes complicated. Experimentally, with the addition of 1.5% Mo, cobalt decreases the M_s temperature when its amount is higher than 15%. Such a tendency can be seen in the curve corresponding to 2% Mo in Fig. 10,

where the increase in M_s with cobalt ceases when cobalt is higher than approximately 9%. From the ANN model, an amount of cobalt above 6% will cause a reduction in M_s when molybdenum is 6%. The increase in molybdenum always suppresses M_s, and the extent of the reduction is enhanced by the increase in the percentage of cobalt in the alloy.

The rate of precipitation is reflected by the increase in hardness at early stages of the aging process. There is a combined influence of cobalt and molybdenum on age-hardening kinetics. The rate of increased hardness for systems with cobalt is greater than for the system

without it (Fig. 11). The hardness increase at the peak position of cobalt-containing systems is much larger than that of the cobalt-free system. This effect can be explained with a metallurgical mechanism. One important role of cobalt in maraging steels is to lower the solubility of molybdenum in martensite, thus producing more densely distributed molybdenum-containing precipitates to increase the strength. The equilibrium fraction of such precipitates increases, promoting the hardening effect. The precipitation of molybdenum is strongly modified by the presence of other elements, most noticeably cobalt. Without cobalt, the
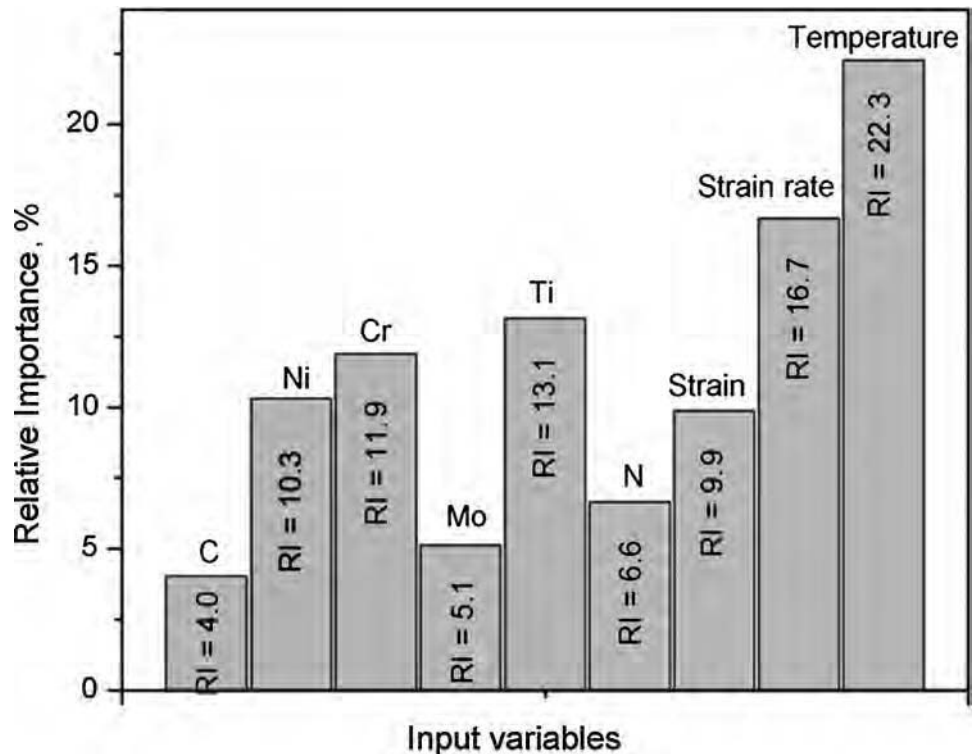
Fig. 9 Relative importance (RI) of individual input variables (composition and external factors) for flow stress predictions in austenitic stainless steels. Source: Ref 25

LIVE GRAPH
Click here to view
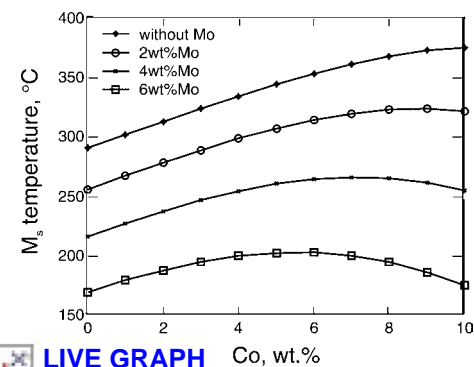
Fig. 10 Influence of cobalt and molybdenum on the martensitic start (M_s) temperature of the Fe-18Ni-Co-Mo system

LIVE GRAPH
Click here to view
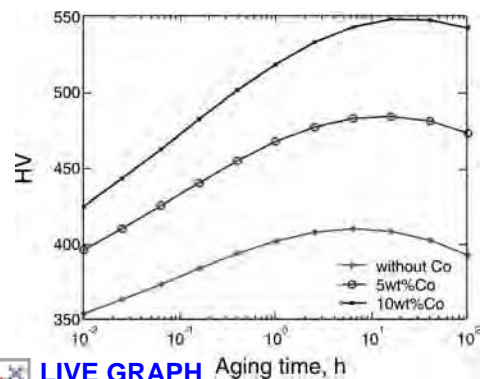
Fig. 11 Influence of cobalt on the precipitation age-hardening kinetics of an Fe-18Ni-Co-4Mo system at 482 °C simulated using the artificial neural-network model. HV, Vickers hardness

precipitation of molybdenum takes place much more slowly. In conclusion, the prediction of the NN model is in agreement with what is expected from a metallurgical viewpoint.

The combined influence of cobalt and molybdenum on other mechanical properties is demonstrated here, using ultimate tensile strength (UTS) and Charpy impact energy (Aκ) as examples. The alloy system is Fe-18Ni-xCo-xMo aged at 482 °C for 4 h without cold deformation and tested at room temperature. The calculated results show that the addition of a small amount of molybdenum decreases the UTS (Fig. 12). Experimentally, the combined effects of cobalt and molybdenum may not be beneficial when the alloying amount of molybdenum is less than 2%. In fact, in commercial maraging steels, the amount of molybdenum is usually higher than 2%.

Without molybdenum, the alloying of cobalt increases the Aκ values (Fig. 13), but such a system will have little aging response. Alloying with 2% Mo significantly increases the Aκ value of the alloy for cobalt concentrations lower than 8%. When the amount of molybdenum is further increased, the Aκ value is significantly decreased. The establishment of Fe-18Ni-9Co-4Mo-0.4Ti as the nominal composition of commercial-grade C250 is of no surprise, because this provides a good combination of strength and toughness. The laboratory-tested UTS and Aκ of C250 alloy (Fe-18Ni-8.5Co-5Mo-0.4Ti) aged at 480 °C for 5 h are 1870 MPa and 37 J, respectively, which are in reasonable agreement with the predicted values shown in Fig. 12 and 13.

The maraging steel 1RK91 (Fe-12Cr-9Ni-4Mo-2Cu-1Ti-0.3Al) was developed by Sandvik and is characterized with good fatigue strength at elevated temperatures. It can reach ultrahigh strength through aging treatment. The agreement between calculation and experimental measurement is acceptable (Fig. 14). The calculated room-temperature UTS value is 2543 MPa (4 h aging at 475 °C), within the range of 2450 to 3000 MPa quoted by Ref 26 for aged products. This alloy differs from many previous maraging grades in that it contains

approximately 2 wt% Cu. Without copper, the alloy may still be age hardened, but the time to achieve considerable hardening will be extraordinarily long, as shown in Fig. 14. In fact, although the good mechanical properties of this alloy are attributed to precipitates rich in nickel, aluminum, and titanium, copper clusters form before these and act as nucleation sites. The application of the NN model to this alloy demonstrates that it can be a good guide in the development of new alloys.

In conclusion, these ANN models can predict well even the complicated cobalt-molybdenum influences on M_s temperature and mechanical properties of maraging steels. The ANN models can also be a guide for new alloy design. Based on these models, optimization of alloy composition and processing parameters can be carried out (Ref 23). For a specified working temperature and the required combination of strength and toughness properties, the model can recommend the most economical composition and processing routes and therefore benefit the industry.

New Alloy Design. Neural-network models are usually used to predict outputs for a user-defined combination of inputs. However, very often in practice, it is necessary to solve the reverse problem, namely, to find combinations of inputs that would produce a desirable output or a combination of outputs. For example, it is of great interest for theory and practice to find the optimal alloy composition, processing, and environmental conditions that would provide a desirable property or combination of properties. The background of new alloy design is the knowledge of the correlations between the alloy composition and the properties. The optimization of one or some of the properties of the material being designed is closely related to the optimization of its composition. However, new materials are still mainly designed empirically, by carrying out expensive and time-consuming experimentation on trial alloys. An NN model has assembled results from hundreds of experiments, and it is trained with them. As a result, it is capable of predicting properties for new alloy composition. It can be used for

optimization of the composition of the alloy being designed to achieve an optimal value for the desired properties. Based on the trained NNs, computer programs for optimization of the input values (the alloy composition and processing) to obtain the desired property (or combination of properties) can be developed by incorporating the trained NN in the body of the program. The optimization criteria can be user-defined. This approach was first applied for the model on mechanical properties of titanium alloys (Ref 23). At first, the quest for the best solution can be based on a simple grid search, but the NN models can also be combined with efficient optimization methods for faster location of the best solution. The user can fix the other inputs (such as temperature, environment, and treatment) on a certain level and vary the composition of the alloy. The alloy composition can be altered by varying a single element concentration and fixing the others on chosen values or by varying the combination and ratios of elements. In this way, the user can analyze the influence of the different elements (or combination of elements) on the output properties and predict the optimal composition of the new alloy that would produce the desired properties. The use of an NN model in this mode would reduce the amount of experimentation for new alloy design.

Examples of this kind of software application include (Ref 23):

- Influence of aluminum and vanadium contents on the tensile properties of Ti-xAl-yV alloys
- Influence of aluminum, vanadium, oxygen, molybdenum, tin, and chromium contents on the isothermal transformation kinetics of various alloys
- Influence of aluminum content on yield strength, impact strength, and hardness of Ti-xAl-6V-2Sn alloys

Here, one of the modules (Fig. 1b, 2b) is used to demonstrate how an NN model can be used for alloy composition optimization. For this
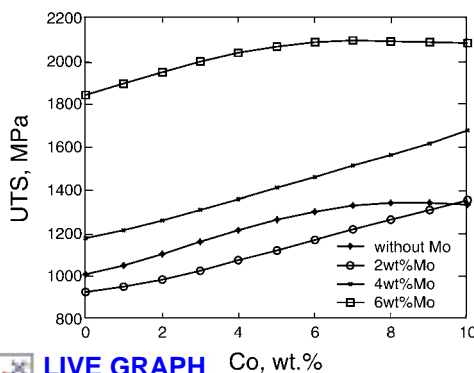
Fig. 12 Influence of cobalt and molybdenum on the ultimate tensile strength (UTS) of the Fe-18Ni-Co-Mo system
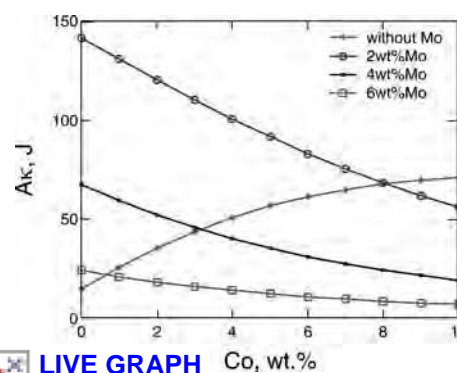
Fig. 13 Influence of cobalt and molybdenum on the Charpy impact energy (Aκ) value of the Fe-18Ni-Co-Mo system
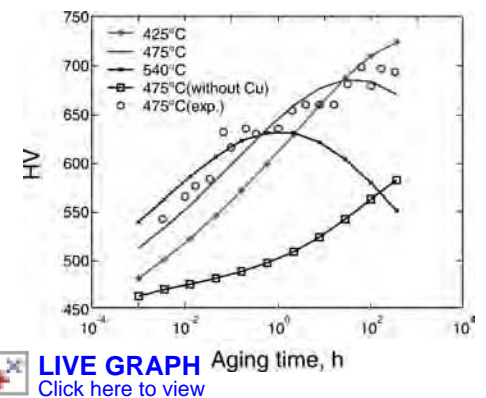
Fig. 14 Age-hardening kinetics curves of 1RK91 maraging steel simulated from a model calculation in comparison with experimental hardness measurement data. HV, Vickers hardness

purpose, the Ti-Al-Fe system is used, which is a base system for designing a new class of low-cost titanium alloys. By varying the alloying element contents, the program can be used to find the optimal alloy composition that would result in maximum yield strength (Fig. 15). Integration of various models would allow simultaneous optimization of different mechanical properties.

Industrial usage and the application of NN in real alloy design are discussed in Ref 1, 5, 27, 28, and 29.

Materials Selection. The materials selection procedure has become of extreme importance for engineering. Many software systems are developed for selection of materials for user-defined requirements. One of the popular systems of this kind is Cambridge Engineering Selector, developed initially as Cambridge Materials Selector (Ref 30). Most of these computer-based systems use the classical approach of searching in a database of materials properties according to user-defined criteria for a combination of properties.

A NN software system can also be used for the materials selection procedure. For this purpose, a special module (Fig. 8) and GUI (such as the one shown in Fig. 16) are needed. The user can use the GUI to define the working conditions (such as temperature and environment) and the requirements for materials properties (such as strength and elasticity). Once these are defined, the user can run the module by clicking on "search." The input information is acquired, organized, and processed to the modules with the trained NNs. Each input is processed only to the modules to which it is relevant. The trained NNs are the core of the computer program that searches for conditions to satisfy the user-defined criteria. The parameters that are fixed by the user are fixed inputs of the trained NN. The other inputs, essentially the material composition, are varied by organized loops, and the trained NNs are used to predict properties at each loop. The predicted combination of properties is compared with the user requirements at each step, and if they are satisfied, the current combination of material composition and processing parameters is suggested as a possible solution. The loops are organized so that all possible combinations between the changeable inputs are attempted. After the completion of this procedure, material compositions that satisfy the user requirements are suggested, if such are found. In the general case, a list of possible material compositions and processing is suggested.

In this mode, a NN model can be used for the materials selection procedure. For example, it can find material compositions that satisfy the user-defined combination of mechanical properties for any work temperature. The approach used here, based on trained ANNs, differs from the approaches that are based on a database search. In the latter, only material that exists in the database can be found and suggested as a material that satisfies the user criteria. In the NN approach, along with the existing materials, a composition of material that satisfies the criteria can be suggested, even if such a material does not exist. It should be restated that for this and all modes, an NN model works only within the ranges of input parameter values that are used for training.

Optimization of the Processing Parameters. Neural-network models can be used for optimization of the processing parameters of a material to achieve a desirable combination of properties at various working conditions. The alloy composition and the working conditions (such as working temperature and environment) should be fixed on the required values. By varying the processing conditions such as heat treatment, time, temperature, and cooling, the user can study their influence on the properties. The user can use the simulations to find optimal processing and heat treatment procedures that would produce the desired combination of properties for various working conditions and alloys. Examples for such usage of the programs are given in Ref 9, where the ANN model is used to study and optimize the influence of the processing parameters (time and temperature) on the microhardness profiles after surface nitrocarburizing of steels. An example is provided here that uses the model to simulate the influence of the processing route on the fatigue strength of corrosion-resistant Ti-6Al-2Sn-4Zr-6Mo alloy (Fig. 17). The alloy composition is fixed. The desired property (in this case, fatigue strength) is selected. This property is studied as a function of the
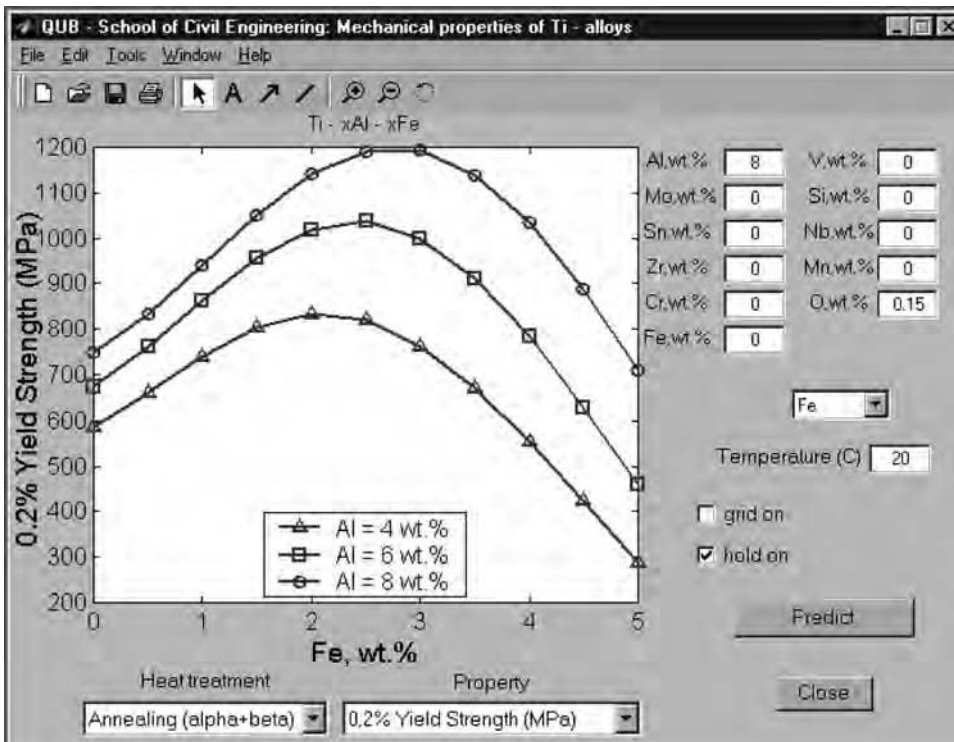
Fig. 15 Use of a neural-network model for optimization of alloy composition in the Ti-Al-Fe system to maximize room-temperature yield strength
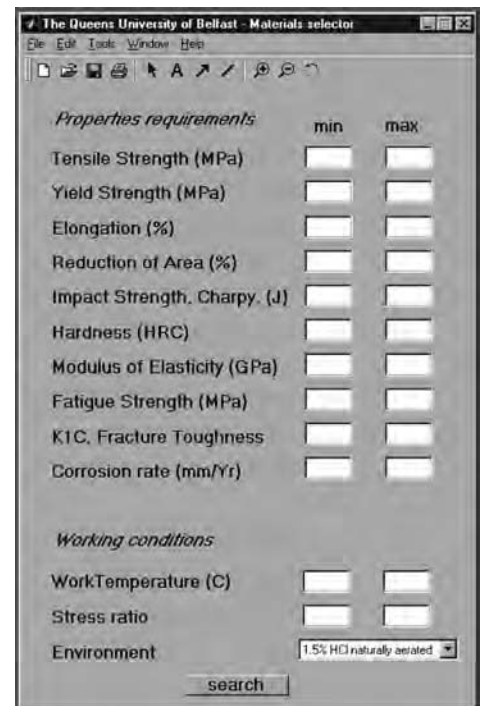
Fig. 16 Graphical user interface of a module for materials selection based on a trained artificial neural network

temperature at various processing procedures, namely, α + β annealing, solution treatment followed by aging, and duplex annealing. In a similar way, the modules can be used to study the influences of the processing route on other selected properties for various alloy compositions.

Apart from heat treatment, many other materials processing techniques can be studied and modeled by using ANNs, such as rolling, milling, and welding. For example, Dutta and Pratihar (Ref 31) have modeled the tungsten inert gas (TIG) welding process using conventional regression analysis and NN-based approaches (Fig. 18). In their work, statistical analyses were carried out on experimental data for the TIG welding process (obtained from published literature) to find its input-output relationships. The performances of the conventional regression analysis approach, a back-propagation NN, and a genetic-neural system were compared on some randomly generated experimental test cases, which are different from the training cases. The NN-based approaches could yield predictions that are more adaptive in nature compared to those of the more conventional regression analysis approach.

Theoretical Studies. Neural networks can also assist theoretical studies in materials science, although this is not as common as application-oriented work. Although often regarded as a black box or incorrectly as a model without relationships being expressed using equations, an NN actually has exact mathematical relationships between the neuron connections. Such relationships can be given explicitly using mathematical equations. An example is the work in fracture mechanics, expressing the explicit formulation of stress-intensity factor (K_I) (Fig. 19) based on NN analysis (Ref 32). In Fig. 19, "Type" refers to crack type, including center-cracked geometry, double-cracked geometry, and single-cracked geometry; w is the crack width; a is the crack length; and $\sigma_{applied}$ is the applied stress. Explicit formulations for K_I values are obtained using the parameters of the trained NN. Some numerical applications are performed to show the generalization capability of the trained NN. Formulation of the stress-intensity factor for three different geometries commonly used in fracture mechanics is obtained. It is shown that the results of the explicit formulation are in good agreement with finite-element calculations, which determine the stress-intensity factor using the displacement extrapolation method, ANSYS, and previous work in the literature for those common cases.

Upgrading Software Systems by Database Enhancement and Retraining

A NN model will work with sufficient accuracy within the range of the dataset used in training. If the user defines input conditions outside this range, a software system will still generate an output, but it may not be correct. In such a case, the program should provide a warning message that the result may not be reliable. Hence, a software system is restricted to predict with a reliable accuracy within certain ranges of input parameter values that are limited by the database used for training. The strategy should be to create general models that will be open for constant upgrade and improvement. Following this goal, modules should be developed that allow the user to input his own data and to retrain the ANNs (left side of Fig. 8).

The input of new data can be organized in two ways. The first option is by simple addition of new data in the spreadsheet files. Since the files are dynamically linked, the new data are automatically updated in the ASCII file and introduced in the matrix **M** that will be used for retraining. This option is more convenient when new data are inputted in batches and the user already has the data in electronic form (as a spreadsheet). The second option is by adding data one by one, using the purposely developed GUI. This option is more convenient when a single data pair is inputted.

Once the new data addition process is completed, the modules should be retrained so they can absorb the information that is contained in the new data. This is organized by the GUI for selection of the retraining parameters. There are three options for retraining:

• Simple retraining with the optimal combination of network architecture and training
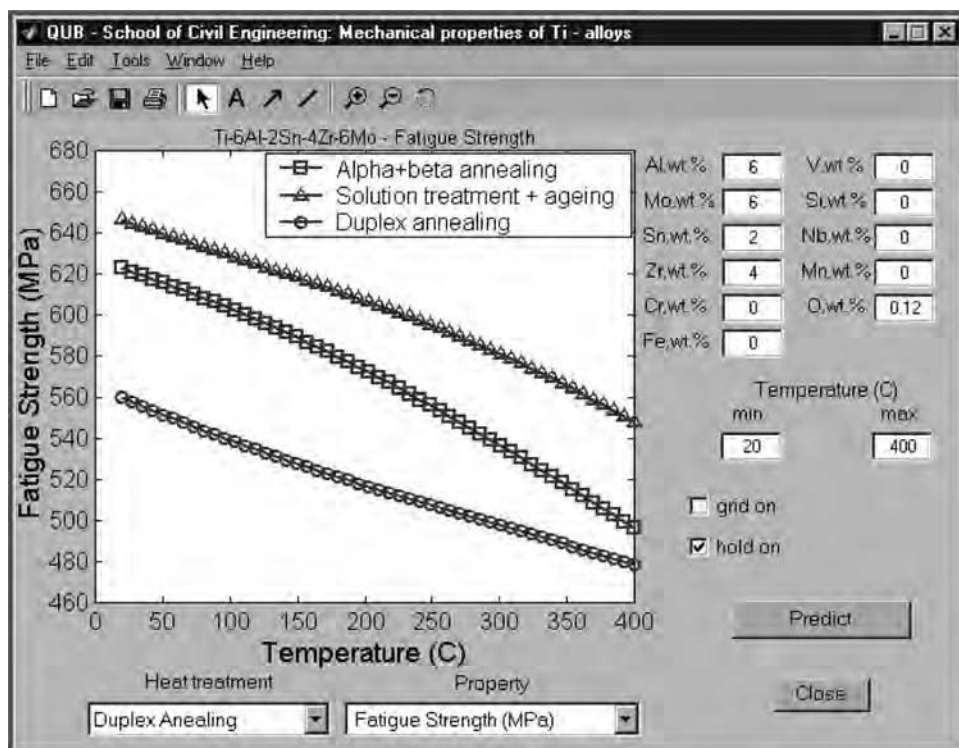
Fig. 17 Use of a neural-network model for optimization of the processing procedure
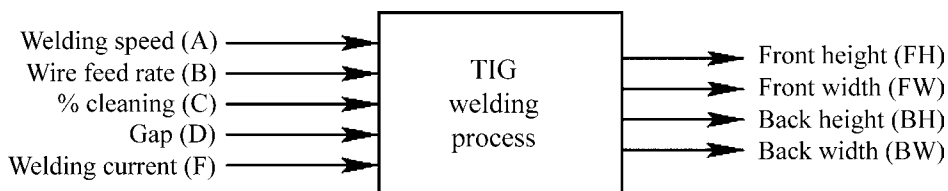
Fig. 18 Input and output variables of the tungsten inert gas (TIG) welding process. Source: Ref 31
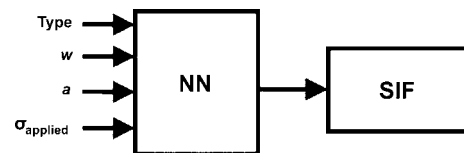
Fig. 19 Neural network (NN) for stress-intensity factor (SIF) calculation. "Type" refers to crack type, including center-cracked geometry, double-cracked geometry, and single-cracked geometry; w, crack width; a, crack length; and $\sigma_{applied}$, applied stress. Source: Ref 32

algorithm found in the sections on training in this article. This is the fastest retraining procedure. However, this option is recommended only in cases where the amount of new data is relatively small. Input of large amounts of new data could require a different NN architecture and training algorithm.

- User-defined parameters for retraining. In this option, the user can define the number of the neurons in the hidden layer, the training algorithm, the transfer functions, and so on. Advanced users who are familiar with ANN modeling should use this option. Some recommendations for selection are given in the section "Training Parameters" in this article.
- Retraining with all possible network architecture and training algorithms for finding the optimal combination. In this option, the program executes the entire procedure described in the section "Algorithm of Computer Program for NN Training" to find the optimal case of trained NN. This is the slowest option that requires significant computational time for retraining, but it is still recommended because, in this option, the user can be sure that the best possible combination in terms of database distribution, network architecture, and learning rule is achieved.

Organized in this way, a software system effectively accumulates and intelligently stores experience and will constantly be improved and expanded to serve the user in the best possible way. The NN developer should consider these options as well when updating a model with new data.

Summary

Neural-network models are powerful tools for simulation and prediction of materials properties and correlations among processing parameters, microstructure, and properties. The applications of NN models include:

- Prediction of properties of existing materials as functions of their chemical composition, heat treatment conditions, and working temperature; easy comparison of properties corresponding to different conditions by plotting them together; and investigation of the influence of various factors on the properties
- Simulation of TTT diagrams as a function of alloy chemical composition and design of diagrams for new alloys, and investigation of the influence of various alloying elements on phase-transformation kinetics
- Simulation of S-N curves under different input conditions and practical applications in solving various problems of fatigue behavior, and investigation of the influence of different factors affecting fatigue
- Significant reduction of the experimental work for measurement of TTT diagrams, S-N fatigue strength diagrams, and other mechanical properties
- New alloys design
- Materials selection
- Practical optimization of the alloy composition and processing parameters

The aforementioned is far from an exhaustive list, because it only contains exemplars. Neural-network models can benefit the industry, especially metals and alloys manufacturing companies. Graphical user interfaces facilitate easy use of the models. Neural-network models are convenient tools to obtain the desired combination of properties at various working temperatures, with minimal investments for experimental and plant trial work.

REFERENCES

1. H.K.D.H. Bhadeshia, Design of Ferritic Creep-Resistant Steels, *ISIJ Int.*, 41 (6), 2001, p 626–640
2. C.Z. Huang, L. Zhang, L. He, J. Sun, B. Fang, B. Zou, Z.Q. Li, and X. Ai, A Study on the Prediction of the Mechanical Properties of a Ceramic Tool Based on an Artificial Neural Network, *J. Mater. Process. Technol.*, 129 (1–3), 2002, p 399–402
3. K.G. Keong, W. Sha, and S. Malinov, Artificial Neural Network Modelling of Crystallization Temperatures of the Ni-P Based Amorphous Alloys, *Mater. Sci. Eng. A*, 365A (1–2), 2004, p 212–218
4. J. Wang, P.J. van der Wolk, and S. van der Zwaag, On the Influence of Alloying Elements on the Bainite Reaction in Low Alloy Steels during Continuous Cooling, *J. Mater. Sci.*, 35 (17), 2000, p 4393–4404
5. D. Cole and H.K.D.H. Bhadeshia, Design of Creep-Resistant Steel Welds, *Mathematical Modelling of Weld Phenomena 5*, H. Cerjak, Ed., London, Institute of Materials, 2001, p 431–448
6. Z. Guo and W. Sha, Modelling the Correlation between Processing Parameters and Properties of Maraging Steels Using Artificial Neural Network, *Comput. Mater. Sci.*, 29 (1), 2004, p 12–28
7. S.H. Lalam, H.K.D.H. Bhadeshia, and D.J.C. Mackay, The Charpy Impact Transition Temperature for Some Ferritic Steel Welds, *Aust. Weld. J.*, 45 (1), 2000, p 33–37
8. S.H. Lalam, H.K.D.H. Bhadeshia, and D.J.C. Mackay, Estimation of Mechanical Properties of Ferritic Steel Welds. I: Yield and Tensile Strength, *Sci. Technol. Weld. Join.*, 5 (3), 2000, p 135–147
9. T. Malinova, S. Malinov, and N. Pantev, Simulation of Microhardness Profiles for Nitrocarburized Surface Layers by Artificial Neural Network, *Surf. Coat. Technol.*, 135 (2–3), 2001, p 258–267
10. T. Malinova, N. Pantev, and S. Malinov, Prediction of Surface Hardness after Ferritic Nitrocarburising of Steels Using Artificial Neural Networks, *Mater. Sci. Technol.*, 17 (2), 2001, p 168–174
11. E.A. Metzbower, J.J. Deloach, S.H. Lalam, and H.K.D.H. Bhadeshia, Neural Network Analysis of Strength and Ductility of Welding Alloys for High Strength Low Alloy Shipbuilding Steels, *Sci. Technol. Weld. Joining*, 6 (2), 2001, p 116–124
12. E.A. Metzbower, J.J. Deloach, S.H. Lalam, and H.K.D.H. Bhadeshia, Analysis of Toughness of Welding Alloys for High Strength Low Alloy Shipbuilding Steels, *Sci. Technol. Weld. Joining*, 6 (6), 2001, p 368–374
13. M.A. Yescas, H.K.D.H. Bhadeshia, and D.J.C. Mackay, Estimation of the Amount of Retained Austenite in Austempered Ductile Irons Using Neural Networks, *Mater. Sci. Eng. A*, 311A (1–2), 2001, 162–173
14. O. Gundersen, A.O. Kluken, O.R. Myhr, J.E. Jones, V. Rhoades, J. Day, J.C. Jones, and B. Krygowski, The Use of an Integrated Multiple Neural Network Structure for Simultaneous Prediction of Weld Shape, Mechanical Properties, and Distortion in 6063-T6 and 6082-T6 Aluminium Assemblies, *Mathematical Modelling of Weld Phenomena 5*, H. Cerjak, Ed., London, Institute of Materials, 2001, p 671–700
15. H.K.D.H. Bhadeshia, Neural Networks in Materials Science, *ISIJ Int.*, 39 (10), 1999, p 966–979
16. A. Zhecheva, S. Malinov, and W. Sha, Simulation of Microhardness Profiles of Titanium Alloys after Surface Nitriding Using Artificial Neural Network, *Surf. Coat. Technol.*, 200 (7), 2005, p 2332–2342
17. Z. Guo and W. Sha, Modelling of Beta Transus Temperature in Titanium Alloys Using Thermodynamic Calculation and Neural Networks, *Titanium'99: Science and Technology, Proceedings of the Ninth World Conference on Titanium*, I.V. Gorynin and S.S. Ushkov, Ed., St. Petersburg, Russia, Central Research Institute of Structural Materials, 2000, p 61–68
18. S. Malinov and W. Sha, Application of Artificial Neural Networks for Modelling Correlations in Titanium Alloys, *Mater. Sci. Eng. A*, 365A (1–2), 2004, p 202–211
19. S. Malinov, W. Sha, and Z. Guo, Application of Artificial Neural Network for Prediction of Time-Temperature-Transformation Diagrams in Titanium Alloys, *Mater. Sci. Eng. A*, 283A (1–2), 2000, p 1–10
20. S. Malinov, W. Sha, and J.J. McKeown, Modelling the Correlation between Processing Parameters and Properties in Titanium Alloys Using Artificial Neural Network, *Comput. Mater. Sci.*, 21 (3), 2001, p 375–394
21. S. Malinov, T. Malinova, and W. Sha, Application of Artificial Neural Networks for Modelling the Mechanical Properties of γ-TiAl Based Alloys, *Proceedings of Annual Scientific Session*, The Technical University of Varna, Bulgaria, 2001, p 35–40

22. S. McShane, S. Malinov, J.J. McKeown, and W. Sha, Simulation of Fatigue Stress Life (S-N) Diagrams for Ti-6Al-4V Alloy by Application of Artificial Neural Network, *Computational Modelling of Materials, Minerals, and Metals Processing*, M. Cross, J.W. Evans, and C. Bailey, Ed., Pennsylvania, TMS, 2001, p 653–662

23. W. Sha and S. Malinov, *Titanium Alloys: Modelling of Microstructure, Properties and Applications*, Cambridge, Woodhead Publishing Limited, 2009

24. Mathworks, Natick, MA, available at http://www.mathworks.com/company, accessed February 2009

25. S. Mandal, P.V. Sivaprasad, S. Venugopal, K.P.N. Murthy, and B. Raj, Artificial Neural Network Modeling of Composition-Process-Property Correlations in Austenitic Stainless Steels, *Mater. Sci. Eng. A*, 485 (1–2), 2008, p 571–580

26. Sandvik Steel, *Data Sheet–Precision Wire Products*, Sandviken, Wire Division, 2001

27. H.K.D.H. Bhadeshia, Design of Creep-Resistant Steels, *Proceedings of Ultra-Steel 2000*, Tsukuba, Japan, National Research Institute for Metals, 2000, p 89–108

28. F. Tancret, Quantitative Modelling Tools for the Design of Ni-Base Superalloys, "Processing for China" Sterling Publications, London, 2000, p 56–58

29. J. Warde and D.M. Knowles, Use of Neural Networks for Alloy Design, *ISIJ Int.*, 39 (10), 1999, p 1015–1019

30. Granta Design, available at http://www.grantadesign.com, accessed February 13, 2009

31. P. Dutta and D.K. Pratihar, Modeling of TIG Welding Process Using Conventional Regression Analysis and Neural Network-Based Approaches, *J. Mater. Process. Technol.,* 184 (1–3), 2007, p 56–68

32. M.A. Kutuk, N. Atmaca, and I.H. Guzelbey, Explicit Formulation of SIF Using Neural Networks for Opening Mode of Fracture, *Eng. Struct.*, 29 (9), 2007, p 2080–2086

Modeling of Microstructure Evolution during the Thermomechanical Processing of Nickel-Base Superalloys

J.P. Thomas, ATI Allvac, an Allegheny Technologies company
F. Montheillet, CNRS, Ecole Nationale Supérieure des Mines de Saint-Etienne
S.L. Semiatin, Air Force Research Laboratory

THE MODELING OF MICROSTRUCTURE EVOLUTION during the processing of nickel-base superalloys aims to address the challenges posed by the variety of manufacturing operations by which parts are produced. These processes are chosen and carefully controlled for a specific end application. For instance, models have been developed to describe microstructure development during the manufacture of directionally solidified superalloy blades for aircraft engines and its subsequent evolution in service as well as the consolidation of powder used to produce engine disks of heavily alloyed grades with very tight tolerances (Ref 1). This article focuses solely on the conventional cast-and-wrought manufacturing route for superalloys, which typically comprises the conversion of large ingots into semifinished mill products that are then made into shaped components via forging, sheet forming, and other metalworking operations (Ref 2). This approach is similar to that used for a large number of metallic alloys and is commonly referred to as thermomechanical processing (TMP) (Ref 3). It is usually the most cost-effective method and is thus used for materials applications in the aerospace market, energy production, chemical industry, oil and gas extraction and refining, and nuclear plants.

In this article, the general features of microstructure evolution during the TMP of nickel-base superalloys and the challenges posed by the modeling of such phenomena are summarized first. In subsequent sections, various modeling methodologies are described. These comprise so-called JMAK (Avrami) models, topological models, and last, mesoscale physics-based models.

Additional related information on the general formulation of models for microstructure evolution is contained in the articles "Models of Recrystallization," "Cellular Automaton Models of Recrystallization," "Monte Carlo Models for Grain Growth and Recrystallization," and "Network and Vertex Models for Grain Growth" in this Volume.

Overview of Microstructure Evolution in Nickel-Base Superalloys during Hot Working

Experimental observations of the various features and mechanisms of microstructure evolution during the TMP of superalloys (and similar materials) form the basis for the different types of models that have been developed to provide quantitative descriptions for scientific and engineering applications. Such observations and the conventional terminology used to describe them are summarized in this section.

Recrystallization Mechanisms. Deformation at high temperature, usually at and above 70% of the melting point (on the absolute temperature scale), triggers a marked transformation of microstructure called recrystallization. When it happens during deformation, it is called dynamic recrystallization. When it continues immediately after deformation due to the energy stored and mechanisms initiated during deformation, it is called metadynamic recrystallization. If it occurs after deformation with some delay (incubation) time or due to mechanisms such as static recovery, it is called static recrystallization (Ref 2, 3).

Recrystallization, be it static or dynamic, can be defined as the generation of new high-angle, mobile grain boundaries and their subsequent migration driven by stored energy in the form of dislocations that are free or arranged in subboundaries. In the wake of such migration, new grains virtually devoid of the defects previously generated by deformation are produced. For dynamic recrystallization, this microstructural process gives rise to a flow curve comprised of an initial strain-hardening transient, during which dislocation multiplication (mitigated partially by dynamic recovery) occurs, a peak stress, and flow softening, during which large numbers of dislocations are annihilated (Fig. 1). Recrystallization has been observed in many crystalline materials, including metals, minerals, and even ice.

Depending on the specific alloy and processing temperature, dynamic recrystallization can actually take one of two forms, denoted as discontinuous or continuous, which are differentiated by their topology and kinetics. Discontinuous dynamic recrystallization (DDRX), or the kind described in the previous paragraph, was observed and quantified first. It occurs in steels forged in the austenitic domain as well as in copper and brass alloys. During DDRX, new grains typically appear first at the periphery of the initial grains through some nucleation mechanism and form a so-called necklace topology (Fig. 2, 3). This necklace of recrystallized grains progresses until the initial grains have been entirely consumed (Fig. 4). After a large enough deformation (i.e., when the microstructure is entirely recrystallized), DDRX enters a dissipative steady-state regime in which the microstructure constantly renews itself by generating new boundaries and accompanying grains to replace those that disappeared, which allows the elimination of defects (dislocations and subboundaries) generated by the continued deformation. The flow stress then reaches a stable (steady-state) value (Fig. 5a). For the specific case in which the DDRX grain size is significantly larger than that of the initial microstructure, typically at least twice as much, DDRX may exhibit a synchronous behavior, because recrystallization is complete before a second generation of nuclei appears in the first recrystallized grains (Fig. 5b).
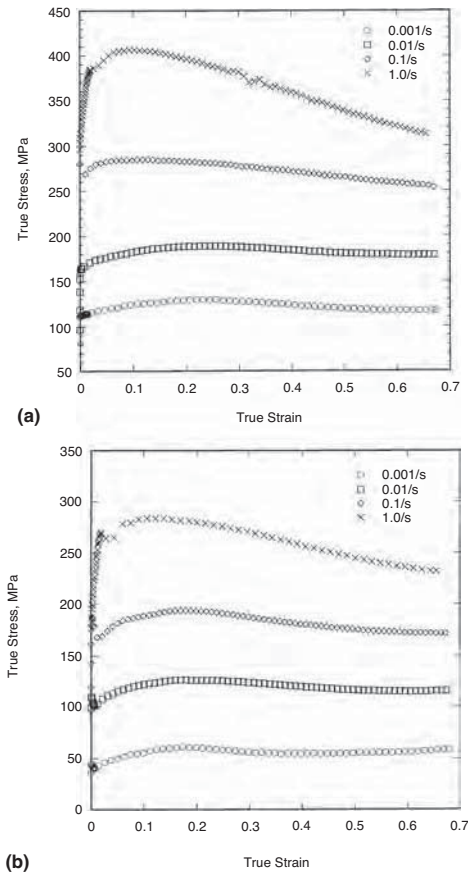
Fig. 1 True stress-true plastic strain curves obtained in compression of alloy 718 at various strain rates and test temperatures of (a) 975 °C or (b) 1050 °C. Source: Ref 4

LIVE GRAPH
Click here to view

LIVE GRAPH
Click here to view

The introduction of the term *discontinuous* to describe dynamic recrystallization was not necessary and did not come into use until relatively recently. It followed the extensive investigation of the recrystallization of minerals, ferritic steels, and aluminum alloys, especially via crystallographic-orientation mapping based on electron backscatter diffraction (EBSD) in scanning electron microscopes. Observations revealed another form of dynamic recrystallization that was termed continuous (CDRX). When CDRX occurs during high-temperature deformation, the formation of new high-angle boundaries happens through the progressive and widespread disorientation of subboundaries within the initial grains. It thus appears to transform the microstructure continuously from initial to recrystallized states. In CDRX, the definition or measurement of a recrystallized fraction, typically done for materials undergoing DDRX, is thus not possible.

Effect of Stacking-Fault Energy on Dynamic Recrystallization. The type of recrystallization exhibited by a specific material is related to its stacking-fault energy (SFE). In materials with high SFE, dynamic recovery occurs readily because dislocations are not
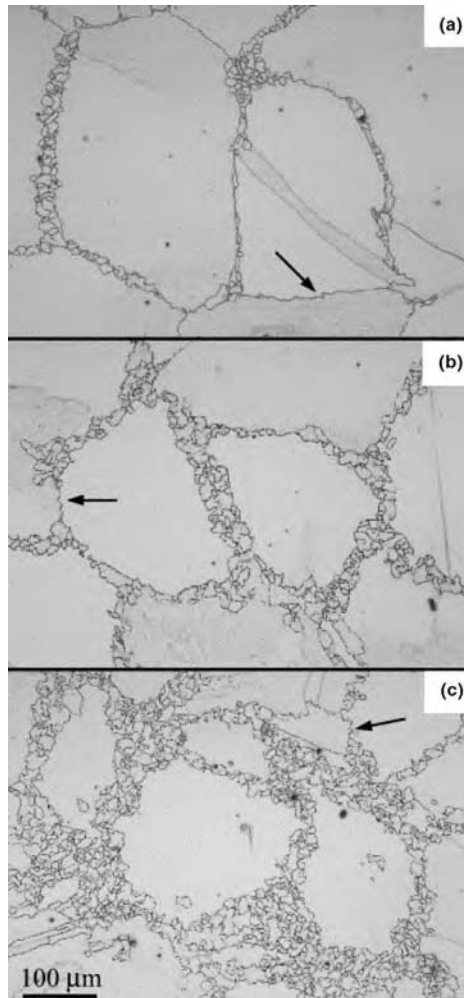
Fig. 2 Microstructures developed in Nimonic 80A deformed at 1055 °C and 0.1 s⁻¹ to strains of (a) 0.37, (b) 0.73, and (c) 1.13. Arrows indicate bulging of the initial grain boundaries. Source: Ref 5

dissociated and therefore may climb readily. Climb enables the annihilation of dislocations of opposite sign, thereby limiting significantly the increase of stored energy. It also promotes the accumulation of geometrically necessary dislocations in subboundaries, which translates in their disorientation. Such processes may delay or inhibit typical nucleation of DDRX (it requires high stored energy to ensure nuclei stability and growth) and hence may activate CDRX instead. Aluminum exhibits a high SFE, especially in comparison to austenitic steels or carbon steels in the austenite phase field (Table 1). For copper, alloying additions dramatically lower its SFE. Pure copper, with a moderate SFE, undergoes DDRX, probably because pure metals exhibit a high grain-boundary mobility, which eases the bulging of initial grain boundaries, one of the precursors of nucleation (Fig. 6).

When SFE is low, twin boundaries (whose surface energy is approximately one-half of

the SFE) (Ref 9) tend to be created behind migrating grain boundaries as a fault when the new crystal grows, because the energy difference between the front of the migrating boundary (containing dislocations) and the rear (a defect-free crystal) is still largely beneficial, even when a faulty twin boundary is formed. Such twin boundaries commonly close the protuberances formed by bulging, which effectively isolates the nuclei (Fig. 6c). Continuing deformation disorients twin boundaries into ordinary mobile grain boundaries. The same reasoning applies for pure nickel to explain its DDRX behavior. Depending on the specific alloying addition (Table 1), however, the SFE of nickel-base alloys may not drop significantly. This tends to suggest that some nickel alloys could exhibit significant dynamic recovery with grain-boundary mobility that is lower than that of pure nickel. Such characteristics would delay nucleation and promote the mechanism represented in Fig. 6(b). As a matter of fact, careful examination of Fig. 2 reveals the presence of nuclei inside the initial grains. These observations suggest that nuclei can be generated in nickel-base superalloys by a subboundary disorientation process through a localized process of continuous recrystallization. The use of crystallographic orientation mapping by EBSD validates such observations (Fig. 7). In the case of very coarse ingot structures, similar behavior is observed. In addition, second-phase particles such as carbides can trigger particle-stimulated nucleation (PSN) (Fig. 8). The PSN can dramatically increase the density of sites at which recrystallization can be initiated and therefore the overall recrystallization kinetics.

In summary, the dynamic recrystallization of nickel-base superalloys exhibits features that belong to both the discontinuous and continuous behaviors. Partially recrystallized microstructures reveal the necklace topology typical of DDRX, often with numerous twin boundaries. In addition, islands of recrystallized grains may develop inside the initial grains due to the local disorientation of subboundaries arising from deformation incompatibility between neighboring grains or between the matrix and coarse particles. These phenomena seem to operate at a level of SFE that separates discontinuous and continuous behaviors (Fig. 9).

Effect of Process Variables on Dynamic Recrystallization. The various factors that control dynamic recrystallization, from dislocation generation and recovery to nucleation and grain-boundary migration, are thermally activated. Therefore, an increase in temperature induces an acceleration of such mechanisms and the overall recrystallization kinetics. Similarly, for a given temperature, if the deformation is applied at a lower strain rate, thermally activated mechanisms are given more time to operate, resulting in higher recrystallized fractions in the material. The combined effect of temperature and strain rate is often quantified by the Zener-Hollomon parameter (Z) (Ref 11):

$$Z = \dot{\varepsilon} \, \exp\left(\frac{Q}{RT}\right) \qquad \text{(Eq 1)}$$

in which Q is the apparent activation energy of deformation, R is the gas constant, and T is the absolute temperature. The value of Q is typically determined from the steady-state flow stress, σ_s, and the expression:

$$\sigma_s = k \, Z^m \qquad \text{(Eq 2)}$$

in which k is a constant for a given material, and m is the strain-rate sensitivity of the fully recrystallized material.

The Zener-Hollomon parameter is extremely important, inasmuch as it describes well the dependence of numerous parameters on strain rate and temperature. For example, the recrystallized grain size during steady-state flow, d_s, can generally be fit by a relation of the form:

$$d_s = k_{ds} \, Z^{-m_{ds}} \qquad \text{(Eq 3)}$$

in which k_{ds} and m_{ds} are material constants. These expressions enable the interpretation of behavior for different alloys using a unified approach. For example, the ratio m/m_{ds} is typically close to $\frac{2}{3}$ for a wide range of materials (Ref 12, 13). Any significant deviation from these expressions should trigger more detailed analysis of measurement techniques and analysis (e.g., accounting for friction, deformation heating, a delay between the end of deformation and the actual quench of samples) (Ref 14) before investigating an atypical microstructural response. When such verifications are done and the reason for any deviation is understood, corrected, or compensated for, it is possible to derive a reliable microstructure-evolution model to replicate experimental observations.

Modeling Challenges

It is important that the time allocated and the goals of microstructure modeling are well established, because each of the various possible approaches presents its own advantages, challenges, and intrinsic limitations. The main challenge in recrystallization modeling lies in the superposition of length scales at which recrystallization happens and expresses itself, namely, the macroscopic, mesoscopic, and microscopic scales. The macroscopic scale is that of the sample or workpiece upon which deformation is applied. The material opposes the applied load through its flow stress integrated over the surface of contact with the tools that impose the deformation. The flow stress is well defined at the mesoscopic scale, that is, that of a grain or a few grains. For small enough samples or regions that deform
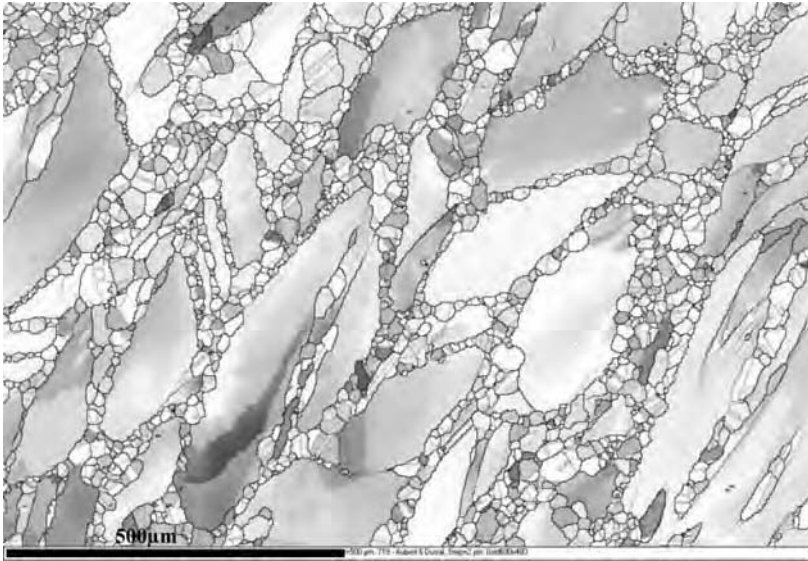
Fig. 3 Electron backscatter diffraction inverse-pole-figure map for alloy 718 with an initial grain size of 200 μm deformed in torsion at 980 °C, 1 s⁻¹ to an effective strain of 0.7 and then held for 140 s. Boundary misorientations are indicated in light gray (2°) to dark gray or red (12°) to black (≥15°); twin boundaries are light gray or green. Source: Ref 6
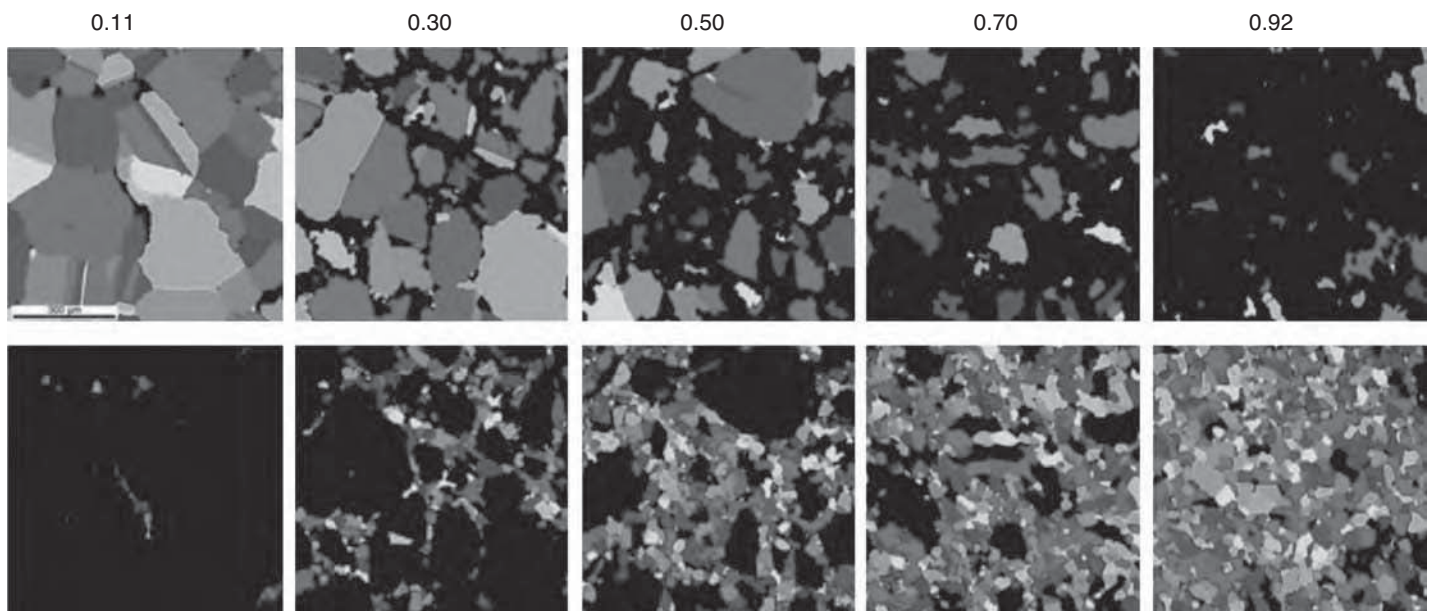
Fig. 4 Electron backscatter diffraction inverse-pole-figure maps for Nimonic 80A showing the deformed (top) and recrystallized (bottom) fraction as a function of the indicated strains. The black areas represent the corresponding second fraction. Coherent twins have been removed. The width of each image is 809 μm. Source: Ref 7
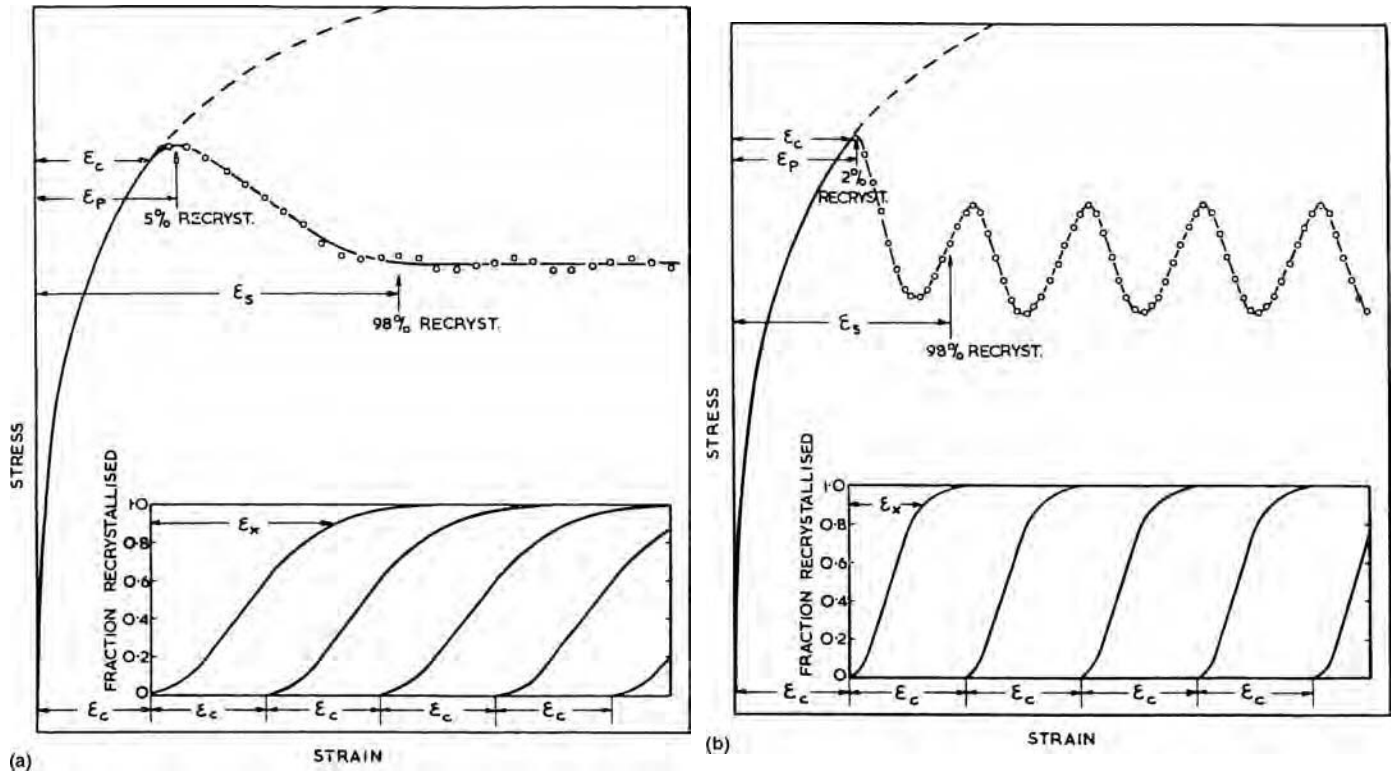
Fig. 5 Calculated stress-strain curves showing successive waves of recrystallization. (a) Asynchronous discontinuous dynamic recrystallization (DDRX). (b) Synchronous DDRX. Source: Ref 8

Table 1 Stacking-fault energies; γ_{SFE}

Metal or alloy	γ_{SFE}, mJ/m^2
Cu	78
Cu-5wt%Al	20
Fe-19wt%Cr-9wt%Ni	21
Fe-18wt%Cr-14wt%Ni	40
Fe-18wt%Cr-19wt%Ni	50
Al	166
Al-1wt%Si	145
Ni	128
Ni-20wt%Co	120
Ni-20wt%Cr	40

Source: Ref 9
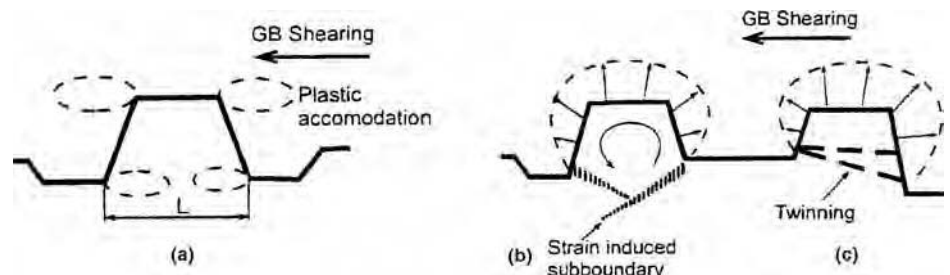
Fig. 6 Schematic representation of nucleation by bulging. (a) A protuberance accentuates local heterogeneities of deformation. It is closed by (b) a subboundary or (c) a twin boundary. GB, grain boundary. Source: Ref 10

homogeneously, the mesoscopic flow stress should be identical to the macroscopic measurement. Nevertheless, one may be interested in the precise distribution of strain and flow stress within each grain to be able to quantify how the incompatibility of deformation between neighboring grains is accommodated. In this case, the strain and flow stress are local, microscopic values.

All mechanisms that govern recrystallization (dislocation generation, recovery, nucleation, grain-boundary migration) as well as the precipitation of secondary phases operate at a microscopic scale. However, their behavior is generally averaged over entire grains in order to deal with them in a mesoscopic fashion. This requires the definition of overall driving forces that depend on mesoscopic quantities such as

grain size, which enables the effect of mechanical interaction with neighboring grains, for instance, to be taken into account, at least approximately. The common expression of these microscopic mechanisms through the progress of recrystallization is inherently mesoscopic because it transforms the grain structure. The modeling of microstructure evolution in general and recrystallization in particular during hot deformation therefore must address two main issues: how to translate microscopic mechanisms into somewhat mesoscopic driving forces, and how to represent combined geometric and topological effects on the evolution of grain structure. The latter is controlled by the sites of nucleation of recrystallized grains and the migration of their boundaries, whether they interface with the initial grains (their behavior then defines the

progress of recrystallization) or with other recrystallized grains (resulting in impingement).

There are three methods to address these challenges: JMAK, topological, and mesoscale physics-based models. The first approach was proposed by Johnson, Mehl, Avrami, and Kolmogorov (Ref 15–19), whose work led to the JMAK, or Avrami, equation, to describe static phase transformations as a function of time. The JMAK equation is readily applied to static recrystallization, for example, by treating the recrystallized grains as a new phase that grows into the initial, strained structure. A number of enhancements to the classical JMAK formalism were proposed throughout the second half of the last century to address the modeling of dynamic and metadynamic recrystallization. With these changes, the
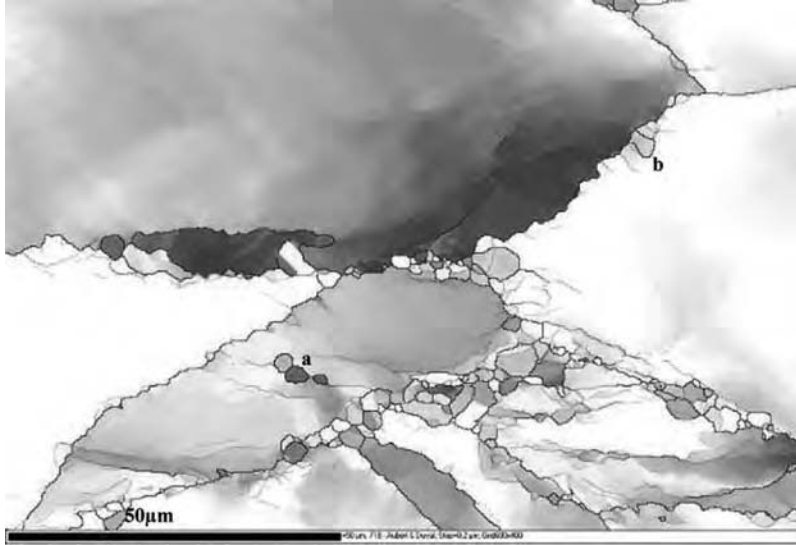
Fig. 7 Electron backscatter diffraction inverse-pole-figure map for alloy 718 with an initial grain size of 50 μm deformed in uniaxial compression at 980 °C, 0.01 s^{-1} to a true strain of 0.4. Boundary misorientations are indicated in light gray (2°) to dark gray or red (12°) to black (≥15°); twin boundaries are light gray or green. Nuclei that appeared inside an initial grain are indicated by the letter "a". A nucleus formed by the bulging of an initial grain boundary is indicated by the letter "b". Source: Ref 6

physical basis of the basic model tends to become more questionable, thus leading to JMAK approaches for dynamic and metadynamic recrystallization that are largely phenomenological. The second method, comprising topological models, includes discrete (cellular automata, Monte Carlo) and continuous (vertex networks) techniques. They can be physics based but tend to be computationally intensive. The third method, mesoscale physics-based models, provides an approach that attempts to combine simplicity and a light computational footprint, characteristic of JMAK approaches, while retaining a sound physical basis.

JMAK Models

The basis of the formulation developed by Johnson, Mehl, Avrami, and Kolmogorov (Ref 15–19) lies in the definition of an extended volume, denoted X_{ext}, as that of the total volume occupied by recrystallized grains should their growth not be constrained by impingement (Ref 20).

Fundamentals of JMAK Models. From a physical standpoint, recrystallized regions cannot grow and increase their volume over regions that are already recrystallized. Hence, at any specific time, only a fraction of the increase of the extended volume translates into an increase of the actual volume of recrystallized regions. Assuming a perfectly random distribution of nucleation sites in the material, this fraction is that of the unrecrystallized regions. In mathematical terms, the rate of increase of the recrystallized fraction X is given by:

$$\dot{X} = \dot{X}_{ext}(1 - X) \tag{Eq 4}$$

Integration with respect to time yields the relation:

$$\ln(1 - X) = -X_{ext} \tag{Eq 5}$$

or

$$X = 1 - \exp(-X_{ext}) \tag{Eq 6}$$

For the case of static recrystallization (SRX) in which the grain-boundary velocity for the growth of recrystallized grains is a constant denoted as v, the recrystallized grain size, D, in the absence of impingement is:

$$D = 2vt \tag{Eq 7}$$

If all the nuclei exist at the end of deformation with a volume density N, the evolution of the extended volume is then given by the relation:

$$X_{ext} = N\frac{\pi D^3}{6} = \frac{4\pi}{3}Nv^3t^3 \tag{Eq 8}$$

However, due to impingement, in accordance with Eq 6, the actual recrystallized fraction X is:

$$X = 1 - \exp\left(-\frac{4\pi}{3}Nv^3t^3\right) \tag{Eq 9}$$

The general form of the JMAK equation for recrystallized fraction as a function of time is:

$$X = 1 - \exp(-kt^n) \tag{Eq 10}$$

The constants k and n depend on the material, temperature, previous strain, initial grain size, and so on. The coefficient n, often called the Avrami exponent, varies between 2 and 5, depending on geometric considerations and the nucleation rate. For example, in the case of three-dimensional growth of pre-existing nuclei, as assumed previously, n is equal to 3. For a two-dimensional growth, n is equal to 2. If nuclei appear continuously during recrystallization, n can be greater than 3, but this is rarely observed experimentally (Ref 20).

In general then, the geometric (impingement-related) aspects of recrystallization are handled in JMAK models by an equation of the form in "1 − exp()"; topological effects are captured somewhat by the Avrami exponent n; and the effects of driving forces on recrystallization are represented by the kinetics of evolution of the extended volume.

JMAK Models for TMP. Specific JMAK-type expressions for dynamic, metadynamic, and static recrystallization during TMP are similar to Eq 10. For DDRX, which occurs during deformation, the JMAK equation is typically written as:

$$X = 1 - \exp\left[-\ln 2\left(\frac{\varepsilon - \varepsilon_c}{\varepsilon_{0.5} - \varepsilon_c}\right)^n\right] \tag{Eq 11}$$

In Eq 11, ε denotes the applied strain; $\varepsilon_{0.5}$ is the strain required at a specific temperature and strain rate for a given initial grain size to recrystallize one-half of the initial structure during deformation; and ε_c is the so-called critical strain at which sufficient energy is stored to initiate nucleation of recrystallization. In some cases, ε_c is not needed or is simply too low to be identified and is then assumed to be equal to zero (Ref 21).

For metadynamic recrystallization (MDRX), a time $t_{0.5}$, defined as the holding time after deformation at which one-half of the starting structure has recrystallized, is used in the corresponding JMAK relation:

$$X = 1 - \exp\left[-\ln 2\left(\frac{t}{t_{0.5}}\right)^n\right] \tag{Eq 12}$$

The value of $t_{0.5}$ for MDRX typically depends on the initial grain size, strain, strain rate, and the temperature at which the predeformation and subsequent holding time have been performed. The corresponding relation for static recrystallization is similar to Eq 12, with the time variables reduced by a factor of t_{inc}, the incubation time for SRX.

The value of the Avrami exponent n may be different for DDRX and MDRX. Moreover, n has a great effect on the shape of the sigmoidal curve described by the JMAK relation. The ability to fit a wide range of experimental observations is thus an important reason for the versatility of JMAK-type models.

For the DDRX of nickel-base superalloys, in particular, the Avrami exponent is usually ~2 to 3 (Tables 2–4). It may be greater for the case of coarse initial structures, as in the case of Waspaloy (UNS N07001) ingot material. For this material, an acceleration of recrystallization has been observed due to the
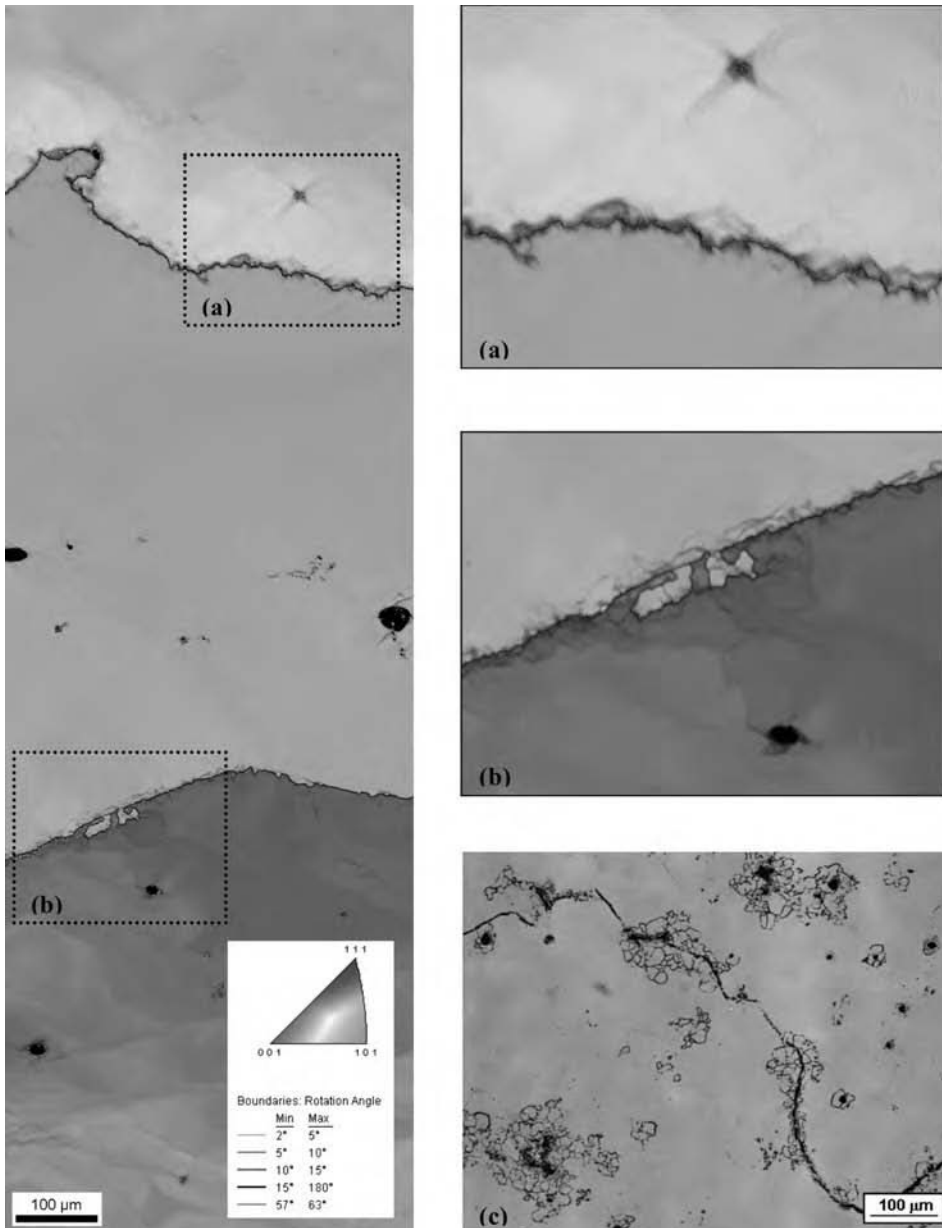
Fig. 8 (a),(b) Electron backscatter diffraction inverse-pole-figure map for Waspaloy ingot material that was deformed in uniaxial compression at 1120 °C, 0.01 s⁻¹ to a true strain of 0.27. (c) Optical micrograph showing Waspaloy ingot material that was deformed at 1120 °C, 0.1 s⁻¹ to a true strain of 1.51. Courtesy of the Air Force Research Laboratory.
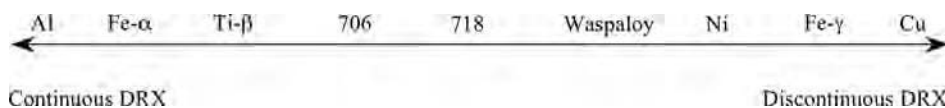
Fig. 9 Relative tendency of different metals and alloys to undergo discontinuous versus continuous dynamic recrystallization (DRX) during hot working

presence of additional nucleation sites within the grains associated with carbides that lead to PSN (Ref 24). It should also be remembered that values of n not readily explained by the original JMAK theory are commonly identified as the best fit. As an example, $n = 1$ is commonly found for MDRX of superalloys (Tables 2–4).

Instead of using $\varepsilon_{0.5}$ or $t_{0.5}$ as normalization factors in the JMAK relations, some models (such as those of Sellars for modeling the recrystallization of steels forged in the austenitic domain (Ref 25) make use of the values of the strain or time at which the microstructure is 95% recrystallized, that is, $\varepsilon_{0.95}$ or $t_{0.95}$. In such cases, the term "ln 2" in Eq 11 and 12 should be replaced by $\ln(1/0.05) \approx 2.996$. This formulation is not recommended for the JMAK modeling of nickel-base superalloys, however. By using such a reference state, one may assume that the structure is essentially fully recrystallized and overlook potentially deleterious unrecrystallized material. Such an assumption may be acceptable for steels. For high-integrity superalloys intended for aerospace applications, however, product specifications define not only the matrix (recrystallized) grain size but also the size of so-called as-large-as (ALA) grains, which are typically the remnants of partially recrystallized grains. Such ALA grains must be considered during component design, for they may seriously degrade properties such as fatigue behavior.

When large strains are applied, a large fraction of the microstructure may be dynamically recrystallized. The application of Eq 12 for subsequent metadynamic evolution typically reinitializes the recrystallized fraction to 0, which is inappropriate. In such cases, an alternate form of the JMAK formalism should be used:

$$X_{mdyn} = X_{dyn}$$
$$+ \left(1 - X_{dyn}\right) \left[1 - \exp\left[-\ln 2 \left(\frac{t}{t_{0.5}}\right)^n\right]\right]$$
(Eq 13)

in which X_{dyn} and X_{mdyn} are the dynamic and metadynamic recrystallized fractions, respectively (Ref 6). For workpieces that are very large, the application of Eq 13 may not yield significantly better results, because the typical times for metadynamic or static recrystallization are long enough and the cooling of the material is slow enough that modeling results should not be needed at, or shortly after, the completion of deformation. By contrast, for hot working operations consisting of cumulative deformations and considerable metadynamic evolution, such as multistand hot rolling, Eq 13 may provide more precise results, even though the strain applied at each pass is typically small and fast enough that the material may not recrystallize much at all during the deformation itself.

The quantitative dependence of $\varepsilon_{0.5}$ and $t_{0.5}$ on process variables is often well described using an Arrhenius relation for temperature, T (introducing an apparent activation energy for recrystallization), and power laws for the effects of strain, ε, strain rate, $\dot{\varepsilon}$, and initial grain size, D_0. From a mathematical standpoint, material-dependent coefficients k_d, a_d, m_d, and Q_d are introduced for dynamic recrystallization, and k_{md}, a_{md}, b_{md}, m_{md}, and Q_{md} for metadynamic recrystallization:

Table 2 JMAK model for dynamic and metadynamic recrystallization and grain growth for wrought Waspaloy

Zener-Hollomon parameter	$Z = \dot{\varepsilon}\,\exp(468{,}000/RT)$
Dynamic recrystallized fraction	$X = 1 - \exp(-\ln 2\,(\varepsilon/\varepsilon_{0.5})^n)$ $n = 3$ for $T < 1010\,^\circ\text{C}$ $n = 2$ for $1010\,^\circ\text{C} < T < 1027\,^\circ\text{C}$ $n = 1.8$ for $1027\,^\circ\text{C} < T$
Strain for dynamic recrystallized fraction of one-half	$\varepsilon_{0.5} = 0.1449\,d_0^{0.32}\,Z^{0.03}$ for $T < 1010\,^\circ\text{C}$ $\varepsilon_{0.5} = 0.056\,d_0^{0.32}\,Z^{0.03}$ for $1010\,^\circ\text{C} < T < 1027\,^\circ\text{C}$ $\varepsilon_{0.5} = 0.035\,d_0^{0.29}\,Z^{0.04}$ for $1027\,^\circ\text{C} < T$
Dynamic recrystallized grain size, μm	$d_{\text{dyn}} = 8103\,Z^{-0.16}$ for $T < 1027\,^\circ\text{C}$ $d_{\text{dyn}} = 108.85\,Z^{-0.0456}$ for $1027\,^\circ\text{C} < T$
Metadynamic recrystallized fraction	$X = 1 - \exp[-\ln 2\,(t/t_{0.5})^n]$ with $n = 1$ for $1027\,^\circ\text{C} < T$ No metadynamic recrystallization otherwise
Time(s) for metadynamic recrystallized fraction of one-half	$t_{0.5} = 4.54 \cdot 10^{-5}\,d_0^{0.51}\,\varepsilon^{-1.28}\,\dot{\varepsilon}^{-0.073}\,\exp(80{,}687/RT)$ for $1027\,^\circ\text{C} < T$ No metadynamic recrystallization otherwise
Metadynamic recrystallized grain size, μm	$d_{\text{m-dyn}} = 14.56\,d_0^{0.33}\,\varepsilon^{-0.44}\,Z^{-0.026}$ for $1027\,^\circ\text{C} < T$ Dynamic recrystallized grain size otherwise
Static grain growth, d, μm; t, s	$d^3 = d_{\text{ini}}^3 + 2 \cdot 10^{26}\,t\,\exp(-595{,}000/RT)$ for $1027\,^\circ\text{C} < T$ No static grain growth otherwise

Source: Ref 21

Table 3 JMAK model of dynamic, metadynamic, and static recrystallization and grain growth for superalloy 718

Zener-Hollomon parameter, (T), K	$Z = \dot{\varepsilon}\,\exp(508{,}000/RT)$
Critical strain	$\varepsilon_c = 2.74 \cdot 10^{-7}\,Z^{0.28}$ for $\dot{\varepsilon} > 0.01\,\text{s}^{-1}$ $\varepsilon_c = 9.112 \cdot 10^{-4}\,Z^{0.0982}$ for $\dot{\varepsilon} \le 0.01\,\text{s}^{-1}$
Dynamic recrystallized fraction	$X_{\text{drx}} = 1 - \exp[-\ln 2\,(\varepsilon - \varepsilon_c/\varepsilon_{0.5})^n]$ with $n = 2$
Strain for dynamic recrystallized fraction of one-half	$\varepsilon_{0.5} = 0.1343\,Z^{0.0515}$
Dynamic recrystallized grain size, μm	$d_{\text{drx}} = 1.0602 \cdot 10^5\,Z^{-0.185}$
Metadynamic recrystallized fraction	$X_{\text{mdrx}} = 1 - \exp[-\ln 2\,(t/t_{0.5})^n]$ with $n = 1$
Time for metadynamic recrystallized fraction of one-half	$t_{0.5} = 1.7 \cdot 10^{-5}\,d_0^{0.5}\,\varepsilon^{-2.0}\,\dot{\varepsilon}^{-0.08}\,\exp(12{,}000/RT)$
Metadynamic recrystallized grain size, μm	$d_{\text{mdrx}} = 8.28\,d_0^{0.29}\,\varepsilon^{-0.14}\,\dot{\varepsilon}^{-0.03}\,\exp(-13{,}440/RT)$
Static recrystallized fraction	$X_{\text{srx}} = 1 - \exp[-\ln 2\,(t/t_{0.5})^n]$ with $n = 1$
Time for static recrystallized fraction of one-half	$t_{0.5} = [-3.92 + 5508/T]\,\varepsilon^{-0.75}\,\exp(74{,}829/RT)$
Static recrystallized grain size, d, μm; T, K; t, s	$d_{\text{srx}} = 28\,(T/1223)^3\exp[3 \cdot 10^{-5}(T - 1223)\,t]$
Static grain growth, D, μm; t, s	$D_1^{15} = D_0^{15} + 5.63 \cdot 10^{20}\,t\,\exp(-115{,}000/RT)$ for $T < 1027\,^\circ\text{C}$, T (K) $D_{\text{tr}} = (D_1 - D_2)\,[\cos(\pi/2 \times T - 1017/1050 - 1017)]^{t/10^4} + D_2$ for $1027\,^\circ\text{C} < T < 1050\,^\circ\text{C}$, T ($^\circ$C) $D_2^{15} = D_0^{15} + 8.05 \cdot 10^{36}\,t\,\exp(-115{,}000/RT)$ for $T > 1050\,^\circ\text{C}$, T (K)
Unrecrystallized fraction	$X_n = 1 - X_{\text{drx}} - X_{\text{mdrx}} - X_{\text{srx}}$
Unrecrystallized grain size, μm	$d_n = d_0\,\exp(-\varepsilon/4)$ in which d_0 is the initial grain size
Average grain size, μm	$1/d_{\text{AV}}^2 = X_{\text{drx}}/d_{\text{drx}}^2 + X_{\text{mdrx}}/d_{\text{mdrx}}^2 + X_{\text{srx}}/d_{\text{srx}}^2 + X_n/d_n^2$
Retained strain	$\varepsilon_{\text{re}} = \varepsilon\,X_n$

Source: Ref 22

Table 4 JMAK model of metadynamic and static recrystallization and grain growth for superalloy 718

Strain at peak stress	$\varepsilon_p = 0.4659 \cdot 10^{-2}\,\dot{\varepsilon}^{0.1238}\,\exp(49{,}520/RT)$
Critical strain	$\varepsilon_c = 5/6\,\varepsilon_p$
Metadynamic or recrystallized fraction	$X_{\text{mdrx}} = 1 - \exp[-\ln 2\,(t/t_{0.5})^n]$ with $n = 1$
Time for metadynamic recrystallized fraction of one-half	$t_{0.5} = 5.043 \cdot 10^{-9}\,\varepsilon^{-1.42}\,\dot{\varepsilon}^{-0.408}\,\exp(196{,}000/RT)$
Metadynamic recrystallized grain size, μm	$d_{\text{mrx}} = 4.85 \cdot 10^{10}\,\varepsilon^{-0.41}\,\dot{\varepsilon}^{-0.028}\,\exp(-240{,}000/RT)$
Static recrystallized fraction	$X_{\text{srx}} = 1 - \exp[-\ln 2\,(t/t_{0.5})^n]$ with $n = 1$
Time for static recrystallized fraction of one-half	$t_{0.5} = 3.16\,\varepsilon^{-0.75}\,\exp(74{,}826/RT)$
Static recrystallized grain size, μm	$d_{\text{srx}} = 6.78 \cdot 10^2\,\exp(-31{,}710/RT)$
Static grain growth, d, μm; t, min	$d_{\text{gg}}^2 = d_0^2 + 9.44 \cdot 10^{19}\,t\,\exp(-467{,}114.7/RT)$

Source: Ref 23

$$\varepsilon_{0.5} = k_{\text{d}}D_0^{a_{\text{d}}}\dot{\varepsilon}^{m_{\text{d}}}\exp\left(\frac{Q_{\text{d}}}{RT}\right) \tag{Eq 14}$$

$$t_{0.5} = k_{\text{md}}D_0^{a_{\text{md}}}\varepsilon^{b_{\text{md}}}\dot{\varepsilon}^{m_{\text{md}}}\exp\left(\frac{Q_{\text{md}}}{RT}\right) \tag{Eq 15}$$

In some cases, the dependence of $\varepsilon_{0.5}$ on strain rate and temperature can be directly related to the Zener-Hollomon parameter by assuming that $Q_{\text{d}} = m_{\text{d}}Q$, thus yielding an expression that is simpler than Eq 14:

$$\varepsilon_{0.5} = k_{\text{d}}D_0^{a_{\text{d}}}Z^{m_{\text{d}}} \tag{Eq 16}$$

Expressions such as Eq 14 to 16 enable the determination of the material coefficients by performing multilinear regression analysis on plots of $\ln(-\ln(1 - X))$ as a function of the logarithm of the initial grain size, strain, strain rate, Zener-Hollomon parameter, and the inverse temperature. For example, for the case of DDRX, neglecting the critical strain ε_c, Eq 11 and 16 yield the following relation:

$$\ln(-\ln(1 - X)) = \ln(\ln 2) \\ + n[\ln \varepsilon - \ln k_{\text{d}} - a_{\text{d}}\ln D_0 - m_{\text{d}}\ln Z] \tag{Eq 17}$$

It is also possible to determine the material coefficients using readily available software tools such as the Solver routine in Microsoft Excel.

Grain-Size Predictions in JMAK Modeling. Similar combinations of Arrhenius and power laws are used in JMAK models to quantify the dependence of recrystallized grain size on process variables and the initial grain size, D_0. An expression for the recrystallized grain size d_s pertaining to the steady-state regime of DDRX was given earlier (Eq 3). While dynamic recrystallization is in progress, the instantaneous recrystallized grain size d is:

$$d = d_{\text{s}}X^\alpha = k_{\text{ds}}\,Z^{-m_{\text{ds}}}X^\alpha \tag{Eq 18}$$

in which α denotes a material constant between 0 and 1.

The size of grains during MDRX is also generally evaluated in two steps. First, their size at the completion of MDRX, d_{mdyn}, is determined using an expression of the form of Eq 19, in which k_{mdd}, p, q, r, and Q_{mdd} are material-dependent parameters. Then, the instantaneous grain size while MDRX progresses is given by Eq 20, in which β is another material constant, whose value is between 0 and 1:

$$d_{\text{mdyn}} = k_{\text{mdd}}D_0^p\varepsilon^q\dot{\varepsilon}^r\exp\left(-\frac{Q_{\text{mdd}}}{RT}\right) \tag{Eq 19}$$

$$d = d_{\text{mdyn}}X^\beta \tag{Eq 20}$$

Static grain growth is usually modeled in accordance with the following expression:

$$d^w = d_{\text{mdyn}}^w + K_{\text{g}}\exp\left(-\frac{Q_{\text{g}}}{RT}\right)t \tag{Eq 21}$$

in which K_g and Q_g are constants, and w is an exponent, typically ~ 2 or 3. Higher values of w are possible, for example, when second-phase particles slow growth. When second-phase particles essentially pin the growing grains, even very high values of w may not be sufficient, and different expressions may be needed altogether (Table 3). An expression of the form of Eq 21 can yield better results than Eq 19 and 20 for grain growth just after deformation, already during metadynamic evolution, for nickel-base superalloys. However, the recrystallized grain size at the end of deformation from Eq 18 should be used as the starting grain size for grain growth instead of d_{mdyn} (Ref 6).

Implementation of JMAK Models in Finite-Element Method Simulations. The workpiece temperature is rarely constant during real-life TMP operations but may vary due to deformation heating, die chilling, and so on. Such temperature transients are readily predicted at each location within a workpiece via finite-element method (FEM) analysis. In such cases, the modeling of microstructure evolution using the classical JMAK approach must be extended from isothermal to transient (nonisothermal) conditions. In these cases, a relatively simple and practical approach consists of introducing a fictitious (virtual) time or strain representative of that which would pertain if the structure had evolved until that moment under isothermal or constant strain-rate conditions, and applying the JMAK model for one (typically short) subsequent strain or time increment. For the case of air cooling a workpiece after forging (in which metadynamic or static evolution applies), for instance, FEM is used to determine the temperature T_i at the ith time step t_i of the simulation at some node of the mesh. Assuming that the recrystallized fraction $X(t_i)$ has been computed up to that moment, one can define a virtual time t_v by rearrangement of:

$$X(t_i) = 1 - \exp\left[-\ln 2\left(\frac{t_v}{t_{0.5}(T_i)}\right)^n\right] \qquad \text{(Eq 22)}$$

and apply the model to reach the next time step, $t_{i+1} = t_i + \delta t_i$, in accordance with the relation:

$$X(t_{i+1}) = 1 - \exp\left[-\ln 2\left(\frac{t_v + \delta t_i}{t_{0.5}(T_i)}\right)^n\right] \qquad \text{(Eq 23)}$$

A similar procedure can be applied during deformation by defining a virtual strain and using the value of $\varepsilon_{0.5}$ that is appropriate for the temperature and strain rate of the considered time step and node. Instead of using the strain rate calculated by the FEM simulation, it is recommended to approximate it as $(\varepsilon(t_{i+1}) - \varepsilon(t_i))/\delta t$ to ensure consistency. In both the dynamic and metadynamic cases, this numerical integration procedure supposes that temperature and strain rate do not vary much from one step to the next. Otherwise, one may use the average temperatures for steps i and $i+1$ to mitigate undesirable numerical effects. Alternatively, one can rerun the FEM simulation to calculate and store the thermomechanical

history of the workpiece during shorter, more frequent time steps.

Industrial metalworking processes such as cogging, tandem rolling, and radial forging usually involve multiple deformations. In such instances, the microstructures may not fully recrystallize by the completion of each deformation. For nickel-base superalloys in particular, dealing appropriately with partially recrystallized structures is of major importance, inasmuch as it is related to the persistence of undesirable (coarse) ALA grains. Thus, it is useful to distinguish at least two families of grains during TMP: one comprising the initial grains that continue to accumulate strain and the other consisting of grains that have been recrystallized at least once, thus being typically finer and totally or partially free of strain (Ref 6). It is thus necessary to determine different values of $\varepsilon_{0.5}$ and $t_{0.5}$, depending on the sizes of the remnants of the initial grains and the previously recrystallized grains. Such analysis of the different generations of recrystallized grains (Fig. 10) allows the application of JMAK analysis to each deformation step and its subsequent metadynamic evolution while tracking specifically the recrystallization of the initial grains.

The analysis of multistep operations should also include the effect of static recovery between successive deformations. The appropriate strain in equations for the initial deformed grains is therefore not equal to the total strain accumulated since the beginning of the metalworking operation. Specifically, if the strain increment applied during the ith deformation is denoted as $\Delta \varepsilon_i$, the strain ε_i to use in the JMAK equations for the strained grains is computed as:

$$\varepsilon_i = \lambda \varepsilon_{i-1} + \Delta \varepsilon_i \qquad \text{(Eq 24)}$$

in which λ is a temperature- and/or time-dependent parameter ranging between 0 and 1. For previously recrystallized grains, λ is equal to zero, because recrystallization can be expected to have eliminated most, if not all, of their stored energy.

In cases requiring less accuracy, partially recrystallized structures may be analyzed by focusing on only one set of "average" grains for each increment of deformation (Table 3, Fig. 11), as is commonly done for steels.

In summary, JMAK models have departed from their initial physics basis but are nonetheless very useful in many instances, such as the design and refinement of forging practices (Ref 26). They make use of a number of intermediate Arrhenius and power laws to capture the combined effects of process parameters and other variables on recrystallization behavior. However, they may not be totally adequate due to their phenomenological nature. In particular, when a JMAK model has been calibrated for a specific superalloy, it is not possible to modify it reliably to address the behavior of other, even similar alloys, because the contributions of all driving forces are mixed in the equations. Thus, other forms of modeling must be explored.

Topological Models

There are two types of topological models that can be used to model microstructure evolution in nickel-base superalloys: discrete and continuous (Ref 27).

Discrete models are based on the mapping of microstructure onto a regular (and usually rather large) grid of cells or lattice points, namely squares or hexagons. Cells are assigned at least one state variable such as a grain number, crystallographic orientation, or dislocation density/level of stored work. Areas formed by cells with an identical grain number or similar state variable(s) define the grains. A set of rules defines the conditions by which cells may switch, that is, may change their state variable(s) to those of a neighboring cell. Grain growth or the migration of a recrystallization front occurs as the cell-switch process is repeated. This repetition is made by picking cells at random; a time step is complete when there has been as many random picks as there are cells in the grid.

There are two main variants of discrete topological models: cellular automata (CA) and Monte Carlo (MC). They are distinguished primarily by the rules that govern the switch of state. Cellular automata are more suitable to represent the progress of recrystallization, while Monte Carlo models are more often applied for static grain growth.

Continuous models rely on networks of interconnected points called vertices. The connections between vertices describe boundaries. In the simplest approaches of this sort, vertices are positioned only at triple junctions. In more complex models, there can be a number of vertices along the boundaries (typically ~ 7) to capture their curvature in addition to the vertices at triple junctions. Grains are then defined as the regions enclosed inside loops of interconnected vertices.
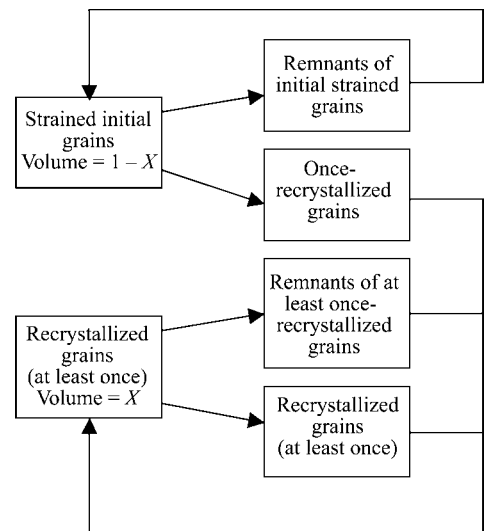
Fig. 10 Model algorithm for treating the evolution of a microstructure comprising initial grains and prior recrystallized grains. Source: Ref 6
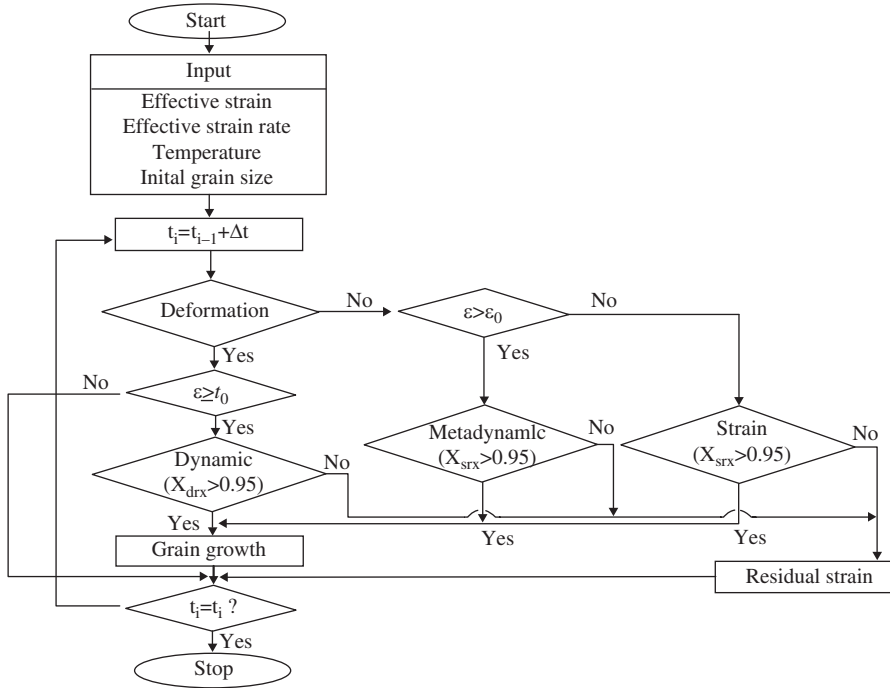
Fig. 11 Model algorithm in which recrystallization is treated by focusing on only one set of "average" grains for each increment of deformation. Source: Ref 22

Fig. 12 Alternating seven-cell neighborhoods used in cellular automata models that lead to octagonal grains in two-dimensional simulations

The movement of vertices under the influence of surface tension forces induces the growth or shrinkage of the grains they surround.

In the sections that follow, the application of the CA technique to model recrystallization is described first. This discussion is followed by a summary of the modeling of grain growth by MC models and network models.

Discrete Topological Models: Cellular Automata. In CA models for recrystallization, each cell is allocated a grain number as well as a variable that indicates its state of recrystallization and/or the recrystallization-wave number to which it belongs. All cells of the initial structure are considered unrecrystallized; they belong to recrystallization wave "0." When a cell is randomly picked, two events can cause it to switch its recrystallization state and grain number. First, if it meets the requirements set for nucleation, its grain number will also be changed to a new value, such as the maximum grain number in the grid incremented by one. Second, if a neighboring cell chosen at random is recrystallized, the cell will change its grain number to that of the recrystallized neighbor, in effect making that grain grow.

Geometric Considerations. The geometry of the array of CA cells plays a key role in the topological evolution of grain structure. For an array of hexagonal cells in two-dimensional (2-D) simulations, grains tend to grow in the shape of hexagons until impingement. Arrays of square cells, while more simple to program, may yield unrealistic grain shapes. If neighboring cells are considered to be only those that share a side (i.e., a so-called von Neumann neighborhood), nuclei will grow into diamond-shaped grains. If all eight surrounding neighbors are considered (i.e., if neighbors are defined as sharing at least a corner, as in a Moore neighborhood), grains will approximate squares. To correct for those shape effects, an alternate seven-cell pattern can be used. Each time a cell is chosen, only six of its neighbors are considered for the random-pick-and-switch analysis, alternating between the two configurations, such as those shown in Fig. 12. This tends to produce more realistic octagonal grains, until they impinge with other recrystallized grains.

The CA models also often assume that the top edge of the array is in contact with the bottom edge and that the right edge is in contact with the left. By this means, a form of spatial periodicity is introduced, thereby preventing boundary effects and essentially extending the size of the model domain.

Driving Forces. The geometric considerations and switching rules described previously are sufficient to reproduce simple recrystallization behaviors via CA. However, it is necessary to add driving forces to create more realistic representations of recrystallization kinetics and to adapt CA to specific materials. This implies the need for additional cell state variables, such as the dislocation density ρ, which are key to defining quantitative rules for nucleation and growth. A typical formulation to introduce driving forces has been used in the CA models developed by Goetz and Seetharaman for the DDRX of wrought structures (Ref 28); the approach was extended by Goetz to address DDRX of ingot structures characterized by the occurrence of PSN of recrystallization at carbide particles (Ref 29).
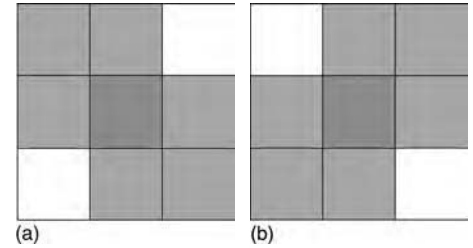
The 2-D CA approach of Goetz and Seerhara-man uses an array of 490 by 490 square cells. At the beginning of each time step, the dislocation density of all cells is uniformly increased at a fixed rate $d\rho/dt$, which is assumed to be proportional to the strain rate. Then, a number of cells N ($N < 490^2$) are randomly picked and updated to account for dynamic recovery by dividing their dislocation density by two (i.e., ρ becomes $\rho/2$ in those cells). This operation introduces a nonuniform dislocation density into the structure. The evolution of cell dislocation density is connected to the flow stress. The average dislocation density, ρ_A, is calculated from:

$$\rho_A = \frac{\sum_{\text{cells}} \rho}{490^2} \tag{Eq 25}$$

and the flow stress is assumed proportional to the square root of ρ_A.

When a steady state of strain hardening and dynamic recovery is reached, the average dislocation density reaches a constant value, ρ_{DR}, because, by definition, there is not net accumulation of dislocations in the structure. This balance of strain hardening and dynamic recovery is described by the following equation:

$$\frac{d\rho}{dt} = \frac{N}{490^2} \frac{\rho_{DR}}{2} \tag{Eq 26}$$

Because the rate at which dislocations are generated ($d\rho/dt$) is considered proportional to the strain rate, the assumption that the stress, σ_{DR}, reached during such steady state exhibits a strain-rate sensitivity, m, is given by:

$$\sigma_{DR} = K \left(\frac{d\rho}{dt}\right)^m \tag{Eq 27}$$

in which K and m are model parameters. Then, one can derive an expression for N:

$$N = 2 \left[\frac{490^2}{K}\right]^2 \left(\frac{d\rho}{dt}\right)^{1-2m} \tag{Eq 28}$$

Nucleation is modeled by randomly picking a fraction of the cells that are on the grain boundaries or next to second-phase particles (to describe

PSN on carbides). This fraction was taken to be $d\rho/dt/1000$ in the work of Goetz and Seetharaman. Nucleation is performed on those cells if their dislocation density has reached a critical value assumed to be ρ_{DR}. For each nucleus, the dislocation density is set back to zero.

The sequential switching trials for 490^2 random cell picks during each CA time step is performed to simulate the growth of recrystallized grains. When a cell is randomly picked, if it is an unrecrystallized cell, the cell grain number will switch to that of a randomly chosen neighbor cell (according to the alternate patterns in Fig. 12) if the latter is recrystallized. After a switch, the cell dislocation density is set back to zero as well.

Output of CA Models. The CA models are commonly used to predict microstructure evolution. The approach is capable of reproducing a number of features commonly observed during DDRX (Fig. 13). For the case of MDRX or SRX, on the other hand, the results may not show such good agreement with observations (e.g., Fig. 14). Specifically, the energy stored in the form of dislocations diminishes when MDRX or SRX is almost complete. Grain-boundary energy, which is not taken into account in CA models, then becomes the prevalent driving force for the evolution of microstructure, which is then typically characterized by isotropic, equiaxed grains. In these instances, MC models are better suited to simulate microstructure evolution.

Discrete Topological Models: Monte-Carlo. The MC models function in essentially the same way as CA models. Cells, or lattice sites as they are commonly known in MC terminology, are picked randomly, and a neighboring lattice site is chosen. The switch, also called a reorientation, is determined by the change in energy, ΔE, it would produce. If the change in energy is negative or equal to zero, the cell reorientation is accepted. If it is positive, it is subjected to a Boltzmann probability, p, given by:

$$p = \exp\left(\frac{\Delta E}{kT}\right) \tag{Eq 29}$$

The energy, E, of the system is evaluated based on grain-boundary energy in the form of a Potts model. In the case of a 2-D hexagonal lattice, such as proposed by Coste et al. (Ref 31) for static grain growth in alloy 718 (UNS No 1718), E is given by:
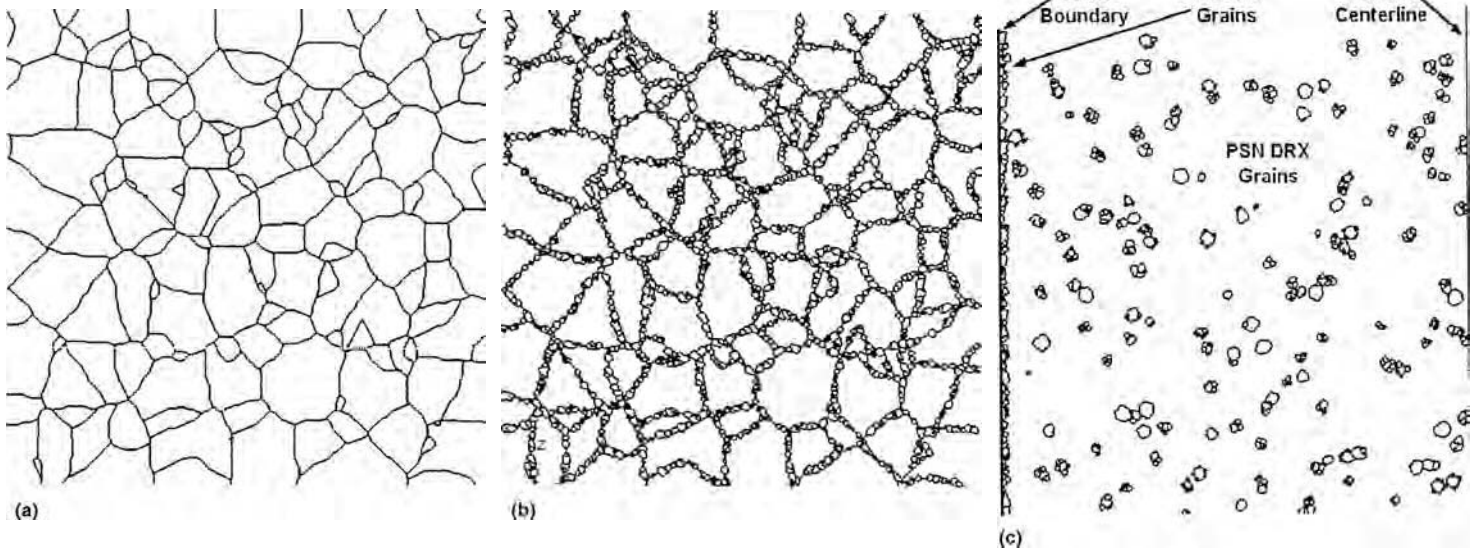
Fig. 13 Cellular automata results for discontinuous dynamic recrystallization (DRX). (a) Initial structure. (b) Necklace structure formed by partial dynamic recrystallization of (a). (c) Partial dynamic recrystallization of an ingot structure with particle-stimulated nucleation (PSN). GB, grain boundary. Source: Ref 29
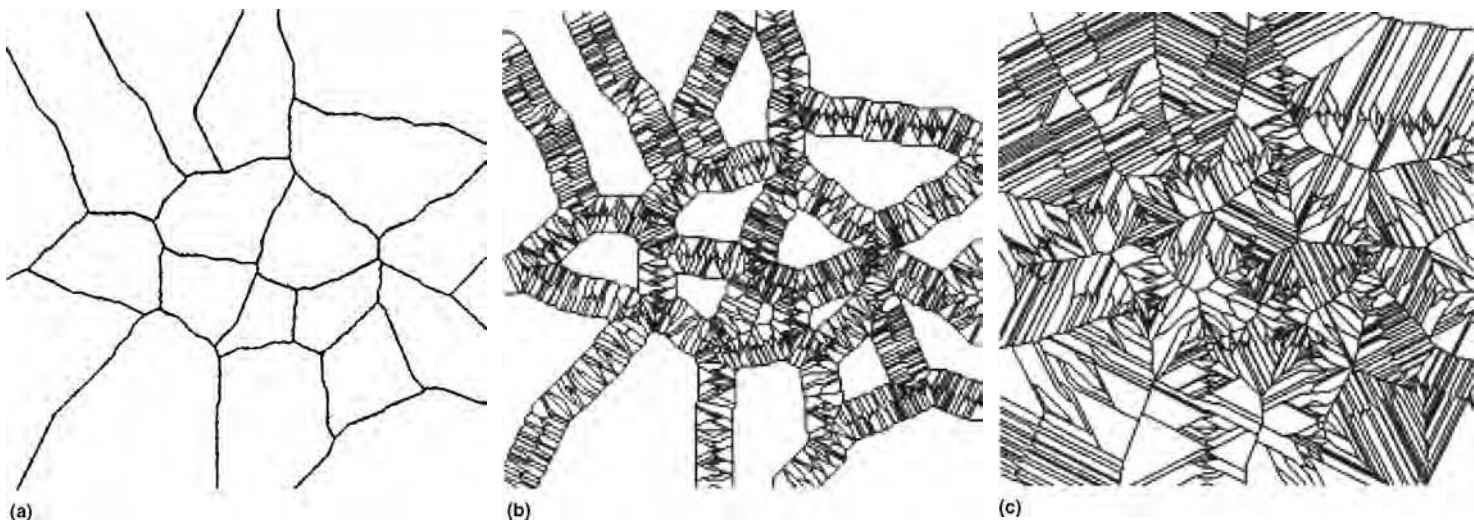
Fig. 14 Cellular automata results for static recrystallization. (a) Initial structure and structures after (b) 50% or (c) 100% static recrystallization. Source: Ref 30

$$E = -\gamma \sum_k \sum_{j=1}^{6} \left[\delta\left(Q_k, Q_{N(k,j)}\right) - 1 \right] \qquad \text{(Eq 30)}$$

in which Q_k is the grain number of lattice point k; $N(k, j)$ for j between 1 and 6 is the index of each neighbor of lattice point k in the grid; γ is the grain-boundary energy; and δ is the Kronecker delta.

In the approach of Coste et al., a specific grain number is introduced for lattice points representing δ-phase precipitates (Ni_3Nb) that may be present in alloy 718 at temperatures below 1020 °C. These precipitate cells cannot be reoriented but are required to reproduce the effect of Zener pinning on grain growth. In alloy 718, regions located within the former dendrite cores tend to exhibit slightly lower niobium contents despite homogenization heat treatment. This may lead to an inhomogeneous distribution of δ precipitates, which translates into different grain growth kinetics and the formation of alternating bands of coarse and fine grains corresponding to zones in which there are few δ precipitates (dendrite cores) or many (interdentritic regions), respectively. Such behavior is very well captured by the MC model of Coste et al. (Fig. 15).

The challenge associated with coupling CA and MC models prevents their extensive use, especially for industrial applications. This is largely due to the uncertainty as to the point during microstructure evolution at which the transition between the two modeling approaches should be made. Some work relevant to this question has been performed by Rollett and Raabe for aluminum alloys (Ref 32). The authors propose a continuous alternation between CA and MC evaluations. However, this approach raises a number of difficulties when it comes to determining the frequency of using one or the other of the two modeling approaches.

The CA and MC models do not provide a clear connection between their time steps and physical time. These models are also computationally intensive. The typical time for one simulation is of the order of minutes, and they require extensive computer memory to store the cells/grid and the values of their associated state variables. These requirements become somewhat unmanageable in view of the fact that such microstructure models would be needed at each of the thousands of nodes comprising an FEM mesh used to simulate an industrial process. In view of such challenges, CA and MC are typically used at selected tracking points to simulate microstructure evolution in regions of interest. These approaches can thus complement JMAK and mesoscale physics-based techniques that provide microstructure throughout a workpiece.

Continuous Topological Models: Network/Vertex Models. Network models based on the migration of vertices are used primarily for simulating static recrystallization and grain growth. They are the most physics-based models of microstructure evolution at high temperature.

In brief, in network models, vertex movement is determined by the evaluation of the dynamic balance of grain-boundary surface tensions, with the decrease of overall grain-boundary energy controlled by a dissipation potential (Ref 27). Such models can represent microstructure using a lighter computational footprint that MC or CA models. However, they require significantly more upfront programming, such as that needed to address topological situations that arise as grains disappear during grain growth, necessitating the reconnection of vertices. Similar difficulties are met when second-phase particles are introduced, because new vertices are needed to represent the contact of grain boundaries with precipitates (Ref 33).

The precision and versatility of network models is better than MC models. For instance, the size of second-phase particles may be limited in MC approaches. Network models can treat virtually any precipitate size and fraction. As a result, they can represent the behavior of banded structures (Fig. 16) more accurately than MC models. Hence, network models can be of great interest for nickel-base superalloys. In addition, these models simulate microstructure evolution in physical time instead of mathematical MC (or CA) time steps. Nevertheless, a comparison of various network and MC models by Maurice (Ref 27) provides a path to resolve the MC time-step issue through a calibration procedure using the network model results for the same material. In this approach, the MC time step is normalized by the ratio of the number of lattice points in the array to the initial number of grains, that is, by the average number of lattice points per grain in the initial configuration.

There are still numerous issues to resolve before topological models can be used routinely to predict microstructure development during the TMP of nickel-base superalloys with the level of precision and reliability that is needed. Nevertheless, progress is expected in academic research that will allow this in the future.

Mesoscale Physics-Based Models

The introduction of a sound physics basis into models of microstructure evolution for industrial applications has been most successful using simple, albeit approximate, mesoscale models.

Early Mesoscale Physics-Based Models. Mesoscale microstructure-evolution models comprise various formulations, each of which attempts to provide a description of the geometry and topology of recrystallization as well as representative equations for the underlying driving forces involved in various complex phenomena. The most common features of these methods are their reliance on an explicit evaluation of dislocation density, similar to that used for CA models, and their treatment of the relationship between grain-boundary velocities and recrystallization. Furthermore, unlike JMAK models, a nonuniform distribution of nucleation sites in the microstructure is also frequently assumed.

The first mesoscale physics-based model for hot working and recrystallization was probably that developed by Sandstrom and Lagneborg in 1975 (Ref 35). Approaches of this type were not pursued for some time thereafter, most likely because of the computing demands and high degree of complexity of these models.

It was not until 1997 that the first mesoscale physics-based model for a nickel-base superalloy was developed and applied successfully by Marty et al. (Ref 36) for the forging of alloy 718 parts. This model includes a simple geometric description of microstructure undergoing recrystallization in the form of a necklace of recrystallized grains surrounding deformed grains of diameter A (Fig. 17). The dislocation density within the initial grains is taken to be ρ, and that of the recrystallized grains is ρ_L. The evolution of ρ is described by the relation:
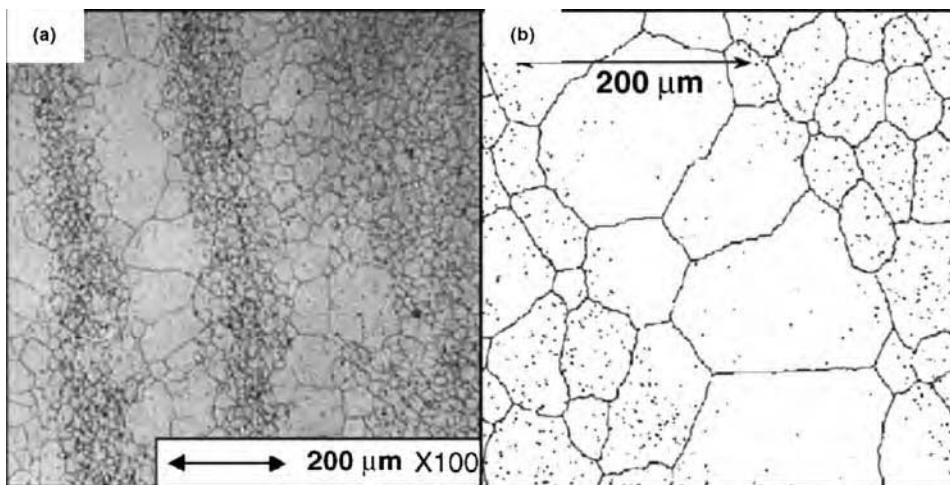
Fig. 15 (a) Micrograph showing a heterogeneous distribution of delta particles (black dots) and banded grain structure. (b) Simulated microstructure after the final step of a Monte Carlo simulation for a material containing an initial distribution of delta particles typical of that in (a). Source: Ref 31
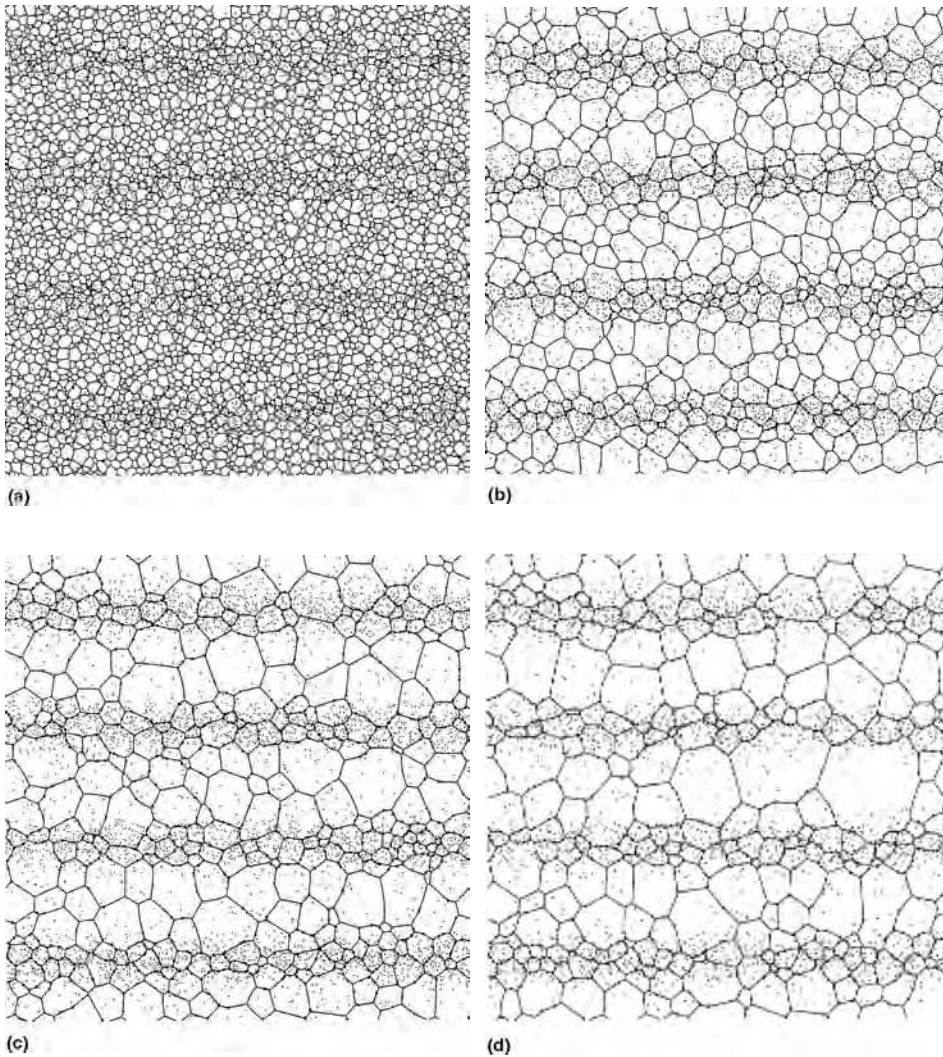
Fig. 16 Vertex model predictions of microstructure evolution from the initial configuration (a) to the final pinned state (d) for the case of a banded distribution of particles. Source: Ref 34
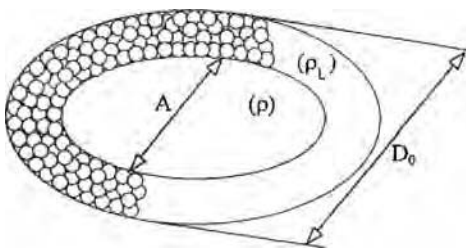
Fig. 17 Schematic representation of the necklace recrystallization structure assumed in the model of Marty et al. Source: Ref 36

$$\frac{d\rho}{dt} = B_1 \left[\sqrt{\frac{\rho}{\rho_0}}\dot{\varepsilon} - B_2 \left(\frac{\rho}{\rho_0} - 1\right)^2 \exp\left(-\frac{T_r}{T}\right) \right]$$

(Eq 31)

in which B_1 and B_2 denote coefficients that depend on temperature and grain size, ρ_0 is

the minimum dislocation density (10^{12} m^{-2}), and T_r is an activation temperature for recovery. The term involving T_r can be easily translated in the more usual form of an activation energy divided by RT, however. In the recrystallized necklace, the dislocation density ρ_L is evaluated through a complex scheme to account for assumed localized adiabatic heating, partly so that the approach can be used to simulate strain-localization effects during rolling.

The temporal evolution of the recrystallized fraction X in the model of Marty et al. (Ref 36) is:

$$\frac{dX}{dt} = A_1 \frac{(1-X)^{2/3}}{D_0} \exp\left(-\frac{T_X}{T}\right)\ (\rho - \rho_L)$$

(Eq 32)

Here, A_1 is a material parameter, D_0 is the initial grain size, and T_X is a so-called activation temperature for recrystallization. The term $(1 - X)^{2/3}/D_0$ is, in effect, proportional to the

surface area of the grain boundaries of the deformed grains divided by the initial grain volume. Therefore, Eq 32 is a geometrically valid expression for the rate at which volume is lost by the deformed grains when their grain boundaries migrate under the influence of a difference in dislocation density between them and their surrounding necklace of recrystallized grains. However, this model accounts for the behavior of recrystallized grains in only an approximate fashion, and it is not clear how microstructures are reinitialized to continue modeling during multiple deformations.

In 1999, a mesoscale model for the steady-state regime of dynamic recrystallization was proposed by Montheillet (Ref 37) and subsequently modified to simulate the recrystallization of alloy 718 (Ref 6). This form of mesoscale physics-based modeling was later extended to ingot microstructures by incorporating the topological and geometric effects of intragranular PSN (Ref 38). It is described in detail in a subsequent section.

Model of Sommitsch et al. In 2002, a dislocation-density-based model for simulating hot forming processes was formulated by Sommitich et al. and applied first to Nimonic 80A (UNS N07080) (Ref 39) and then to alloy 718 (UNS N07718) (Ref 7). This model puts a strong emphasis on the formulation of the driving forces that control nucleation and growth. The formulation draws from the concepts introduced by Sandstrom and Lagneborg and results in similarly complex expressions (Ref 40).

The Sommitsch et al. model (Ref 39) starts with expressions commonly used as the foundation of other mesoscale physics-based models. Strain hardening and recovery are described by the same equation used by Sandstrom and Lagneborg (Ref 35):

$$\frac{d\rho}{dt} = \frac{\dot{\varepsilon}}{\mathbf{b}l} - 2M\tau\rho^2$$

(Eq 33)

In this equation, $\mathbf{b}$ denotes the Burgers vector, l is the mean free path of dislocations, M is a so-called mobility of recovery, and τ is the average energy per unit length of dislocations. The latter is generally assumed to be $\sim 0.5 G\mathbf{b}^2$, in which G is the shear modulus. The recovery term in Eq 33 (second term on the right side) is similar to that used by Marty et al. (Ref 36) in that it is proportional to the square of the dislocation density. On the other hand, the strain-hardening term (first term on the right side of Eq 33) is different. Nevertheless, a similarity to the form assumed by Marty et al. (Ref 36) is obtained if the mean free path of dislocations, l, is related to the average distance between dislocations. The latter is proportional to the inverse square root of the dislocation density. Whether l depends only on strain rate and temperature (e.g., subgrain size) or is related to the free dislocation density should not affect the steady-state values, because it is always possible to adjust the recovery term to

compensate. However, it does change the initial shape of stress-strain curves, which depend on the square root of the dislocation density. A constant mean free path, l, produces an initial linear increase in dislocation density, leading to a rounded (square-root) shape for the stress-strain curve. In contrast, if l depends on the square root of the dislocation density, the initial portion of the stress-strain curve would increase linearly.

As in other models, the velocity of grain boundaries in the Sommitsch et al. model is derived from Turnbull's equation:

$$v = m\Delta P = m(\tau\Delta\rho - P_Z)\,P_S \qquad \text{(Eq 34)}$$

in which m is the grain-boundary mobility, and ΔP is the difference in stored energy between the two sides of the boundary (Ref 41). Sommitsch et al. chose to define ΔP to account not only for the effect of a difference in dislocation density, $\Delta\rho$, but also for additional effects such as Zener pinning due to second-phase particles (P_Z) and solute drag (P_S).

The analytical integration of expressions such as Eq 33 and 34, combined with nucleation and growth, leads to complex equations describing the evolution of grains undergoing a wave of recrystallization.

In contrast to the detailed analysis of driving forces, the representation of the geometry of the grain structure in the Sommitsch et al. model is rather limited, inasmuch as it only allows for the first two waves of recrystallization (as represented in Fig. 5a), although the approach could be extended to deal with subsequent ones as well. Furthermore, the location of nucleation sites is not explicitly addressed. With regard to interactions between recrystallized grains, the Sommitsch et al. model assumes strict impingement; that is, there is no migration of the boundaries formed by the contact between recrystallized grains that belong to the same wave of recrystallization. Specifically, the fraction Ψ of recrystallized grain boundaries for which migration stops due to impingement is introduced:

$$\Psi(X) = 1 - \left(\frac{X - X_C}{1 - X_C}\right)^n S(X, X_C) \qquad \text{(Eq 35)}$$

The value of Ψ depends on the recrystallized fraction X_C at which impingement is assumed to start, a constant n, and a step function S that is equal to 0 if $X < X_C$, and 1 otherwise. This equation may be somewhat approximate in view of its underlying assumptions. In particular, the coefficients X_C and n cannot be related to any geometric value yet probably have a great impact on the outcome of the model, similar to that of the Avrami exponent in JMAK formulations.

The Sommitsch et al. model was implemented into the FEM software DEFORM. An example of the output of the coupled microstructure-evolution/FEM model for the industrial processing of Nimonic 80A is shown in Fig. 18.
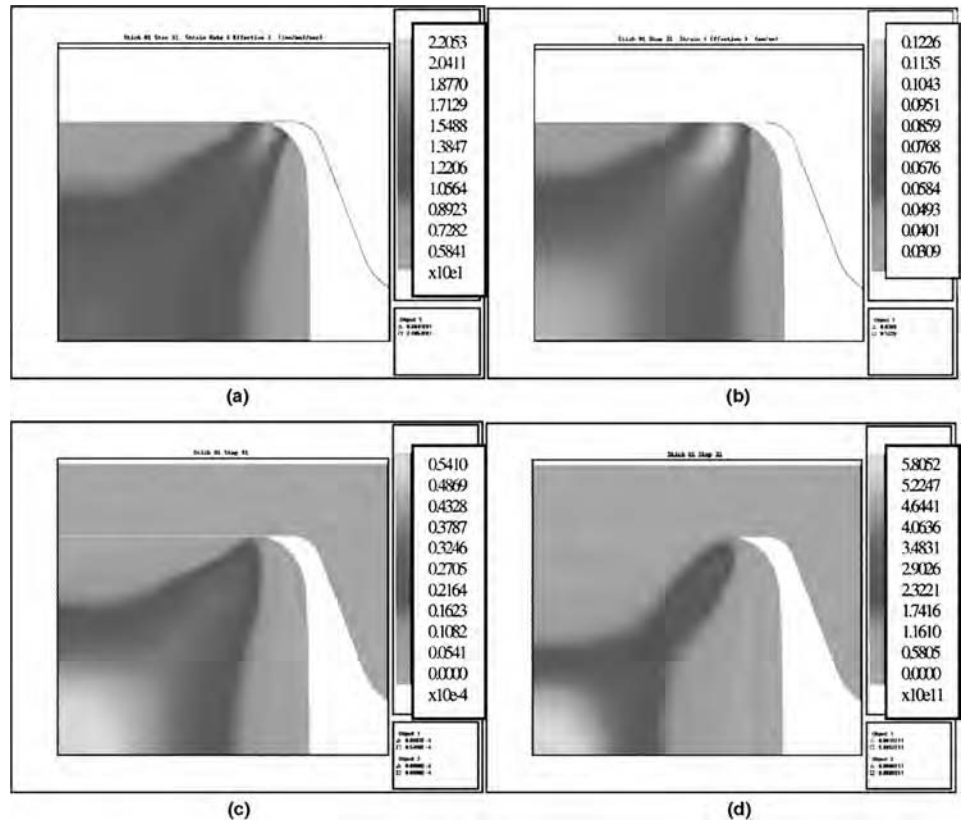
Fig. 18 Model predictions for slab rolling. (a) Strain rate (s^{-1}). (b) Strain. (c) Recrystallized fraction. (d) Dislocation density during the first recrystallization cycle. Due to symmetry, only a quarter section of the slab is shown. Source: Ref 39

Model of Montheillet and Thomas. The geometry-related shortcomings of the Sommitich et al. model (Ref 39, 40) are addressed in detail in the approach formulated by Montheillet and Thomas (Ref 37, 38, 42). Their approach comprises a rigorous framework to describe both the geometric/topological aspects and driving forces that pertain to recrystallization problems, such as those of importance for superalloys.

Geometric Framework. The modeling approach originally introduced by Montheillet (Ref 37) incorporates a rigorous geometric description of a microstructure undergoing the steady state of DDRX. It provided the basis for the development of two subsequent mesoscale physics-based models that were applied to the microstructure evolution for alloy 718 (Ref 6) and Waspaloy (Ref 38). These models were implemented using RX-MOD, a computer program dedicated to the modeling of microstructure evolution during hot working. The geometric part of these models evolved progressively into a framework capable of describing microstructures undergoing recrystallization via both a necklace of recrystallized grains on grain boundaries and intragranular recrystallized regions (Ref 42).

The geometric framework is based on so-called mesostructure units (MSUs) that can be viewed as aggregates of grains that exhibit similar properties, such as grain size and dislocation density. Each MSU is assigned its own index. In its simplest version, a model of discontinuous recrystallization requires two MSUs: one for the initial grains and the other for the recrystallized grains. In the specific model developed for Waspaloy, five MSUs are allocated. The first two represent the initial grains, distinguished by their Taylor factor. Similarly, two are used to store the grains that recrystallized during previous deformations. The final MSU (of highest index) consists of the recrystallized grains formed during the latest or current deformation step.

The framework is self-consistent in that it not only describes the microstructural state at any given time but also implements the rules that govern geometric changes in the grain structure. These rules are activated in reaction to the two types of input derived from the driving-force module of the model, namely, the rate at which nuclei are formed in each MSU and the velocity of grain boundaries between each pair of MSUs. Initially, the rates at which volume is lost through nucleation by each MSU and its grains are evaluated, assuming an identical nucleus size for all MSUs. The total volume of nuclei is added to that of the MSU of highest index, because this MSU consists of the grains recrystallized during the

current deformation step. Next, the rates of volume variation of each MSU and its grains due to grain-boundary migration are calculated based on the probabilities of contact between grains and in a way that naturally ensures volume conservation of the overall microstructure.

Volume conservation is a very important feature of the method introduced by Montheillet and therefore its geometric framework. To adequately address this requirement, MSUs are geometrically defined by five primary variables, three of which are volumes, denoted V_i, v_{ei}, and v_i for each MSU of index i. These denote, respectively, the total volume of MSU i, the volume enclosed in the outer grain boundary (envelope) of each grain of MSU i, and the actual volume of the crystal that constitutes each grain of MSU i. The term v_i is allowed to become less than v_{ei} to account for the development of intragranular recrystallized regions. Specifically, in the model formulation for the recrystallization of Waspaloy ingots, this feature is activated for the first two MSUs when they represent an initial structure of coarse columnar grains containing carbide particles that serve as PSN sites. The two other principal variables (α_{xyi} and α_{xzi}) are the $y{:}x$ and $z{:}x$ aspect ratios of the grains of MSU i, respectively. When the grains are spherical, α_{xyi} and α_{xzi} are equal to unity.

A number of other (secondary) variables of the geometric description can be derived from the five primary ones. For example, the grain density n_i is the ratio of V_i to v_i. In the case of spherical grains, the grain diameter D_i is equal to $(6v_{ei}/\pi)^{1/3}$, the boundary surface area of each grain of MSU i is $s_{ei} = \pi D_i^2$; and the total grain-boundary surface area of MSU i is $S_{ei} = n_i \cdot s_{ei}$. The density of grain boundaries in the volume of MSU i is $S_{vei} = 0.5 s_{ei}/v_{ei}$. (The coefficient 0.5 in this relation accounts for the fact that a grain boundary separates two grains, but it should contribute to an evaluation of grain-boundary density only once.)

The rate at which the volume, v, of a grain changes is defined by the surface area, s, of its boundary and the velocity, $\dot{u}$, at which its boundary migrates under the effect of driving force:

$$\dot{v} = s\,\dot{u} \tag{Eq 36}$$

Because the grains of any MSU i can be in contact with grains of different MSUs, Eq 36 must be refined to account for the variety of possible interactions in accordance with the following equation:

$$\dot{v}_{ei} = \sum_{j=1}^{N} \dot{v}_{eij} = \sum_{j=1}^{N} s_{ij}\,\dot{u}_{ij} \tag{Eq 37}$$

Here, $\dot{u}_{ij}$ is the grain-boundary velocity between MSUs i and j (provided as input to the framework), $s_{e\,ij}$ is the average surface area of grains of MSU i in contact with grains of MSU j, and N is the total number of MSUs. Assuming a uniform probability of contact

with grains of any MSU j in the structure, one can define S_{total} as the total density of grain boundaries in the structure and the probability, q_j, of contact of any grain boundary in the structure with that of a grain of MSU j such that:

$$s_{ij} = q_j\,s_{ei} \tag{Eq 38}$$

$$q_j = \frac{n_j\,s_{ej}}{\sum_{k=1}^{N} n_k\,s_{ek}} = \frac{S_{ej}}{S_{\text{total}}} \tag{Eq 39}$$

Equations 37 to 39 are useful to determine the rate of volume variation due to interactions between grains of different MSUs. However, they are not applied in the case of interactions between grains of the same MSU. In such a case, grain growth arises from the elimination of grain boundaries as they migrate and meet each other. It is evaluated through the grain-boundary density, S_{vei}, and the volume it sweeps under the effect of its migration, $S_{vei} q_i \dot{u}_{ii}$. Grain boundaries present in the volume swept are eliminated, and therefore, the grain-boundary density decreases at a rate $dS_{vei}/dt = S_{vei}^2 q_i \dot{u}_{ii}$. The solution of this differential equation gives the contribution of intra-MSU interactions to the rate at which the volume of its grains varies:

$$\frac{\dot{v}_{e\,ii}}{v_{ei}} = 3\,q_i\,S_{vei}\,\dot{u}_{ii} \tag{Eq 40}$$

It may be shown that the incorporation of the nuclei generated by all MSUs of the structure into the MSU of highest index N is given by the expression

$$\frac{\dot{v}_{eN}^{\text{nucl}}}{v_{eN}} = -\frac{\dot{n}_{\text{total}}^{\text{nucl}}}{n_N}\left(1 - \frac{v^{\text{nucl}}}{v_{eN}}\right) \tag{Eq 41}$$

in which v^{nucl} is the volume of a nucleus, and $\dot{n}_{\text{total}}^{\text{nucl}}$ is the total number of nuclei generated in the structure. Because the size of nuclei can reasonably be assumed to be finer than that of growing recrystallized grains, the incorporation of nuclei in the MSU of highest index contributes as a negative term to the rate of volume variation of the grains of that MSU, that is, leads to their refinement. When all other MSUs have been consumed by recrystallization, $q_N = 1$. Then, the addition of the positive term given by Eq 40 (with $i = N$) and the negative term provided by Eq 41 in order to calculate the rate of volume variation of the grains of that MSU shows that a steady state of dynamic recrystallization is reached, assuming an appropriate nucleation-rate input is provided to the geometric framework by the driving-force side of the model.

The aforementioned geometric framework may appear to be an oversimplified representation of the renewal of grains in a recrystallized structure, especially with regard to the typical understanding of DDRX and its steady state. In actuality, however, it is a very reasonable approach for nickel-base superalloys, considering

that the recrystallized grains in these materials undergo cycles of growth and fragmentation governed by the gradual disorientation of their subboundaries and twin boundaries. In fact, when the geometric framework is tested with rather simple inputs, it does indeed reproduce the expected range of behaviors (e.g., Fig. 19, 20).

The geometric framework also treats the development of intragranular recrystallized regions. The calculation in this case relies on specific probabilities of interaction, identified by a similar reasoning as before. In brief, if a volume-nucleation rate is provided as input to the framework for the first two MSUs (for coarse ingot grains), it induces a faster diminution of v_i than v_{ei}, because the latter is defined as the volume enclosed in the outer envelope of grains and is therefore not affected by intragranular nucleation, whereas v_i always is, by definition. For a given density of second-phase particles, typically carbides, this volume difference is translated in the average volume surrounding particles occupied by nuclei/
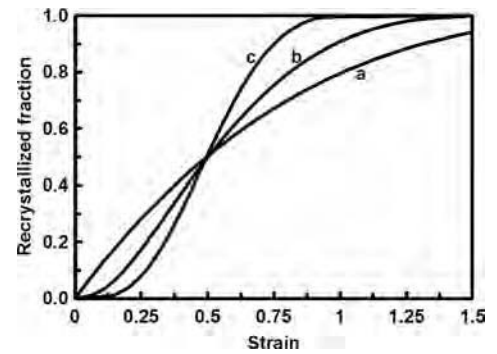
Fig. 19 Evolution of recrystallized fraction predicted by the mesoscale geometric framework for two-mesostructure unit simulations with nucleation rates of one nucleus per (a) 5, (b) 100, and (c) 5000 μm² of boundary and per unit strain and grain-boundary velocities of (a) 0, (b) 42, and (c) 105 μm per unit strain. The volume of the nucleus was assumed to be 100 μm³. Source: Ref 42
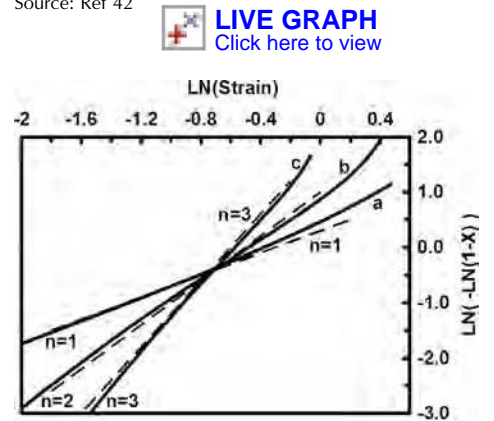
Fig. 20 Avrami analysis of the recrystallized-fraction curves in Fig. 19. X denotes the recrystallized fraction; the slope of the curves is the Avrami exponent n. Source: Ref 42

recrystallized grains. It then leads to the evaluation of the diameter of those regions and of their surface area (S_{bi}), where the coarse initial grains can be consumed from within by the growth of recrystallized grains through a new, intragranular form of interaction. The surface area of those regions provides the starting point to evaluate the various probabilities of interaction needed to translate grain-boundary velocity inputs into the rates of volume variation of grains and MSUs.

Driving Forces. In addition to the primary and secondary geometric variables of the geometric framework, MSUs are provided with variables for the modeling of driving forces. At a minimum, these driving-force variables include a dislocation density and various values that describe the substructure and/or twin boundaries that serve as nucleation sites, which may vary among different model formulations.

In the model developed for Waspaloy, the rate of dislocation generation is taken to be:

$$\dot{\rho}^{+} = \dot{\varepsilon}\frac{\sqrt{\rho}}{n\,\mathbf{b}} \qquad \text{(Eq 42)}$$

in which n is an estimate of the average number of forest dislocations a new, mobile dislocation can cross before it is stopped. Assuming the flow stress is expressed by:

$$\sigma = \sigma_0 + \alpha\,G\,\mathbf{b}\,\sqrt{\rho} \qquad \text{(Eq 43)}$$

the apparent recovery rate can be derived from a measured stress-strain curve using:

$$\frac{d\rho^{-}}{d\varepsilon} = \frac{\sigma - \sigma_0}{\alpha G\mathbf{b}^2}\left(\frac{1}{n} - \frac{2}{\alpha G}\frac{d\sigma}{d\varepsilon}\right) \qquad \text{(Eq 44)}$$

As an example, Fig. 21(a) shows actual stress-strain curves for a lot of wrought Waspaloy. The results in Fig. 21(b) are the apparent recovery rates for the two 0.1 s^{-1} increments of deformation (calculated in accordance with Eq 44) as a function of the instantaneous dislocation density obtained by rearrangement of Eq 43. The initial linear response shows that the rate of dynamic recovery is directly proportional to the dislocation density. The additional softening evidenced by the differences between the calculated apparent recovery rates and the corresponding initial linear responses is a result of DDRX. Therefore, the mesoscale model for Waspaloy should assume that dislocation densities for each MSU evolve according to an equation of the form:

$$\dot{\rho} = \dot{\varepsilon}\left(\frac{\sqrt{\rho}}{n\,\mathbf{b}} - r\,\rho\right) \qquad \text{(Eq 45)}$$

in which n and r depend on temperature, strain rate, and so on.

With regard to the modeling of nucleation, nickel-base superalloys exhibit generation and disorientation of subboundaries and, to a lesser extent, the formation and disorientation of twin boundaries. When a subboundary reaches a critical disorientation, it becomes mobile and thus serves a viable nucleus. The tracking of the kinetics of disorientation provides a quantitative description of nucleation, but it increases computation time.
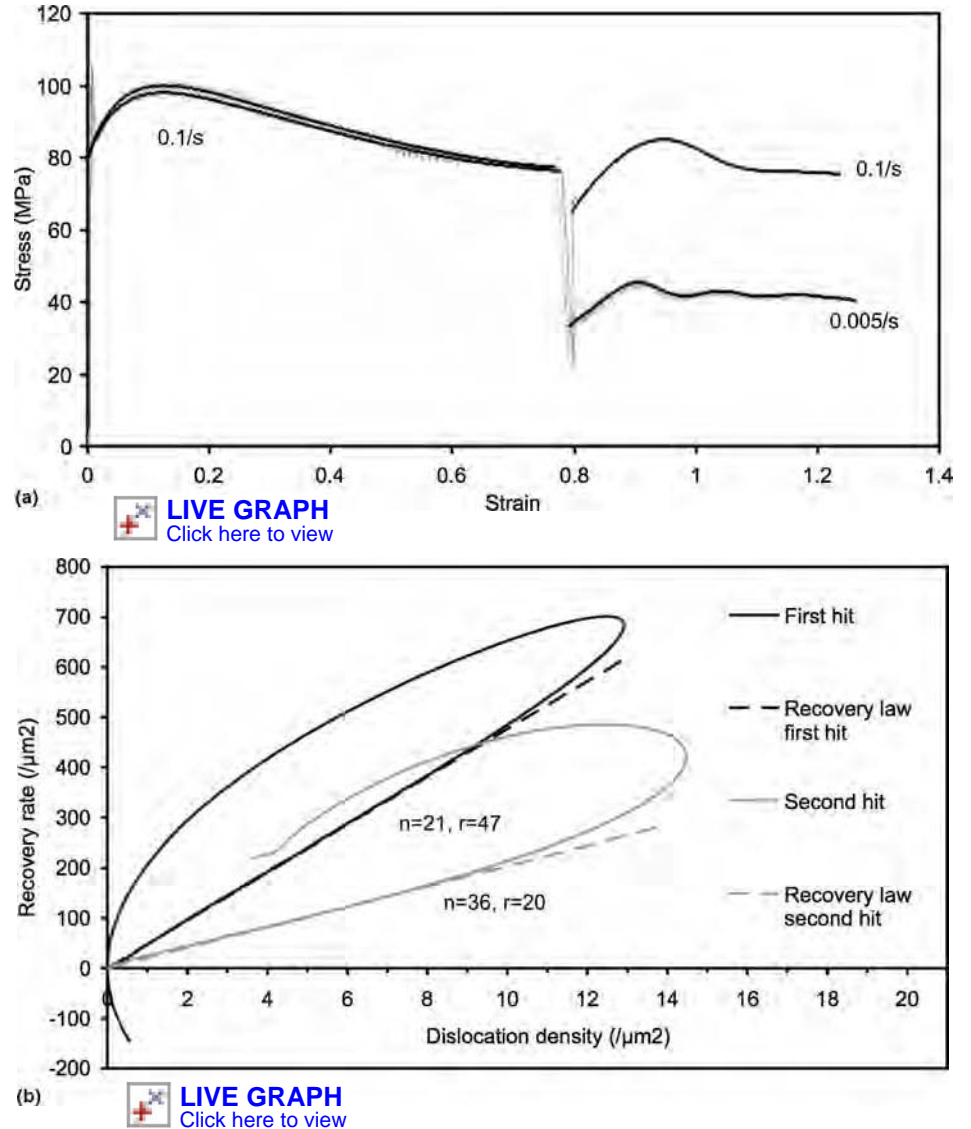
In the case of the model developed for alloy 718 (Ref 6), subboundaries and twin boundaries are represented by two distributions of boundary densities with disorientations ranging from 1 to 15° or from 60 to 55°, respectively. The calculation of the evolution of these distributions requires careful integration with fine time steps to prevent numerical instabilities and requires rather large amounts of computer memory. As a result, the approach for nucleation cannot be used at every node of a large three-dimensional FEM mesh.

The disorientation and nucleation model implemented for the dynamic recrystallization of Waspaloy assumes a simpler description of subboundaries and neglects twin boundaries altogether. It distinguishes between two types of subboundaries: those generated at the periphery of grains and those inside the grains, such as in the vicinity of coarse carbide particles. For each kind of subboundary, the model uses a density (S_{sb} and S_{sbV}) and a phenomenological value representing the overall level of disorientation (x and x_v). The first subboundaries control the nucleation of recrystallized grains at the surface/periphery of grains, that is, the formation of the necklace. The second subboundaries drive nucleation within the volume of grains and hence the development of intragranular recrystallized regions if the MSU to which they belong allows it; otherwise, they are simply added to the first.

The equations used for volume subboundaries in the model for Waspaloy ingot material serve as an example for the approach. The

Fig. 21 Constitutive analysis for wrought Waspaloy deformed in uniaxial compression at 1177 °C. (a) Measured stress-strain curves and (b) apparent recovery rate for the two increments of deformation imposed at a strain rate of 0.1 s^{-1}. Source: Ref 38

generation rate of subboundaries within the volume of grains is given by:

$$\dot{S}_{sbV}^{+} = k_V \, \exp\!\left(\frac{Q_V}{RT}\right)\dot{\varepsilon}\, F_V\!\left(M_{Taylor}\right) \qquad \text{(Eq 46)}$$

A critical strain for the activation of nucleation (and of recrystallization) is calculated from:

$$\varepsilon_{cV} = k_{cV} \, \exp\!\left(\frac{Q_{cV}}{RT}\right) F_{cV}\!\left(M_{Taylor}\right) \qquad \text{(Eq 47)}$$

This leads to the evaluation of the rate of increase of x or x_V:

$$\dot{x}_V^{+} = \frac{\dot{\varepsilon}}{\varepsilon_{cV}} \qquad \text{(Eq 48)}$$

Finally, the rate of transformation of subboundaries into grain boundaries is:

$$\dot{S}_{sbV}^{-} = \dot{S}_{gbV}^{+} = \left(k_{NV} + k_{NV}' \, S_b^p\right)\dot{\varepsilon}\,(x_v - 1)\,S_{sbV} \qquad \text{(Eq 49)}$$

Nucleation rates are derived from the rates of generation of new grain boundaries ($\dot{S}_{gb}^{+}$ and $\dot{S}_{gbV}^{+}$) by dividing them by the surface area of a nucleus. Equation 49 shows that nucleation starts when x_v (or x) becomes greater than one, which happens when the strain exceeds the critical value given by Eq 48. S_b is the surface area of intragranular regions, one of the secondary variables characterizing MSUs in the geometric framework, as mentioned previously. When the development of intragranular regions is activated for an MSU, such as for coarse ingot grains, S_b expands throughout recrystallization and provides increasing numbers of sites for subsequent nucleation in the form of a positive feedback loop. This is in part responsible for the acceleration of recrystallization observed when PSN is activated. F_V and F_{cV} are functions that provide acceleration or deceleration of PSN, depending on the Taylor factor of ingot structures. The Taylor factor itself is a function of the orientation of the fiber-textured as-cast grains with regard to the axis of compression. The functions F_V and F_{cV} are set equal to unity for wrought structures that typically exhibit a random texture. In most cases, even for fiber-textured ingot grains, the orientation dependence is relatively weak and may be neglected to a first order.

The coupling of mesoscale physics-based models with FEM simulations allows the prediction of microstructure evolution throughout a thermomechanical process. As an example, Fig. 22 shows the evolution of the size of recrystallized grains predicted by the mesoscale model developed for Waspaloy when it is coupled with DEFORM-3D for the FEM simulation of a billet-cogging process.

Current Status and Future Outlook

There are three main forms of models for the prediction of microstructure evolution during the TMP of nickel-base superalloys. First, JMAK models have demonstrated their reliability and usefulness throughout years of development and application in both academia and industry. Second, topological models primarily address more research-oriented issues than full-scale industrial applications, but this may change thanks to future improvements, especially in terms of computing power. Third, mesoscale physics-based models have become increasingly relevant and applicable since the end of the 1990s. They are capable of addressing the complex industrial issues posed by the TMP of nickel-base superalloys. They combine the advantages of a relatively light computer footprint, similar to that of JMAK models, with a firm physical basis reminiscent of topological models.

REFERENCES

1. D.U. Furrer and H. Fecht, Ni-Base Superalloys for Turbine Discs, *JOM,* Jan 1999, p 14–17
2. R.M. Forbes-Jones and L.A. Jackman, The Structural Evolution of Superalloy Ingots during Hot Working, *JOM,* Jan 1999, p 27–31
3. D.U. Furrer and S.L. Semiatin, Forging of Nickel-Base Superalloys, *Metalworking: Bulk Forming,* Vol 14A, *ASM Handbook,* ASM International, 2005, p 324–330
4. S.C. Medeiros, Y.V.R.K. Prasad, W.G. Frazier, and R. Srinivasan, Microstructural Modeling of Metadynamic Recrystallization in Hot Working of IN718 Superalloy, *Mater. Sci. Eng. A,* Vol 293, 2000, p 198–207
5. B. Tian, C. Lind, E. Schafler, and O. Paris, Evolution of Microstructures during Dynamic Recrystallization and Dynamic Recovery in Hot Deformed Nimonic 80A, *Mater. Sci. Eng. A,* Vol 367, 2004, p 198–204
6. J.P. Thomas, E. Bauchet, C. Dumont, and F. Montheillet, EBSD Investigation and Modeling of the Microstructural Evolutions of Superalloy 718 during Hot Deformation, *Superalloys 2004,* K.A. Green, T.M. Pollock, H. Harada, T.E. Howson, R.C. Reed, J.J. Schirra, and S. Waltson, Ed., TMS, 2004, p 959–968
7. P. Poelt, C. Sommitsch, S. Mitsche, and M. Walter, Dynamic Recrystallization of Ni-Base Alloys—Experimental Results and Comparisons with Simulations, *Mater. Sci. Eng. A,* Vol 420, 2006, p 306–314
8. M.J. Luton and M. Sellars, Dynamic Recrystallization in Nickel and Nickel-Iron Alloys during High Temperature Deformation, *Acta Metall.,* Vol 17, 1969, p 1033–1043
9. L.E. Murr, *Interfacial Phenomena in Metals and Alloys,* Addison-Wesley Publishing Company, 1975
10. T. Sakai, Microstructural Development under Dynamic Recrystallization of Multi-Crystalline Materials, *J.J. Jonas Symposium on Thermomechanical Processing* (Ottawa, Canada), 2000
11. C. Zener and J.H. Hollomon, Effect of Strain Rate upon Plastic Flow of Steel, *J. Appl. Phys.,* Vol 15, 1944, p 22–32
12. B. Derby, The Dependence of Grain Size on Stress during Dynamic Recrystallization, *Acta Metall. Mater.,* Vol 39, 1991, p 955–962
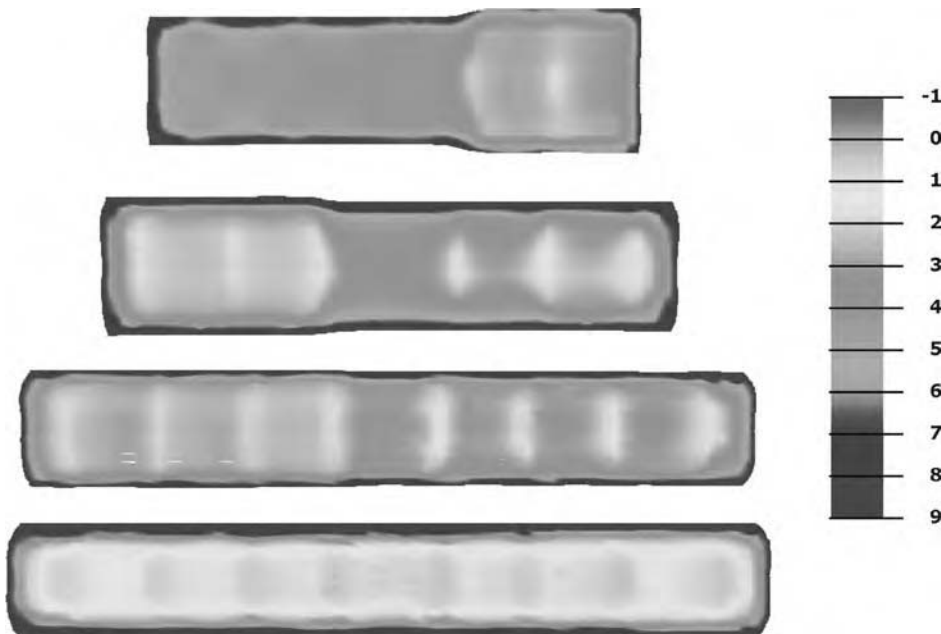
Fig. 22 Model predictions of the evolution of recrystallized grain size in the longitudinal section of a Waspaloy billet during the cogging process. Courtesy of the Air Force Research Laboratory

13. B. Derby, Dynamic Recrystallization: The Steady State Grain Size, *Scr. Metall.,* Vol 27, 1992, p 1581–1585

14. M. Aghaie-Khafri and N. Golarzi, Dynamic and Metadynamic Recrystallization of Hastelloy X Superalloy, *J. Mater. Sci.,* Vol 43, 2008, p 3717–3724

15. A.N. Kolmogorov, *Izv. Akad. Nauk. SSSR, Ser. Materm.,* Vol 3, 1937, p 335–359

16. W.A. Johnson and R.F. Mehl, *Trans. AIME,* Vol 135, 1939, p 416–443

17. M. Avrami, Granulation, Phase Change, and Microstructure, Kinetics of Phase Change I, *J. Chem. Phys.,* Vol 7, 1939, p 1103–1112

18. M. Avrami, Granulation, Phase Change, and Microstructure, Kinetics of Phase Change II, *J. Chem. Phys.,* Vol 8, 1940, p 212–224

19. M. Avrami, Granulation, Phase Change, and Microstructure, Kinetics of Phase Change III, *J. Chem. Phys.,* Vol 9, 1941, p 177–184

20. T. Furu, K. Marthinsen, and E. Nes, Modeling Recrystallization, *Mater. Sci. Technol.,* Vol 6, 1990, p 1093–1102

21. G. Shen, S.L. Semiatin, and R. Shivpuri, Modeling Microstructural Development during the Forging of Waspaloy, *Metal. Mater. Trans. A,* Vol 26, 1995, p 1795–1803

22. J.T. Yeom, C.S. Lee, J.H. Kima, and N.K. Park, Finite-Element Analysis of Microstructure Evolution in the Cogging of an Alloy 718 Ingot, *Mater. Sci. Eng. A,* Vol 449–451, 2007, p 722–726

23. D. Huang, W.T. Wu, D. Lambert, and S.L. Semiatin, Computer Simulation of Microstructure Evolution during Hot Forging of Waspaloy and Nickel Alloy 718, *Microstructure Modeling and Prediction during Thermomechanical Processing,* R. Srinivasan, S.L. Semiatin, A. Beaudoin, S. Fox, and Z. Jin, Ed., TMS, 2001, p 137–146

24. S.L. Semiatin, D.S. Weaver, P.L. Fagin, M.G. Glavicic, R.L. Goetz, and N.D. Frey, Deformation and Recrystallization Behavior during Hot Working of a Coarse-Grain, Nickel-Base Superalloy Ingot Material, *Metall. Mater. Trans. A,* Vol 35, 2004, p 679–693

25. A.J. Brand, K. Karhausen, and R. Kopp, Microstructural Simulation of Nickel Base Alloy Inconel 718 in Production of Turbine Discs, *Mater. Sci. Tech.,* Vol 12, 1996, p 963–969

26. B. Antolovich and M. Evans, Predicting Grain Size Evolution of UDIMET Alloy 718 during the 'Cogging' Process through the Use of Numerical Analysis, *Superalloys 2000,* T.M. Pollock, R.D. Kissinger, R.R. Bowman, K.A. Green, M. McLean, S. Olson, and J.J. Shirra, Ed., TMS, 2000, p 39–47

27. C. Maurice, Numerical Modelling of Grain Growth: Current Status, *Recrystallization and Grain Growth,* G. Gottstein and D.A. Molodov, Ed., Springer-Verlag, 2001, p 123–134

28. R.L. Goetz and V. Seetharaman, Modeling Dynamic Recrystallization Using Cellular Automata, *Scr. Mater.,* Vol 38 (No. 3), 1998, p 405–413

29. R.L. Goetz, Particle Stimulated Nucleation during Dynamic Recrystallization Using a Cellular-Automata Model, *Scr. Mater.,* Vol 52, 2005, p 851–856

30. R.L. Goetz and V. Seetharaman, Static Recrystallization Kinetics with Homogeneous and Heterogeneous Nucleation Using a Cellular-Automata Model, *Metall. Mater. Trans. A,* Vol 29, 1998, p 2307–2321

31. S. Coste, E. Andrieu, and J. Huez, Effect of a Heterogeneous Distribution of Particles on the Formation of Banded Grain Structure in Wrought Alloy 718, *Mater. Sci. Eng. A,* Vol 396, 2005, p 92–98

32. A.D. Rollett and D. Raabe, A Hybrid Model for Mesoscopic Simulation of Recrystallization, *Comput. Mater. Sci.,* Vol 21, 2001, p 69–78

33. D. Weygand, Y. Bréchet, and J. Lépinoux, Zener Pinning and Grain Growth: A Two-Dimensional Vertex Computer Simulation, *Acta Mater.,* Vol 47 (No. 3), 1999, p 961–970

34. D. Weygand, Y. Bréchet, and J. Lépinoux, Inhibition of Grain Growth by Particle Distribution: Effect of Spatial Heterogeneities and of Particle Strength Dispersion, *Mater. Sci. Eng. A,* Vol 292, 2000, p 34–39

35. R. Sandström and R. Lagneborg, A Model for Hot Working Occurring by Recrystallization, *Acta Metall.,* Vol 23, 1975, p 387–398

36. B. Marty, J.Y. Guedou, P. Gergaud, and J.L. Lebrun, Recrystallization and Work-Hardening Prediction during Forging Process of Inconel 718, *Superalloys 718, 625, 706 and Various Derivatives,* E.A. Loria, Ed., TMS, 1997, p 331–342

37. F. Montheillet, Modeling the Steady State Regime of Discontinuous Dynamic Recrystallization, *Fourth Int. Conf. on Recrystallization and Related Phenomena,* 1999, p 651–658

38. J.P. Thomas and S.L. Semiatin, Mesoscale Modeling of the Recrystallization of Waspaloy and Application to the Simulation of the Ingot-Cogging Process, *Materials Science and Technology (MS&T) 2006: Processing,* TMS, Warrendale, PA, 2006, p 609–619

39. C. Sommitsch, V. Wieser, and S. Kleber, A Dislocation Density Model for the Simulation of Hot Forming Processes, *J. Mater. Proc. Technol.,* Vol 125–126, 2002, p 130–137

40. C. Sommitsch and W. Mitter, On Modeling of Dynamic Recrystallization of FCC Materials with Low Stacking Fault Energy, *Acta Mater.,* Vol 54, 2006, p 357–375

41. D. Turnbull, Theory of Grain Boundary Migration Rates, *Trans. Metall. Soc. AIME,* Vol 191, 1951, p 661–665

42. J.P. Thomas, F. Montheillet, and S.L. Semiatin, A Geometric Framework for Mesoscale Models of Recrystallization, *Metall. Mater. Trans. A,* Vol 38, 2007, p 2095–2109

مرجع مهندسی مواد و متالورژی

Physical Data on the Elements and Alloys

Periodic Table of Elements

FUNDAMENTAL TO MODELING is the fact that engineering materials are composed of chemical elements, and the performance of engineering materials reflects the chemistry of these elements. Basic information is provided here on the chemical elements and their arrangement into a periodic table (Fig. 1) based on recurring similarities in the fundamental nature of the elements.

Elements are the basic substances that cannot be decomposed further by chemical means or made by chemical union. These elements follow a periodic pattern related to the electron configuration that allows them to be arranged into the convenient periodic table. In the periodic table, the elements are arranged in the order of increasing atomic number (the number of protons in the nucleus), but those that have the same number of electrons inhabiting their outermost electron shells (and therefore having similar chemical behavior) are grouped under each other; hence, the vertical columns of related elements in the periodic table are called groups. As atomic number increases, more orbital electrons must be accommodated in the shells, and as each outermost shell is filled, a new outermost shell (and a new period of the table) is begun. Because the additional shells are larger than the preceding shells, there is room for more electrons, and the periods get longer: from two (period 1) to eight and eight (periods 2 and 3), then to eighteen and eighteen (periods 4 and 5), and finally to thirty-two and (probably) thirty-two (periods 6 and 7).

Over the years, various conventions have been used to identify each group in the periodic table. For example, the 1948 edition of the *Metals Handbook* used a system similar to that adopted in 1970 by the International Union of Pure and Applied Chemistry (IUPAC) in which the left-side and right-side elements in periods 4 through 7 are distinguished by the addition of the letters "A" and "B," respectively, the only difference being that the Handbook used small (lower-case) letters with the Roman numerals rather than capital (upper-case) letters. In contrast to the 1970 IUPAC system, which was commonly used in Europe, the Chemical Abstracts Service (CAS) of the American Chemical Society and many other U.S. chemists applied the "B" to the additional groups (groups 3 through 12) in periods 4 through 7 and "A" to the rest. Other conventions that have been occasionally followed in the past are to use "0" instead of VIIIA or VIIIB to identify group 18, and to use Arabic numerals instead of Roman numerals. The most recent (1988) IUPAC recommendation is to eliminate both practices and to use Arabic numerals 1 through 18 instead.

Another common practice is to identify groups of similar elements with collective names. For example, the elements lithium, sodium, potassium, rubidium, cesium, and francium in group 1 are called alkali metals from the Arabic for "the ashes" (hydrogen, which is not normally a solid, is not included), and all the elements in group 2 are called alkaline earth metals. While not yet approved by IUPAC, the term *pnicogens* or *pnictides* has been applied to the elements in group 15; the term is from the Greek for "stifle, choke, or strangle," which alludes to nitrogen, that is, "burnt air." The term *chalcogens* has been applied to group 16 elements; the term is from the Greek for "copper," which first alluded to the common occurrence of oxygen and sulfur in many copper ores but later referred to all these elements that were "ore formers." The term *halogens* (from the Greek for "salt") has been applied to group 17, and the elements in group 18 are called noble gases or inert gases. Sometimes, hydrogen is positioned at the top of the halogen group in addition to its normal position at the top of group 1.

Rows of some similar elements are similarly named. For example, the elements in rows 3 through 10, whose atoms have an incomplete *d* subshell of electrons, are called transition elements or transition metals. IUPAC 1988, however, also includes the atoms in row 11 in the group of transition elements, while other organizations include all the elements between rows 2 and 12. The elements lanthanum through lutetium (atomic numbers 57 through 71) are termed lanthanide elements, lanthanide metals, or simply lanthanides. The lanthanides together with group 3 elements yttrium (atomic number 39) and scandium (atomic number 21) are called rare earth elements, rare earth metals, or simply rare earths. The elements actinium through lawrencium (atomic numbers 89 through 103) are similarly termed actinide elements, actinide metals, or simply actinides (or sometimes second-series rare earths), while those from thorium (atomic number 90) through lawrencium (atomic number 103) are termed actinoids.

Most of the chemical elements are metals. A metal is any element that tends to lose electrons from the outer shells of its atoms, with the resulting positive ions held together in a unique crystal structure by the cloud of these free electrons in a mechanism that has been called metallic bonding. This type of atomic bonding is in contrast to ionic bonding and to covalent bonding. In ionic bonding, transfer of valence (outer shell) electrons between dissimilar atoms produces stable outer shells in each and results in positive and negative ions that are mutually attracted by coulombic forces but do not form a crystal. In covalent bonding, valence electrons are shared rather than exchanged, and a nonmetallic crystal can form, for example, the diamond form of carbon.

The cloud of free electrons in metallic crystals gives rise to the three physical characteristics typical of solid metals: metals are good conductors of electricity, good conductors of heat, and they have a lustrous appearance. In addition, most elemental metals are malleable, ductile, generally denser than other elemental substances, and usually form positive ions. Those elements that do not display the characteristics of the metallic elements are called nonmetals. However, there are some elements that act as metals under some circumstances and like nonmetals under different circumstances and are called semimetal. As shown, the boundaries separating the regions in the periodic table covered by the three classes of elements are not distinct, except that nonmetals never form positive ions.

Metals — Semi-metals — Nonmetals

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
IUPAC 1988	IA	IIA	IIIA	IVA	VA	VIA	VIIA		VIIIA		IB	IIB	IIIB	IVB	VB	VIB	VIIB	VIIIB	IUPAC 1970
Previous CAS	IA	IIA	IIIB	IVB	VB	VIB	VIIB		VIIIB		IB	IIB	IIIA	IVA	VA	VIA	VIIA	VIIIA	

Key to chart

Atomic number → 50 +2 ← Oxidation states
Symbol → Sn +4
Atomic weight → 118.71
-18-18-4 ← Electron configuration

Transition elements

Period 1

1 +1 / -1																	2 0
H																	He
1.00794 / 1																	4.002602 / 2

K

Period 2

3 +1	4 +2											5 +3	6 +2 +4 -4	7 +1 +5 +2 -1 +3 -2 -4 +4 -3	8 -2	9 -1	10 0
Li	Be											B	C	N	O	F	Ne
6.941 / 2-1	9.0122 / 2-2											10.811 / 2-3	12.0107 / 2-4	14.00674 / 2-5	15.9994 / 2-6	18.998403 / 2-7	20.1797 / 2-8

K-L

Period 3

11 +1	12 +2											13 +3	14 +2 +4 -4	15 +3 +5 -3	16 +4 +6 -2	17 +1 +5 +7 -1	18 0
Na	Mg											Al	Si	P	S	Cl	Ar
22.98977 / 2-8-1	24.3050 / 2-8-2											26.98154 / 2-8-3	28.0856 / 2-8-4	30.97376 / 2-8-5	32.066 / 2-8-6	35.4527 / 2-8-7	39.948 / 2-8-8

K-L-M

Period 4

19 +1	20 +2	21 +3	22 +2 +3 +4	23 +2 +3 +4 +5	24 +2 +3 +6	25 +2 +3 +4 +7	26 +2 +3	27 +2 +3	28 +2 +3	29 +1 +2	30 +2	31 +3	32 +2 +4	33 +3 +5 -3	34 +4 +6 -2	35 +1 +5 -1	36 0
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
39.0983	40.078	44.95591	47.967	50.9415	51.9961	54.9380	55.845	58.9332	59.6934	63.546	65.39	69.723	72.61	74.9216	78.96	79.904	83.80
-8-8-1	-8-8-2	-8-9-2	-8-10-2	-8-11-2	-8-13-1	-8-13-2	-8-14-2	-8-15-2	-8-16-2	-8-18-1	-8-18-2	-8-18-3	-8-18-4	-8-18-5	-8-18-6	-8-18-7	-8-18-8

-L-M-N

Period 5

37 +1	38 +2	39 +3	40 +4	41 +3 +5	42 +6	43 +4 +6 +7	44 +3	45 +3	46 +2 +4	47 +1	48 +2	49 +3	50 +2 +4	51 +3 +5 -3	52 +4 +6 -2	53 +1 +5 +7 -1	54 0
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
85.4678	87.62	88.90585	91.224	92.90638	95.94	(98)	101.07	102.9055	106.42	107.8682	112.411	114.818	118.71	121.76	127.60	126.90447	131.29
-18-8-1	-18-8-2	-18-9-2	-18-10-2	-18-12-1	-18-13-1	-18-13-2	-18-15-1	-18-16-1	-18-18-0	-18-18-1	-18-18-2	-18-18-3	-18-18-4	-18-18-5	-18-18-6	-18-18-7	-18-18-8

-M-N-O

Period 6

55 +1	56 +2	La-Lu	72 +4	73 +5	74 +6	75 +4 +6 +7	76 +3 +4	77 +3 +4	78 +2 +4	79 +1 +3	80 +1 +2	81 +1 +3	82 +2 +4	83 +3 +5	84 +2 +4	85	86 0
Cs	Ba		Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
132.90545	17.327		178.49	180.9479	183.84	186.207	190.73	192.217	195.078	196.9665	200.59	204.3833	207.2	208.98038	[209]	[210]	[222]
-18-8-1	-18-8-2		-32-10-2	-32-11-2	-32-12-2	-32-13-2	-32-14-2	-32-15-2	-32-16-2	-32-17-1	-32-18-2	-32-18-3	-32-18-4	-32-18-5	-32-18-6	-32-18-7	-32-18-8

-N-O-P

Period 7

87 +1	88 +2	Ac-Lr	104 +4	105	106	107	108	109	110	111							
Fr	Ra		Rf	Db	Sg	Bh	Hs	Mt	Uun	Uuu							
[223]	[226]		[261]	[262]	[263]	[262]	[265]	[266]	[269]	[272]							
-18-8-1	-18-8-2		-32-10-2	-32-11-2	-32-12-2	-32-13-2	-32-14-2	-32-15-2	-32-16-2	-32-18-1							

-O-P-Q

Alkali metals · Alkaline earth metals · Pnicogens · Chalcogens · Halogens · Noble gases · Electron orbit

Lanthanides

57 +3	58 +3 +4	59 +3	60 +3	61 +3	62 +2 +3	63 +2 +3	64 +3	65 +3	66 +3	67 +3	68 +3	69 +3	70 +2 +3	71 +3
La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
138.9055	140.116	140.9077	144.24	[145]	150.36	151.964	157.25	158.92534	162.50	164.9303	167.26	168.93421	173.04	174.967
-18-9-2	-20-8-2	-21-8-2	-22-8-2	-23-8-2	-24-8-2	-25-8-2	-25-9-2	-27-8-2	-28-8-2	-29-8-2	-30-8-2	-31-8-2	-32-8-2	-32-9-2

-N-O-P

Actinides

89 +3	90 +4	91 +5 +4	92 +3 +4 +5 +6	93 +3 +4 +5 +6	94 +3 +4 +5 +6	95 +3 +4 +5 +6	96 +3	97 +3 +4	98 +3	99 +3	100 +3	101 +2 +3	102 +2 +3	103 +3
Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr
[227]	232.0381	231.0359	238.0289	[237]	[244]	[243]	[247]	[247]	[251]	[252]	[257]	[258]	[259]	[262]
-18-9-2	-18-10-2	-20-9-2	-21-9-2	-22-9-2	-24-8-2	-25-8-2	-25-9-2	-27-8-2	-28-8-2	-29-8-2	-30-8-2	-31-8-2	-32-8-2	-32-9-2

-O-P-Q

Fig. 1 Periodic Table of Elements

Periodic System for Ferrous Metallurgists

The basic principles of alloying were applied to develop the periodic system shown here, which illustrates the fundamental alloying nature of iron. The solid solubility of each element in iron can be resolved accurately with few exceptions by considering only atomic size. Alloying valence, crystal structure and electronegativity are useful supplementary factors of varying significance. The tendency to form compounds, intermetallic or ionic, is related to the difference in electronegativity which, in general, increases in a sweep from lower left to upper right of the periodic system.

Adapted Primarily for Ferrous Metallurgists

Atomic size factors (in parentheses) are % smaller (–) or larger (+) than gamma (FCC) iron at 75 F. Lattice environment (Coordination No.) is taken into account; CN is 12 except 6 for interstitials H, B, C, N & O. Groups VI, VIb, VII and VIIb form ionic compounds with the metals.

0	I	II	III		IV	V	VI	VII
				H-1 ▲ (–58) ⊗ XX				
He-2 FCC (Others)	Li-3 ⊗ (+23) BCC* HCP†	Be-4 ● (–11) HCP* BCC	B-5 ▲ (–29) ⊗ XX		C-6 ▲ (–34) ⊗ XX	N-7 ▲ (–36) ⊗ XX	O-8 ▲ (–33) ⊗ XX	F-9
Ne-10 FCC	Na-11 ⊗ (+50) BCC* HCP	Mg-12 ⊗ (+27) HCP	Al-13 ◑ (+14) FCC		Si-14 ● (+7) XX	P-15 ● (+2) XX	S-16 ● (+1) XX	Cl-17 XX

0	Ia	IIa	IIIa	IVa	Va	VIa	VIIa	VIII			Ib	IIb	IIIb	IVb	Vb	VIb	VIIb
Ar-18 FCC	K-19 ⊗ (+86) BCC	Ca-20 ⊗ (+56) FCC* BCC	Sc-21 ⊗ (+29) HCP* BCC	Ti-22 ◑ (+16) HCP* BCC	V-23 ● (+6) BCC	Cr-24 ● (+1) BCC	Mn-25 XX* FCC ‡	Fe-26 ● (0) BCC* FCC	Co-27 ● (–1) HCP* FCC	Ni-28 ● (–1) FCC	Cu-29 ● (+1) FCC	Zn-30 ● (+6) HCP	Ga-31 ● (+12) XX	Ge-32 ● (+9) XX	As-33 ● (+11) XX	Se-34 ● (+11) XX	Br-35 XX
Kr-36 FCC	Rb-37 ⊗ (+97) BCC	Sr-38 ⊗ (+71) FCC* HCP ‡	Y-39 ⊗ (+42) HCP* BCC	Zr-40 ⊗ (+27) HCP* BCC	Cb-41 ◑ (+15) BCC	Mo-42 ◑ (+10) BCC	Tc-43 ● (+8) HCP	Ru-44 ● (+6) HCP	Rh-45 ● (+6) FCC	Pd-46 ● (+9) FCC	Ag-47 ◑ (+14) FCC	Cd-48 ⊗ (+20) HCP	In-49 ⊗ (+25) XX	Sn-50 ⊗ (+23) XX	Sn-51 ⊗ (+27) XX	Te-52 ⊗ (+27) XX	I-53 XX
Xe-54 FCC	Cs-55 ⊗ (+112) BCC	Ba-56 ⊗ (+76) BCC	La-57 ⊗ (+48) HCP* FCC ‡	Hf-72 ⊗ (+26) HCP* BCC	Ta-73 ◑ (+16) BCC	W-74 ◑ (+16) BCC	Re-75 ● (+11) BCC	Os-76 ● (+7) HCP	Ir-77 ● (+8) FCC	Pt-78 ● (+10) FCC	Au-79 ◑ (+14) FCC	Hg-80 ⊗ (+25) XX	Tl-81 ⊗ (+36) HCP* BCC	Pb-82 ⊗ (+39) FCC	Bi-83 ⊗ (+35) XX	Po-84 ⊗ (+40) XX	At-85
Rn-86	Fr-87	Ra-88	Ac-89 ⊗ (+49) FCC				Note 1: The rare-earth (lanthanide, 58-71) and actinide (90-103) series are omitted. Note 2: Valence is 4 for C; 3 for N and P. Note 3: (1) and (2) are not alloying valences.										
Alloying Valence	1	2	3	4	5	6	6	6	6	6	5.56	4.56	3.56	2.56 Note 2	1.56 Note 2	(2) Note 3	(1) Note 3

Substitutional solid solutions

- ● FAVORABLE SIZE FACTOR: 0 TO ± 13%
- ◑ BORDERLINE SIZE FACTOR: ± 14 TO ± 16%
- ⊗ UNFAVORABLE SIZE FACTOR: > ± 16%

Interstitial solid solutions

- ▲ FAVORABLE SIZE FACTOR: > (–40%)
- ▲ BORDERLINE SIZE FACTOR: (–30) TO (–40%)
- ▲ UNFAVORABLE SIZE FACTOR: < (–30%)

Structure

BCC – BODY CENTERED CUBIC
FCC – FACE CENTERED CUBIC
HCP – HEXAGONAL CLOSE PACKED
XX – NOT BCC, FCC OR HCP USUALLY MORE COMPLEX
* – STRUCTURE AT 75 F
† – ALSO FCC   ‡ – ALSO BCC

Type of gamma iron (FCC) field if alloyed with iron

- GAMMA LOOP, LIKE Cr
- LIMITED GAMMA LOOP, LIKE B
- OPEN GAMMA REGION, LIKE Ni
- LIMITED GAMMA REGION, LIKE C

Fig. 1 Source: *ASM Metals Reference Book,* 2nd ed., American Society for Metals, 1983, p 83

Physical Constants and Physical Properties of the Elements

Table 1 Fundamental physical constants

Quantity	Symbol	Numerical value(a)	Units
Speed of light (in vacuum)	c	299,792,458 (exact)	$m \cdot s^{-1}$
Electronic charge	e	$1.602176487\ (40) \times 10^{-19}$	C
Planck constant	h	$6.62606896\ (33) \times 10^{-34}$	$J \cdot s$
Pianck constant/2π	$h(bar)$	$1.054571628\ (53) \times 10^{-34}$	$J \cdot s$
Avogadro constant (number)	N_A	$6.02214179\ (30) \times 10^{23}$	mol^{-1}
Atomic mass unit	amu or u	$1.660538782\ (83) \times 10^{-27}$	kg
Electron rest mass	m_e	$9.10938215\ (45) \times 10^{-31}$	kg
Proton rest mass	m_p	$1.672621637\ (83) \times 10^{-27}$	kg
Neutron rest mass	m_n	$1.674927211\ (84) \times 10^{-27}$	kg
Faraday constant ($N_A e$)	F	9.64853399 (24)	$C \cdot mol^{-1}$
Electron magnetic moment	μ_e	$-9.28476377\ (23) \times 10^{-26}$	$J \cdot T^{-1}$
Molar gas constant	R	8.314472 (15)	$J \cdot mol^{-1} \cdot K^{-1}$
Molar value of ideal gas at STP(b)	V_m	$22.710981\ (40) \times 10^{-3}$	$m^3\ mol^{-1}$
Boltzmann constant (R/N_A)	k	$1.3806504\ (24) \times 10^{-23}$	$J \cdot K^{-1}$
Stefan-Boltzmann constant	σ	$5.670400\ (40) \times 10^{-8}$	$W\ m^{-2}\ K^{-4}$
Standard gravity (gravitational acceleration)	g	9.80665 (exact)	$m \cdot s^{-2}$
Energy equivalent, thermochemical calorie	cal	4.184 (exact)	J
Absolute temperature	T_0 °C	273.150 ± 0.010	K

(a) The number in parenthesis is the estimated one-standard-deviation uncertainty in the last digits of the quoted value. (b) STP = standard temperature and pressure (273.15 K, 100 kPa). Source: NIST, current (2006) CODATA recommended values, accessed June 2, 2009

Table 2   Standard atomic weights of the elements

Symbol	Atomic number	Atomic weight	Symbol	Atomic number	Atomic weight	Symbol	Atomic number	Atomic weight
Ac	89	227.0278	He	2	4.00260	Ra	88	226.0254
Al	13	26.98154	Ho	67	164.9304	Rn	86	(222)
Am	95	(243)	H	1	1.0079	Re	75	186.207
Sb	51	121.75*	In	49	114.82	Rh	45	102.9055
Ar	18	39.948	I	53	126.9045	Rb	37	85.4678*
As	33	74.9216	Ir	77	192.22*	Ru	44	101.07*
At	85	(210)	Fe	26	55.847*	Sm	62	150.36*
Ba	56	137.33	Kr	36	83.80	Sc	21	44.9559
Bk	97	(247)	La	57	138.9055*	Se	34	78.96*
Be	4	9.01218	Lr	103	(260)	Si	14	28.0855*
Bi	83	208.9804	Pb	82	207.2	Ag	47	107.868
B	5	10.81	Li	3	6.941*			
Br	35	79.904	Lu	71	174.967*	Na	11	22.98977
Cd	48	112.41	Mg	12	24.305	Sr	38	87.62
Cs	55	132.9054	Mn	25	54.9380	S	16	32.06
Ca	20	40.08	Md	101	(258)	Ta	73	180.9479
Cf	98	(251)	Hg	80	200.59*	Tc	43	(98)
C	6	12.011	Mo	42	95.94	Te	52	127.60*
Ce	58	140.12	Nd	60	144.24*	Tb	65	158.9254
Cl	17	35.453	Ne	10	20.179	Tl	81	204.383
Cr	24	51.996	Np	93	237.0482	Th	90	232.0381
Co	27	58.9332	Ni	28	58.69	Tm	69	168.9342
Cu	29	63.546*	Nb	41	92.9064	Sn	50	118.69*
Cm	96	(247)	N	7	14.0067	Ti	22	47.88*
Dy	66	162.50*	No	102	(259)			
Es	99	(252)	Os	76	190.2	W	74	183.85*
Er	68	167.26*	O	8	15.9994*	(Unh)	106	(263)
Eu	63	151.96	Pd	46	106.42	(Unp)	105	(262)
Fm	100	(257)	P	15	30.97376	(Unq)	104	(261)
F	9	18.998403	Pt	78	195.08*	U	92	238.0289
Fr	87	(223)	Pu	94	(244)	V	23	50.9415
Gd	64	157.25*	Po	84	(209)	Xe	54	131.29*
Ga	31	69.72	K	19	39.0983	Yb	70	173.04*
Ge	32	72.59*	Pr	59	140.9077	Y	39	88.9059
Au	79	196.9665	Pm	61	(145)	Zn	30	65.38
Hf	72	178.49*	Pa	91	231.0359	Zr	40	91.22

Note: The atomic weights are in atomic mass units (amu) relative to $^{12}C = 12$. The atomic weights of many elements are not invariant but depend on the origin and treatment of the material. The values given apply to elements as they exist naturally on earth and to certain artificial elements. They are considered reliable to ± 1 in the last digit or ± 3 when followed by an asterisk (*). Values in parentheses are used for radioactive elements whose atomic weights cannot be quoted precisely without knowledge of the origin of the elements; the value given is the atomic mass number of the isotope of that element of longest known half-life. Source: M. Bauccio, Ed., *ASM Metals Reference Book*, 3rd ed., ASM International, 1993

Table 3   Melting points of the elements

Symbol	Atomic number	Melting point K	Melting point °C	Estimated error; footnotes	Symbol	Atomic number	Melting point K	Melting point °C	Estimated error; footnotes
Ac	89	1324	1051	$\pm$50	Mn	25	1519	1246	$\pm$5
Ag	47	1235.08	961.93	(a)	Mo	42	2896	2623	(b)
Al	13	933.602	660.452	(b)	N	7	63.1458	−210.0042	$\pm$0.0002; (c, d)
Am	95	1449	1176	(e)	Na	11	371.0	97.8	$\pm$0.1
Ar	18	83.798	−189.352	(a, c)	Nb	41	2742	2469	(b)
As	33	876	603	(f)	Nd	60	1294	1021	(e)
At	85	(575)	(302)	est.; (g)	Ne	10	24.563	−248.587	$\pm$0.002; (b, c)
Au	79	1337.58	1064.43	(a)	Ni	28	1728	1455	(b)
B	5	2365	2092	(g)	No	102	(1100)	(827)	est.; (g)
Ba	56	1002	729	$\pm$2	Np	93	910	637	$\pm$2
Be	4	1562	1289	$\pm$5	O	8	54.361	−218.789	(a, c)
Bi	83	544.592	271.442	(b)	Os	76	3306	3033	$\pm$20
Bk	97	1256	983	(e)	P	15	317.29	44.14	$\pm$0.1; (h)
Br	35	265.90	−7.25	(c)	Pa	91	1848	1575	(g)
C	6	4100	3826	(f)	Pb	82	600.652	327.502	(b)
Ca	20	1113	840	$\pm$2	Pd	46	1828	1555	$\pm$0.4; (d)
Cd	48	594.258	321.108	(b)	Pm	61	1315	1042	(e)
Ce	58	1071	798	$\pm$3; (e)	Po	84	527	254	(g)
Cf	98	1213	940	(e)	Pr	59	1204	931	(e)
Cl	17	172.18	−100.97	(c)	Pt	78	2042.1	1769.0	(b)
Cm	96	1613	1340	(g)	Pu	94	913	640	$\pm$1
Co	27	1768	1495	(b)	Ra	88	973	700	(g)
Cr	24	2133	1860	$\pm$20	Rb	37	312.63	39.48	$\pm$0.5
Cs	55	301.54	28.39	$\pm$0.05	Re	75	3459	3186	$\pm$20
Cu	29	1358.02	1084.87	$\pm$0.04; (d)	Rh	45	2236	1963	(b)
Dy	66	1685	1412	(e)	Rn	86	202	−71	...
Er	68	1802	1529	(e)	Ru	44	2607	2334	$\pm$10; (d)
Es	99	1093	820	(e)	S	16	388.37	115.22	...
Eu	63	1095	822	(e)	Sb	51	903.905	630.755	(b)
F	9	53.48	−219.67	(c)	Sc	21	1814	1541	(e)
Fe	26	1808	1535	(b)	Se	34	494	221	...
Fm	100	(1800)	(1527)	est.; (g)	Si	14	1687	1414	$\pm$2
Fr	87	(300)	(27)	est.; (g)	Sm	62	1347	1074	(e)
Ga	31	302.9241	29.7741	$\pm$0.001; (c,d)	Sn	50	505.1181	231.9681	(a)
Gd	64	1586	1313	(e)	Sr	38	1042	769	...
Ge	32	1211.5	938.3	...	Ta	73	3293	3020	...
H	1	13.81	−259.34	(a, c)	Tb	65	1629	1356	(e)
He	2	4.215	−268.935	(j)	Tc	43	2477	2204	$\pm$50
Hf	72	2504	2231	$\pm$20	Te	52	722.72	449.57	$\pm$0.3
Hg	80	234.314	−38.836	(b)	Th	90	2031	1758	$\pm$10
Ho	67	1747	1474	(e)	Ti	22	1943	1670	$\pm$6; (d)
I	53	386.7	113.6	(c)	Tl	81	577	304	$\pm$2
In	49	429.784	156.634	(b)	Tm	69	1818	1545	(e)
Ir	77	2720	2447	(b)	U	92	1407	1134	...
K	19	336.34	63.19	$\pm$0.5	V	23	2202	1929	$\pm$6
Kr	36	115.765	−157.385	$\pm$0.001; (d)	W	74	3695	3422	(b)
La	57	1191	918	(e)	Xe	54	161.3918	−111.7582	$\pm$0.0002; (c, d)
Li	3	453.7	180.6	$\pm$0.5	Y	39	1795	1522	(e)
Lr	103	(1900)	(1627)	est.; (g)	Yb	70	1092	819	(e)
Lu	71	1936	1663	(e)	Zn	30	692.73	419.58	(a)
Md	101	(1100)	(827)	est.; (g)	Zr	40	2128	1855	$\pm$5
Mg	12	922	649	$\pm$0.5					

Note: The melting points, except those footnoted to indicate otherwise, are derived from R. Hultgren, P.D. Desai, D.T. Hawkins, M. Gleiser, K.K. Kelley and D.D. Wagman, *Selected Values of the Thermodynamic Properties of the Elements*, American Society for Metals, 1973, which are based on the 1948 International Practical Temperature Scale. Values have been corrected to the 1968 scale (IPTS-68). Except for triple points, values are for a pressure of 1 atm. (Note that melting and freezing points should be identical for pure elements.) (a) Defined fixed point on 1968 International Practical Temperature Scale (IPTS-68): Amended Edition of 1975, *Metrologia*, Vol 12, 1976, p 7–17 (b) Secondary reference point in Extended List of Secondary Reference Points on 1968. International Practical Temperature Scale (IPTS-68), *Metrologia*, Vol 13, 1977, p 197–206 (c) Triple point. (d) Secondary reference point 1980 supplement to 1977 Extended List of Secondary Reference Points on 1968 International Practical Temperature Scale (IPTS-68): Amended Edition of 1975, Report 5 of Working Group II, April 1980. (e) From *Metals Handbook*, 9th ed., Vol 2, American Society for Metals, 1979. (f) Sublimation point at atmospheric pressure. (g) R.H. Lamoreaux Melting Point, Gram-Atomic Volumes and Enthalpies of Atomization for Liquid Elements, nLBL Report 4995, 1976. (h) Melting point for white α-P; red P sublimes without melting at atmospheric pressure and has a triple point of 862.8 K (589.7 °C). (j) Boiling point at 1 atm; there are various triple points; see Hultgren et al., 1973

Table 4   Atomic size parameters for the elements

Element	Atomic number	Pearson symbol	Atomic volume (Ω), nm³	Interatomic distance (S₀), nm	Equivalent atomic radius (r₀), nm	Notes	Element	Atomic number	Pearson symbol	Atomic volume (Ω), nm³	Interatomic distance (S₀), nm	Equivalent atomic radius (r₀), nm	Notes
Ac	89	cF4	0.037451	0.3755	0.2076	...	βNa	11	cl2	0.039493	0.3716	0.2113	...
Ag	47	cF4	0.017056	0.2889	0.1597	...	Nb	41	cl2	0.017980	0.2859	0.1625	...
Al	13	cF4	0.016603	0.2864	0.1583	...	αNd	60	hP4	0.034179	0.3322	0.2013	...
αAm	95	hP4	0.029271	0.3451	0.1911	...	Ne	10	cF4	0.022212	0.3155	0.1744	(a)
Ar	18	cF4	0.037473	0.3756	0.2076	(a)	Ni	28	cF4	0.010942	0.2492	0.1377	...
αAs	33	hR2	0.021518	0.2517	0.1726	...	αNp	93	oP8	0.019224	0.2599	0.1662	...
Au	79	cF4	0.016959	0.2884	0.1594	...	αO	8	mC4	0.017360	(0.115)	(0.1606)	(a, b)
γB	5	tP50	0.007786	(0.1624)	(0.1230)	(c)	Os	76	hP2	0.013988	0.2735	0.1495	...
Ba	56	cl2	0.063367	0.4350	0.2473	...	P (black)	15	oC8	0.018993	(0.2224)	(0.1655)	(c)
αBe	4	hP2	0.008108	0.2225	0.1246	...	αPa	91	tl2	0.025212	0.3214	0.1819	...
αBi	83	hR2	0.035384	0.3071	0.2037	...	Pb	82	cF4	0.030326	0.3500	0.1935	...
αBk	97	hP4	0.027965	0.3398	0.1883	...	Pd	46	cF4	0.014717	0.2751	0.1520	...
Br	35	oC8	0.03277	(0.227)	(0.199)	(a, b)	Pm	61	hP4	0.03360	0.330	0.200	...
C(graph)	6	hP4	0.008800	(0.1421)	(0.1281)	(c)	αPo	84	cP1	0.038137	0.3366	0.2088	...
αCa	20	cF4	0.043631	0.3952	0.2184	...	αPr	59	hP4	0.034545	0.3338	0.2020	...
Cd	48	hP2	0.021581	0.2979	0.1727	...	Pt	78	cF4	0.015097	0.2774	0.1533	...
αCe	58	cF4	0.034367	0.3650	0.2017	...	αPu	94	mP16	0.019998	0.257	0.1684	...
Cl	17	oC8	0.02886	(0.198)	(0.190)	(a,b)	Ra	88	cl2	0.068216	0.4459	0.2535	...
αCm	96	hP4	0.029984	0.3479	0.1927	...	Rb	37	cl2	0.092743	0.4939	0.2808	...
αCo	27	hP2	0.011076	0.2497	0.1383	...	Re	75	hP2	0.014713	0.2740	0.1520	...
Cr	24	cl2	0.012003	0.2498	0.1420	...	Rh	45	cF4	0.013753	0.2689	0.1486	...
Cs	55	cl2	0.115794	0.5318	0.3024	...	Ru	44	hP2	0.013568	0.2650	0.1480	...
Cu	29	cF4	0.011809	0.2238	0.1413	...	αS	16	oF128	0.025754	(0.2037)	(0.1832)	(c)
αDy	66	hP2	0.031558	0.3504	0.1960	...	αSb	51	hR2	0.030201	0.2908	0.1932	...
αEr	68	hP2	0.030636	0.3467	0.1941	...	αSc	21	hP2	0.024974	0.3254	0.1813	...
Eu	63	cl2	0.048121	0.3969	0.2256	...	γSe	34	hP3	0.027274	0.2374	0.1867	...
αF	9	mC8	0.01605	(0.149)	(0.197)	(a, b)	Si	14	cF8	0.020020	0.2352	0.1684	...
αFe	26	cl2	0.011777	0.2483	0.1411	...	αSm	62	hR3	0.033202	0.3587	0.1994	...
αGa	31	oC8	0.019580	0.2484	0.1672	...	βSn	50	tl4	0.027049	0.3022	0.1862	...
αGd	64	hP2	0.033050	0.3572	0.1991	...	αSr	38	cF4	0.056299	0.4302	0.2378	...
αGe	32	cF8	0.022634	0.2450	0.1755	...	Ta	73	cl2	0.018019	0.2861	0.1626	...
αH	1	hP2	0.037882	(0.3768)	(0.2083)	(a,b)	αTb	65	hP2	0.032066	0.3528	0.1971	...
αHe	2	hP2	0.032367	0.3577	0.1977	(a)	Te	43	hP2	0.014264	0.2707	0.1505	...
αHf	72	hP2	0.022321	0.3127	0.1747	...	αTe	52	hP3	0.033969	0.2834	0.2009	...
αHg	80	hR1	0.023354	0.2993	0.1773	(a)	αTh	90	cF4	0.032873	0.3596	0.1987	...
αHo	67	hP2	0.031139	0.3487	0.1952	...	αTi	22	hP2	0.017653	0.2986	0.1615	...
I	53	oC8	0.042696	(0.269)	(0.2168)	(b)	αTl	81	hP2	0.028586	0.3408	0.1897	...
In	49	tl2	0.026158	0.3252	0.1842	...	αTm	69	hP2	0.030006	0.3447	0.1928	...
Ir	77	cF4	0.014146	0.2715	0.1500	...	αU	92	oC4	0.020747	0.2753	0.1705	...
K	19	cl2	0.075327	0.4608	0.2620	...	V	23	cl2	0.013824	0.2619	0.1489	...
Kr	36	cF4	0.044992	0.3992	0.2206	(a)	W	74	cl2	0.015844	0.2741	0.1558	...
αLa	57	hP4	0.037532	0.3456	0.2077	...	Xe	54	cF4	0.057463	0.4336	0.2396	(a)
βLi	3	cl2	0.021609	0.3309	0.1728	...	αY	39	hP2	0.033033	0.3557	0.1991	...
αLu	71	hP2	0.029524	0.3434	0.1917	...	αYb	70	cF4	0.041250	0.3878	0.2143	...
Mg	12	hP2	0.023239	0.3197	0.1770	...	Zn	30	hP2	0.015214	0.2664	0.1537	...
αMn	25	cl58	0.012245	0.2258	0.1430	...	αZr	40	hP2	0.023279	0.3179	0.1771	...
Mo	42	cl2	0.015583	0.2745	0.1550	...							
αN	7	cP8	0.022653	(0.1098)	(0.1755)	(a, b)							

Note: The atomic size parameters were derived from the crystal structure data for the Elements given in *Bull. Alloy Phase Diagrams*, Vol 2 (No.3), 1981, p 401–402. The volume per atom of the structure, Ω, was derived from the room-temperature lattice parameter data by calculating the volume of the unit cell and dividing by the number of atoms contained within the unit cell, which is given by the numerals in the Pearson symbol. The closest distance of approach, S_0, was derived from the minimum interatomic distances in the unit cell, except for the diatomic gases and the nonmetallic elements that exist in molecular form. The atomic radius, r_0, was derived from the volume per atom data, using the relationship $4/3\pi r_0^3$ = volume per atom. (a) These elements are gaseous, or liquid, at room temperature. The structural data for Ar, H, Kr, Ne, and Xe refer to 4.2 K, whereas that for Br, Cl, He, Hg, N, and O refer to 123 K, 113 K, 1.5 K, 227 K, 20 K, and 23 K, respectively. (b) These elements form diatomic gases, and the basis of the crystal structure is therefore the molecular unit R_2. The volume-per-atom data are thus more meaningful if considered in terms of two times the volume per molecule, whereas the values of the parameter r_0 (listed in parentheses) should not be equated with atomic radii in this context. The values of S_0 (also listed in parentheses) refer to interatomic distances in the covalently bonded molecule R_2, rather than distance between equipositioned neighbors in a crystal. (c) The chemistry of these elements permits them to form a number of allotropes at room temperature, with crystal structures based on different molecular bases. The previous comments on volume per atom and r_0 are thus also applicable to these elements, because they form crystal structures based on chains or network layers

Table 5 Heats of transition of the elements

Element	Atomic number	Transformation	Enthalpy (ΔH), J/mol	Temperature, °C	Element	Atomic number	Transformation	Enthalpy (ΔH), J/mol	Temperature, °C
Ag	47	L ↔ S	11,300	961.93(a)	Nd	60	β ↔ α	3030	855
Al	13	L ↔ S	10,700	660.457(b)	Ne	10	L ↔ S	331.7	24.561 K(c)
Am	95	L ↔ γ	14,395	1,176	Ni	28	L ↔ S	17,470	1,455(b)
		γ ↔ β	5,860	1,077	Np	93	L ↔ γ	5,190	639
		β ↔ α	775	650			γ ↔ β	5,270	576
Ar	18	L ↔ S	1,190	83.798 K(d)			β ↔ α	5,605	280
Au	79	L ↔ S	13,000	1,064.43(a)	O	8	L ↔ γ	223	54.361 K(d)
B	5	L ↔ β	50,200	2,077			γ ↔ β	371.3	43.801 K
Ba	56	L ↔ S	7,120	727			β ↔ α	48.4	23.867 K
Be	4	L ↔ β	(12,600)	1,287	Os	76	L ↔ S	(31,800)	3,025
		β ↔ α	(2,100)	1,277	P(white α)	15	L ↔ α	629	44
Bi	83	L ↔ S	11,300	271.442(c)	Pa	91	L ↔ β	12,340	1,572
Br	35	L ↔ S	5,286	265.9 K			β ↔ α	6,640	1,170
Ca	20	L ↔ β	8,540	842	Pb	82	L ↔ S	4,800	327.502(b)
		β ↔ α	842	443	Pd	46	L ↔ S	(17,560)	1,554(b)
Cd	48	L ↔ S	6,200	321.108(b)	Pm	61	L ↔ S	(7,550)	...
Ce	58	L ↔ δ	5,460	800			B ↔ A	(2,900)	...
		δ ↔ γ	2,990	725	Pr	59	L ↔ β	6,890	930
		γ ↔ β	190	...			β ↔ α	3,170	795
		β ↔ α	1,950	...	Pt	78	L ↔ S	(19,650)	1,769(b)
Cl	17	L ↔ S	3,203	172.16 K	Pu	94	L ↔ ε	2,825	640
Cm	96	L ↔ β	14,645	1,345			ε ↔ δ′	1,840	479
		β ↔ γ	3,245	1,277			δ′ ↔ δ	80	457
Co	27	L ↔ β	16,200	1,495(b)			δ ↔ γ	585	315
		β ↔ α	450	427			γ ↔ β	565	207
Cr	24	L ↔ S	(20,500)	1,857			β ↔ α	3,375	122
Cs	55	L ↔ S	2,090	28.44	Rb	37	L ↔ S	2,190	39.32
Cu	29	L ↔ S	13,050	1,084.88(b)	Re	75	L ↔ S	(33,230)	3,180
Dy	66	L ↔ β	11,060	1,409	Rh	45	L ↔ S	(21,490)	1,963(b)
		β ↔ α	4,160	1,385	Rn	86	L ↔ S	(2,890)	−71
Er	68	L ↔ S	19,900	1,522	Ru	44	L ↔ S	(24,280)	2,250
Eu	63	L ↔ S	9,210	817	S	16	L ↔ β	1,718	115
F	9	L ↔ β	255	53.48 K			β ↔ α	402	95
		β ↔ α	364	45.55 K	Sb	51	L ↔ S	19,900	630.775(b)
Fe	26	L ↔ δ	13,800	1,535(d)	Sc	21	L ↔ β	14,100	1,539
		δ ↔ γ	840	1,392			β ↔ α	4,010	1,335
		γ ↔ α	900	911	Se	34	L ↔ S	6,700	220
Ga	31	L ↔ S	5,565	29.771(b)	Si	14	L ↔ S	50,210	1,417
Gd	64	L ↔ β	10,050	1,312	Sm	62	L ↔ β	8,620	1,072
		β ↔ α	3,910	1,260			β ↔ α	3,110	917
Ge	32	L ↔ S	37,030	937	Sn	50	L ↔ β	7,195	231.9681(a)
H	1	L ↔ S	58.68	13.81 K(d)	Sr	38	L ↔ γ	7,431	777
Hf	72	L ↔ S	(29,300)	2,227			γ ↔ α	837	547
		β ↔ α	(5,910)	1,781	Ta	73	L ↔ S	36,570	2,985
Hg	80	L ↔ α	2,295	−38.836(b)	Tb	65	L ↔ β	10,800	1,355
Ho	67	L ↔ β	(16,900)	1,470			β ↔ α	5,020	1,285
1	53	L ↔ S	7,820	113.5	Te	52	L ↔ S	17,490	449.5
In	49	L ↔ S	3,280	156.634(b)	Th	90	L ↔ β	13,807	1,750
Ir	77	L ↔ S	(26,140)	2,447(d)			β ↔ α	3,599	1,360
K	19	L ↔ S	2,320	63.71	Ti	22	L ↔ β	14,150	1,663
Kr	36	L ↔ S	1,638	115.770 K(c)			β ↔ α	4,170	893
La	57	L ↔ γ	6,200	920	TI	81	L ↔ β	4,200	303
		γ ↔ β	3,120	860			β ↔ α	360	234
		β ↔ α	360	275	Tm	69	L ↔ S	16,840	1,545
Li	3	L ↔ β	3,000	180.54	U	92	L ↔ γ	9,142	1,135
Lu	71	L ↔ S	(18,650)	1,663			β ↔ γ	4,757	776
Mg	12	L ↔ S	8,477	650			α ↔ β	2,791	669
Mn	25	L ↔ δ	(12,060)	1,245	V	23	L ↔ S	22,845	1,917
		δ ↔ γ	1,880	1,135	W	74	L ↔ S	46,000	3,422(b)
		γ ↔ β	2,120	1,085	Xe	54	L ↔ S	2,300	161.388 K(c)
		β ↔ α	2,230	700	Y	39	L ↔ β	11,400	1,525
Mo	42	L ↔ S	35,980	2,623(b)			β ↔ α	4,990	1,480
N	7	L ↔ β	360.4	63.146 K(c)	Yb	70	L ↔ β	7,660	824
		β ↔ α	116	35.61 K			β ↔ α	1,750	760
Na	11	L ↔ β	2,600	97.86	Zn	30	L ↔ S	7,320	419.58(a)
Nb	41	L ↔ S	(26,900)	2473(b)	Zr	40	L ↔ β	20,920	1,855(b)
Nd	60	L ↔ β	7,140	1,015			β ↔ α	4,015	862

(a) Melting points or freezing points, which are defined fixed points of IPTS-68. (b) Melting points or freezing points, which are secondary reference points of IPTS-68. (c) Triple-point values, which are secondary reference points of IPTS-68. (d) Triple-point values, which are defined fixed points of IPTS-68.

Table 6 Thermal properties of the elements

Symbol	Coefficient of linear thermal expansion(a), 10^{-6}/°C (10^{-6}/°F)	Thermal conductivity(a), cal/(cm · s · °C)	Specific heat(b) cal/g · °C (J/kg · K)	Heat of fusion, cal/g (Btu/lb)
Ac	… (…)	…	… (…)	… (…)
Al	23.6(c) (13.1)(c)	0.53	0.215 (900)	94.5 (170)
Am	… (…)	…	… (…)	… (…)
Sb	8.5 to 10.8(d) (4.7 to 6)(d)	0.045	0.049 (205)	38.3 (68.9)
A	… (…)	0.406×10^{-4}	0.125 (523)	6.7 (12)
As	4.7 (2.6)	…	0.082 (343)	88.5 (159.3)
At	… (…)	…	… (…)	… (…)
Ba	… (…)	…	0.068 (285)	… (…)
Bk	… (…)	…	… (…)	… (…)
Be	11.6(e) (6.4)(e)	0.35	0.45 (190)	260 (470)
Bi	13.3 (7.4)	0.020	0.0294 (123)	12.5 (22.5)
B	8.3(f) 4.6(f)	…	0.309 (1290)	… (…)
Br	… (…)	…	0.070 (290)	16.2 (29.2)
Cd	29.8 (16.55)	0.22	0.055 (230)	13.2 (23.8)
Ca	22.3(g) (12.4)(g)	0.3	0.149(h) (624)(h)	52 (93.6)
Cf	… (…)	…	… (…)	… (…)
C	0.6 to 4.3(c) (0.3 to 2.4)(c)	0.057	0.165 (691)	… (…)
Ce	8 (4.44)	0.026(i)	0.045 (190)	8.5 (15.9)
Cs	97(j) (54)(j)	…	0.04817 (201.7)	3.8 (6.8)
Cl	… (…)	0.172×10^{-4}	0.116 (486)	21.6 (38.9)
Cr	6.2 (3.4)	0.16	0.11 (460)	96 (173)
Co	13.8 (7.66)	0.165	0.099 (410)	58.4 (105)
Cu	16.5 (9.2)	0.941 ± 0.005	0.092 (380)	50.6 (91.1)
Cm	… (…)	…	… (…)	… (…)
Dy	9 (5) (…)	0.024(i)	0.041 (170) (…)	25.2 (45.4) (…)
Er	9 (5)	0.023(i)	0.040 (170)	24.5 (44.1)
Eu	26 (14.44)	…	0.039 (160)	16.5 (29.6)
Fm	… (…)	…	… (…)	… (…)
F	… (…)	…	0.18 (750)	10.1 (18.2)
Fr	… (…)	…	… (…)	… (…)
Gd	4(k) (2.22)(k)	0.021(i)	0.071 (360)	23.5 (42.4)
Ga	18(l) (10)(l)	0.07 to 0.09(m)	0.079 (330)	19.16 (34.49)
Ge	5.75	0.14	0.073	…

(continued)

(a) Near 20 °C (68 °F). (b) At 20 °C (68 °F). (c) From 20 to 100 °C (68 to 212 °F). (d) From 20 to 60 °C (68 to 140 °F). (e) From 25 to 100 °C (77 to 212 °F) (f) From 20 to 750 °C (68 to 1380 °F). (g) For α at 0 to 400 °C (32 to 750 °F). (h) From 0 to 100 °C (32 to 212 °F). (i) At 28 °C (82 °F). (j) From 0 to 26 °C (32 to 70 °F). (k) Near 40 °C (105 °F); the coefficient of expansion of gadolinium changes rapidly between −100 and +100 °C (−150 and +212 °F). (l) From 0 to 30 °C (32 to 86 °F). (m) At melting point. (n) At 18 °C (64 °F). (o) From 20 to 200 °C (68 to 390 °F). (p) W/cm/°C at 50 °C (120 °F). (q) At 25 °C (77 °F) for high-purity k iron. (r) For ingot iron at 0 °C (32 °F). (s) Annealed. (t) From 17 to 100 °C (63 to 212 °F). (u) At 0 °C (32 °F). (v) Along a-axis; 24.3 along c-axis. (w) α; γ, 14; both from 0 to 100 °C (32 to 212 °F). (x) α; γ, 0.120; both at 25.2 °C (77.3 °F). (y) Estimated (z) At −2.22 °C (28 °F). (aa) At 25 °C (77 °F). (bb) At 50 °C (122 °F), parallel to a-axis, mean value; parallel to c-axis at 50 °C (122 °F), 5.8. (cc) At 17 °C (63 °F). (dd) From 21 to 104 °C (70 to 219 °F). (ee) For α at 25 °C (77 °F). (ff) From 20 to 500 °C (68 to 930 °F). (gg) Calculated. (hh) From 25 to 1000 °C (77 to 1830 °F), for iodide thorium. (ii) At 100 °C (212 °F). (jj) From 0 to 100 °C (32 to 212 °F), for polycrystalline metal. (kk) Btu · ft/h · ft2 · °F at −400 °F. (ll) Rolled rods. (mm) At 70 °C (158 °F). (nn) At 27 °C (80 °F). (oo) From 23 to 100 °C (73 to 212 °F). (pp) From 20 to 250 °C (68 to 480 °F), for polycrystalline metal. (qq) α, polyerystal-line. (rr) W/cm/°C at 27 °C (80.6 °F)

Table 6 (continued)

Symbol	Coefficient of linear thermal expansion(a), 10^{-6}/°C (10^{-6}/°F)	Thermal conductivity(a), cal/(cm · s · °C)	Specific heat(b), cal/g · °C (J/kg · K)	Heat of fusion, cal/g (Btu/lb)
	(3.19)		(310)	(...)
Au	14.2 (7.9)	0.71	0.0312(n) (131)(n)	16.1 (29.0)
Hf	519(o) (288)(o)	0.223(p)	0.0351 (147)	... (...)
He	... (...)	3.32×10^{-4}	1.25 (5230)	... (...)
Ho	... (...)	...	0.039 (160)	24.9 (44.7)
H	... (...)	4.06×10^{-4}	3.45 (14.400)	15.0 (27.0)
In	33 (18)	0.057	0.057 (240)	6.8 (12.2)
I	93 (52)	10.4×10^{-4}	0.052 (220)	14.2 (25.6)
Ir	6.8 (3.8)	0.14	0.0307 (129)	... (...)
Fe	11.76(q) (6.53)(q)	0.18(r)	0.11 (460)	65.5 (117.9)
Kr	... (...)	0.21×10^{-4}	... (...)	... (...)
La	5 (2.77)	0.033(s)	0.048 (200)	17.3 (31.1)
Lw	... (...)	...	... (...)	... (...)
Pb	29.3(t) (16.3)(t)	0.083(u)	0.0309(u) (129)(u)	6.26 (11.27)
Li	56 (31)	0.17	0.79 (3300)	104.2 (187.6)
Lu	... (...)	...	0.037 (150)	26.29 (47.32)
Mg	27.1(v) (15.05)(v)	0.367	0.245 (1030)	88 ± 2 (158 ± 4)
Mn	22(w) (12.22)(w)	...	0.115(x) (481)(x)	63.7 (114.7)
Mv	... (...)	...	... (...)	... (...)
Hg	... (...)	0.0196(u)	0.033 (140)	2.8 (5.0)
Mo	4.9(c) (2.7)(c)	0.34	0.066 (280)	69.8(y) (125.6)(y)
Nd	6 (3.33)	0.031(z)	0.045 (190)	11.78 (21.20)
Ne	... (...)	0.00011	... (...)	... (...)
Np	... (...)	...	... (...)	... (...)
Ni	13.3(h) (7.39)(h)	0.22(aa)	0.105 (440)	73.8 (132.8)
Nb	7.31 (4.06)	0.125(u)	0.065(u) (270)(u)	69 (124.2)
N	... (...)	0.000060	0.247 (1030)	6.2 (11.2)
No	... (...)	...	... (...)	... (...)
Os	4.6(bb) (2.6)(bb)	...	0.031 (130)	... (...)
O	... (...)	0.000059	0.218 (913)	3.3 (5.9)
Pd	11.76 (6.53)	1.68(n)	0.0584(u) (245)(u)	34.2 (61.6)
P	125 (70)	...	0.177 (741)	5.0 (9.0)
Pt	8.9 (4.9)	0.165(cc)	0.0314(u) (131)(u)	26.9 (48.4)
Pu	55(dd) (30.55)(dd)	0.020(aa)	0.033(ee) (140)(ee)	... (...)

(continued)

(a) Near 20 °C (68 °F). (b) At 20 °C (68 °F). (c) From 20 to 100 °C (68 to 212 °F). (d) From 20 to 60 °C (68 to 140 °F). (e) From 25 to 100 °C (77 to 212 °F) (f) From 20 to 750 °C (68 to 1380 °F). (g) For α at 0 to 400 °C (32 to 750 °F). (h) From 0 to 100 °C (32 to 212 °F). (i) At 28 °C (82 °F). (j) From 0 to 26 °C (32 to 70 °F). (k) Near 40 °C (105 °F); the coefficient of expansion of gadolinium changes rapidly between −100 and +100 °C (−150 and +212 °F). (l) From 0 to 30 °C (32 to 86 °F). (m) At melting point. (n) At 18 °C (64 °F). (o) From 20 to 200 °C (68 to 390 °F). (p) W/cm/°C at 50 °C (120 °F). (q) At 25 °C (77 °F) for high-purity k iron. (r) For ingot iron at 0 °C (32 °F). (s) Annealed. (t) From 17 to 100 °C (63 to 212 °F). (u) At 0 °C (32 °F). (v) Along a-axis; 24.3 along c-axis. (w) α; γ, 14; both from 0 to 100 °C (32 to 212 °F). (x) α; γ, 0.120; both at 25.2 °C (77.3 °F). (y) Estimated (z) At −2.22 °C (28 °F). (aa) At 25 °C (77 °F). (bb) At 50 °C (122 °F), parallel to a-axis, mean value; parallel to c-axis at 50 °C (122 °F), 5.8. (cc) At 17 °C (63 °F). (dd) From 21 to 104 °C (70 to 219 °F). (ee) For α at 25 °C (77 °F). (ff) From 20 to 500 °C (68 to 930 °F). (gg) Calculated. (hh) From 25 to 1000 °C (77 to 1830 °F), for iodide thorium. (ii) At 100 °C (212 °F). (jj) From 0 to 100 °C (32 to 212 °F), for polycrystalline metal. (kk) Btu · ft/h · ft2 · °F at −400 °F. (ll) Rolled rods. (mm) At 70 °C (158 °F). (nn) At 27 °C (80 °F). (oo) From 23 to 100 °C (73 to 212 °F). (pp) From 20 to 250 °C (68 to 480 °F), for polycrystalline metal. (qq) α, polyerystal-line. (rr) W/cm/°C at 27 °C (80.6 °F)

Table 6 (continued)

Symbol	Coefficient of linear thermal expansion(a), 10^{-6}/°C (10^{-6}/°F)	Thermal conductivity(a), cal/(cm · s · °C)	Specific heat(b) cal/g · °C (J/kg · K)	Heat of fusion, cal/g (Btu/lb)
Po	. . .	. . .	. . .	. . .
	(. . .)		(. . .)	(. . .)
K	83	0.24	0.177	14.6
	(46)		(741)	(26.3)
Pr	4	0.28(z)	0.045	11.71
	(2.22)		(188)	(21.08)
Pm	. . .	. . .	. . .	. . .
	(. . .)		(. . .)	(. . .)
Pa	. . .	. . .	. . .	. . .
	(. . .)		(. . .)	(. . .)
Ra	. . .	. . .	. . .	. . .
	(. . .)		(. . .)	(. . .)
Rn	. . .	. . .	. . .	. . .
	(. . .)		(. . .)	(. . .)
Re	6.7(ff)	C.17	0.033	. . .
	(3.7)(ff)		(140)	(. . .)
Rh	8.3	0.21(cc)	0.059(u)	. . .
	(4.6)		(250)(u)	(. . .)
Rb	90	. . .	0.080	6.5
	(50)		(330)	(11.79)
Ru	9.1	. . .	0.057(u)	. . .
	(5.1)		240(u)	(. . .)
Sm	. . .	. . .	0.042(gg)	17.29
	(. . .)		(180)(gg)	(31.12)
Sc	. . .	. . .	0.134	84.52
	(. . .)		(561)	(152.14)
Se	37	7 to 18.3×10^{-4}	0.084(s)	16.4
	(21)		(350)(s)	(29.5)
Si	2.8 to 7.3	0.20	0.162(u)	432
	(1.6 to 4.1)		(678)(u)	(778)
Ag	19.68(h)	1.0(u)	0.0559(u)	25
	(10.9)(h)		(234)(u)	(45)
Na	71	0.32	0.295	27.5
	(39)		(1240)	(49.5)
Sr	. . .	. . .	0.176	25
	(. . .)		(737)	(45)
S	64	6.31×10^{-4}	0.175	9.3
	(36)		(733)	(16.7)
Ta	6.5	0.130	0.034(aa)	38
	(3.6)		(140)	(68)
Tc	. . .	. . .	. . .	. . .
	(. . .)		(. . .)	(. . .)
Te	16.75	0.014	0.047	32
	(9.3)		(200)	(58)
Tb	7	. . .	0.044	24.54
	(3.88)		(180)	(44.17)
Tl	28	0.093	0.031	5.04
	(16)		(130)	(9.07)
Th	12.5(hh)	0.090(ii)	0.034	<19.82
	(6.9)(hh)		(140)	(<35.68)
Tm	. . .	. . .	0.038	26.04
	(. . .)		(160)	(46.87)
Sn	23(jj)	1.50(d)	0.054	14.5
	(13)(jj)		(230)	(26.1)
Ti	8.41	6.6(kk)	0.124	104(y)
	(4.67)		(519)	(188)(y)
W	4.6	0.397(d)	0.033	44
	(2.55)		(140)	(70)
U	6.8 to 14.1(ll)	0.07(mm)	0.02709(nn)	. . .
	(3.8 to 7.8)(ll)		(113.4)(nn)	(. . .)
V	8.3(oo)	0.074(ii)	0.119(g)	. . .
	(4.6)(oo)		(498)(g)	(. . .)
Xe	. . .	1.24×10^{-1}	. . .	. . .
	(. . .)		(. . .)	(. . .)
Yb	25	. . .	0.035	12.71
	(13.9)		(150)	(22.88)
Y	. . .	0.035(z)	0.071	46

(continued)

(a) Near 20 °C (68 °F). (b) At 20 °C (68 °F). (c) From 20 to 100 °C (68 to 212 °F). (d) From 20 to 60 °C (68 to 140 °F). (e) From 25 to 100 °C (77 to 212 °F). (f) From 20 to 750 °C (68 to 1380 °F). (g) For α at 0 to 400 °C (32 to 750 °F). (h) From 0 to 100 °C (32 to 212 °F). (i) At 28 °C (82 °F). (j) From 0 to 26 °C (32 to 70 °F). (k) Near 40 °C (105 °F); the coefficient of expansion of gadolinium changes rapidly between −100 and +100 °C (−150 and +212 °F). (l) From 0 to 30 °C (32 to 86 °F). (m) At melting point. (n) At 18 °C (64 °F). (o) From 20 to 200 °C (68 to 390 °F). (p) W/cm/°C at 50 °C (120 °F). (q) At 25 °C (77 °F) for high-purity k iron. (r) For ingot iron at 0 °C (32 °F). (s) Annealed. (t) From 17 to 100 °C (63 to 212 °F). (u) At 0 °C (32 °F). (v) Along a-axis; 24.3 along c-axis. (w) γ, 14; both from 0 to 100 °C (32 to 212 °F). (x) α; γ, 0.120; both at 25.2 °C (77.3 °F). (y) Estimated (z) At −2.22 °C (28 °F). (aa) At 25 °C (77 °F). (bb) At 50 °C (122 °F), parallel to a-axis, mean value; parallel to c-axis at 50 °C (122 °F), 5.8. (cc) At 17 °C (63 °F). (dd) From 21 to 104 °C (70 to 219 °F). (ee) For α at 25 °C (77 °F). (ff) From 20 to 500 °C (68 to 930 °F). (gg) Calculated. (hh) From 25 to 1000 °C (77 to 1830 °F), for iodide thorium. (ii) At 100 °C (212 °F). (jj) From 0 to 100 °C (32 to 212 °F), for polycrystalline metal. (kk) Btu · ft/h · ft2 · °F at −400 °F. (ll) Rolled rods. (mm) At 70 °C (158 °F). (nn) At 27 °C (80 °F). (oo) From 23 to 100 °C (73 to 212 °F). (pp) From 20 to 250 °C (68 to 480 °F), for polycrystalline metal. (qq) α, polyerystal-line. (rr) W/cm/°C at 27 °C (80.6 °F)

Table 6 (continued)

Symbol	Coefficient of linear thermal expansion(a), 10^{-6}/°C (10^{-6}/°F)	Thermal conductivity(a), cal/(cm · s · °C)	Specific heat(b) cal/g · °C (J/kg · K)	Heat of fusion, cal/g (Btu/lb)
	(. . .)		(300)	(83)
Zn	39.7(pp)	0.27(aa)	0.0915	24.09
	(22.0)(pp)		(383)	(43.36)
Zr	5.85(qq)	0.211(rr)	0.067 + 0.001	60(y)
	(3.2)(qq)		(280 + 4)	(110)(y)

(a) Near 20 °C (68 °F). (b) At 20 °C (68 °F). (c) From 20 to 100 °C (68 to 212 °F). (d) From 20 to 60 °C (68 to 140 °F). (e) From 25 to 100 °C (77 to 212 °F) (f) From 20 to 750 °C (68 to 1380 °F). (g) For α at 0 to 400 °C (32 to 750 °F). (h) From 0 to 100 °C (32 to 212 °F). (i) At 28 °C (82 °F). (j) From 0 to 26 °C (32 to 70 °F). (k) Near 40 °C (105 °F); the coefficient of expansion of gadolinium changes rapidly between −100 and +100 °C (−150 and +212 °F). (l) From 0 to 30 °C (32 to 86 °F). (m) At melting point. (n) At 18 °C (64 °F). (o) From 20 to 200 °C (68 to 390 °F). (p) W/cm° at 50 °C (120 °F). (q) At 25 °C (77 °F) for high-purity k iron. (r) For ingot iron at 0 °C (32 °F). (s) Annealed. (t) From 17 to 100 °C (63 to 212 °F). (u) At 0 °C (32 °F). (v) Along a-axis; 24.3 along c-axis. (w) α; γ, 14; both from 0 to 100 °C (32 to 212 °F). (x) α; γ, 0.120; both at 25.2 °C (77.3 °F). (y) Estimated (z) At −2.22 °C (28 °F). (aa) At 25 °C (77 °F). (bb) At 50 °C (122 °F), parallel to *a*-axis, mean value; parallel to *c*-axis at 50 °C (122 °F), 5.8. (cc) At 17 °C (63 °F). (dd) From 21 to 104 °C (70 to 219 °F). (ee) For α at 25 °C (77 °F). (ff) From 20 to 500 °C (68 to 930 °F). (gg) Calculated. (hh) From 25 to 1000 °C (77 to 1830 °F), for iodide thorium. (ii) At 100 °C (212 °F). (jj) From 0 to 100 °C (32 to 212 °F), for polycrystalline metal. (kk) Btu · ft/h · ft^2 · °F at −400 °F. (ll) Rolled rods. (mm) At 70 °C (158 °F). (nn) At 27 °C (80 °F). (oo) From 23 to 100 °C (73 to 212 °F). (pp) From 20 to 250 °C (68 to 480 °F), for polycrystalline metal. (qq) α, polyerystalline. (rr) W/cm/°C at 27 °C (80.6 °F)

Table 7 Temperature-dependent allotropic structures of the elements

Allotrope	Pearson symbol	a	b	c	c/a, or α or β	Stability range(a)	Allotrope	Pearson symbol	a	b	c	c/a, or α or β	Stability range(a)
αAm	hP4	0.3468	...	1.1241	2 × 1.621	RT	βNa	cl2	0.42096	...	...	...	RT
βAm	cF4	0.4894	...	...	...	>605 °C	αNd	hP4	0.36582	...	1.17966	2×1.6124	RT
αAr	cF4	0.5312	...	...	...	<83.8 K	βNd	cl2	0.413	...	...	...	>862 °C
βAr	hP2	0.3760	...	0.6141	1.633	<83.8 K	αNp	oP8	0.6683	0.4723	0.4887	...	RT
αAs	hR2	0.41320	...	...	α = 54.12 °	RT	βNp	tP4	0.4896	...	0.3387	0.692	>280 °C
εAs	oC8	0.362	1.085	0.448	...	>448 °C	γNp	cl2	0.352	...	...	...	>577 °C
αBe	hP2	0.22857	...	0.35839	1.5680	RT	αO	mC4	0.5403	0.3429	0.5086	β = 132.53 °	4.2 K
βBe	cl2	0.25515	...	...	...	>1250 °C	βO	hR2	0.4210	...	...	α = 46.27 °	>23.9 K
αBk	hP4	0.3416	...	1.1069	2 × 1.620	RT	γO	cP16	0.683	...	...	...	>43.6 K
βBk	cF4	0.4997	...	...	...	~RT	αPa	tl2	0.3945	...	0.3242	0.822	RT
αCa	cF4	0.55884	...	...	...	RT	βPa	cl2	0.381	...	...	...	>1170 °C
γCa	cl2	0.4480	...	...	...	>737 °C	αPo	cP1	0.3366	...	...	...	RT
αCe	cF4	0.51610	...	...	...	RT	βPo	hR1	0.3373	...	...	α = 98.08 °	>54 °C
βCe	hP4	0.3673	...	1.1802	2×1.607	<263 K	αPr	hP4	0.36721	...	1.18326	2×1.6111	RT
γCe	cF4	0.485	...	...	...	<95 K	βPr	cl2	0.413	...	...	...	>821 °C
αCm	hP4	0.3496	...	1.1331	2×1.621	RT	αPu	mP16	0.6183	0.4822	1.0968	α = 101.78 °	RT
βCm	cF4	0.4382	...	...	...	−RT	βPu	ml34	0.9284	1.0463	0.7859	α = 92.13 °	>122 °C
αCo	hP2	0.25071	...	0.40694	1.6232	RT	γPu	oF8	0.31587	0.57682	1.0162	...	>235 °C
βCo	cF4	0.35445	...	...	...	>388 °C	δPu	cF4	0.46371	...	0.3279	...	>319 °C
αDy	hP2	0.35915	...	0.56501	1.5732	RT	δ'Pu	tl2	0.33261	...	0.44630	1.3418	>450 °C
βDy	cl2	?	...	...	...	>970 °C	εPu	cl2	0.5703	...	...	...	>471 °C
γDy	oF4	0.3595	0.6184	0.5678	...	<86 K	αSc	hP2	0.33088	...	0.52680	1.5921	RT
αEr	hP2	0.35592	...	0.55850	1.5692	RT	βSc	cl2	?	...	...	...	>1334 °C
βEr	cl2	?	...	...	...	HT	αSe	mP32	0.9054	0.9083	0.2336	β = 90.82 °	RT
αF	mC8	0.550	0.338	0.728	β = 102.17 °	4.2 K	βSe	mP64	1.5018	1.4713	0.8879	β = 93.6 °	RT
βF	cP16	0.667	...	...	...	>45.6 K	γSe	hP3	0.43655	...	0.49576	1.1356	RT
αFe	cl2	0.28665	...	...	...	RT	αSm	hR3	0.36290	...	2.6207	4.5×1.6084	RT
γFe	cF4	0.36467	...	...	...	>910 °C	βSm	cl2	?	...	...	...	>917 °C
δFe	cl2	0.29135	...	...	...	>1390 °C	αSn	cF8	0.64892	...	...	...	<18 °C
αGd	hP2	0.36336	...	0.57810	1.5910	RT	βSn	tl4	0.58316	...	0.31815	0.5456	RT
βGd	cl2	0.406	...	...	...	>1262 °C	αSr	cF4	0.6084	...	...	...	RT
αH	hP2	0.3771	...	0.6152	1.631	4.2 K	βSr	hP2	0.428	...	0.705	1.647	>213 °C
βH	cF4	0.5334	...	...	...	<1.3 K	γSr	cl2	0.487	...	...	...	>605 °C
αHf	hP2	0.31946	...	0.50511	1.5811	RT	αTb	hP2	0.36055	...	0.56966	1.5800	RT
βHf	cl2	0.3610	...	...	...	>1995 °C	βTb	cl2	?	...	...	...	>1316 °C
αHo	hP2	0.35778	...	0.56178	1.5702	RT	αTh	cF4	0.50851	...	...	...	RT
βHo	cl2	?	...	...	...	HT	βTh	cl2	0.411	...	...	...	>1400°C
αLa	hP4	0.37740	...	1.2171	2×1.6125	RT	αTi	hP2	0.29503	...	0.46836	1.5875	RT
βLa	cF4	0.53045	...	...	...	>340°C	βTi	cl2	0.33065	...	...	...	>900 °C
γLa	cl2	0.4265	...	...	...	>868 °C	αT1	hP2	0.34563	...	0.55263	1.5989	RT
αLi	hP2	0.3111	...	0.5093	1.637	<72 K	βT1	cl2	0.3879	...	...	...	>230 °C
βLi	cl2	0.35093	...	...	...	RT	αTm	hP2	0.3575	...	0.55540	1.5700	RT
αLu	hP2	0.35052	...	0.55494	1.5832	RT	βTm	cl2	?	...	...	...	RT
βLu	cl2	?	...	...	...	HT	αU	oC4	0.28538	0.58680	0.49557	...	RT
αMn	cl58	0.89219	...	...	...	RT	βU	tP30	1.0759	...	0.5654	0.526	>662 °C
βMn	cP20	0.63152	...	...	...	>727 °C	γU	cl2	0.3524	...	...	...	>772 °C
γMn	cF4	0.38624	...	...	...	>1095 °C	αYb	cF4	0.54848	...	...	...	RT
δMn	cl2	0.30806	...	...	...	>1135 °C	βYb	cl2	0.444	...	...	...	>732 °C
αN	cP8	0.5659	...	...	...	4.2 K	γYb	hP2	0.38799	...	0.63859	1.6459	<270 K
βN	hP4	0.4046	...	0.6629	1.638	>35.6 K	αZr	hP2	0.32317	...	0.51476	1.5928	RT
αNa	hP2	0.3767	...	0.6154	1.634	<36 K	βZr	cl2	0.3609	...	...	...	>865 °C

The accuracy of the data in this table is considered to be reliable to +2 in the last reported digit. (a) RT, room temperature; HT, high temperature

Table 8 Pressure-dependent allotropic structures of the elements

Allotrope	Pearson symbol	a	b	c	c/a, or α or β	Stability range(a)
Al-I	cF4	0.40496	...	...	...	RTP
Al-II	hP2	0.2693	...	0.4398	1.633	>205 kB
αAm	hP4	0.3468	...	1.1241	2×1.621	RTP
γAm	oC4	0.3063	0.5968	0.5169	...	>150 kB
αBa	cI2	0.5023	...	...	...	RTP
βBa	hP2	0.3901	...	0.6154	1.578	>53.3 kB
γBa	?	...	...	...	...	>230 kB
αBi	hR2	0.4760	...	...	α = 57.23°	RTP
βBi	?	...	...	...	...	>28 kB
γBi	mP3	0.605	0.420	0.465	β = 85.33°	>30 kB
δBi	?	...	...	...	...	>43 kB
εBi	?	...	...	...	...	>65 kB
ζBi	cI2	3.800	...	...	...	>90 kB
C (graph)	hP4	0.24612	...	0.67090	2.7259	RTP
C (dia)	cF8	0.35669	...	...	...	600 kB
αCe	cF4	0.51610	...	...	...	RTP
α′Ce	cF4	0.482	...	...	...	>15 kB
Ce-III	mI2	0.4762	0.3170	0.3169	β = 91.7°	>51 kB
αCr	cI2	0.28847	...	...	...	RTP
α′Cr	tI2	0.2882	...	0.2887	1.002	HP
Cs-I	cI2	0.6141	...	...	...	RTP
Cs-II	cF4	0.5984	...	...	...	>23.7 kB
Cs-III	cF4	0.5800	...	...	...	>42.2 kB
αDy	hP2	0.35915	...	0.56501	1.5731	RTP
γDy	hR3	0.3436	...	2.4830	4.5×1.606	>75 kB
αFe	cI2	0.28665	...	...	...	RTP
εFe	hP2	0.2485	...	0.3990	1.606	>130 kB
αGa	oC8	0.45192	0.76586	0.45258	...	RTP
βGa	tI2	0.2808	...	0.4458	1.587	>12 kB
γGa	oC40	1.0593	1.3523	0.5203	...	>30 kB; 220 K
αGd	hP2	0.36336	...	0.57810	1.5910	RTP
γGd	hR3	0.361	...	2.603	4.5×1.60	>30 kB
αGe	cF8	0.56574	...	...	...	RTP
βGe	tI4	0.4884	...	0.2692	0.551	>120 kB
γGe	tP12	0.593	...	0.698	0.18	Decompressed βGe
δGe	cI16	0.692	...	...	...	>120 kB
αHe	hP2	0.3577	...	0.5842	1.633	4.2 K
βHe	cF4	4.240	...	...	...	1.25 kB; 1.6 K
γHe	cI2	4.110	...	...	...	0.3 kB; 1.73 K
αHg	hR1	0.3005	...	...	α = 70.53°	227 K
βHg	tI2	0.3995	...	0.2825	0.707	HP; 77 K
αHo	hP2	0.35778	...	0.56178	1.5702	RTP
γHo	hR3	0.334	...	2.45	4.5×1.63	>40 kB
K-I	cI2	0.5321	...	...	...	RTP
K-II	?	...	...	...	...	280 kB; 77 K
K-III	?	...	...	...	...	360 kB; 77 K
αLa	hP4	0.37740	...	1.2171	2×1.6125	RTP
β′La	cF4	0.517	...	...	...	>20 kB
αLi	cI2	0.35093	...	...	...	RTP
γLi	cF4	0.4388	...	...	...	CW at <72 K
αN	cP8	0.5659	...	...	...	4.2 K
N-II	tP4	0.3957	...	5.101	1.289	>33 kB; 20 K
αNd	hP4	0.36582	...	1.17966	2×1.6124	RTP
γNd	cF4	0.480	...	...	...	>50 kB
Pb-I	cF4	0.49502	...	...	...	RTP
Pb-II	hP2	0.3265	...	0.5387	1.653	>103 kB
αPr	hP4	0.36721	...	1.18326	2×1.6111	RTP
γPr	cF4	0.488	...	...	...	>40 kB
αRb	cI2	0.5703	...	...	...	RTP
βRb	?	...	...	...	...	>10.8 kB
γRb	?	...	...	...	...	>20.5 kB
Sb-I	hR2	0.45065	...	...	α = 57.11°	RTP
Sb-II	cP1	0.2992	...	...	...	>50 kB
Sb-III	hP2	0.3376	...	0.5341	1.582	>75 kB
Sb-IV	mP4	0.556	0.404	0.422	β = 86.0°	>140 kB
αSi	cF8	0.54306	...	...	...	RTP
βSi	tI4	0.4686	...	0.2585	0.551	>95 kB
γSi	cI16	0.636	...	...	...	>160 kB
δSi	hP4	0.380	...	0.628	1.635	Decompressed βSi
αSm	hR3	0.36290	...	2.607	4.5×1.6048	RTP
γSm	hP4	0.3618	...	1.166	2×1.611	>40 kB
βSn	tI4	0.58316	...	0.31815	0.5456	RTP
γSn	tI2	0.370	...	0.337	0.911	>90 kB
αSr	cF4	0.6084	...	...	...	RTP
Sr-II	cI2	0.4437	...	...	...	>35 kB
αTb	cP2	0.36055	...	0.56966	1.5800	RTP
Tb-II	hR3	0.341	...	2.45	4.5×1.60	>60 kB
αTe	hP3	0.44561	...	0.59271	1.3301	RTP
βTe	hR2	0.469	...	...	α = 53.30°	>30 kB
γTe	hR1	0.3002	...	...	α = 103.3°	>70 kB
αTi	hP2	0.34563	...	0.55263	1.5989	RTP
ωTi	hP2	0.4625	...	0.2813	0.608	Decompressed
αTl	hP2	0.34563	...	0.55540	1.5700	RTP
γTl	cF4	...	...	...	...	HP
αZr	hP2	0.32217	...	0.51476	1.5928	RTP
ωZr	hP2	0.506	...	0.3109	0.617	Decompressed

The accuracy of the data in this table is considered to be reliable to ±2 in the last reported digit. (a) RTP, room temperature and pressure; HP, high pressure. High-pressure data refer to pressures within ±1 of the last reported digit in kilobars.

Table 9   Magnetic phase transition temperatures of the elements

Chemical symbol	Atomic number	Allotrope (a)	Phase transition temperature (T_c), K	Type of magnetic ordering (b)	Phase transition temperature (T_{c2}), K	Type of magnetic ordering(b)	Phase transition temperature (T_{c3}), K	Type of magnetic ordering(b)	Saturation magnetic moment, μ_B
Ce(c)	58	β- dcph	13.7	AC?	12.5	AC?	...	...	2.61
		γ- fcc	14.4	AC?	...	...	...	...	~2.5
Cm	96	...	52	AC	...	...	...	...	...
Co	27	fcc	1388	FM	...	...	...	...	1.715
Cr	24	...	312.7	AI	...	...	...	...	0.45
Dy	66	α-cph	179.0	AI	89.0	FM	...	...	10.33
Er	68	...	85.0	AI	53	AC	20.0	CF	9.1
Eu	63	...	90.4	AC	...	...	...	...	5.9
Fe	26	α-bcc	1044	FM	...	...	...	...	2.216
		γ- fcc	67	AC	...	...	...	...	0.75
Gd	64	α- cph	293.4	FM	...	...	...	...	7.63
Ho	67	...	132.0	AI	20.0	CF	...	...	10.34
Mn	25	α- bcc	100	AC	...	...	...	...	(d)
Nd	60	α- dcph	19.9	AI	7.5	AC	...	...	1.84
Ni	28	...	627.4	FM	...	...	...	...	0.616
Pm	61	α- dcph	98	FM?	...	...	...	...	0.24
Pr	59	α- dcph	0.06	AC	...	...	...	...	0.36
Sm	62	α-rhomb.	106	h, A(e)	13.8	c, A(e)	...	...	0.1
Tb	65	α-cph	230.0	AI	219.5	FM	...	...	9.34
Tm	69	...	58.0	AI	40 to 32	FI	...	...	7.14

(a) dcph, double close-packed hexagonal; fcc, face-centered cubic; cph close-packed hexagonal; bcc, body-centered cubic. (b) FM, transition from paramagnetic to ferromagnetic state; AC, transition to periodic (antiferromagnetic) state that is commensurate with the lattice periodicity (e.g., spins on three atom layers directed up followed by three layers down, etc.); AI, transition to periodic (antiferromagnetic) state that is generally not commensurate with the lattice periodicity (e.g., helical spin ordering); CF, transition to conical ferromagnetic state (combination of planar helical antiferromagnet plus ferromagnetic component); FI, transition to ferromagnetic periodic structure (unequal number of up and down spin layers). (c) Ce exists in five crystal structures, two of which are magnetic (γ, fcc; and β, dcph). γCe is estimated to be antiferromagnetic below 14.4 K by extrapolation from fcc Ce-La alloys. (αCe does not exist in pure form below ~100 K.) βCe is thought to exhibit antiferromagnetism on the hexagonal lattice sites below 13.7 K and on the cubic sites below 12.5 K. (d) The magnetic moment assignments of Mn are complex: see R.S. Tebble and D.J. Craik *Magnetic Materials*, Wiley Interscience, London, 1969, p 60–62. (e) h, A;c, A indicate that sites of hexagonal and cubic point symmetry order antiferromagnetically but at different temperatures

Density of Metals and Alloys

Table 1   Density of metals and alloys

Metal or alloy	Density g/cm³	lb/in.³	Metal or alloy	Density g/cm³	lb/in.³	Metal or alloy	Density g/cm³	lb/in.³
Aluminum and aluminum alloys			**Wrought alloys**			76Cu-2.5Sn-6.5Pb-15Zn	8.77	0.317
Aluminum (99.996%)	2.6989	0.0975	Gilding, 95%	8.86	0.320	72Cu-1Sn-3Pb-24Zn	8.50	0.307
Wrought alloys			Commercial bronze, 90%	8.80	0.318	67Cu-1Sn-3Pb-29Zn	8.45	0.305
EC, 1060 alloys	2.70	0.098	Jewelry bronze, 87.5%	8.78	0.317	61Cu-1Sn-1Pb-37Zn	8.40	0.304
1100	2.71	0.098	Red brass, 85%	8.75	0.316	Manganese bronze		
2011	2.82	0.102	Low brass, 80%	8.67	0.313	60 ksi	8.2	0.30
2014	2.80	0.101	Cartridge brass, 70%	8.53	0.308	65 ksi	8.3	0.30
2024	2.77	0.100	Yellow brass	8.47	0.306	90 ksi	7.9	0.29
2218	2.81	0.101	Muntz metal	8.39	0.303	110 ksi	7.7	0.28
3003	2.73	0.099	Leaded commercial bronze	8.83	0.319	Aluminum bronze		
4032	2.69	0.097	Low-leaded brass (tube)	8.50	0.307	Alloy 9A	7.8	0.28
5005	2.70	0.098	Medium-leaded brass	8.47	0.306	Alloy 9B	7.55	0.272
5050	2.69	0.097	High-leaded brass (tube)	8.53	0.308	Alloy 9C	7.5	0.27
5052	2.68	0.097	High-leaded brass	8.50	0.307	Alloy 9D	7.7	0.28
5056	2.64	0.095	Extra-high-leaded brass	8.50	0.307	Nickel silver		
5083	2.66	0.096	Free-cutting brass	8.50	0.307	12% Ni	8.95	0.323
5086	2.65	0.096	Leaded Muntz metal	8.41	0.304	16% Ni	8.95	0.323
5154	2.66	0.096	Forging brass	8.44	0.305	20% Ni	8.85	0.319
5357	2.70	0.098	Architectural bronze	8.47	0.306	25% Ni	8.8	0.32
5456	2.66	0.096	Inhibited admiralty	8.53	0.308	Silicon bronze	8.30	0.300
6061, 6063	2.70	0.098	Naval brass	8.41	0.304	Silicon brass	8.30	0.300
6101, 6151	2.70	0.098	Leaded naval brass	8.44	0.305	**Iron and iron alloys**		
7075	2.80	0.101	Manganese bronze (A)	8.36	0.302	Pure iron	7.874	0.2845
7079	2.74	0.099	Phosphor bronze, 5% (A)	8.86	0.320	Ingot iron	7.866	0.2842
7178	2.82	0.102	Phosphor bronze, 8% (C)	8.80	0.318	Wrought iron	7.7	0.28
			Phosphor bronze, 10% (D)	8.78	0.317	Gray cast iron	7.15(a)	0.258(a)
Casting alloys			Phosphor bronze, 1.25%	8.89	0.321	Malleable iron	7.27(b)	0.262(b)
A13	2.66	0.096	Free-cutting phosphor bronze	8.89	0.321	0.06% C steel	7.871	0.2844
43	2.69	0.097	Cupro-nickel, 30%	8.94	0.323	0.23% C steel	7.859	0.2839
108, A108	2.79	0.101	Cupro-nickel, 10%	8.94	0.323	0.435% C steel	7.844	0.2834
A132	2.72	0.098	Nickel silver, 65-18	8.73	0.315	1.22% C steel	7.830	0.2829
D132	2.76	0.100	Nickel silver, 55-18	8.70	0.314			
F132	2.74	0.099	High-silicon bronze (A)	8.53	0.308	**Low-carbon chromium-molybdenum steels**		
138	2.95	0.107	Low-silicon bronze (B)	8.75	0.316	0.5% Mo steel	7.86	0.283
142	2.81	0.101	Aluminum bronze, 5% Al	8.17	0.294	1Cr-0.5Mo steel	7.86	0.283
195, B195	2.81	0.101	Aluminum bronze, (3)	7.78	0.281	1.25Cr-0.5Mo steel	7.86	0.283
214	2.65	0.096	Aluminum-silicon bronze	7.69	0.278	2.25Cr-1.0Mo steel	7.86	0.283
220	2.57	0.093	Aluminum bronze, (1)	7.58	0.274	5Cr-0.5Mo steel	7.78	0.278
319	2.79	0.101	Aluminum bronze, (2)	7.58	0.274	7Cr-0.5Mo steel	7.78	0.278
355	2.71	0.098	Beryllium copper	8.23	0.297	9Cr-1Mo steel	7.67	0.276
356	2.68	0.097						
360	2.64	0.095	**Casting alloys**			**Medium-carbon alloy steels**		
380	2.71	0.098	Chromium copper (1% Cr)	8.7	0.31	1Cr-0.35Mo-0.25 V steel	7.86	0.283
750	2.88	0.104	88Cu-10Sn-2Zn	8.7	0.31	H11 die steel (5Cr-1.5Mo-0.4V)	7.79	0.281
40E	2.81	0.101	88Cu-8Sn-4Zn	8.8	0.32			
			89Cu-11Sn	8.78	0.317	**Other iron-base alloys**		
			88Cu-6Sn-1.5Pb-4.5Zn	8.7	0.31	A-286	7.94	0.286
Copper and copper alloys			87Cu-8Sn-1Pb-4Zn	8.8	0.32	16-25-6 alloy	8.08	0.292
Wrought coppers			87Cu-10Sn-1Pb-2Zn	8.8	0.32	RA-330	8.03	0.290
Pure copper	8.96	0.324	80Cu-10Sn-10Pb	8.95	0.323	Incoloy	8.02	0.290
Electrolytic tough			83Cu-7Sn-7Pb-3Zn	8.93	0.322	Incoloy T	7.98	0.288
pitch copper (ETP)	8.89	0.321	85Cu-5Sn-9Pb-1Zn	8.87	0.320	Incoloy 901	8.23	0.297
Deoxidized copper, high residual			78Cu-7Sn-15Pb	9.25	0.334	T1 tool steel	8.67	0.313
phosphorus (DHP)	8.94	0.323	70Cu-5Sn-25Pb	9.30	0.336	M2 tool steel	8.16	0.295
Free-machining copper			85Cu-5Sn-5Pb-5Zn	8.80	0.318	H41 tool steel	7.88	0.285
0.5% Te	8.94	0.323	83Cu-4Sn-6Pb-7Zn	8.6	0.31	20W-4Cr-2V-12Co steel	8.89	0.321
1.0% Pb	8.94	0.323	81Cu-3Sn-7Pb-9Zn	8.7	0.31	Invar (36% Ni)	8.00	0.289
						Hipernik (50% Ni)	8.25	0.298

(continued)

(a) 6.95 to 7.35 g/cm³ (0.251 to 0.265 lb/in.³). (b) 7.20 to 7.34 g/cm³ (0.260 to 0.265 lb/in.³). (c) Face-centered cubic. (d) Hexagonal. (e) Body-centered cubic. (f) Close-packed hexagonal. (g) Rhombohedral. Source: M. Bauccio, *ASM Metals Reference Book*, 3rd ed., ASM International, 1993

Table 1   (continued)

Metal or alloy	Density g/cm³	lb/in.³
4% Si	7.6	0.27
10.27% Si	6.97	0.252
Stainless steels and heat-resistant alloys		
Corrosion-resistant steel castings		
CA-15	7.612	0.2750
CA-40	7.612	0.2750
CB-30	7.53	0.272
CC-50	7.53	0.272
CE-30	7.67	0.277
CF-8	7.75	0.280
CF-20	7.75	0.280
CF-8M, CF-12M	7.75	0.280
CF-8C	7.75	0.280
CF-16F	7.75	0.280
CH-20	7.72	0.279
CK-20	7.75	0.280
CN-7M	8.00	0.289
Heat-resistant alloy castings		
HA	7.72	0.279
HC	7.53	0.272
HD	7.58	0.274
HE	7.67	0.277
HF	7.75	0.280
HH	7.72	0.279
HI	7.72	0.279
HK	7.75	0.280
HL	7.72	0.279
HN	7.83	0.283
HT	7.92	0.286
HU	8.04	0.290
HW	8.14	0.294
HX	8.14	0.294
Wrought stainless and heat-resisting steels		
Type 301	7.9	0.29
Type 302	7.9	0.29
Type 302B	8.0	0.29
Type 303	7.9	0.29
Type 304	7.9	0.29
Type 305	8.0	0.29
Type 308	8.0	0.29
Type 309	7.9	0.29
Type 310	7.9	0.29
Type 314	7.72	0.279
Type 316	8.0	0.29
Type 317	8.0	0.29
Type 321	7.9	0.29
Type 347	8.0	0.29
Type 403	7.7	0.28
Type 405	7.7	0.28
Type 410	7.7	0.28
Type 416	7.7	0.28
Type 420	7.7	0.28
Type 430	7.7	0.28
Type 430F	7.7	0.28
Type 431	7.7	0.28
Types 440A, 440B, 440C	7.7	0.28
Type 446	7.6	0.27
Type 501	7.7	0.28
Type 502	7.8	0.28
19-9DL	7.97	0.29
Precipitation-hardening stainless steels		
PH15-7 Mo	7.804	0.2819
17-4 PH	7.8	0.28
17-7 PH	7.81	0.282
Nickel-base alloys		
D-979	8.27	0.299
Nimonic 80A	8.25	0.298
Nimonic 90	8.27	0.299
M-252	8.27	0.298
Inconel	8.51	0.307
Inconel "X" 550	8.30	0.300
Inconel 700	8.17	0.295
Inconel "713C"	7.913	0.2859
Waspaloy	8.23	0.296

Metal or alloy	Density g/cm³	lb/in.³
René 41	8.27	0.298
Hastelloy alloy B	9.24	0.334
Hastelloy alloy C	8.94	0.323
Hastelloy alloy X	8.23	0.297
Udimet 500	8.07	0.291
GMR-235	8.03	0.290
Cobalt-chromium-nickel-base alloys		
N-155 (HS-95)	8.23	0.296
S-590	8.36	0.301
Cobalt-base alloys		
S-816	8.68	0.314
V-36	8.60	0.311
HS-25	9.13	0.330
HS-36	9.04	0.327
HS-31	8.61	0.311
HS-21	8.30	0.300
Molybdenum-base alloy		
Mo-0.5Ti	10.2	0.368
Lead and lead alloys		
Chemical lead (99.90+% Pb)	11.34	0.4097
Corroding lead (99.73 +% Pb)	11.36	0.4104
Arsenical lead	11.34	0.4097
Calcium lead	11.34	0.4097
5-95 solder	11.0	0.397
20-80 solder	10.2	0.368
50-50 solder	8.89	0.321
Antimonial lead alloys		
1% antimonial lead	11.27	0.407
Hard lead (96Pb-4Sb)	11.04	0.399
Hard lead (94Pb-6Sb)	10.88	0.393
8% antimonial lead	10.74	0.388
9% antimonial lead	10.66	0.385
Lead-base Babbitt alloys		
Lead-base Babbitt		
SAE 13	10.24	0.370
SAE 14	9.73	0.352
Alloy 8	10.04	0.363
Arsenical lead		
Babbitt (SAE 15)	10.1	0.365
"G" Babbitt	10.1	0.365
Magnesium and magnesium alloys		
Magnesium (99.8%)	1.738	0.06279
Casting alloys		
AM100A	1.81	0.065
AZ63A	1.84	0.066
AZ81A	1.80	0.065
AZ91A, B, C	1.81	0.065
AZ92A	1.82	0.066
HK31A	1.79	0.065
HZ32A	1.83	0.066
ZH42, ZH62A	1.86	0.067
ZK51A	1.81	0.065
ZE41A	1.82	0.066
EZ33A	1.83	0.066
EK30A	1.79	0.065
EK41A	1.81	0.065
Wrought alloys		
M1A	1.76	0.064
A3A	1.77	0.064
AZ31B	1.77	0.064
PE	1.76	0.064
AZ61A	1.80	0.065
AZ80A	1.80	0.065
ZK60A, B	1.83	0.066
ZE10A	1.76	0.064
HM21A	1.78	0.064
HM31A	1.81	0.065
Nickel and nickel alloys		
Nickel (99.95% Ni + Co)	8.902	0.322
"A" Nickel	8.885	0.321
"D" Nickel	8.78	0.317
Duranickel	8.26	0.298
Cast nickel	8.34	0.301
Monel	8.84	0.319

Metal or alloy	Density g/cm³	lb/in.³
"K" Monel	8.47	0.306
Monel (cast)	8.63	0.312
"H" Monel (cast)	8.5	0.31
"S" Monel (cast)	8.36	0.302
Inconel	8.51	0.307
Inconel (cast)	8.3	0.30
Ni-o-nel	7.86	0.294
Nickel-molybdenum-chromium-iron alloys		
Hastelloy B	9.24	0.334
Hastelloy C	8.94	0.323
Hastelloy D	7.8	0.282
Hastelloy F	8.17	0.295
Hastelloy N	8.79	0.317
Hastelloy W	9.03	0.326
Hastelloy X	8.23	0.297
Nickel-chromium-molybdenum-copper alloys		
Illium G	8.58	0.310
Illium R	8.58	0.310
Electrical resistance alloys		
80Ni-20Cr	8.4	0.30
60Ni-24Fe-16Cr	8.247	0.298
35Ni-45Fe-20Cr	7.95	0.287
Constantan	8.9	0.32
Tin and tin alloys		
Pure tin	7.3	0.264
Soft solder (30% Pb)	8.32	0.301
Soft solder (37% Pb)	8.42	0.304
Tin Babbitt		
Alloy 1	7.34	0.265
Alloy 2	7.39	0.267
Alloy 3	7.46	0.269
Alloy 4	7.53	0.272
Alloy 5	7.75	0.280
White metal	7.28	0.263
Pewter	7.28	0.263
Titanium and titanium alloys		
99.9% Ti	4.507	0.1628
99.2% Ti	4.507	0.1628
99.0% Ti	4.52	0.163
Ti-6A1-4V	4.43	0.160
Ti-5A1-2.5Sn	4.46	0.161
Ti-2Fe-2Cr-2Mo	4.65	0.168
Ti-8Mn	4.71	0.171
Ti-7Al-4Mo	4.48	0.162
Ti-4Al-4Mn	4.52	0.163
Ti-4Al-3Mo-1V	4.507	0.1628
Ti-2.5Al-16V	4.65	0.168
Zinc and zinc alloys		
Pure zinc	7.133	0.2577
AG40A alloy	6.6	0.24
AC41A alloy	6.7	0.24
Commercial rolled zinc		
0.08% Pb	7.14	0.258
0.06 Pb, 0.06 Cd	7.14	0.258
0.3 Pb, 0.3 Cd	7.14	0.258
Copper-hardened, rolled zinc (1% Cu)	7.18	0.259
Rolled zinc alloy (1 Cu, 0.010 Mg)	7.18	0.259
Zn-Cu-Ti alloy (0.8 Cu, 0.15 Ti)	7.18	0.259
Precious metals		
Silver	10.49	0.379
Gold	19.32	0.698
70Au-30Pt	19.92	. . .
Platinum	21.45	0.775
Pt-3.5Rh	20.9	. . .
Pt-5Rh	20.65	. . .
Pt-10Rh	19.97	. . .
Pt-20Rh	18.74	. . .
Pt-30Rh	17.62	. . .
Pt-40Rh	16.63	. . .
Pt-5Ir	21.49	. . .
Pt-10Ir	21.53	. . .
Pt-15Ir	21.57	. . .
Pt-20Ir	21.61	. . .
Pt-25Ir	21.66	. . .
Pt-30Ir	21.70	. . .

(continued)

(a) 6.95 to 7.35 g/cm³ (0.251 to 0.265 lb/in.³). (b) 7.20 to 7.34 g/cm³ (0.260 to 0.265 lb/in.³). (c) Hexagonal. (e) Body-centered cubic. (f) Close-packed hexagonal. (g) Rhombohedral. Source: M. Bauccio, *ASM Metals Reference Book*, 3rd ed., ASM International, 1993

مرجع مهندسی مواد و متالورژی

Table 1   (continued)

Metal or alloy	Density g/cm³	lb/in.³	Metal or alloy	Density g/cm³	lb/in.³	Metal or alloy	Density g/cm³	lb/in.³
Pt-35Ir	21.79	...	Chromium	7.19	0.260	Uranium	19.07	0.689
Pt-5Ru	20.67	...	Cobalt	8.85	0.322	Vanadium	6.1	0.22
Pt-10Ru	19.94	...	Gallium	5.907	0.213	Zirconium	6.5	0.23
Palladium	12.02	0.4343	Germanium	5.323	0.192	**Rare earth metals**		
60Pd-40Cu	10.6	0.383	Hafnium	13.1	0.473	Cerium	8.23(c)	...
95.5Pd-4.5Ru	12.07(a)	...	Indium	7.31	0.264	Cerium	6.66(d)	...
95.5Pd-4.5Ru	11.62(b)	...	Iridium	22.5	0.813	Cerium	6.77(e)	...
Permanent magnet materials			Lithium	0.534	0.019	Dysprosium	8.55(f)	...
Cunico	8.30	0.300	Manganese	7.43	0.270	Erbium	9.15(f)	...
Cunife	8.61	0.311	Mercury	13.546	0.489	Europium	5.245(e)	...
Comol	8.16	0.295	Molybdenum	10.22	0.369	Gadolinium	7.86(f)	...
Alnico I	6.89	0.249	Niobium	8.57	0.310	Holmium	6.79(f)	...
Alnico II	7.09	0.256	Osmium	22.583	0.816	Lanthanum	6.19(d)	...
Alnico III	6.89	0.249	Plutonium	19.84	0.717	Lanthanum	6.18(c)	...
Alnico IV	7.00	0.253	Potassium	0.86	0.031	Lanthanum	5.97(e)	...
Alnico V	7.31	0.264	Rhenium	21.04	0.756	Lutetium	9.85(f)	...
Alnico VI	7.42	0.268	Rhodium	12.44	0.447	Neodymium	7.00(d)	...
Barium ferrite	4.7	0.17	Ruthenium	12.2	0.441	Neodymium	6.80(e)	...
Vectolite	3.13	0.113	Selenium	4.79	0.174	Praseodymium	6.77(d)	...
Pure metals			Silicon	2.33	0.084	Praseodymium	6.64(e)	...
Antimony	6.62	0.239	Silver	10.49	0.379	Samarium	7.49(g)	...
Beryllium	1.848	0.067	Sodium	0.97	0.035	Scandium	2.99(f)	...
Bismuth	9.80	0.354	Tantalum	16.6	0.600	Terbium	8.25(f)	...
Cadmium	8.65	0.313	Thalium	11.85	0.428	Thulium	9.31(f)	...
Calcium	1.55	0.056	Thorium	11.72	0.423	Ytterbium	6.96(c)	...
Cesium	1.903	0.069	Tungsten	19.3	0.697	Yttrium	4.47(f)	...

(a) 6.95 to 7.35 g/cm³ (0.251 to 0.265 lb/in.³). (b) 7.20 to 7.34 g/cm³ (0.260 to 0.265 lb/in.³). (c) Face-centered cubic. (d) Hexagonal. (e) Body-centered cubic. (f) Close-packed hexagonal. (g) Rhombohedral. Source: M. Bauccio, *ASM Metals Reference Book*, 3rd ed., ASM International, 1993

Linear Thermal Expansion of Metals and Alloys

Table 1 Linear thermal expansion of selected metals and alloys

Metal or alloy	Temp., °C	Coefficient of expansion, $10^{-6}/°C$	Metal or alloy	Temp., °C	Coefficient of expansion, $10^{-6}/°C$	Metal or alloy	Temp., °C	Coefficient of expansion, $10^{-6}/°C$
Aluminum and aluminum alloys			Red brass, 85%	20 to 300	18.7	**Iron and iron alloys**		
Aluminum (99.996%)	20 to 100	23.6	Low brass, 80%	20 to 300	19.1	Pure iron	20	11.7
Wrought alloys			Cartridge brass, 70%	20 to 300	19.9	Fe-C alloys		
EC. 1060, 1100	20 to 100	23.6	Yellow brass	20 to 300	20.3	0.06% C	20 to 100	11.7
2011, 2014	20 to 100	23.0	Muntz metal	20 to 300	20.8	0.22% C	20 to 100	11.7
2024	20 to 100	22.8	Leaded commercial bronze	20 to 300	18.4	0.40% C	20 to 100	11.3
2218	20 to 100	22.3	Low-leaded brass	20 to 300	20.2	0.56% C	20 to 100	11.0
3003	20 to 100	23.2	Medium-leaded brass	20 to 300	20.3	1.08% C	20 to 100	10.8
4032	20 to 100	19.4	High-leaded brass	20 to 300	20.3	1.45% C	20 to 100	10.1
5005, 5050, 5052	20 to 100	23.8	Extra-high-leaded brass	20 to 300	20.5	Invar (36% Ni)	20	0.2
5056	20 to 100	24.1	Free-cutting brass	20 to 300	20.5	13Mn-1.2C	20	18.0
5083	20 to 100	23.4	Leaded Muntz metal	20 to 300	20.8	13Cr-0.35C	20 to 100	10.0
5086	60 to 300	23.9	Forging brass	20 to 300	20.7	12.3Cr-0.4Ni-0.09C	20 to 100	9.8
5154	20 to 100	23.9	Architectural bronze	20 to 300	20.9	17.7Cr-9.6Ni-0.06C	20 to 100	16.5
5357	20 to 100	23.7	Inhibited admiralty	20 to 300	20.2	18W-4Cr-1V	0 to 100	11.2
5456	20 to 100	23.9	Naval brass	20 to 300	21.2	Gray cast iron	0 to 100	10.5
6061, 6063	20 to 100	23.4	Leaded naval brass	20 to 300	21.2	Malleable iron (pearlitic)	20 to 400	12
6101, 6151	20 to 100	23.0	Manganese bronze (A)	20 to 300	21.2	**Lead and lead alloys**		
7075	20 to 100	23.2	Phosphor bronze, 5% (A)	20 to 300	17.8	Corroding lead (99.73 + % Pb)	17 to 100	29.3
7079, 7178	20 to 100	23.4	Phosphor bronze, 8% (C)	20 to 300	18.2	5-95 solder	15 to 110	28.7
Casting alloys			Phosphor bronze, 10% (D)	20 to 300	18.4	20-80 solder	15 to 110	26.5
A13	20 to 100	20.4	Phosphor bronze, 1.25%	20 to 300	17.8	50-50 solder	15 to 110	23.4
43 and 108	20 to 100	22.0	Free-cutting phosphor bronze	20 to 300	17.3	1% antimonial lead	20 to 100	28.8
A108	20 to 100	21.5	Cupro-nickel, 30%	20 to 300	16.2	Hard lead (96Pb-4Sb)	20 to 100	27.8
A132	20 to 100	19.0	Cupro-nickel, 10%	20 to 300	17.1	Hard lead (94Pb-6Sb)	20 to 100	27.2
D132	20 to 100	20.05	Nickel silver, 65–18	20 to 300	16.2	8% antimonial lead	20 to 100	26.7
F132	20 to 100	20.7	Nickel silver, 55–18	20 to 300	16.7	9% antimonial lead	20 to 100	26.4
138	20 to 100	21.4	Nickel silver, 65–12	20 to 300	16.2	Lead-base babbitt		
142	20 to 100	22.5	High-silicon bronze (a)	20 to 300	18.0	SAE 14	20 to 100	19.6
195	20 to 100	23.0	Low-silicon bronze (b)	20 to 300	17.9	Alloy 8	20 to 100	24.0
B195	20 to 100	22.0	Aluminum bronze (3)	20 to 300	16.4	**Magnesium and magnesium alloys**		
214	20 to 100	24.0	Aluminum-silicon bronze	20 to 300	18.0	Magnesium (99.8%)	20	25.2
220	20 to 100	25.0	Aluminum bronze (1)	20 to 300	16.8	**Casting alloys**		
319	20 to 100	21.5	Beryllium copper	20 to 300	17.8	AM100A	18 to 100	25.2
355	20 to 100	22.0	**Casting alloys**			AZ63A	20 to 100	26.1
356	20 to 100	21.5	88Cu-8Sn-4Zn	21 to 177	18.0	AZ91A, B, C	20 to 100	26
360	20 to 100	21.0	89Cu-11Sn	20 to 300	18.4	AZ92A	18 to 100	25.2
750	20 to 100	23.1	88Cu-6Sn-1.5Pb-4.5Zn	21 to 260	18.5	HZ32A	20 to 200	26.7
40E	21 to 93	24.7	87Cu-8Sn-1Pb-4Zn	21 to 177	18.0	ZH42	20 to 200	27
Copper and copper alloys			87Cu-10Sn-1Pb-2Zn	21 to 177	18.0	ZH62A	20 to 200	27.1
Wrought coppers			80Cu-10Sn-10Pb	21 to 204	18.5	ZK51A	20	26.1
Pure copper	20	16.5	78Cu-7Sn-15Pb	21 to 204	18.5	EZ33A	20 to 100	26.1
Electrolytic tough pitch copper (ETP)	20 to 100	16.8	85Cu-5Sn-5Pb-5Zn	21 to 204	18.1	EK30A, EK41A	20 to 100	26.1
Deoxidized copper, high residual phosphorus (DHP)	20 to 300	17.7	72Cu-1Sn-3Pb-24Zn	21 to 93	20.7	**Wrought alloys**		
Oxygen-free copper	20 to 300	17.7	67Cu-1Sn-3Pb-29Zn	21 to 93	20.2	M1A, A3A	20 to 100	26
Free machining copper, 0.5% Te or 1% Pb	20 to 300	17.7	61Cu-1Sn-1Pb-37Zn	21 to 260	21.6	AZ31B, PE	20 to 100	26
Wrought alloys			Manganese bronze			AZ61A, AZ80A	20 to 100	26
Gilding, 95%	20 to 300	18.1	60 ksi	21 to 204	20.5	ZK60A, B	20 to 100	26
Commerical bronze, 90%	20 to 300	18.4	65 ksi	21 to 93	21.6	HM31A	20 to 93	26.1
Jewelry bronze, 87.5%	20 to 300	18.6	110 ksi	21 to 260	19.8	**Nickel and nickel alloys**		
			Aluminum bronze			Nickel (99.95% Ni + Co)	0 to 100	13.3
			Alloy 9A	. . .	17	Duranickel	0 to 100	13.0
			Alloy 9B	20 to 250	17	Monel	0 to 100	14.0
			Alloys 9C, 9D	. . .	16.2			

(continued)

(a) Longitudinal; 23.4 transverse. (b) Longitudinal; 21.1 transverse. (c) Longitudinal; 19.4 transverse. Source: M. Bauccio, *ASM Metals Reference Book*, 3rd ed., ASM International, 1993

Table 1 (continued)

Metal or alloy	Temp., °C	Coefficient of expansion, 10^{-6}/°C	Metal or alloy	Temp., °C	Coefficient of expansion, 10^{-6}/°C	Metal or alloy	Temp., °C	Coefficient of expansion, 10^{-6}/°C
Monel (cast)	25 to 100	12.9	Solder (63Sn-37Pb)	15 to 110	24.7	Cadmium	20	29.8
Inconel	20 to 100	11.5	**Titanium and titanium alloys**			Calcium	0 to 400	22.3
Ni-o-nel	27 to 93	12.9	99.9% Ti	20	8.41	Chromium	20	6.2
Hastelloy B	0 to 100	10.0	99.0% Ti	93	8.55	Cobalt	20	13.8
Hastelloy C	0 to 100	11.3	Ti-5Al-2.5Sn	93	9.36	Gold	20	14.2
Hastelloy D	0 to 100	11.0	Ti-8Mn	93	8.64	Iridium	20	6.8
Hastelloy F	20 to 100	14.2	**Zinc and zinc alloys**			Lithium	20	56
Hastelloy N	21 to 204	10.4	Pure zinc	20 to 250	39.7	Manganese	0 to 100	22
Hastelloy W	23 to 100	11.3	AG40A alloy	20 to 100	27.4	Palladium	20	11.76
Hastelloy X	26 to 100	13.8	AC41A alloy	20 to 100	27.4	Platinum	20	8.9
Illium G	0 to 100	12.19	**Commercial rolled zinc**			Rhenium	20 to 500	6.7
Illium R	0 to 100	12.02	0.08 Pb	20 to 40	32.5	Rhodium	20 to 100	8.3
80Ni-20Cr	20 to 1000	17.3	0.3 Pb, 0.3 Cd	20 to 98	33.9(a)	Ruthenium	20	9.1
60Ni-24Fe-16Cr	20 to 1000	17.0	Rolled zinc alloy	20 to 100	34.8(b)	Silicon	0 to 1400	5
35Ni-45Fe-20Cr	20 to 500	15.8	(1 Cu, 0.010 Mg)			Silver	0 to 100	19.68
Constantan	20 to 1000	18.8	Zn-Cu-Ti alloy (0.8 Cu, 0.15 Ti)	20 to 100	24.9(c)	Tungsten	27	4.6
Tin and tin alloys			**Pure metals**			Vanadium	23 to 100	8.3
Pure tin	0 to 100	23	Beryllium	25 to 100	11.6	Zirconium	. . .	5.85
Solder (70Sn-30Pb)	15 to 110	21.6						

(a) Longitudinal; 23.4 transverse. (b) Longitudinal; 21.1 transverse. (c) Longitudinal; 19.4 transverse. Source: M. Bauccio, *ASM Metals Reference Book*, 3rd ed., ASM International, 1993

Thermal Conductivity of Metals and Alloys

Table 1   Thermal conductivity of selected metals and alloys

Metal or alloy	Thermal conductivity near room temperature, cal/(cm · s · °C)
Aluminum and aluminum alloys	
Wrought alloys	
EC(O)	0.57
1060(O)	0.56
1100	0.53
2011 (T3)	0.34
2014 (O)	0.46
2024 (O)	0.45
2218 (T72)	0.37
3003 (O)	0.46
4032 (O)	0.37
5005	0.48
5050 (O)	0.46
5052 (O)	0.33
5056 (O)	0.28
5083	0.28
5086	0.30
5154	0.30
5357	0.40
5456	0.28
6061 (O)	0.41
6063 (O)	0.52
6101 (T6)	0.52
6151 (O)	0.49
7075 (T6)	0.29
7079 (T6)	0.29
7178	0.29
Casting alloys	
A13	0.29
43 (F)	0.34
108 (F)	0.29
A108	0.34
A132 (T551)	0.28
D132 (T5)	0.25
F132	0.25
138	0.24
142 (T21, sand)	0.40
195 (T4, T62)	0.33
B195 (T4, T6)	0.31
214	0.33
200 (T4)	0.21
319	0.26
355 (T51, sand)	0.40
356 (T51, sand)	0.40
360	0.35
380	0.23
750	0.44
40E	0.33
Copper and copper alloys	
Wrought coppers	
Pure copper	0.941
Electrolytic tough pitch copper (ETP)	0.934
Deoxidized copper, high residual phosphorus (DHP)	0.81

Metal or alloy	Thermal conductivity near room temperature, cal/(cm · s · °C)
Free-machining copper (0.5% Te)	0.88
Free-machining copper (1% Pb)	0.92
Wrought alloys	
Gilding, 95%	0.56
Commercial bronze, 90%	0.45
Jewelry bronze, 87.5%	0.41
Red brass, 85%	0.38
Low brass, 80%	0.33
Cartridge brass, 70%	0.29
Yellow brass	0.28
Muntz metal	0.29
Leaded-commercial bronze	0.43
Low-leaded brass (tube)	0.28
Medium-leaded brass	0.28
High-leaded brass (tube)	0.28
High-leaded brass	0.28
Extra-high-leaded brass	0.28
Leaded Muntz metal	0.29
Forging brass	0.28
Architectural bronze	0.29
Inhibited admiralty	0.26
Naval brass	0.28
Leaded naval brass	0.28
Manganese bronze (A)	0.26
Phosphor bronze, 5% (A)	0.17
Phosphor bronze, 8% (C)	0.15
Phosphor bronze, 10% (D)	0.12
Phosphor bronze, 1.25%	0.49
Free-cutting phosphor bronze	0.18
Cupro-nickel, 30%	0.07
Cupro-nickel, 10%	0.095
Nickel silver, 65-18	0.08
Nickel silver, 55-18	0.07
Nickel silver, 65-12	0.10
High-silicon bronze (A)	0.09
Low-silicon bronze (B)	0.14
Aluminum bronze, 5% Al	0.198
Aluminum bronze, (3)	0.18
Aluminum-silicon bronze	0.108
Aluminum bronze, (1)	0.144
Aluminum bronze, (2)	0.091
Beryllium copper	0.20(a)
Casting alloys	
Chromium copper (1% Cr)	0.4(a)
89cu-11Sn	0.121
88Cu-6Sn-1.5Pb-4.5Zn	(b)
87Cu-8Sn-1Pb-4Zn	(c)
87Cu-10Sn-1Pb-2Zn	(c)
80Cu-10Sn-10Pb	(c)
Manganese bronze, 110 ksi	(d)
Aluminum bronze	
Alloy 9A	(e)
Alloy 9B	(f)
Alloy 9C	(b)
Alloy 9D	(c)
Propeller bronze	(g)

Metal or alloy	Thermal conductivity near room temperature, cal/(cm · s · °C)
Nickel silver	
12% Ni	(h)
16% Ni	(h)
20% Ni	(j)
25% Ni	(k)
Silicon bronze	(h)
Iron and iron alloys	
Pure iron	0.178
Cast iron (3.16 C, 1.54 Si, 0.57 Mn)	0.112
Carbon steel (0.23 C, 0.64 Mn)	0.124
Carbon steel (1.22 C, 0.35 Mn)	0.108
Alloy steel (0.34 C, 0.55 Mn, 0.78 Cr, 3.53 Ni, 0.39 Mo, 0.05 Cu)	0.079
Type 410	0.057
Type 304	0.036
T1 tool steel	0.058
Lead and lead alloys	
Corroding lead (99.73 + % Pb)	0.083
5-95 solder	0.085
20-80 solder	0.089
50-50 solder	0.111
1% antimonial lead	0.080
Hard lead (96Pb-4Sb)	0.073
Hard lead (94Pb-6Sb)	0.069
8% antimonial lead	0.065
9% antimonial lead	0.064
Lead-base babbitt (SAE 14)	0.057
Lead-base babbitt (alloy 8)	0.058
Magnesium and magnesium alloys	
Magnesium (99.8%)	0.367
Casting alloys	
AM100A	0.17
AZ63A	0.18
AZ81A (T4)	0.12
AZ91A, B, C	0.17
AZ92A	0.17
HK31A (T6, sand cast)	0.22
HZ32A	0.26
ZH42	0.27
ZH62A	0.26
ZK51A	0.26
ZE41A (T5)	0.27
EZ33A	0.24
EK30A	0.26
EK41A (T5)	0.24
Wrought alloys	
M1A	0.33
AZ31B	0.23
AZ61A	0.19
AZ80A	0.18
ZK60A, B (F)	0.28
ZE10A (O)	0.33
HM21A (O)	0.33

(continued)

(a) Depends on processing. (b) 18% of Cu. (c) 12% of Cu. (d) 9.05% of Cu. (e) 15% of Cu. (f) 16% of Cu. (g) 11% of Cu. (h) 7% of Cu. (j) 6% of Cu. (k) 6.5% of Cu. To convert cal/(cm · s · °C) to w/m · k, multiply by 418%. Source: M. Bauccio, *ASM Metals Reference Book*, 3rd ed., ASM International, 1993.

Table 1 (continued)

Metal or alloy	Thermal conductivity near room temperature, cal/(cm · s · °C)	Metal or alloy	Thermal conductivity near room temperature, cal/(cm · s · °C)	Metal or alloy	Thermal conductivity near room temperature, cal/(cm · s · °C)
HM31A	0.25	Ti-5Al-2.5Sn	0.019	Lithium	0.17
Nickel and nickel alloys		Ti-2Fe-2Cr-2Mo	0.028	Molybdenum	0.34
Nickel (99.95% Ni + Co)	0.22	Ti-8Mn	0.026	Niobium	0.13
"A" nickel	0.145			Palladium	0.168
"D" nickel	0.115	**Zinc and zinc alloys**		Platinum	0.165
Monel	0.062	Pure zinc	0.27	Plutonium	0.020
"K" Monel	0.045	AG40A alloy	0.27	Rhenium	0.17
Inconel	0.036	AC41A alloy	0.26	Rhodium	0.21
Hastelloy B	0.027	Commercial rolled zinc		Silicon	0.20
Hastelloy C	0.03	0.08 Pb	0.257	Silver	1.0
Hastelloy D	0.05	0.06 Pb, 0.06 Cd	0.257	Sodium	0.32
Illium G	0.029	Rolled zinc alloy (1 Cu, 0.010 Mg)	0.25	Tantalum	0.130
Illium R	0.031	Zn-Cu-Ti alloy (0.8 Cu, 0.15 Ti)	0.25	Thallium	0.093
60Ni-24Fe-16Cr	0.032			Thorium	0.090
35Ni-45Fe-20Cr	0.031	**Pure metals**		Tungsten	0.397
Constantan	0.051	Beryllium	0.35	Uranium	0.071
Tin and tin alloys		Cadmium	0.22	Vanadium	0.074
Pure tin	0.15	Chromium	0.16	Yttrium	0.035
Soft solder (63Sn-37Pb)	0.12	Cobalt	0.165		
Tin foil (92Sn-8Zn)	0.14	Germanium	0.14		
Titanium and titanium alloys		Gold	0.71		
Titanium (99.0%)	0.043	Indium	0.057		
		Iridium	0.14		

(a) Depends on processing. (b) 18% of Cu. (c) 12% of Cu. (d) 9.05% of Cu. (e) 15% of Cu. (f) 16% of Cu. (g) 11% of Cu. (h) 7% of Cu. (j) 6% of Cu. (k) 6.5% of Cu. To convert cal/(cm · s · °C) to w/m · k, multiply by 418%, Source: M. Bauccio, *ASM Metals Reference Book*, 3rd ed., ASM International, 1993.

Electrical Conductivity of Metals and Alloys

Table 1   Electrical conductivity and resistivity of selected metals and alloys at ambient temperature

Metal or alloy	Conductivity, % IACS	Resistivity μΩ • cm	Metal or alloy	Conductivity, % IACS	Resistivity μΩ • cm	Metal or alloy	Conductivity, % IACS	Resistivity μΩ • cm
Aluminum and aluminum alloys			**Platinum and platinum alloys**			Molybdenum disilicide, MoSi$_2$	4.5	37.24
Aluminum (99.996%)	64.95	2.65	Platinum	16	10.6	Graphite	. . .	910.1
EC (O, H19)	62	2.8	95Pt-5Ir	9	19			
5052 (O, H38)	35	4.93	90Pt-10Ir	7	25	**Instrument and control alloys**		
5056 (H38)	27	6.4	85Pt-15Ir	6	28.5	**Cu-Ni alloys**		
6101 (T6)	56	3.1	80Pt-20Ir	5.6	31	98Cu-2Ni	35	4.99
			75Pt-25Ir	5.5	33	94Cu-6Ni	17	9.93
Copper and copper alloys			70Pt-30Ir	5	35	89Cu-11Ni	11	14.96
Wrought copper			65Pt-35Ir	5	36	78Cu-22Ni	5.7	29.92
Pure copper	103.06	1.67	95Pt-5Ru	5.5	31.5	55Cu-45Ni (constantan)	3.5	49.87
Electrolytic (ETP)	101	1.71	90Pt-10Ru	4	43	**Cu-Mn-Ni alloys**		
Oxygen-free copper (OF)	101	1.71	89Pt-11Ru	4	43	87Cu-13Mn (manganin)	3.5	48.21
Free-machining copper			86Pt-14Ru	3.5	46	83Cu-13Mn-4Ni (manganin)	3.5	48.21
0.5% Te	95	1.82	96Pt-4W	5	36	85Cu-10Mn-4Ni (shunt manganin)	4.5	38.23
1.0% Pb	98	1.76	**Palladium and palladium alloys**			70Cu-20Ni-10Mn	3.6	48.88
Wrought alloys			Palladium	16	10.8	67Cu-5Ni-27Mn	1.8	99.74
Cartridge brass, 70%	28	6.2	95.5Pd-4.5Ru	7	24.2	**Ni-base alloys**		
Yellow brass	27	6.4	90Pd-10Ru	6.5	27	99.8 Ni	23	7.98
Leaded commercial bronze	42	4.1	70Pd-30Ag	4.3	40	71Ni-29Fe	9	19.95
Phosphor bronze, 1.25%	48	3.6	60Pd-40Ag	4.0	43	80Ni-20Cr	1.5	112.2
Nickel silver, 55-18	5.5	31	50Pd-50Ag	5.5	31.5	75Ni-20Cr-3Al+Cu or Fe	1.3	132.98
Low-silicon bronze (B)	12	14.3	72Pd-26Ag-2Ni	4	43	76Ni-17Cr-4Si-3Mn	1.3	132.98
Beryllium copper	22 to 30(a)	5.7 to 7.8(a)	60Pd-40Cu	5	35(c)	60Ni-16Cr-24Fe	1.5	112.2
Casting alloys			45Pd-30Ag-20Au-5Pt	4.5	39	35Ni-20Cr-45Fe	1.7	101.4
Chromium copper (1% Cr)	80 to 90(a)	2.10	35Pd-30Ag-14Cu-10Pt-10Au-1Zn	5	35	**Fe-Cr-Al alloy**		
88Cu-8Sn-4Zn	11	15	**Gold and gold alloys**			72Fe-23Cr-5Al-0.5Co	1.3	135.48
87Cu-10Sn-1Pb-2Zn	11	15	Gold	75	2.35	**Pure metals**		
			90Au-10Cu	16	10.8	Iron (99.99%)	17.75	9.71
Electrical contact materials			75Au-25Ag	16	10.8	**Thermostat metals**		
Copper alloys			72.5Au-14Cu-8.5Pt-4Ag-1Zn	10	17	75Fe-22Ni-3Cr	3	78.13
0.04 oxide	100	1.72	69Au-25Ag-6Pt	11	15	72Mn-18Cu-10Ni	1.5	112.2
1.25 Sn + P	48	3.6	41.7Au-32.5Cu-18.8Ni-7Zn	4.5	39	67Ni-30Cu-1.4Fe-1Mn	3.5	56.52
5 Sn + P	18	11				75Fe-22Ni-3Cr	12	15.79
8 Sn + P	13	13	**Electrical heating alloys**			66.5Fe-22Ni-8.5Cr	3.3	58.18
15 Zn	37	4.7	**Ni-Cr and Ni-Cr-Fe alloys**					
20 Zn	32	5.4	78.5Ni-20Cr-1.5Si (80-20)	1.6	108.05	**Permanent magnet materials**		
35 Zn	27	6.4	73.5Ni-20Cr-5Al-1.5Si	1.2	137.97	Carbon steel (0.65% C)	9.5	18
2 Be + Ni or Co(b)	17 to 21	9.6 to 11.5	68Ni-20Cr-8.5Fe-2Si	1.5	116.36	Carbon steel (1% C)	8	20
Silver and silver alloys			60Ni-16Cr-22.5Fe-1.5Si	1.5	112.20	Chromium steel (3.5% Cr)	6.1	29
Fine silver	106	1.59	35Ni-20Cr-43.5Fe-1.5Si	1.7	101.4	Tungsten steel (6% W)	6	30
92.5Ag-7.5Cu	85	2	**Fe-Cr-Al alloys**			Cobalt steel (17% Co)	6.3	28
90Ag-10Cu	85	2	72Fe-23Cr-5Al	1.3	138.8	Cobalt steel (36% Co)	6.5	27
72Ag-28Cu	87	2	55Fe-37.5Cr-7.5Al	1.2	166.23	**Intermediate alloys**		
72Ag-26Cu-2Ni	60	2.9	**Pure metals**			Cunico	7.5	24
85Ag-15Cd	35	4.93	Molybdenum	34	5.2	Cunife	9.5	18
97Ag-3Pt	50	3.5	Platinum	16	10.64	Comol	3.6	45
97Ag-3Pd	60	2.9	Tantalum	13.9	12.45	**Alnico alloys**		
90Ag-10Pd	30	5.3	Tungsten	30	5.65	Alnico I	3.3	75
90Ag-10Au	40	4.2	**Nonmetallic heating element materials**					
60Ag-40Pd	8	23	Silicon carbide, SiC	1 to 1.7	100 to 200			
70Ag-30Pd	12	14.3						

(continued)

(a) Precipitation hardened; depends on processing. (b) A heat-treatable alloy. (c) Annealed and quenched. (d) At low field strength and high electrical resistance. (e) At higher field strength; annealed for optimum magnetic properties. Source: M. Bauccio, *ASM Metals Reference Book*, 3rd ed., ASM International, 1993.

Table 1 (continued)

Metal or alloy	Conductivity, % IACS	Resistivity μΩ • cm	Metal or alloy	Conductivity, % IACS	Resistivity μΩ • cm	Metal or alloy	Conductivity, % IACS	Resistivity μΩ • cm
Alnico II	3.3	65	4-79 Moly Permalloy, Hymu 80	3	58	2.5% Si	4	41
Alnico III	3.3	60				3% Si	3.5	48
Alnico IV	3.3	75	Mumetal	3	60	3% Si, grain-oriented	3.5	48
Alnico V	3.5	47	1040 alloy	3	56	4% Si	3	50
Alnico VI	3.5	50	High Permalloy 49, A-L 4750, Armco 48	3.6	48	**Stainless steels**		
Magnetically soft materials						Type 410	3	57
			45 Permalloy	3.6	45	Type 416	3	57
Electrical steel sheet						Type 430	3	60
M-50	9.5	18	**High-permeability materials(e)**			Type 443	3	68
M-43	6 to 9	20 to 28	Supermendur	4.5	40	Type 446	3	61
M-36	5.5 to 7.5	24 to 33	2V Permendur	4.5	40	**Nickel irons**		
M-27	3.5 to 5.5	32 to 47	35% Co, 1% Cr	9	20	50% Ni	3.5	48
M-22	3.5 to 5	41 to 52	Ingot iron	17.5	10	78% Ni	11	16
M-19	3.5 to 5	41 to 56	0.5% Si steel	6	28	77% Ni (Cu, Cr)	3	60
M-17	3 to 3.5	45 to 58	1.75% Si steel	4.6	37	79% Ni (Mo)	3	58
M-15	3 to 3.5	45 to 69	3.0% Si steel	3.6	47	**Stainless and heat-resisting alloys**		
M-14	3 to 3.5	58 to 69	Grain-oriented 3.0% Si steel	3.5	50	Type 302	3	72
M-7	3 to 3.5	45 to 52				Type 309	2.5	78
M-6	3 to 3.5	45 to 52	Grain-oriented 50% Ni iron	3.6	45	Type 316	2.5	74
M-5	3 to 3.5	45 to 52	50% Ni iron	3.5	50	Type 317	2.5	74
			Relay steels and alloys after annealing			Type 347	2.5	73
Moderately high-permeability materials(d)						Type 403	3	57
Thermenol	0.5	162	**Low-carbon iron and steel**			Type 405	3	60
16 Alfenol	0.7	153				Type 501	4.5	40
Sinimax	2	90	Low-carbon iron	17.5	10	HH	2.5	80
Monimax	2.5	80	1010 steel	14.5	12	HK	2	90
Supermalloy	3	65	**Silicon steels**			HT	1.7	100
			1% Si	7.5	23			

(a) Precipitation hardened; depends on processing. (b) A heat-treatable alloy. (c) Annealed and quenched. (d) At low field strength and high electrical resistance. (e) At higher field strength; annealed for optimum magnetic properties. Source: M. Bauccio, *ASM Metals Reference Book*, 3rd ed., ASM International, 1993.

Vapor Pressures of the Elements

Table 1   Vapor pressures of selected metallic elements

| | Pressure, atm | | | | | | | | | | |
| | 0.0001 | | 0.001 | | 0.01 | | 0.1 | | 0.50 | | 1.0 | |
Element	°C	°F	°C	°F	°C	°F	°C	°F	°C	°F	°C	°F
Aluminum	1110	2030	1263	2305	1461	2662	1713	3115	1940	3524	2056	3733
Antimony	759	1398	872	1602	1013	1855	1196	2185	1359	2478	1440	2624
Arsenic	308	586	363	685	428	802	499	930	578	1072	610	1130
Bismuth	914	1677	1008	1846	1121	2050	1254	2289	1367	2493	1420	2588
Cadmium	307(a)	585(a)	384(b)	723(b)	471	880	594	1101	708	1306	765	1409
Calcium	688	1270	802(c)	1476(c)	958(b)	1756(b)	1175	2147	1380	2516	1487	2709
Carbon	3257	5895	3547	6417	3897	7047	4317	7803	4667	8433	4827	8721
Chromium	1420(a)	2588(a)	1594(b)	2901(b)	1813	3295	2097	3807	2351	4264	2482	4500
Copper	1412	2574	1602	2916	1844	3351	2162	3924	2450	4442	2595	4703
Gallium	1178	2152	1329	2424	1515	2759	1751	3184	1965	3569	2071	3760
Gold	1623	2953	1839	3342	2115	3839	2469	4476	2796	5065	2966	5371
Iron	1564	2847	1760	3200	2004	3639	2316	4201	2595	4703	2735	4955
Lead	815	1499	953	1747	1135	2075	1384	2523	1622	2952	1744	3171
Lithium	592	1098	707	1305	858	1576	1064	1947	1266	2311	1372	2502
Magnesium	516	961	608(a)	1126(a)	725(b)	1337(b)	886	1627	1030	1886	1107	2025
Manganese	1115(d)	2039(d)	1269(b)	2316(b)	1476	2889	1750	3182	2019	3666	2151	3904
Mercury	77.9(b)	172.2(b)	120.8	249.4	176.1	349.0	251.3	484.3	321.5	610.7	357	675
Molybdenum	2727	4941	3057	5535	3477	6291	4027	7281	4537	8199	4804	8679
Nickel	1586	2887	1782	3240	2025	3677	2321	4210	2593	4699	2732	4950
Platinum	2367	4293	2687	4869	3087	5589	3637	6579	4147	7497	4407	7965
Potassium	261	502	332	630	429	804	565	1051	704	1299	774	1425
Rubidium	223	433	288	550	377	711	497	927	617	1143	679	1254
Selenium	282	540	347	657	430	806	540	1004	634	1173	680	1256
Silicon	1572	2862	1707	3105	1867	3393	2057	3735	2217	4023	2287	4149
Silver	1169	2136	1334	2433	1543	2809	1825	3317	2081	3778	2212	4014
Sodium	349	660	429	804	534	993	679	1254	819	1506	892	1638
Strontium	...	...	(a)	(a)	877(b)	1629(b)	1081	1978	1279	2334	1384	2523
Tellurium	(a)	(a)	509(b)	948(b)	632	1170	810	1490	991	1816	1087	1989
Thallium	692	1277	809	1488	962	1764	1166	2131	1359	2478	1457	2655
Tin	...	...	...	...	...	...	1932(b)	3510(b)	2163	3925	2270	4118
Tungsten	3547	6417	3937	7119	4437	8019	5077	9171	5647	10197	5927	10701
Zinc	399(a)	750(a)	477(b)	891(b)	579	1074	717	1323	842	1548	907	1665

(a) In the solid state. (b) In the liquid state. (c) β. (d) γ. Source: M. Bauccio, *ASM Metals Reference Book*, 3rd ed., ASM International, 1993.

Reference Information

Metric Conversion Guide

This Section is intended as a guide for expressing weights and measures in the Système International d'Unités (SI). The purpose of SI units, developed and maintained by the General Conference of Weights and Measures, is to provide a basis for worldwide standardization of units and measure. For more information on metric conversions, the reader should consult the following references:

- *The International System of Units*, SP 330, 1991, National Institute of Standards and Technology. Order from Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402-9325
- *Metric Editorial Guide,* 5th ed. (revised), 1993, American National Metric Council, 4340 East West Highway, Suite 401, Bethesda, MD 20814-4411

- "Standard for Use of the International system of Units (SI): The Modern Metric System," IEEE/ASTM SI 10-1997 Institute of Electrical and Elcectronics Enigineers, 345 East 47th Street, New York, NY 10017, USA
- *Guide for the Use of the Internaitional System of Units (SI)*, SP811, 1995, National lInstitute of Standers and Technolohy, U.S. Government Printing Office, Washington, DC 20402

Base, supplementary, and derived SI units

Measure	Unit	Symbol	Measure	Unit	Symbol
Base units			Entropy	joule per kelvin	J/K
			Force	newton	N
Amount of substance	mole	mol	Frequency	hertz	Hz
Electric current	ampere	A	Heat capacity	joule per kelvin	J/K
Length	meter	m	Heat flux density	watt per square meter	W/m^2
Luminous intensity	candela	cd	Illuminance	lux	lx
Mass	kilogram	kg	Inductance	henry	H
Thermodynamic temperature	kelvin	K	Irradiance	watt per square meter	W/m^2
Time	second	s	Luminance	candela per square meter	cd/m^2
			Luminous flux	lumen	lm
Supplementary units			Magnetic field strength	ampere per meter	A/m
Plane angle	radian	rad	Magnetic flux	weber	Wb
Solid angle	steradian	sr	Magnetic flux density	tesla	T
			Molar energy	joule per mole	J/mol
			Molar entropy	joule per mole kelvin	J/mol · K
Derived units			Molar heat capacity	joule per mole kelvin	J/mol · K
Absorbed does	gray	Gy	Moment of force	newton meter	N · m
Acceleration	meter per second squared	m/s^2	Permeability	henry per meter	H/m
Activity (of radionuclides)	becquerel	Bq	Permittivity	farad per meter	F/m
Angular acceleration	radian per second squared	rad/s^2	Power, radiant flux	watt	W
Angular velocity	radian per second	rad/s	Pressure, stress	pascal	Pa
Area	square meter	m^2	Quantity of electricity, electric charge	coulomb	C
Capacitance	farad	F	Radiance	watt per square meter steradian	W/m^2 · sr
Concentration (of amount of substance)	mole per cubic meter	mol/m^3	Radiant intensity	watt per steradian	W/sr
Current density	ampere per square meter	A/m^2	Specific heat capacity	joule per kilogram kelvin	J/kg · K
Density, mass	kilogram per cubic meter	kg/m^3	Specific energy	joule per kilogram	J/kg
Does equivalent, dose equivalent index	sievert	Sv	Specific entropy	joule per kilogram kelvin	J/kg · K
Electric change density	coulomb per cubic meter	C/m^3	Specific volume	cubic meter per kilogram	m^3/kg
Electric conductance	siemens	S	Specific volume	cubic meter per kilogram	m^3/kg
Electric field strength	volt per cubic meter	V/m	Surface tension	newton per meter	N/m
Electric flux density	coulomb per square meter	C/m^2	Thermal conductivity	watt per meter kelvin	W/m · K
Electric potential, potential difference, electromotive force	volt	V	Velocity	meter per second	m/s
			Viscosity, dynamic	pascal second	Pa · s
Electric resistance	ohm	Ω	Viscosity, kinematic	square meter per second	m^2/s
Energy, work, quantity of heat	joule	J	Volume	cubic meter	m^3
Energy density	joule per cubic meter	J/m^3	Wavenumber	1 per meter	1/m

Conversion factors

To convert from	to	multiply by	To convert from	to	multiply by	To convert from	to	multiply by
Angle			**Impact energy per unit area**			**Specific heat**		
degree	rad	1.745 329 E − 02	ft.lbf/ft^2	J/m^2	1.459 002 E + 01	Btu/lb · °F	J/kg · K	4.186 800 E + 03
Area			**Length**			cal/g · °C	J/kg · K	4.186 800 E + 03
in.^{2}	mm^2	6.451 600 E + 02	Å	nm	1.000 000 E − 01	**Stress (force per unit area)**		
in.^{2}	cm^2	6.451 600 E + 00	μin.	μm	2.540 000 E − 02	tonf/in.^{2} (tsi)	MPa	1.378 951 E + 01
in.^{2}	m^2	6.451 600 E − 04	mil	μm	2.540 000 E + 01	kgf/mm^2	MPa	9.806 650 E + 00
ft^2	m^2	9.290 304 E − 02	in.	mm	2.540 000 E + 01	ksi	MPa	6.894 757 E + 00
Bending moment or torque			in.	cm	2.540 000 E + 00	lbf/in.^{2} (psi)	MPa	6.894 757 E − 03
lbf · in.	N · m	1.129 848 E − 01	ft	m	3.048 000 E − 01	MN/m^2	MPa	1.000 000 E + 00
lbf · ft	N · m	1.355 818 E + 00	yd	m	9.144 000 E − 01	**Temperature**		
kgf · m	N · m	9.806 650 E + 00	mile, international	km	1.609 344 E + 00	°F	°C	5/9 · (°F − 32)
ozf · in.	N · m	7.061 552 E − 03	mile, nautical	km	1.852 000 E + 00	°R	°K	5/9
Bending moment or torque per unit length			mile, U. S. statute	km	1.609 347 E + 00	K	°C	K−273.15
lbf · in./in.	N · m/m	4.448 222 E + 00	**Mass**			**Temperature interval**		
lbf · ft/in.	N · m/m	5.337 866 E + 01	oz	kg	2.834 952 E − 02	°F	°C	5/9
Current density			lb	kg	4.535 924 E − 01	**Thermal conductivity**		
A/cm^2	A/cm^2	1.550 003 E − 01	ton (short, 2000 lb)	kg	9.071 847 E + 02	Btu · in./s · ft^2 · °F	W/m · K	5.192 204 E + 02
A/in.^{2}	A/mm^2	1.550 003 E − 03	ton (short, 2000 lb)	kg × 10^3(a)	9.071 847 E − 01	Btu/ft · h · °F	W/m · K	1.730 735 E + 00
A/ft^2	A/m^2	1.076 400 E + 01	ton (long, 2240 lb)	kg	1.016 047 E + 03	Btu · in./h · ft^2 · °F	W/m · K	1.442 279 E − 01
Electricity and magnetism			**Mass per unit area**			cal/cm · s · °C	W/m · K	4.184 000 E + 02
gauss	T	1.000 000 E − 04	oz/in.^{2}	kg/m^2	4.395 000 E + 01	**Thermal expansion(b)**		
maxwell	μWb	1.000 000 E − 02	oz/ft^2	kg/m^2	3.051 517 E − 01	in./in. · °C	m/m · K	1.000 000 E + 00
mho	S	1.000 000 E + 00	oz/yd^2	kg/m^2	3.390 575 E − 02	in./in. · °F	m/m · K	1.800 000 E + 00
Oersted	A/m	7.957 700 E + 01	lb/ft^2	kg/m^2	4.882 428 E + 00	**Velocity**		
Ω · cm	Ω · m	1.000 000 E − 02	**Mass per unit length**			ft/h	m/s	8.466 667 E − 05
Ω circular−mil/ft	μΩ · m	1.662 426 E − 03	lb/ft	kg/m	1.488 164 E + 00	ft/min	m/s	5.080 000 E − 03
Energy (impact, other)			lb/in.	kg/m	1.785 797 E + 01	ft/s	m/s	3.048 000 E − 01
ft · lbf	J	1.355 818 E + 00	**Mass per unit time**			in./s	m/s	2.540 000 E − 02
Btu (thermochemical)	J	1.054 350 E + 03	lb/h	kg/s	1.259 979 E − 04	km/h	m/s	2.777 778 E − 01
cal (thermochemical)	J	4.184 000 E + 00	lb/min	kg/s	7.559 873 E − 03	mph	km/h	1.609 344 E + 00
kW · h	J	3.600 000 E + 06	lb/s	kg/s	4.535 924 E − 01	**Velocity of rotation**		
W · h	J	3.600 000 E + 03	**Mass per unit volume (includes density)**			rev/min (rpm)	rad/s	1.047 164 E − 01
Flow rate			g/cm^3	kg/m^3	1.000 000 E + 03	rev/s	rad/s	6.283 185 E + 00
ft^3/h	L/min	4.719 475 E − 01	lb/ft^3	g/cm^3	1.601 846 E − 02	**Viscosity**		
ft^3/min	L/min	2.831 000 E + 01	lb/ft^3	kg/m^3	1.601 846 E + 01	poise	Pa · s	1.000 000 E − 01
gal/h	L/min	6.309 020 E − 02	lb/in.^{3}	g/cm^3	2.767 990 E + 01	stokes	m^2/s	1.000 000 E − 04
gal/min	L/min	3.785 412 E + 00	lb/in.^{3}	kg/m^3	2.767 990 E + 04	ft^2/s	m^2/s	9.290 304 E − 02
Force			**Power**			in.^{2}/s	mm^2/s	6.451 600 E + 02
lbf	N	4.448 222 E + 00	Btu/s	kW	1.055 056 E + 00	**Volume**		
kip (1000 lbf)	N	4.448 222 E + 03	Btu/min	kW	1.758 426 E − 02	in.^{3}	m^3	1.638 706 E − 05
tonf	kN	8.896 443 E + 00	Btu/h	W	2.928 751 E − 01	ft^3	m^3	2.831 685 E − 02
kgf	N	9.806 650 E + 00	erg/s	W	1.000 000 E − 07	fluid oz	m^3	2.957 353 E − 05
Force per unit length			ft · lbf/s	W	1.355 818 E + 00	gal (U.S. liquid)	m^3	3.785 412 E − 03
lbf/ft	N/m	1.459 390 E + 01	ft · lbf/min	W	2.259 697 E − 02	**Volume per unit time**		
lbf/in.	N/m	1.751 268 E + 02	ft · lbf/h	W	3.766 161 E − 04	ft^3/min	m^3/s	4.719 474 E − 04
Fracture toughness			hp (550 ft · lbf/s)	kW	7.456 999 E − 01	ft^3/s	m^3/s	2.831 685 E − 02
ksi $\sqrt{\text{in.}}$	MPa $\sqrt{\text{m}}$	1.098 800 E + 00	hp (electric)	kW	7.460 000 E − 01	in.^{3}/min	m^3/s	2.731 177 E − 07
Heat content			**Power density**					
Btu/lb	kJ/kg	2.326 000 E + 00	W/in.^{2}	W/m^2	1.550 003 E + 03			
cal/g	kJ/kg	4.186 800 E + 00	**Pressure (fluid)**					
Heat input			atm (standard)	Pa	1.013 250 E + 05			
J/in.	J/m	3.937 008 E + 01	bar	Pa	1.000 000 E + 05			
kJ/in.	kJ/m	3.937 008 E + 01	in. Hg (32 °F)	Pa	3.386 380 E + 03			
			in. Hg (60 °F)	Pa	3.376 850 E + 03			
			lbf/in.^{2} (psi)	Pa	6.894 757 E + 03			
			torr (mm Hg, 0 °C)	Pa	1.333 220 E + 02			

(a) kg × 10^3 = 1 metric ton or 1 megagram (Mg). (b) Preferred expression is 10^{-6}/K or 10^{-6}/F as length units are unnecessary.

Thermodynamics*

First Law of Thermodynamics— Conservation of Energy

One of the fundamental laws of thermodynamics is that energy cannot be created or destroyed. Energy can only be converted from one form to another, for example, from potential to kinetic energy, heat to work, internal energy to kinetic energy, or transferred from one mass to another as heat or work.

Closed-System Formulation. The general energy equation is $Q - W = E_f = E_i$, where:

Q = heat transfer across the system boundary as the system proceeds from state i to state f. Heat is an energy transfer across a system boundary due to a temperature difference. Heat transfer to the system is positive, and heat transfer from the system is negative

W = work transfer across the system boundary as the system proceeds from state i to state f. Work is an energy transfer across a system boundary due to a property difference other than a temperature difference. Work done by the system is positive, and work done on the system is negative. The criterion for work being done by a system is that the system could pass through the same sequence of states as the system passes from state i to state f, and the sole effect on the surroundings could be the raising of a weight. The criterion for work done on the system is obtained by interchanging the system and the surroundings and applying the criterion to the new system.

$E_f - E_i$ = energy change of the system and includes changes in internal, kinetic, potential, electrical, chemical, nuclear, or surface energies

The reduced energy equation is

$Q - W' = U_f - U_i$, where:

W' = same as defined previously, except W' includes only work transfers due to a deformation of the control volume

$U_f - U_i$ = internal energy change of the system boundary

The reduced energy equation is a form of the principle of conservation of energy that is written by an observer on a coordinate system fixed on the center of mass of the closed system. This observer does not record any change in the kinetic and potential energies of the mass or any translation work on the mass.

Control-Volume Formulation. The general energy equation for a control volume is:

$$Q - \dot{W} = \frac{d}{dt}\int_{V}\rho\left(u + \frac{V^2}{2} + gz\right)dV$$
$$+ \int_{A}\rho\left(h + \frac{V^2}{2}\right) + gz\right)(\vec{u}\cdot\hat{\mathbf{n}})dA \quad \text{(Eq 1)}$$

where:

$\dot{Q}$ = rate of heat transfer across the control surface (or open system boundary) without a mass transfer and due to a temperature difference. Heat transfer to the material inside the control volume is positive, and heat transfer from the material inside the control volume is negative

$\dot{W}$ = rate of work (or power) transfer across the control surface without a mass transfer and due to a property difference other than a temperature difference. Power transfer from the material in the control volume is positive, and power transfer to the material in the control volume is negative. The criterion for a power transfer by or on the material in the control volume is similar to the work criterion for a closed system (or systems). $\dot{W}$ would include the rate of work done on or by the material in the control volume also when the control volume changes size and shape.

t = time

ρ = mass density

u = internal energy per unit mass. u could be interpreted as including surface energy and electromagnetic energies, if involved in the process.

V = velocity measured relative to a convenient frame of reference

g = local acceleration of gravity

z = elevation above some reference datum

V = volume of control volume

h = enthalpy = $u + p/\rho$, where p is the absolute pressure

$\vec{u}$ = velocity of fluid leaving or entering the control volume and measured relative to the local control surface

$\hat{\mathbf{n}}$ = unit outward normal vector at the local control surface

A = control volume surface area

The energy-equation for uniform properties throughout the control volume and at each inlet and outlet is given by:

$$\dot{Q} - \dot{W} = \frac{d}{dt}\left[m_{CV}\left(u + \frac{V^2}{2} + gz\right)_{CV}\right] +$$
$$\sum_{out}\dot{m}\left(h + \frac{V^2}{2} + gz\right) - \sum_{in}\dot{m}\left(h + \frac{V^2}{2} + g\right)$$
$$\text{(Eq 2)}$$

where:

m_{CV} = mass in the control volume at any instant of time t

$\dot{m}$ = mass flow rate into or out of the control volume at each inlet or outlet

The time derivative d/dt in Eq 2 is used if the control volume is moving relative to the coordinate system, and $\partial/\partial t$ is used if the control volume is fixed relative to the coordinate system. When a control volume is fixed relative to the coordinate system and steady-state conditions and uniform properties exist:

$$\dot{Q} - \dot{W} = \sum_{out}\dot{m}\left(h + \frac{1}{2}V^2 + gz\right)$$
$$- \sum_{in}\dot{m}\left(h + \frac{1}{2}V^2 + gz\right) \quad \text{(Eq 3)}$$

The values $\dot{Q}, \dot{W}$, and $\dot{m}$ do not change with time.

Gibbs' equations may be applied to a process followed by a closed system or may be applied to a fixed amount of mass as it flows through a control volume. They may be applied regardless of whether the fixed amount of mass undergoes a reversible or irreversible process as long as the end points of such a process are true thermodynamic states; that is, properties are uniform throughout the fixed amount of mass. Gibbs' equations can then be applied between

---

*Adapted from *ASM Handbook of Engineering Mathematics*, American Society for Metals, 1983

the inlet and outlet of a control volume for steady-state conditions. Gibbs' equations are:

$$Tds = du + pdv \quad Tds = dh - vdp$$

where T is absolute temperature, s is entropy per unit mass, p is absolute pressure, and v is specific volume.

Thermodynamic Processes for a Control Volume. A steady-state process is characterized by the condition where the mass flow rates and all properties of each point in the control volume do not change with time, that is, $\partial/\partial t = 0$. Consider a control volume moving relative to one coordinate system (XYZ) and fixed relative to another (xyz), as shown in Fig. 1. The xyz observer may see a steady-state condition in his coordinate system but the XYZ observer may not. For steady-state conditions seen by the xyz observer, use Eq 1 (with $d/dt = \partial/\partial t = 0$) if nonuniform properties exist throughout the control volume and at the inlet and outlet flow areas, and Eq 3 if uniform properties exist throughout the control volume and at the inlet and outlet flow areas.

A steady-flow process is characterized by the condition where $\dot{Q}, \dot{W}$ (Eq 1), the mass flow rate, and all properties at each point are constant in time. For example, consider an air compressor with constant $\dot{Q}$ and $\dot{W}$. An observer on the compressor casing does not observe a steady-flow condition because there are flow-rate pulsations at the inlet and outlet. If the control surface is sufficiently far from the compressor, so that flow and property fluctuations have been damped out, the process is considered a steady-flow process.

Many authors consider the terms *steady flow* and *steady state* as being equivalent and the same as a steady-state, steady-flow process described as follows. Use Eq 1 if properties are nonuniform and Eq 2 if properties are uniform. The steady-state, steady-flow process is characterized by:

- A control volume that is fixed relative to the coordinate system
- A condition where mass flow rate as well as all properties at each point of each inlet and outlet flow area are constant in time
- A condition where mass flow rate and all properties at each point inside the control volume are constant in time
- A condition where $\dot{Q}$ and $\dot{W}$ are constant in time

Equation 1 (with $\partial/\partial t = 0$) is used if properties are nonuniform, and Eq 3 is used if properties are uniform.

The uniform state, uniform flow process is characterized by:

- A control volume that is fixed relative to the coordinate system
- A condition where properties of the mass flow rate are uniform over each inlet and outlet flow area are constant in time, that is, the mass flow rates may change with time but the properties will not
- A condition where properties of the mass inside the control volume are uniform throughout the control volume at each instant of time but may change with time

Equation 2 is used under these conditions.

Work Equations

Closed System—Work of Deformation. The general equation for the work of deformation, W', for a reversible process is given by:

$$W' = \int_i^f pd\Psi$$

where the system proceeds from the initial state, i, to the final state, f. The work of deformation, w', per unit mass is:

$$w' = \int_i^f pdv$$

where p is the absolute pressure used to calculate the work done by the system or the gage pressure used to calculate the net useful work, that is, system work minus work done on the atmosphere (Fig. 2).

The following equations in cases 1 through 6 are used to calculate the work of deformation for specific reversible processes.

Case 1: Constant pressure (isobaric), reversible processes:

$$W' = p(\Psi_f - \Psi_i) \quad w' = p(v_f - v_i)$$

where p is the absolute pressure or gage pressure, depending on whether system work or net work is being calculated.

Case 2: Constant pressure (isobaric), reversible process for an ideal gas:

$$W' = mR(T_f - T_i) \quad w' = R(T_f - T_i)$$

Case 3: Constant volume (isometric), reversible process:

$$W' = 0 \quad w' = 0$$

Case 4: Constant temperature (isothermal), reversible process for an ideal gas:

$$W' = mRT_i \ln \frac{v_f}{v_i}$$

$$W' = mRT_i \ln \frac{\Psi_f}{\Psi_i}$$

$$W' = mRT_i \ln \frac{p_i}{p_f}$$

$$w' = \frac{W'}{m}$$

Case 5: Adiabatic or adiabatic and reversible process (isentropic process):

$$W' = m(u_i - u_f) \quad w' = (u_i - u_f)$$

An adiabatic and reversible process is isentropic, but an isentropic process need not necessarily be both adiabatic and reversible.

Case 6: Polytropic, reversible process:

$$W' = \frac{(p_f\Psi_f - p_i\Psi_i)}{1 - n} \quad w' = \frac{p_fv_f - p_iv_i}{1 - n}$$

Control Volume, Steady-State Conditions (Constant g). The general equations for work

Fig. 1 Relative control-volume movement. Control volume fixed relative to xyz, which moves with velocity $\vec{V}$ relative to XYZ

$$W'_{sys} = \int_i^f p_{abs} \, dV$$

$$W'_{net} = \int_i^f p_{gage} \, dv$$

$$W'_{net} = \int_i^f p_{gage} \, d\Psi$$

$$W'_{svs} = \int_i^f p_{abs} \, d\Psi$$
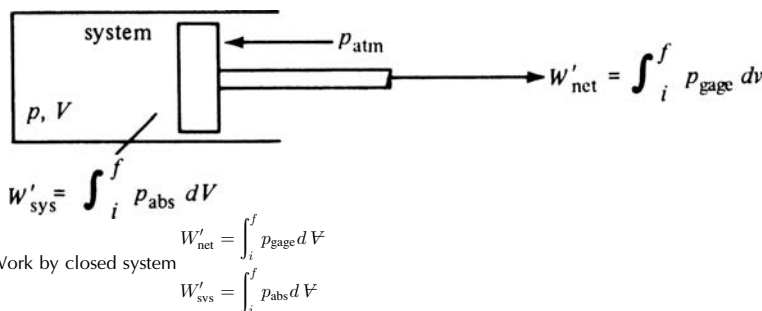
Fig. 2 Work by closed system

for a reversible process under steady-state conditions are:

$$w = -\left[\int_1^2 v\, dp + \frac{V_2^2 - V_1^2}{2} + g(z_2 - z_1)\right]$$
$$\dot{W} = \dot{m}w$$

where p is the absolute pressure. Point 1 represents the inlet state and point 2 the outlet state.

The following equations in cases 1 through 7 are used to calculate the control-volume work or power for specific reversible processes.

Case 1: Constant pressure (isobaric), reversible process:

$$w = \frac{V_1^2 - V_2^2}{2} + g(z_1 - z_2)$$
$$\dot{W} = \dot{m}w$$

Case 2: Constant specific volume (isometric), reversible process:

$$w = v(p_1 - p_2) + \frac{V_1^2 - V_2^2}{2} + g(z_1 - z_2)$$
$$\dot{W} = \dot{m}w$$

Case 3: Constant specific volume (isometric), reversible process for an ideal gas:

$$w = R(T_1 - T_2) + \frac{V_1^2 - V_3^2}{2} + g(z_1 - z_2)$$
$$\dot{W} = \dot{m}w$$

Case 4: Constant temperature (isothermal), reversible process for an ideal gas:

$$w = RT_1 \ln\frac{p_1}{p_2} + \frac{V_1^2 - V_2^2}{2} + g(z_1 - z_2)$$
$$w = RT_1 \ln\frac{v_2}{v_1} + \frac{V_1^2 - V_2^2}{2} + g(z_1 - z_2)$$
$$\dot{W} = \dot{m}w$$

Case 5: Adiabatic or adiabatic and reversible process (isentropic process):

$$\dot{W} = -\dot{m}\left[(h_2 - h_1) + \frac{V_2^2 - V_1^2}{2} + g(z_2 - z_1)\right]$$
$$w = -\left[(h_2 - h_1) + \frac{V_2^2 - V_1^2}{2} + g(z_2 - z_1)\right]$$

Case 6: Polytropic, reversible process:

$$w = \frac{-n}{n-1}(p_2 v_2 - p_1 v_1) + \frac{V_1^2 - V_2^2}{2} + g(z_1 - z_2)$$
$$\dot{W} = \dot{m}w$$

Case 7: Adiabatic, reversible process for an ideal gas with constant specific heat:

$$w = \frac{kRT_1}{k-1}\left[1 - \left(\frac{p_2}{p_1}\right)^{\frac{k-1}{k}}\right] + \frac{V_1^2 - V_2^2}{2} + g(z_1 - z_2)$$

Heat-Transfer Equations

Closed System. The general equation for the heat transfer, Q, for a reversible process is given by:

$$Q = \int_i^f T\, dS$$

where Q is the total heat transfer as the system proceeds from the initial state, i, to the final state, f; T is the absolute temperature (K or °R); and S is the total entropy of the mass, m. The heat transfer per unit mass, q, is:

$$q = \int_i^f T\, ds$$

Also:

$$Q = mq = m\int_i^f T\, ds$$

where s is the entropy per unit mass.

The following equations in cases 1 through 7 are used to calculate the heat transfer for specific reversible process.

Case 1: Constant pressure (isobaric), reversible process:

$$Q = m(h_f - h_i) \quad q = h_f - h_i$$

Case 2: Constant pressure (isobaric), reversible process for an ideal gas with constant specific heat:

$$Q = mC_p \ln\frac{T_f}{T_i} \quad q = C_p \ln\frac{T_f}{T_i}$$

Case 3: Constant volume (isometric), reversible process:

$$Q = m(u_f - u_i) \quad q = u_f - u_i$$

Case 4: Constant volume (isometric), reversible process for an ideal gas with constant specific heat:

$$Q = mC_v \ln\frac{T_f}{T_i} \quad q = C_v \ln\frac{T_f}{T_i}$$

Case 5: Constant temperature (isothermal), reversible process for an ideal gas:

$$Q = T_i(S_f - S_i) = W$$
$$Q = mRT_i \ln\frac{v_f}{v_i}$$
$$Q = mRT_i \ln\frac{V_f}{V_i}$$
$$Q = mRT_i \ln\frac{p_i}{p_f}$$
$$q = \frac{Q}{m} = w$$

Case 6: Adiabatic or adiabatic and reversible process (isentropic process):

$$Q = 0 \quad q = 0$$

Case 7: Polytropic, reversible process:

$$q = h_f - h_i - \frac{n}{n-1}(p_f v_f - p_i v_i)$$
$$q = u_f - u_i - \frac{1}{n-1}(p_f v_f - p_i v_i)$$
$$Q = mq$$

Control-Volume, Steady-State Conditions (Constant q). The general equations for heat transfer for a reversible process are given by:

$$q = \int_1^2 T ds \quad \dot{Q} = \dot{m}q$$

The following equations in cases 1 through 7 are used to calculate the heat transfer for specific reversible processes.

Case 1: Constant pressure (isobaric), reversible process:

$$q = (h_2 - h_1) \quad \dot{Q} = \dot{m}q$$

Case 2: Constant specific volume (isometric), reversible process:

$$q = (u_2 - u_1) \quad \dot{Q} = \dot{m}q$$

Case 3: Constant specific volume (isometric), reversible process for an ideal gas with constant specific heat:

$$q = C_v(T_2 - T_1) \quad \dot{Q} = \dot{m}q$$

Case 4: Constant temperature (isothermal), reversible process for an ideal gas:

$$q = h_2 - h_1 - RT_1 \ln\frac{p_1}{p_2}$$
$$q = h_2 - h_1 - RT_1 \ln\frac{v_2}{v_1}$$
$$q = T_1(s_2 - s_1)$$
$$\dot{Q} = \dot{m}q$$

Case 5: Constant temperature (isothermal), reversible process for an ideal gas with constant specific heat:

$$q = C_p(T_2 - T_1) - RT_1 \ln\frac{p_1}{p_2}$$
$$q = C_p(T_2 - T_1) - RT_1 \ln\frac{v_2}{v_1}$$
$$q = T_1(s_2 - s_1)$$
$$\dot{Q} = \dot{m}q$$

Case 6: Adiabatic or adiabatic and reversible process (isentropic process):

$$q = 0 \quad \dot{Q} = 0$$

Case 7: Polytropic, reversible process:

$$q = (h_2 - h_1) - \frac{n}{n-1}(p_2 v_2 - p_1 v_1)$$
$$\dot{Q} = \dot{m}q$$

Property Relations

Entropy Change Equations—Closed System or Control Volume, Steady-State Conditions. For a reversible process, the general equations for the entropy change of a pure substance are given by:

$$s_2 - s_1 = \int_1^2 \frac{du}{T} + \int_1^2 \frac{p}{T}dv$$
$$s_2 - s_1 = \int_1^2 \frac{dh}{T} - \int_1^2 \frac{v}{T}dp$$
$$s_2 - s_1 = \int_1^2 C_v \frac{dT}{T} + \int_1^2 \left(\frac{\partial p}{\partial T}\right)_v dv$$
$$s_2 - s_1 = \int_1^2 C_p \frac{dT}{T} - \int_1^2 \left(\frac{\partial v}{\partial T}\right)_p dp$$

where T is absolute temperature, v is specific volume, C_v is specific heat at constant volume, and C_p is specific heat at constant pressure. For a closed system, 1 represents the initial state and 2 the final state. For a control volume, 1 represents the inlet and 2 the outlet.

The following equations in cases 1 through 9 are used to calculate the entropy change of a pure substance for specific reversible processes.

Case 1: Reversible process for an ideal gas:

$$s_2 - s_1 = \int_1^2 C_v \frac{dT}{T} + R \ln \frac{v_2}{v_1}$$

$$s_2 - s_1 = \int_1^2 C_p \frac{dT}{T} - R \ln \frac{p_2}{p_1}$$

where R is the specific gas constant.

Case 2: Reversible process for an ideal gas with constant specific heats:

$$s_2 - s_1 = C_v \ln \frac{T_2}{T_1} + R \ln \frac{v_2}{v_1}$$

$$s_2 - s_1 = C_p \ln \frac{T_2}{T_1} - R \ln \frac{p_2}{p_1}$$

$$s_2 - s_1 = C_p \ln \frac{v_2}{v_1} + C_v \ln \frac{p_2}{p_1}$$

Case 3: Constant pressure (isobaric), reversible process:

$$s_2 - s_1 = \int_1^2 \frac{dh}{T}$$

Case 4: Constant pressure, reversible process for an ideal gas:

$$s_2 - s_1 = \int_1^2 C_p \frac{dT}{T}$$

Case 5: Constant specific volume (isometric), reversible process:

$$s_2 - s_1 = \int_1^2 \frac{du}{T}$$

Case 6: Constant specific volume, reversible process for an ideal gas:

$$s_2 - s_1 = \int_1^2 C_v \frac{dT}{T}$$

Case 7: Adiabatic, reversible process:

$$s_2 - s_1 = 0$$

Case 8: Isentropic process:

$$s_2 - s_1 = 0$$

Case 9: Mixing of ideal gases at constant temperature and constant total pressure, p:

$$\Delta S = \sum_{\ell=1}^{N} n_i \overline{R} \ln \frac{p_i}{p}$$

where:

p_i = partial pressure of component i
p = total pressure of mixture
$\sum_{\ell=1}^{N} X_i p_i$ = initial pressure of each component before mixing

N = number of components
n_i = moles of component i
X_i = mole fraction of component i

Process Equations. The following equations in cases 1 through 6 relate the properties of a pure substance during specific reversible processes.

Case 1: Constant pressure (isobaric), reversible process for an ideal gas:

$$\frac{v}{T} = \text{Constant} = \frac{R}{p}$$

Case 2: Constant specific volume (isometric), reversible process for an ideal gas:

$$\frac{p}{T} = \text{Constant} = \frac{R}{v}$$

Case 3: Constant temperature (isothermal), reversible process for an ideal gas:

$$pv = \text{Constant} = RT_1$$

Case 4: Polytropic, reversible process:

$$pv^n = \text{Constant} \quad n = \text{Constant} = \frac{\ln\left(\frac{p_2}{p_1}\right)}{\ln\left(\frac{v_1}{v_2}\right)}$$

Case 5: Polytropic, reversible process for an ideal gas:

$$pv^n = \text{Constant}$$
$$Tp^{\frac{1-n}{n}} = \text{Constant}$$
$$Tv^{n-1} = \text{Constant}$$
$$n = \text{Constant} = \frac{\ln\left(\frac{p_2}{p_1}\right)}{\ln\left(\frac{v_1}{v_2}\right)}$$

Case 6: Isentropic process for an ideal gas with constant specific heats:

$$pv^k = \text{Constant}$$
$$Tp^{\frac{1-k}{k}} = \text{Constant}$$
$$Tv^{k-1} = \text{Constant}$$
$$k = \frac{C_p}{C_v} = \text{Constant}$$

Equations of State. The following equations relate the properties of a pure substance (Table 1 and 2). In all equations:

T = absolute temperature T_c = critical point temperature
p = absolute pressure p_c = critical point pressure
v = mass specific volume v_c = critical point specific volume
$\overline{v}$ = molar specific volume
R = specific gas constant = $\frac{\overline{R}}{M}$ M = molecular weight
$\overline{R}$ = universal gas constant = $0.08205 \frac{\ell \cdot \text{atm}}{\text{gmol} \cdot \text{K}} = 8.315 \frac{\text{J}}{\text{gmol} \cdot \text{K}}$
$\overline{R} = 0.08315 \frac{\text{bar} \cdot \text{m}^3}{\text{kgmol} \cdot \text{K}} = 1545 \frac{\text{ft} \cdot \text{lbf}}{\text{lbmol} \cdot {}^\circ \text{R}} = 1.986 \frac{\text{kcal}}{\text{kmol} \cdot \text{K}}$
$\overline{R} = 0.730 \frac{\text{atm} \cdot \text{ft}^3}{\text{lbmol} \cdot {}^\circ \text{R}} = 10.73 \frac{\text{psia} \cdot \text{ft}^3}{\text{kmol} \cdot \text{K}}$

The ideal gas law (perfect gas law) is expressed as:

$$pv = RT$$
$$p\overline{v} = \overline{R}T$$

The van der Waals equation is:

$$\left(p + \frac{a}{v^2}\right)(v - b) = RT$$

where:

$$a = \frac{27}{64} \frac{R^2 T_c^2}{p_c}$$
$$b = \frac{RT_c}{8p_c}$$
$$v_c = \frac{3RT_c}{8p_c}$$

The constants a and b are found in Tables 1 and 2. The compressibility factor equation is expressed as:

$$Z = \frac{pv}{RT}$$

where Z is the compressibility factor, p_R is the reduced pressure $= \frac{p}{p_c}$; T_R is the reduced temperature $= \frac{T}{T_c}$, and v_R is the reduced specific volume $= \frac{vp_c}{RT_c}$.

Table 1   Critical properties and van der Waals constants, SI units

Substance	T_c, K	p_c, bar	v_c, m³/kmol	$Z_c = \frac{P_c v_c}{RT_c}$	van der Waals a, bar $\left(\frac{\text{m}^3}{\text{kmol}}\right)^2$	b, $\frac{\text{m}^3}{\text{kmol}}$
Acetylene (C_2H_2)	309	62.8	0.112	0.274	4.410	0.0510
Air (equivalent)	133	37.7	0.0829	0.284	1.358	0.0364
Ammonia (NH_3)	406	112.8	0.0723	0.242	4.233	0.0373
Benzene (C_6H_6)	562	49.3	0.256	0.274	18.63	0.1181
n-Butane (C_4H_{10})	425.2	38.0	0.257	0.274	13.80	0.1196
Carbon dioxide (CO_2)	304.2	73.9	0.0941	0.276	3.643	0.0427
Carbon monoxide (CO)	133	35.0	0.0928	0.294	1.463	0.0394
Refrigerant 12 (CCl_2F_2)	385	40.1	0.214	0.270	10.78	0.0998
Ethane (C_2H_6)	305.4	48.8	0.221	0.273	5.575	0.0650
Ethylene (C_2H_4)	283	51.2	0.143	0.284	4.563	0.0574
Helium (He)	5.2	2.3	0.0579	0.300	0.0341	0.0234
Hydrogen (H_2)	33.2	13.0	0.0648	0.304	0.247	0.0265
Methane (CH_4)	190.7	46.4	0.0991	0.290	2.285	0.0427
Nitrogen (N_2)	126.2	33.9	0.0897	0.291	1.361	0.0385
Oxygen (O_2)	154.4	50.5	0.0741	0.290	1.369	0.0315
Propane (C_3H_8)	370	42.7	0.195	0.276	9.315	0.0900
Sulfur dioxide (SO_2)	431	78.7	0.124	0.268	6.837	0.0568
Water (H_2O)	647.3	220.9	0.0558	0.230	5.507	0.0304

Modified critical temperature, T_c', and pressure, p_c', are defined by:

$$p_c' = p_c + 8 \quad T_c' = T_c + 8$$

and are used for hydrogen and helium, where T_c is in K and p_c is in atmospheres. Figure 3 shows the generalized compressibility factor in graphic form.

The virial equations of state are expressed as:

$$pv = a + bp + cp^2 + dp^3 + \cdots$$
$$pv = a\left(1 + \frac{b}{v} + \frac{c}{v^2} + \frac{d}{v^3} + \cdots\right)$$

where:

$a = a(T)$
$b = b(T)$
$c = c(T)$ and so on

All sets of coefficients ($a, b, c, d, \ldots$) are virial coefficients and can be found from experimental data or statistical mechanics.

The Berthelot equation of state is expressed as:

$$p = \frac{RT}{v-b} - \frac{a}{Tv^2}$$

The constants a and b must be found from experimental data for a particular fluid.

The Dieterici equation of state is expressed as:

$$p = \frac{RT}{v-b}e^{-a/RTv}$$

The constants a and b must be found from experimental data for a particular fluid.

The Clausius equation of state is expressed as:

$$p(\mathcal{V} - b') = mRT$$

where:

$$b' = \frac{16}{3}N\pi r^3$$

r = one-half the center-to-center distance of two colliding molecules
N = number of molecules in volume V

The Redlich-Kwong equation of state is expressed as:

$$p = \frac{RT}{v-b} - \frac{a}{T^{1/2}v(v+b)}$$

where:

$a = 0.4275\frac{R^2 T_c^{2.5}}{p_c}$
$b = 0.0867\frac{RT_c}{p_c} = 0.26v_c$

Table 3 gives numerical values of Redlich-Kwong constants for selected gases.

The Benedict-Webb-Rubin equation of state, which is especially applicable to hydrocarbons, is expressed as:

$$p = \frac{RT}{v}\left(B_0 RT - A_0 - \frac{C_0}{T^2}\right)\frac{1}{v^2}$$
$$+ \frac{bRT - a}{v^3} + \frac{a\alpha}{v^6} + \frac{c}{v^3 T^2}\left(1 + \frac{\gamma}{v^2}\right)e^{(-\gamma/v^2)}$$

Table 2   Critical properties and van der Waals constants, English engineering units

Substance	T_c, °R	p_c, atm	$\bar{v}_c$, ft³/(lbmol)	$Z_c = \frac{p_c v_c}{RT_c}$	van der Waals a, atm [ft³/(lbmol)]²	b, ft³/(lbmol)
Acetylene ($C_2 H_2$)	556	62	1.80	0.274	1121	0.818
Air (equivalent)	239	37.2	1.33	0.284	345.2	0.585
Ammonia (NH_3)	730	111.3	1.16	0.242	1076	0.598
Benzene ($C_6 H_6$)	1013	48.7	4.11	0.274	4736	1.896
n-Butane ($C_4 H_{10}$)	765	37.5	4.13	0.274	3508	1.919
Carbon dioxide (CO_2)	548	72.9	1.51	0.276	926	0.686
Carbon monoxide (CO)	239	34.5	1.49	0.294	372	0.632
Refrgerant 12 ($CCl_2 F_2$)	693	39.6	3.43	0.270	2718	1.595
Ethane ($C_2 H_6$)	549	48.2	3.55	0.273	1410	1.041
Ethylene ($C_2 H_4$)	510	50.5	2.29	0.284	1158	0.922
Helium (He)	9.33	2.26	0.93	0.300	8.66	0.376
n-Heptane ($C_7 H_{16}$)	972	27	6.86	0.26	7866	3.298
Hydrogen (H_2)	59.8	12.8	1.04	0.304	62.8	0.426
Methane (CH_4)	344	45.8	1.59	0.290	581	0.685
Methyl chloride ($CH_3 Cl$)	749	65.8	2.29	0.276	1917	1.040
Nitrogen (N_2)	227	33.5	1.44	0.291	346	0.618
Nonane ($C_9 H_{20}$)	1071	22.86	8.86	0.250		
n-Octane ($C_8 H_{18}$)	1025	24.6	7.82	0.258	9601	3.76
Oxygen (O_2)	278	49.8	1.19	0.290	348	0.506
Propane ($C_3 H_8$)	666	42.1	3.13	0.276	2368	1.445
Sulfur dioxide (SO_2)	775	77.7	1.99	0.268	1738	0.911
Water ($H_2 O$)	1165	218.2	0.896	0.230	1400	0.488
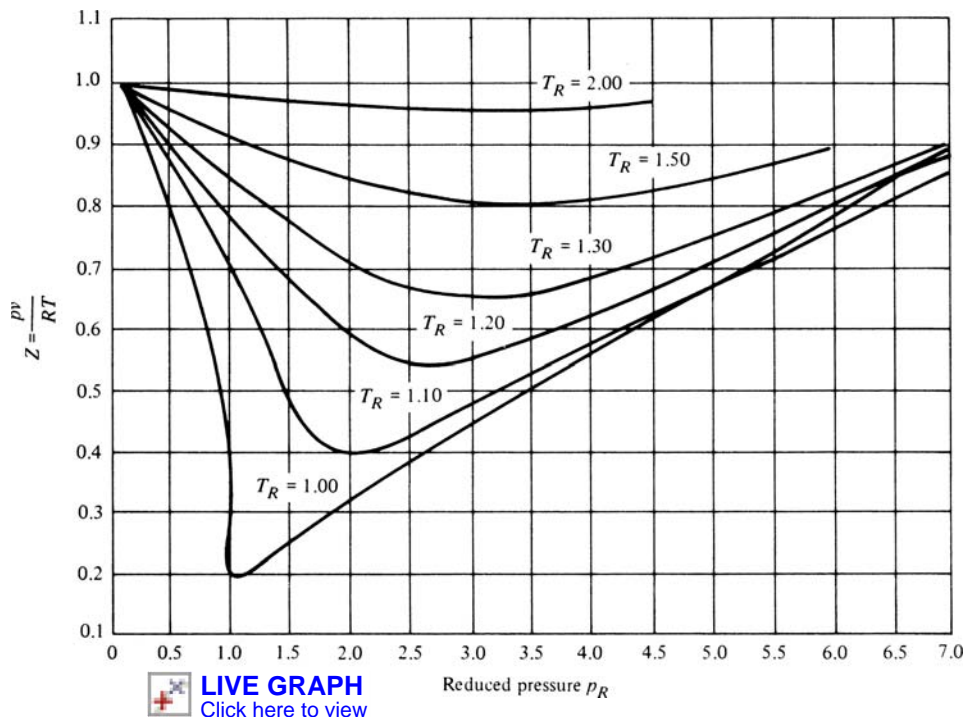
Fig. 3 Generalized compressibility factor, Z

Table 3   Redlich-Kwong constants

Substance	Constants for p in bar(s), v in m³/(kmol), and T in K a	b	Constants for p in atm, v in ft³/(lbmol), and T in °R a	b
Carbon dioxide, CO_2	64.64	0.02969	21,970	0.4757
Carbon monoxide, CO	17.26	0.02743	5,870	0.4395
Methane, CH_4	32.19	0.02969	10,930	0.4757
Nitrogen, N_2	15.59	0.02681	5,300	0.4294
Oxygen, O_2	17.38	0.02199	5,900	0.3522
Propane, $C_3 H_8$	183.07	0.06269	62,190	1.0040
Refrigerant 12, $CCl_2 F_2$	214.03	0.06913	72,710	1.1080
Sulfur dioxide, SO_2	144.49	0.03939		
Water, $H_2 O$	142.64	0.02110	48,460	0.3381

The constants in the equation are defined in Table 4 for various gases.

Specific Internal Energy Change—Closed System or Steady-State Control Volume. The following equations are used to calculate the internal energy change of a pure substance. The term *specific* refers to the property per unit mass or per unit mole. For a closed system, 1 represents the initial state and 2 the final state. For a control volume, 1 represents the inlet and 2 the outlet.

For a reversible process:

$$u_2 - u_1 = \int_1^2 C_v dT + \int_1^2 \left[T\left(\frac{\partial p}{\partial T}\right)_v - p \right] dv$$

For a constant-volume (isometric) reversible process:

$$u_2 - u_1 = \int_1^2 C_v dT$$

For a reversible process for an ideal gas:

$$u_2 - u_1 = \int_1^2 C_u dT$$

where $C_v = C_v (T)$.

Specific Enthalpy Change—Closed System or Steady-State Control Volume. The following equations are used to calculate the enthalpy change of a pure substance. The term *specific* refers to the property per unit mass or per unit mole. For a closed system, 1 represents the initial state and 2 the final state. For a control volume, 1 represents the inlet and 2 the outlet.

For a reversible process:

$$h_2 - h_1 = \int_1^2 C_p dT + \left[v - T\left(\frac{\partial v}{\partial T}\right)_p \right] dp$$

For a constant pressure (isobaric), reversible process:

$$h_2 - h_1 = \int_1^2 C_p dT$$

For a reversible process for an ideal gas:

$$h_2 - h_1 = \int_1^2 C_p dT$$

where $C_p = C_p (T)$.

Second Law of Thermodynamics

The second law states that it is not possible to construct a device whose sole effect is the extraction of heat from a heat source and the conversion of this heat completely into mechanical work. The second law means that some energy is always "lost"; no heat engine is 100% efficient. The following equations give the cycle thermodynamic efficiency for various devices operating under steady-state conditions.

Cycle Thermodynamic Efficiency and Coefficient of Performance. The efficiency, η, of a Carnot heat engine, shown schematically in Fig. 4, is given by:

$$\eta = \frac{W}{Q_H} = 1 - \frac{T_L}{T_H} \qquad \frac{Q_L}{Q_H} = \frac{T_L}{T_H}$$

The coefficient of performance (COP) of a Carnot refrigerator, shown schematically in Fig. 5, is given by:

$$(COP)_{Heating} = \frac{Q_H}{W} = \frac{T_H}{T_H - T_L}$$

$$(COP)_{Cooling} = \frac{Q_L}{W} = \frac{T_L}{T_H - T_L}$$

$$\frac{Q_L}{Q_H} = \frac{T_L}{T_H}$$

where:

T_H = absolute temperature of high-temperature reservoir
Q_H = heat transfer from high-temperature reservoir

W = work done by engine
Q_L = heat transfer to low-temperature reservoir
T_L = absolute temperature of low-temperature reservoir

The efficiency, η, of an air standard Carnot cycle (Fig. 6) is given by:

$$\eta = 1 - \frac{1}{(CR)^{k-1}} = 1 - \frac{T_1}{T_3} = 1 - \frac{T_4}{T_2}$$

where, for all air standard cycles:

CR = compression ratio
$k = C_p/C_v$, constant C_p and C_v
T = absolute temperature
p = absolute pressure

The efficiency, η, of an air standard Otto cycle (Fig. 7) is given by:

$$\eta = 1 - \frac{1}{(CR)^{k-1}} = 1 - \frac{T_1}{T_2}$$

The efficiency, η, of an air standard Diesel cycle (Fig. 8) is given by:

$$\eta = 1 - \frac{(T_4 - T_1)}{k(T_3 - T_2)} = 1 - \frac{1}{k(CR)^{k-1}} \left[\frac{\left(\frac{v_3}{v_2}\right)^k - 1}{\left(\frac{v_3}{v_2}\right) - 1} \right]$$

The efficiency, η, of an air standard Stirling cycle (Fig. 9) is given by:

Fig. 4 Carnot heat engine

Fig. 5 Carnot refrigeration

Table 4 Benedict-Webb-Rubin

	p in bar(s), v in m³/(kmol), T in K				
Constants	*n*-Butane, C_4H_{10}	Carbon dioxide, CO_2	Carbon monoxide, CO	Methane, CH_4	Nitrogen, N_2
a	1.9068	0.1386	0.0371	0.0500	0.0254
A_0	10.216	2.7730	1.3587	1.8791	1.0673
b	0.039998	0.007210	0.002632	0.003380	0.002328
B_0	0.12436	0.04991	0.05454	0.04260	0.04074
c	3.205×10^5	1.511×10^4	1.054×10^3	2.578×10^3	7.379×10^2
C_0	1.006×10^6	1.404×10^5	8.673×10^3	2.286×10^4	8.164×10^3
α	1.101×10^{-3}	8.470×10^{-5}	1.350×10^{-4}	1.244×10^{-4}	1.272×10^{-4}
γ	0.0340	0.00539	0.0060	0.0060	0.0053
	p in atm, v in ft³/(lb mol), T in °R				
Constants	*n*-Butane, C_4H_{10}	Carbon dioxide, CO_2	Carbon monoxide, CO	Methane, CH_4	Nitrogen, N_2
a	7747	563.1	150.7	203.1	103.2
A_0	2590	703.0	344.5	476.4	270.6
b	10.27	1.852	0.676	0.868	0.598
B_0	1.993	0.7998	0.8740	0.6827	0.6529
c	4.219×10^9	1.989×10^8	1.387×10^7	3.393×10^7	9.713×10^6
C_0	8.263×10^8	1.153×10^8	7.124×10^6	1.878×10^7	6.706×10^6
α	4.531	0.3486	0.5556	0.5120	0.5235
γ	8.732	1.384	1.541	1.541	1.361

Fig. 6 Air standard Carnot cycle

Fig. 7 Air standard Otto cycle

Fig. 8 Air standard Diesel cycle

Fig. 9 Air standard Stirling cycle

$$\eta = 1 - \frac{T_2}{T_3} = 1 - \frac{T_1}{T_4}$$

The efficiency, η, of an air standard Ericsson cycle (Fig. 10) is given by:

$$\eta = 1 - \frac{T_2}{T_3} = 1 - \frac{T_1}{T_4}$$

The efficiency, η, of an air standard Atkinson cycle (Fig. 11) is given by:

$$\eta = 1 - \frac{k}{(\text{CR})^{k-1}} \left[\frac{\left(\frac{p_3}{p_2}\right)^{1/k} - 1}{\left(\frac{p_3}{p_2}\right) - 1} \right]$$

The efficiency, η, of an air standard closed Brayton cycle (Fig. 12) is given by:

$$\eta = 1 - \frac{1}{\left(\frac{p_2}{p_1}\right)^{\frac{k-1}{k}}}$$

The efficiency, η, of an air standard open Brayton cycle (Fig. 13) is given by:

$$\eta = 1 - \frac{T_4 - T_1}{T_3 - T_2}$$

The efficiency, η, of an air standard open Brayton cycle with regeneration (Fig. 14) is given by:

$$\eta = \frac{(T_3 - T_2) - (T_4 - T_1)}{T_3 - T_5}$$

Process Efficiency. The following equations give the efficiencies of various processes for steady-state conditions. Nozzle efficiency, η, is given by:

$$\eta_{\text{nozzle}} = \frac{V_2^2}{V'^2_2}$$
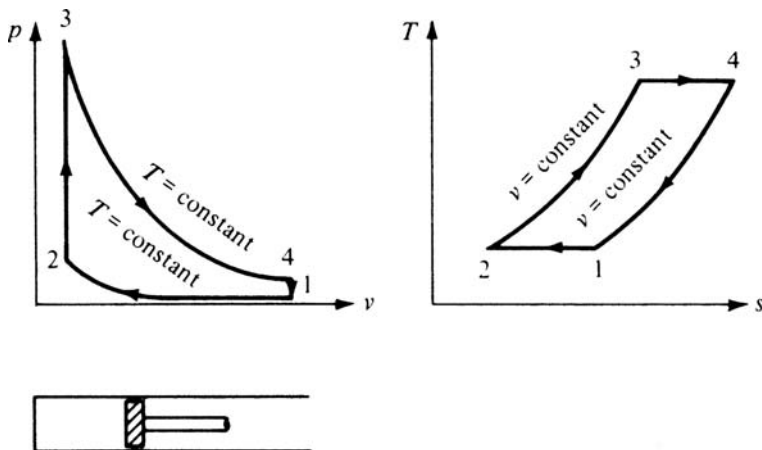
Fig. 10 Air standard Ericsson cycle

Fig. 11 Air standard Atkinson cycle

for a system shown in Fig. 15, where V'_2 is the ideal outlet velocity for the condition where the outlet entropy, s_2, equals the inlet entropy, s_1, and the actual exit (or outlet) pressure is p_2.

Turbine isentropic efficiency, η, is given by:

$$\eta_{\text{turbine}} = \frac{\text{(Actual work per unit mass)}}{\left(\begin{array}{l}\text{Isentropic work per unit}\\\text{mass for actual inlet}\\\text{conditions and actual}\\\text{exit pressure}\end{array}\right)}$$

Compressor isentropic efficiency, η, is given by:

$$\eta_{\text{compressor}} = \frac{\left(\begin{array}{l}\text{Isentropic work per unit}\\\text{mass for actual inlet con-}\\\text{ditions and actual exit}\\\text{pressure}\end{array}\right)}{\text{(Actual work per unit mass)}}$$

Compressor isothermal efficiency, η, is given by:

$$\eta_{\text{compressor}} = \frac{\left(\begin{array}{l}\text{Isothermal work per unit}\\\text{mass for actual inlet con-}\\\text{ditions and actual exit}\\\text{pressure}\end{array}\right)}{\text{(Actual work per unit mass)}}$$

Boiler or burner combustion efficiency, η, is given by:

$$\eta_{\text{boiler}} = \frac{\left(\begin{array}{l}\text{Rate of heat transferred}\\\text{to working fluid(s)}\end{array}\right)}{\left(\begin{array}{l}\text{Fuel mass flow rate times}\\\text{fuel heating value}\end{array}\right)} \quad \text{(Eq 4)}$$

In Eq 4, either the higher heating value or the lower heating value could be used; by common practice, the higher heating value is normally used.

Increase of Entropy Principle. Unlike energy, entropy is not conserved. For example, the entropy of an adiabatic, closed system, if left to itself, could increase in time as the system approached equilibrium. Mathematically, the entropy of this system would attain its maximum possible value when equilibrium is reached. One

interpretation of entropy is that it represents the degree of "disorder" of the molecules of a system. The system is in disorder if the various molecules of a multicomponent system are uniformly distributed throughout the system. The system is most ordered if each type of molecule is separated from the others. Another interpretation of entropy is that it is a measure of the fraction of the system energy that cannot be converted into work. The larger entropy means that less work can be obtained from the energy of the system.

In the following equations, the inequality (>0) applies to an irreversible process, while the equality ($=0$) applies to a reversible process. The subscript i refers to the initial state and f to the final state. S represents the entropy of either a system or its surroundings.

For a closed system and its surroundings:

$$dS_{\text{SYSTEM}} + dS_{\text{SURR}} \geq 0$$
$$S_{f,\text{SYSTEM}} - S_{i,\text{SYSTEM}} + S_{f,\text{SURR}} - S_{i,\text{SURR}} \geq 0$$

For an adiabatic, closed system:

$$dS \geq 0$$
$$S_f - S_i \geq 0$$

For an isolated system:

$$dS > 0$$
$$S_f - S_i \geq 0$$

For a control volume:

$$\frac{dS_{\text{cv}}}{dt} + \sum_{\text{out}} \dot{m}s - \sum_{\text{in}} \dot{m}s \geq \int_{\substack{A \text{ control} \\ \text{surface}}} \frac{\dot{q}}{T} dA$$

where:

T = absolute temperature
$\dot{q}$ = heat-transfer rate per unit area of the control surface and added to or lost from the control volume
cv = control volume

For a control volume and its surroundings:

$$\frac{dS_{\text{cv}}}{dt} + \frac{dS_{\text{SURR}}}{dt} \geq 0$$

Reversible Power, Irreversibility, and Availability. Irreversibility is the difference between the maximum power obtainable (reversible power) and the actual power obtainable from the control volume; availability is a mathematical definition and represents the reversible work per unit mass when a unit mass proceeds from the state of interest to a state of zero velocity, ambient temperature and ambient pressure.

Point i represents the initial state, f the final state, 1 the inlet state, and 2 the outlet state. Uniform properties are assumed over each flow area and throughout the control volume. Equations 5 through 10 apply to a control volume. The corresponding equations applicable to a closed system are obtained by setting each mass flow rate equal to zero.
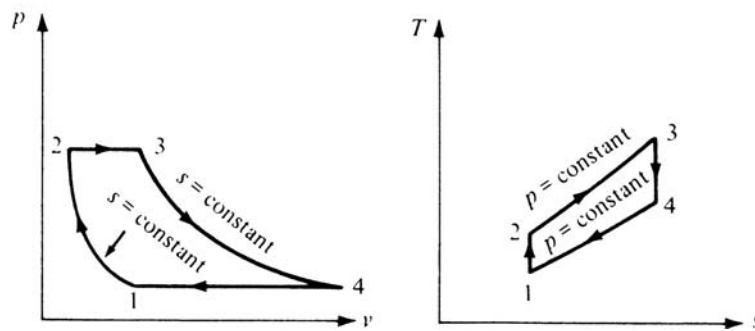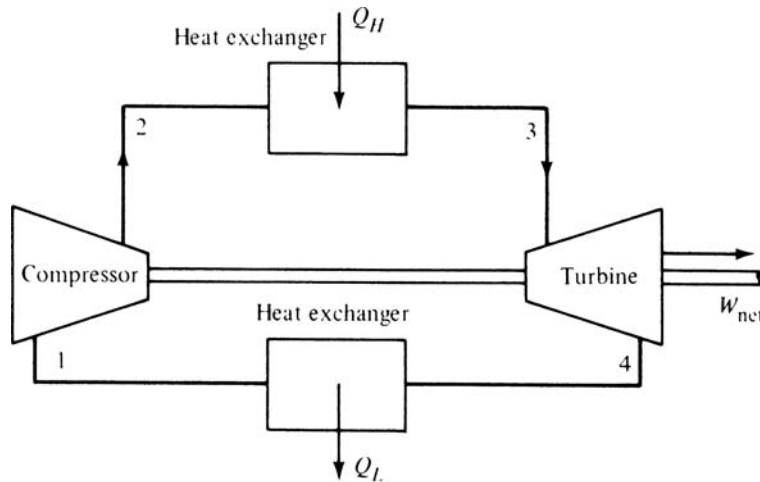
Fig. 12 Air standard closed Brayton cycle

Reversible power for a steady-state control volume is given by:

$$\dot{W}_{rev} = \sum_{inlet} \dot{m}_1 \left(h_1 - T_0 s_1 + \frac{V_1^2}{2} + g z_1 \right)$$
$$- \sum_{outlet} \dot{m}_2 \left(h_2 - T_0 s_2 + \frac{V_2^2}{2} + g z_2 \right)$$
$$+ m_i \left(u_i - T_0 s_i + \frac{V_i^2}{2} + g z_i \right)$$
$$- m_f \left(u_f - T_0 s_f + \frac{V_f^2}{2} + g z_f \right)$$

(Eq 5)

Irreversibility for a steady-state process is given by:

$$\dot{I} = \dot{W}_{rev} - \dot{W}_{actual}$$

(Eq 6)

where $\dot{W}_{actual}$ is obtained from the first law of thermodynamics applied to the actual control volume; also by:

$$\dot{I} = T_0 \left(\sum_{outlet} \dot{m}_2 s_2 - \sum_{inlet} \dot{m}_1 s_1 + m_f s_f - m_i s_i \right) - \dot{Q}_{cv}$$

(Eq 7)

where $\dot{Q}_{cv}$ is the rate of heat transfer into the control volume of interest; also by:

$$\dot{I} = T_0 \left(\frac{dS_{control\ volume}}{dt} + \frac{dS_{surroundings}}{dt} \right)$$

(Eq 8)

and by:

$$\dot{I} = T_0 \left(\frac{dS_{control\ volume}}{dt} + \sum \dot{m}_2 s_2 - \Sigma \dot{m}_1 s_1 - \frac{Q_{cv}}{T_0} \right)$$

(Eq 9)

where T_0 is the ambient absolute temperature.

Availability for a steady-state control volume is given by:

$$\psi = \left(h - T_0 s + \frac{V^2}{2} + g z \right) - (h_0 - T_0 s_0 + g z_0)$$

(Eq 10)

where the zero subscript represents the state in which the fluid is in equilibrium with the surroundings (at ambient temperature, T_0, and pressure, P_0).

Mixtures and Solutions

The following equations apply to a system consisting of several homogeneous phases.

Chemical Reactions. For chemical reactions of ideal gas mixtures at temperature T and total pressure p, the reaction equation is:

$$v_A A + v_B B \rightleftharpoons v_C C + v_D D$$
(Reactants)   (Products)
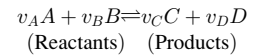
The law of mass action is expressed as:

$$K = \frac{p_C^{v_C} p_D^{v_D}}{p_A^{v_A} p_B^{v_B}} = \frac{x_C^{v_C} x_D^{v_D}}{x_A^{v_A} x_D^{v_B}} p^{v_C + v_D - v_A - v_B}$$
$$= \frac{v_A \bar{g}_A^0 + v_B \bar{g}_B^0 - v_C \bar{g}_C^0 - v_D \bar{g}_D^0}{\bar{R} T}$$

where:

K = equilibrium constant
$\bar{g}_i^0$ = Gibbs function per unit mole of component i at absolute temperature T and 1.0 atm pressure
p = absolute pressure of mixture in atmospheres
x_i = mole fraction of component i
v_i = coefficient in reaction equation ($i = A, B, C,$ or D)
R = universal gas constant
T = absolute temperature

Table 5 gives equilibrium constants for various simple reactions.

The van't Hoff equation is given as:

$$\frac{d(\ln K)}{d(1/T)} = \frac{v_A \bar{h}_A^0 + v_B \bar{h}_B^0 - v_C \bar{h}_C^0 - v_D \bar{h}_D^0}{\bar{R}}$$

where $\bar{h}_i^0$ is the enthalpy per unit mole of component i at temperature T, and superscript zero means 1 atm pressure.

Equilibrium of a heterogeneous system is expressed by the Gibbs-Duhem equation:

$$S\,dT + \Psi dp = \sum_{i=1}^{N} n_i d\bar{\mu}_i$$

where:

$\bar{\mu}_i$ = chemical potential of component i at the partial pressure p_i of component i and the temperature T of the mixture
S = entropy of mixture
T = absolute temperature
Ψ = volume of mixture
p = total pressure of mixture
n_i = moles of component i
x_i = mole fraction of component i
N = number of components

Ideal Gas Equations. For all equations:

N = number of components
x_i = mole fraction of component i
n_i = moles of component i
n = total moles of mixture
y_i = mass fraction of component i
m_i = mass of component i
m = mass of mixture
M_i = molecular weight of component i
M = mixture molecular weight
R = universal gas constant
R = average specific gas constant
R_i = specific gas constant of component i
Ψ_i = partial volume of component i
Ψ = volume of mixture
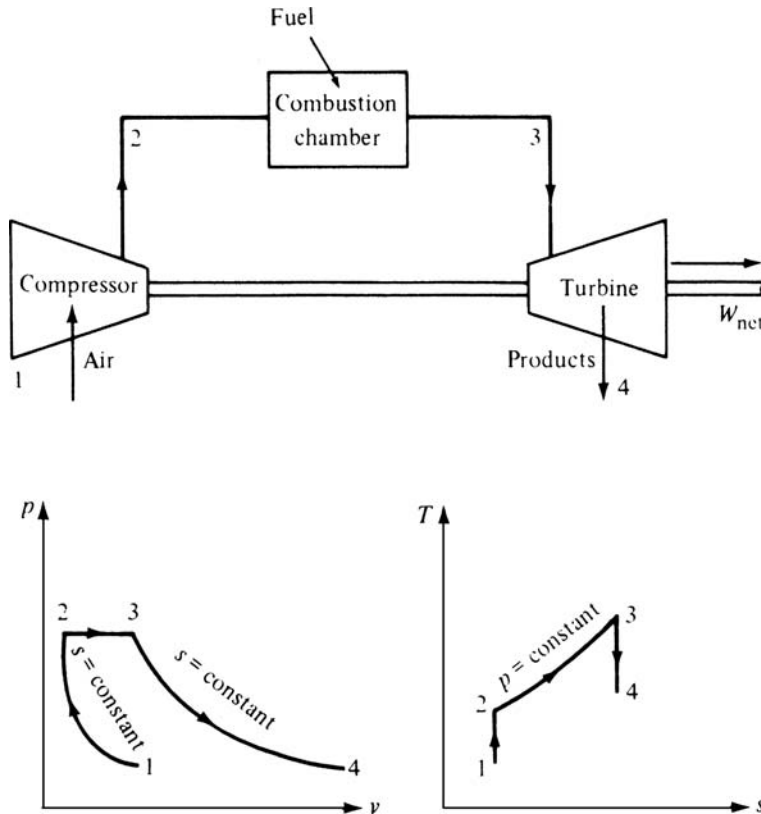
Fig. 13 Air standard Brayton cycle
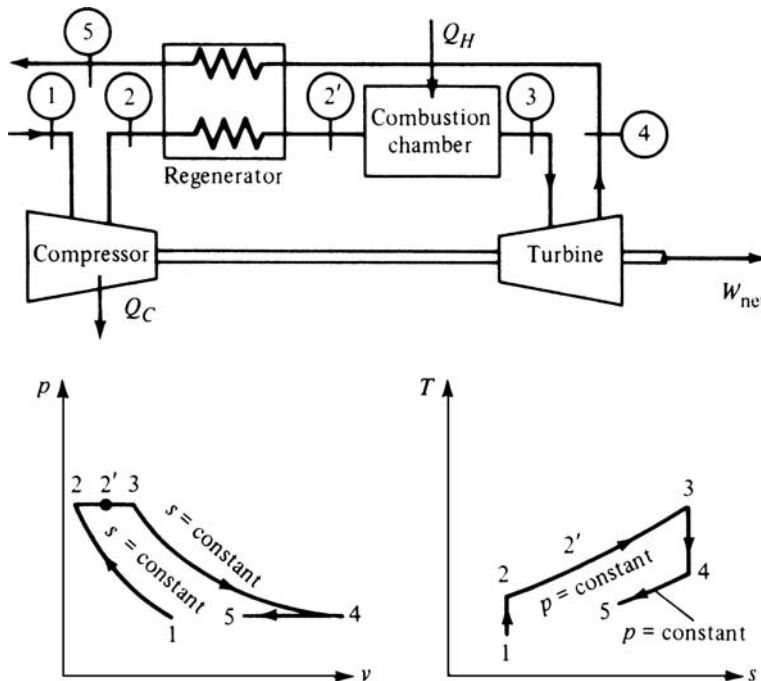
Fig. 14 Air standard open Brayton cycle

p_i = partial pressure of component i
p = total pressure of mixture

Dalton's law of partial pressures is expressed as:

$$p = \sum_{i=1}^{N} p_i$$

Amagat's law of partial volumes is expressed as:

$$V = \sum_{i=1}^{N} V_i$$

Mixture molecular weight is found by:

$$M = \frac{\sum_{i=1}^{N} m_i}{\sum_{i=1}^{N} n_i} = \sum_{i=1}^{N} x_i M_i$$

Mixture average specific gas constant is found by:

$$R = \sum_{i=1}^{N} y_i R_i$$

Mixture average specific heat (either C_P or C_v) is found by:

$$C_{\text{ave}} = \sum_{i=1}^{N} y_i C_i \quad \overline{C}_{\text{ave}} = \sum_{i=1}^{N} x_i \overline{C}_i$$

where C has dimensions of energy/(mass • temp), and $\overline{C}$ has dimensions of energy/(mole • temp) (Table 6).

Mixture average internal energy is found by:

$$u_{\text{ave}} = \sum_{i=1}^{N} y_i u_i \quad \overline{u}_{\text{ave}} = \sum_{i=1}^{N} x_i \overline{u}_i$$

and the mixture average enthalpy is found by:

$$h_{\text{ave}} = \sum_{i=1}^{N} y_i h_i \quad \overline{h}_{\text{ave}} = \sum_{i=1}^{N} x_i \overline{h}_i$$

where u and h have dimensions of energy/mass, and $\overline{u}$ and $\overline{h}$ have dimensions of energy/mole. Mixture average entropy is found by:

$$s_{\text{ave}} = \sum_{i=1}^{N} y_i S_i \quad \overline{S}_{\text{ave}} = \sum_{i=1}^{N} x_i \overline{S}_i$$

where s has dimensions of energy/(mass • temp), and $\overline{s}$ has dimensions of energy/(mole • temp).

The Clausius-Clapeyron equation applies to a pure substance with one phase in equilibrium with another. Absolute temperature and pressure are denoted by T and p. The general form for the Clausius-Clapeyron equation is:
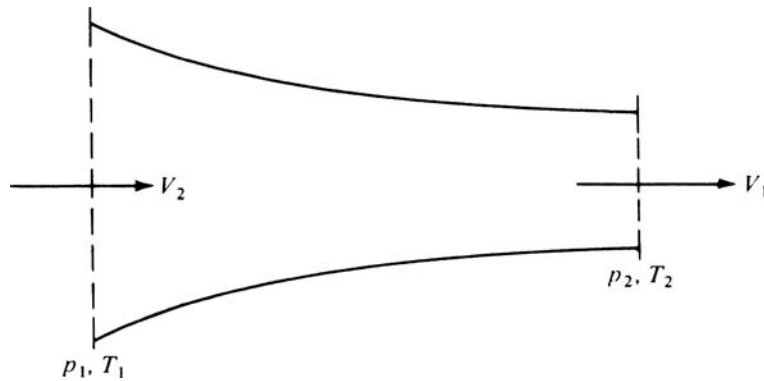
Fig. 15 Nozzle schematic

$$\left(\frac{dp}{dT}\right)_{\text{sat}} = \frac{s'' - s'}{v'' - v'} = \frac{h'' - h'}{T(v'' - v')}$$

The integrated form for the solid-vapor region, which assumes ideal gas vapor and h_{ig} = constant, is:

$$\ln\frac{p_B}{p_A} = \frac{h_{ig}}{R}\left(\frac{1}{T_A} - \frac{1}{T_B}\right)$$

The integrated form for the liquid-vapor region, which assumes ideal gas vapor and h_{fg} = constant, is:

$$\ln\frac{p_B}{p_A} = \frac{h_{fg}}{R}\left(\frac{1}{T_A} - \frac{1}{T_B}\right)$$

The integrated form for the solid-vapor region, which assumes constant h_{if} and v_{if}, is:

$$p_B - p_A = \frac{h_{if}}{v_{if}}\ln\frac{T_B}{T_A}$$

where:

p = absolute pressure
T = absolute temperature
h = enthalpy (per unit mass)
$h_{if} = h_f - h_i$
$h_{ig} = h_g - h_i$
$h_{fg} = h_g - h_f$
R = specific gas constant
v = specific volume
$v_{if} = v_f - v_i$
s = entropy (per unit mass)
Subscripts: i = ice
f = liquid
g = vapor (gas)

Psychrometrics—Dry Air and Water Vapor Considered Ideal Gases. Psychrometrics is the study of the thermodynamic properties of moist air and the analysis of conditions and processes involving moist air.

Humidity ratio, W, is expressed as:

$$W = \frac{\text{Mass of water vapor in air sample}}{\text{Mass of dry air in same air sample}}$$
$$= 0.622\frac{p_w}{p - p_w}$$

where p is the total absolute pressure, and p_w is the partial absolute pressure of water vapor.

Relative humidity, ϕ, is expressed as:

$$\phi = \frac{\text{Partial pressure of water vapor in air}}{\left(\begin{array}{c}\text{Saturation pressure of water vapor}\\\text{at dry bulb temperature}\end{array}\right)}$$
$$= \frac{p_w}{(p_{w,\text{sat}})_{\text{dbt}}}$$

where $(p_{w,\text{sat}})_{\text{dbt}}$ is the saturation pressure of water vapor at the dry bulb temperature.

Specific humidity, q, is expressed as:

$$q = \frac{\text{Mass of water vapor in air sample}}{\text{Total mass of same air sample}} = \frac{W}{1 + W}$$

Degree of saturation, μ, is expressed as:

Table 5   Natural logarithms of the equilibrium constant K (for 1.0 atm pressure)

Temp, K	$H_2 \rightleftharpoons 2H$	$O_2 \rightleftharpoons 2O$	$N_2 \rightleftharpoons 2N$	$2H_2O \rightleftharpoons 2H_2 + O_2$
298	−164.005	−186.975	−367.480	−184.416
500	−92.827	−105.630	−213.372	−105.382
1000	−39.803	−45.150	−99.127	−46.326
1200	−30.874	−35.005	−80.011	−36.364
1400	−24.463	−27.742	−66.329	−29.218
1600	−19.637	−22.285	−56.055	−23.842
1800	−15.866	−18.030	−48.051	−19.652
2000	−12.840	−14.622	−41.645	−16.290
2200	−10.353	−11.827	−36.391	−13.536
2400	−8.276	−9.497	−32.011	−11.238
2600	−6.517	−7.521	−28.304	−9.296
2800	−5.002	−5.826	−25.117	−7.624
3000	−3.685	−4.357	−22.359	−6.172
3200	−2.534	−3.072	−19.937	−4.902
3400	−1.516	−1.935	−17.800	−3.782
3600	−0.609	−0.926	−15.898	−2.784
3800	0.202	−0.019	−14.199	−1.890
4000	0.934	0.796	−12.660	−1.084
4500	2.486	2.513	−9.414	0.624
5000	3.725	3.895	−6.807	1.992
5500	4.743	5.023	−4.666	3.120
6000	5.590	5.963	−2.865	4.064

Temp, K	$2H_2O \rightleftharpoons H_2 + 2OH$	$2CO_2 \rightleftharpoons 2CO + O_2$	$N_2 + O_2 \rightleftharpoons 2NO$
298	−212.416	−207.524	−70.104
500	−120.562	−115.232	−40.590
1000	−52.068	−47.058	−18.776
1200	−40.566	−35.742	−15.138
1400	−32.198	−27.684	−12.540
1600	−26.132	−21.660	−10.588
1800	−21.314	−16.994	−9.072
2000	−17.456	−13.270	−7.862
2200	−14.296	−10.240	−6.866
2400	−11.664	−7.720	−6.038
2600	−9.438	−5.602	−5.342
2800	−7.526	−3.788	−4.742
3000	−5.874	−2.222	−4.228
3200	−4.424	−0.858	−3.776
3400	−3.152	0.338	−3.380
3600	−2.176	1.402	−3.026
3800	−1.002	2.352	−2.712
4000	−0.088	3.198	−2.432
4500	1.840	4.980	−1.842
5000	3.378	6.394	−1.372
5500	4.636	7.542	−0.994
6000	5.686	8.490	−0.682

Table 6 Constant pressure specific heats at zero pressure

Gas	$\overline{C}_p, \dfrac{kcal}{kmol \cdot K}, \dfrac{Btu}{lb\ mol \cdot °R}$	Range, °R	Max error, %	$\overline{C}_p, \dfrac{kJ}{kmol \cdot K}$	Range, K	Max error, %
N_2	$9.3355 - 122.56\ \theta^{-1.5} + 256.38\ \theta^{-2} - 196.08\ \theta^{-3}$	540–6300	0.43	$39.060 - 512.79\ \theta^{-1.5} + 1072.7\ \theta^{-2} - 820.40\ \theta^{-3}$	300–3500	0.43
O_2	$8.9465 + 4.8044 \times 10^{-3}\ \theta^{1.5} - 42.679\ \theta^{-1.5} + 56.615\ \theta^{-2}$	540–6300	0.30	$37.432 + 0.020102\ \theta^{1.5} - 178.57\ \theta^{-1.5} + 236.88\ \theta^{-2}$	300–3500	0.30
H_2	$13.505 - 167.96\ \theta^{-0.75} + 278.44\ \theta^{-1} - 134.01\ \theta^{-1.5}$	540–6300	0.60	$56.505 - 702.74\ \theta^{-0.75} + 1165.0\ \theta^{-1} - 560.70\ \theta^{-1.5}$	300–3500	0.60
CO	$16.526 - 0.16841\ \theta^{0.75} - 47.985\ \theta^{-0.5} + 42.246\ \theta^{-0.75}$	540–6300	0.42	$69.145 - 0.70463\ \theta^{0.75} - 200.77\ \theta^{-0.5} + 176.76\ \theta^{-0.75}$	300–3500	0.42
OH	$19.490 - 14.185\ \theta^{0.25} + 4.1418\ \theta^{0.75} - 1.0196\ \theta$	540–6300	0.43	$81.546 - 59.350\ \theta^{0.25} + 17.329\ \theta^{0.75} - 4.2660\ \theta$	300–3500	0.43
NO	$14.169 - 0.40861\ \theta^{0.5} - 16.877\ \theta^{-0.5} + 17.899\ \theta^{-1.5}$	540–6300	0.34	$59.283 - 1.7096\ \theta^{0.5} - 70.613\ \theta^{-0.5} + 74.889\ \theta^{-1.5}$	300–3500	0.34
H_2O	$34.190 - 43.868\ \theta^{0.25} + 19.778\ \theta^{0.5} - 0.88407\ \theta$	540–6300	0.43	$143.05 - 183.54\ \theta^{0.25} + 82.751\ \theta^{0.5} - 3.6989\ \theta$	300–3500	0.43
CO_2	$-0.89286 + 7.2967\ \theta^{0.5} - 0.98074\ \theta + 5.7835 \times 10^{-3}\ \theta^2$	540–6300	0.19	$-3.7357 + 30.529\ \theta^{0.5} - 4.1034\ \theta + 0.024198\ \theta^2$	300–3500	0.19
NO_2	$11.005 + 51.650\ \theta^{-0.5} - 86.916\ \theta^{-0.75} + 55.580\ \theta^{-2}$	540–6300	0.26	$46.045 + 216.10\ \theta^{-0.5} - 363.66\ \theta^{-0.75} + 232.550\ \theta^{-2}$	300–3500	0.26
CH_4	$-160.82 + 105.10\ \theta^{0.25} - 5.9452\ \theta^{0.75} + 77.408\ \theta^{-0.5}$	540–3600	0.15	$-672.87 + 439.74\ \theta^{0.25} - 24.875\ \theta^{0.75} + 323.88\ \theta^{-0.5}$	300–2000	0.15
C_2H_4	$-22.800 + 29.433\ \theta^{0.5} - 8.5185\ \theta^{0.75} + 43.683\ \theta^{-3}$	540–3600	0.07	$-95.395 + 123.15\ \theta^{0.5} - 35.641\ \theta^{0.75} + 182.77\ \theta^{-3}$	300–2000	0.07
C_2H_6	$1.648 + 4.124\ \theta - 0.153\ \theta^2 + 1.74 \times 10^{-3}\ \theta^3$	540–2700	0.83	$6.895 + 17.26\ \theta - 0.6402\ \theta^2 + 0.00728\ \theta^3$	300–1500	0.83
C_3H_8	$-0.966 + 7.279\ \theta - 0.3755\ \theta^2 + 7.58 \times 10^{-3}\ \theta^3$	540–2700	0.40	$-4.042 + 30.46\ \theta - 1.571\ \theta^2 + 0.03171\ \theta^3$	300–1500	0.40
C_4H_{10}	$0.945 + 8.873\ \theta - 0.438\ \theta^2 + 8.36 \times 10^{-3}\ \theta^3$	540–2700	0.54	$3.954 + 37.12\ \theta - 1.833\ \theta^2 + 0.03498\ \theta^3$	300–1500	0.54

$$\theta = \frac{T(\text{Kelvin})}{100} = \frac{T(\text{Rankine})}{180} \qquad\qquad \theta = \frac{T(\text{Kelvin})}{100}$$

The equations in Table 6 are used to calculate the specific heats, $\overline{C}_p$, of ideal gases. The specific heat at constant volume, $\overline{C}_v$, is calculated by $\overline{C}_v = \overline{C}_p - \overline{R}$, where $\overline{R}$ is the universal gas constant.

$$\mu = \frac{\text{Humidity ratio}}{\left(\begin{array}{c}\text{Humidity ratio of saturated}\\ \text{air at the dry bulb temperature}\end{array}\right)}$$

$$= \left[\frac{p_w}{p - p_w}\right]\left[\frac{p - (p_{w,\text{sat}})_{\text{dbt}}}{(p_{w,\text{sat}})_{\text{dbt}}}\right]$$

Specific volume is expressed by:

$$v = \frac{\text{Volume of sample}}{\text{Mass of dry air in sample}}$$
$$= \frac{R_{\text{da}}T}{p}(1 + 1.608W)$$

where R_{da} is the specific gas constant for dry air.

Enthalpy is expressed as:

$$h = h_{\text{da}} + Wh_g$$

where:

W = humidity ratio (lbm water/lbm dry air)
h_{da} = enthalpy of dry air per unit mass of dry air

$h_{\text{da}} = 0$ at 0 °F
$h_{\text{da}} = 0.240T$ (Btu/lbm dry air with T in °F)
h_g = enthalpy of saturated water vapor at the dry bulb temperature (Btu/lbm water)
$h_g = 0$ for liquid water at 32.0 °F and 14.696 psia
$h_g = 1061 + 0.444T$ (Btu/lbm water with T in °F)

The relationship of humidity ratio and dry bulb temperature, T, and wet bulb temperature, T':

$$W = \frac{(1093 - 0.556T')W' - 0.240(T - T')}{1093 + 0.444T - T'}$$

with T and T' in °F, W and W' in lbm water/lbm dry air, and W' = humidity ratio of saturated air at T'.

SELECTED REFERENCES

- F.H. Crawford, and W.D. van Vorst, *Thermodynamics for Engineers*, Harcourt, Brace, and World, Inc., New York, 1968
- F.F. Huang, *Engineering Thermodynamics*, Macmillan, New York, 1976
- J.F. Lee, F.W. Sears, and D.L. Turcotte, *Statistical Thermodynamics*, Addison-Wesley, Reading, MA, 1963
- W.C. Reynolds, *Thermodynamics*, McGraw-Hill, New York, 1968
- M.A. Saad, *Thermodynamics for Engineers*, Prentice Hall, Englewood Cliffs, NJ, 1966
- R.E. Sonntag, and G.J. Van Wylen, *Introduction to Thermodynamics: Classical and Statistical*, 2nd ed., John Wiley & Sons, New York, 1982
- G.J. Van Wylen and R.E. Sonntag, *Fundamentals of Classical Thermodynamics*, 2nd ed., John Wiley & Sons, New York, 1973
- K. Wark, *Thermodynamics*, McGraw-Hill, New York, 1977
- B.D. Wood, *Applications of Thermodynamics*, 2nd ed., Addison-Wesley, Reading, MA, 1982

Heat-Transfer Equations*

HEAT TRANSFER is energy in transit as a result of temperature difference. Whenever a temperature difference exists in a medium or between media, heat transfer must take place. Mechanical engineers deal with problems of heat transfer in the fields of internal combustion engines, power generation, refrigeration, heating, and ventilation. Metallurgical and ceramic engineers must control temperatures accurately during the heat treatment of various metals and ceramics to achieve the desired properties of the heat treated material. In all fields of engineering, problems are confronted concerning the most effective transmission of heat or the protection of a construction most effectively against heat losses or gains.

Three recognized modes of heat transfer are conduction, convection, and thermal radiation. They differ entirely in physical mechanism and governing laws. In conduction, heat flows from a high-temperature region to a region of lower temperature within a medium or between different media in direct physical contact. Conduction is the only mechanism by which heat can flow in opaque solids. Convective heat transfer occurs when a fluid is mixed due to gross movement of fluid mass. Energy transfer from one fluid particle to another remains a process of conduction, but the energy may be transferred from one point in space to another by the fluid displacement. Therefore, convection is the most important mechanism of energy transfer between a solid and a liquid or a gas. Thermal radiation is a process of heat transfer from one body to another by electromagnetic wave motion. The transmission of radiant energy does not require a carrying medium.

The basic law of heat conduction, proposed by Fourier, states that the rate of heat flow in a material is proportional to the area perpendicular to heat flow and the temperature gradient at that section. The proportion constant is called the thermal conductivity of the material. In general, the thermal conductivity varies with temperature. Some industrial problems involving heat conduction are the annealing of castings, the vulcanizing of rubber, and the heating or cooling of the walls of buildings, furnaces, and ovens.

The prediction of the rates at which heat is transferred by convection from a solid surface by an ambient fluid involves a thorough understanding of principles of heat conduction, fluid dynamics, and boundary-layer theory. The complexities involved in an analytical approach may be combined in terms of a single parameter by introduction of Newton's law of cooling, which states that the heat flux due to convection is the product of film coefficient (or heat-transfer coefficient) and the temperature difference between the surface and ambient fluid. The heat-transfer coefficient is not a material property but a complex function of system geometry, fluid flow, and fluid properties. Heat transfer by convection occurs on walls of rooms, on the outside of cold and warm pipes, and between surfaces and fluid of all types of heat exchangers. Boiling and condensation are also classified as convection heat transfer.

All bodies emit radiant energy. An ideal radiator, called a black body, emits radiant energy at a rate proportional to the fourth power of the absolute temperature of the body. This is known as Stefan-Boltzmann law of thermal radiation. Other surfaces, such as a glossy painted surface, do not radiate as much energy as the black body. To describe the gray nature of such a surface, the term *emissivity* is introduced. Emissivity is the ratio of radiation of the gray surface to an ideal black surface. The importance of radiation becomes intensified at high temperature levels. Consequently, radiation contributes substantially to heat transfer in boilers, furnaces, combustion chambers, nuclear power plants, and rocket nozzles. Solar radiation plays an important part in the design of heating and ventilating systems. Radiation can also be of importance in some instances when free convection is present, even though the temperature levels are not elevated.

Most industrial problems dealing with heat exchange do not involve a single mechanism of heat transfer but include a combination of two or more. In a steam condenser, heat transfer occurs through a series of conduction and convection processes between the condensing steam and the cooling water. In a furnace of large steam generators, heat is transferred simultaneously by radiation, convection, and conduction.

Heat Conduction

Conduction heat transfer involves energy transfer by physical interaction between adjacent molecules of substances at different temperatures. Heat flow within a homogeneous material is analogous to the flow of electricity in a conductor. The quantity flowing per unit of time is proportional to the:

- Conductivity of the material
- Area of the conductor perpendicular to the path
- Potential temperature gradient
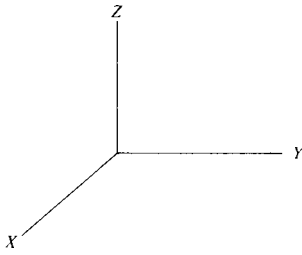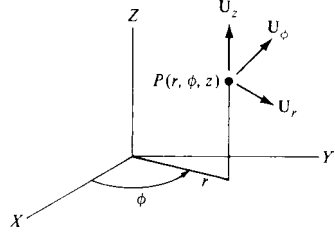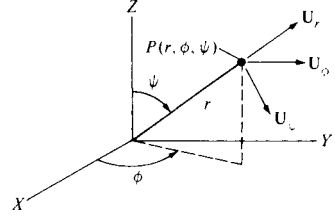
If the heat flowing into a body is exactly equal to the heat flowing out, a steady-state condition exists. This condition refers only to cases where the temperature at any given point within the body is independent of the time. If the inflow and outflow of heat are not equal, and the temperature at a given point and the heat content of the body vary with time, then heat is said to be flowing in an unsteady state.

Fourier Law, Heat Conduction Equations, and Boundary Conditions

The basic law of heat conduction was first presented by a French physicist, Fourier, and is referred to as Fourier's law of heat conduction. Table 1 presents the general form related to various coordinate systems. From Fourier's law and the first law of thermodynamics, the differential equations for heat conduction in solids can be derived (Table 2). To determine temperature distribution due to heat conduction, boundary conditions must be known. Frequently encountered boundary conditions in conduction are summarized in Table 3.

---

*Adapted from *ASM Handbook of Engineering, Mathematics,* American Society for Metals, 1983

Table 1 Fourier's law of heat conduction

General form	$\mathbf{q}(\mathbf{v}, \theta) = -k \, \nabla \, (\mathbf{r}, \theta)$

Cartesian system

$q_x = -k\dfrac{\partial T}{\partial x}$

$q_y = -k\dfrac{\partial T}{\partial y}$

$q_z = -k\dfrac{\partial T}{\partial z}$

Cylindrical system
$x = r \cos \phi$
$y = r \sin \phi$

$z = z$

$q_r = -k\dfrac{\partial T}{\partial r}$

$q_\phi = -k\dfrac{1}{r}\dfrac{\partial T}{\partial \phi}$

$q_z = -k\dfrac{\partial T}{\partial z}$

Spherical system
$x = r \sin \psi \sin \phi$

$y = r \sin \psi \cos \phi$

$z = r \cos \psi$

$q_r = -k\dfrac{\partial T}{\partial r}$

$q_\phi = -k\dfrac{1}{r}\dfrac{\partial T}{\partial \phi}$

$q_\phi = -k\dfrac{1}{r\sin\phi}\dfrac{\partial T}{\partial \psi}$

Table 2 Differential equations for heat conduction in solids

General form with variable thermal properties	$\rho c_p \dfrac{\partial T(\mathbf{r}, \theta)}{\partial \theta} = \nabla \cdot [k\nabla T(\mathbf{r}, \theta)] + g(\mathbf{r}, \theta)$
General form with constant thermal properties	$\dfrac{1}{a}\dfrac{\partial T(\mathbf{r}, \theta)}{\partial \theta} = \nabla^2 T(\mathbf{r}, \theta) + \dfrac{g(\mathbf{r}, \theta)}{k}$
General form, constant properties, without heat source	$\dfrac{1}{a}\dfrac{\partial T(\mathbf{r}, \theta)}{\partial \theta} = \nabla^2 T(\mathbf{r}, \theta)$ (Fourier equation)
General form, constant properties, steady state	$\nabla^2 T(\mathbf{r}) + \dfrac{g\mathbf{r}}{k} = 0$ (Poisson equation)
General form, constant properties, steady state without heat source	$\nabla^2 T(\mathbf{r}) = 0$ (Laplace equation)
Cartesian system, constant properties	$\dfrac{\partial^2 T}{\partial x^2} + \dfrac{\partial^2 T}{\partial y^2} + \dfrac{\partial^2 T}{\partial z^2} + \dfrac{g(x,y,z,\theta)}{k} = \dfrac{1}{a}\dfrac{\partial T}{\partial \theta}$
Cylindrical system, constant properties	$\dfrac{1}{r}\dfrac{\partial}{\partial r}\left(r\dfrac{\partial T}{\partial r}\right) + \dfrac{1}{r^2}\dfrac{\partial^2 T}{\partial \phi^2} + \dfrac{\partial^2 T}{\partial z^2} + \dfrac{g(r,\phi,z,\theta)}{k} = \dfrac{1}{\alpha}\dfrac{\partial T}{\partial \theta}$
Spherical system, constant properties	$\dfrac{1}{r^2}\dfrac{\partial}{\partial r}\left(r^2\dfrac{\partial T}{\partial r}\right) + \dfrac{1}{r^2 \sin\psi}\dfrac{\partial}{\partial \psi}\left(\sin\psi\dfrac{\partial T}{\partial \psi}\right) +$ $\dfrac{1}{r^2 \sin^2\psi}\dfrac{\partial^2 T}{\partial \phi^2} + \dfrac{g(r,\phi,\psi,\theta)}{k} = \dfrac{1}{\alpha}\dfrac{\partial T}{\partial \theta}$

Thermal Conductivity of Substances

Thermal conductivity is a thermophysical property of the conducting medium that represents the rate of conduction heat transfer per unit area for a unit temperature gradient. Table 4 lists thermal conductivity values for several materials. In general, thermal conductivity is strongly temperature dependent. Variations in thermal conductivity are shown in Fig. 1 for several representative substances. Values of thermal conductivity, k, for a substance also depend on chemical composition, on the physical state and texture, and on pressure.

One-Dimensional Steady-State Heat Conduction

Neither the heat flow rate nor the temperature distribution in a one-dimensional system varies with time. Furthermore, heat flow is in one direction. Certain practical problems, such as heat conduction through a slab or through the wall of a pipe or a spherical vessel, fall into this category. The Fourier law and the differential equation of heat conduction for this case are reduced to:

$$Q = -kA\frac{dT}{dx} \text{ or } Q = -kA\frac{dT}{dr}$$

and

$$\frac{1}{r^n}\frac{d}{dr}\left(r^n\frac{dT}{dr}\right) = 0$$

with r replacing x for $n = 0$, and with $n = 0$, 1, 2 representing slab, cylinder, and sphere, respectively. Table 5 presents temperature distribution and heat-transfer rate for some simple systems in one-dimensional steady-state conduction.

One-Dimensional Composite Structure at Steady State

Heat transfer in composite regions consisting of several layers has practical applications, such as rocket thrust chamber liners, fuel elements of nuclear reactors, and space re-entry bodies. Consider N layers of material having different thicknesses and thermal conductivities (Fig. 2). The contact resistance between layers is negligible. To determine the heat flow rate, Q, and temperature profile, T, of the structure, assuming one-dimensional heat conduction, the analogy between the diffusion and electric current can be extended to obtain:

$$\frac{1}{U_i A_i} = \frac{1}{U_0 A_0} = R_i + \sum_{n=1}^{N} R_n + R_0 = \frac{T_i - T_0}{Q}$$

where R_i, R_0, and R_n are the thermal resistances at the internal surface, external surface, and the nth layer, respectively; U_i and U_0 are the overall heat-transfer coefficients at the internal and external surfaces, respectively; and A_i and A_0 are the cross-sectional areas perpendicular to heat flow at the internal and the external surfaces. The explicit forms for the inverse of overall heat-transfer coefficient at the external surface, $1/U$, for various geometries are presented in Table 6.

To obtain the temperature distribution in the structure, Q is expressed in terms of the temperature difference $T - T_0$ and associated resistances (from the series R_n, R_{n+1}, R_{n+2}, ..., R_N). The result is:

$$Q = (T - T_0)\Big[(1/k_n)\int_{s}^{s_{n+1}} ds/A(s) + \sum (1/k_m)\cdot\int_{s_m}^{s_{m+1}} ds/A(s) + 1/h_0 A\Big]^{-1}$$

where T represents the temperature of the location s (Fig. 2). The dimensionless temperature profile can be expressed in terms of U_0 for

Table 3 Boundary conditions of heat conduction in solids

Description of system	Schematic diagram	Boundary condition	Description of system	Schematic diagram	Boundary condition
Prescribed surface temperature		$T = T_s$ at $n = s$	Interface of two continua of different conductivities at perfect contact		$-k_1 \dfrac{\partial T_1}{\partial n} =$ $-k_2 \dfrac{\partial T_2}{\partial n}$ at $n = s$ $T_1 = T_2$ at $n = s$
Prescribed heat flux at the surface		(a) $-k\left(\dfrac{\partial T}{\partial n}\right) = +q''$ at $n = s$ (b) $-k\left(\dfrac{\partial T}{\partial n}\right) = -q''$ at $n = s$	Two solid continua in relative motion with pressure on interface p, the coefficient of dry friction, μ, and the relative velocity, V		$-k_1 \dfrac{\partial T_1}{\partial n} + \mu p V =$ $-k_2\left(\dfrac{\partial T_2}{\partial n}\right)$ at $n = s$
Insulation at the surface		$\dfrac{\partial T}{\partial n} = 0$ at $n = s$			
Heat transfer to the ambient with temperature T_∞ by convection		$-k\dfrac{\partial T}{\partial n} = h(T - T_\infty)$ at $n = s$	Ablation or sublimation of solid with melted material removed upon its formation		$f(\theta) + k_s \dfrac{\partial T}{\partial n} = \rho L \dfrac{ds(\theta)}{d\theta}$ at $n = s(\theta)$
Heat transfer to environment at temperature T_e by radiation		$-k\dfrac{\partial T}{\partial n} = \sigma\varepsilon(T^4 - T_e^{\ 4})$ at $n = s$	Solid-liquid interface for one-dimensional solidification and melting problems		$\left[k_s\left(\dfrac{\partial T_s}{\partial n}\right) - k_l\left(\dfrac{\partial T_l}{\partial n}\right)\right]$ $= \rho L \dfrac{ds}{d\theta}$ at $n = s(\theta)$
Transfer heat to the ambient by convection while receiving heat flux from a distance		$-k\dfrac{\partial T}{\partial n} + q'' =$ $h(T - T_\infty)$ at $n = s$			

(a) Solidification

(b) Melting

L = latent heat of fusion; s = melting, solidification, or surface recession distance, or interface; $ds/d\theta$ = melting, solidification, or ablation rate; $f(\theta)$ = time-dependent surface heat flux; subscript s = solid phase; subscript l = liquid phase; ρ = density

various geometric systems (Table 7). These formulations are restricted to the steady-state one-dimensional heat conduction. In practice, the combination of series-and parallel-connected structures is also important, especially in Cartesian geometry. Figure 3(a) illustrates a parallel-series composite wall that is represented by the thermal circuit shown in Fig. 3(b). The corresponding overall heat-transfer coefficient is given by:

$$U = \frac{1}{(b_1 + b_2)(R_1 + R_2 + R_3)}$$

$$= \frac{1}{\dfrac{L_1}{k_1} + \dfrac{b_1 + b_2}{(k_1 b_2/L_2) + (k_2 b_1/L_2)} + \dfrac{L_3}{k_3}}$$

Table 4 Thermal conductivity of various substances at room temperature

Substance	k W/(m·°C)	k Btu/(h·ft·°F)
Metals		
Silver	420	240
Copper	390	230
Gold	320	180
Aluminum	200	120
Silicon	150	87
Nickel	91	53
Chromium	90	52
Iron (pure)	80	46
Germanium	60	35
Carbon steel (1% C)	54	31
Alloy steel (18% Cr, 8% Ni)	16	9.2
Nonmetal solids		
Diamond, type 2A	2300	1300
Diamond, type 1	900	520
Sapphire (Al_2O_3)	46	27
Limestone	1.5	0.87
Glass (Pyrex 7740)	1.0	0.58
Teflon (Duroid 5600)	0.40	0.23
Brick (B & W K-28)	0.25	0.14
Plaster	0.13	0.075
Cork	0.040	0.023
Liquids		
Mercury	8.7	5.0
Water	0.6	0.35
Freon F-12	0.08	0.046
Gases		
Hydrogen	0.18	0.10
Air	0.026	0.015
Nitrogen	0.026	0.015
Steam	0.018	0.01
Freon F-12	0.0097	0.0056

Note: (1 W/(m · °C) = 0.5778 Btu/(h · ft · °F). Source: Ref 1

Fig. 1 Variation of thermal conductivity, k, with temperature, T, for representative substances

LIVE GRAPH
Click here to view

Critical Radius for Insulated Pipes and Spheres

An application of one-dimensional composite structure formulas having practical significance involves insulation of small pipes, electrical wires, and spheres. Given a pipe or sphere of fixed outside radius, as shown in Fig. 4, the insulation thickness that yields the optimum insulating effect must be determined. As insulation is added, the outer surface temperature decreases, but the surface area for convective heat dissipation increases. Some optimum thickness of insulation may exist due to these opposing effects. At a certain radius of insulation, the heat loss is maximum instead of minimum, which is known as critical radius, $r*_0$. Table 8 lists governing equations for determining the critical radius of insulated pipe and sphere for these cases of constant heat-transfer coefficients and variable heat-transfer coefficients with and without the effect of thermal radiation. When the surface radiation is included, the determination of the critical radius requires the solutions of nonlinear systems of equations, because both surface temperature, T_0, and critical radius, $r*_0$, are unknown under this condition.

Steady-State Heat Transfer from Extended Surfaces

One of the important applications of heat transfer is the thermal analysis of extended surfaces or fins to increase the heat transfer between the structure and an ambient fluid. Fins on one side of a wall separating two heat-exchanging fluids are most effective when made part of the face on which the thermal surface resistivity is greatest. Among considerations in fin design and placement are maximum cooling efficiency; minimum material for cost, weight, and space; minimum resistance to the flow of ambient cooling medium; adequate strength; and ease of manufacture. Table 9 presents formulas for temperature profiles, total heat-transfer rate, and fin efficiencies for nine different fin types. The results are based on the assumptions that thermal properties are constant; radiation is not considered; heat transfer at fin tip is negligibly small, and heat conduction is one dimensional. The fin efficiency is defined as the ratio of the actual heat transfer to the maximum possible heat transfer from the fin. Figures 5 and 6 show the comparisons of fin efficiencies of four longitudinal fins and four spines or pin fins, respectively. Table 10 gives optimum dimensions of some common longitudinal and pin fins in terms of fin profile area, or fin volume, heat-transfer coefficient, and thermal conductivity.

Multidimensional Steady-State Heat Conduction

The problems of steady-state heat conduction in more than one dimension can be solved by analytical, graphical, experimental-analogic, and numerical means. The analytical approach is recommended in dealing with systems that are simple in geometry and boundary conditions.

Systems of complex geometry with isothermal and insulated boundaries are readily solved by graphical or analogical methods. When the boundary conditions involve heat transfer through surface conductance, these methods become less useful, and the numerical approach is recommended. This method has sufficient flexibility for problems with variable physical properties and nonuniform boundary conditions. Numerical solutions can be conveniently obtained by a digital computer.
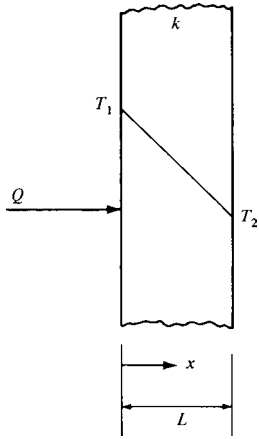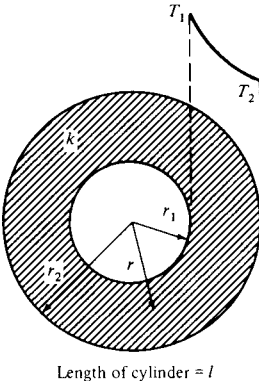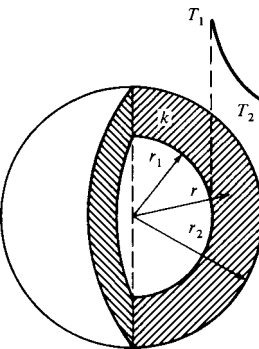
In a two-dimensional system where only two temperature limits are involved, a conduction shape factor, S, may be defined in such a way that $Q = kS \Delta T_{overall}$. The values of S for several geometries are summarized in Table 11.

Unsteady-State Heat Conduction

If a solid body is rapidly subjected to a change in environment, some time must elapse before temperature equilibrium or steady state prevails. Analyses must be modified to account for the change in internal energy of the body with time, and boundary conditions must be adjusted to match the physical situation in the unsteady-state heat-transfer problem. Unsteady heat flow problems also include periodic variations of temperature and heat flow, such as found in internal combustion engines, air conditioning, instrumentation, and process control.

System with Negligible Internal Thermal Resistance. Although no materials possess an infinite thermal conductivity, many transient heat-flow problems may be solved with acceptable accuracy by assuming that the internal conductive resistance is so small that the system temperature is substantially uniform at any instant. This assumption is justified when the external thermal resistance between the surface and the surrounding medium is so large compared to the internal thermal resistance that it controls the heat-transfer process. Analysis based on this assumption is called lumped heat-capacity analysis. A measure of the thermal resistance within a solid body is the ratio of the internal to the external resistance. This ratio can be written in dimensionless form as hL/k, the Biot number, where h is the average unit-surface conductance, L is a significant length dimension obtained by dividing the volume of the body by its surface area, and k is the thermal conductivity of the solid body. When the Biot number is less than 0.1, the lumped capacity type of analysis is expected to yield a

Table 5 Temperature distribution and heat-transfer rate through slab, hollow cylinder, and hollow sphere

Conducting body	Schematic diagram	Temperature distribution, $\frac{T-T_2}{T_1-T_2}$	Thermal conductance, $\frac{Q}{T_1-T_2}$
Slab		$\left(1+\frac{x}{d}\right)$	$\frac{kA}{L}$
Circular cylinder	Length of cylinder = l	$\frac{\ln(r-r_2)}{\ln(r_1-r_2)}$	$\frac{2\pi kl}{\ln(r_2/r_1)}$
Hollow sphere		$\left(\frac{\frac{1}{r}-\frac{1}{r_2}}{\frac{1}{r_1}-\frac{1}{r_2}}\right)$	$\left(\frac{4\pi k}{\frac{1}{r_1}-\frac{1}{r_2}}\right)$

reasonable estimate. For a solid with a uniform initial temperature, T_i, suddenly placed in an environment with a uniform temperature, T_∞, if the heat-transfer coefficient and specific heat are constant, the lumped capacity analysis leads to the following expression for temperature history of the solid:

$$\frac{T-T_\infty}{T_i-T_\infty} = \exp(-kA\theta/\rho\,cV)$$
$$= \exp(-F_0 B_i)$$

where B_i and F_0 are the Biot number and Fourier number, respectively ($F_0 = a\,\theta/L^2$). The instantaneous transfer rate at any time, θ, is:

$$q = \rho\,cV\frac{dT}{d\theta} = kA(T_\infty - T_i)\exp(-B_i \cdot F_0)$$

The amount of heat transfer in time interval, θ, which equals the change of internal energy of the lumped system, becomes:

$$Q = \int_0^0 q\,d\theta = (T_\infty - T_i)\rho\,cV[1 - \exp(B_i \cdot F_0)]$$

A summary of temperature histories of some simple systems on the basis of lumped capacity analysis is presented in Table 12.

System with Negligible Surface Thermal Resistance. When the internal thermal resistance of a solid is substantially higher than the surface resistance, the heat-transfer coefficient may be considered infinite ($h \rightarrow \infty$) so that the initial body surface temperature, T_0, is suddenly changed to and maintained at a constant ambient temperature, T_∞ (i.e., $T_0 = T_\infty$). For this case, the determination of temperature distribution requires the solution of partial differential equations. Analytical solutions are available for simple geometries with constant thermal properties k, ρ, and c (Table 13). The corresponding graphical representations are shown in Fig. 7 through 10.

System with Finite Surface and Internal Thermal Resistance. In most practical conduction heat-transfer problems, both the thermal conductivity of the solid and the surface heat-transfer coefficient are finite. Therefore, the convective boundary conditions must be employed in the solution of Fourier differential equations. A large number of analytical solutions are available in the literature. Table 14 illustrates a few simple systems encountered frequently in engineering practice, with corresponding temperature charts presented in Fig. 11 through 20 (Heisler and Gröber charts).

Two- and Three-Dimensional Transient Systems. Many practical problems involve two- and three-dimensional heat flow. The solution of such problems is often achieved from a product solution. Several one-dimensional transient solutions may be combined to obtain the solution to a two- or three-dimensional transient problem. Figure 21 shows the required product solution for the geometries indicated. In such cases, the body is initially at a uniform temperature equal to T_i and is instantaneously placed in a convective environment at temperature T_∞. The following notations are employed in Fig. 21:

$$C(r,\theta) = \frac{T(r,\theta) - T_\infty}{T_i - T_\infty}$$
$$P(x,\theta) = \frac{T(x,\theta) - T_\infty}{T_i - T_\infty}$$
$$S(x,\theta) = \frac{T(x,\theta) - T_\infty}{T_i - T_\infty}$$

where $C(r,\theta)$ represents a transient solution for a cylindrical geometry, $P(x,\theta)$ represents a transient solution for a plane wall, and $S(x,\theta)$ represents a transient solution for a semi-infinite body.

Convection Heat Transfer

Convection heat transfer occurs between a fluid and a solid surface in contact with the fluid. When fluids are everywhere at rest,
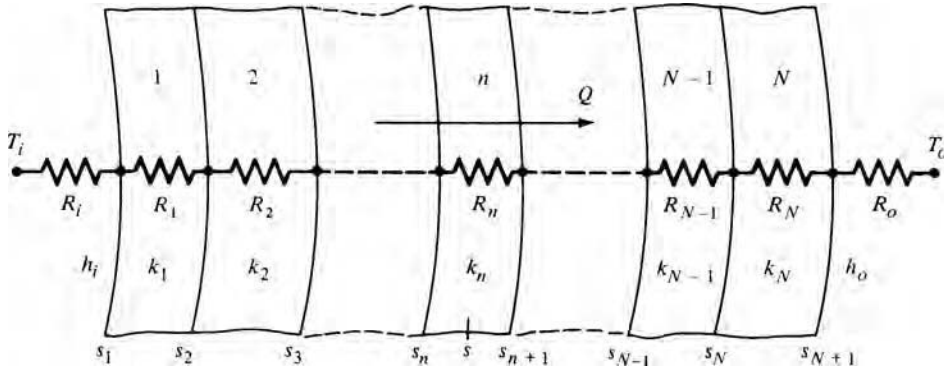
Fig. 2 One-dimensional composite structure

Table 6 Inverse of overall heat-transfer coefficient on the external surface, $1/U_0$, of one-dimensional composites

General formula	$\dfrac{A(s_{N+1})/A(s_1)}{h_i} + A(s_{N+1}) \displaystyle\sum_{n=1}^{N} \frac{1}{k_n} \int_{s_n}^{s_{n+1}} \frac{ds}{A(s)} + \frac{1}{h_0}$
Cartesian system	$\dfrac{1}{h_i} + \displaystyle\sum_{n=1}^{N} \frac{L_n}{k_n} + \frac{1}{h_0}$
Cylindrical system	$\dfrac{(r_{N+1}/r_1)}{h_i} + r_{N+1} \displaystyle\sum_{n=1}^{N} \frac{1}{k_n} \ln\left(\frac{r_{n+1}}{r_n}\right) + \frac{1}{h_0}$
Spherical system	$\dfrac{(r_{N+1}/r_1)^2}{h_i} + r_{N+1}^2 \displaystyle\sum_{n=1}^{N} \frac{1}{k_n} \left(\frac{1}{r_n} - \frac{1}{r_{n+1}}\right) + \frac{1}{h_0}$

Note: $A(s_n)$ = cross-sectional area of nth layer composite at the location s_n; L_n = thickness of nth layer; see Fig. 2 for other notations.

Table 7 Dimensionless temperature profile in one-dimensional composites, $\dfrac{T-T_0}{T_i-T_0}$

General formula	$U_0 \left[\dfrac{A(s_{N+1})}{k_n} \displaystyle\int_s^{s_{N+1}} \frac{ds}{A(s)} + A(s_{N+1}) \times \right.$ $\left. \displaystyle\sum_{m=n+1}^{N} \frac{1}{k_m} \int_{s_m}^{s_{m+1}} \frac{ds}{A(s)} + \frac{1}{h_0} \right]$
Cartesian system	$U_0 \left[\dfrac{x_{n+1}-x}{k_n} + \displaystyle\sum_{m=n+1}^{N} \frac{x_{m+1}-x_m}{k_m} + \frac{1}{h_0} \right]$
Cylindrical system	$U_0 \left[\dfrac{r_{N+1}}{k_n} \ln\left(\frac{r_{n+1}}{r}\right) + r_{N+1} \times \right.$ $\left. \displaystyle\sum_{m=n+1}^{N} \frac{1}{k_m} \ln\left(\frac{r_{m+1}}{r_m}\right) + \frac{1}{h_0} \right]$
Spherical system	$U_0 \left[\dfrac{r_{N+1}}{k_n} \left(\frac{r_{N+1}}{r} - \frac{r_{N+1}}{r_{n+1}}\right) + r_{N+1}^2 \times \right.$ $\left. \displaystyle\sum_{m=n+1}^{N} \frac{1}{k_m} \left(\frac{1}{r_m} - \frac{1}{r_{m+1}}\right) + \frac{1}{h_0} \right]$

Note: x_n = distance of inner surface of nth composite layer; r_n = inner radius of nth composite layer; see Fig. 2 for other notations.
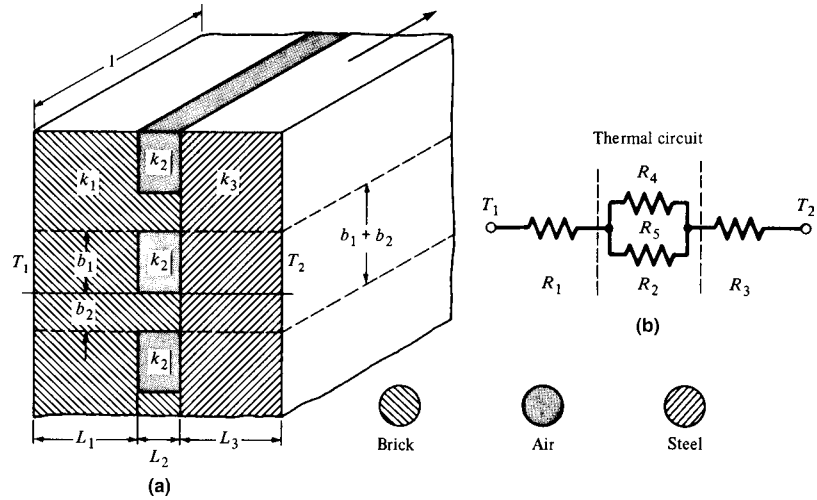
Fig. 3 Parallel-series composite and its equivalent thermal circuit
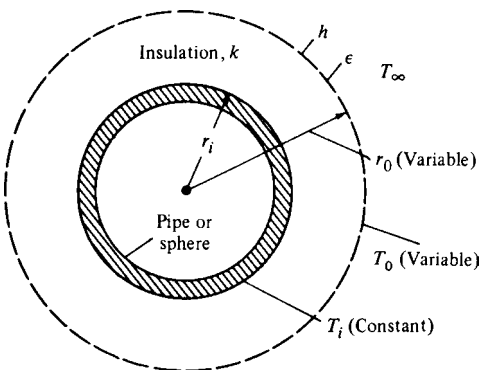
Fig. 4 Critical thickness of pipe or sphere insulation

Table 8 Equations for critical radius, r_0^*, in radial conduction

Boundary condition	Cylindrical system	Spherical system
Constant heat-transfer coefficient without radiation	$r_0^* = \dfrac{k}{h}$	$r_0^* = \dfrac{2k}{h}$
Variable heat-transfer coefficient with $h = \dfrac{(T_0 - T_\infty)^n}{r_0^m}$ and without radiation	$r_0^* = \dfrac{k(1-m)}{h(1+n)}$	$r_0^* = \dfrac{k(2-m)}{h(1+n)}$
Constant heat-transfer coefficient but with radiation	$r_0^* = \dfrac{k}{h + 4\varepsilon\sigma T_0^3}$ $\dfrac{k(T_i - T_0)}{\ln \frac{r_0^*}{r_i}} = hr_0^*(T_0 - T_\infty) + r_0^*\sigma\varepsilon(T_0^4 - T_\infty^4)$	$r_0^* = \dfrac{2k}{h + 4\sigma\varepsilon T_0^3}$ $\dfrac{kr_i(T_i - T_0)}{r_0^* - r_i} = hr_0^*(T_0 - T_\infty) + \sigma\varepsilon r_0^*(T_0^4 - T_\infty^4)$
Variable heat-transfer coefficient with $h = (T_0 - T_\infty)^n/r_0^m$ and with radiation	$r_0^* = \dfrac{k[h(1-m)(T_0 - T_\infty) + \varepsilon\sigma(T_0^4 - T_\infty^4)]}{[(1+n)h + 4\varepsilon\sigma T_0^3][h(T_0 - T_\infty) + \sigma\varepsilon(T_0^4 - T_\infty^4)]}$ and $\dfrac{k(T_i - T_0)}{\ln \frac{r_0^*}{r_i}} = hr_0^*(T_0 - T_\infty) + r_0^*\sigma\varepsilon(T_0^4 - T_\infty^4)$	$r_0^* = \dfrac{k[(2-m)h(T_0 - T_m) + 2\sigma\varepsilon(T_0^4 - T_\infty^4)]}{[(1+n)h + 4\sigma\varepsilon T_0^3][h(T_0 - T_\infty) + \sigma\varepsilon(T_0^4 - T_\infty^4)]}$ and $\dfrac{kr_i(T_i - T_0)}{r_0^* - r_i} = hr_0^*(T_0 - T_\infty) + \sigma\varepsilon r_0^*(T_0^4 - T_\infty^4)$

Table 9(a) Heat-transfer characteristics of various types of fin

Type of fin	Schematic diagram	Type of fin	Schematic diagram
Longitudinal fin of rectangular profile		Radial fin of hyperbolic profile	
Longitudinal fin of triangular profile		Cylindrical spine	
Longitudinal fin of concave parabolic profile		Conical spine	
Longitudinal fin of convex parabolic profile		Spine of concave parabolic profile	
Radial fin of rectangular profile		Spine of convex parabolic profile	

Source: Ref 2

contact with the fluid, and h is the heat-transfer coefficient. The heat-transfer coefficient is a complicated function of the fluid flow, the thermal properties of the fluid, and the geometry of the system. Its numerical value, in general, depends on the location where the fluid temperature, T_∞, is measured. Four general methods are used to evaluate convection-transfer coefficients:

- Dimensional analysis combined with experiments
- Exact mathematical solutions of the boundary-layer equations
- Approximate analyses of the boundary layer by integral methods
- The analogy between heat, mass, and momentum transfer

In the most practical case, the heat-transfer coefficients are evaluated from empirical equations. These equations are obtained by correlating experimental results coupling with dimensional analysis. For most engineering applications, average values are generally of interest. Table 15 gives typical values of the order of magnitude of average convection heat-transfer coefficients encountered in engineering practice. The heat-transfer coefficients are usually given by a relationship between one dependent nondimensional group, the Nusselt number, Nu, and three other nondimensional groups, the Reynolds number, Re; the Prandtl number, Pr; and the Grashof number, Gr, depending on whether the case involves free convection or forced convection. Table 16 gives some of dimensionless groups relevant to convection problems along with their physical interpretations.

Exact mathematical analyses of convective heat transfer require the simultaneous solution of the equations describing the fluid motion and the energy transfer in the moving fluid. Hence, the knowledge of fluid flow is essential to the fundamental study of the convective heat-transfer processes.

Differential Equations of Fluid Flow. Table 17 summarizes these equations for laminar flow with constant thermal properties. The important differential equations of fluid flow are the:

- Continuity equation, based on the law of conservation of mass
- Momentum equation, based on Newton's second law of motion
- Energy equations, based on the law of conservation of energy

Analogy between Heat Transfer and Momentum Transfer

In hydrodynamic analysis of both forced and free convection, it is important to establish whether the flow is laminar or turbulent. Laminar forced or natural convection flows exist when individual elements of fluid follow a

the problem becomes one of either simple heat conduction or simple diffusion. If fluid motion is induced by such means as a pump, blower, wind, or vehicle motion, the process is generally called forced convection. If the fluid motion arises from external force fields such as gravity, acting on a density gradient induced by the transport process itself, the process is referred to as free convection. The rate of heat transfer by convection between a solid and fluid is evaluated by Newton's law of cooling, $Q = hA(T_s - T_\infty)$, where A is the surface area in

Table 9(b) Heat-transfer characteristics of various types of fins

Type of fin	Temperature profile, $\frac{T-T_\infty}{T_b-T_\infty}=\frac{\Theta}{\Theta_b}$	Rate of total heat dissipation, Q
Longitudinal fin of rectangular profile	$\dfrac{\cosh mx}{\cosh mb}$	$k\delta_o m\theta_b\tanh mb$
Longitudinal fin of triangular profile	$\dfrac{I_0(2m\sqrt{bx})}{I_0(2mb)}$	$\dfrac{2h\Theta_b I_1(2mb)}{mI_0(2mb)}$
Longitudinal fin of concave parabolic profile	$\left(\dfrac{x}{b}\right)^{P_1}$	$\dfrac{k\delta_o\Theta_b}{2b}\left[-1+\sqrt{1+(2mb)^2}\right]$
Longitudinal fin of convex parabolic profile	$\left(\dfrac{x}{b}\right)^{1/4}\dfrac{I_{-1/3}\left(\frac{4}{3}mb^{1/4}x^{3/4}\right)}{I_{-1/3}\left(\frac{4}{3}mb\right)}$	$k\delta_o\Theta_b m\dfrac{I_{2/3}\left(\frac{4}{3}mb\right)}{I_{-1/3}\left(\frac{4}{3}mb\right)}$
Radial fin of rectangular profile	$\dfrac{[K_1(mr_e)I_0(mr)+I_1(mr_e)K_0(mr)]}{I_0(mr_o)K_1(mr_e)+I_1(mr_e)K_0(mr_o)}$	$2\pi r_o\delta_o km\Theta_b\left[\dfrac{I_1(mr_e)K_1(mr_o)-K_1(mr_e)I_1(mr_o)}{I_0(mr_o)K_1(mr_e)+I_1(mr_e)K_0(mr_o)}\right]$
Radial fin of hyperbolic profile	$\left(\dfrac{r}{r_o}\right)^{1/2}\left\{\dfrac{I_{2/3}\left(\frac{2}{3}Mr_e^{3/2}\right)I_{1/3}\left(\frac{2}{3}Mr^{3/2}\right)-I_{-2/3}\left(\frac{2}{3}Mr_e^{3/2}\right)I_{-1/3}\left(\frac{2}{3}Mr^{3/2}\right)}{I_{2/3}\left(\frac{2}{3}Mr_e^{3/2}\right)I_{1/3}\left(\frac{2}{3}Mr_o^{3/2}\right)-I_{-2/3}\left(\frac{2}{3}Mr_e^{3/2}\right)I_{-1/3}\left(\frac{2}{3}Mr_o^{3/2}\right)}\right\}$	$2\pi k r_o\delta_o\Theta_b M\sqrt{r_o\psi}$
Cylindrical spine	$\dfrac{\cosh mx}{\cosh mb}$	$\dfrac{\pi}{4}kd^2 m\Theta_b\tanh mb$
Conical spine	$\left(\dfrac{b}{x}\right)^{1/2}\dfrac{I_1(2M\sqrt{x})}{I_1(2M\sqrt{b})}$	$\dfrac{\pi k\delta_o^2\Theta_b M}{4\sqrt{b}}\left[\dfrac{I_2(2M\sqrt{b})}{I_1(2M\sqrt{b})}\right]$
Spine of concave parabolic profile	$\left(\dfrac{x}{b}\right)^{P_1}$	$\dfrac{\pi k\delta_o^2\Theta_b[-3+(9+4M^2)^{1/2}]}{8b}$
Spine of convex parabolic profile	$\dfrac{I_0\left(\frac{4}{3}\sqrt{2}mb^{1/4}x^{3/4}\right)}{I_0\left(\frac{4}{3}\sqrt{2}mb\right)}$	$\dfrac{\sqrt{2}}{4}k\pi\delta_o^2\Theta_b m\dfrac{I_1\left(\frac{4}{3}\sqrt{2}mb\right)}{I_0\left(\frac{4}{3}\sqrt{2}mb\right)}$
Longitudinal fin of rectangular profile	$\dfrac{\tanh mb}{mb}$	$m=(2h/k\delta_o)^{1/2}$
Longitudinal fin of triangular profile	$\dfrac{I_1(2mb)}{mbI_0(2mb)}$	$m=(2h/k\delta_o)^{1/2}$
Longitudinal fin of concave parabolic profile	$\dfrac{2}{1+\sqrt{1+(2mb)^2}}$	$P_1=-\dfrac{1}{2}+\dfrac{1}{2}(1+4m^2b^2)^{1/2}$
Longitudinal fin of convex parabolic profile	$\dfrac{1}{mb}\dfrac{I_{2/3}\left(\frac{4}{3}mb\right)}{I_{-1/3}\left(\frac{4}{3}mb\right)}$	$m=\left(\dfrac{2h}{k\delta_0}\right)^{1/2}$
Radial fin of rectangular profile	$\dfrac{2r_o}{m(r_e^2-r_o^2)}\left[\dfrac{I_1(mr_e)K_1(mr_o)-K_1(mr_e)I_1(mr_o)}{I_0(mr_o)K_1(mr_e)+I_1(mr_e)K_0(mr_o)}\right]$	$m=\left(\dfrac{2h}{k\delta_o}\right)^{1/2}$
Radial fin of hyperbolic profile	$\dfrac{2r_o\psi}{m(r_e^2-r_o^2)}$	$\psi=\dfrac{\left[I_{2/3}\left(\frac{2}{3}Mr_e^{3/2}\right)I_{-2/3}\left(\frac{2}{3}Mr_o^{3/2}\right)-I_{-2/3}\left(\frac{2}{3}Mr_e^{3/2}\right)I_{2/3}\left(\frac{2}{3}Mr_o^{3/2}\right)\right]}{\left[I_{-2/3}\left(\frac{2}{3}Mr_e^{3/2}\right)I_{-1/3}\left(\frac{2}{3}Mr_o^{3/2}\right)-I_{2/3}\left(\frac{2}{3}Mr_e^{3/2}\right)I_{1/3}\left(\frac{2}{3}Mr_o^{3/2}\right)\right]}$ $M^2=m^2/r_o=\dfrac{2h}{k\delta_o r_o}$
Cylindrical spine	$\dfrac{\tanh mb}{mb}$	$m=(4h/kd)^{1/2}$
Conical spine	$\dfrac{\sqrt{2}I_2(2\sqrt{2}mb)}{(mb)I_1(2\sqrt{2}mb)}$	$M=(2m^2b)^{1/2}$ and $m=(2h/k\delta_o)^{1/2}$
Spine of concave parabolic profile	$\dfrac{2}{1+(1+\frac{8}{9}m^2b^2)^{1/2}}$	$M=\sqrt{2}mb$ and $m=(2h/k\delta_o)^{1/2}$ $P_1=-\dfrac{3}{2}+\dfrac{1}{2}\sqrt{9+4M^2}$
Spine of convex parabolic profile	$\dfrac{1}{(2\sqrt{2}/3)mb}\dfrac{I_1\left(\frac{4}{3}\sqrt{2}mb\right)}{I_o\left(\frac{4}{3}\sqrt{2}mb\right)}$	$m=(2h/k\delta_o)^{1/2}$

Source: Ref 2

smooth, streamlined path. The flow is turbulent when the movement of elements of fluid is unsteady and random. The dimensionless group known as Reynolds number is the criterion for determining whether laminar or turbulent flow is the stable form under given conditions. Complete mathematical equations describing the fluid flow and heat-transfer mechanism can be written for laminar flow, although the equations can be solved analytically only for a number of simple systems, such as flow over a flat plate or through a circular tube. Our knowledge of turbulent exchange mechanism is not sufficient to write equations describing the temperature distribution directly. A useful tool for analyzing the turbulent heat-transfer process is the concept of analogy between heat and momentum; that is, the heat-transfer coefficient is expressed in terms of friction coefficient, f, which is available from experiments. Table 18 presents a few well-known analogy formulas for turbulent pipe flow of gases and liquids. The Reynolds analogy is satisfactory for gases only (Pr $\approx$ 1), but the Colburn analogy can be used for a fluid having Prandtl numbers ranging from 0.6 to 50.

To apply the analogy between heat and momentum transfers, it is necessary to know the
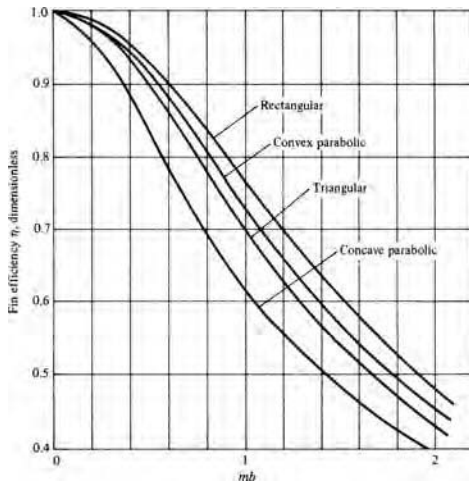
Fig. 5 Comparison of fin efficiencies of four longitudinal fins

LIVE GRAPH
Click here to view
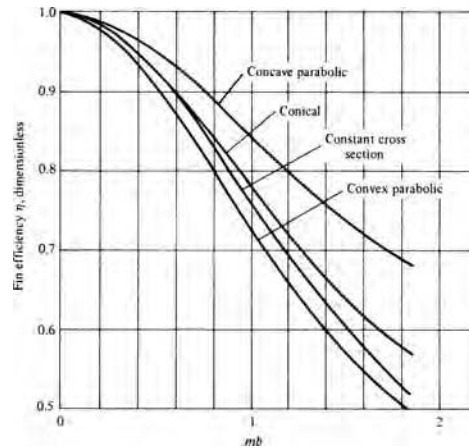
Fig. 6 Comparison of fin efficiencies of four spines

LIVE GRAPH
Click here to view

Table 10   Optimum dimensions of some longitudinal fins and spines

Fin profile	Dimension Fin width, δ_0	Fin height, b
Longitudinal fin of rectangular profile	$0.791\left[A_p^2\left(\frac{2h}{k}\right)\right]^{1/3}$	$1.262\left(\frac{kA_p}{2h}\right)^{1/3}$
Longitudinal fin of triangular profile	$1.328\left[A_p^2\left(\frac{2h}{k}\right)\right]^{1/3}$	$1.506\left(\frac{A_pk}{2h}\right)^{1/3}$
Longitudinal fin of concave parabolic profile	$1.651\left[A_p^2\left(\frac{2h}{k}\right)\right]^{1/3}$	$1.817\left(\frac{A_pk}{2h}\right)^{1/3}$
Longitudinal fin of convex parabolic profile	$1.071\left[A_p^2\left(\frac{2h}{k}\right)\right]^{1/3}$	$1.401\left(\frac{kA_p}{2h}\right)^{1/3}$
Cylindrical spine	$1.308\left[V^2\left(\frac{2h}{k}\right)\right]^{1/5}$	$0.744\left[V\left(\frac{k}{2h}\right)^2\right]^{1/5}$
Conical spine	$1.701\left[V^2\left(\frac{2h}{k}\right)\right]^{1/5}$	$1.320\left[V\left(\frac{k}{2h}\right)^2\right]^{1/5}$
Spine of concave parabolic profile	$1.825\left[V^2\left(\frac{2h}{k}\right)\right]^{1/5}$	$1.911\left[V\left(\frac{k}{2h}\right)^2\right]^{1/5}$
Spine of convex parabolic profile	$1.564\left[V^2\left(\frac{2h}{k}\right)\right]^{1/5}$	$1.041\left[V\left(\frac{k}{2h}\right)^2\right]^{1/5}$

Source: Ref 3, 4

friction factor. Table 19 summarizes the equations available for the prediction of friction factors for both laminar and turbulent flows in various geometries. Nomenclature for flow across tube banks is shown in Fig. 22. In addition, the friction factor chart for pipe flow is shown in Fig. 23, which is known as the Moody Chart.

Forced Convection of External Flow

External flow includes flow along a flat plate; flow over a bluff body such as a sphere, wire, or tube; and flow normal to tube bundles. The important difference between flows over a plate and over a bluff body lies in the behavior of the boundary layer. On a streamlined body, the flow separation, if it takes place at all, occurs near the rear. On a bluff body, the point of separation of the boundary layer often lies not far from the leading edge.

The flat plate is the simplest geometry to analyze. It has been thoroughly studied, and the results are very useful. For several geometries, Table 20 shows a summary of the dimensionless mean heat-transfer coefficient, Nu, for external flow. Heat transfer in flow across tube banks is of particular importance in the design of heat exchangers. The correlations for the average Nusselt numbers for this situation have the following form:

$$\mathrm{Nu} = C\mathrm{Re}_{max}\mathrm{Pr}^{1/3}$$

where Re_{max} is the Reynolds number based on the largest velocity. All fluid properties in the various dimensionless groups are evaluated at the film temperature. The values of C and n are presented in Table 21 for tube banks of 10 rows or more. If the number of rows is less than 10, the Nu value obtained from Table 21 should be multiplied by an appropriate factor from Table 22.

Forced Convection of Internal Flow
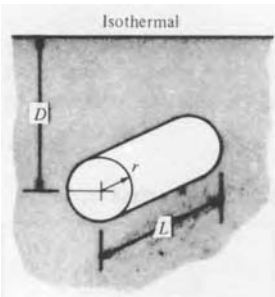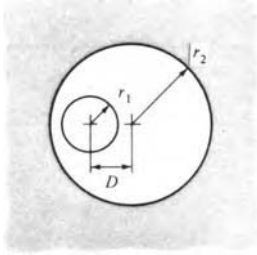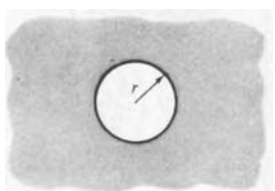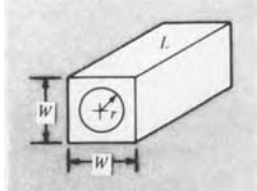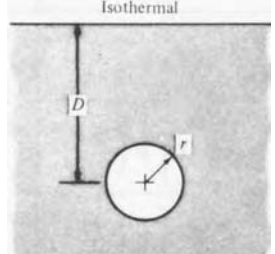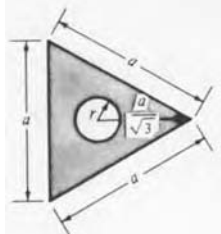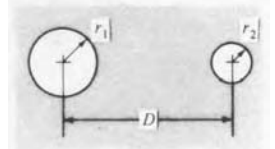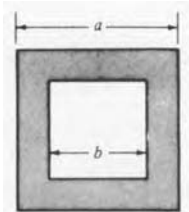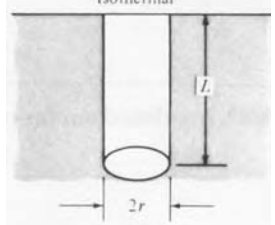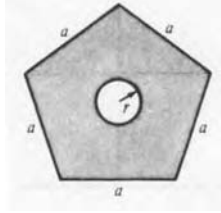
The heating and cooling of fluids flowing in conduits are important heat-transfer processes in engineering. In convective heat-transfer problems of pipe flow, two types of boundary conditions encountered most frequently are uniform wall temperature (UWT) and uniform wall heat flux (UHF). In either case, the Nusselt number is expressed in terms of the Reynolds number and Prandtl number. Table 23 summarizes the correlations for forced convection of internal flow under fully developed conditions.

When the fluid enters the pipe, a certain distance, called entrance length, is required for the velocity to be fully developed. After the entrance length, the velocity is independent of axial position. Similarly, the thermal entrance length is the distance required for the temperature profile to become fully developed. The heat-transfer coefficient decreases as fluid moves farther from the entrance. Table 24 provides the values of local Nusselt numbers for laminar flow through a circular tube subject to constant wall temperature and constant wall flux conditions. For the case of turbulent flow, the local Nusselt numbers in the combined thermal and hydrodynamic entry length of circular tubes with UHF and UWT are shown in Fig. 24 and 25, respectively.

Free Convection

Free convection currents transfer internal energy stored in the fluid in essentially the same manner as forced convection currents. However, the intensity of the mixing motion is generally less in free convection; consequently, the heat-transfer coefficients are lower than that of forced convection. Free convection is the dominant heat-flow mechanism from steam radiators, walls of a building, the human body, transmission lines, transformers, and electrically heated wires to a quiescent atmosphere. The determination of heat load on air conditioning or refrigeration equipment requires a knowledge of free convection heat-transfer coefficients. Gravity is not the only body force that can produce free convection. Centrifugal forces and coriolis forces also influence free convective heat transfer, particularly in rotating systems. The Nusselt number associated with free convection from an isothermal wall is related to the Grashof number and Prandtl number by the expression $\mathrm{Nu} = C(\mathrm{GrPr})^n R$, in which a correction function, R, is introduced to account for counteracting effects and to cover a wider range of parameters. Results of free convective heat-transfer coefficients of

Table 11 Shape factor, S, for various systems in steady-state heat conduction, $S = \dfrac{Q}{k(T_1 - T_2)}$

Schematic diagram	Shape factor	Restrictions	Schematic diagram	Shape factor	Restrictions
Isothermal cylinder of radius r buried in semi-infinite medium having isothermal surface			**Eccentric cylinders of length L**		
	$\dfrac{2\pi L}{\cosh^{-1}(D/r)}$	$L \gg r$		$\dfrac{2\pi L}{\cosh^{-1}\left(\frac{r_1^2 + r_2^2 - D^2}{2r_1 r_2}\right)}$	$L \gg r_2$
	$\dfrac{2\pi L}{\ln(2D/r)}$	$L \gg r$			
	$\dfrac{2\pi L}{\frac{\ln L}{r\left\{1 - \frac{\ln L/(2D)}{\ln(L/r)}\right\}}}$	$D > 3r$ $D \gg r$ $L \gg D$			
Isothermal sphere of radius r buried in infinite medium			**Cylinder centered in a square of length L**		
	$4\pi r$			$\dfrac{2\pi L}{\ln(0.54\, W/r)}$	$L \gg W$
Isothermal sphere of radius r buried in semi-infinite medium having isothermal surface			**Pipe in a triangular body**		
	$\dfrac{4\pi r}{1 - r/2D}$			$\dfrac{2\pi L}{\ln\left(\frac{0.327a}{r}\right)}$	
Conduction between two isothermal cylinders buried in infinite medium			**Square pipe**		
	$\dfrac{2\pi L}{\cosh^{-1}\left(\frac{D^2 - r_1^2 - r_2^2}{2r_1 r_2}\right)}$	$L \gg r$ $L \gg D$		$\dfrac{6.791\, L}{\ln\left(\frac{a}{b}\right) - 0.054}$	
Isothermal cylinder of radius r placed in semi-infinite medium as shown			**Pipe in a pentagonal body**		
	$\dfrac{2\pi L}{\ln(2L/r)}$	$L \gg 2r$		$\dfrac{2\pi L}{\ln\left(\frac{0.724a}{r}\right)}$	

(continued)

T_1 and T_2 are two isothermal surface temperatures of the body summarized from Ref 5 and 6.
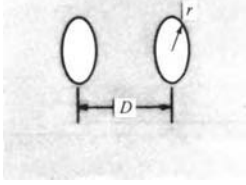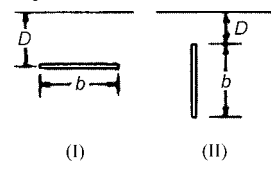
Table 11   (continued)

Schematic diagram	Shape factor	Restrictions	Schematic diagram	Shape factor	Restrictions
Isothermal rectangular parallelepiped buried in semi-infinite medium having isothermal surface Isothermal	$1.685 L\left[\log\left(1+\dfrac{b}{a}\right)\right]^{-0.59}\left(\dfrac{b}{c}\right)^{-0.078}$		**Pipe in a hexagonal body**	$\dfrac{2\pi L}{\ln\left(\frac{0.898a}{r}\right)}$	
Thin horizontal disk buried in semi-infinite medium with isothermal surface Isothermal	$4r$ $8r$	$D = 0$ $D \gg 2r$	**Confocal ellipses**	$\dfrac{2\pi L}{\ln\left(\frac{c+D}{a+D}\right)}$	
Hemisphere buried in semi-infinite medium Insulated	$2\,\pi\,r$		**Strip in an elliptical body**	$\dfrac{2\pi L}{\ln\left(\frac{D+c}{a}\right)}$	
Isothermal sphere buried in semi-infinite medium with insulated surface	$\dfrac{4\pi r}{1 + r/2D}$		**Rectangular duct buried in a solid**	$\dfrac{\left(5.7 + \frac{6}{2a}\right)L}{\ln\left(\frac{3.5D}{b^{1/4}a^{3/4}}\right)}$	
Two isothermal spheres buried in infinite medium	$\dfrac{4\pi}{\dfrac{r_2}{r_1}\left[1 - \dfrac{(r_1/D)^4}{1-(r_2/D)^2}\right] - \dfrac{2r_2}{D}}$	$D > 5r$	**Row of pipes buried in a solid (for any one pipe)**	$\dfrac{2\pi L}{\ln\left[\frac{e}{\pi r}\sinh\left(\frac{2\pi D}{e}\right)\right]}$ Pipes in semi-infinite solid $\dfrac{2\pi L}{\ln\left[\frac{e}{\pi r}\sinh\left(\frac{\pi D}{e}\right)\right]}$ Pipes in midplane of solid	
Thin rectangular plate of length L buried in semi-infinite medium having isothermal surface Isothermal	$\dfrac{\pi W}{\ln(4W/L)}$ $\dfrac{2\pi W}{\ln(4W/L)}$	$D = 0$ $D \gg W$	**Toroidal body buried in a soil**	$4\pi^2 R/\ln\left[8\dfrac{R}{r}\left(\dfrac{\ln(4R/D)}{\ln(8R/r)}\right)+1\right]$ $4\pi^2 R/\ln(8\,R/r)$	$r \ll D \ll R$ $D \gg R$

(continued)

T_1 and T_2 are two isothermal surface temperatures of the body summarized from Ref 5 and 6.

Table 11   (continued)

Schematic diagram	Shape factor	Restrictions	Schematic diagram	Shape factor	Restrictions
Parallel disks buried in infinite medium	$\dfrac{4\pi}{2\left[\frac{\pi}{2}-\tan^{-1}(r/D)\right]}$	$D > 5r$	**Strip buried in a solid** (I) (II)	(I) $\dfrac{2.94\,L}{(D/b)^{0.32}}$ (II) $\dfrac{2.38\,L}{(D/b)^{0.24}}$	

T_1 and T_2 are two isothermal surface temperatures of the body summarized from Ref 5 and 6.

Table 12   Temperature history of simple systems of negligible internal thermal resistance ($k \rightarrow \infty$)

System	Temperature history, $\frac{T-T_\infty}{T_i-T_\infty}$
Infinite plate of thickness L	$e^{-(2h/L\rho c)\theta}$
Infinite cylinder of radius r_0	$e^{-(2h/r_0\rho c)\theta}$
Sphere of radius r_0	$e^{-(3h/r_0\rho c)\theta}$
Infinite square rod of side a	$e^{-(4h/a\rho c)\theta}$
Cube of side a	$e^{-(6h/a\rho c)\theta}$
A thin plate submerged in a fluid with temperature that varies linearly with time, i.e., $T_\infty = a + b\theta$	$\theta = (a+b\theta) - \dfrac{\rho cVb}{hA} +$ $\left(T_i - a + \dfrac{\rho cVb}{hA}\right)e^{-(hA/\rho cV)\theta}$

Note: ρ = density; c = specific heat; V = volume of the body; A = surface area

various geometrical systems are presented in Table 25. The rate of free convection heat transfer between the solid surface and the surrounding fluid, apart from other factors, depends on whether the flow is laminar or turbulent. For an isothermal wall, laminar free convection occurs when $10^4 < \text{PrGr} < 10^9$, and turbulent free convection occurs when $\text{PrGr} > 10^9$. The free convection effect can be neglected when $\text{PrGr} < 10^3$.

Another interesting phenomenon of complex systems involving free convection is heat transfer inside an enclosed space, such as heat transfer in two isothermal vertical plates with two different temperatures. For this case, the heat flux is conveniently expressed in terms of effective or apparent thermal conductivity, k_e, that is:

$$\frac{Q}{A} = k_e \frac{\Delta T_w}{\delta}$$

where ΔT_w is the temperature difference of two surfaces, and δ is the distance between the two isothermal surfaces. The apparent conductivity, which takes into account the free convection effect, is expressed in a general form:

$$\frac{k_e}{k} = C(\text{Gr}_\delta\, \text{Pr})^n \left(\frac{L}{\delta}\right)^m$$

Table 26 lists values of C, n, and m for a number of physical systems.

Mixed Convection and Convection from Rotating Bodies

In some forced convection situations, velocities of forced flow are comparable with the velocities of natural convective currents, for example, when the air velocity is of the order 1 ft/s. Under such conditions, a super-position of forced and natural convection, called mixed convection, must be considered. If the forced convection effects are very large, the influence of natural convection current may be negligible, and similarly when the natural convection forces are very strong, the forced convection effect may be neglected. The ratio of Gr/Re^2 gives a qualitative indication of the influence of buoyancy on forced convection. When $0.1 < \text{Gr}/\text{Re}^2 < 10$, mixed convection becomes significant. Regimes of free, forced, and mixed convection for flows through horizontal or vertical tubes are graphically shown in Fig. 26 and 27, respectively.

Heat transfer in a rotating system provides typical examples of mixed convection. When the rotational speed is low or when the difference between the surface temperature and fluid bulk temperature is large, free convection may dominate. When rotational speed is sufficiently large, the influence of free convection becomes relatively small, and the predominant mode of heat transfer is by forced convection. When Gr is of the same order of magnitude as Re^2, the combined effects of free and forced convection must be taken into account. Table 27 gives the formulas for determining the heat transfer of various rotating systems.

Thermal Radiation

The emission of thermal radiation is governed by the temperature of the emitting body. Although the physical mechanism of radiation is not completely understood, radiant energy is sometimes visualized as transported by electromagnetic waves or by photons. Neither viewpoint completely describes the nature of all observed phenomena.

Electromagnetic phenomena encompass many types of radiation, from short-wavelength cosmic and γ rays to long-wavelength radio waves (Fig. 28). Thermal radiation is the portion of the spectrum between wave-lengths 10^{-7} and 10^{-4} m. The visible spectrum runs from 3.9×10^{-7} to 7.8×10^{-7} m. The propagation velocity for all types of electromagnetic radiation in a vacuum is $c = \lambda\, v = 3 \times 10^8$ m/s, where λ is the wavelength and v is the frequency of the radiation. Radiation heat transfer differs in two important respects from the conductive and convective modes: no transfer medium is required, and the energy transferred is proportional to the fourth or fifth power of the temperatures of the emitting bodies.

Blackbody and the Basic Laws of Thermal Radiation

When radiant energy falls on a body, part may be absorbed, part reflected, and the remainder transmitted through the body. In mathematical form $\alpha + \rho + \tau = 1$, where α, ρ, and τ are absorptance, reflectance, and transmittance, respectively. They represent the fraction of total energy absorbed, reflected, and transmitted, respectively. For the majority of opaque solid materials encountered in engineering, except for extremely thin layers, practically none of the radiant energy is transmitted through the body. Under this condition, $\alpha + \rho = 1$. An arrangement that absorbs all radiant energy at all wavelengths and reflects none is called a blackbody. Although no material with $\alpha = 1$ and $\rho = 0$ exists, it is used as a standard of comparison for a real body absorbing and emitting radiation.

For a blackbody, total emissive power is described by Stefan-Boltzmann's law, and its directional distribution of radiation intensity is described by Lambert's cosine law. The spectral distribution of radiation intensity is given by Planck's distribution law and the wavelength for the maximum spectral radiation intensity by Wien's displacement law. Thermal radiation emitted from a black surface flows uniformly into the whole hemispherical space, as depicted in Fig. 29. A blackbody of surface area A, with uniform

Table 13 Temperature distribution and heat transfer in time interval θ

Schematic diagram	Temperature distribution, $\frac{T-T_0}{T_i-T_0}$	Initial energy storage, Q_i	Heat transfer in time interval θ, $\frac{Q}{Q_i}$
Semi-infinite solid	$\mathrm{erf}\left(\dfrac{x}{2\sqrt{a\theta}}\right)$	$\rho c A (T_i - T_0)$	$-\dfrac{2}{\sqrt{\pi}}\sqrt{a\theta}$
Infinite cylinder of radius r_0	$\displaystyle\sum_{n=1}^{\infty}\frac{1}{\beta_n}\exp(-\beta_n^2 a\theta/r_0^2)$ $J_0[\beta_n r/r_0]/J_1(\beta_n)$ β_n are the roots of $J_0(\beta)=0$	$\rho c \pi r_0^2 (T_i - T_0)$	$\displaystyle\sum_{n=1}^{\infty}\frac{-1}{\beta_n^2}\{1-\exp[-\beta_n^2 a\theta/r_0^2]\}$
Infinite plate of thickness $2L$	$\dfrac{2}{\pi}\displaystyle\sum_{n=0}^{\infty}\exp\left\{-\left[\left(n+\frac{1}{2}\right)\pi\right]^2\frac{a\theta}{L^2}\right\}$ $\dfrac{(-1)^n}{n+\frac{1}{2}}\cos\left(n+\frac{1}{2}\right)\dfrac{\pi x}{L}$	$\rho c A L (T_i - T_0)$	$\dfrac{2}{\pi^2}\displaystyle\sum_{n=0}^{\infty}\frac{-1}{\left(n+\frac{1}{2}\right)^2}$ $\left[1-\exp\left\{-\left[\left(n+\frac{1}{2}\right)\pi\right]^2\frac{a\theta}{L^2}\right\}\right]$
Sphere of radius r_0	$\dfrac{2r_0}{\pi r}\displaystyle\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\exp\left[-\frac{n^2\pi^2 a\theta}{r_0^2}\right]$ $\sin\left(\dfrac{n\pi r}{r_0}\right)$	$\dfrac{4}{3}\pi r_0^3\,\rho c(T_i - T_0)$	$\dfrac{6}{\pi^2}\displaystyle\sum_{n=1}^{\infty}\frac{-1}{n^2}\left[1-\exp\left(-\frac{n^2\pi^2 a\theta}{r_0^2}\right)\right]$

Note: T_i = initial temperature, $T_0 = T_\infty$, surface temperature; J_0 = Bessel function of the first kind with zero order; J_1 = Bessel function of the first kind with first order; erf = error function

Fig. 7 Temperature response, temperature gradient, and heating rate in a semi-infinite solid, $x \geq 0$, after sudden change in surface temperature from T_i when $\theta < 0$ to T_0 for $\theta \geq 0$, ($T_0 = T_\infty$)

LIVE GRAPH
Click here to view

amount it can emit. This law applies only to a few types of surfaces bearing gray surface characteristics.

Lambert's Cosine Law. Radiation intensity in a given direction is proportional to the cosine of the angle formed with the normal of the surface of emission:

$$i_{b\theta} = i_{bn}\cos\theta$$

When the surface radiation intensity follows the same cosine law of a black-body radiation, the normal emissivity is independent of the angle θ, and it is identical with the hemispherical emissivity, $\varepsilon_n = \varepsilon$.

Planck's Distribution Law. Spectral distribution of the radiation intensity of a blackbody is given by the equation:

$$i_{\lambda b} = \frac{2C_1}{\lambda^5(e^{C_2/\lambda T}-1)}$$

which gives the magnitude of the emitted energy at each wavelength in the radiation spectrum. This equation can be rewritten as:

$$\frac{i_{\lambda b}}{T^5} = \frac{2C_1}{(\lambda T)^5(e^{C_2/\lambda T}-1)}$$

which states that for a given value of (λT), the ratio of $i_{\lambda b}/T^5$ is the same at all temperatures, and their relationship can be represented by a single curve.

Stefan-Boltzmann's Law. Total emissive power of a blackbody is proportional to the fourth power of the absolute temperature of the surface. The hemispherical total emissive power is given by:

$$e_b = \int_0^\infty e_{\lambda b}d\lambda = \int_0^\infty \pi i_{\lambda b}d\lambda = \sigma T^4$$

where σ is the Stefan-Boltzmann constant.

Wien's Displacement Law. As temperature increases, the maximum blackbody radiation intensity shifts in the direction of the shorter wave-lengths. The wavelength for the maximum radiation intensity for temperature T can

temperature T_1, loses heat to its surroundings (assumed black) with uniform temperature T_2, at a rate given by the equation $Q = A\,\sigma\,(T_1^4 - T_2^4)$.

The blackbody hemispherical emissive power is π times the directional emissive power, which is normal to the surface and equal to the radiation intensity. A summary of the blackbody thermal properties is given in Table 28. Some basic laws of thermal radiation are briefly presented as follows.

Kirchhoff's Law. The ratio of the emissive power of a body in thermal equilibrium to that of a blackbody at the same temperature is equal to its absorptivity, expressed as:

$$\frac{e}{e_b} = \alpha \quad \text{or} \quad \varepsilon = \alpha$$

At a given temperature, the amount of incident radiation a body can absorb equals the

Fig. 8 Temperature distribution and heat flow in an infinite slab of thickness 2L at a temperature T_i that has its surface temperature suddenly changed to T_0. (x is measured from the surface.)
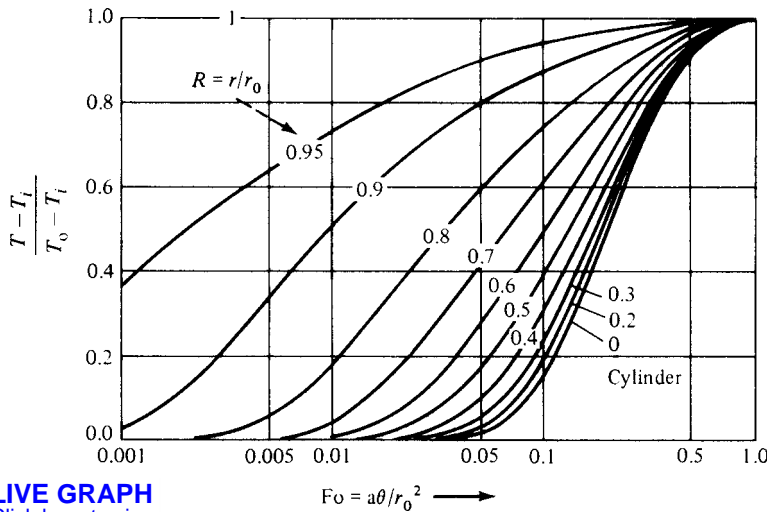
Fig. 9 Temperature response of a cylinder, $0 \leq r \leq r_0$, after sudden change in external surface temperature at $r = r_0$ from T_i when $\theta < 0$ to T_0 for $\theta \geq 0$
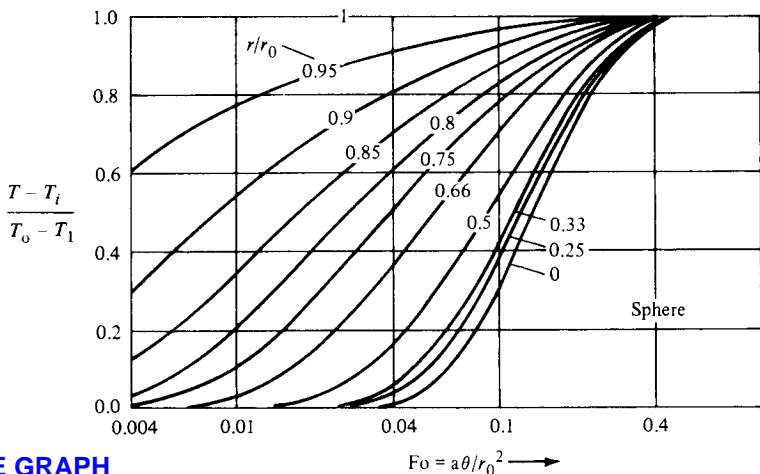
Fig. 10 Temperature response of a sphere, $0 \leq r \leq r_0$, after sudden change in external surface temperature at $r = r_0$ from T_i when $\theta < 0$ to T_0 for $\theta \geq 0$

be calculated from the equation $\lambda_{max} T = C_3$. The numerical values for the various radiation constants mentioned previously are given in Table 29.

Radiation Shape Factor

In most practical problems involving radiation, the intensity of thermal radiation moving between surfaces is not greatly affected by intervening media. Unless the temperature is sufficient to cause ionization or dissociation, monatomic and most diatomic gases as well as air are transparent. Most industrial surfaces can be considered diffuse emitters and radiation reflectors in a heat-transfer analysis. Thus, a key problem in calculating radiation heat transfer is to determine the fraction of the total diffuse radiation leaving one surface and intercepted by another. The fraction of diffusely distributed radiation leaving a surface A_i and reaching surface A_j is the radiation shape factor F_{i-j}. The first subscript denotes the emitting surface, while the second subscript denotes the receiving surface.

For two black surfaces A_1 and A_2, as shown in Fig. 30, the net rate of transfer between A_1 and A_2 can be written as:

$$Q_{1-2} = \sigma(T_1^4 - T_2^4)A_1 F_{1-2}$$
$$= \sigma(T_1^4 - T_2^4)A_2 F_{2-1}$$

where:

$$F_{1-2} = \frac{1}{\pi A_1}\int_{A_1}\int_{A_2}\frac{\cos\theta_1\cos\theta_2}{S^2}dA_1 dA_2$$
$$F_{2-1} = \frac{1}{\pi A_2}\int_{A_1}\int_{A_2}\frac{\cos\theta_1\cos\theta_2}{S^2}dA_1 dA_2$$

The shape factor F_{i-j} for diffuse radiation is a geometric property of two surfaces involved. One of the important properties of the shape factor is known as the reciprocal relationship, $A_1 F_{1-2} = A_2 F_{2-1}$. A summary of view-factor and energy-exchange relationships is presented in Table 30. Radiation view factors between two differential areas are described in Fig. 31. View factors for finite areas are given in Ref 6, 10–13, 50, 59, 66.
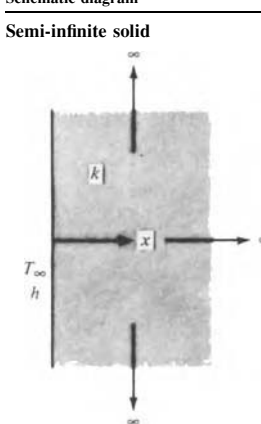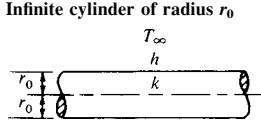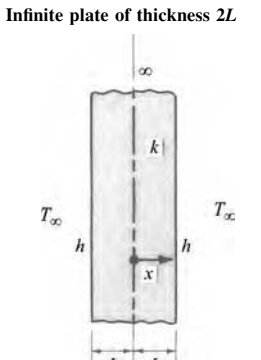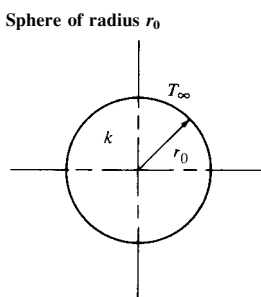
Radiation Exchange in Black Enclosure

Consider the energy transfer within an enclosure composed of N black surfaces that are individually isothermal (Fig. 32). The net energy loss due to radiation at the kth surface or the energy required from an external source through conduction, convection, or internal generation to maintain the kth surface at T_k is:

$$Q_k = \sigma A_k \sum_{j=1}^{N}(T_k^4 - T_j^4)F_{k-j}$$

which is the net energy transferred from A_k to each surface.

Table 14 Temperature distribution and heat transfer in solids with finite conductivity and transfer coefficient

Schematic diagram	Temperature distribution, $\frac{T-T_\infty}{T_i-T_\infty}$	Initial energy storage per unit area, Q_i	Heat transfer in time interval θ, $\frac{Q}{Q_i}$
Semi-infinite solid 	$\operatorname{erf}\left(\dfrac{x}{2\sqrt{a\theta}}\right) + \exp\left[\dfrac{hx}{k} + \dfrac{h^2 a\theta}{k^2}\right]$ $\left[1 - \operatorname{erf}\left(\dfrac{x}{2\sqrt{a\theta}} + \dfrac{h}{k}\sqrt{a\theta}\right)\right]$		
Infinite cylinder of radius r_0 	$2\sum_{n=1}^{\infty}\dfrac{J_1(\beta_n)J_0(\beta_n r/r_0)}{J_0^2(\beta_n)+J_1^2(\beta_n)}$ $\exp(-\beta_n^2 a\theta/r_0^2)$ where β_n are the roots of $\beta = \dfrac{hr_0 J_0(\beta)}{kJ_1(\beta)}$	$\rho c\pi r_0^2(T_i - T_\infty)$	$4\sum_{n=1}^{\infty}\dfrac{1}{\beta_n^2}\dfrac{-J_1^2(\beta_n)}{J_0^2(\beta_n)+J_1^2(\beta_n)}$ $[1-\exp(-\beta_n^2 a\theta/r_0^2)]$
Infinite plate of thickness $2L$ 	$4\sum_{n=1}^{\infty}\left(\dfrac{\sin\ \beta_n}{2\beta_n+\sin\ 2\beta_n}\right)e^{-\beta_n^2 a\theta/L^2}$ $\cos\left(\dfrac{\beta_n x}{L}\right)$ where β_n are the roots of $\beta\tan\beta = hL/k$	$\rho c_\rho L(T_i - T_\infty)$	$4\sum_{n=1}^{\infty}\dfrac{-\sin^2\ \beta_n}{2\beta_n^2+\beta_n\sin\ 2\ \beta_n}$ $[1-\exp(-\beta_n^2 a\theta/L^2)]$
Sphere of radius r_0 	$4\sum_{n=1}^{\infty}\dfrac{\sin\ \beta_n - \beta_n\cos\ \beta_n}{2\beta_n-\sin\ 2\beta_n}$ $\exp\left[-\beta_n^2 a\theta/r_0^2\right]$ $\dfrac{\sin(\beta_n r/r_0)}{(\beta_n r/r_0)}$ where β_n are the roots of $\beta\cos\beta = 1 - \dfrac{hr_0}{k}$	$\dfrac{4}{3}\pi r_0^3\rho c(T_i - T_\infty)$	$12\sum_{n=1}^{\infty}\dfrac{-(\sin\ \beta_n - \beta_n\cos\ \beta_n)^2}{\beta_n^3(2\beta_n-\sin\ 2\beta_n)}$ $[1-\exp(-\beta_n^2 a\theta/r_0^2)]$

solution of N systems of equations. Table 32 gives some formulas for diffuse radiant heat exchange for systems commonly encountered in engineering applications.

Radiation Shields

Radiation heat transfer can be substantially reduced by means of radiation shields, usually consisting of a surface opaque to radiant transmission with high thermal conductivity and low emissivity. Shielding effectiveness is increased if the shields are placed between mutually radiating surfaces or over a surface against radiation heating. The configuration in Fig. 34 shows two large parallel walls separated by N radiation shields. The heat-flow rate by radiation per unit area, q, at steady state is given by:

$$q = \frac{\sigma(T_1^4 - T_2^4)}{1/\varepsilon_1 + 1/\varepsilon_2 - 1 + \sum_{n=1}^{N}(1/\varepsilon_{n1} + 1/\varepsilon_{n2} - 1)}$$

where ε_{n1} and ε_{n2} are the emissivities on both sides of a typical shield, n. In most instances, the ε is the same on both sides of each shield, and all the shields have the same ε. If all the shield emissivities are represented as ε_s, then q becomes:

$$q = \frac{\sigma(T_1^4 - T_2^4)}{1/\varepsilon_1 + 1/\varepsilon_2 - 1 + N(2/\varepsilon_s - 1)}$$

If wall emissivities are the same as the shield emissivities, that is, $\varepsilon_1 = \varepsilon_2 = \varepsilon_s$, then the previous expression reduces to:

$$q = \frac{\sigma(T_1^4 - T_2^4)}{(N+1)(2/\varepsilon_s - 1)}$$

where q decreases as $1/(N+1)$ as the number of shields increases. The expression for heat flow through a series of concentric cylindrical or spherical radiation shields is similar to that for flat plates, as shown in Fig. 35. Thus:

$$Q = \frac{A_1\sigma(T_1^4 - T_2^4)}{1/\varepsilon_1 + (A_1/A_2)(1/\varepsilon_2 - 1) + \sum_{n=1}^{N}(A_1/A_{sn})(1/\varepsilon_{n1} + 1/\varepsilon_{n2} - 1)}$$

where the walls A_1 and A_2 and all the shields A_{sn} are diffuse. If all the shield emissivities are the same and equal to ε_s, then:

$$Q = \frac{A_1\sigma(T_1^4 - T_2^4)}{1/\varepsilon_1 + 1/\varepsilon_s - 1 + \sum_{n=1}^{N-1}(A_1/A_{sn})(2/\varepsilon_s - 1) + (A_1/A_{sN})(1/\varepsilon_s + 1/\varepsilon_2 - 1)}$$

REFERENCES

1. L.C. Thomas, *Fundamentals of Heat Transfer,* Prentice Hall, Englewood Cliffs, NJ, 1980
2. D.Q. Kern and A.D. Kraus, *Extended Surface Heat Transfer,* 2nd ed., McGraw-Hill, New York, 1972
3. V.S. Arpaci, *Conduction Heat Transfer*, Addison-Wesley, Reading, MA, 1966

Radiation Exchange in Gray Enclosure

An enclosure with N discrete surface areas is shown in Fig. 33. A complex radiative exchange occurs inside the enclosure when radiation leaves a surface, is partially reflected by other surfaces, and is then re-reflected many times within the enclosure, with partial absorption at each surface contact. An analysis can be formulated with the net-radiation method. Based on this method, the relationships between energy flux at temperature in diffuse gray enclosures are obtained. They are summarized in Table 31 under various boundary conditions.

As can be seen from Table 31, the desired quantities of the external heat input, Q_k, the radiosity, J_k, and the surface temperature, T_k, at a particular surface k are obtained from the

Fig. 11 Dimensionless temperature distribution in a semi-infinite slab subjected to a sudden change in environmental temperature

10. R. Siegel and J.R. Howell, *Thermal Radiation Heat Transfer,* 2nd ed., McGraw-Hill, New York, 1980
11. B.T.F. Chung and M.H.N. Naraghi, A Simpler Formulation for Radiative View Factors from Spheres to a Class of Axisymmetric Bodies, *J. Heat Transf.,* Vol 104, 1982, p 201
12. M.H.N. Naraghi and B.T.F. Chung, Radiation Configuration Factor Between Disks and a Class of Axisymmetric Bodies, *J. Heat Transf.,* Vol 104, 1982, p 426
13. *ASM Handbook of Engineering Mathematics,* American Society for Metals, 1983

SELECTED REFERENCES

- J.T. Anderson and O.A. Saunders, Convection from an Isolated Heated Horizontal Cylinder Rotating about its Axis, *Proc. Roy. Soc.,* Vol 217A, 1953, p 555
- C.K. Brown and W.H. Gauvin, Combined Free and Forced Convection, Parts I & II, *Can. J. Chem. Eng.,* Vol 43, 1965, p 306, 313
- A.J. Buschman and C.M. Pittmann, "Configuration Factors for Exchange of Radiant Energy between Axisymmetric Sections of Cylinders, Cones and Hemispheres and Their Base," NASA TN D-944, 1961
- A.J. Chapman, *Heat Transfer,* 3rd ed., MacMillan, New York, 1974
- B.T.F. Chung and P.S. Sumitra, Radiation Shape Factor from Plane Point Sources, *J. Heat Transf.,* Vol 94, 1972, p 328
- B.T.F. Chung and M.H.N. Naraghi, Some Exact Solutions for Radiation View Factors for Spheres, *AIAA J.,* Vol 19, 1981, p 1077
- B.T.F. Chung, M.M. Kermani and M.H.N. Naraghi, "A Formulation of Radiation View Factors from Conical Surfaces," AIAA Paper 83–0156, 1983
- E.C. Cobb and O.A. Saunders, Heat Transfer from a Rotating Disc, *Proc. Rov. Soc.,* Vol 236A, 1956, p 343
- R.G. Deissler Analysis of Turbulent Heat Transfer, Mass Transfer and Friction in Smooth Tubes at High Prandtl Numbers," NASA TR 1210, 1955
- D.F. Dipprey and R.H. Sabersky, Heat and Momentum Transfer in Smooth and Rough Tubes at Various Prandtl Numbers, *Int. J. Heat Mass Transf.,* Vol 6, 1963, p 329
- E. Eckert and T. Jackson, "Analysis of Turbulent Free Convection Boundary Layer on Flat Plate," NASA TN 2207, 1950
- E.R.G. Eckert and R.M. Drake, *Heat and Mass Transfer,* 2nd ed., McGraw Hill Book Co., New York, 1959
- A.J. Ede, Advances in Free Convection, *Advances in Heat Transfer,* Vol 4, Academic Press, 1967
- R.M. Frand, Heat Transfer by Forced Convection from a Cylinder to Water in

4. B.T.F. Chung, "Heat Transfer Processes Notes," University of Akron, unpublished work
5. J.P. Holman, *Heat Transfer,* 5th ed., McGraw-Hill, St. Louis, 1981
6. H.Y. Wong, *Handbook of Essential Formulas and Data on Heat Transfer for Engineers,* Longman Group Ltd., London, 1977
7. B.V. Karlekar and R.M. Desmond, *Engineering Heat Transfer,* West Publishing, New York, 1977
8. M.W. Kays and M.E. Crawford, *Convective Heat and Mass Transfer,* 2nd ed., McGraw-Hill, New York, 1980
9. R.V. Andrews, Solving Conductive Heat Transfer Problems with Electrical-

Analogue Shape Factors, *Chemical Engineering Progress,* Vol 5 (No. 2), 1955, p 67

Fig. 12 Midplane temperature for an infinite plate of thickness 2L

LIVE GRAPH
Click here to view

$\frac{k}{hL}$
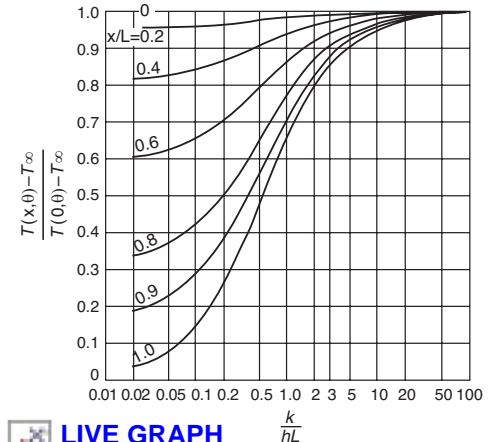
Fig. 13 Temperature as a function of center temperature in an infinite plate of thickness 2L

LIVE GRAPH
Click here to view

$\frac{ha\theta}{k^2}$
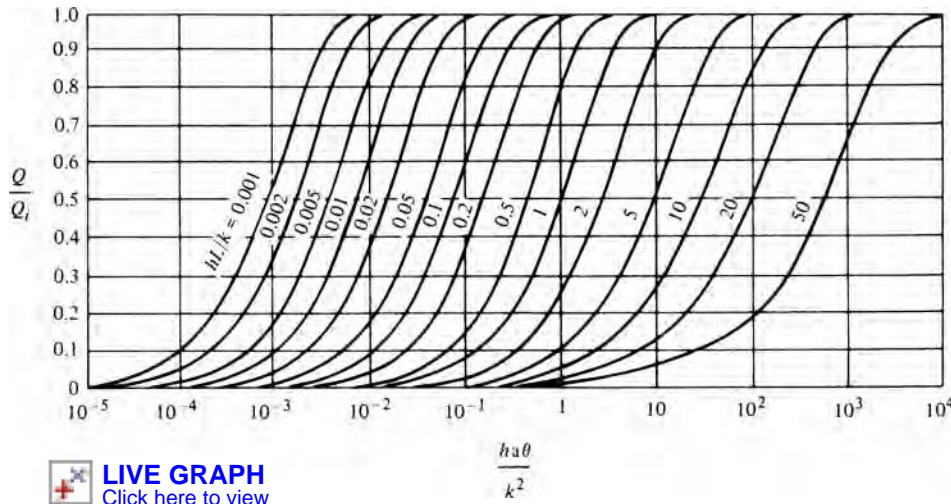
Fig. 14 Dimensionless heat loss Q/Q_i of an infinite plate of thickness 2L with time

LIVE GRAPH
Click here to view

$\frac{k}{hr_0}$
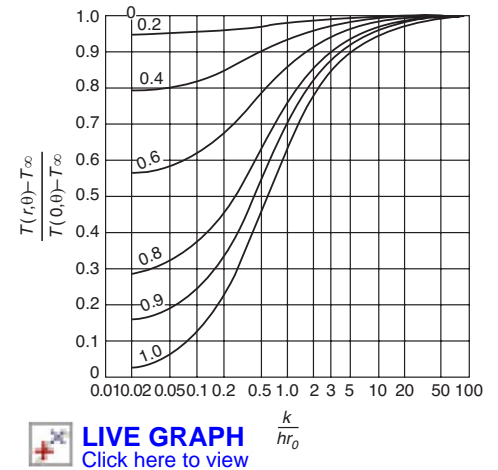
Fig. 16 Temperature as a function of axis temperature in an infinite cylinder of radius r_0
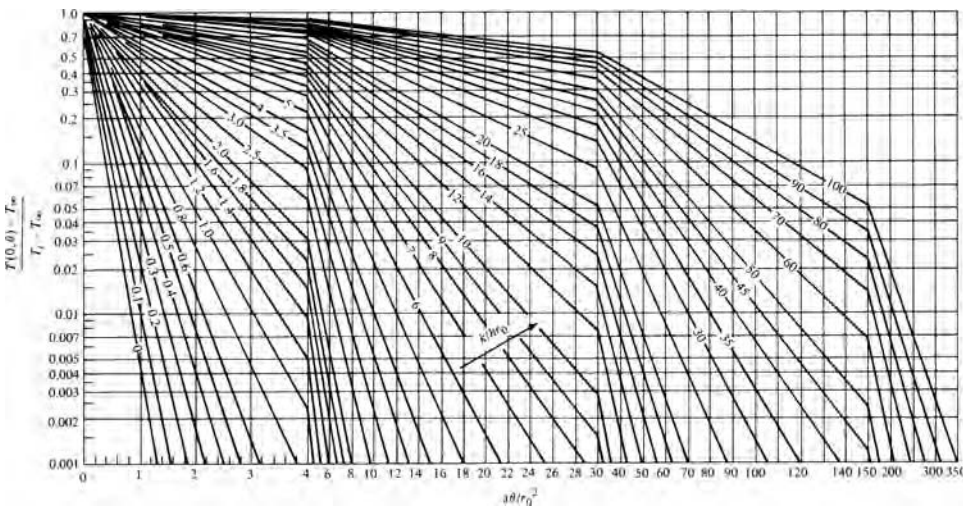
Fig. 15 Axis temperature for an infinite cylinder of radius r_0

Crossflow, *Int. J. Heat Mass Transf.*, Vol 8, 1965, p 995
- A. Feingold and K.G. Gupta, New Analytical Approach to the Evaluation of Configuration Factors in Radiation from Spheres and Infinitely Long Cylinders, *J. Heat Transf.*, Vol 92, 1970, p 67
- M. Fishenden and O.A. Saunders *An Introduction to Heat Transfer,* Oxford University Press, 1950
- E.D. Grimson, Correlation and Utilization of New Data on Flow Resistance and Heat Transfer for Cross Flow of Gases over Tube Banks, *Trans. ASME,* Vol 59, 1937, p 583
- H. Grober, S. Erk and U. Grigull *Fundamentals of Heat Transfer,* McGraw-Hill, New York, 1961
- E. Hahne and U. Grigull, Formfakter and Formweider stand der Stationaun Mehrdimensionalen Warmeleitung, *Int. J. Heat Mass Transf.*, Vol 18, 1975, p 251
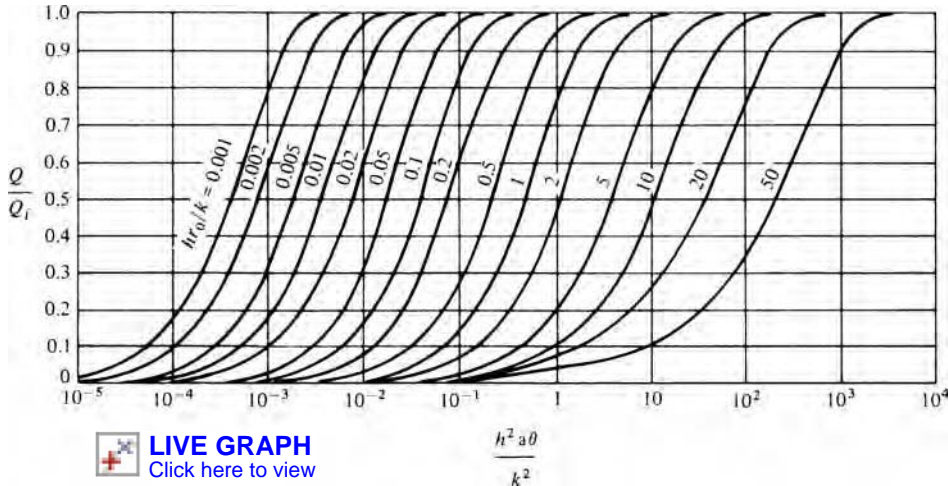
Fig. 17 Dimensionless heat loss Q/Q_i of an infinite cylinder of radius r_0 with time

Fig. 18 Center temperature for a sphere of radius r_0

• D.C. Hamilton and W.R. Morgan "Radiant Interchange Configuration Factors," NASA TN 2836, 1952
• M.P. Heisler, Temperature Charts for Induction and Constant Temperature Heating, *Trans. ASME,* Vol 69, 1947, p 227
• R.G. Hering and R.J. Grosh, Laminar Combined Convection from a Rotating Cone, *J. Heat Transf.*, Vol 85, 1963, p 29
• H.C. Hottel, Radiant Heat Transmis Between Surfaces Separated by Non-absorbing Media, *Trans. ASME,* Vol 53, 1931, p 265
• F.P. Incropera and D.P. Dewitt, *Fundamentals of Heat Transfer,* John Wiley & Sons, New York, 1981
• M. Jakob, *Heat Transfer,* Vol 1, John Wiley & Sons, New York, 1957
• M. Jakob, *Heat Transfer,* Vol 2, John Wiley & Sons, New York, 1957
• M. Jakob, Heat Transfer and Flow Resistance in Cross Flow of Gases over Tube Banks, *Trans. ASME,* Vol 60, 1938, p 384
• W.M. Kays and R.K. Lo, "Basic Heat Transfer and Flow Friction Data for Gas Flow Normal to Banks of Staggered Tubes: Use of a Transient Technique, Stanford University," Tech. Rep. 15, Navy Contract N6-ONR 251 T.O.6, 1952
• W.M. Kays and I.S. Bjorklund, Heat Transfer from a Rotating Cylinder with and without Cross Flow, *Trans. ASME*, Vol 80, 1958, p 70
• J.G. Knudsen and D.L. Katz, *Fluid Dynamics and Heat Transfer,* McGraw-Hill, St. Louis, 1958
• H. Kramers, Heat Transfer from Spheres to Flowing Media, *Physica,* Vol 12, 1946, p 61
• F. Kreith, L.R. Roberts, J.A. Sullivan, and S. N. Sinha, Convection Heat Transfer and Flow Phenomena of Rotating Spheres, *Int. J. Heat Transf.*, Vol 6, 1963, p 881
• F. Kreith, Convective Heat Transfer in Rotating Systems, *Advances in Heat Transfer,* Vol 5, Academic Press, 1968
• F. Kreith, *Principles of Heat Transfer,* 3rd ed., Harper & Row, New York, 1973

Fig. 19 Temperature as a function of center temperature for a sphere of radius r_0

Fig. 20 Dimensionless heat loss Q/Q_i of a sphere of radius r_0 with time

$$\frac{T(x, y, \theta) - T_\infty}{T_i - T_\infty} = S(x, \theta)P(v, \theta)$$

(a) Semi-infinite plate

$$\frac{T(x, y, \theta) - T_\infty}{T_i - T_\infty} = P(x, \theta)P(v, \theta)$$

(b) Infinite rectangular bar

$$\frac{T(x, y, z, \theta) - T_\infty}{T_i - T_\infty} = S(x, \theta)P(y, \theta)P$$

(c) Semi-infinite rectangular bar

$$\frac{T(x, y, z, \theta) - T_\infty}{T_i - T_\infty} = P(x, \theta)P(y, \theta)P(z, \theta)$$

(d) Rectangular parallelpiped

$$\frac{T(r, x, \theta) - T_\infty}{T_i - T_\infty} = S(x, \theta)C(r, \theta)$$

(e) Semi-infinite cylinder

$$\frac{T(r, x, \theta) - T_\infty}{T_i - T_\infty} = P(x, \theta)C(r, \theta)$$

(f) Short cylinder

$$\frac{T(x, y, \theta) - T_\infty}{T_i - T_\infty} = S(x, \theta)S(v, \theta)$$

(g) Quarter-infinite solid

$$\frac{T(x, y, z, \theta) - T_\infty}{T_i - T_\infty} = S(x, \theta)S(y, \theta)S(z, \theta)$$

(h) Eighth-infinite solid

Fig. 21 Product solutions for temperatures in multidimensional systems

- S.S. Kutateladze and V.M. Borishanskii, *A Concise Encyclopedia of Heat Transfer,* Pergamon Press, 1966
- E.J. LeFevre and A.J. Ede, Laminar Free Convection from the Outer Surface of a Vertical Circular Cylinder, *Proc. Ninth Congress, Applied Mech. Paper 1167,* 1956
- H. Leuenberger and R.A., Person, "Compilation of Radiation Shape Factors for Cylindrical Assemblies," ASME Paper 56-A-144, 1956

- J.H. Lienhard, *A Heat Transfer Textbook,* Prentice Hall, Englewood Cliffs, NJ, 1981
- W.H. McAdams, *Heat Transmission,* 3rd ed., McGraw-Hill, New York, 1954
- B. Metais and E.R.G. Eckert, Forced, Mixed and Free Convection Regimes, *ASME J. Heat Transf.*, Vol 86, 1964, p 295
- V.T. Morgan, The Overall Convective Heat Transfer from Smooth Circular Cylinders, *Advances in Heat Transfer,* Vol 11, Vol Academic Press, 1975

- M.N. Ozisik, *Boundary Value Problems of Heat Conduction,* International Textbook Co., Scranton, PA, 1968
- S. Ostrach, "An Analysis of Laminar Free Convective Flow and Heat Transfer about a Flat Plate Parallel to the Direction of the Generating Body Force," NASA TN 2635, 1952
- B.S. Petukhov and V.N. Popov, Theoretical Calculation of Heat Exchange and Frictional Resistance in Turbulent Flow in Tubes of an Incompressible Fluid with Variable

Table 15 Order of magnitude of convective heat-transfer coefficients, *h*

Means of heat transfer	Btu/h·ft²·°F	W/m²·K
Air, free convection	1–5	6–30
Superheated steam or air, forced convection	5–50	30–300
Oil, forced convection	10–300	60–1800
Water, forced convection	50–2000	300–6000
Water, boiling	500–10000	3000–60000
Steam, condensing	1000–20000	6000–120,000

Table 16 Physical interpretation of some dimensionless groups

Name of the dimensionless group	Group	Physical interpretation
Biot (Bi)	$\dfrac{hL_c}{k}$	$\dfrac{\text{Internal resistance to heat conduction}}{\text{External resistance to heat conduction}}$
Eckert (E)	$\dfrac{u_\infty^2}{c_p(T_w - T_\infty)}$	$\dfrac{\text{Kinetic energy}}{\text{Thermal energy}}$
Euler (Eu)	$\dfrac{p}{\rho u_\infty^2}$	$\dfrac{\text{Pressure forces}}{\text{Inertia forces}}$
Fourier (Fo)	$\dfrac{k\theta}{\rho c_p L_c^2}$	$\dfrac{\text{Dimensionless time for transient}}{\text{Conduction}}$
Froude (Fr)	$\dfrac{u_\infty^2}{L_c g}$	$\dfrac{\text{Inertia forces}}{\text{Gravity forces}}$
Grashof (Gr)	$\dfrac{gL_c^3\beta(T_w - T_\infty)}{v^2}$	$\dfrac{\text{(Buoyancy forces) (Inertia forces)}}{(\text{Viscous forces})^2}$
Lewis (Le)	$\dfrac{D\rho_{cp}}{k}$	$\dfrac{\text{Mass diffusivity}}{\text{Thermal diffusivity}}$
Mach (M)	$\dfrac{u_\infty}{u_c}$	$\dfrac{\text{Velocity}}{\text{Sonic velocity}}$
Nusselt (Nu)	$\dfrac{hL_c}{k}$	Ratio of temperature gradients
Peclet (Pe)	$\dfrac{c_p\rho u_\infty L_c}{k}$	$\dfrac{\text{Convective heat transfer}}{\text{Conductive heat transfer}}$
Prandtl (Pr)	$\dfrac{\mu c_p}{k}$	$\dfrac{\text{Momentum diffusivity}}{\text{Thermal diffusivity}}$
Rayleigh (Ra)	$\dfrac{gL_c^3\beta(T_w - T_\infty)}{\text{nu}\alpha}$	$\dfrac{\text{Forces due to buoyancy and inertia}}{\text{Forces due to viscosity and thermal diffusion}}$
Reynolds (Re)	$\dfrac{\rho\mu_\infty L_c}{\mu}$	$\dfrac{\text{Inertia forces}}{\text{Viscous forces}}$
Schmidt (Sc)	$\dfrac{\mu}{\rho D}$	$\dfrac{\text{Momentum diffusivity}}{\text{Mass diffusivity}}$
Sherwood (Sh)	$\dfrac{h_D L_c}{D}$	Ratio of concentration gradients
Stanton (St)	$\dfrac{h}{c_p\rho u_\infty}$	$\dfrac{\text{Heat transfer at wall}}{\text{Convective heat transfer}}$

Table 17 Key differential equations of fluid flow

Continuity equation	$\nabla \cdot \mathbf{V} = 0$

Equation of motion

Free convection

$$\rho\frac{d\mathbf{V}}{d\theta} = \mu\nabla^2\mathbf{V} - \rho g\beta(T - T_\infty)$$

Forced convection

$$\rho\frac{d\mathbf{V}}{d\theta} = -\nabla_p + \mu\nabla^2\mathbf{V} + \rho g$$

Energy equation

$$\rho c_p\frac{dT}{d\theta} = k\nabla^2 T + \mu\overline{\Phi}$$

$\dfrac{d}{d\theta}$ = substantial time derivative, $\dfrac{d}{d\theta} = \dfrac{\partial}{\partial\theta} + \mathbf{V}\cdot\nabla$

$\mathbf{V} = \mathbf{u_i} + \mathbf{v_j} + \mathbf{w_k}$, velocity field

$\overline{\Phi}$ = dissipation function,

e.g., $\overline{\Phi} = 2\left[\left(\dfrac{\partial u}{\partial x}\right)^2 + \left(\dfrac{\partial v}{\partial y}\right)^2 + \left(\dfrac{\partial w}{\partial z}\right)^2\right] + \left(\dfrac{\partial v}{\partial x} + \dfrac{\partial u}{\partial y}\right)^2 + \left(\dfrac{\partial w}{\partial y} + \dfrac{\partial v}{\partial z}\right)^2 + \left(\dfrac{\partial u}{\partial z} + \dfrac{\partial w}{\partial x}\right)^2 - \dfrac{2}{3}\left(\dfrac{\partial u}{\partial x} + \dfrac{\partial v}{\partial y} + \dfrac{\partial w}{\partial z}\right)^2$ for Cartesian coordinates

μ = dynamic viscosity; β = coefficient of volumetric thermal expansion

Physical Properties, *Trans. High Temp.* Vol 17 (No. 1), 1963

- R.W. Powell, C.Y. Ho, and P.E. Lidey, "Thermal Conductivity of Selected Materials," NSRDS-NBS 8, Washington, D.C., U.S. Department of Commerce, 1966
- W.M. Rohsenow and J.P. Hartnett, *Handbook of Heat Transfer,* McGraw-Hill, New York, 1973
- R. Rudenberg, Die Ausbreitung der Luft-und Erdfelder um Hocha-pannungaleitungen besonders bei Erd und Kurzschterssen, *Electrotech. Z.*, Vol 46, 1945, p 1342
- P.J. Schneider, *Conduction Heat Transfer,* Addison-Wesley, Reading, MA, 1955
- P.J. Schneider, *Temperature Response Charts,* John Wiley & Sons, New York, 1963
- E.M. Sparrow, T.M. Hallman, and R. Seigel, Turbulent Heat Transfer in the Thermal Entrance Region of a Pipe with Uniform Heat Flux, *Appl. Sci. Res.,* Vol A7, 1957, p 37
- E.M. Sparrow and R.D. Cess, *Radiation Heat Transfer,* Wadsworth Publishing Co., Englewood Cliffs, NJ, 1966
- Y.S. Touloukian, R.W. Powell, C.Y. Ho, and P.G. Klemens, *Thermophysical Properties of Matter,* Vol 1–3, IFI/Plenum Data Corporation, New York, 1970
- R.L. Webb, A Critical Evaluation of Analytical Solutions and Reynolds Analogy Equations for Turbulent Heat and Mass Transfer in Smooth Tubes, *Warme-und Stoffubertragung,* Bd. 4, 1971, p 197
- S. Whitaker, Forced Convection Heat Transfer Correlation for Flow in Pipes, Past Flat Plate, Single Cylinders, Single Spheres and Flow in Packed Beds and Tube Bundles, *AICHE J.*, Vol 18, 1972, p 361

Table 18 Analogy formula for Stanton number of tube flow

Type of analogy	Stanton No., $St = \frac{Nu}{RePr}$
Reynolds	$\dfrac{f}{8}$
Prandtl	$\dfrac{\frac{f}{8}}{1 + 5\sqrt{\frac{f}{8}}(Pr - 1)}$
Von Karman	$\dfrac{\frac{f}{8}}{1 + 5\sqrt{\frac{f}{8}}\left\{Pr - 1 + \ln\left[1 + \frac{5}{6}(Pr - 1)\right]\right\}}$
Colburn's	$\dfrac{f}{8} Pr^{-2/3}$
Deissler and Webb	$\dfrac{\frac{f}{8}}{1.07 + 9\sqrt{\frac{f}{8}}(Pr - 1)\, Pr^{-1/4}}$
Petukhov and Popov	$\dfrac{\frac{f}{8}}{1.07 + 12.7\sqrt{\frac{f}{8}}(Pr^{2/3} - 1)}$

The friction factor here is defined as $f = \tau_w / (1/2\rho\, u_b^2)$. Another form of friction factor, known as Faning friction factor, is encountered frequently in engineering. It is one-quarter of the present value.
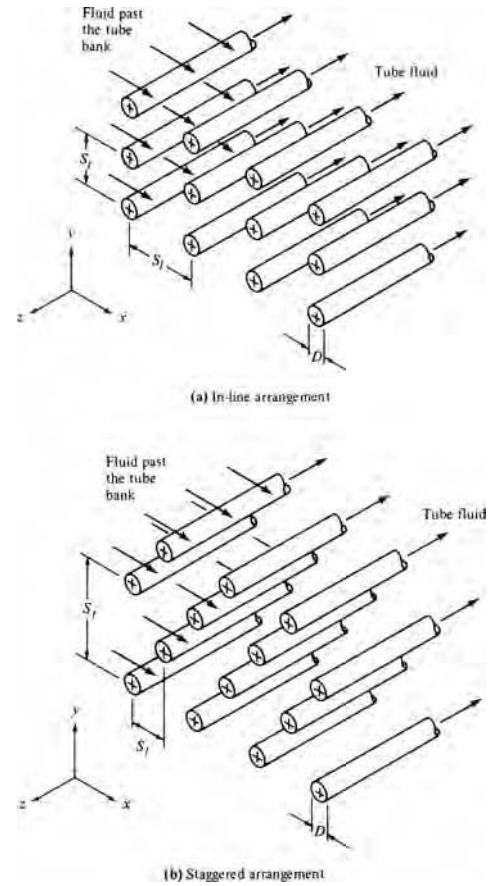
Fig. 22 Nomenclature for flow across tube banks

Table 19 Equations for friction coefficient at various flow and geometric conditions

Flow and geometry	Friction coefficient equation	Restriction and remarks
Laminar flow in either smooth or rough pipe	$f = \dfrac{64}{Re}$	$Re < 2{,}000$
Turbulent flow in smooth pipe and parallel planes	$\dfrac{1}{\sqrt{\frac{f}{4}}} = 4.0\log\left(Re\sqrt{\frac{f}{4}}\right) - 0.4$	$Re > 3{,}000$
Turbulent flow in rectangular, triangular, and trapezoidal conduit	$\dfrac{f}{4} = 0.079\, Re^{-0.25}$	$Re < 100{,}000$
Fully turbulent flow in rough pipe	$\dfrac{1}{\sqrt{\frac{f}{4}}} = 4\log\dfrac{D}{e} + 2.28$	$\dfrac{D/e}{Re\sqrt{\frac{f}{4}}} > 0.01$
Transition flow in rough pipe	$\dfrac{1}{\sqrt{\frac{f}{4}}} = 4\log\dfrac{D}{e} + 2.28 - 4\log$ $\left(1 + 4.67\dfrac{D/e}{Re\sqrt{\frac{f}{4}}}\right)$	
Laminar flow over flat plate	$f_x = 0.664\, Re_x^{-0.5}$	$Re_x < 5 \times 10^5$
Turbulent flow over flat plate	$f_x = 0.0592\, Re_x^{-0.2}$	$5 \times 10^5 < Re_x < 10^7$
Flow across tube banks in line arrangement	$f = 4\left[0.044 + \dfrac{0.08(S_l/D)}{[(S_t - D)/D]^{[10.43 + (1.13D/S_l)]}}\right]Re_{max}^{-0.15}$	$2{,}000 < Re < 40{,}000$(a)
Flow across tube bank in staggered arrangement	$f = 4\left[0.25 + \dfrac{0.1175}{\left(\frac{S_t - D}{D}\right)^{1.08}}\right]Re_{max}^{-0.16}$	$2{,}000 < Re < 40{,}000$(a)

Note: Re_{max} = maximum Reynolds number = $\dfrac{\rho u_{max} D}{\mu}$. (a) See Fig. 22 for notations.

Fig. 23 Friction factor

LIVE GRAPH
Click here to view
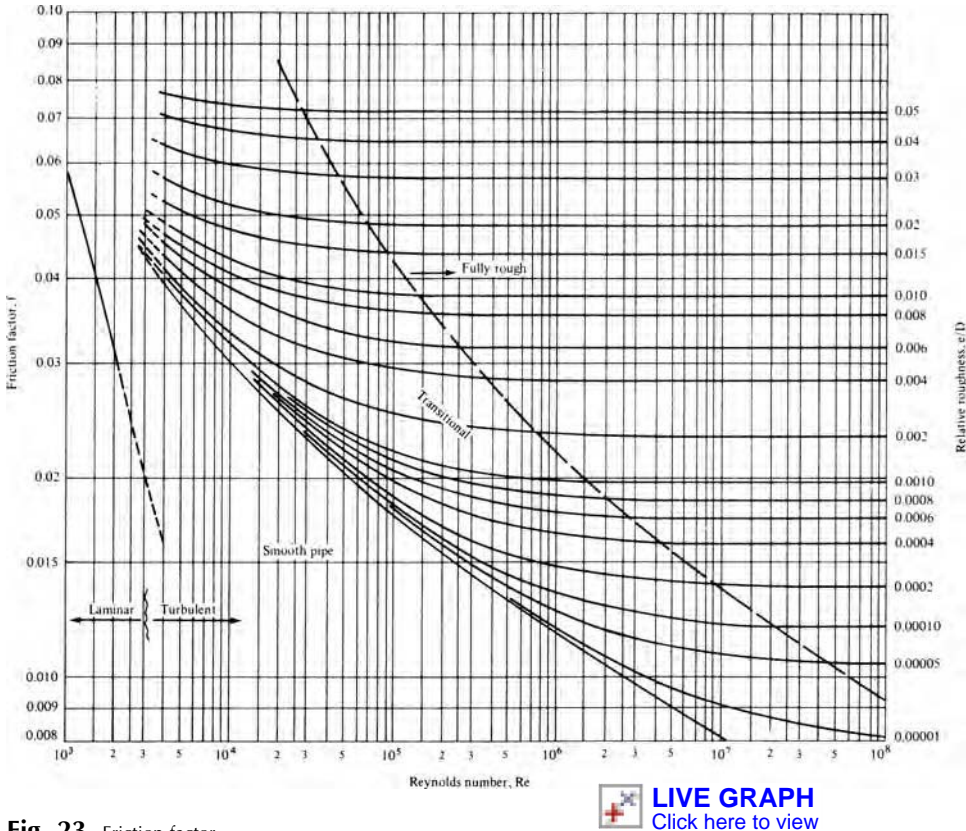
Table 20   Correlations for forced convection from external flow

Geometry	Correlation for Nu	Restrictions
		Laminar
Flow over flat plate	$0.664 \, Re_L^{1/2} \, Pr^{1/3}$	$Re < 5 \times 10^5$
	$(0.036 \, Re_L^{4/5} - 836)Pr^{1/3}$	$Re > 5 \times 10^5$
	$0.036 \, (Re_L^{0.8} \, Pr^{0.43} - 17,400) + 297 \, Pr^{1/3}$	$10^5 < Re_L < 5.5 \times 10^6 \quad 0.7 < Pr < 380$
	$0.036 \, Pr^{0.42} \, (Re_L^{0.8} - 9,200)(\mu_\infty/\mu_w)^{0.14}$	$0.26 < \mu_\infty/\mu_w < 3.5$
Inclined plate	$0.86 \, Re^{1/2} \, Pr^{1/3}$	
Flow across cylinder	$[0.8327 - 0.4 \ln (RePr)]^{-1}$	$RePr < 0.2$
	$0.3 + 0.62\phi \, [1 + 3.92 \times 10^{-4} \, Re^{5/8}]^{4/5}$	
	$\phi = Re^{1/2} \, Pr^{1/3} \, [1 + (0.4/Pr)^{2/3}]^{-1/4}$	$0.2 < RePr$
	$0.989 \, Re^{0.33} \, Pr^{1/3}$	$0.4 < Re < 4$
	$0.911 \, Re^{0.385} \, Pr^{1/3}$	$4 < Re < 40$
	$0.683 \, Re^{0.466} \, Pr^{1/3}$	$40 < Re < 4000$
	$0.193 \, Re^{0.618} \, Pr^{1/3}$	$4000 < Re < 40,000$
	$0.0266 \, Re^{0.805} \, Pr^{1/3}$	$40,000 < Re < 400,000$
	$(0.35 + 0.56 \, Re^{0.62}) \, Pr^{0.3}$	$0.1 < Re < 100,000$
	$(0.4 \, Re^{1/2} + 0.06 \, Re^{2/3}) \, Pr^{0.4} \, (\mu_b/\mu_w)^{1/4}$	$10 < Re < 100,000$
Flow across square tube	$0.246 \, Re^{0.58} \, Pr^{1/3}$	$5000 < Re < 100,000$
	$0.102 \, Re^{0.675} \, Pr^{1/3}$	$5000 < Re < 100,000$
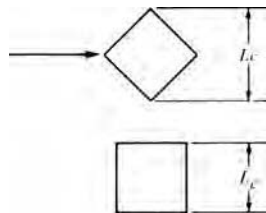
Note: All fluid properties are evaluated at the film temperature, $T_f = (T_w + T_\infty)/2$. The Reynolds number and the Prandtl number are evaluated at film temperature. Source: Ref 7

Table 21 Values of *C* and *n* in the correlation for heat transfer in flow across tube banks of 10 rows or more

Nu = $C\,\mathrm{Re}^n\mathrm{Pr}^{1/3}$ (2000 < Re_{max} < 40,000)

		S_t/D							
		1.25		1.5		2.0		3.0	
S_t/D		*C*	*n*	*C*	*n*	*C*	*n*	*C*	*n*
In-line:	1.250	0.348	0.592	0.275	0.608	0.100	0.704	0.0633	0.752
	1.500	0.367	0.586	0.250	0.620	0.101	0.702	0.0678	0.744
	2.000	0.418	0.570	0.299	0.602	0.229	0.632	0.198	0.648
	3.000	0.290	0.601	0.357	0.584	0.374	0.581	0.286	0.608
Staggered:	0.600	…	…	…	…	…	…	0.213	0.636
	0.900	…	…	…	…	0.446	0.571	0.401	0.581
	1.000	…	…	0.497	0.558	…	…	…	…
	1.125	…	…	…	…	0.478	0.565	0.518	0.560
	1.250	0.518	0.556	0.505	0.554	0.519	0.556	0.522	0.562
	1.500	0.451	0.568	0.460	0.562	0.452	0.568	0.488	0.568
	2.000	0.404	0.572	0.416	0.568	0.482	0.556	0.449	0.570
	3.000	0.310	0.592	0.356	0.580	0.440	0.562	0.421	0.574

Note: See Fig. 22 for geometry and notations.

Table 22 Ratio of (Nu) for *N* rows deep to (Nu) for 10 rows deep

N	1	2	3	4	5	6	7	8	9	10
Ratio for in-line tubes	0.64	0.80	0.87	0.90	0.92	0.94	0.96	0.98	0.99	1.0
Ratio for staggered tubes	0.68	0.75	0.83	0.89	0.92	0.95	0.97	0.98	0.99	1.0

Table 23 Correlations for forced convection from internal flow

Flow geometry	Correlation for Nu	Restrictions
Flow through a smooth circular tube $\mathrm{Re}_{cr}=2300$	UWT: $1.86(\mathrm{RePr})^{1/3}\,(D/C)^{1/3}\,(\mu/\mu_w)^{0.14}$ UWT: $3.66+\dfrac{0.0668(D/L)\mathrm{RePr}}{1+0.04[(D/L)\mathrm{RePr}]^{2/3}}$	Laminar flow $\mathrm{RePr}(D/L)>10$ Laminar flow
	Fully developed	Laminar flow
	UWT: Nu = 3.658 UHF: Nu = 4.364	
	UWT: $0.021\mathrm{Re}^{0.8}\mathrm{Pr}^{0.6}$ UHF: $0.022\mathrm{Re}^{0.8}\mathrm{Pr}^{0.\varepsilon}$	Turbulent flow $0.5<\mathrm{Pr}<1.0$
	UHF: $\dfrac{\left(\frac{f}{8}\right)\mathrm{RePr}(\mu/\mu_w)^n}{1.07+12.7\sqrt{\frac{f}{8}}(\mathrm{Pr}^{2/3}-1)}$	Turbulent flow $2<\mathrm{Pr}<140$
	$f=[1.82\log_{10}\mathrm{Re}-1.64]^{-2}$	$5\times10^3<\mathrm{Re}<1.25\times10^5$ $0.08<$ $(\mu/\mu_w)<40$
	$n=0.11$ for heating; 0.25 for cooling	
	UWT: $5.0+0.025(\mathrm{RePr})^{0.8}$ UHF: $4.82+0.0185(\mathrm{RePr})^{0.827}$	Liquid metal $1000<\mathrm{RePr}$ Liquid metal $10^2<\mathrm{RePr}<10^4$ $3.6\times10^3<\mathrm{Re}<9.05\times10^5$
Flow through a rough circular tube	UWT: $\dfrac{\left(\frac{f}{8}\right)\mathrm{RePr}}{1+1.5\mathrm{Re}^{-1/8}\,\mathrm{Pr}^{-1/6}[(f/f_s)\,\mathrm{Pr}-1]}$	$\mathrm{Pr}<1$ $500<\mathrm{Re}<8\times10^4$ $1.2<\mathrm{Pr}<5.9$
	UHF: $\left(\dfrac{f}{8}\right)\mathrm{RePr}\left\{1+\left[5.19\,\mathrm{Pr}^{0.44}(e^*)^{0.2}-8.48\right]\sqrt{\dfrac{f}{8}}\right\}$	$6\times10^4<\mathrm{Re}<5\times10^5$
	$e^*=\mathrm{Re}\sqrt{\dfrac{f}{8}}\times\dfrac{\varepsilon}{D}$	$0.0024<\dfrac{\varepsilon}{D}<0.049$
Flow between plates	UWT: $12+0.03\mathrm{Re}^m\mathrm{Pr}^n$ $m=0.88-\dfrac{0.24}{3.6+\mathrm{Pr}}$ $n=0.33+0.5e^{-0.6\mathrm{Pr}}$ $8.3+0.02\mathrm{Re}^{0.82}\mathrm{Pr}^n$ $n=0.52+\dfrac{0.0096}{0.02+\mathrm{Pr}}$	$0.1<\mathrm{Pr}<10^4$ $10^4<\mathrm{Re}<10^6$ $0.004<\mathrm{Pr}<0.1$ $10^4<\mathrm{Re}<10^6$
Helically coiled tube; coil dia., *D*; tube dia., *d*	UWT: $\dfrac{0.32+3(d/D)}{0.86-0.8(d/D)}$	Laminar flow
	$\mathrm{Re}^{0.5}\mathrm{Pr}^{0.33}(d/L)^{(0.14+0.8(d/D))}$ UHF: $1.268\mathrm{Re}^{0.26}\mathrm{Pr}^{1/6}$	$20<\mathrm{Re}\,(d/D)^{1/2}<830$ $30<\mathrm{Pr}<450$ $0.01<(d/D)<0.08$

Note: UHF = uniform heat flux at the boundary; UWT = uniform wall temperature; thermal properties are evaluated at bulk temperature except μ_w, which is evaluated at the wall temperature; f = friction factor for a smooth tube; ε/D = relative roughness of wall. Source Ref 7

Fig. 24 Nusselt number in the combined thermal and hydrodynamic entry length of a circular tube with uniform heat flux at wall

LIVE GRAPH
Click here to view
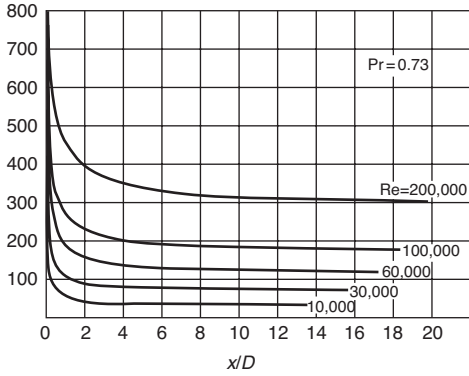
Table 24 Local laminar Nusselt number, for circular tube with uniform wall heat flux (UHF) and wall temperature (UWT)

$x^* = \frac{x}{r}\frac{1}{RePr}$	Nu_{x^*} for UHF	Nu_{x^*} for UWT
0.001	. . .	12.86
0.002	12.0	. . .
0.004	9.93	7.91
0.010	7.49	5.99
0.020	6.14	. . .
0.040	5.19	4.18
0.080	. . .	3.79
0.100	4.51	3.71
0.200	. . .	3.66
∞	4.36	3.66

Source: Ref 8

LIVE GRAPH
Click here to view
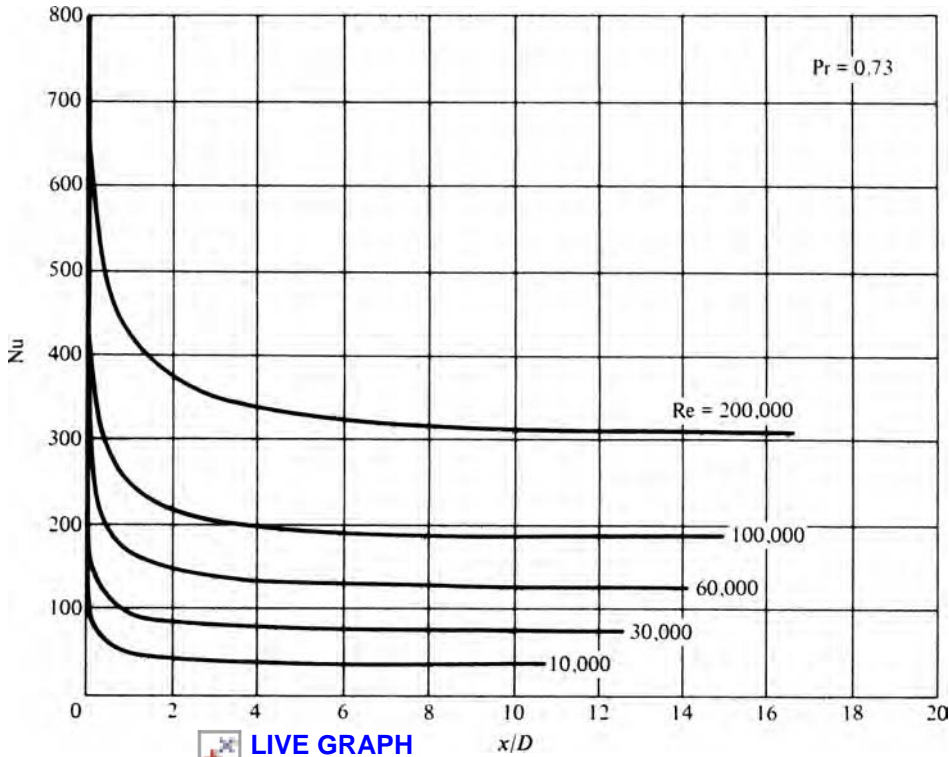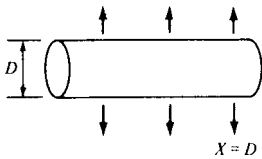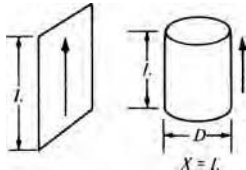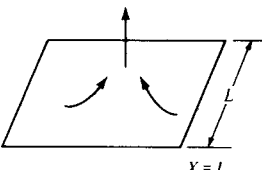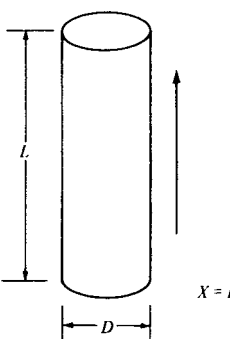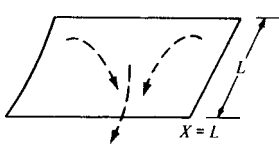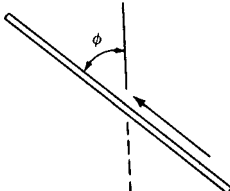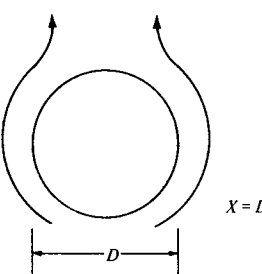
Fig. 25 Nusselt number in the combined thermal and hydrodynamic entry length of a circular tube with constant wall temperature
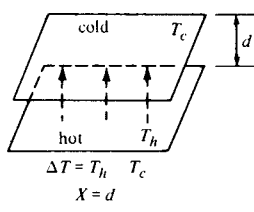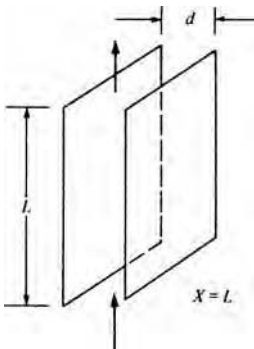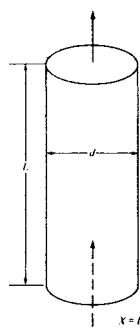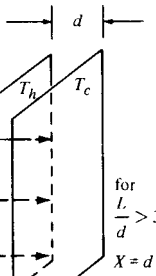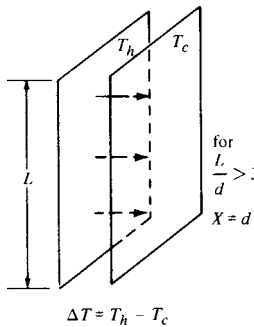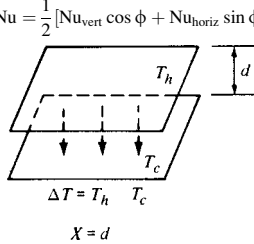
Table 25 Free convection heat-transfer coefficient for various systems, Nu = $C(GrPr)^n$R

System	Schematic diagram	C	n	R	Operating conditions
Exposed surfaces					
Horizontal cylinder		0.53	$\frac{1}{4}$	1	Laminar flow
	$X = D$	0.13	$\frac{1}{3}$	1	Turbulent flow
Vertical plate and vertical cylinder with large diameter		0.8	$\frac{1}{4}$	$\left[1 + \left(1 + \frac{1}{\sqrt{Pr}}\right)^2\right]^{-1/4}$	Laminar flow; to obtain local Nu, use $C = 0.6$, $X = x$; formula applicable to vertical cylinder when $\frac{D}{L} \geq 38(Gr)^{-1/4}$
	$X = L$	0.0246	$\frac{2}{5}$	$[Pr^{1/6}/(1 + 0.494Pr^{2/3})]^{2/5}$	Turbulent flow; to obtain local Nu use $C = 0.0296$ and $X = x$
Heated horizontal plate facing upward		0.54	$\frac{1}{4}$	1	Laminar flow; for circular disc of diameter D, use $X = 0.9D$
	$X = L$	0.14	$\frac{1}{3}$	1	Turbulent flow
Vertical cylinder with small diameter	$X = L$	0.686	$\frac{1}{4}$	$[Pr/(1 + 1.05Pr)]^{1/4}$	Laminar flow; $Nu_{total} = Nu + 0.52\frac{L}{D}$
Heated horizontal plate facing downward	$X = L$	0.27	$\frac{1}{4}$	1	Laminar flow only
Moderately inclined plate		0.8	$\frac{1}{4}$	$\left[\frac{\cos\phi}{1 + \left(1 + \frac{1}{\sqrt{Pr}}\right)}\right]^{1/4}$	Laminar flow (multiply Gr by cos φ in the formula for inclined plate)
Sphere	$X = D$	0.49	$\frac{1}{4}$	1	Laminar flow (air)

(continued)

Note: X = characteristic length of the system; $\Delta T = T_w - T_\infty$, T_h = temperature of the hot surface, T_c = temperature of the cold surface; β = coefficient of volumetric thermal expansion for fluids; all properties are evaluated at film temperature except β, which is computed from T_∞ or T_b. Source: 1, 6

Table 25 (continued)

System	Schematic diagram	C	n	R	Operating conditions
Enclosed spaces					
Two horizontal parallel plates; cold plate uppermost		0.195	$\frac{1}{4}$	$Pr^{-1/4}$	Laminar (air) $10^4 < Gr < 4 \times 10^5$
		0.068	$\frac{1}{3}$	$Pr^{-1/3}$	Turbulent (air) $Gr > 4 \times 10^5$
Two vertical parallel plates at the same temperature		0.04	1	$(d/L)^3$	Air layer
Hollow vertical cylinder with open ends		0.01	1	$(d/L)^3$	Air column
Two vertical parallel plates at different temperatures (*h* for both surfaces)		0.18	$\frac{1}{4}$	$(L/d)^{-1/9}(Pr)^{-1/4}$	Laminar (air) $2 \times 10^4 < Gr < 2 \times 10^5$
		0.065	$\frac{1}{3}$	$(L/d)^{-1/9}(Pr)^{-1/3}$	Turbulent (air) $2 \times 10^5 < Gr < 10^7$
Two inclined parallel plates					
Two horizontal parallel plates; hot plate uppermost	$Nu = \frac{1}{2}[Nu_{vert} \cos\phi + Nu_{horiz} \sin\phi]$	...	...	...	Pure conduction $q = k\frac{(T_h - T_c)}{d}$
		0.27	$\frac{1}{4}$	1	Laminar (air) $3 \times 10^5 < Gr \cdot Pr < 3 \times 10^{10}$

(continued)

Note: X = characteristic length of the system; $\Delta T = T_w - T_\infty$, T_h = temperature of the hot surface, T_c = temperature of the cold surface; β = coefficient of volumetric thermal expansion for fluids; all properties are evaluated at film temperature except β, which is computed from T_∞ or T_b. Source: 1, 6

Table 25 (continued)

System	Schematic diagram	C	n	R	Operating conditions
Rectangular solid with uniform temperature		0.55	$\frac{1}{4}$	1	Laminar flow $\dfrac{1}{X} = \dfrac{1}{c} + \dfrac{1}{(a+b)/2}$

Note: X = characteristic length of the system; $\Delta T = T_w - T_\infty$, T_h = temperature of the hot surface, T_c = temperature of the cold surface; β = coefficient of volumetric thermal expansion for fluids; all properties are evaluated at film temperature except β, which is computed from T_∞ or T_b. Source: 1, 6

Table 26 Summary of empirical relations for free convection in enclosures in the form $e/k = C(Gr_\delta \, Pr)^n (L/\delta)^m$

Fluid	Geometry	$Gr_\delta Pr$	Pr	L/δ	C	n	m
Gas	Vertical plate, isothermal	<2000	$k_e/k = 1.0$				
		6000–200,000	0.5–2	11–42	0.197	$\frac{1}{4}$	$-\frac{1}{9}$
		200,000–1.1×10^7	0.5–2	11–42	0.073	$\frac{1}{3}$	$-\frac{1}{9}$
	Horizontal plate, isothermal, heated from below	<1700	$k_e/k = 1.0$				
		1700–7000	0.5–2	. . .	0.059	0.4	0
		7000–3.2×10^5	0.5–2	. . .	0.212	$\frac{1}{4}$	0
		>3.2×10^5	0.5–2	. . .	0.061	$\frac{1}{3}$	0
Liquid	Vertical plate, constant heat flux or isothermal	10^4–10^7	1–20,000	10–40	$0.42(Pr^{0.012})$	$\frac{1}{4}$	−0.3
		10^6–10^9	1–20	1–40	0.046	$\frac{1}{3}$	0
	Horizontal plate, isothermal, heated from below	<1700	$k_e/k\pi1.0$	. . .			
		1700–6000	1–5000	. . .	0.012	0.6	0
		6000–37,000	1–5000	. . .	0.375	0.2	0
		37,000–10^8	1–20	. . .	0.13	0.3	0
		>10^8	1–20		0.057	$\frac{1}{3}$	0
Gas or liquid	Vertical annulus	Same as vertical plates					
	Horizontal annulus, isothermal	6000–10^6	1–5000	. . .	0.11	0.29	0
		10^6–10^8	1–5000	. . .	0.40	0.20	0
	Spherical annulus	120–1.1×10^9	0.7–4000	. . .	0.228	0.226	0

Source: Ref 9
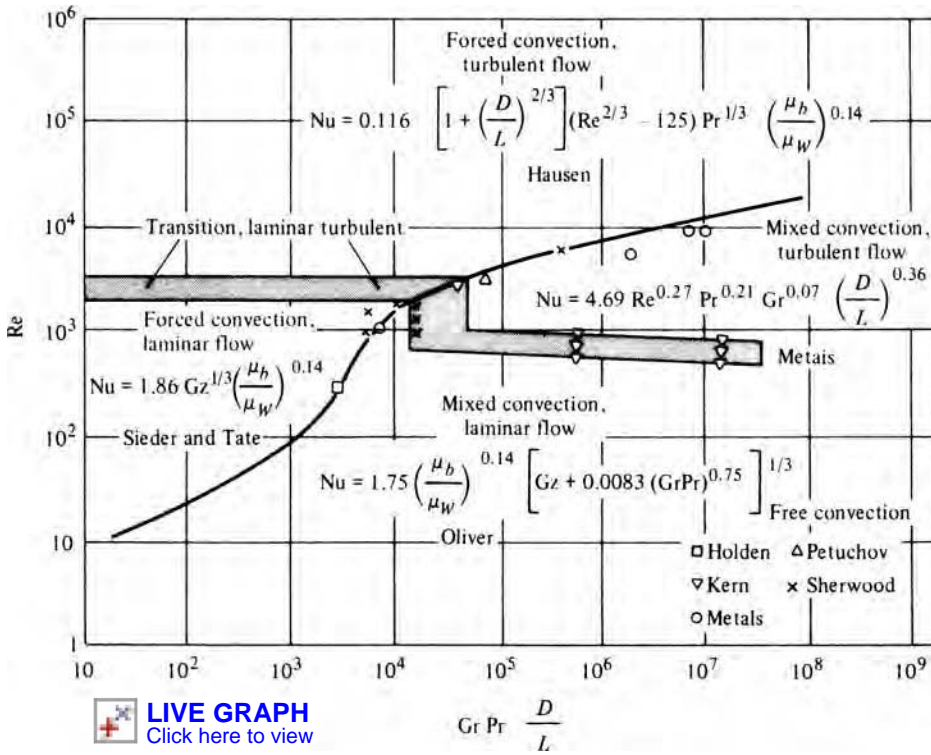
Fig. 26 Regimes of free, forced, and mixed convection for flow through horizontal tubes
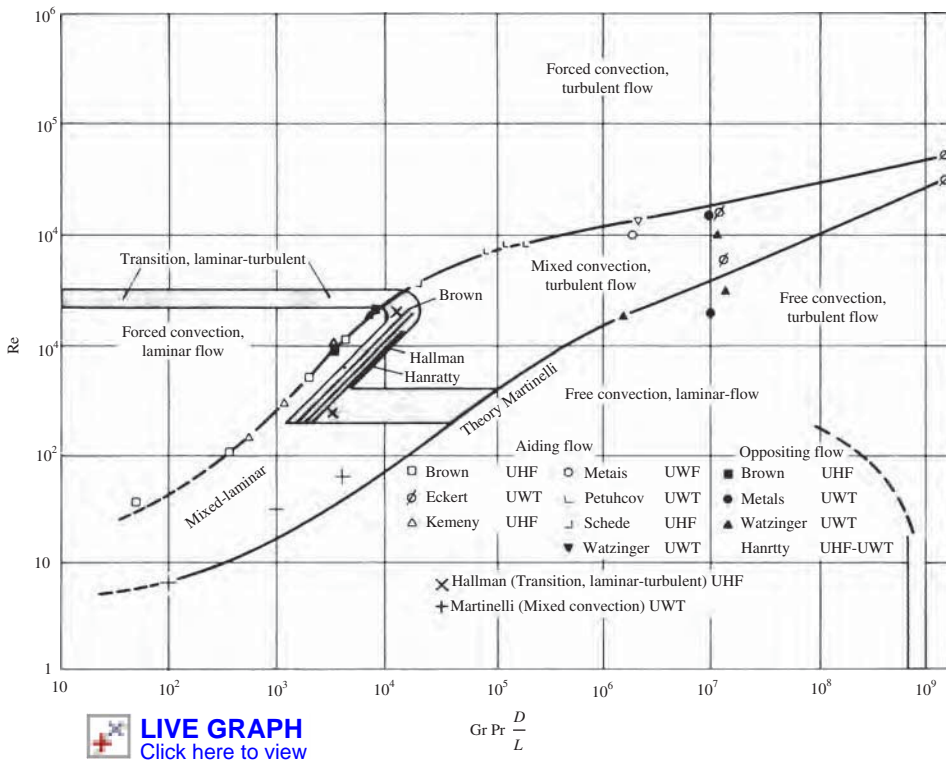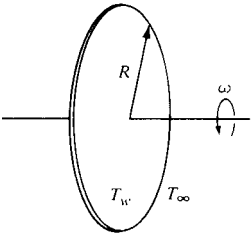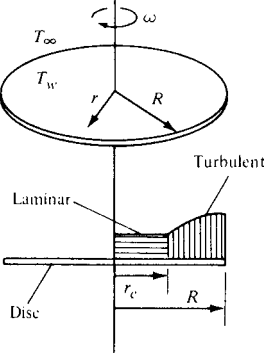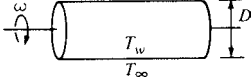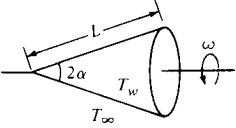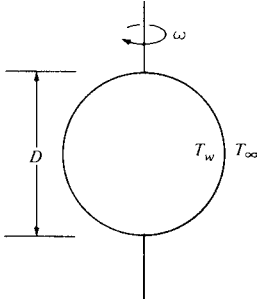
Fig. 27 Regimes of free, forced, and mixed convection for flow through vertical tubes

Table 27 Heat-transfer coefficient of rotating bodies

Formula	Conditions	Formula	Conditions
Rotating disc		**Rotating cylinder**	

Rotating disc

$Nu = (0.277 + 0.105\,Pr)Re^{0.5}$ — Laminar flow, $Re < 2.5 \times 10^5$, $0.7 < Pr < 5.0$

$Nu = 1.1\,Re^{0.5}$ — Laminar flow, $Re < 2.5 \times 10^5$, $Pr = 10$

$Nu = 0.015\,Re^{0.8}$ — Turbulent flow, $Re > 2.5 \times 10^5$, $Pr = 0.72$

$Nu = 0.015\,Re^{0.8} - 100\left(\dfrac{r_c}{R}\right)^2$ — Laminar flow between $r = 0$ and $r = r_c$, turbulent flow between $r = r_c$ and $r = R$ where $r_c = (2.5 \times 10^5\,v/\omega)^{1/2}$, $Pr = 0.72$

$Nu = 0.4(Re^2 + Gr)^{0.25}$

where

$Nu = \dfrac{hR}{k}$, $Re = \dfrac{\omega R^2}{v}$,

$Gr = \dfrac{\beta g R^3 \pi^{3/2} \Delta T}{v^2}$

Combined effects of free convection and rotation in laminar flow (axis horizontal)

Rotating cone

$Nu = 0.515\,(Gr)^{0.25}$ — Laminar free convection, $Pr = 0.72$, $Gr/Re^2 > 2.0$

$Nu = 0.33\,Re^{0.5}$ — Forced convection, $Pr = 0.72$, $Gr/Re^2 < 0.05$

$Nu = Re^{0.5}[0.331 + 0.412\,(Gr/Re^2) + \ldots]$ — Combined free and forced convection, $Pr = 0.72$, $0.2 < Gr/Re^2 < 1.0$

where

$Nu = \dfrac{hL}{k}$, $Re = \dfrac{\omega L^2 \sin \alpha}{v}$,

$Gr = \dfrac{\beta g L^3 \cos \alpha \Delta T}{v^2}$

Rotating cylinder

$Nu = 0.456\,(GrPr)^{0.26}$ — Free convection, $Re < (Gr/Pr)^{0.5}$

$Nu = 0.18[(0.5\,Re^2 + Gr)Pr]^{0.315}$ — Combined free and forced convection, $Re \leq 5 \times 10^4$

$Nu = \dfrac{Re \cdot Pr \sqrt{C_D/2}}{5\,Pr + 5\ln(3\,Pr + 1)} + \sqrt{2/C_D} - 12$ — Forced convection, $Re > 10^5$

C_D from:

$\dfrac{Re}{B} = -1.828 + 1.77\ln B$ for $B > 950$

$\dfrac{Re}{B} = -3.68 + 2.04\ln B$ for $B < 950$

where $B = Re\sqrt{C_D}$

$Nu = 0.135[(0.5Re^2 + Re_f^2 + Gr)\,Pr]^{0.33}$ — Combined effects of rotation, free convection, and crossflow, $Re_f < 1.5 \times 10^4$, $0.6 < Pr < 15$, $10^3 < Re < 5 \times 10^4$, value in square bracket [ ] $< 10^9$

$Nu = \dfrac{hD}{k}$, $Re = \dfrac{\omega D^2}{v}$,

$Re_f = \dfrac{y_\infty D}{v}$, $Gr = \dfrac{\beta g D^3 \Delta T}{v^2}$

Rotating sphere

$Nu = 0.43\,Re^{0.5}\,Pr^{0.4}$ — Laminar flow, $Gr/Re^2 < 0.1$, $Re < 5 \times 10^4$, $0.7 < Pr < 217$

$Nu = 0.066\,Re^{0.67}\,Pr^{0.4}$ — Turbulent flow, Gr/Re^2 0.1, $5 \times 10^4 < Re < 7 \times 10^5$, $0.7 < Pr < 7$

$Nu = 2(Re^2 + Gr)^{0.164}$ — Combined free and forced convection, $Gr/Re^2 > 0.1$, $10^3 < Re < 2 \times 10^4$, $4 \times 10^6 < Gr < 2 \times 10^7$

where

$Nu = \dfrac{hD}{k}$, $Re = \dfrac{\omega D^2}{v}$

$Gr = \dfrac{\beta g D^3 \Delta T}{v^2}$

Note: The fluid properties are taken at the film temperature; ω = angular velocity of rotation; X = characteristic length; $\Delta T = T_w - T_\infty$ C_D = surface drag coefficient Source: Ref 6
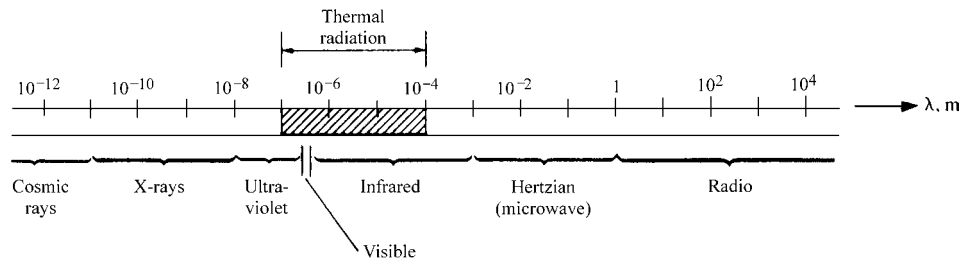
Fig. 28 Radiation spectrum
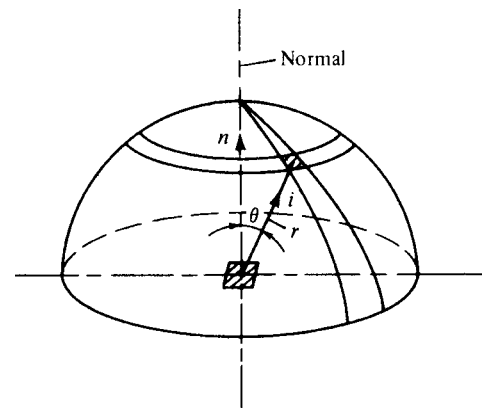
Fig. 29 Diffuse radiation
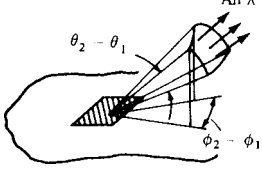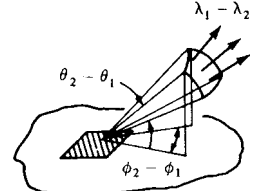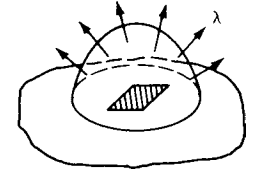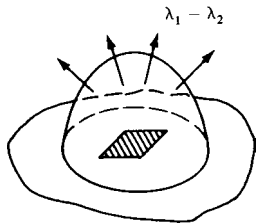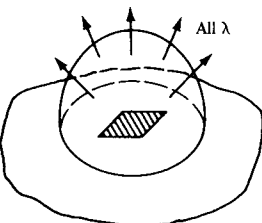
Table 28   Blackbody thermal properties

Symbol	Name	Definition	Geometry	Formula
$i'_{\lambda,b}(\lambda, T)$	Spectral intensity	Emission in any direction per unit of projected area normal to that direction, and per unit time, wavelength interval about λ, and solid angle		$\dfrac{2C_1}{\lambda^5(e^{C_2/\lambda T} - 1)}$ Planck's law
$i'_b(T)$	Total intensity	Emission, including all wavelengths, in any direction per unit of projected area normal to that direction, and per unit time and solid angle		$\dfrac{\sigma T^4}{\pi}$
$e'_{\lambda,b}(\lambda, \theta, T)$	Directional spectral emissive power	Emission per unit solid angle in direction θ per unit surface area, wavelength interval, and time		$i'_{\lambda,b}\cos\theta$ Lambert's cosine law
$e'_b(\theta, T)$	Directional total emissive power	Emission, including all wavelengths, in direction θ per unit surface area, solid angle, and time		$\dfrac{\sigma T^4}{\pi}\cos\theta$ Lambert's cosine law
$e_{\lambda,b}(\lambda, \theta_1 - \theta_2, \varphi_1 - \varphi_2, T)$	Finite solid-angle spectral emissive power	Emission in solid angle $\theta_1 \leq \theta \leq \theta_2$, $\varphi_1 \leq \varphi \leq \varphi_2$ per unit surface area, wavelength interval, and time		$i'_{\lambda,b}\dfrac{\sin^2\theta_2 - \sin^2\theta_1}{2}(\varphi_2 - \varphi_1)$

(continued)

Note: Spectral denotes dependence on wavelength for any radiation quantity; superscript prime ($'$) implies the directional quantity; $F_{0-\lambda}$ represents fraction of total blackbody intensity or emissive power lying in spectral region $0 - \lambda$, that is, $F_{0-\lambda} = \int_0^\lambda \frac{e_{\lambda,b}(\lambda)d\lambda}{\sigma T^4}$. [1]

Table 28 (continued)

Symbol	Name	Definition	Geometry	Formula
$e\,(\theta_1-\theta_2,\varphi_1-\varphi_2,T)$	Finite solid-angle total emissive power	Emission, including all wave-lengths, in solid angle $\theta_1\le\theta\le\theta_2,\varphi_1\le\varphi\le\varphi_2$ per unit surface area, and time		$\dfrac{\sigma T^4}{\pi}(\varphi_2-\varphi_1)\dfrac{\sin^2\theta_2-\sin^2\theta_1}{2}$
$e_{\lambda,b}\,(\lambda_1-\lambda_2,\theta_1-\theta_2,\varphi_1-\varphi_2,T)$	Finite solid-angle band emissive power	Emission in solid angle $\theta_1\le\theta\le\theta_2,\varphi_1\le\varphi\le\varphi_2$, and wavelength band $\lambda_1-\lambda_2$ per unit surface area and time		$\dfrac{\sigma T^4}{\pi}(\varphi_2-\varphi_1)\dfrac{\sin^2\theta_2-\sin^2\theta_1}{2}$ $\times(F_{0-\lambda_2}-F_{0-\lambda_1})$
$e_{\lambda,b}\,(\lambda,T)$	Hemispherical spectral emissive power	Emission into hemispherical solid angle per unit surface area, wavelength interval, and time		$\pi i'_{\lambda,b}$
$e_{\lambda,b}\,(\lambda_1-\lambda_2,T)$	Hemispherical band emissive power	Emission in wavelength band $\lambda_1-\lambda_2$ into hemispherical solid angle per unit surface area and time		$(F_{0-\lambda_2}-F_{0-\lambda_1})\sigma T^4$
$e,(T)$	Hemispherical total emissive power	Emission, including all wavelengths, into hemispherical solid angle per unit surface area and time		σT^4 Stefan-Boltzmann's law

Note: Spectral denotes dependence on wavelength for any radiation quantity; superscript prime (′) implies the directional quantity; $F_{0-\lambda}$ represents fraction of total blackbody intensity or emissive power lying in spectral region $0-\lambda$, that is, $F_{0-\lambda}=\displaystyle\int_0^\lambda\dfrac{e_{\lambda,b}(\lambda)d\lambda}{\sigma T^4}$. Source: Ref 10

Table 29 Radiation constants in Planck's, Stefan-Boltzmann's, and Wien's equations

Constant	Definition	Value
C_1	Planck's spectral energy distribution first constant	0.595×10^{-8} W·m^2
C_2	Planck's spectral energy distribution second constant	1.438×10^{-2} m·K
C_3	Wien's displacement law	0.289×10^{-2} m.K
σ	Stefan-Boltzmann's constant	5.669×10^{-8} W/m^2·K^4
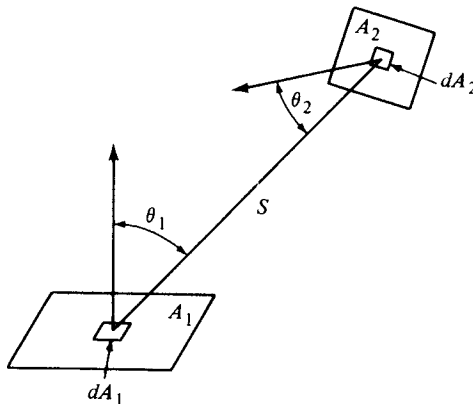
Fig. 30 Geometrical shape-factor notation

Table 30 Summary of configuration-factor and energy-exchange relations

Net energy transfer	Configuration factor	Reciprocity
Elemental area to elemental area $d^2Q_{d1\leftrightarrow d2} = \sigma(T_1{}^4 - T_2{}^4)dA_1 dF_{d1-d2}$	$dF_{d1-2} = \dfrac{\cos\theta_1\cos\theta_2}{\pi S^2}dA_2$	$dA_1 dF_{d1-d2} = dA_2 dF_{d2-d1}$
Elemental area to finite area $dQ_{d1\rightarrow 2} = \sigma(T_1{}^4 - T_2{}^4)dA_1 F_{d1-2}$	$F_{d1-2} = \displaystyle\int_{A_2}\dfrac{\cos\theta_1\cos\theta_2}{\pi S^2}dA_2$	$dA_1 F_{d1-2} = A_2 dF_{2-d1}$
Finite area to finite area $Q_{1\leftrightarrow 2} = \sigma(T_1{}^4 - T_2{}^4)A_1 F_{1-2}$	$F_{1-2} = \dfrac{1}{A_1}\displaystyle\int_{A_1}\int_{A_2}\dfrac{\cos\theta_1\cos\theta_2}{\pi S^2}dA_2 dA_1$	$A_1 F_{1-2} = A_2 F_{2-1}$

Fig. 32 Enclosure composed of N black isothermal surfaces

$$dF_{d1-d2} = \frac{\cos\varphi}{2}d\varphi = \frac{1}{2}d(\sin\varphi)$$

(a)

$$dF_{d1-d2} = \frac{\cos\varphi}{\pi}d\varphi\tan^{-1}\frac{b}{r}$$

(b)

$$X = \frac{x}{2r}$$

$$dF_{d1-d2} = \left[1 - \frac{2X^3 + 3X}{2(X^2+1)^{3/2}}\right]dX_2$$

(c)
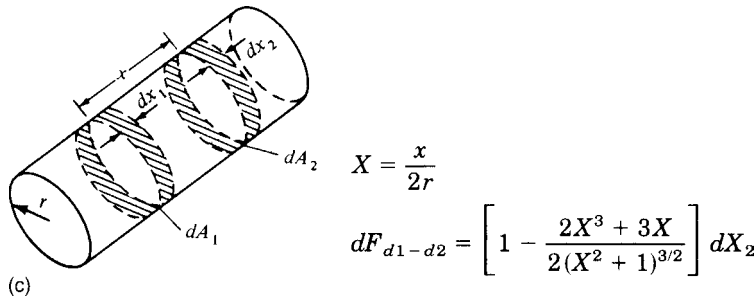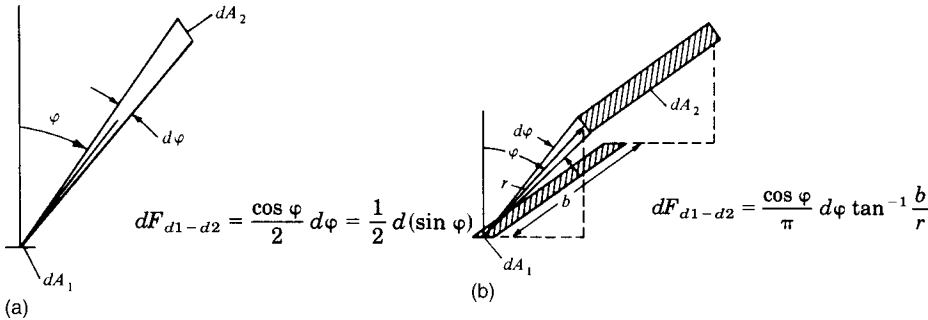
Fig. 31 (a) Area dA_1 of differential width and any length, to infinitely long strip dA_2 of differential width and with parallel generating line to dA_1. (b) Strip of infinite length b and of differential width, to differential strip of same length on parallel generating line. (c) Two ring elements on the interior of a right circular cylinder.

Fig. 33 Enclosure composed of N discrete gray surface areas with typical surfaces j and k

Table 31 Relationship between energy flux and temperature in diffuse gray enclosures

Known boundary conditions	Desired quantities	Relationship
T_k on all surfaces	Q_k	$\displaystyle\sum_{j=1}^{N}\left(\dfrac{\delta_{kj}}{\varepsilon_j} - F_{k-j}\dfrac{1-\varepsilon_j}{\varepsilon_j}\right)\dfrac{Q_j}{A_j} = \sum_{j=1}^{N}(\delta_{kj} - F_{k-j})\sigma T_j{}^4$
$1 \le k \le N$	J_k	$\delta_{kj} = \begin{cases} 1, \text{when } k=j \\ 0, \text{when } k \ne j \end{cases}$
		$\displaystyle\sum_{j=1}^{N}[\delta_{kj} - (1-\varepsilon_k)F_{k-j}]J_j = \varepsilon_k\sigma T_k{}^4$ $k = 1, 2, \ldots N$
Q_k on all surfaces	T_k	$\displaystyle\sum_{j=1}^{N}(\delta_{kj} - F_{k-j})\sigma T_j{}^4 = \sum_{j=1}^{N}\left(\dfrac{\delta_{kj}}{\varepsilon_j} - F_{k-j}\dfrac{1-\varepsilon_j}{\varepsilon_j}\right)\dfrac{Q_j}{A_j}$
$1 \le k \le N$	J_k	$\left(J_k - \displaystyle\sum_{j=1}^{N}F_{k-j}J_j\right) = \dfrac{Q_k}{A_k}$
T_k for $1 \le k \le m$	Q_k for $1 \le k \le m$	$\displaystyle\sum_{j=1}^{N}[\delta_{kj} - (1-\varepsilon_k)F_{k-j}]J_j = \varepsilon_k\sigma T_k{}^4 \quad 1 \le k \le m$
Q_k for $m+1 \le k \le N$	T_k for $m+1 \le k \le N$	$\displaystyle\sum_{j=1}^{N}(\delta_{kj} - F_{k-j})J_j = \dfrac{Q_k}{A_k} \quad m+1 \le k \le N$
	J_k for $1 \le k \le N$	Solve for J_j from above N system equations first, then: $Q_k = \dfrac{A_k\varepsilon_k}{1-\varepsilon_k}(\sigma T_k{}^4 - J_k)\, 1 \le k \le m$ $T_k = \left[\left(\dfrac{Q_k(1-\varepsilon_k)}{A\varepsilon_k} + J_k\right)\dfrac{1}{\sigma}\right]^{1/4}\, m+1 \le k \le N$

Note: Q_k equals external heat input to kth surface; J_k is radiosity, or energy leaving kth surface per unit area and unit time.

Table 32 Rate of radiant energy exchange between diffuse surfaces

Schematic diagram	Radiation heat-exchange rate
Gray surface to blackbody surroundings	$Q_{1-2} = \varepsilon_1 \sigma (T_1^4 - T_2^4) A_1$
Two arbitrary surfaces	$Q_{1-2} = \dfrac{\sigma \varepsilon_1 \varepsilon_2 A_1 F_{1-2}(T_1^4 - T_2^4)}{\varepsilon_1 \varepsilon_2 + \varepsilon_2 F_{1-2}(1-\varepsilon_1) + \varepsilon_1 F_{2-1}(1-\varepsilon_2)}$
Two infinitely large parallel planes	$Q_{1-2} = \dfrac{1}{\frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_2} - 1} \sigma (T_1^4 - T_2^4) A_1$
A small enclosed body and the enclosure	$Q_{1-2} = \varepsilon_1 A_1 \sigma (T_1^4 - T_2^4)$
Infinitely long concentric cylinders, concentric spheres, or an arbitrary enclosed body and its enclosure	$Q_{1-2} = \dfrac{1}{\frac{1}{\varepsilon_1} + \frac{A_1}{A_2}\left(\frac{1}{\varepsilon_2}-1\right)} \sigma (T_1^4 - T_2^4) A_1$
Two black surfaces in the presence of reradiating surfaces	$Q_{\underset{r}{1-2}} = r(T_1^4 - T_2^4) A_1 F_{1r^2}$ where $F_{\underset{r}{1-2}} = F_{1-2} + \dfrac{1}{\frac{1}{F_{1-r}} + \frac{A_1}{A_2 F_{2-r}}}$

Schematic diagram	Radiation heat-exchange rate
Two black surfaces that do not "see" themselves in the presence of reradiating surfaces	$Q_{1-2} = \sigma(T_1^4 - T_2^4) A_1 F_{1-2}$ where $F_{1-2} = \dfrac{A_2 - A_1 F_{1-2}^2}{A_1 + A_2 - 2A_1 F_{1-2}}$
Two arbitrary gray surfaces in the presence of reradiating surfaces	$Q_{\underset{r}{1-2}} = \sigma A_1 F_{1-2}(T_1^4 - T_2^4)$ where $F_{1-2} = \dfrac{1}{\frac{1}{F_{1r2}} + \frac{1}{\varepsilon_1} - 1 + \frac{A_1}{A_2}\left(\frac{1}{\varepsilon_2}-1\right)}$ and $F_{\underset{r}{1-2}} = F_{1-2} + \dfrac{1}{\frac{1}{F_{1-r}} + \frac{A_1}{A_2 F_{2-r}}}$
A gas and a black bounding surface	$Q_{A-g} = \sigma A_1 (\varepsilon_g T_g^4 - \alpha_g T_w^4)$
A gas and a black bounding surface in the presence of a reradiating bounding surface	$Q_{\underset{r}{A-g}} = \overline{\varepsilon}_g A_1 (T_g^4 - T_w^4)$ where $\overline{\varepsilon}_g = \left[1 + \dfrac{A_r/A_1}{1 + \frac{\varepsilon_g}{(1-\varepsilon_g)F_{r-1}}}\right]\varepsilon_g = \text{equivalent emissivity}$
A gas and a gray bounding surface in the presence of a reradiating bounding surface	$Q_{\underset{r}{A-c}} = \dfrac{\sigma A_1 (T_g^4 - T_w^4)}{\frac{1}{\varepsilon_g} + \frac{1}{\varepsilon_1} - 1}$ where $\overline{\varepsilon}_g = \left[1 + \dfrac{A_r/A_1}{1 + \frac{\varepsilon_g}{(1-\varepsilon_g)F_{r-1}}}\right]\varepsilon_g$

Fig. 34 Parallel walls separated by *N* radiation shields

Fig. 35 Radiation shields between concentric cylinders or spheres

Fluid Dynamic Equations*

A FLUID cannot resist shear stress by static deformation in the manner of a solid. Any shear stress applied to a fluid results in motion of that fluid, which continues as long as the shear stress is applied. A fluid at rest must be in a state of zero shear stress, a state known as the hydrostatic stress condition. A liquid, which is composed of relatively close-packed molecules with strong cohesive forces, retains its volume and forms a free surface in a gravitational field if the upper surface is not confined. Gases, with widely spaced molecules and negligible cohesive forces, are free to expand until encountering confining walls.

Properties of Fluids

The Continuum

Certain properties are defined on the basis of fluid particles being considered as a continuous medium. A fluid particle is composed of many finite-sized molecules with finite distance between the molecules. For air under standard conditions, there are 2×10^{19} molecules in 1 cm^3, with a mean distance travelled between each molecular collision of 6.35×10^{-6} cm. Therefore, the overall motion of the fluid is of interest, not the motion of individual molecules, and the fluid can be treated as a continuous medium, also called a continuum.

Pressure

Pressure is defined at a point in a continuum as:

$$p = \lim_{A \to A'} \frac{F}{A} \qquad \text{(Eq 1)}$$

where F is the force normal to the surface A, and A' is the smallest area surrounding the point that is consistent with continuum approach. The mean pressure over a plane area in a fluid is the ratio of the normal force acting on an area to the area. The pressure at a point is the limit approached by the mean pressure as the area is reduced to a very small size around the point. When pressure is measured relative to zero pressure, it is called absolute pressure. When pressure is measured relative to surrounding atmospheric pressure, it is called gage pressure.

Density

Density at a point in a continuum is defined as:

$$\rho = \lim_{V \to V'} \frac{M}{V} \qquad \text{(Eq 2)}$$

where M is the mass in volume, V, and V' is the smallest volume surrounding the point that is consistent with the continuum approach. Mean density is the ratio of the fluid mass to the volume. Density at a point is the limit approached by the mean density as the volume is reduced to a very small size around the point.

Coefficient of Compressibility

This property describes the change of volume with applied pressure at a given temperature:

$$\beta = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_T \qquad \text{(Eq 3)}$$

Its reciprocal, the isothermal bulk modulus κ, is defined as:

$$\kappa = -V\left(\frac{\partial p}{\partial V}\right)_T \qquad \text{(Eq 4)}$$

Viscosity

In a solid, shear stress is generally proportional to shear strain, and the material ceases to deform when equilibrium is reached. In a viscous fluid, however, shear stress is proportional to the time rate of strain. The proportionality factor for the viscous fluid is the dynamic, or absolute, viscosity:

$$\tau = \mu \frac{du}{dy} \qquad \text{(Eq 5)}$$

where τ is the shear stress, μ is the dynamic viscosity, u is the velocity and du/dy is the time rate of strain or velocity gradient (Fig. 1). This linear variation of shear stress with rate of strain describes the behavior of a large class of fluids called Newtonian fluids. The velocity gradient at the boundary must be finite, because the shear stress cannot be infinite. When the velocity profile is such that the gradient becomes smaller farther away from the boundary, the shear stress will reach a maximum at the boundary and will decrease with distance from the wall. Another form of the coefficient of viscosity is the kinematic viscosity, defined as:

$$\nu = \frac{\mu}{\rho}$$

Variation of Viscosity with Temperature. Temperature has a strong effect on viscosity, while pressure exerts a moderate effect. The viscosity of gases and most liquids increases slowly with increases in pressure. Gas viscosity increases with temperature, the functional relationship being approximated by the two common equations listed as follows:
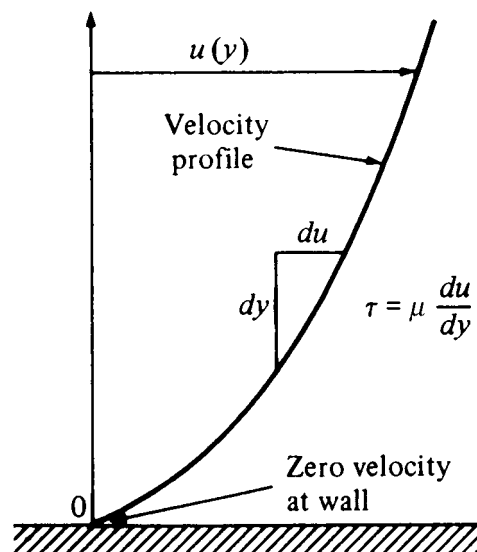
Fig. 1 Velocity gradient

$$\frac{\mu}{\mu_0} \approx \begin{cases} (T/T_0)^n & \text{Power law} \\ [(T/T_0)^{3/2}(T_0+S)]/T+S & \text{Sutherland law} \end{cases}$$

(Eq 6)

where μ_0 is a known viscosity at a known absolute temperature, T_0, and n and S are constants. For air, $n \approx 0.67$ and $S \approx 110$ K.

Liquid viscosity decreases with temperature in a manner roughly described by the exponential function, $\mu \approx ae^{-bT}$. Preferred, however, is the empirical result:

$$\ln\frac{\mu}{\mu_0} \approx a + b\left(\frac{T_0}{T}\right) + c\left(\frac{T_0}{T}\right)^2$$

(Eq 7)

For water, with $T_0 = 273.16$ K and $\mu_0 = 0.001792$ kg/(m · s), acceptable values are $a = -1.94$, $b = -4.80$, and $c = 6.74$, with accuracy of approximately $\pm 1\%$.

Non-Newtonian Fluids. Fluids that do not follow the linear law of Eq 5 are called non-Newtonian and are analyzed as problems in rheology. Figure 2 shows how the shear stress varies with respect to the strain rate for some different types of fluids as compared to Newtonian fluids.

Surface Tension

Liquid surfaces exert a tension on portions of surfaces and objects that are in contact with the liquid, even in the absence of motion. This tension acts in the plane of the surface, and the magnitude of the force per unit length is defined as the surface tension, σ.

Vapor Pressure

Molecules escape from a liquid surface until the pressure of the space in contact with the surface reaches a value such that there is no net exchange of molecules between the liquid and vapor. This pressure is called the saturated vapor pressure, p_v. The vapor pressure is a function of the temperature of the liquid.

Fluid Statics

Pressure

The pressure intensity, or pressure, is a scalar quantity and was previously defined in Eq 1. Figure 3 shows a free-body diagram of forces on a fluid element in a static fluid. Analysis of these forces shows that pressure at a point in a static fluid acts with the same magnitude in all directions:

$$p_n = p_x = p_y = p_z$$

(Eq 8)

Variation of Pressure with Elevation

Considering the forces on the element in Fig. 4, it can be shown that by applying the equations of equilibrium, the pressure variation becomes:

$$\frac{dp}{dl} = -\gamma \sin \alpha$$

Noting that $\sin \alpha = dz/dl$:

$$\frac{dp}{dl} = -\gamma \frac{dz}{dl}$$

(Eq 9)

or:

$$\frac{dp}{dz} = -\gamma$$

(Eq 10)

where $\gamma = pg$ (specific weight). This is the basic equation for the variation of hydrostatic pressure with elevation. For a fluid with a uniform density, this pressure variation becomes:

$$\frac{p_1}{\gamma} + z_1 = \frac{p_2}{\gamma} + z_2$$

(Eq 11)

or:

$$\Delta p = -\gamma \Delta z$$

(Eq 12)

The sum shown in Eq 11, $p/\gamma + z$, is called the piezometric head.

The static pressure variation in a compressible fluid can be determined using the ideal gas equation of state:

$$\rho = \frac{p}{RT} = \frac{1}{v}$$

(Eq 13)

where $v \equiv$ specific volume. This can be expressed in the form:

$$\gamma = \frac{pg}{RT}$$

(Eq 14)

where R = gas constant, T = absolute temperature, and p = absolute pressure. The temperature variation in the troposphere is given by:
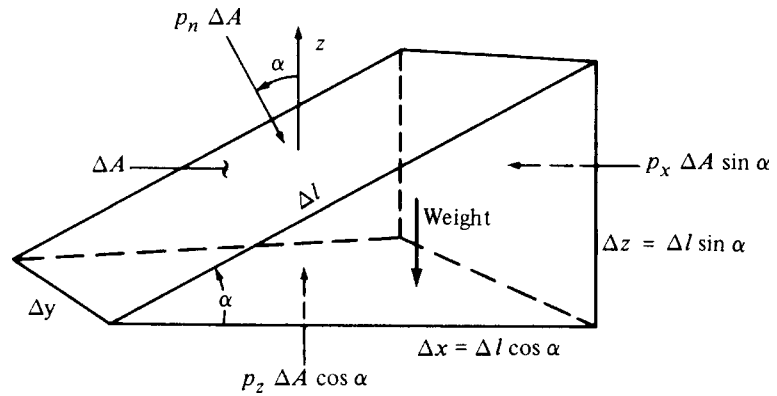
Fig. 3 Pressure forces on a fluid element

Fig. 2 Rheological behavior of viscous materials. Stress versus strain rate
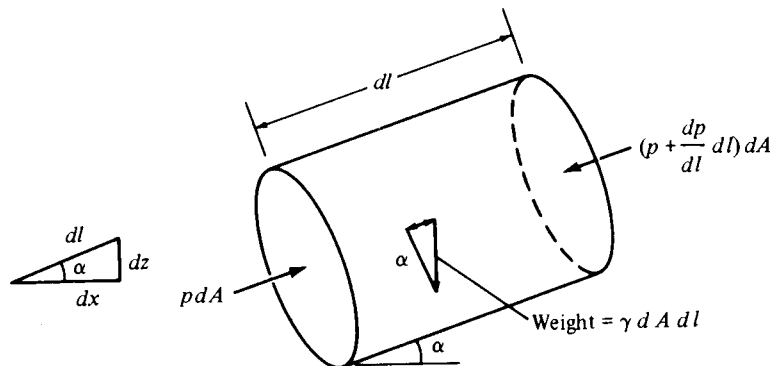
Fig. 4 Pressure variation with elevation

$$T = T_0 - \alpha(z - z_0) \tag{Eq 15}$$

where T_0 is the temperature at a reference level where the pressure is known, and α is the lapse rate. Using Eq 14, the pressure relation becomes:

$$\frac{dp}{dz} = -\frac{pg}{RT} \tag{Eq 16}$$

For the stratosphere, the temperature is assumed to be constant. Therefore:

$$\ln p = -\frac{zg}{RT} + C \tag{Eq 17}$$

Fig. 5 Open manometer for measuring p_A relative to atmospheric pressure
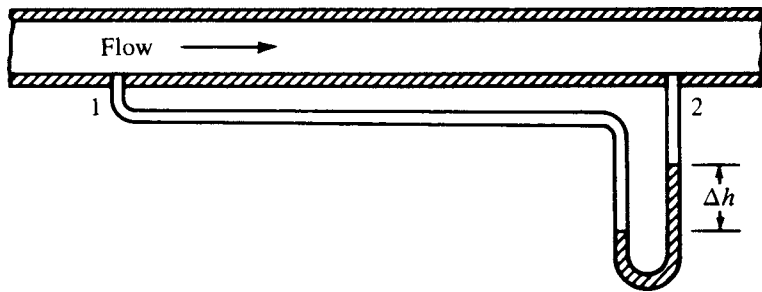
Fig. 6 Differential manometer
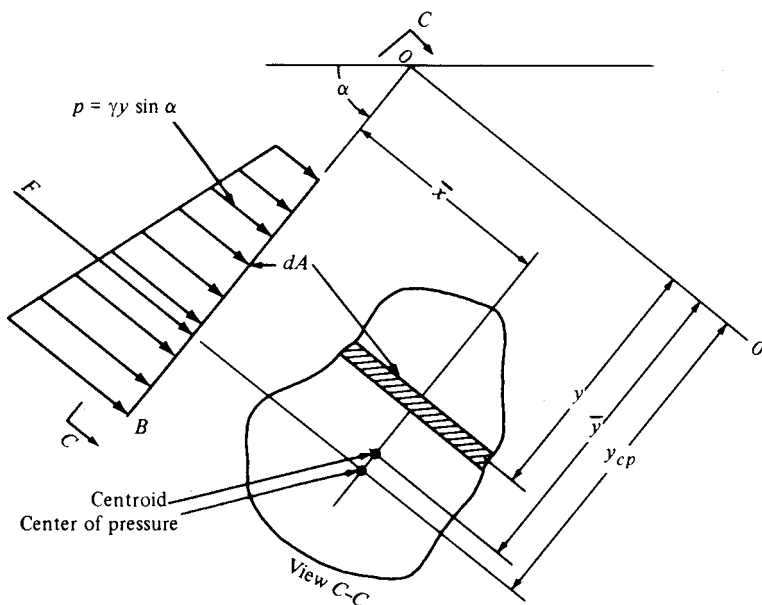
Fig. 7 Hydrostatic pressure distribution on a plane surface

At $z = z_0$, $p = p_0$ and:

$$p = p_0 e^{-(z-z_0)g/RT} \tag{Eq 18}$$

Manometers

The basic law governing manometers is Pascal's law: Any two points at the same elevation in a continuous mass of the same static fluid will be at the same pressure. For an open manometer (Fig. 5), the pressures are related by:

$$p_A = p_a - \rho_1 g(z_A - z_1) - \rho_2 g(z_1 - z_2) \tag{Eq 19}$$

where p_a is the atmospheric pressure.

For differential manometers, with a gas flowing, the pressure difference between points 1 and 2 is given by $\Delta p = \gamma_m \Delta h$, where γ_m is the specific weight of the manometer liquid, and Δh is the deflection of this liquid (Fig. 6). For a more dense fluid in the pipe:

$$\Delta p = (\gamma_m - \gamma_f)\Delta h \tag{Eq 20}$$

where γ_f = specific weight of fluid in the pipe and γ_m = specific weight of manometer fluid.

Forces on Submerged Surfaces

Plane (Flat) Surfaces. For the flat plate shown in Fig. 7, the differential force on the differential area, dA, is:

$$dF = p\, dA \text{ or } dF = \gamma y \sin \alpha dA$$

Total force on the area becomes:

$$F = \int_A p\, dA = \int_A \gamma y \, \sin \alpha \, dA \tag{Eq 21}$$

For constant γ and $\sin \alpha$, Eq 21 is integrated to give:

$$F = (\gamma y \, \sin \alpha)A \tag{Eq 22}$$

Therefore, the resultant hydrostatic force on a plane surface is the product of the pressure at the centroid of the surface and the area of the surface:

$$F = \overline{p}A \tag{Eq 23}$$

The position of the center of pressure (Fig. 7) is given by:

$$y_{cp} = \overline{y} + \frac{\overline{I}}{\overline{y}A} \tag{Eq 24}$$

where $\overline{y}$ is the centroid of the flat area and $\overline{I}$ = moment of inertia about the axis through the centroid and parallel to the surface of the fluid.

The location of the x-coordinate of the center of pressure is:

$$x_{cp} = \frac{I_{xy}}{\overline{y}A} + \overline{x} \tag{Eq 25}$$

where I_{xy} = product of inertia of the plate about the centroidal axes and $x_{cp} = \overline{x}$ if the plate is symmetric about one of the centroidal axes.

Curved Surfaces. Vertical forces on a fluid volume in contact with a curved surface (Fig. 8) are given by:

$$F_{Ry} = F_V + F_W \qquad \text{(Eq 26)}$$

where F_V is due to the pressure of the fluid along OB, F_W is due to the weight of the fluid in the volume AOB, and F_{Ry} is the vertical reaction of surface AB. The reaction F_{Rx} serves to
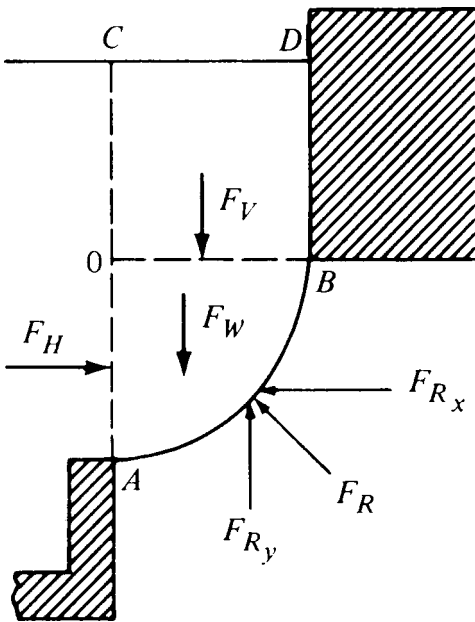
Fig. 8 Analysis of hydrostatic force on a curved surface

balance the pressure force caused by the fluid along OA. Forces on the curved surface can therefore be analyzed by applying the equations of equilibrium to the fluid volume (AOB) supported by the surface.

Buoyancy. The two basic laws concerning buoyancy are:

- A body immersed in a fluid experiences a vertical buoyant force equal to the weight of the fluid it displaces. The net vertical force on a submerged body is:

$$F_B = (\rho g) \text{ (body volume)} \qquad \text{(Eq 27)}$$

- A floating body displaces its own weight in the fluid in which it floats. The net vertical force on a floating body is given by:

$$F_B = (\rho g) \text{ (displaced volume)} \qquad \text{(Eq 28)}$$

Fluid Motion

Kinematic Relationships—Velocity and Acceleration

Velocity is a vector function of position and time with three components, u, v, and w, each of which is a scalar field in itself:

$$\mathbf{V}(x,y,z,t) = u(x,y,z,t)\mathbf{i} + v(x,y,z,t)\mathbf{j} + w(x,y,z,t)\mathbf{k} \qquad \text{(Eq 29)}$$

By mathematically mainpulating the velocity-field function, kinematic properties such as the acceleration vector, the local angular-velocity vector, and the volume flux through a surface can be derived.

The acceleration vector is found from the time derivative of the velocity vector, that is:

$$\mathbf{a} = \frac{d\mathbf{V}}{dt} = \frac{\partial \mathbf{V}}{\partial t} + \frac{\partial \mathbf{V}}{\partial x}\frac{dx}{dt} + \frac{\partial \mathbf{V}}{\partial y}\frac{dy}{dt} + \frac{\partial \mathbf{V}}{\partial z}\frac{dz}{dt} \qquad \text{(Eq 30)}$$

where $dx = u\,dt$ $dy = v\,dt$ and $dz = w\,dt$. Therefore:

$$\frac{d\mathbf{V}}{dt} = \frac{\partial \mathbf{V}}{\partial t} + \left(u\frac{\partial \mathbf{V}}{\partial x} + v\frac{\partial \mathbf{V}}{\partial y} + w\frac{\partial \mathbf{V}}{\partial z}\right) \qquad \text{(Eq 31)}$$

where the first term on the right side is called the local acceleration, and the three terms in the parentheses are called the convective acceleration. The total acceleration (Eq 31) of a particle can also be written:

$$\mathbf{a} = \frac{d\mathbf{V}}{dt} = \frac{\partial \mathbf{V}}{\partial t} + (\mathbf{V} \cdot \nabla)\mathbf{V} \qquad \text{(Eq 32)}$$

where:

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}$$

The operator, $d/dt = \partial/\partial t + (\mathbf{V} \cdot \nabla)$, may be applied to any fluid property, either scalar or vector, and is given the name substantial or material derivative.

Rotation of Fluid and Vorticity

Considering rotation about the z-axis in Fig. 9, the angular velocity is given by:

$$\omega_z = \frac{\omega_{AB} + \omega_{AC}}{2} \qquad \text{(Eq 33)}$$

where:

$$\omega_{AB} = \frac{\partial v}{\partial x}; \omega_{AC} = -\frac{\partial u}{\partial y}$$

Therefore:

$$\omega_z = \frac{1}{2}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) \qquad \text{(Eq 34)}$$

The rates of rotation about the x- and y-axes are:

$$\omega_x = \frac{1}{2}\left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\right) \qquad \text{(Eq 35)}$$

$$\omega_y = \frac{1}{2}\left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}\right) \qquad \text{(Eq 36)}$$

The vorticity vector for Cartesian coordinates is:

$$\mathbf{\Omega} = \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\right)\mathbf{i} + \left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}\right)\mathbf{j} + \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)\mathbf{k}$$
$$= 2 \cdot \omega = 2(\nabla \times \mathbf{V}) \qquad \text{(Eq 37)}$$

For incompressible flow:

$$\nabla \cdot \mathbf{V} = 0 \qquad \text{(Eq 38)}$$
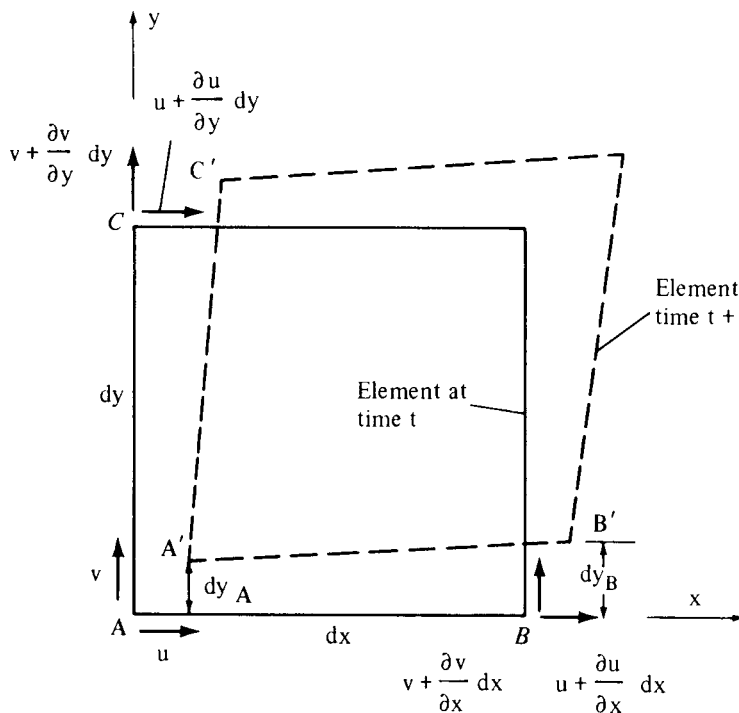
Fig. 9 Deformation of fluid element

Flow with negligible local angular velocity or vorticity is called irrotational, and:

$$\nabla \times \mathbf{V} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ u & v & w \end{vmatrix} = 0 \quad \text{(Eq 39)}$$

Equation 39 implies that for irrotational flow:

$$\frac{\partial w}{\partial y} = \frac{\partial v}{\partial z} \quad \frac{\partial u}{\partial z} = \frac{\partial w}{\partial x} \quad \frac{\partial v}{\partial x} = \frac{\partial u}{\partial y}$$

Flow Rate Through a Surface

As shown in Fig. 10, the volume that leaves through area dA in time dt is:

$$dV = V \, dt \, dA \, \cos\theta = (\mathbf{V}\cdot\mathbf{n})dA \, dt \quad \text{(Eq 40)}$$

The total volume rate of flow Q through the surface S is given by:

$$Q = \int_S (\mathbf{V}\cdot\mathbf{n})dA = \int_S V_n dA \quad \text{(Eq 41)}$$

where V_n is the component of $\mathbf{V}$ normal to dA, and $\mathbf{n}$ is the outward normal unit vector. The mass flow is expressed as:

$$\dot{m} = \int_S \rho(\mathbf{V}\cdot\mathbf{n})dA = \int_S \rho V_n dA \quad \text{(Eq 42)}$$

For constant density, it becomes:

$$\dot{m} = \rho Q \quad \text{(Eq 43)}$$

The average velocity passing through the surface can be computed from Q by:

$$V_{\text{avg}} = \frac{Q}{A} = \frac{\int_S V_n dA}{\int_S dA} \quad \text{(Eq 44)}$$
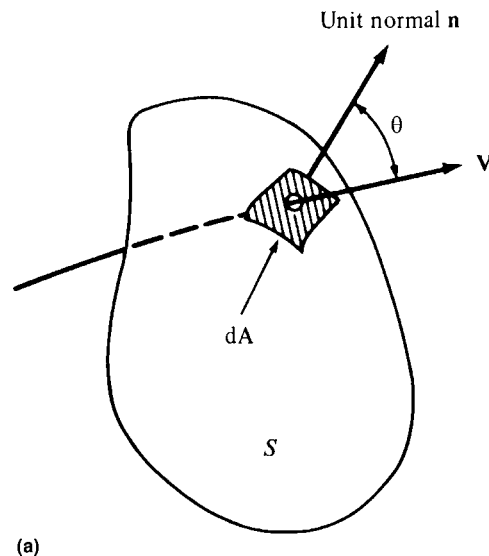
Streamlines

Streamlines are used to visualize the flow patterns in a fluid. For any given instant, they represent lines in the flow field across which no fluid passes. The velocity of every fluid particle on the streamline is in the direction tangent to the line, and when a sufficient number of such streamlines in the flow field are known, the flow pattern is determined, as shown in Fig. 11.

Mathematically, this means that every vector are length $d\mathbf{r}$ along a streamline must be tangent to $\mathbf{V}$, and their respective components must be in exact proportion. Therefore:

$$\frac{dx}{u} = \frac{dy}{v} = \frac{dz}{w} = \frac{dr}{V} \quad \text{(Eq 45)}$$

and:

$$\frac{dx}{ds} = u \quad \frac{dy}{ds} = v \quad \frac{dz}{ds} = w \quad \text{(Eq 46)}$$

where ds is a parameter equal to the ratios in Eq 45. These equations can be integrated to find streamline equations.

Variation of Pressure due to Rigid Body Motion

For a tank (Fig. 12) undergoing uniform acceleration, the angle of the liquid surface is:

$$\tan\alpha = \frac{a_x}{g} \quad \text{(Eq 47)}$$

Pressure can be easily determined using this angle to find the fluid depth.

For a tank (Fig. 13) undergoing rotation, the pressure variation becomes:

$$\frac{d}{dr}(p + \gamma z) = \rho r \omega^2 \quad \text{(Eq 48)}$$
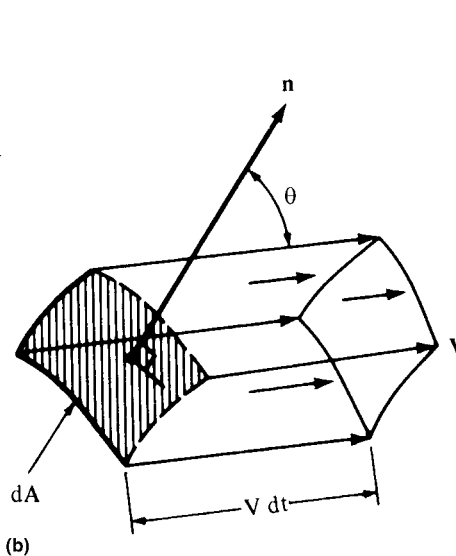
Integration produces:

$$p + \gamma z = \frac{\rho r^2 \omega^2}{2} + \text{constant}$$

or:

$$\frac{p}{\gamma} + z - \frac{V^2}{2g} = \text{constant} \quad \text{(Eq 49)}$$

Concept of the Control Volume

Eulerian and Lagrangian Views

In analyzing problems in mechanics, two approaches have evolved. In the Lagrangian method, equations of motion are written for a moving particle. Variations in a fluid property are monitored for each individual particle as it moves through the flow field, and the property, say pressure, is considered a function of time, $p = p(t)$.

The Eulerian approach requires analysis of the fluid particles as they pass given locations in the flow field. The fluid property is considered a function of the spatial coordinates as well as time and $p = p(x, y, z, t)$. This is often called the control-volume viewpoint and is the one most often used in fluid mechanics.

Control-Volume Transport Equation

The symbol B is used to represent a general extensive property, and β is used for the corresponding intensive property. The volume rate of flow past a given area A can be written as $Q = \mathbf{V} \cdot \mathbf{A}$ if the velocity vector $\mathbf{V}$ is considered uniform over the vector area $\mathbf{A}$. For the control volume (cv) of Fig. 14, the net volumetric flow rate out of the control volume is:

$$Q_{\text{out}} - Q_{\text{in}} = \mathbf{V}_2 \cdot \mathbf{A}_2 + \mathbf{V}_1 \cdot \mathbf{A}_1 \quad \text{(Eq 50)}$$
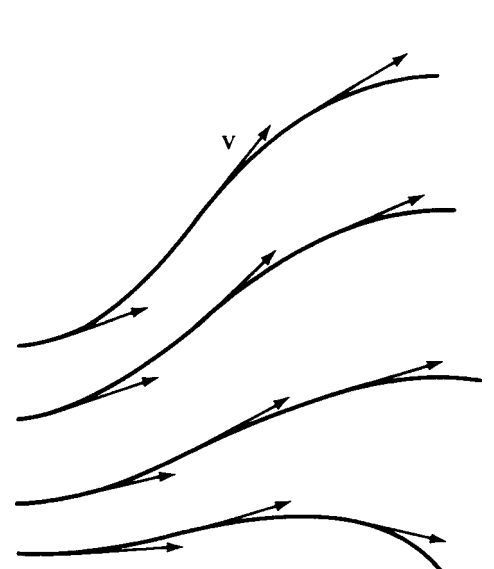
Fig. 10 Volume rate of flow through an arbitrary surface. (a) An elemental area dA on the surface. (b) The incremental volume moving through dA equals $VdtdA \cos\theta$

Fig. 11 Streamlines describing a flow pattern

The net mass flow rate then becomes:

$$\dot{m} = \sum_{cs} \rho \mathbf{V} \cdot \mathbf{A} \qquad \text{(Eq 51)}$$

where the summation indicates that several inflow and outflow areas can be considered. This is similar to Eq 42.

For the rate of flow of an extensive property, B, out of the control volume, the mass rate is multiplied by the intensive property, β:

$$\dot{B} = \sum_{cs} \beta \rho \mathbf{V} \cdot \mathbf{A} \qquad \text{(Eq 52)}$$

Generally, this equation is used in integral form:

$$\dot{B} = \int_{cs} \beta \rho \mathbf{V} \cdot d\mathbf{A} \qquad \text{(Eq 53)}$$

The transport equation relates the rate of change of property B for a fixed amount of mass (system) to the change in the property as it relates to the control volume. The equation is:

$$\frac{dB_{\text{sys}}}{dt} = \frac{\partial}{\partial t} \int_{cv} \beta \rho dV + \int_{cs} \beta \rho \mathbf{V} \cdot d\mathbf{A} \qquad \text{(Eq 54)}$$

where the left side is the rate of change of the extensive property of the system. The right side refers to accumulation of the property in the control volume and flow of the property across the control surface (cs). For steady flow, the equation becomes:
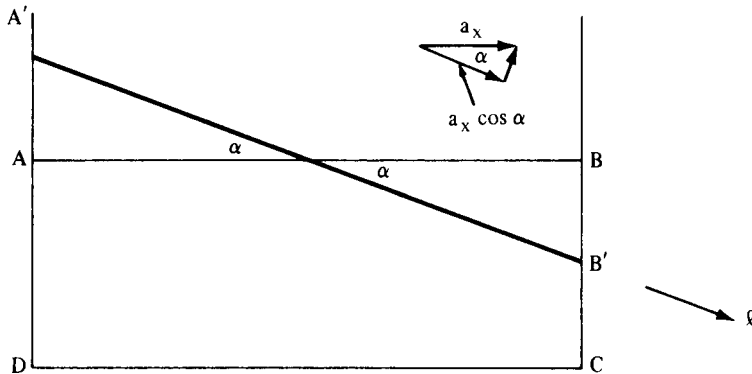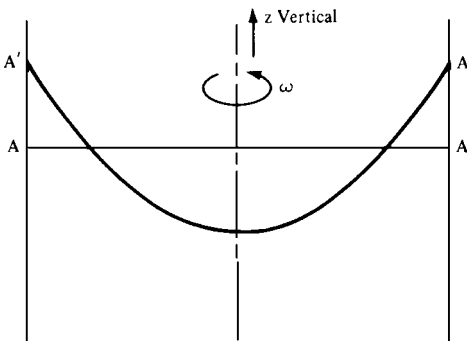
$$\frac{dB_{\text{sys}}}{dt} = \int_{cs} \beta \rho \mathbf{V} \cdot d\mathbf{A} \qquad \text{(Eq 55)}$$

Note that B can be either a scalar or vector quantity.

For a control volume moving at constant velocity with respect to an inertial frame:

$$\mathbf{V}_r = \mathbf{V} - \mathbf{V}_s \qquad \text{(Eq 56)}$$

and:

$$\frac{d}{dt}(B_{\text{sys}}) = \int_{cv} \frac{\partial}{\partial t}(\beta \rho)dV + \int_{cs} \beta \rho (\mathbf{V}_r \cdot \mathbf{n})dA \qquad \text{(Eq 57)}$$

where $\mathbf{V}_r$ = fluid velocity relative to the control surface, $\mathbf{V}_s$ = velocity of the control volume, and $\mathbf{V}$ = absolute fluid velocity.

For the most general situation, the control volume is both moving and deforming arbitrarily. The control surface has a deformation, its velocity $\mathbf{V}_s = \mathbf{V}_s(\mathbf{r}, t)$, so that the relative velocity becomes $\mathbf{V}_r = \mathbf{V}(\mathbf{r}, t) - \mathbf{V}_s(\mathbf{r}, t)$. This may be a complicated function. The volume integral in Eq 57 must allow the volume elements to distort with time, and therefore, the time derivative must be applied after integration. The transport theorem becomes:

$$\frac{d}{dt}(B_{\text{sys}}) = \frac{\partial}{\partial t}\left(\int_{cv} \beta \rho dV\right) + \int_{cs} \beta \rho (\mathbf{V}_r \cdot \mathbf{n})dA \qquad \text{(Eq 58)}$$

Laws to which the Control-Volume Concept is Applied

The time rate of change of total mass, M, of a system of particles (conservation of mass) is:

$$\frac{dM}{dt} = 0 \qquad \text{(Eq 59)}$$

The time rate of change of momentum of a system of particles is equal to the sum of the externally applied forces (Newton's second law):

$$\sum \mathbf{F} = \frac{d(M\mathbf{V})}{dt} \qquad \text{(Eq 60)}$$

The rate of change of total energy (E) of a system of particles is equal to the rate of addition of heat ($\tilde{Q}$) less the rate of work (W) done by the system (conservation of energy):

$$\frac{dE}{dt} = \frac{d\tilde{Q}}{dt} - \frac{dW}{dt} \qquad \text{(Eq 61)}$$

These laws express the time rate of change of fluid quantities for a system of particles and must now be applied to the control volume.

Continuity Equation

Integral Formulation

The general form of the continuity equation (conservation of mass) is:

$$\int_{cs} \rho \mathbf{V} \cdot d\mathbf{A} = -\frac{\partial}{\partial t} \int_{cv} \rho dV \qquad \text{(Eq 62)}$$

This states that the net rate of mass outflow from the control volume equals the rate of decrease of mass within the control volume. For steady flow:

$$\int_{cs} \rho \mathbf{V} \cdot d\mathbf{A} = 0 \qquad \text{(Eq 63)}$$

If the flow is steady and the velocity is constant across several inflow and outflow sections:

$$\sum \rho \mathbf{V} \cdot \mathbf{A} = 0$$

For one inflow and one outflow, such as seen in Fig. 14:

Fig. 12 Uniform acceleration of a tank of liquid

Fig. 13 Rigid body rotation of a liquid in a tank
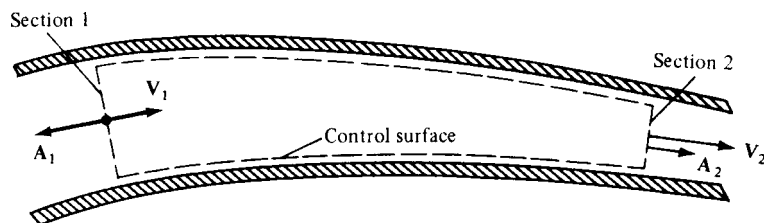
Fig. 14 Control-volume example

$$\rho_1 V_1 A_1 = \rho_2 V_2 A_2 \qquad \text{(Eq 64)}$$

If ρ is constant, then:

$$V_1 A_1 = V_2 A_2 \qquad \text{(Eq 65)}$$

Differential Formulation

The differential formulation may be derived by applying the transport equation to a differential control volume. For Cartesian coordinates, the continuity equation is:

$$\frac{\partial}{\partial x}(\rho u) + \frac{\partial}{\partial y}(\rho v) + \frac{\partial}{\partial z}(\rho w) = -\frac{\partial \rho}{\partial t} \qquad \text{(Eq 66)}$$

If the flow is steady, this becomes:

$$\frac{\partial}{\partial x}(\rho u) + \frac{\partial}{\partial y}(\rho u) + \frac{\partial}{\partial z}(\rho w) = 0 \qquad \text{(Eq 67)}$$

For an incompressible fluid, then:

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} + \frac{\partial w}{\partial z} = 0 \qquad \text{(Eq 68)}$$

for both steady and unsteady flow. In vector notation, Eq 68 becomes:

$$\nabla \cdot \mathbf{V} = 0 \qquad \text{(Eq 69)}$$

Momentum Equation

Integral Form—Linear Momentum

From Newton's second law, it can be shown that the summation of all external forces on a system equals the net rate at which momentum crosses the control surface plus the rate at which momentum is accumulated inside the control volume:

$$\Sigma \mathbf{F} = \int_{cs} \mathbf{V}_\rho \mathbf{V} \cdot d\mathbf{A} + \frac{\partial}{\partial t} \int_{cv} \mathbf{V}\rho \, dV \qquad \text{(Eq 70)}$$

The two types of forces usually considered are surface forces and body forces:

$$\sum \mathbf{F} = \sum \mathbf{F}_S + \Sigma \mathbf{F}_B \qquad \text{(Eq 71)}$$

In Cartesian coordinates, the momentum equation for each direction is given as follows:
x-direction:

$$\sum F_x = \sum_{cs} u(\rho \mathbf{V} \cdot \mathbf{A}) + \frac{\partial}{\partial t} \int_{cv} u\rho \, dV \qquad \text{(Eq 72)}$$

y-direction:

$$\sum F_y = \sum_{cs} v(\rho \mathbf{V} \cdot \mathbf{A}) + \frac{\partial}{\partial t} \int_{cv} v\rho \, dV \qquad \text{(Eq 73)}$$

z-direction:

$$\sum F_z = \sum_{cs} w(\rho \mathbf{V} \cdot \mathbf{A}) + \frac{\partial}{\partial t} \int_{cv} w\rho \, dV \qquad \text{(Eq 74)}$$

The term $\sum\limits_{cs}$:

used in Eq 72 through 74 indicates that the velocities are uniformly distributed across the areas.

The use of an average (uniform) velocity across a surface results in an error. The simple one-dimensional momentum flux calculation:

$$\int_{cs} \mathbf{V}(\rho \mathbf{V} \cdot \mathbf{n})dA = \int \rho V^2 dA = \rho A \bar{V}^2$$

where $\bar{V}$ represents the average velocity, should be modified using the term $\alpha \rho A \bar{V}^2$. The factor α is the dimensionless momentum flux correction factor, $\alpha \geq 1$, defined as:

$$\alpha = \int \left(\frac{V}{\bar{V}}\right)^2 dA \qquad \text{(Eq 75)}$$

where V is the actual velocity (spatially varying).

Integral Form—Angular Momentum (Moment of Momentum)

The extensive property describing angular momentum for the system is:

$$B_{\text{sys}} = \int (\mathbf{r} \times \mathbf{V})\rho \, dV$$
$$= \text{angular momentum of the system} \qquad \text{(Eq 76)}$$

The general form of the angular momentum equation becomes:

$$\sum \mathbf{M} = \int_{cs} (\mathbf{r} \times \mathbf{V})\rho \mathbf{V} \cdot d\mathbf{A} + \frac{\partial}{\partial t} \int_{cv} (\mathbf{r} \times \mathbf{V})\rho \, dV \qquad \text{(Eq 77)}$$

For uniform velocity across the inlet and outlet areas:

$$\sum \mathbf{M} = \sum_{cs} (\mathbf{r} \times \mathbf{V})\rho \mathbf{V} \cdot \mathbf{A} + \frac{\partial}{\partial t} \int_{cv} (\mathbf{r} \times \mathbf{V})\rho \, dV \qquad \text{(Eq 78)}$$

Differential Form—Linear Momentum

As previously stated, the forces in a linear momentum system are of two types: body forces and surface forces. Body forces are due to external fields (gravity, magnetism, electric potential) that act on the entire mass within the element. The only body force considered here is gravity. The surface forces are caused by stresses on a fluid element, the general state of stress for an element being shown in Fig. 15. The stress tensor is defined as:

$$\tau_{ij} = \begin{vmatrix} -p + \tau'_{xx} & \tau_{yx} & \tau_{zx} \\ \tau_{xy} & -p + \tau'_{yy} & \tau_{zy} \\ \tau_{xz} & \tau_{yz} & -p + \tau'_{zz} \end{vmatrix} \qquad \text{(Eq 79)}$$

The basic differential momentum equation for an infinitesimal element now becomes:

$$\rho\mathbf{g} - \nabla p + \nabla \cdot \tau'_{ij} = \rho \frac{d\mathbf{V}}{dt} \qquad \text{(Eq 80)}$$
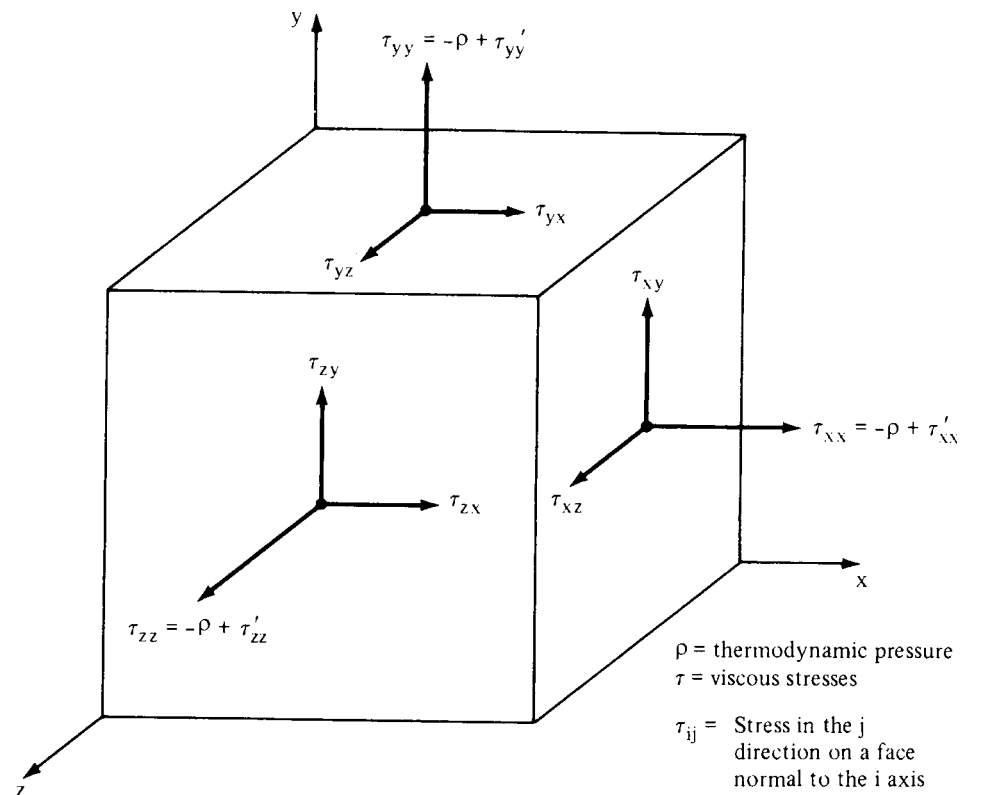
Fig. 15 Notation for stresses on a fluid element

where:

$$\frac{d\mathbf{V}}{dt} = \frac{\partial \mathbf{V}}{\partial t} + u\frac{\partial \mathbf{V}}{\partial x} + v\frac{\partial \mathbf{V}}{\partial y} + w\frac{\partial \mathbf{V}}{\partial z}$$

and:

$$\tau'_{ij} = \begin{vmatrix} \tau'_{xx} & \tau_{yx} & \tau_{zx} \\ \tau_{xy} & \tau'_{yy} & \tau_{zy} \\ \tau_{xz} & \tau_{yz} & \tau_{zz} \end{vmatrix}$$

This last tensor is called the viscous stress tensor and results from removing the thermodynamic pressure from Eq 79. The third term in Eq 80 represents the viscous force per unit volume exerted on the fluid element.

In Cartesian coordinates, the three components of the momentum equation are:

$$\rho g_x - \frac{\partial p}{\partial x} + \frac{\partial \tau'_{xx}}{\partial x} + \frac{\partial \tau_{yx}}{\partial y} + \frac{\partial \tau_{zx}}{\partial z}$$
$$= \rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}\right)$$
$$\rho g_y - \frac{\partial p}{\partial y} + \frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \tau'_{yy}}{\partial y} + \frac{\partial \tau_{zy}}{\partial z}$$
$$= \rho\left(\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z}\right)$$
$$\rho g_z - \frac{\partial p}{\partial z} + \frac{\partial \tau_{xz}}{\partial x} + \frac{\partial \tau_{yz}}{\partial y} + \frac{\partial \tau'_{zz}}{\partial z}$$
$$= \rho\left(\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z}\right) \quad \text{(Eq 81)}$$

For frictionless flow, $\tau'_{ij} = 0$, and the equation becomes:

$$\rho\mathbf{g} - \nabla p = \rho\frac{d\mathbf{V}}{dt} \quad \text{(Eq 82)}$$

This is Euler's equation for inviscid flow. It can be integrated along a streamline to get the Bernoulli equation.

For a Newtonian fluid, the viscous stresses are proportional to the strain rates of the fluid element and the coefficient of viscosity. If the fluid is considered to be both isotropic and incompressible, the terms of the stress tensor become:

$$\tau'_{xx} = 2\mu\frac{\partial u}{\partial x} \quad \tau'_{yy} = 2\mu\frac{\partial v}{\partial y} \quad \tau'_{zz} = 2\mu\frac{\partial w}{\partial z}$$
$$\tau_{xy} = \tau_{yx} = \mu\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) \quad \tau_{xz} = \tau_{zx} = \mu\left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right)$$
$$\tau_{yz} = \tau_{zy} = \mu\left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right)$$
$$\text{(Eq 83)}$$

where μ is the viscosity coefficient. The momentum equation for a Newtonian fluid with constant density and viscosity is now written as:

$$\rho g_x - \frac{\partial p}{\partial x} + \mu\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) = \rho\frac{du}{dt}$$
$$\rho g_y - \frac{\partial p}{\partial y} + \mu\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2}\right) = \rho\frac{dv}{dt}$$
$$\rho g_z - \frac{\partial p}{\partial z} + \mu\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}\right) = \rho\frac{dw}{dt}$$
$$\text{(Eq 84)}$$

These are referred to as the Navier-Stokes equations. In vector form, this set has the form:

$$\rho\frac{d\mathbf{V}}{dt} = \rho\mathbf{g} - \nabla p + \mu\nabla^2\mathbf{V} \quad \text{(Eq 85)}$$

The d/dt terms in Eq 80, 84, and 85 refer to the substantial derivative.

Energy Equation

Integral Form

The symbol E refers to the total energy of the system, and e is the energy per unit mass. From the first law of thermodynamics, $dE/dt = \dot{Q} - \dot{W}$ and the energy equation becomes:

$$\dot{Q} - \dot{W} = \frac{\partial}{\partial t}\int_{cv} e\rho\, dV + \sum_{cs} e\rho\mathbf{V}\cdot\mathbf{A} \quad \text{(Eq 86)}$$

where the velocity and the energy crossing the control surface are considered constant over a flow area. The $\dot{Q}$ and $\dot{W}$ indicate rates for heat transfer and work, respectively.

The term e can be replaced by its equivalent:

$$e = \hat{u} + \frac{V^2}{2} + zg$$

These terms represent the internal energy, kinetic energy, and potential energy per unit mass. The energy equation becomes:

$$\dot{Q} - \dot{W} = \frac{\partial}{\partial t}\int_{cv}\left(\frac{V^2}{2} + gz + \hat{u}\right)\rho\, dV$$
$$+ \sum_{cs}\left(\frac{V^2}{2} + gz + \hat{u}\right)\rho\mathbf{V}\cdot\mathbf{A} \quad \text{(Eq 87)}$$

Work is the sum of the shaft work and the flow work.

Flow work is the work done by pressure forces as the system moves through space. This is expressed for all streams passing through the control surface as:

$$\dot{W}_f = \sum_{cs} p\mathbf{V}\cdot\mathbf{A} \quad \text{(Eq 88)}$$

where area vector $\mathbf{A}$ is considered to be directed outward. Shaft work is defined as any work other than flow work.

Using flow work and shaft work, the energy equation takes the form:

$$\dot{Q} - \dot{W}_s = \frac{\partial}{\partial t}\int_{cv}\left(\frac{V^2}{2} + gz + \hat{u}\right)\rho\, dV +$$
$$\sum_{cs}\left(\frac{p}{\rho} + \frac{V^2}{2} + gz + \hat{u}\right)\rho\mathbf{V}\cdot\mathbf{A} \quad \text{(Eq 89)}$$

If the velocity or the other properties are not uniform across the section, then:

$$\dot{Q} - \dot{W}_s = \frac{\partial}{\partial t}\int_{cv}\left(\frac{V^2}{2} + gz + \hat{u}\right)\rho\, dV +$$
$$\int_{cs}\left(\frac{p}{\rho} + \frac{V^2}{2} + gz + \hat{u}\right)\rho\mathbf{V}\cdot d\mathbf{A} \quad \text{(Eq 90)}$$

For steady flow:

$$\dot{Q} - \dot{W}_s = \sum_{cs}\left(\frac{V^2}{2} + gz + h\right)\rho\mathbf{V}\cdot\mathbf{A} \quad \text{(Eq 91)}$$

where the property of enthalpy, $h = p/\rho + \hat{u}$, has been employed. With a single inflow and outflow, this equation reduces to:

$$\frac{1}{\dot{m}}(\dot{Q} - \dot{W}_s) + gz_1 + h_1 + \frac{V_1^2}{2} = gz_2 + h_2 + \frac{V_2^2}{2}$$
$$\text{(Eq 92)}$$

The kinetic energy terms in Eq 92 are not exact, because V_1 and V_2 are assumed to be constant across the input and output areas. If we consider for the moment that $\bar{V}_1$ and $\bar{V}_2$ represent these average (uniform) velocities, then the proper kinetic energy terms are:

$$\lambda_1\frac{\bar{V}_1^2}{2} \text{ and } \lambda_2\frac{\bar{V}_2^2}{2}$$

where λ_1 and λ_2 are the kinetic energy correction factors. The term λ is defined by:

$$\lambda = \frac{1}{A}\int\left(\frac{V}{\bar{V}}\right)^3 dA \quad \text{(Eq 93)}$$

where V is the actual velocity (spatially varying) at the control surface. Some common values for λ are 1.0, constant velocity over the section; 2.0, parabolic laminar flow; and 1.05, turbulent flow. The momentum correction factors of Eq 75 do not vary from 1.0 as much as the values of λ. The use of the correction factors do not account for the rather remote possibility that enthalpy varies across the section.

Equation 92 may be rewritten in the following form:

$$\frac{p_1}{\gamma} + \frac{V_1^2}{2g} + z_1 = \frac{p_2}{\gamma} + \frac{V_2^2}{2g} + z_2 + \frac{\dot{W}}{\dot{m}g} + h_L \quad \text{(Eq 94)}$$

This arrangement assigns units of length (feet, meters) to each term. The symbol h_L represents losses between points 1 and 2 (head loss) and includes heat transfer from the control volume resulting from viscous dissipation and any loss of mechanical energy in general.

For a stream tube (a filament of fluid of infinitesimal cross section bounded by streamlines) containing an inviscid fluid and without any shaft work, Eq 94 leads to Bernoulli's equation:

$$\frac{p_1}{\gamma} + \frac{V_1^2}{2g} + z_1 = \frac{p_2}{\gamma} + \frac{V_2^2}{2g} + z_2 \quad \text{(Eq 95)}$$

Equation 95 can be derived either from the energy equation, as given previously, by integrating Euler's equation (Eq 82) along a streamline. It is thus considered valid if applied along a streamline in a frictionless, incompressible, steady flow with no shaft work, and also may be shown to apply between any two points in an incompressible, irrotational, steady flow field.

Differential Formulation

The heat transfer across the control surface is considered here to be by conduction only. Expressed in vector form, Fourier's law shows that:

$$\dot{Q} = \nabla\cdot(k\nabla T)dx\,dy\,dz \tag{Eq 96}$$

where k is the thermal conductivity. The net viscous-work rate can be shown to be:

$$\dot{W}_v = -\left[\frac{\partial}{\partial x}\left(u\tau'_{xx} + v\tau_{xy} + w\tau_{xz}\right)\right.$$
$$+ \frac{\partial}{\partial y}\left(u\tau_{yx} + v\tau'_{yy} + w\tau_{yz}\right)$$
$$+ \left.\frac{\partial}{\partial z}\left(u\tau_{zx} + v\tau_{zy} + w\tau'_{zz}\right)\right]dx\,dy\,dz$$
$$= -\nabla\cdot\left(\mathbf{V}\cdot\tau'_{ij}\right)dx\,dy\,dz \tag{Eq 97}$$

The resulting form of the differential energy equation is:

$$\rho\frac{de}{dt} + \mathbf{V}\cdot\nabla p = \nabla\cdot(k\,\nabla\,T) + \nabla\cdot\left(\mathbf{V}\cdot\tau'_{ij}\right) \tag{Eq 98}$$

where $e = \hat{u} + 1/2V^2 + gz$. The viscous-work term can be separated as follows:

$$\nabla\cdot\left(\mathbf{V}\cdot\tau'_{ij}\right) \equiv \mathbf{V}\cdot\left(\nabla\cdot\tau'_{ij}\right) + \Phi \tag{Eq 99}$$

where Φ is the viscous dissipation function. For an incompressible Newtonian fluid:

$$\Phi = \mu\left[2\left(\frac{\partial u}{\partial x}\right)^2 + 2\left(\frac{\partial v}{\partial y}\right)^2 + 2\left(\frac{\partial w}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)^2\right.$$
$$\left. + \left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}\right)^2 + \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right)^2\right] \tag{Eq 100}$$

The viscous dissipation is always positive.

The usual vector form of the general differential energy equation that can be obtained using Eq 98, 99, and 100 is:

$$\rho\frac{\partial\hat{u}}{\partial t} + p(\nabla\cdot\mathbf{V}) = \nabla\cdot(k\nabla T) + \Phi \tag{Eq 101}$$

This form neglects radiation and internal energy generation. Using the approximation:

$$d\hat{u} \approx c_v dT$$

where c_v (constant volume specific heat), μ, k, and ρ are constants, the energy equation takes the form:

$$\rho c_v\frac{dT}{dt} = k\nabla^2 T + \Phi \tag{Eq 102}$$

where:

$$\frac{dT}{dt} = \frac{\partial T}{\partial T} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} + w\frac{\partial T}{\partial z}$$

Boundary Conditions for the Differential Equations

Typical boundary conditions are shown in Fig. 16 and described subsequently.

Initial Condition. For unsteady flow, there must be an initial condition or a known initial spatial distribution for each variable: ρ, $\mathbf{V}$, p, $\hat{u}$, and T.

Solid Wall. For a solid, impermeable wall, there is no velocity slip and no temperature jump in a viscous heat-conducting fluid:

$$\mathbf{V}_{\text{fluid}} = \mathbf{V}_{\text{wall}} \qquad T_{\text{fluid}} = T_{\text{wall}} \tag{Eq 103}$$

An exception to Eq 103 exists in an extremely rarefied gas flow, where slippage can occur.

Inlet or Outlet of Control Volume. At any inlet or outlet section of the flow, complete distribution of velocity, pressure, and temperature must be known for all times.

Liquid/Gas Interface. The interface is denoted by $z = \eta(x, y, t)$. There is equality of vertical velocity across the interface, so that no holes appear between liquid and gas and:

$$w_{\text{liq}} = w_{\text{gas}} = \frac{d\eta}{dt} = \frac{\partial\eta}{\partial t} + u\frac{\partial\eta}{\partial x} + v\frac{\partial\eta}{\partial y} \tag{Eq 104}$$

This is the kinematic boundary condition. The viscous-shear stresses must also balance:

$$(\tau_{zy})_{\text{liq}} = (\tau_{zy})_{\text{gas}} \qquad (\tau_{zx})_{\text{liq}} = (\tau_{zx})_{\text{gas}} \tag{Eq 105}$$

The pressures must balance at the interface except for surface-tension effects:

$$p_{\text{liq}} = p_{\text{gas}} - \Gamma(R_x^{-1} + R_y^{-1}) \tag{Eq 106}$$

where:

$$R_x^{-1} + R_y^{-1} = \frac{\partial^2\eta/\partial x^2 + \partial^2\eta/\partial y^2}{[1 + (\partial\eta/\partial x)^2 + (\partial\eta/\partial y)^2]^{3/2}}$$

and Γ is the surface tension coefficient.

The heat transfer must be the same on both sides of the interface:

$$(\tilde{Q}_z)_{\text{liq}} = (\tilde{Q})_{\text{gas}} \tag{Eq 107a}$$

or:

$$\left(k\frac{\partial T}{\partial z}\right)_{\text{liq}} = \left(k\frac{\partial T}{\partial z}\right)_{\text{gas}} \tag{Eq 107b}$$

Equation 107(b) accounts for only conduction at the surface.

Simplified Free Surface. It is assumed that the upper fluid only exerts pressure on the lower surface, and shear, heat transfer, and non-linear slope terms are neglected. The conditions become:

$$p_{\text{liq}} \approx p_{\text{gas}} - \Gamma\left(\frac{\partial^2\eta}{\partial x^2} + \frac{\partial^2\eta}{\partial y^2}\right) \tag{Eq 108}$$
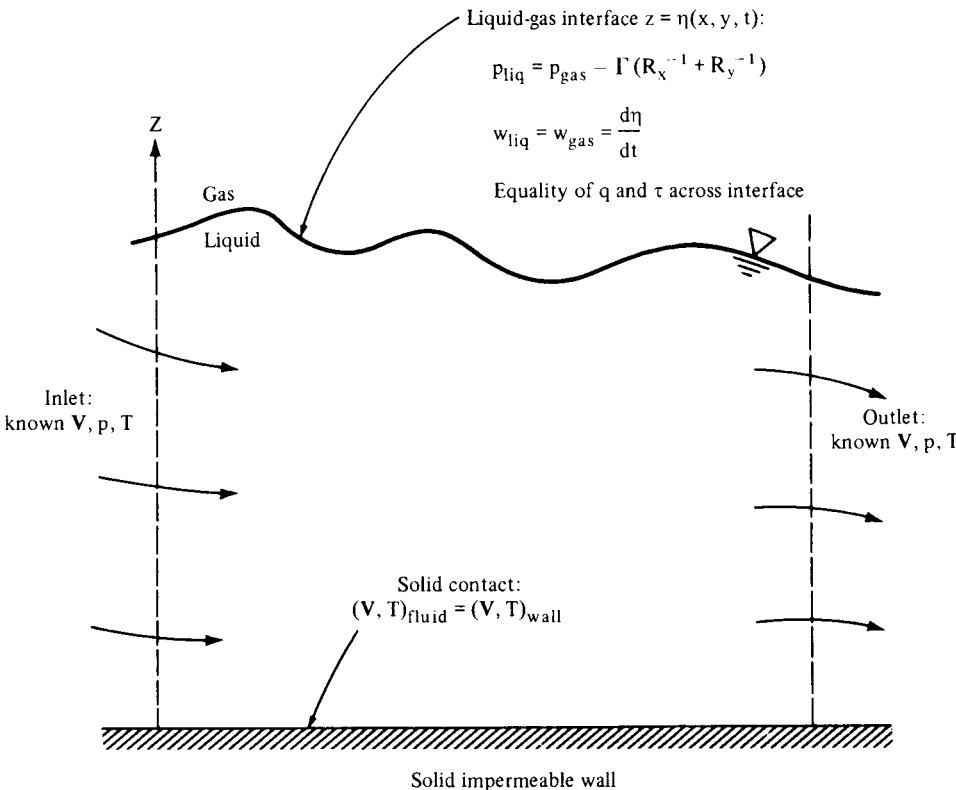
Fig. 16 Typical boundary conditions in viscous heat-conducting fluid flow. Source: Ref 1 with permission

where:

$$w_{\text{liq}} \approx \frac{\partial \eta}{\partial t}; \left(\frac{\partial V}{\partial z}\right)_{\text{liq}} \approx 0; \left(\frac{\partial T}{\partial z}\right)_{\text{liq}} \approx 0$$

For open-channel flow, surface tension is also neglected and:

$$p_{\text{liq}} \approx p_{\text{atm}} \qquad \text{(Eq 109)}$$

Inviscid Flow. Only the normal velocities must match at any solid surface:

$$(V_n)_{\text{fluid}} = (V_n)_{\text{wall}} \qquad \text{(Eq 110)}$$

Dimensional Analysis

Dimensional analysis enables the formulation of functional relationships in a set of nondimensional groups composed of physical variables, with the groups numbering less than the variables. For dimensional analysis, physical variables must be known or assumed. An equation expressing a relationship between flow quantities must be dimensionally homogeneous. Physical quantities used in fluid-flow problems can be expressed in terms of the fundamental units of mass (M), length (L), and time (T). Temperature is not required in basic fluid problems.

The dimensions of various physical quantities used in fluid mechanics are presented in Table 1. This table is based on the M, L, T primary system.

The Pi Theorem

This theorem is critical to all applications of dimensional analysis and can be stated as follows: if a dimensional quantity u, an unknown,

Table 1 Dimensions of various physical quantities

Quantity	Symbol	Dimensions (M, L, T)
Length	l	L
Time	t	T
Mass	m	M
Force	F	ML/T^2
Velocity (linear)	V	L/T
Acceleration (linear)	a	L/T^2
Area	A	L^2
Volume	V	L^3
Pressure	p	M/LT^2
Density	ρ	M/L^3
Acceleration due to gravity	g	L/T^2
Dynamic viscosity	μ	M/LT
Kinematic viscosity	v	L^2/T
Surface tension	σ	M/T^2
Angle (radians)	θ	No dimensions
Velocity (angular)	ω	$1/T$
Acceleration (angular)	α	$1/T^2$
Torque or moment	T_o	ML^2/T^2
Work, energy	W	ML^2/T^2
Momentum (linear)	mV	ML/T
Volume flow rate	Q	L^3/T
Mass flow rate	$\dot{m}$	M/T
Power	P	ML^2/T^3
Moment of inertia	I	ML^2
Momentum (angular)	$I\omega$	ML^2/T

Source: Ref 2 with permission

is expressed in terms of n other dimensional quantities $x_1, \ldots, x_n$ by:

$$u = f(x_1, \ldots, x_n) \qquad \text{(Eq 111)}$$

and if the first m variables $x_1, \ldots, x_m$ have independent dimensions, then Eq 111 can be written as:

$$\Pi = f(1, 1, \ldots, \Pi_1, \ldots, \Pi_{n-m}) \qquad \text{(Eq 112)}$$

The terms Π, Π_1, Π_2, ..., Π_{n-m} are forms of u, $x_{m+1}, \ldots, x_n$, respectively, made dimensionless by forming combinations with $x_1, \ldots, x_m$.

Any dimensional relationship that expresses one physical variable, the unknown, in terms of n other physical variables, of which m have independent dimensions, can be arranged as a relationship between the unknown in dimensionless form and $n - m$ dimensionless combinations of the remaining variables. The pi theorem is valid only if the relationship in Eq 111 is true regardless of what system of units is being used.

Power Product Method

This method is a practical application of the pi theorem, and it is best described using an example taken from Ref 2 and used by permission. To determine the frictional force exerted by a flowing fluid on a smooth pipe, assume that the significant physical quantities are the fluid density, ρ, the average fluid velocity, V, the interior surface area of pipe, S, the dynamic viscosity, μ, the internal diameter of pipe, D, and the frictional force, F, exerted on the pipe by the fluid. This relationship takes the form:

$$F = F(\rho, V, S, \mu, D) \qquad \text{(Eq 113)}$$

This type of expression can generally be expanded into a power series in the form:

$$F = c_1 \rho^{\alpha_1} V^{\alpha_2} S^{\alpha_3} \mu^{\alpha_4} D^{\alpha_5} + c_2 \rho^{\beta_1} V^{\beta_2} S^{\beta_3} \mu^{\beta_4} D^{\beta_5} + \ldots \qquad \text{(Eq 114)}$$

where the α and β terms represent numerical exponents that depend on the nature of the function F, and the c's represent dimensionless numerical coefficients. The terms on the right side of Eq 114 are added, so every term must have the same dimensions.

Substituting the physical dimensions from Table 1 into Eq 114 produces:

$$\frac{ML}{T^2} = \left(\frac{M}{L^3}\right)^{\alpha_1} \left(\frac{L}{T}\right)^{\alpha_2} (L^2)^{\alpha_3} \left(\frac{M}{LT}\right)^{\alpha_4} L^{\alpha_5} + \left(\frac{M}{L^3}\right)^{\beta_1} \left(\frac{L}{T}\right)^{\beta_2} (L^2)^{\beta_3} \left(\frac{M}{LT}\right)^{\beta_4} L^{\beta_5} + \ldots$$

For each term, like exponents for the primary dimensions M, L, and T are collected to obtain three algebraic equations with five unknowns. For the first term, we obtain:

For M : $1 = \alpha_1 + \alpha_4$

For L : $1 = -3\alpha_1 + \alpha_2 + 2\alpha_3 - \alpha_4 + \alpha_5$

For T : $-2 = -\alpha_2 - \alpha_4$

For the second term we obtain:

For M : $1 = \beta_1 + \beta_4$

For L : $1 = -3\beta_1 + \beta_2 + 2\beta_3 - \beta_4 + \beta_5$

For T : $-2 = -\beta_2 - \beta_4$

and so on. Solving these equations in terms of two of the unknowns, say, α_4, α_5 and β_4, β_5, we obtain:

$$\begin{aligned} \alpha_1 &= 1 - \alpha_4 & \beta_1 &= 1 - \beta_4 \\ \alpha_2 &= 2 - \alpha_4 & \text{and} \quad \beta_2 &= 2 - \beta_4 \\ \alpha_3 &= 1 - \frac{\alpha_4}{2} - \frac{\alpha_5}{2} & \beta_3 &= 1 - \frac{\beta_4}{2} - \frac{\beta_5}{2} \end{aligned}$$

Substituting these values into Eq 114 yields:

$$F = \rho V^2 S \left[c_1 \left(\frac{\mu}{\rho D V}\right)^{\alpha_4} \left(\frac{D^2}{S}\right)^{(\alpha_4 + \alpha_5)/2} + c_2 \left(\frac{\mu}{\rho D V}\right)^{\beta_4} \left(\frac{D^2}{S}\right)^{(\beta_4 + \beta_5)/2} + \ldots \right] \qquad \text{(Eq 115a)}$$

The right-hand series is a function of $\mu/\rho D V$ and D^2/S, and F can be written symbolically as:

$$F = \rho V^2 S f\left(\frac{\mu}{\rho D V}, \frac{D^2}{S}\right) \qquad \text{(Eq 115b)}$$

Only one term need be considered in Eq 115 (a) to obtain the functional relationship between the physical quantities. In addition, the functional relationship between F, ρ, V, S, μ, and D is equivalent to a relationship between the three nondimensional products, expressed as:

$$\frac{F}{\frac{1}{2}\rho V^2 S} = f\left(\frac{\rho V D}{\mu}, \frac{D^2}{S}\right) \qquad \text{(Eq 116)}$$

where the product $F/\rho V^2 S$ is multiplied by 2. Therefore, in an experimental investigation of the relationship between the force F and the other physical parameters, only the relation between the following nondimensional parameters needs to be determined:

$\dfrac{F}{\frac{1}{2}\rho V^2 S}$ (called the force coefficient)

$\dfrac{\rho V D}{\mu}$ (called the Reynolds number)

$\dfrac{D^2}{S}$ (a ratio of cross-sectional area and surface area)

Table 2 gives important dimensionless groups used in fluid mechanics.

Nondimensionalization of the Basic Equations

A very powerful technique for finding dimensionless parameters involves the basic

Table 2 Dimensionless groups in fluid mechanics

Parameter	Definition	Qualitative ratio of effects	Importance
Reynolds number	$\mathrm{Re} = \dfrac{\rho VL}{\mu}$	$\dfrac{\text{Inertia}}{\text{Viscosity}}$	General flow
Mach number	$\mathrm{M} = \dfrac{V}{a}$	$\dfrac{\text{Flow speed}}{\text{Sound speed}}$	Compressible flow
Froude number	$\mathrm{Fr} = \dfrac{V^2}{gL}$	$\dfrac{\text{Inertia}}{\text{Gravity}}$	Free-surface flow
Weber number	$\mathrm{We} = \dfrac{\rho V^2 L}{\Gamma}$	$\dfrac{\text{Inertia}}{\text{Surface tension}}$	Free-surface flow
Cavitation number (Euler number)	$\mathrm{Ca} = \dfrac{p - p_v}{\rho V^2}$	$\dfrac{\text{Pressure}}{\text{Inertia}}$	Cavitation
Prandtl number	$\mathrm{Pr} = \dfrac{\mu c_p}{k}$	$\dfrac{\text{Dissipation}}{\text{Conduction}}$	Heat convection
Eckert number	$\mathrm{Ec} = \dfrac{V^2}{c_p T_0}$	$\dfrac{\text{Kinetic energy}}{\text{Enthalpy}}$	Dissipation
Specific heat ratio	$\kappa = \dfrac{c_p}{c_v}$	$\dfrac{\text{Enthalpy}}{\text{Internal energy}}$	Compressible flow
Strouhal number	$\mathrm{St} = \dfrac{\omega L}{V}$	$\dfrac{\text{Oscillation}}{\text{Mean speed}}$	Oscillating flow
Roughness ratio	$\dfrac{\varepsilon}{L}$	$\dfrac{\text{Wall roughness}}{\text{Body length}}$	Turbulent, rough walls
Grashof number	$\mathrm{Gr} = \dfrac{\beta \Delta T g L^3 \rho^2}{\mu^2}$	$\dfrac{\text{Buoyancy}}{\text{Viscosity}}$ (β = coeff. of volumetric expansion)	Natural convection
Temperature ratio	$\dfrac{T_w}{T_o}$	$\dfrac{\text{Wall temperature}}{\text{Stream temperature}}$	Heat transfer

Source: Ref 1 with permission

Fig. 17 Boundary layer on flat plate

equations of flow. A brief example concerns the incompressible flow equations with constant viscosity. Rewriting the continuity and momentum equations previously presented:

$$\nabla \cdot \mathbf{V} = 0 \tag{Eq 69}$$

$$\rho \frac{d\mathbf{V}}{dt} = \rho \mathbf{g} - \nabla p + \mu \nabla^2 \mathbf{V} \tag{Eq 85}$$

Equations 69 and 85 contain the three basic dimensions *MLT*. Variables p, $\mathbf{V}$, x, y, z, and t can be made nondimensional using the density and two reference constants that may be characteristic of the particular fluid flow:

Reference velocity $= U$ Reference length $= L$

The inlet or freestream velocity could be used for U, and L could be the diameter of a body immersed in the stream. Relevant dimensionless variables can be defined as:

$$\mathbf{V}' = \frac{\mathbf{V}}{U} \quad x' = \frac{x}{L} \quad y' = \frac{y}{L} \quad z' = \frac{z}{L}$$
$$t' = \frac{tU}{L} \quad p' = \frac{p + \rho gz}{\rho U^2}$$

All these are fairly obvious except for p', where the gravity effect is included.

Because ρ, U, and L are all constants, the derivatives in Eq 69 and 85 can be handled in dimensionless form with dimensional coefficients. The resulting dimensionless equations of motion are as follows:

Continuity:

$$\nabla' \cdot \mathbf{V}' = 0 \tag{Eq 117}$$

Momentum:

$$\frac{d\mathbf{V}'}{dt'} = -\nabla' p' + \frac{\mu}{\rho UL} \nabla'^2 (\mathbf{V}') \tag{Eq 118}$$

The Reynolds number is introduced by the process:

$$\mathrm{Re} = \frac{\rho UL}{\mu}$$

Boundary-Layer Flow

High Reynolds-number flow, which describes most flow situations for common low-viscosity fluids such as air and water, implies that inertial forces in the fluid are predominant over viscous forces. The effect of viscosity is small over most of the flow field. The layer of fluid at the wall, however, has zero velocity relative to the surface, and the resulting large velocity gradients occurring near a surface, indicate that viscous effects cannot be neglected in this region. Away from the wall, the velocity gradients are small, and the viscous forces are negligible in comparison with the inertial forces.

The effects of viscosity can therefore be confined to a thin layer in the vicinity of the surface, called the boundary layer. Outside the boundary layer, the flow can be treated as nonviscous (Fig. 17). Although the boundary layer is thin, it gives rise to viscous drag caused by the shear stresses and dominates the transfer of heat from the surface.

Laminar and Turbulent Boundary Layers

Near the leading edge of the flat plate shown in Fig. 18, the boundary layer is laminar. Farther along the plate, transition occurs to a turbulent boundary layer where mixing results in a more uniform velocity profile. The previously mentioned viscous or laminar sublayer also develops somewhere downstream of the transition region. Shear stress in the laminar region of the boundary layer is given by the familiar:

$$\tau = \mu \frac{du}{dy} \tag{Eq 119}$$

In the turbulent portion of the boundary layer, a turbulent kinematic viscosity, or eddy viscosity, ε, is defined, so the turbulent shear stress becomes:

$$\tau_t = \rho\varepsilon\frac{du}{dy} \qquad \text{(Eq 120)}$$

Equations of Motion—Laminar Boundary Layer

The boundary-layer flow is assumed to be steady, two-dimensional, incompressible, and with no variation of μ. Gravitational forces are neglected. The radius of curvature of the surface is assumed large, and the pressure does not vary in the y-direction (Fig. 19). The boundary-layer thickness, δ, is normally defined as the distance from the surface where the velocity, u, is 99% of the free stream velocity, u_∞. The continuity equation for this system is:

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} = 0 \qquad \text{(Eq 121)}$$

and the momentum equation reduces to the form:

$$-\frac{\partial p}{\partial x} + \mu\frac{\partial^2 u}{\partial y^2} = \rho\left(u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y}\right) \qquad \text{(Eq 122)}$$

where the boundary conditions are:

$$u = v = 0 \text{ at } y = 0$$
$$u = u_\infty \text{ at } y = \delta$$

The free stream velocity, u_∞, is given by the potential flow solution about the body. The pressure gradient is also available from the potential flow solution by applying Bernoulli's equation. For a flat plate, the free stream velocity is constant along the plate, and the Bernoulli equation shows there is no pressure variation outside the boundary layer. Thus, for a flat plate boundary layer, the momentum equation is:

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = \nu\frac{\partial^2 u}{\partial y^2} \qquad \text{(Eq 123)}$$

Laminar Flow Solutions

Blasius Solution (Flat Plate Aligned with Flow). Blasius (1908) arrived at a solution to the flat plate equations by assuming that the shape of a nondimensional velocity distribution did not vary from section to section along the plate (similarity assumption). The velocity ratio (u/u_∞) was considered to be a function of y/δ, which did not change with x. The solution, obtained using infinite series or numerical means, is shown in Fig. 20. The dimensionless velocity is plotted against a dimensionless distance from the wall.

From the Blasius curve in Fig. 20, the boundary-layer thickness (where $u = 0.99u_\infty$) is computed to be:

$$\delta = \frac{5.0x}{\sqrt{\mathrm{Re}_x}} \qquad \text{(Eq 124)}$$

where:

$$\mathrm{Re}_x = \frac{\rho u_\infty x}{\mu}$$

and x is the distance from the leading edge of the plate. Other methods have been used to define a boundary-layer thickness, one being the displacement thickness, δ^*, which represents the displacement of the external flow due to the boundary layer. The displacement thickness is defined as:

$$\delta^* = \int_0^\infty \left(1 - \frac{u}{u_\infty}\right)dy \qquad \text{(Eq 125)}$$

Using the Blasius solution, this becomes:

$$\delta^* = \frac{1.73x}{\sqrt{\mathrm{Re}_x}} \qquad \text{(Eq 126)}$$
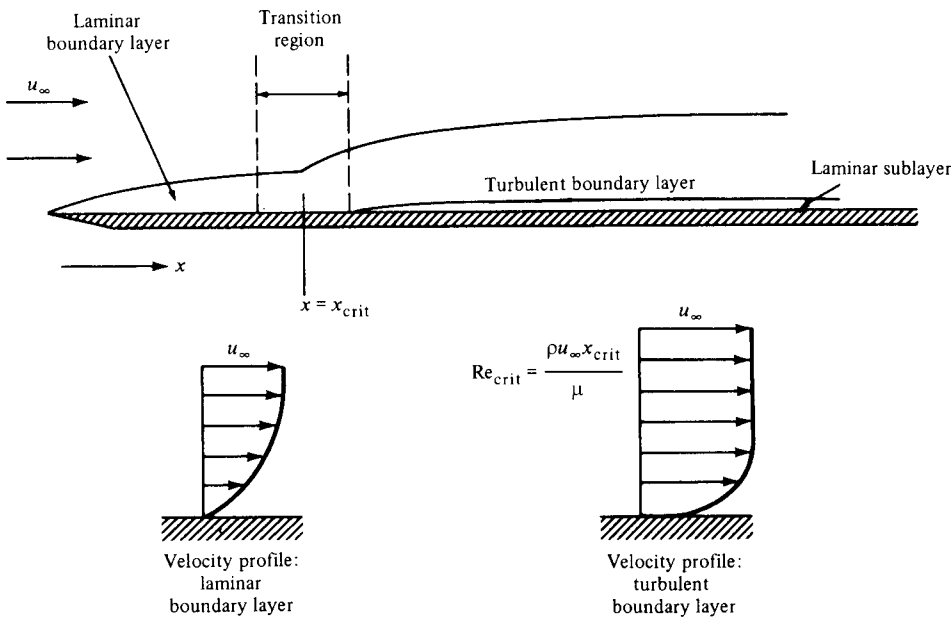
Fig. 18 Development of laminar and turbulent boundary layers on a flat plate. Source: Ref 2 with permission
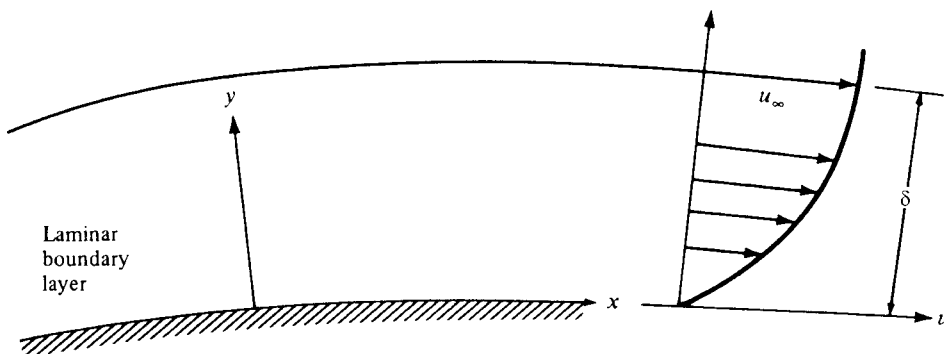
Fig. 19 Laminar boundary layer
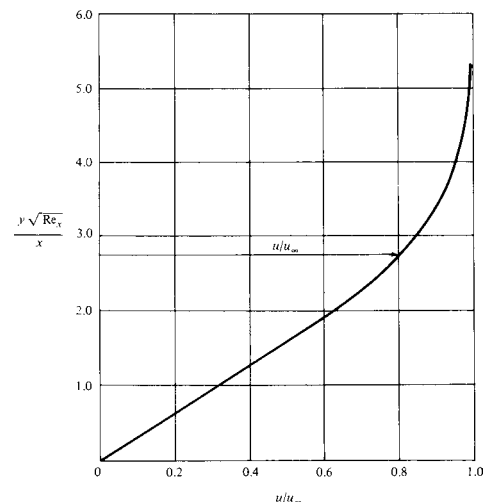
Fig. 20 Velocity distribution in laminar boundary layer—Blasius profile

The wall shear stress on the flat plate is obtained by:

$$\tau_w = 0.332\mu \frac{u_\infty}{x}\sqrt{\mathrm{Re}_x} \qquad \text{(Eq 127)}$$

The local skin friction coefficient, C_{fx}, then becomes:

$$C_{fx} = \frac{\tau_w}{\frac{1}{2}\rho u_\infty^2} = \frac{0.664}{\sqrt{\mathrm{Re}_x}} \qquad \text{(Eq 128)}$$

and total drag on the plate is found to be:

$$D = \frac{0.664 b \rho u_\infty^2 L}{\sqrt{\mathrm{Re}_L}} \qquad \text{(Eq 129)}$$

where:

$$\mathrm{Re}_L = \frac{\rho u_\infty L}{\mu}$$

and L is the total length of the plate. The dimensionless skin friction drag coefficient, C_{D_f}, is:

$$C_{D_f} = \frac{1.328}{\sqrt{\mathrm{Re}_L}} \qquad \text{(Eq 130)}$$

Relative Movement of Parallel Plates. For the steady-state, incompressible flow between two plates that occurs when one plate moves with a velocity u_0, the resulting velocity profile (Fig. 21) is:

$$u = \frac{u_0}{h}y - \frac{h^2}{2\mu}\frac{dp}{dy}\frac{y}{h}(1 - y/h) \qquad \text{(Eq 131)}$$

For no pressure gradient, this becomes:

$$u = \frac{u_0}{h}y \qquad \text{(Eq 132)}$$

which is defined as Couette flow.

Turbulent Boundary Layer

Momentum Integral Method. The Karman momentum integral method requires writing the continuity and momentum equations for a control volume extending from the wall surface to the outer edge of the boundary layer (Fig. 22). Analyzing this control volume gives the momentum integral equation for two-dimensional, incompressible steady flow:
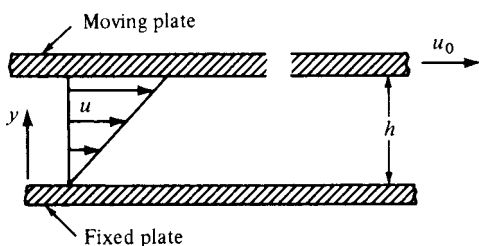
$$\rho\frac{d}{dx}\left(\int_0^\delta u^2 dy\right) - \rho u_\infty\frac{d}{dx}\left(\int_0^\delta u dy\right) = -\delta\frac{dp}{dx} - \tau_w \qquad \text{(Eq 133)}$$

This is valid for both laminar and turbulent boundary-layer flows but does not yield any information about the details of the flow at a particular point. A functional relationship of u and y must be assumed so wall shear stress can be computed. A rough assumption can provide relatively good results. The pressure gradient is dealt with knowing that:

$$\frac{dp}{dx} = -\frac{u_\infty}{\rho}\frac{du_\infty}{dx}$$

a relationship obtainable from the Bernoulli equation.

Assuming a linear profile on a flat plate, as shown in Fig. 23, and using the integral equation gives the following results. For the assumed profile, shear stress at the wall is:

$$\tau_w = \mu\frac{u_\infty}{\delta} \qquad \text{(Eq 134)}$$

The boundary-layer thickness becomes:

$$\delta = \frac{3.46x}{\sqrt{\mathrm{Re}_x}} \qquad \text{(Eq 135)}$$

and the displacement thickness is:

$$\delta^* = \frac{1.73x}{\sqrt{\mathrm{Re}_x}} \qquad \text{(Eq 136)}$$

The skin friction coefficient is given by:

$$C_{fx} = \frac{0.578}{\sqrt{\mathrm{Re}_x}} \qquad \text{(Eq 137)}$$

In Eq 137, the result can be seen to be reasonably close to that for the Blasius solution.

A profile based on the power-law equation:

$$\frac{u}{u_\infty} = \left(\frac{y}{\delta}\right)^{1/7} \qquad \text{(Eq 138)}$$

gives a very good correlation with experimental data over a wide range of turbulent Reynolds numbers on a flat plate. This profile is not valid in the immediate vicinity of the wall, and experimental data must be used to obtain an expression for τ_w. For the Reynolds number range $5\times10^5 < \mathrm{Re}_x < 10^7$, the following formula correlates well with data:

$$\frac{\tau_w}{\rho u_x^2} = 0.0225\left(\frac{v}{u_\infty\delta}\right)^{1/4} \qquad \text{(Eq 139)}$$

The boundary-layer thickness becomes:

$$\delta = 0.37x\left(\frac{u_\infty x}{v}\right)^{-1/5} \qquad \text{(Eq 140)}$$

with the displacement thickness being:

$$\delta^* = 0.046x(\mathrm{Re}_x)^{-1/5} \qquad \text{(Eq 141)}$$

The local skin friction coefficient is computed by:

$$C_{fx} = \frac{0.058}{\mathrm{Re}_x^{1/5}} \qquad \text{(Eq 142)}$$

The total drag caused by the turbulent boundary layer over a flat plate of length L and width b is:

$$D = 0.036\rho u_\infty^2\frac{bL}{\mathrm{Re}_L^{1/5}} \qquad \text{(Eq 143)}$$

and the dimensionless skin friction drag coefficient becomes:

$$C_{D_f} = \frac{0.072}{\mathrm{Re}^{1/5}} \qquad \text{(Eq 144)}$$

Experiments have shown that a constant of 0.074 fits the data better. This expression is valid over the Reynolds number range $5\times10^5 < \mathrm{Re}_L < 10^7$.

For a flat plate with a laminar boundary layer starting at the leading edge and a turbulent layer that begins some distance beyond the leading edge, the drag can be computed as follows:

$$\begin{aligned}D = & \text{ Turbulent drag (from } x = 0 \text{ to } x = L)\\ & - \text{ Turbulent drag(from } x = 0 \text{ to } x = x_{\text{crit}})\\ & + \text{ Laminar drag(from } x = 0 \text{ to } x = x_{\text{crit}})\end{aligned}$$

where x_{crit} is the location of the transition from laminar to turbulent. The total drag coefficient, C_D, is:

Fig. 21 Flow between fixed plate and moving plate

Fig. 22 Control volume for integral analysis
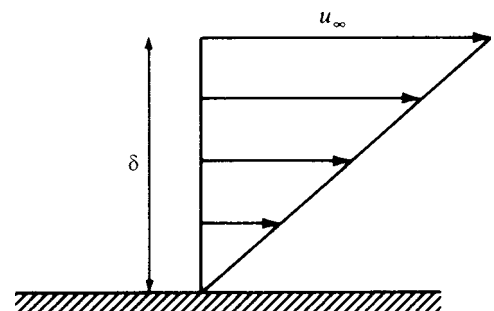
Fig. 23 Linear profile on flat plate

$$C_D = C_{D_t} - \frac{1700}{\text{Re}_L} \qquad \text{(Eq 145)}$$

where C_{D_t} is the turbulent drag coefficient. This equation is valid up to $\text{Re}_L = 1 \times 10^7$. An expression for C_{D_f} that seems to fit experimental data over a wider range of Reynolds numbers is:

$$C_{D_f} = \frac{0.455}{(\log \text{Re}_L)^{2.58}} - \frac{1700}{\text{Re}_L} \qquad \text{(Eq 146)}$$

This equation is valid for turbulent Reynolds numbers up to 10^9.

The analysis of flat-plate boundary layers seems rather specialized, but many thin, nonflat shapes can be considered in the same way because the basic differential equation (Eq 122) and integral equation (Eq 133) do not change. The variation of free stream velocity along the body must be known to arrive at a solution, because this allows determination of the pressure gradient. The pressence of an adverse pressure gradient that accompanies decelerating flow can lead to separation of the boundary layer. This compounds the problems in analyzing the flow. More detail on boundary-layer solutions is given in Ref 3 and 4.

REFERENCES

1. F.M. White, *Fluid Mechanics,* McGraw-Hill, New York, 1979
2. J.E.A. John, and W.L. Haberman, *Introduction to Fluid Mechanics,* 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1980
3. H. Schlichting, *Boundary Layer Theory,* 7th ed., McGraw-Hill, New York, 1979
4. V.L. Streeter, Ed., *Handbook of Fluid Dynamics,* 1st ed., McGraw-Hill, New York, 1961

SELECTED REFERENCES

- R.W. Fox, and A.T. McDonald, *Introduction to Fluid Mechanics,* John Wiley &Sons, New York, 1973
- J.A. Roberson, and C.T. Crowe, *Engineering Fluid Mechanics,* 2nd ed., Houghton Mifflin, Boston, 1980
- I.R. Shames, *Mechanics of Fluids,* 2nd ed., McGraw-Hill, New York, 1982
- V.L. Streeter, and E.B. Wylie, *Fluid Mechanics,* 6th ed., McGraw-Hill, New York, 1975
- F.M. White, *Viscous Fluid Flow,* McGraw-Hill, New York, 1974

Differential Calculus and Equations*

MODELS OF PHYSICAL PHENOMENA involve solution of differential equations, which include either ordinary or partial differential equations, as described in this article. The basic concepts of differential calculus also are reviewed in this article, with a table of power series for common analytical functions.

Basic Concepts of Differential Calculus

When variables x and y are related in such a way that for each value of x in a given domain there is a corresponding value of y, then y is a function of x. That is, x is the independent variable and y the dependent variable. Such symbols as $F(x)$, $f(x)$, and $\phi(x)$ denote functions of x, and the symbol $f(a)$ denotes the value of $f(x)$ for $x = a$.

Differentiation formulas for various algebraic and transcendental functions are listed in Table 1.

Limits. The function $f(x)$ approaches the limit 1 as x approaches a if, within a sufficiently small interval with a as midpoint, the difference $|f(x) - 1|$ can be made arbitrarily small for all values of x except a. Symbolically, this limitation is expressed as:

$$\lim_{x \to a} f(x) = 1$$

The symbols:

$$\lim_{x \to a} f(x) = \infty \text{ or } \lim_{x \to a} f(x) = -\infty$$

mean that $f(x)$ can be made arbitrarily large positively or negatively, respectively, for all values of x except a within a small interval with a as midpoint.

The symbols:

$$\lim_{x \to \infty} f(x) = 1 \text{ or } \lim_{x \to -\infty} f(x) = 1$$

mean that the difference $|f(x) - 1|$ can be made arbitrarily small for all sufficiently large values of x, positively or negatively, respectively.

Changes in x are called increments of x and are denoted by Δx, and the corresponding changes in y are denoted by Δy. If the relationship:

$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

exists, it is called the derivative of y with respect to x and is denoted by:

$$dy/dx, f'(x), \text{ or } D_x y$$

Table 1 Differentiation Formulas

Algebraic functions (a)(b)

$$\frac{d}{dx} a = 0$$

$$\frac{d}{dx} au = a \frac{du}{dx}$$

$$\frac{d}{dx}(u + v + w + \cdots) = \frac{du}{dx} + \frac{dv}{dx} + \frac{dw}{dx} + \cdots$$

$$\frac{d}{dx} uv = u \frac{dv}{dx} + v \frac{du}{dx}$$

$$\frac{d}{dx}(uvw \cdots) = \left(\frac{1}{u}\frac{du}{dx} + \frac{1}{v}\frac{dv}{dx} + \frac{1}{w}\frac{dw}{dx} + \cdots\right)(uvw \cdots)$$

$$\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$$

$$\frac{d}{dx} u^n = nu^{n-1} \frac{du}{dx}$$

Transcendental functions (a)(b)(c)

$$\frac{d}{dx} \log_e u = \frac{1}{u}\frac{du}{dx}$$

$$\frac{d}{dx} \log_{10} u = \frac{1}{u}\frac{du}{dx}\log_{10} e = (0.4343)\frac{1}{u}\frac{du}{dx}$$

$$\frac{d}{dx} e^u = e^u \frac{du}{dx}$$

$$\frac{d}{dx} u^v = vu^{v-1}\frac{du}{dx} + u^v\frac{dv}{dx}\log_e u$$

$$\frac{d}{dx} f(u) = \frac{df(u)}{du} \cdot \frac{du}{dx}$$

Transcendental functions (a)(b)(c)

$$\frac{d^2 f(u)}{dx^2} = \frac{df(u)}{du} \cdot \frac{d^2 u}{dx^2} + \frac{d^2 f(u)}{du^2}\left(\frac{du}{dx}\right)^2$$

$$\frac{d}{dx} \sin u = \cos u \frac{du}{dx}$$

$$\frac{d}{dx} \cos u = -\sin u \frac{du}{dx}$$

$$\frac{d}{dx} \tan u = \sec^2 u \frac{du}{dx}$$

$$\frac{d}{dx} \cot u = -\csc^2 u \frac{du}{dx}$$

$$\frac{d}{dx} \sec u = \sec u \tan u \frac{du}{dx}$$

$$\frac{d}{dx} \csc u = -\csc u \cot u \frac{du}{dx}$$

$$\frac{d}{dx} \sin^{-1} u = \frac{1}{\sqrt{1-u^2}}\frac{du}{dx}\left(-\frac{\pi}{2} \leq \sin^{-1} u \leq \frac{\pi}{2}\right)$$

$$\frac{d}{dx} \cos^{-1} u = \frac{1}{\sqrt{1-u^2}}\frac{du}{dx}(0 \leq \cos^{-1} u \leq \pi)$$

$$\frac{d}{dx} \tan^{-1} u = \frac{1}{1+u^2}\frac{du}{dx}$$

$$\frac{d}{dx} \cot^{-1} u = -\frac{1}{1+u^2}\frac{du}{dx}$$

$$\frac{d}{dx} \sec^{-1} u = \frac{1}{u\sqrt{u^2-1}}\frac{du}{dx}(d)$$

$$\frac{d}{dx} \csc^{-1} u = -\frac{1}{u\sqrt{u^2-1}}\frac{du}{dx}(d)$$

Transcendental functions (a)(b)(c)

$$\frac{d}{dx} \sinh u = \cosh u \frac{du}{dx}$$

$$\frac{d}{dx} \cosh u = \sinh u \frac{du}{dx}$$

$$\frac{d}{dx} \tanh u = \text{sech}^2 u \frac{du}{dx}$$

$$\frac{d}{dx} \coth u = -\text{csch}^2 u \frac{du}{dx}$$

$$\frac{d}{dx} \text{sech } u = -\text{sech } u \tanh u \frac{du}{dx}$$

$$\frac{d}{dx} \text{csch } u = -\text{csch } u \coth u \frac{du}{dx}$$

$$\frac{d}{dx} \sinh^{-1} u = \frac{1}{\sqrt{u^2+1}}\frac{du}{dx}$$

$$\frac{d}{dx} \cosh^{-1} u = \frac{1}{\sqrt{u^2-1}}\frac{du}{dx}$$

$$\frac{d}{dx} \tanh^{-1} u = \frac{1}{1-u^2}\frac{du}{dx}$$

$$\frac{d}{dx} \coth^{-1} u = \frac{1}{1-u^2}\frac{du}{dx}$$

$$\frac{d}{dx} \text{sech}^{-1} u = -\frac{1}{u\sqrt{1-u^2}}\frac{du}{dx}$$

$$\frac{d}{dx} \text{csch}^{-1} u = -\frac{1}{u\sqrt{u^2+1}}\frac{du}{dx}$$

(a) Functions of x are represented by u, v, and w. (b) a and n are constants. (c) e is the base of the natural or Napierian logarithms; $e = 2.7183$ (d) For angles in the first and third quadrants. Use the opposite sign in the second and fourth quadrants.

*Adapted from *ASM Handbook of Engineering Mathematics,* American Society for Metals, 1983

Geometric Interpretation of the Derivative. If $y = f(x)$ is represented by a graph in Cartesian co-ordinates, as shown in Fig. 1, then $f'(x) = \tan \alpha$, where α lies between the x-axis and the line tangent to the curve in the given point, with the angle measured counterclockwise.

The derivative $f'(x)$ exists for these values of x for which the function $f(x)$ is defined and continuous, and the considered ratio has a finite limit. If, for a value x_1, the derivative does not exist, no definite tangent line exists at the corresponding point of the graph of the function, or the tangent is perpendicular to the x-axis. In the latter case, the limit is infinite and is expressed as $f'(x_1) = \infty$; that is, the derivative becomes infinite. For example, the derivative does not exist at $x = 0$ for the function:

$$f(x) = \sqrt[3]{x}$$

because the derivative becomes infinite at the point 0 (Fig. 2).

Derivative of a Function of a Function. If y is a function of a variable u, and if u is a function of another variable x, then:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

Derivative of an Inverse Function. If $y = f_1(x)$ is solved for x, then the solution of $x = f_2(y)$ is termed its inverse. For example, the expression $x = \sin^{-1} y$ is the inverse of $y = \sin x$ and $x = \pm\sqrt{y}$ is the inverse of $y = x^2$. Symbolically:

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} \text{ or } f_2'(y) = \frac{1}{f_1'(x)}$$

Derivatives of Second and Higher Order. The derivative of the first derivative of y with respect to x is termed the second derivative of y with respect to x. If $y = f(x)$, the second derivative with respect to x is denoted variously by:

$$\frac{d^2y}{dx^2} \quad D_{x^2}y \quad \frac{dy'}{dx} \quad f''(x) \quad y'' \quad \ddot{y}$$

Similarly, the derivative with respect to x of the second derivative is called the third derivative and may be denoted by:

$$\frac{d^3y}{dx^3} \quad D_{x^3}y \quad f^{(3)}(x) \quad y^{(3)}$$
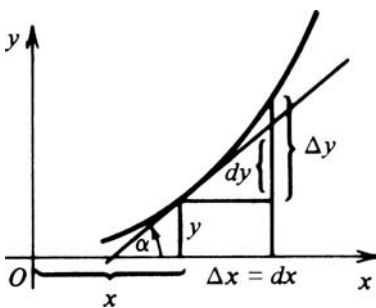
For example, to find the mth derivative of x^n, note the following sequence of derivatives:

$$y' = nx^{n-1}$$
$$y'' = n(n-1)x^{n-2}$$
$$y^{(3)} = n(n-1)(n-2)x^{n-3}$$

It can be seen that:

$$y^{(m)} = n(n-1)(n-2)\cdots(n-m+1)x^{n-m}$$

Differentials. If the curve AB in Fig. 3 is defined by $y = f(x)$, and $P(x, y)$ and $Q(x + \Delta x, y + \Delta y)$ are nearby points, then as Q is moved toward P, Δx and Δy approach zero. The limiting position of the line PQ is the tangent at P. The slope of the tangent is dy/dx, which is the limit of the ratio $\Delta y/\Delta x$ as $\Delta x \to 0$. That is, dy represents the increment of the ordinate of the tangent at P for a given increment dx at P, whereas Δy is the increment of the ordinate to the curve AB.

The increments dx and dy are termed differentials. For a function $y = f(x)$, the differential equals its derivative multiplied by the differential of the independent variable, $dy = f'(x)dx$. The length of the line PQ in Fig. 3 is given by $(\Delta s)^2 = (\Delta x)^2 + (\Delta y)^2$. Similarly, the differential length of curve is given as:

$$ds^2 = dx^2 + dy^2$$
$$ds = \sqrt{1 + y'^2} \quad dx = \sqrt{1 + x'^2}dy$$

In polar coordinates:

$$ds = \sqrt{d\rho^2 + \rho^2 d\theta} = \sqrt{\rho^2 + \left(\frac{d\rho}{d\theta}\right)^2}d\theta$$

Implicit Functions. The expression $F(x,y) = 0$ defines y as an implicit function of x, and x as an implicit function of y. When an equation is solved for y in terms of x, $y = f(x)$, then y becomes an explicit function of x. For example, as an implicit function, $F(x, y) = x^2 + y^2 - r^2 = 0$, and as an explicit function:

$$y = \pm\sqrt{r^2 - x^2}$$

To find dy/dx, differentiate $y = f(x)$ or use as an alternative method:

$$\frac{dy}{dx} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}\left(\frac{\partial F}{\partial y} \neq 0\right)$$

and:

$$\frac{d^2y}{dx^2} = -\frac{\frac{\partial^2 F}{\partial x^2}\left(\frac{\partial F}{\partial y}\right)^2 - 2\frac{\partial^2 F}{\partial x \partial y}\frac{\partial F}{\partial x}\frac{\partial F}{\partial y} + \frac{\partial^2 F}{\partial y^2}\left(\frac{\partial F}{\partial x}\right)^2}{\left(\frac{\partial F}{\partial y}\right)^3}$$

$$\left(\frac{\partial F}{y} \neq 0\right)$$

Differentiation of Parametric Functions. If a function is given in parametric form, to find the derivatives of y with respect to x if $y = y(t)$ and $x = x(t)$ use:

$$y' = \frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} \qquad y'' = \frac{d^2y}{dx^2} = \frac{\frac{dy'}{dt}}{\frac{dx}{dt}}$$
$$y^{(n)} = \frac{d^n y}{dx^n} = \frac{\frac{dy^{(n-1)}}{dt}}{\frac{dx}{dt}}$$

To find the derivatives of y with respect to x for the ellipse $x = a\cos t$, $y = b\sin t$, use the following relationships:

$$y' = \frac{dy}{dx} = \frac{b\cos t}{-a\sin t} = -\frac{b}{a}\cot t$$
$$y'' = \frac{dy'}{dx} = \frac{\frac{b}{a}\csc^2 t}{-a\sin t} = -\frac{b}{a^2}\csc^3 t$$
$$y''' = \frac{dy''}{dx} = \frac{\frac{3b}{a^2}\csc^3 t \cot t}{-a\sin t} = -\frac{3b}{a^3}\csc^4 t \cot t$$

Logarithmic Differentiation for Products and Quotients. If:

$$y = \frac{u^l v^m}{w^n}$$

express both sides of the equation in terms of logarithms before differentiating:

$$\ln y = l\ln u + m\ln v - n\ln w$$
$$\frac{1}{y}\frac{dy}{dx} = \frac{l}{u}\frac{du}{dx} + \frac{m}{v}\frac{dv}{dx} - \frac{n}{w}\frac{dw}{dx}$$
$$\frac{dy}{dx} = y\left(\frac{l}{u}\frac{du}{dx} + \frac{m}{v}\frac{dv}{dx} - \frac{n}{w}\frac{dw}{dx}\right)$$

Fig. 1 Geometric interpretation of a derivative (dy/dx) as the tangent of the angle α

Fig. 2 Function without a derivative at $x = 0$
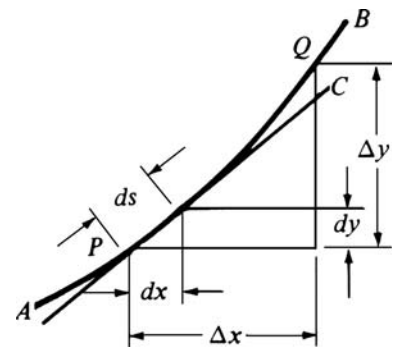
Fig. 3 Differential length (ds) of a curve

For example, to find $\dfrac{dy}{dx}$ if:

$$y = \frac{\sqrt{x^2 - 25}}{(x-1)^3(x+5)^2}$$

use the following relationships:

$$\ln y = \tfrac{1}{2}\ln(x^2 - 25) - 3\ln(x-1) - 2\ln(x+5)$$
$$\frac{1}{y}\frac{dy}{dx} = \frac{2x}{2(x^2 - 25)} - \frac{3}{x-1} - \frac{2}{x+5}$$
$$\frac{dy}{dx} = \frac{y(-4x^2 + 11x + 65)}{(x^2 - 25)(x-1)}$$

Partial Derivatives

If three variables $f(x, y)$, x, y are related in such a way that for each set of values of x and y in a given domain there exists a corresponding value of $f(x, y)$, then $f(x, y)$ is a function of x and y. If x is the only variable and y is a constant, then the derivative of $f(x, y)$ with respect to x becomes the partial derivative of f with respect to x and is given as:

$$\frac{\partial f}{\partial x} = f_x = \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

Similarly, the partial derivative of f with respect to y is obtained when y varies and x is constant:

$$\frac{\partial f}{\partial y} = f_y = \lim_{\Delta y \to 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}$$

Partial Derivatives of Second or Higher Order. If $\partial f/\partial x$ and $\partial f/\partial y$ are again differentiable, the partial derivatives of the second order may be found:

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = f_{xx} \qquad \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2} = f_{yy}$$
$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y} = f_{yx} \qquad \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} = f_{xy}$$

When the derivatives are continuous, the order of differentiation is immaterial, that is:

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$$

The third and higher partial derivatives of $f(x, y)$ may be determined in a similar manner. If continuous, the third partial derivatives are the following four:

$$\frac{\partial}{\partial x}\left(\frac{\partial^2 f}{\partial x^2}\right) = \frac{\partial^3 f}{\partial x^3} \qquad \frac{\partial}{\partial x}\left(\frac{\partial^2 f}{\partial y^2}\right) = \frac{\partial}{\partial y}\left(\frac{\partial^2 f}{\partial x \partial y}\right)$$
$$= \frac{\partial^2}{\partial y^2}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^3 f}{\partial x \partial y^2}$$
$$\frac{\partial}{\partial y}\left(\frac{\partial^2 f}{\partial y^2}\right) = \frac{\partial^3 f}{\partial y^3} \qquad \frac{\partial}{\partial y}\left(\frac{\partial^2 f}{\partial x^2}\right) = \frac{\partial}{\partial x}\left(\frac{\partial^2 f}{\partial x \partial y}\right)$$
$$= \frac{\partial^2}{\partial x^2}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^3 f}{\partial x^2 \partial y}$$

Functions of _N_ Variables. The partial derivative formulas may be generalized to the form

that has f as a function of more than two variables; that is, there corresponds a value of $f(x, y, z, \cdots)$ to every set of values of $x, y, z, \cdots$.

If the increments $\Delta x, \Delta y, \Delta z, \cdots$ are assigned to the variables $x, y, z, \cdots$ in the function $f(x, y, z, \cdots)$, the total increment of f is:

$$\Delta f = f(x + \Delta x, y + \Delta y, z + \Delta z, \cdots) - f(x, y, z, \cdots)$$

The total differential of f becomes:

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz + \cdots$$

The second total differential of f becomes:

$$d^2 f = \frac{\partial^2 f}{\partial x^2}(dx)^2 + \frac{\partial^2 f}{\partial y^2}(dy)^2 + \frac{\partial^2 f}{\partial z^2}(dz)^2 + \cdots$$
$$+ 2\frac{\partial^2 f}{\partial x \partial y}dx dy + \cdots$$

The general form is:

$$d^n f = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy + \frac{\partial}{\partial z}dz + \cdots\right)^n f(x, y, z, \cdots)$$

Exact Differential. For the expression $P(x, y)\, dx + Q(x, y)\, dy$ to be the exact or complete differential of a function of two variables, the following condition is necessary:

$$\frac{\partial Q}{\partial x} = \frac{\partial P}{\partial y}$$

For three variables, $P\, dx + Q\, dy + R\, dz$, the required conditions are:

$$\frac{\partial Q}{\partial z} = \frac{\partial R}{\partial y}$$
$$\frac{\partial R}{\partial x} = \frac{\partial P}{\partial z}$$
$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$$

Composite Functions. If $u = f(x, y, z, \cdots w)$, and $x, y, z, \cdots w$ are functions of a single variable t, then:

$$\frac{du}{dt} = \frac{\partial u}{\partial x}\frac{dx}{dt} + \frac{\partial u}{\partial y}\frac{dy}{dt} + \cdots + \frac{\partial u}{\partial w}\frac{dw}{dt}$$

This expression represents the total derivative of u with respect to t.

Infinite Series

If a_1, a_2, ..., a_n, ... is a number sequence formed according to some rule, the indicated sum is an infinite series:

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + \cdots + a_n + \cdots$$

If the partial sums, s_n, in the expression $s_n = a_1 + a_2 + \cdots + a_n$ approach a limit S as $n \to \infty$, the series is convergent; S is the sum

of the series, or its value. A series that does not converge is divergent.

If the series of absolute values $|a_1| + |a_2| + \cdots + |a_n| + \cdots$ is convergent, then the aforementioned infinite series is absolutely convergent. If the series converges, but not absolutely, it is conditionally convergent. An absolutely convergent series does not charge is sum by rearrangement of its terms. The convergence of a series can be verified by testing.

Comparison Test. For a series of positive terms, such as $c_1 + c_2 + \cdots + c_n + \cdots$, for which $a_n \overset{\le}{=} c_n$ for every n from some term onward, the infinite series converges. If there is a divergent series of positive terms $d_1 + d_2 + \cdots + d_n + \cdots$, for which $a_n \overset{\ge}{=} d_n$ for every n from some term onward, then the infinite series diverges. A useful comparison series is the geometric series $a + ar + ar^2 + \cdots + ar^{n-1} + \cdots$, which converges for $|r| < 1$ and diverges for $|r| \overset{\ge}{=} 1$.

Ratio Test. In the expression:

$$L = \lim_{n \to \infty}\left|\frac{a_n + 1}{a_n}\right|$$

if $L < 1$, the infinite series converges absolutely; if L does not exist or if $L > 1$, the series diverges; if $L = 1$, the test fails.

Root Test. For the expression:

$$L = \lim_{n \to \infty}|a_n|^{1/n}$$

if $L < 1$, the infinite series converges; if $L > 1$, the series diverges; if $L = 1$, the test fails.

Integral Test. For the function $f(n) = a_n$, if $f(x)$ is positive and non-increasing for $x > k$, then the infinite series converges or diverges with the improper integral:

$$\int_k^{\infty} f(x)dx$$

Raabe's Test. For the expression:

$$L = \lim_{n \to \infty} n\left(\frac{a_n}{a_n + 1} - 1\right)$$

if $L > 1$, the infinite series converges; if $L < 1$, the infinite series diverges; if $L = 1$, the test fails.

Convergence of an Alternating Series. In an alternating series:

$$a_1 - a_2 + a_3 - + \cdots + (-1)^{n+1}a_n + \cdots$$

the terms are alternately positive and negative. The series converges if, from some term onward, $|a_{n+1}| \overset{\le}{=} |a_n|$ and $a_n \to 0$ as $n \to \infty$. The sum of the first n terms differs numerically from the sum of the series by less than $|a_{n+1}|$.

Expansion of a Function into a Power Series

Table 2 gives the power series of various functions. A power series takes the form:

Table 2 Expansion of functions into power series

(log = log_e)

Function	Series expansion	Domain of convergence	Function	Series expansion	Domain of convergence
$(a+x)^n$	$a^n + na^{n-1}x + \dfrac{n(n-1)}{2!}a^{n-2}x^2 + \dfrac{n(n-1)(n-2)}{3!}a^{n-3}x^3 + \cdots$	$(x^2 < a^2)$	$\sin^{-1}$	$x + \dfrac{x^3}{2\cdot3} + \dfrac{3x^5}{2\cdot4\cdot5} + \dfrac{3\cdot5x^7}{2\cdot4\cdot6\cdot7} + \cdots$	$(-1 \leqq x \leqq 1)$
e^x	$1 + x + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \dfrac{x^4}{4!} + \cdots$	$(-\infty < x < \infty)$	$\cos^{-1} x$	$\dfrac{\pi}{2} - \sin^{-1} x$	
a^x	$1 + x\log a + \dfrac{(x\log a)^2}{2!} + \dfrac{(x\log a)^3}{3!} + \cdots$	$(-\infty < x < \infty)$	$\tan^{-1} x$	$\dfrac{\pi}{2} - \dfrac{1}{x} + \dfrac{1}{3x^3} - \dfrac{1}{5x^5} + \cdots$	$(x^2 \geqq 1)$
e^{-x^2}	$1 - x^2 + \dfrac{x^4}{2!} - \dfrac{x^6}{3!} + \dfrac{x^8}{4!} - \cdots$	$(-\infty < x < \infty)$		$x - \dfrac{x^3}{3} + \dfrac{x^5}{5} - \dfrac{x^7}{7} + \cdots$	$(-1 \leqq x \leqq 1)$
$e^{\sin x}$	$1 + x + \dfrac{x^2}{2!} - \dfrac{3x^4}{4!} - \dfrac{8x^5}{5!} - \dfrac{3x^6}{6!} - \dfrac{56x^7}{7!} + \cdots$	$(-\infty < x < \infty)$	$\cot^{-1} x$	$\dfrac{\pi}{2} - \tan^{-1} x$	
$e^{\cos x}$	$e\left(1 - \dfrac{x^2}{2!} + \dfrac{4x^4}{4!} - \dfrac{31x^6}{6!} + \cdots\right)$	$(-\infty < x < \infty)$	$\sec^{-1} x$	$\dfrac{\pi}{2} - \dfrac{1}{x} - \dfrac{1}{2\cdot3x^3} - \dfrac{3}{2\cdot4\cdot5x^5} - \dfrac{3\cdot5}{2\cdot4\cdot6\cdot7x^7} - \cdots$	$(x^2 > 1)$
$e^{\tan x}$	$1 + x + \dfrac{x^2}{2!} + \dfrac{3x^3}{3!} + \dfrac{9x^4}{4!} + \dfrac{37x^5}{5!} + \cdots$	$\left(-\dfrac{\pi}{2} < x < \dfrac{\pi}{2}\right)$	$\csc^{-1} x$	$\dfrac{\pi}{2} - \sec^{-1} x$	
$\log x$	$\dfrac{x-1}{x} + \dfrac{1}{2}\left(\dfrac{x-1}{x}\right)^2 + \dfrac{1}{3}\left(\dfrac{x-1}{x}\right)^3 + \cdots$	$\left(x > \dfrac{1}{2}\right)$	$\sinh x$	$x + \dfrac{x^3}{3!} + \dfrac{x^5}{5!} + \dfrac{x^7}{7!} + \cdots$	$(-\infty < x < \infty)$
$\log x$	$2\left[\dfrac{x-1}{x+1} + \dfrac{1}{3}\left(\dfrac{x-1}{x+1}\right)^3 + \dfrac{1}{5}\left(\dfrac{x-1}{x+1}\right)^5 + \cdots\right]$	$(x > 0)$	$\cosh x$	$1 + \dfrac{x^2}{2!} + \dfrac{x^4}{4!} + \dfrac{x^6}{6!} + \dfrac{x^8}{8!} + \cdots$	$(-\infty < x < \infty)$
$\log(1+x)$	$x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \dfrac{x^4}{4} + \cdots$	$(-1 < x < 1)$	$\tanh x$	$x - \dfrac{x^3}{3} + \dfrac{2x^5}{15} - \dfrac{17x^7}{315} + \cdots$	$\left(-\dfrac{\pi}{2} < x < \dfrac{\pi}{2}\right)$
$\log\left(\dfrac{1+x}{1-x}\right)$	$2\left[x + \dfrac{x^3}{3} + \dfrac{x^5}{5} + \dfrac{x^7}{7} + \cdots\right]$	$(-1 < x < 1)$	$\coth x$	$\dfrac{1}{x} + \dfrac{x}{3} - \dfrac{x^3}{45} + \dfrac{2x^5}{945} - \dfrac{x^7}{4725} + \cdots$	$(-\pi < x < \pi)$
$\log\left(\dfrac{x+1}{x-1}\right)$	$2\left[\dfrac{1}{x} + \dfrac{1}{3x^3} + \dfrac{1}{5x^5} + \cdots\right]$	$(x^2 > 1)$	$\operatorname{sech} x$	$1 - \dfrac{x^2}{2!} + \dfrac{5x^4}{4!} - \dfrac{61x^6}{6!} + \dfrac{1385x^8}{8!} - \cdots$	$\left(-\dfrac{\pi}{2} < x < \dfrac{\pi}{2}\right)$
$\log \sin x$	$\log x - \dfrac{x^2}{6} - \dfrac{x^4}{180} - \dfrac{x^6}{2835} - \cdots$	$(-\pi < x < \pi)$	$\operatorname{csch} x$	$\dfrac{1}{x} - \dfrac{x}{6} + \dfrac{7x^3}{360} - \dfrac{31x^5}{15,120} + \cdots$	$(-\pi < x < \pi)$
$\log \cos x$	$-\dfrac{x^2}{2} - \dfrac{x^4}{12} - \dfrac{x^6}{45} - \dfrac{17x^8}{2520} - \cdots$	$\left(-\dfrac{\pi}{2} < x < \dfrac{\pi}{2}\right)$	$\sinh^{-1} x$	$x - \dfrac{x^3}{2\cdot3} + \dfrac{3x^5}{2\cdot4\cdot5} - \dfrac{3\cdot5x^7}{2\cdot4\cdot6\cdot7} + \cdots$	$(-1 < x < 1)$
$\log \tan x$	$\log x + \dfrac{x^2}{3} + \dfrac{7x^4}{90} + \dfrac{62x^6}{2835} + \cdots$	$\left(-\dfrac{\pi}{2} < x < \dfrac{\pi}{2}\right)$	$\sinh^{-1} x$	$\log 2x + \dfrac{1}{2\cdot2x^2} - \dfrac{3}{2\cdot4\cdot4x^4} + \dfrac{3\cdot5}{2\cdot4\cdot6\cdot6x^6} + \cdots$	$(x^2 > 1)$
$\sin x$	$x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} - \dfrac{x^7}{7!} + \cdots$	$(-\infty < x < \infty)$	$\cosh^{-1} x$	$\pm\left(\log 2x - \dfrac{1}{2\cdot2x^2} - \dfrac{1\cdot3}{2\cdot4\cdot4x^4} - \dfrac{1\cdot3\cdot5}{2\cdot4\cdot6\cdot6x^6} - \cdots\right)$	$(x > 1)$
$\cos x$	$1 - \dfrac{x^2}{2!} + \dfrac{x^4}{4!} - \dfrac{x^6}{6!} + \cdots$	$(-\infty < x < \infty)$	$\tanh^{-1} x$	$x + \dfrac{x^3}{3} + \dfrac{x^5}{5} + \dfrac{x^7}{7} + \cdots$	$(-1 < x < 1)$
$\tan x$	$x + \dfrac{x^3}{3} + \dfrac{2x^5}{15} + \dfrac{17x^7}{315} + \dfrac{62x^9}{2835} + \cdots$	$\left(-\dfrac{\pi}{2} < x < \dfrac{\pi}{2}\right)$	$\coth^{-1} x$	$\dfrac{1}{x} + \dfrac{1}{3x^3} + \dfrac{1}{5x^5} + \dfrac{1}{7x^7} + \cdots$	$(x^2 > 1)$
$\cot x$	$\dfrac{1}{x} - \dfrac{x}{3} - \dfrac{x^3}{45} - \dfrac{2x^5}{945} - \dfrac{x^7}{4275} - \cdots$	$(-\pi < x < \pi)$	$\operatorname{sech}^{-1} x$	$\pm\left(\log\dfrac{2}{x} - \dfrac{1}{2\cdot2}x^2 - \dfrac{1\cdot3}{2\cdot4\cdot4}x^4 - \dfrac{1\cdot3\cdot5}{2\cdot4\cdot6\cdot6}x^6 - \cdots\right)$	$(0 < x < 1)$
$\sec x$	$1 + \dfrac{x^2}{2!} + \dfrac{5x^4}{4!} + \dfrac{61x^6}{6!} + \cdots$	$\left(-\dfrac{\pi}{2} < x < \dfrac{\pi}{2}\right)$	$\operatorname{csch}^{-1} x$	$\dfrac{1}{x} - \dfrac{1}{2\cdot3x^3} + \dfrac{3}{2\cdot4\cdot5x^5} - \dfrac{3\cdot5}{2\cdot4\cdot6\cdot7x^7} + \cdots$	$(x^2 > 1)$
$\csc x$	$\dfrac{1}{x} + \dfrac{x}{3!} + \dfrac{7x^3}{3\cdot5!} + \dfrac{31x^5}{3\cdot7!} + \cdots$	$(-\pi < x < \pi)$			

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots$$

If:

$$\lim_{n\to\infty}\left|\frac{a_{n-1}}{a_n}\right| = r$$

the series converges absolutely for all values of x in the interval $-r < x < r$. For $|x| = r$, one of the convergence tests for a series of numerical terms must be used. For example, for a series in the form:

$$1 - \frac{x}{1\cdot2} + \frac{x^2}{2\cdot2^2} - \frac{x^3}{3\cdot2^3} + \cdots + (-1)^n\frac{x^n}{n\cdot2^n} + \cdots$$

Because:

$$\lim_{n\to\infty}\frac{n\cdot2^n}{(n-1)2^{n-1}} = 2$$

the interval of convergence is $-2 < x < 2$. For $x = 2$, the series is a convergent alternating series. For $x = -2$, it is a divergent p-series.

Taylor's Series. A function $y = f(x)$, continuous and with all derivatives at $x = a$, can be expressed as:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots$$
$$+ \frac{f^{(n-1)}(a)}{(n-1)!}(x - a)^{n-1} + \cdots$$

The expression for the remainder after n terms is:

$$R_n = \frac{f^{(n)}(\xi)}{n!}(x - a)^n \quad \xi = a + \theta(x - a)$$
$$0 < \theta < 1$$

Another form of Taylor's series is:

$$f(x + h) = f(x) + \frac{h}{1!}f'(x) + \frac{h^2}{2!}f''(x) + \cdots$$
$$+ \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(x) + \cdots$$

where the remainder after n terms is:

$$R_n = \frac{h^n}{n!}f^{(n)}(\xi) \quad \xi = x + \theta h \quad 0 < \theta < 1$$

Maclaurin's Series. The formula for expansion of a function $f(x)$ in powers of x, obtained from Taylor's series for $a = 0$, is:

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots$$
$$+ \frac{f^{(n-1)}(0)}{(n-1)!}x^{n-1} + \cdots$$

where the remainder after n terms is:

$$R_n = \frac{f^{(n)}(\xi)}{n!}x^n \quad \xi = \theta x \quad 0 < \theta < 1$$

A Taylor's or Maclaurin's series represents a function in an interval only when $R_n \to 0$ as $n \to \infty$. For example, to expand e^{ax} in powers of x, let:

$$f(x) = e^{ax} \quad f'(x) = ae^{ax} \quad f''(x) = a^2 e^{ax}$$
$$f'''(x) = a^3 e^{ax}, \cdots$$
$$f(0) = 1 \quad f'(0) = a \quad f''(0) = a^2$$
$$f'''(0) = a^3, \cdots$$
$$f(x) = e^{ax} = 1 + \frac{a}{1!}x + \frac{a^2}{2!}x^2 + \frac{a^3}{3!}x^3 + \cdots$$

Because:

$$\lim_{n \to \infty} \frac{\frac{a^{n-1}}{(n-1)!}}{\frac{a^n}{n!}} = \lim_{n \to \infty} \frac{n}{a} = \infty$$

the series converges for all values of x.

Taylor's series for a function of two variables is expressed in the following:

$$f(x + h, y + k) = f(x, y)\,\frac{1}{1!}\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)$$
$$f(x, y) + \frac{1}{2!}\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^2 f(x, y) + \cdots +$$
$$\frac{1}{(n-1)!}\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^{n-1} f(x, y) + \cdots$$

where the remainder is:

$$R_n = \frac{1}{n!}\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^n f(x + \theta h, y + \theta k)$$
$$0 < \theta < 1$$

Fourier Series. If a function $f(x)$ varies over an interval of length $2l$ and can be expressed as the difference of two nondecreasing or nonincreasing bounded functions, then:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos\frac{n\pi x}{l} + b_n \sin\frac{n\pi x}{l}\right)$$
$$= \frac{a_0}{2} + a_1 \cos\frac{\pi x}{l} + a_a \cos\frac{2\pi x}{l} + \cdots +$$
$$b_1 \sin\frac{\pi x}{l} + b_2 \sin\frac{2\pi x}{l} + \cdots\Big)$$

for which:

$$a_n = \frac{1}{l}\int_k^{k+2l} f(x)\cos\frac{n\pi x}{l}\,dx$$
$$b_n = \frac{1}{l}\int_k^{k+2l} f(x)\sin\frac{n\pi x}{l}\,dx$$
$$n = 0, 1, 2, \cdots$$

In exponential form, the function is:

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{\frac{in\pi x}{l}}$$
$$c_n = \frac{1}{2l}\int_k^{k+2l} f(x)e^{\frac{-in\pi x}{l}}\,dx$$
$$n = \cdots, -2, -1, 0, 1, 2, \cdots$$

At a point of discontinuity, a Fourier series produces the value at the midpoint of the jump. For example, when e^x is expanded in the interval 0 to 2π:

$$a_0 = \frac{1}{\pi}\int_0^{2\pi} e^x dx = \frac{1}{\pi}(e^{2\pi} - 1)$$
$$a_n = \frac{1}{\pi}\int_0^{2\pi} e^x \cos nx\,dx = \frac{e^{2\pi} - 1}{\pi(n^2 + 1)}$$
$$b_n = -\frac{n(e^{2\pi} - 1)}{\pi(n^2 + 1)}$$

Thus:

$$e^x = \frac{1}{\pi}(e^{2\pi} - 1)\left[\frac{1}{2} + \frac{1}{1^2 + 1}\cos x + \frac{1}{2^2 + 1}\cos 2x\right.$$
$$+ \frac{1}{3^2 + 1}\cos 3x + \cdots\Big] - \frac{1}{\pi}(e^{2\pi} - 1)\left[\frac{1}{1^2 + 1}\right.$$
$$\sin x + \frac{2}{2^2 + 1}\sin 2x + \frac{3}{3^2 + 1}\sin 3x + \cdots\Big]$$

Only in the interval from 0 to 2π does the expansion have validity; beyond that interval, the series repeats itself, as a result of the periodic behavior of $\sin nx$ and $\cos nx$. In a Fourier series, if $f(-x) = f(x)$, it is an even function. Then:

$$a_n = \frac{2}{l}\int_0^l f(x)\cos\frac{n\pi x}{l}\,dx \quad n = 0, 1, 2, \cdots$$

and:

$$b_n = 0$$

If $f(-x) = -f(x)$, it is an odd function, $a_n = 0$, and:

$$b_n = \frac{2}{l}\int_0^l f(x)\sin\frac{n\pi x}{l}\,dx, n = 0, 1, 2, \cdots$$

For example, if the function $f(x) = x$ is expanded in a cosine series in the interval $(0, \pi)$:

$$\frac{1}{2}a_0 = \frac{1}{\pi}\int_0^{\pi} x\,dx = \frac{\pi}{2}$$

and:

$$a_n = \frac{2}{\pi}\int_0^{\pi} x\cos nx dx = \frac{2}{\pi}\left\{\left[\frac{x\sin nx}{n}\right]_0^{\pi} - \int_0^{\pi}\frac{\sin nx}{n}dx\right\}$$
$$= \frac{2}{\pi}\left[\frac{1}{n^2}\cos nx\right]_0^{\pi} = \frac{2}{\pi n^2}(\cos n\pi - 1)$$

Thus:

$$x = \frac{\pi}{2} - \frac{4}{\pi}\left[\cos x + \frac{\cos 3x}{3^2} + \frac{\cos 5x}{5^2} + \cdots\right]$$
$$(0 < x < \pi)$$

For $x = 0$, the sum of the series is 0, and for $x = \pi$, the sum of the series is π.

Ordinary Differential Equations

An equation involving derivatives or differentials is a differential equation. A differential equation is of order n if the equation involves a derivative of order n and contains no higher-order derivative.

If the equation has derivatives with respect to one variable only, it is an ordinary differential equation; otherwise, it is a partial differential equation. For example, the equations:

$$\frac{d^2y}{dx^2} + k^2 y = 0 \tag{Eq 1}$$

$$\frac{d^2y}{dx^2} = \sqrt{1 + y^2 + \frac{dy}{dx}} \tag{Eq 2}$$

$$y - x\frac{dy}{dx} + 3\frac{dx}{dy} = 0 \tag{Eq 3}$$

are ordinary differential equations. The equation:

$$y\frac{\partial^2 z}{\partial x^2} + zx\frac{\partial^2 z}{\partial x \partial y} - \frac{\partial z}{\partial y} = xyz \tag{Eq 4}$$

is a partial differential equation. The degree of a differential equation refers to the greatest power to which the highest-order derivative occurs in an equation that has been made rational and integral in its derivatives. For the ordinary differential equation of order n:

$$F(x, y, y', y'', \ldots, y^{(n)}) = 0 \tag{Eq 5}$$

a solution, or integral, is a function $y = \phi(x)$ for which F is transformed into a function identically zero for all values of x, when $\phi(x)$, $\phi'(x)$, ..., $\phi^{(n)}(x)$ are substituted for y, y', ..., $y^{(n)}$, respectively. The general solution of Eq 5 has the form:

$$y = \Phi(x, c_1, c_2, \ldots, c_n) \qquad \text{(Eq 6)}$$

where c_1, c_2, ..., c_n are arbitrary constants. If Eq 6 is differentiated n times with respect to x, and if the n constants c_i that appear in the resulting n equations and Eq 6 are eliminated, the resulting differential equation is Eq 5. Equation 6 is the primitive of Eq 5.

A particular solution is a solution obtained by giving specific values to the constants in Eq 6. Singular solutions to Eq 5 are solutions that may exist but cannot be obtained by giving the constants c_i special values.

A differential equation is considered solved when the solution has been reduced to quadratures or if y has been obtained merely as an implicit function of x by $\psi(x, y, c_1, \ldots, c_n) = 0$.

First-Order Equations

A first-order differential equation in the form:

$$\frac{dy}{dx} = f(x, y) \qquad \text{(Eq 7)}$$

has a solution $y = y(x)$ through every point ($x = x_0$, $y = y_0$) having a neighborhood in which $f(x, y)$ is continuous. For example, if D is a region of points (x, y), (x, η) where $f(x, y)$ is continuous. For example, if D is a region of points (x, y), (x, η) where $f(x, y)$ is single-valued, bounded, and continuous, and the condition $|f(x, y) - f(x, \eta)| \leq M|y - \eta|$ holds for some real M independent of y and η, then Eq 7 has a unique solution $y = y(x)$ through every point ($x = x_0$, $y = y_0$) of D. The solution $y(x)$ is a continuous function of the given value $y_0 = y(x_0)$. Each solution extends to the boundary of D.

Methods of Solution. The following special types of first-order equations can be solved with relative ease:

- Equations with variables that are separable: $y' = f_1(x)/f_2(y)$. The solution can be obtained from $\int f_2(y)dy = \int f_1(x)dx + C$.
- Homogeneous first-order equations: $y' = f(y/x)$. Introduce $\overline{y} = y/x$ to reduce the equation to one in which variables can be separated.

Separation of Variables. For the differential equation:

$$M\,dx + N\,dy = 0 \qquad \text{(Eq 8)}$$

where $M(x, y)$ and $N(x, y)$ are functions of x and y. If M and N are products of factors, each being a function either of x alone or of y alone, then Eq 8 takes the form:

$$A(x)P(y)dx + B(x)Q(y)dy = 0 \qquad \text{(Eq 9)}$$

where A and B are functions of x alone, and P and Q are functions of y alone. The general solution of Eq 9 is:

$$\int \frac{A(x)}{B(x)}dx + \int \frac{Q(y)}{P(y)}dy = c \qquad \text{(Eq 10)}$$

with c a constant.

The expression $dy/dx = -x/y$ can be written as $x\,dx + y\,dy = 0$. It has the solution $\int x\,dx + \int y\,dy = x^2/2 + y^2/2 = c$. If $c = \frac{r^2}{2}$, then $x^2 + y^2 = r^2$, is a set of concentric circles. An infinite number of solutions exists, depending on the value of r, and through each point in the plane there passes only one circle.

Homogeneous Equations. If $f(kx, ky) = k^n f(x, y)$, a function $f(x, y)$ is homogeneous of the nth degree in x and y. An equation:

$$P(x, y)dx + Q(x, y)dy = 0 \qquad \text{(Eq 11)}$$

is homogeneous when functions $P(x, y)$ and $Q(x, y)$ are homogeneous in x and y. The variables can be separated by substituting $y = vx$.

For example, to solve $(x^2 + y^2)dx - 2\,xy\,dy = 0$, which is of the form $P(x, y)\,dx + Q(x, y)\,dy = 0$, where P and Q are homogeneous functions of the second degree, substitute $y = vx$. The equation becomes $(1 + v^2)\,dx - 2v(x\,dv + v\,dx) = 0$. Separating variables produces:

$$\frac{dx}{x} - \frac{2v}{1 - v^2}\,dv = 0$$

By integrating, $\log_e x(1 - v^2) = \log_e c$; by replacing $v = y/x$, $\log(1 - y^2/x^2)x = \log_e c$; and by taking exponentials, $x^2 - y^2 = cx$.

Exact Differential Equation. The equation:

$$P(x, y)dx + Q(x, y)dy = 0 \qquad \text{(Eq 12)}$$

is an exact differential equation when the expression on the left side is an exact differential, $d\phi(\partial P/\partial y = \partial Q/\partial x)$. The solution is obtained from:

$$\phi(x, y) = \int P(x, y)dx + \int \left[Q(x, y) - \frac{\partial}{\partial y}\int P(x, y)dy \right]dy = C \quad \text{(Eq 13)}$$

If the expression on the left of Eq 12 is not an exact differential ($\partial P/\partial y \neq \partial Q/\partial x$), an integrating factor $\mu = \mu(x, y)$ may be found to permit multiplication of Eq 12 by $\mu(x, y)$ to yield an exact differential equation. The partial differential equation:

$$\mu\left(\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x}\right) = Q\frac{\partial \mu}{\partial x} - P\frac{\partial \mu}{\partial y} \qquad \text{(Eq 14)}$$

is satisfied by the integrating factor $\mu(x, y)$.

Linear Differential Equation. The linear first-order equation $y' + a(x)y = f(x)$ can be modified by the integrating factor $\mu = \mu(x) = \exp \int a(x)dx$. The complete primitive becomes:

$$y = \frac{1}{\mu(x)}\left[\int f(x)\mu(x)dx + C\right] \qquad \text{(Eq 15)}$$

Many first-order equations can be reduced by transformation of variables. In particular:

- $y' = f(\alpha x + \alpha y)$ is reduced to an equation with variables that are separable by introducing $\overline{y} = \alpha x + \beta y$.
- $y' = \alpha_1 x + \beta_1 y + \gamma_1/\alpha_2 x + \beta_2 y + \gamma_2$ is reduced to a homogeneous first-order equation by a coordinate translation if $\alpha_1 \beta_2 - \alpha_2 \beta_1 \neq 0$; or by introducing $\overline{y} = \alpha_2 x + \beta_2 y + \gamma_2$ to separate the variables.

Bernoulli's Differential Equation. The Bernoulli equation is:

$$dy/dx + P(x)y = Q(x)y^n \qquad \text{(Eq 16)}$$

in which $n \neq 1$, is reduced to a linear equation by substituting $z = y^{1-n}$. The general solution is:

$$y = e^{-\int P(x)dx}\left[(1 - n)\int e^{(1-n)\int P(x)dx}Q(x)dx + c\right]^{\frac{1}{1-n}} \qquad \text{(Eq 17)}$$

For example, the equation $dy/dx - xy = xy^2$ can be solved by substituting $z = y^{-1}$ to obtain $dy/dx + xz = -x$. The general integral is:

$$z = ce^{-x^2/2} - 1 \text{ or } y = \frac{1}{ce^{-x^2/2}-1}$$

Riccati Equations. The general Riccati differential equation:

$$y' = a(x)y^2 + b(x)y + c(x) \qquad \text{(Eq 18)}$$

can be simplified by the transformation $y = 1/\overline{y}$. The transformations:

$$x = \overline{x} \;\; y = \frac{-\overline{y}'}{a(x)\overline{y}}$$

lead to a homogeneous second-order equation for $\overline{y} = \overline{y}(x)$:

$$\overline{y}'' - \left[\frac{a'(x)}{a(x)} + b(x)\right]\overline{y}' + a(x)c(x)\overline{y} = 0 \qquad \text{(Eq 19)}$$

If a particular integral $y_1(x)$ of Eq 18 is known, the transformation:

$$y = y_1(x) + 1/\overline{y}$$

produces a linear differential equation. If two particular integrals y_1, y_2 or three particular integrals y_1, y_2, y_3 are known, then:

$$y = y_1 + \frac{y_2 - y_1}{1 + C \, \exp \int a(x)(y_2 - y_1)dx}$$

$$y = \frac{y_1(y_2 - y_3) + Cy_2(y_1 - y_3)}{y_2 - y_3 + C(y_1 - y_3)} \qquad \text{(Eq 20)}$$

Second-Order Equations

The equation:

$$F\left(x, y, \frac{dy}{dx}, \frac{d^2y}{dx^2}\right) = 0 \qquad \text{(Eq 21)}$$

is a second-order differential equation. The method of solution is straightforward if some of these variables are missing. With y and $\frac{dy}{dx}$ missing, for example, the equation:

$$\frac{d^2y}{dx^2} = f(x) \qquad \text{(Eq 22)}$$

has the solution $y = \int dx \int f(x)dx + cx + c_1$. For the equation:

$$\frac{d^2y}{dx^2} = f(y) \qquad \text{(Eq 23)}$$

with x and dy/dx missing, both sides can be multiplied by $2dy/dx$ to obtain:

$$x = \int \frac{dy}{\sqrt{c + 2\int f(y)dy}} + c_1 \qquad \text{(Eq 24)}$$

In the equation:

$$\frac{d^2y}{dx^2} = f\left(\frac{dy}{dx}\right) \qquad \text{(Eq 25)}$$

with x and y missing, using the transformations:

$$\frac{dy}{dx} = p \qquad \frac{d^2y}{dx^2} = \frac{dp}{dx} \qquad \text{(Eq 26)}$$

produces

$$x = \int \frac{dp}{f(p)} + c$$

This equation can be solved for p; p can be replaced by dy/dx; the resulting first-order equation can then be solved.

Homogeneous linear second-order equations in the form:

$$Lw \equiv \frac{d^2w}{dz^2} + a_1(z)\frac{dw}{dz} + a_2(z)w = 0 \qquad \text{(Eq 27)}$$

are equivalent to:

$$\frac{d}{dz}\left[p(z)\frac{dw}{dz}\right] + q(z)w = 0 \qquad \text{(Eq 28)}$$

with:

$$p(z) = \exp\int a_1(z)dz \quad q(z) = a_2(z)p(z)$$

When a solution $w_1(z)$ of Eq 27 or 28 is known, the complete primitive is:

$$w(z) = w_1(z)\left[C_1 + C_2\int\frac{dz}{w_1^2(z)p(z)}\right] \qquad \text{(Eq 29)}$$

Homogeneous Differential Equation with Constant Coefficients. A solution of the equation:

$$\frac{d^ny}{dx^n} + a_1\frac{d^{n-1}y}{dx^{n-1}} + \cdots + a_{n-1}\frac{dy}{dx} + a_ny = 0 \qquad \text{(Eq 30)}$$

becomes:

$$y_k = ce^{r_kx} \qquad \text{(Eq 31)}$$

if r_k is a root of the algebraic equation:

$$r^n + a_1r^{n-1} + \cdots + a_{n-1}r + a_n = 0 \qquad \text{(Eq 32)}$$

If n roots $r_1, r_2, \ldots, r_n$ of Eq 32 differ, then:

$$y = c_1e^{r_1x} + c_2e^{r_2x} + \cdots + c_ne^{r_nx} \qquad \text{(Eq 33)}$$

becomes a general solution of Eq 30. If k roots are equal, $r_1 = r_2 = \cdots = r_k$ while $r_{k+1}, \ldots, r_n$ differ, then:

$$y = (c_1 + c_2x + \cdots + c_kx^{k-1})e^{r_1x} + c_{k+1}e^{r_{k+1}x} + \cdots + c_ne^{r_nx} \qquad \text{(Eq 34)}$$

becomes a general solution. If $r_1 = p + iq$, $r_2 = p - iq$ are conjugate complex roots of Eq 32, then:

$$c_1e^{r_1x} + c_2e^{r_2x} = e^{px}(C_1\cos qx + C_2\sin qx) \qquad \text{(Eq 35)}$$

Nonhomogeneous Differential Equations with Constant Coefficients. The general solution of nonhomogeneous differential equations with constant coefficients is:

$$Ly \equiv a_0\frac{d^ry}{dt^r} + a_1\frac{d^{r-1}}{dt^{r-1}} + \cdots + a_ry = f(t) \qquad \text{(Eq 36)}$$

If $f(t) = 0$ for $t \leq 0$, the particular integral $y = y_N(t)$ of Eq 36 with $y_N = y'_N = y''_N = \cdots = y_N^{(r-1)} = 0$ for $t \leq 0$ is the normal response to the given forcing function $f(t)$.

Each function $y_k = y_k(t)$ of a system of linear differential equations with constant coefficients:

$$\phi_{j1}\left(\frac{d}{dt}\right)y_1 + \phi_{j2}\left(\frac{d}{dt}\right)y_2 + \cdots + \phi_{jn}\left(\frac{d}{dt}\right)y_n = f_j(t)$$
$$(j = 1, 2, \ldots, n) \qquad \text{(Eq 37)}$$

is the sum of the corresponding solution function of a complementary homogeneous system and a particular solution function.

Euler's Homogeneous Equation. For the equation:

$$x^n\frac{d^ny}{dx^n} + ax^{n-1}\frac{d^{n-1}y}{dx^{n-1}} + \cdots + a_{n-1}x\frac{dy}{dx} + a_ny = 0 \qquad \text{(Eq 38)}$$

place $x = e^t$ because:

$$x\frac{dy}{dx} = \frac{dy}{dt} \qquad x^2\frac{d^2y}{dx^2} = \left[\frac{d}{dt}\left(\frac{d}{dt} - 1\right)\right]y$$
$$x^3\frac{d^3y}{dx^3} = \left[\frac{d}{dt}\left(\frac{d}{dt} - 1\right)\left(\frac{d}{dt} - 2\right)\right]y, \ldots \qquad \text{(Eq 39)}$$

then Eq 38 becomes a linear homogeneous differential equation with constant coefficients.

Depression of Order. When, for a linear homogeneous differential equation, a particular integral is known, the order of the equation can be lowered. If y_1 is a particular integral of:

$$\frac{d^ny}{dx^n} + P_1(x)\frac{d^{n-1}y}{dx^{n-1}} + \cdots + P_{n-1}(x)\frac{dy}{dx} + P_n(x) = 0 \qquad \text{(Eq 40)}$$

$y = y_1 z$ can be substituted. The coefficient of z is zero. Placing $dz/dx = u$ reduces the equation to the $(n - 1)$ st order. For example, for the equation:

$$\frac{d^2y}{dx^2} + p(x)\frac{dy}{dx} + q(x)y = 0$$

let $y = y_1z$. Then:

$$\frac{dy}{dx} = y_1\frac{dz}{dx} + z\frac{dy_1}{dx}$$
$$\frac{d^2y}{dx^2} = y_1\frac{d^2z}{dx^2} + 2\frac{dy_1}{dx}\frac{dz}{dx} + z\frac{d^2y_1}{dx^2}$$

Substituting in the original equation produces:

$$y_1\frac{d^2z}{dx^2} + 2\frac{dy_1}{dx}\frac{dz}{dx} + z\frac{d^2y_1}{dx^2} + p\left[y_1\frac{dz}{dx} + z\frac{dy_1}{dx}\right] + qy_1z = 0$$

Because the coefficient of z is zero, this expression reduces to:

$$y_1\frac{d^2z}{dx^2} + \left(2\frac{dy_1}{dx} + py_1\right)\frac{dz}{dx} = 0$$

By writing:

$$\frac{dz}{dx} = u, \frac{du}{u} + \left(2\frac{dy_1}{dx} + py_1\right)\frac{dx}{y_1} = 0$$

Integrating produces:

$$\log_e u + \int p\,dx + \log_e y_1^2 = \log_e c \quad \text{or}$$
$$u = \frac{c}{|y_1^2|}e^{-\int p\,dx}$$

Integrating again gives z, and:

$$y = y_1\int\frac{c}{y_1^2}e^{-\int p\,dx}dx + c_1$$

Pfaffian Differential Equations

A Pfaffian differential equation, or first-order linear total differential equation:

$$P(x, y, z)dx + Q(x, y, z)dy + R(x, y, z)dx = 0 \qquad \text{(Eq 41)}$$

with continuously differentiable coefficients P, Q, R, can be interpreted geometrically as a condition $\mathbf{P} \cdot d\mathbf{r} = 0$ on the tangent vector $d\mathbf{r} \equiv (dx, dy, dz)$ of

an integral curve. The curve is described by two equations $f(x, y, z) = 0$, $g(x, y, z, C) = 0$, with C a constant of integration. To determine the integral curves on an arbitrary regular surface:

$$f(x, y, z) = 0 \qquad \text{(Eq 42)}$$

solve the ordinary differential equation obtained by eliminating z and dz from Eq 41 and solve the equation $df(x, y, z) = 0$.

Equation 41 is integrable if there exists an integrating factor $\mu = \mu(x, y, z)$ for which $\mu(P\,dx + Q\,dy + R\,dz)$ is an exact differential $d\phi(x, y, z)$; that is, if:

$$P\left(\frac{\partial Q}{\partial z} - \frac{\partial R}{\partial y}\right) + Q\left(\frac{\partial R}{\partial x} - \frac{\partial P}{\partial z}\right) + R\left(\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x}\right) = 0 \tag{Eq 43}$$

In such a case, each curve on an integral surface, $\phi(x, y, z) = C$, orthogonal to a series of curves described by $dx/P = dy/Q = dz/R$ is a solution.

Linear Differential Equations with Constant Coefficients

For n linear equations with constant coefficients and n dependent variables and one independent variable t, a solution may be reached with the symbolic algebraic method. If $n = 2$:

$$\left.\begin{aligned}
(D^n + a_1 D^{n-1} + \cdots + a_n) \\
x + (D^m + b_1 D^{m-1} + \cdots + b_m)y = R(t) \\
(D^p + c_1 D^{p-1} + \cdots + c_p) \\
x + (D^q + d_1 D^{q-1} + \cdots + d_q)y = S(t)
\end{aligned}\right\} \tag{Eq 44}$$

where $D = d/dt$. The equations may be expressed as:

$$P_1(D)x + Q_1(D)y = R \quad P_2(D)x + q_2(D)y = S$$

Either x or y can be eliminated to solve the equation obtained by treating these as algebraic equations. For example, the system of equations:

$$\frac{dx}{dt} + \frac{dy}{dt} + 2x + y = 0 \tag{Eq 45a}$$

$$\frac{dy}{dt} + 5x + 3y = 0 \tag{Eq 45b}$$

can be written:

$$(D+2)x + (D+1)y = 0 \quad 5x + (D+3)y = 0$$

by using the symbol D. Eliminating x, $(D^2 + 1)y = 0$. Using Eq 35, Eq 45(a) has the solution $y = c_1 \cot t + c_2 \sin t$. Substituting this result in Eq 45(b):

$$x = -\frac{3c_1 + c_2}{5}\cos t + \frac{c_1 - 3c_2}{5}\sin t$$

Simultaneous Total Differential Equations

The system:

$$\begin{cases} P_1 dx + Q_1 dy + R_1 dz = 0 \\ P_2 dx + Q_2 dy + R_2 dz = 0 \end{cases} \tag{Eq 46}$$

where P_1, Q_1, R_1, P_1, are functions of x, y, and z, may be written:

$$\frac{dx}{P} = \frac{dy}{Q} = \frac{dz}{R} \tag{Eq 47}$$

where:

$$P = \begin{vmatrix} Q_1 & R_1 \\ Q_2 & R_2 \end{vmatrix} \quad Q = \begin{vmatrix} R_1 & P_1 \\ R_2 & P_2 \end{vmatrix} \quad R = \begin{vmatrix} P_1 & Q_1 \\ P_2 & Q_2 \end{vmatrix} \tag{Eq 48}$$

The general solution of Eq 46 consists of relationships involving two arbitrary constants:

$$\phi_1(x, y, z, c_1) = 0 \quad \phi_2(x, y, z, c_2) = 0 \tag{Eq 49}$$

Partial Differential Equations

Partial differential equations involve partial derivatives, and common types of linear partial differential equations are listed in Table 3. Explicit solutions can be reached in only a relatively few cases and usually involve arbitrary functions, in a manner similar to the solutions of ordinary differential equations, which involve arbitrary constants. In practice, a problem usually involves determining a function that satisfies the differential equation and meets specific initial or boundary conditions.

First-Order Partial Differential Equations

Interpreted geometrically for a first-order partial differential equation:

$$\begin{aligned} F(x, y, z, p, q) &= 0 \\ \left(p \equiv \frac{\partial z}{\partial x}, q \equiv \frac{\partial z}{\partial y}; F_{p^2} + F_{q^2} \neq 0\right) \end{aligned} \tag{Eq 50}$$

with an unknown function $z = z(x, y)$, assume the given function F is single-valued and twice continuously differentiable and x, y, z as rectangular Cartesian coordinates. Every solution $z = z(x,y)$ of Eq 50 represents a surface with a normal having the direction numbers p, q, -1 at every surface point (x, y, z). The solution surface is defined by:

$$dx : dy : dz = F_p : F_q : (pF_p + qF_q)$$

at every point (x, y, z). A set of values (x, y, z, p, q) describes a planar element associating the direction numbers p, q, -1 with a point (x, y, z).

Characteristic Equations. A set of differentiable functions:

$$x = x(t) \quad y = y(t) \quad z = z(t) \quad p = p(t) \quad q = q(t) \tag{Eq 51}$$

represents the points and tangent planes along a strip of a regular surface if the functions (Eq 51) satisfy the strip condition:

$$\frac{dz}{dt} = p\frac{dx}{dt} = q\frac{dy}{dt}$$

Every set of functions (Eq 51) that satisfies the characteristic ordinary differential equations associated with Eq 50, that is:

$$\left.\begin{aligned}
\frac{dz}{dt} &= p\frac{dx}{dt} + q\frac{dy}{dt} & \frac{dx}{dt} &= F_p & \frac{dy}{dt} &= F_q \\
\frac{dp}{dt} &= -(pF_z + F_x) & \frac{dq}{dt} &= -(qF_z + F_y)
\end{aligned}\right\} \tag{Eq 52}$$

together with Eq 50 describe a characteristic strip.

Linear First-Order Partial Equation. The equation:

$$Ap + Bq = C \tag{Eq 53}$$

for which A, B, and C are functions of x, y, and z and:

$$p = \frac{\partial z}{\partial x} = z_x \quad q = \frac{\partial z}{\partial y} = z_y$$

has the general solution $\Phi(u, v) = 0$. In this general solution, Φ is an arbitrary function, $u(x, y, z) = c_1$ and $v(x, y, z) = c_2$. These elements form the general solution of the differential equations:

Table 3   Linear partial differential equations

Physical background	One-dimensional	Multidimensional
Parabolic(a) Heat conduction, diffusion	$\frac{\partial^2 \Phi}{\partial x^2} - \frac{1}{\gamma^2}\frac{\partial \Phi}{\partial t} = f(x, t)$	$\nabla^2 \Phi - \frac{1}{\gamma^2}\frac{\partial \Phi}{\partial t} = f(\mathbf{r}, t)$
Hyperbolic(b) Waves (strings, membranes, fluids, electromagnetic)	$\frac{\partial^2 \Phi}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 \Phi}{\partial t^2} = f(x, t)$	$\nabla^2 \Phi - \frac{1}{c^2}\frac{\partial^2 \Phi}{\partial t^2} = f(\mathbf{r}, t)$
Damped waves, transmission lines	$\frac{\partial^2 \Phi}{\partial x^2} - a_0\frac{\partial^2 \Phi}{\partial t^2} - a_1\frac{\partial \Phi}{\partial t} - a_2\Phi = f(x, t)$	$\nabla^2 \Phi - a_0\frac{\partial^2 \Phi}{\partial t^2} - a_1\frac{\partial \Phi}{\partial t} - a_2\Phi = f(\mathbf{r}, t)$
Elliptic(c) Static case	$\frac{\partial^2 \Phi}{\partial x^2} = f(x)$	$\nabla^2 \Phi = f(\mathbf{r})$
4th order(b) Elastic vibrations	$\frac{\partial^4 \Phi}{\partial x^4} + \frac{1}{c^2}\frac{\partial^2 \Phi}{\partial t^2} = f(x, t)$	$\nabla^2\nabla^2\Phi + \frac{1}{c^2}\frac{\partial^2 \Phi}{\partial t^2} = f(\mathbf{r}, t)$
4th order(c) Static case	$\frac{\partial^4 \Phi}{\partial x^4} = f(x)$	$\nabla^2 \nabla^2 \Phi = f(\mathbf{r})$

(a) Boundary conditions; initial conditions on Φ. (b) Boundary conditions; initial conditions on Φ and $\partial\Phi/\partial t$. (c) Boundary conditions only

$$\frac{dx}{A} = \frac{dy}{B} = \frac{dz}{C} \tag{Eq 54}$$

The normal to the surface $\Phi(u, v) = 0$ is perpendicular to the curves of Eq 54. The characteristic curves of Eq 53 are $u = c_1$ and $v = c_2$.

General Solution. For the partial differential equation $F(x, y, z, p, q) = 0$, because z is a function of x and y, then $dz = p\,dx + q\,dy$. If another relationship can be found among x, y, z, p, q, so as to produce $f(x, y, z, p, q) = 0$, then p and q can be eliminated. The solution of the resulting ordinary differential equation involving x, y, z satisfies the given equation, $F(x, y, z, p, q) = 0$. The following linear partial differential equation must be satisfied by the unknown function f:

$$\frac{\partial F}{\partial p}\frac{\partial f}{\partial x} + \frac{\partial F}{\partial q}\frac{df}{\partial y} + \left(p\frac{\partial F}{\partial p} + q\frac{\partial F}{\partial q}\right)\frac{\partial f}{\partial z}$$
$$- \left(\frac{\partial F}{\partial x} + p\frac{\partial F}{\partial z}\right)\frac{\partial f}{\partial p} - \left(\frac{\partial F}{\partial y} + q\frac{\partial F}{\partial z}\right)\frac{\partial F}{\partial q} = 0 \tag{Eq 55}$$

This equation is satisfied by any of the solutions of the system:

$$\frac{\partial x}{\frac{\partial F}{\partial p}} = \frac{\partial y}{\frac{\partial F}{\partial q}} = \frac{dz}{\frac{\partial F}{\partial p} + \frac{q\partial F}{\partial q}} = \frac{-dp}{\frac{\partial F}{\partial x} + \frac{p\partial F}{\partial z}} = \frac{-dq}{\frac{\partial F}{\partial y} + \frac{q\partial F}{\partial z}} \tag{Eq 56}$$

Second-Order Partial Differential Equations

A linear second-order partial differential equation with two independent variables has the form:

$$L = Ar + 2Bs + Ct + Dp + Eq + Fz = f(x, y) \tag{Eq 57}$$

for which:

$$r = \frac{\partial^2 z}{\partial x^2} \quad s = \frac{\partial^2 z}{\partial x \partial y} \quad t = \frac{\partial^2 z}{\partial y^2} \quad p = \frac{\partial z}{\partial x} \quad q = \frac{\partial z}{\partial y}$$

Coefficients $A, \ldots, F$ are real continuous functions of the real variables x and y. For the following homogeneous partial differential equation of the first order:

$$Ap^2 + 2Bpq + Cq^2 = 0 \tag{Eq 58}$$

assume that two solutions are $\xi = \xi(x, y)$, $\eta = \eta(x, y)$.

The homogeneous form of Eq 57, $L = 0$, is the parabolic type, if $B^2 - AC = 0$, and has the normal form:

$$\frac{\partial^2 z}{\partial \xi^2} + a\frac{\partial z}{\partial \xi} + b\frac{\partial z}{\partial \eta} + cz = 0 \tag{Eq 59}$$

where a, b, c are functions of ξ and η.

If $B^2 - AC > 0$ in Eq 58, the homogeneous form of Eq 57 is hyperbolic, which has two normal forms:

$$\frac{\partial^2 z}{\partial \xi \partial \eta} + a\frac{\partial z}{\partial \xi} + b\frac{\partial z}{\partial \eta} + cz = 0 \tag{Eq 60}$$

$$\frac{\partial^2 z}{\partial \xi^2} - \frac{\partial^2 z}{\partial \eta^2} + a\frac{\partial z}{\partial \xi} + b\frac{\partial z}{\partial \eta} + cz = 0 \tag{Eq 61}$$

The equation is an elliptic type that has the normal form:

$$\frac{\partial^2 z}{\partial \xi^2} + \frac{\partial^2 z}{\partial z^2} + a\frac{\partial z}{\partial \xi} + b\frac{\partial z}{\partial \eta} + cz = 0 \tag{Eq 62}$$

when $B^2 - AC < 0$.

Laplace's Equation. The general solution of the equation:

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0 \tag{Eq 63}$$

is of the form:

$$\phi = f_1(x + iy) + f_2(x - iy) \quad i^2 = -1 \tag{Eq 64}$$

with f_1 and f_2 being arbitrary functions. In practice, this solution is too general, because determining this function so as to satisfy given boundary conditions is difficult. A function that satisfies Eq 63 is a harmonic function.

A useful method depends on the assumption that a particular solution is a product of functions each containing only one of the variables. Combining a number of such solutions often results in a sufficiently general solution. For example, assuming:

$$\phi = X(x) \cdot Y(y) \tag{Eq 65}$$

is a solution of Eq 63, then $(X''/X) + (Y''/Y) = 0$. From this:

$$\frac{1}{X}\frac{d^2 X}{dx^2} = -\omega^2 \quad \frac{1}{Y}\frac{d^2 Y}{dy^2} = \omega^2 \tag{Eq 66}$$

where ω is a constant. Solutions to Eq 66 are:

$$X = c_1 \cos \omega x + c_2 \sin \omega x \quad Y = c_3 e^{\omega y} + c_4 e^{-\omega y}$$

in which c_1, c_2, c_3, c_4 are arbitrary constants. Equation 65 becomes:

$$\phi = \phi(\omega)$$
$$= e^{\omega y}(A_\omega \cos \omega x + B_\omega \sin \omega x)$$
$$+ e^{-\omega y}(C_\omega \cos \omega x + D_\omega \sin \omega x)$$

in which A_ω, B_ω, C_ω, D_ω are arbitrary constants. Generally:

$$\phi = \sum_{\omega = 0}^{\infty} \phi(\omega) \tag{Eq 67}$$

may be assumed to be a solution to Eq 63, with the constants determined so that Eq 67 will satisfy initial or boundary conditions of a particular problem.

Equations Linear in the Second Derivatives. The general second-order equation linear in the second derivative may be expressed as:

$$Ar + Bs + Ct = V \tag{Eq 68}$$

for which A, B, C, V are functions of x, y, z, p, q. From the equations:

$$A\,dy^2 - B\,dx\,dy + C\,dx^2 = 0 \tag{Eq 69}$$

$$A\,dp\,dy + C\,dq\,dx - V\,dx\,dy = 0 \tag{Eq 70}$$

$$p\,dx + q\,dy = dz \tag{Eq 71}$$

one or two relationships may be derived between x, y, z, p, q. These are intermediary integrals. From these, the solution of Eq 68 may be deduced. To obtain an intermediary integral, resolve Eq 69, assuming that the left member is not a perfect square, into the equations $dy - n_1\,dx = 0$ and $dy - n_2\,dx = 0$. From the first and from Eq 70, combined with Eq 71, if necessary, obtain the two integrals $u_1(x, y, z, p, q) = a$ and $v_1(x, y, z, p, q) = b$. Then $u_1 = f_1(v_1)$, with f_1 an arbitrary function, becomes an intermediary integral. In the same way, from $dy - n_2 dx = 0$, another pair of integrals $u_2 = a_1$, $v_2 = b_1$ is obtained. Then $u_2 = f_2(v_2)$ is an intermediary integral. If $n_1 = n_2$, the intermediary integral may be integrated to determine the final integral. If $n_1 \neq n_2$, the two intermediary integrals are solved for p and q and substituted in $p\,dx + q\,dy = dz$. The result is integrated.

Homogeneous Equation with Constant Coefficients. The equation:

$$\frac{\partial^2 z}{\partial x^2} + A_1\frac{\partial^2 z}{\partial x \partial y} + A_2\frac{\partial^2 z}{\partial y^2} = 0 \tag{Eq 72}$$

is equivalent to:

$$\left(\frac{\partial}{\partial x} - m_1\frac{\partial}{\partial y}\right)\left(\frac{\partial}{\partial x} - m_2\frac{\partial}{\partial y}\right)z = 0 \tag{Eq 73}$$

where m_1 and m_2 are roots of the auxiliary equation $X^2 + A_1 X + A_2 = 0$. The general solution of Eq 73 is:

$$z = f_1(y + m_1 z) + f_2(y + m_2 x) \tag{Eq 74}$$

Euler's Equation. The general Euler's equation:

$$a\frac{\partial^2 z}{\partial x^2} + 2b\frac{\partial^2 z}{\partial x \partial y} + c\frac{\partial^2 z}{\partial y^2} = 0 \tag{Eq 75}$$

for which a, b, and c are constants, has the general solutions:

$$z = \phi(x + \lambda_1 y) + \psi(x + \lambda_2 y) \quad \text{if } b^2 \neq ac \tag{Eq 76}$$

$$z = \phi(x + \lambda_1 y) + (\gamma x + \delta y)\psi(x + \lambda_1 y) \quad \text{if } b^2 = ac \tag{Eq 77}$$

where λ_1 and λ_2 are the roots of $a + 2b\lambda + c\lambda^2 = 0$. ϕ and ψ are arbitrary functions, and γ and δ are arbitrary constants. The general solution of $\partial^2 x/\partial x\,\partial y = 0$ is $z = \phi(x) + \psi(y)$; ϕ and ψ are arbitrary functions. The general solution of $\partial^2 u/\partial t^2 = a^2 \partial^2 u/\partial x^2$ is $u = \phi(x + at) + \psi(x - at)$; ϕ and ψ are arbitrary functions.

Separation of Variables

In many important applications, a solution of the form:

$$\Phi = \Phi(x_1, x_2, \ldots, x_n) = \phi_1(x_1)\phi_0(x_2, x_3, \ldots, x_n)$$

(Eq 78)

can be reached by rewriting a partial differential equation in the separated form:

$$F_1\left(x_1, \phi_1, \frac{d\phi_1}{dx_1}, \frac{d^2\phi_1}{dx_1^2}, \ldots\right)$$
$$= F_0\left(x_2, x_3, \ldots, x_n; \phi_0; \frac{\partial\phi_0}{\partial x_2}, \frac{\partial\phi_0}{\partial x_3}, \ldots\right)$$

The unknown functions $\phi_1(x_1)$ and $\phi_0(x_2, x_3, \ldots, x_n)$ must satisfy the differential equations:

$$F_1\left(x_1, \phi_1, \frac{d\phi_1}{dx_1}, \frac{d^2\phi_1}{dx_1^2}, \ldots\right) = C$$

(Eq 79)

$$F_2\left(x_2, x_3, \ldots, x_n; \phi_0; \frac{\partial\phi_0}{\partial x_2}, \frac{\partial\phi_0}{\partial x_3}, \ldots\right) = C$$

(Eq 80)

where C is a constant of integration, or separation constant, to be determined according to boundary or other conditions. Equation 79 is an ordinary differential equation for the unknown function $\phi_1(x_1)$. With Eq 80, the separation process may be repeated.

Integral Calculus*

THE BASIC CONCEPT OF INTEGRATION is the continuous summing of differentials, and the symbol for the process of integration ($\int$) is analogous to the symbol Σ for the summation of discrete elements. For example, for any function u of variable x with the differential of du, then:

$$\int du = u + C$$

$$\int \frac{du}{dx} dx = u + C$$

$$\frac{d}{dx} \int u \, dx = u + C$$

where C is an arbitrary constant for a family of curves (Fig. 1). Each of the curves is equally valid when the function u is determined from the integration of the differential du. Likewise, a constant factor (k) can be moved from the integrand to outside the integral sign, as follows:

$$\int k \, du = k \int du$$

The integral of the algebraic sum of a finite number of functions is the algebraic sum of the integrals of the functions. If u, v, and w are functions of a single variable, for example:

$$\int (du + dv - dw) = \int du + \int dv - \int dw$$

Integration Methods

Table 1 lists some standard elementary integrals. The first step in integration is to simplify the integral, if necessary. If the integrand is a sum of functions, it may be rewritten as the sum of the integrals of the functions. When the integrand is a rational fraction with numerator of degree equal to or greater than that of the denominator, it may be transformed until the remainder is of lower degree than the denominator. The remainder may be broken into partial fractions for further simplification. Substitution of another variable for part of the integrand may simplify integration.

The original or simplified integrand can be compared with the standard elementary forms or with solutions in more elaborate tables of integrals, such as presented later in this article. If the solution does not appear with inspection,

algebraic or trigonometric manipulation may be used to put the integral in a recognizable form for solution with a table. Substitution of another variable may also be tried.

If an integral contains:	Try substituting:
$f(\sqrt{a^2 - u^2})$	$u = a \sin \theta$
$f(\sqrt{a^2 + u^2})$	$u = a \tan \theta$
$f(\sqrt{u^2 - a^2})$	$u = a \sec \theta$
$f(x^{p/q})$	$x = z^n$

where n is the least common denominator of the fractional exponents of x.

If an integral contains:	Try substituting:
$f[x, (a + bx)^{p/q}]$	$a + bx = z^n$

where n represents the least common denominator of the fractional exponents of $a + bx$. For integrals in the form:

$$x^m (a + bx^n)^{p/q}$$

where $q > 0$, $n \neq 0$, and m, n, p, q are integers, try:

$$z^q = a + bx^n \quad \text{if } \frac{m+1}{n} \text{ is zero or an integer}$$

or:

$$z^q x^n = a + bx^n \quad \text{if } \frac{m+1}{n} + \frac{p}{q} \text{ is zero or an integer.}$$

Thus, the fundamental rules of integration are those properties of indefinite integrals that

enable transformation of the integral of a given function to the integrals of other functions.

Table 1 Standard elementary indefinite integrals

$$\int dx = x + C$$

$$\int u^n du = \frac{u^{n+1}}{n+1} + C \quad n \neq -1$$

$$\int \frac{du}{u} = \log_e u + C = \log_e cu$$

$$\int e^u \, du = e^u + C$$

$$\int a^u \, du = \frac{a^u}{\log_e a} + C$$

$$\int \sin u \, du = -\cos u + C$$

$$\int \cos u \, du = \sin u + C$$

$$\int \tan u \, du = \log_e \sec u + C$$

$$\int \cot u \, du = \log_e \sin u + C$$

$$\int \sec u \, du = \log_e (\sec u + \tan u) C$$

$$\int \csc u \, du = \log_e (\csc u - \cot u) C$$

$$\int \sec^2 u \, du = \tan u + C$$

$$\int \csc^2 u \, du = -\cot u + C$$

$$\int \sec u \tan u \, du = \sec u + C$$

$$\int \csc u \cot u \, du = -\csc u + C$$

$$\int \frac{du}{u^2 + a^2} = \frac{1}{a} \arctan \frac{u}{a} + C$$

$$\int \frac{du}{u^2 - a^2} = \frac{1}{2a} \log_e \frac{a - u}{a + u} C$$

$$\int \frac{du}{a^2 - u^2} = \frac{1}{2a} \log_e \frac{a + u}{a - u} C$$

$$\int \frac{du}{\sqrt{u^2 \pm a^2}} = \log_e (u + \sqrt{u^2 \pm a^2}) C$$

$$\int \frac{au}{\sqrt{a^2 - u^2}} = \arcsin \frac{u}{a} + C$$

$$\int \sqrt{a^2 - u^2} du = \frac{u}{2} \sqrt{a^2 - u^2} + \frac{a^2}{2} \arcsin \frac{u}{a} + C$$

$$\int \sqrt{u^2 \pm a^2} du = \frac{u}{2} \sqrt{u^2 \pm a^2} \pm \frac{a^2}{2} \log_e (u + \sqrt{u^2 \pm a^2}) C$$

$$\int \sinh u \, du = \cosh u + C$$

$$\int \cosh u \, du = \sinh u + C$$

$$\int \tanh u \, du = \log_e \cosh u + C$$

$$\int \coth u \, du = \log_e |\sinh u| + C$$

Fig. 1 Function (u) with different integration constants (C_1, C_2, C_3) from integration of the differential du

Integration by Parts. If u and v are functions of a single variable:

$$\int u\,dv = uv - \int v\,du$$

Hence, $u\,dv$ can be integrated if $v\,du$ can be integrated.

Integration by Series. If the integrand can be expressed as a uniformly convergent series of powers of x, within its interval of convergence, and if term-by-term integration of this series also produces a uniformly convergent series, the sum of this series is also the value of the integral.

Integration of Trigonometric Functions. A trigonometric integrand may be reduced to a standard form in several ways.

If an integral contains $f(\sin\theta, \cos\theta)$, substitution of $0 = 2\arctan x$ may be used, so that $x = \tan(\theta/2)$, $\sin\theta = 2x/(1+x^2)$, $\cos\theta = (1-x^2)/(1+x^2)$ and $d\theta = 2dx/(1+x^2)$.

If an integral has the form:

$$\int \sin^{2m+1}\theta \cos^n\theta\,d\theta$$

where m is any integer or zero, the integrand may be rewritten in the form:

$$\sin^{2m}\theta \cos^n\theta \sin\theta = (1-\cos^2\theta)^m \cos^n\theta \sin\theta$$

After multiplication, the integrand becomes a sum of powers of $\cos\theta$ multiplied by $\sin\theta$; it then can be integrated term by term.

For an integral in the form:

$$\int \sin^p\theta \cos^{2r+1}\theta\,d\theta$$

where r is any integer or zero, the integrand can be rewritten as:

$$\sin^p\theta \cos^{2r}\theta \cos\theta = \sin^p\theta(1-\sin^2\theta)^r \cos\theta$$

After multiplication, the integrand becomes a sum of powers of $\sin\theta$ multiplied by $\cos\theta$; it then can be integrated term by term.

For an integral in the form:

$$\int \sin^{2m}\theta \cos^{2n}\theta d\theta$$

where m and n are integers or zero, use the identities:

$$\sin^2\theta = 1/2(1-\cos 2\theta)$$
$$\cos^2\theta = 1/2(1+\cos 2\theta)$$
$$\sin\theta\cos\theta = \frac{\sin 2\theta}{2}$$

For an integral in the form:

$$\int \tan^m\theta \sec^n\theta d\theta \quad \text{or} \quad \int \cot^m\theta \csc^n\theta d\theta$$

the identities $\sec^2\theta = 1 + \tan^2\theta$ or $\csc^2\theta = 1 + \cot^2\theta$ may be tried. The form:

$$\int \sin^m\theta \cos^n\theta d\theta$$

also may be integrated by the use of reduction formulas, which contain an integral of lower degree in $\sin\theta$ or $\cos\theta$. Successive reduction may produce the required integral.

Integration of Rational Fractions. If the quotient of two polynomials:

$$R(x) = \frac{P_n(x)}{P_d(x)}$$

is not a proper fraction, $R(x)$ can be changed, by dividing, to the sum of integrable polynomial and a proper fraction. If the proper fraction cannot be integrated by reference to a table of integrals, resolve it, if possible, into partial fractions. These can be integrated from a table.

Integration of Irrational Functions. Irrational functions are not always integrable in terms of elementary functions. In simplest cases, integrals of irrational functions can be reduced to those of rational functions by means of the substitutions in Table 2.

Elliptic Integrals. Integrals such as:

$$\int R\left(x, \sqrt{ax^3 + bx^2 + cx + \partial}\right)dx$$

and:

$$\int R\left(x, \sqrt{ax^4 + bx^3 + cx^2 + \partial x + e}\right)dx$$

cannot, in general, be expressed in terms of elementary functions. In such cases, they are called elliptic integrals.

Integrals of this type, which are not expressable in terms of elementary functions, can be reduced by transformation to elementary functions and to the following types of integrals:

$$\int \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}$$
$$\int \frac{t^2\,dt}{\sqrt{(1-t^2)(1-k^2t)^2}}$$
$$\int \frac{dt}{(1+ht^2)\sqrt{(1-t^2)(1-k^2t^2)}}$$

where $0 < k < 1$. By substituting $t = \sin\varphi$ ($0 < \varphi < \tfrac{1}{2}\pi$), these transformed integrals can be reduced to the following form of Legendre:

$$\int \frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}} \quad \text{(elliptic integral of the 1st kind)}$$
$$\int \sqrt{1-k^2\sin^2\varphi}\,d\varphi \quad \text{(elliptic integral of the 2nd kind)}$$
$$\int \frac{d\varphi}{(1+h\sin^2\varphi)\sqrt{1-k^2\sin^2\varphi}}$$
(elliptic integral of the 3rd kind)

The corresponding definite integrals with the lower limit of integration equal to zero are denoted as follows:

(I) $$\int_0^\varphi \frac{d\psi}{\sqrt{1-k^2\sin^2\psi}} = F(k,\varphi)$$

(II) $$\int_0^\varphi \sqrt{1-k^2\sin^2\psi}\,d\psi = E(k,\varphi)$$

(III) $$\int_0^\varphi \frac{d\psi}{(1+h\sin^2\psi)\sqrt{1-k^2\sin^2\psi}} = \Pi(h,k,\varphi)$$

where $k < 1$.

These are incomplete elliptic integrals of the first, second, and third kind, respectively. When $\varphi = \tfrac{1}{2}\pi$, (I) and (II) are called complete elliptic integrals and denoted by:

$$K = F\left(k,\frac{\pi}{2}\right) = \int_0^{\pi/2} \frac{d\psi}{\sqrt{1-k^2\sin^2\psi}}$$
$$E = E\left(k,\frac{\pi}{2}\right) = \int_0^{\pi/2} \sqrt{1-k^2\sin^2\psi}\,d\psi$$

Definite Integrals

If $f(x)$ is continuous in the closed interval $[a, b]$ and this interval is divided into n equal parts by the points $a, x_1, x_2,\ldots,x_{n-1}, b$ in such a way that $\Delta x = (b-a)/n$, then the definite integral of $f(x)$ with respect to x between the limits $x = a$ to $x = b$, is expressed as:

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_1^n f(X_i)\Delta x$$
$$= \left[\int f(x)dx\right]_a^b = [F(x)]_a^b = F(b) - F(a)$$

where $F(x)$ is a function and $f(x)$ is the derivative of the function with respect to x. The numbers a and b are called, respectively, the lower and upper limits of integration, and $[a, b]$ is called the range of integration. Geometrically, the definite integral of $f(x)$ with respect to x, between limits $x = a$ to $x = b$, is the area bounded by $f(x)$, the X-axis, and the verticals through the end points of a and b.

Table 2 Substitutions for integration of irrational functions

Form	Substitution
$f[(ax+b)^{p/q}]dx$	let $ax + b = y^q$
$f[(ax+b)^{p/q}(ax+b)^{r/s}]dx$	let $ax + b = y^n$, where n is the least common multiple of q, s
$f[x, \sqrt{x^3 + ax + b}]dx$	let $\sqrt{x^2 + ax + b} = y - x$
$f[x, \sqrt{-x^2 + ax + b}]dx$	let $\sqrt{-x^2 + ax + b} = \sqrt{(\alpha - x)(\beta + x)} = (\alpha - x)y$ or $= (\beta + x)y$
$f[\sin x, \cos x]dx$	let $\tan\dfrac{x}{2} = y$
$f[x, \sqrt{a^2 - x^2}]dx$	let $x = a\sin y$
$f[x, \sqrt{x^2 - a^2}]dx$	let $x = a\sec y$ or $x = a\cosh y$
$f[x, \sqrt{x^2 + a^2}]dx$	let $x = a\tan y$ or $x = a\sinh y$

Figure 2 shows the division of intervals and the area boundaries of a generalized definite integral. General rules of limits are:

$$\int_a^b = -\int_b^a$$

$$\int_a^b + \int_b^c = \int_a^c$$

$$\int_a^c - \int_b^c = \int_a^b$$

$$\int_a^a = 0$$

Fundamental Properties

The integral of an algebraic sum of several functions is equal to the sum of the integrals of these functions:

$$\int_a^b (f(x) + \varphi(x) - \psi(x))dx$$
$$= \int_a^b f(x)dx + \int_a^b \varphi(x)dx - \int_a^b \psi(x)dx$$

A constant factor can be brought out from under the integral sign:

$$\int_a^b cf(x)dx = c \int_a^b f(x)dx$$

Mean Value Theorem. If the function $f(x)$ is continuous in the interval $[a, b]$, then there exists at least one point ξ inside the interval $[a, b]$, such that:

$$\int_a^b f(x)dx = (b - a)f(\xi)$$

The geometric significance of this theorem is shown in Fig. 3; a point ξ exists between a and b that defines the area of $ABCD$ equal to the area of the rectangle $AB'C'D$.

For the integral of a product of two functions $f(x)$ and $\varphi(x)$, where $f(x)$ is continuous and $\varphi(x)$ has a constant sign in the interval $[a, b]$, there exists inside the interval $[a, b]$ at least one number ξ for which:

$$\int_a^b f(x)\varphi(x)dx = f(\xi)\int_a^b \varphi(x)dx$$

Estimation of the Integral. The value of a definite integral is contained between the products of the least and the greatest value of the integrand multiplied by the length of the interval of integration:
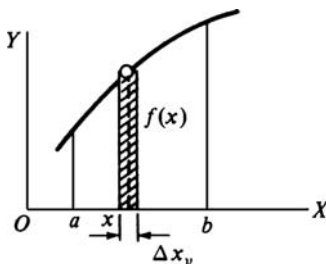
$$m(b - a) \leq \int_a^b f(x)dx \leq M(b - a)$$

where m is the least and M is the greatest value of $f(x)$ in the interval $[a, b]$. The significance of this theorem is illustrated in Fig. 4.

Evaluation of Definite Integrals

The fundamental method of evaluating definite integrals is based on replacing the definite integral by an indefinite integral:

$$\int_a^b f(x)dx = \left[\int f(x)dx\right]_a^b$$

In this case, to evaluate the definite integral, a primitive function of $f(x)$ should be found. Definite and indefinite integrals can be transformed one into another by the following rules.

Substitution. By introducing an auxiliary function $x = \varphi(t)$, where the new variable t is a single-valued function $t = \psi(x)$ of x in the interval $[a, b]$, the integral can be transformed into the form:

$$\int_a^b f(x)dx = \int_{\psi(a)}^{\psi(b)} f(\varphi(t))\varphi'(t)dt$$

Using this formula, the inverse substitution in evaluating the indefinite integral can be avoided. For example:

$$\int_0^a \sqrt{a^2 - x^2}dx = \int_{\arcsin 0}^{\arcsin 1} a^2\sqrt{1 - \sin^2 t}d\sin t$$
$$= a^2 \int_0^{\pi/2} \cos^2 t \, dt$$
$$= a^2 \int_0^{\pi/2} \frac{1}{2}(1 + \cos 2t)dt$$
$$= \frac{a^2}{2}[t]_0^{\pi/2} + \frac{a^2}{4}\int_0^{\pi} \cos z \, dz$$
$$= \frac{\pi a^2}{4} + \frac{a^2}{4}[\sin z]_0^{\pi} = \frac{\pi a^2}{4}$$

Integration by Parts. The expression $f(x)dx$ can be written in an arbitrary way, in the form $u \, dv$; du can be found by differentiation and v by integration. The definite integral this way can be transformed into the form:
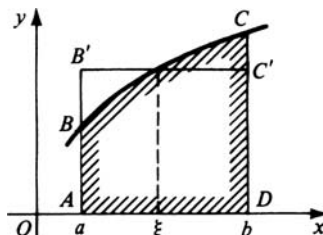
$$\int_a^b f(x)dx = \int_a^b u \, dv = [uv]_a^b - \int_a^b v \, du$$

For example:

$$\int_0^1 xe^x dx = [xe^x]_0^1 - \int_0^1 e^x dx = e - (e - 1) = 1$$

Integration by Expansion into a Series. If $f(x)$ is represented in the interval of integration by a uniformly convergent series of functions:

$$f(x) = \varphi_1(x) + \varphi_2(x) + \cdots + \varphi_n(x) + \cdots$$

then the following relationship holds:

$$\int f(x)dx = \int \varphi_1(x)dx + \int \varphi_2(x)dx + \cdots$$
$$+ \int \varphi_n(x)dx + \cdots$$

Thus, the definite integral can be represented as a convergent series of numbers:

$$\int_a^b f(x)dx = \int_a^b \varphi_1(x)dx + \int_a^b \varphi_2(x)dx + \cdots$$
$$+ \int_a^b \varphi_n(x)dx + \cdots$$

If the functions $\varphi_1(x)$ can be easily integrated, as with an expansion of $f(x)$ into a power series uniformly convergent in the interval $[a, b]$, then the integral $\int_a^b f(x)dx$ can be evaluated with an arbitrary accuracy.

Improper Integrals. If one limit is infinite, then:

$$\int_a^{\infty} f(x)dx = \lim_{b \to \infty} \int_a^b f(x)dx$$

The integral exists, or converges, if there exist numbers $k > 1$ and M independent of x in a relationship $x^k |f(x)| < M$ for arbitrarily large values of x. If $x|f(x)| > m$, an arbitrary positive number, for sufficiently large values of x, the interval diverges. For example, the integral:

$$\int_0^{\infty} \frac{xdx}{(x + x^2)^{1/2}}$$

exists because, for $k = 2$ and $M = 1$:

$$x^2 \left| \frac{x}{(x + x^2)^{1/2}} \right| = \left(\frac{x^2}{x + x^2} \right)^{1/2} < 1$$

regardless of how large x becomes. If the integrand is infinite at the upper limit, then:

Fig. 2 Area interval of summation of a definite integral

Fig. 3 Mean value theorem
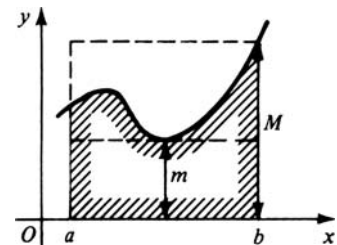
Fig. 4 Method of estimation

$$\int_a^b f(x)dx = \lim_{\varepsilon \to 0} \int_a^{b-\varepsilon} f(x)dx, \text{ with } 0 < \varepsilon < (b-a)$$

The integral exists if numbers $k < 1$ and M independent of x exist in the relationship $(b-x)^k \, |f(x)| < M$ for $a \le x < b$. If there is a number $k \ge 1$ and a number m such that $(b-x)^k \, |f(x)| > m$ for $a \le x < b$, the integral diverges. For example, the integral:

$$\int_0^1 \frac{dx}{1-x}$$

diverges, because for $k = 1$ and $m = \frac{1}{2}$

$$\frac{1-x}{1-x} = 1 > \frac{1}{2}$$

The tests are analogous if the integrand is infinite at the lower limit. If the integrand is infinite at an intermediate point, the point is used to divide the interval into two subintervals, and the aforementioned tests are applied.

Multiple Integrals. Let $f(x, y)$ be defined in the region R of the xy plane. Divide R into subregions ΔR_1, ΔR_2, $\ldots$, ΔR_n of areas ΔA_1, ΔA_2, $\ldots$, ΔA_n. Let (ξ_i, η_i) be any point in ΔR_i. If the sum:

$$\sum_{i=1}^n f(\xi_i, \eta_i)\Delta A_i$$

has a limit as $n > \infty$ and the maximum diameter of the subregions ΔR_i approaches 0, then:

$$\int_R f(x,y)dA = \lim_{n\to\infty} \sum_{i=1}^n f(\xi_i, \eta_i)\Delta A_i$$

The double integral is evaluated by two successive single integrations. It is first evaluated with respect to y holding x constant, between variable limits of integration, and then with respect to x between constant limits, as shown in Fig. 5. If $f(x, y)$ is continuous, the order of integration can be reversed:

$$\int_R f(x,y)dA = \int_a^b \int_{y_1(x)}^{y_2(x)} f(x,y)dy\ dx$$
$$= \int_c^d \int_{x_1(y)}^{x_2(y)} f(x,y)dx\ dy$$

In polar coordinates:

$$\int_R F(r,\theta)dA = \int_\alpha^\beta \int_{r_1(\theta)}^{r_2(\theta)} F(r,\theta)r\ dr\ d\theta$$
$$= \int_k^l \int_{\theta_1(r)}^{\theta_2(r)} F(r,\theta)r\ d\theta\ dr$$

Similarly, triple integrals are evaluated by three single integrations. In rectangular coordinates:

$$\int_R f(x,y,z)dV = \iiint f(x,y,z)dx\ dy\ dz$$

In spherical coordinates:

$$\int_R F(r,\theta,\phi)dV = \iiint F(r,\theta,\phi)r^2 \sin\theta dr\ d\theta\ d\phi$$

In cylindrical coordinates:

$$\int_R G(\rho,\phi,z)dv = \iiint G(\rho,\phi,z)\rho\ d\rho\ d\phi\ dz$$

Integrals Depending on a Parameter. The definite integral:

$$\int_a^b f(x,y)dx = F(y)$$

is a function of the variable y called a parameter.

The function $F(y)$ often is not an elementary function. The integral may be an ordinary integral, an improper integral, or an integral of a discontinuous function $f(x, y)$. For example:

$$\Gamma(y) = \int_0^\infty x^{y-1}e^{-x}dx$$

with the integral convergent for $y > 0$. This is the Gamma function or Euler's integral of the second kind.

Differentiation Under the Integral Sign. If the function:

$$\int_a^b f(x,y)dx = F(y)$$

is defined in the interval $c \le y \le e$ and the function $f(x, y)$ is continuous in the rectangle $a \le x \le b$, $c \le y \le e$ with a continuous partial derivative $\frac{\partial f}{\partial y}$ in this domain, then, for every y in the interval $[c, e]$, there is:

$$\frac{d}{dy}\int_a^b f(x,y)dx = \int_a^b \frac{\partial f(x,y)}{\partial y}dx$$

For example, in an arbitrary interval, for $y > 0$:

$$\frac{d}{dy}\int_0^1 \arctan\frac{x}{y}dx = \int_0^1 \frac{\partial}{\partial y}\left(\arctan\frac{x}{y}\right)dx =$$
$$-\int_0^1 \frac{x}{x^2+y^2}dx = \frac{1}{2}\ln\frac{y^2}{1+y^2}$$

Expressed another way:

$$\int_0^1 \arctan\frac{x}{y}dx = \arctan\frac{1}{y} + \frac{1}{2}y\ \ln\frac{y^2}{1+y^2}$$
$$\frac{d}{dy}\left(\arctan\frac{1}{y} + \frac{1}{2}y\ \ln\frac{1}{1+y^2}\right) = \frac{1}{2}\ \ln\frac{y^2}{1+y^2}$$

Integration Under the Integral Sign. If the function:

$$\int_a^b f(x,y)dx = F(y)$$

is defined in the interval $[c, e]$ and the function $f(x, y)$ is continuous in the rectangle $a \le x \le b$, $c \le y \le e$, then:

$$\int_c^e \left(\int_a^b f(x,y)dx\right)dy = \int_a^b \left(\int_c^e f(x,y)dy\right)dx$$

Line, Surface, and Volume Integrals

Line Integrals. The functions $P(x, y)$ and $Q(x, y)$ are continuous at all points of a continuous curve C joining the points A and B in the xy plane, as shown in Fig. 6. The curve C is divided into n arbitrary parts Δs_v by the points (x_v, y_v), and (ξ_v, η_v) is an arbitrary point on Δs_v. The increments Δx_v and Δy_v are projections of ΔS_v on the x- and the y-axes. The line integral is:

Fig. 5 Limits for double integration of example in text
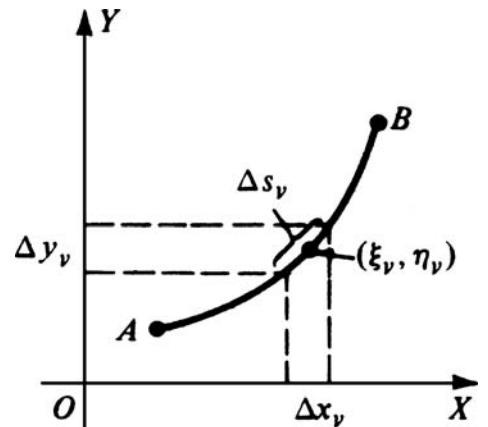
Fig. 6 Elements of a curve for a line integral

$$\int_A^B [P(x,y)dx + Q(x,y)dy] =$$

$$\lim_{\substack{n \to \infty \\ \max \Delta x_\nu, \Delta y_\nu \to 0}} \sum_{\nu-1}^u [P(\xi_\nu, \eta_\nu)\Delta x_\nu + Q(\xi_\nu, \eta_\nu)\Delta y_\nu]$$

If the equation of the curve C is $y = f(x)$, $x = \phi(y)$, or the parametric equations $x = x(t)$, $y = y(t)$, the line integral may be evaluated as a definite integral in variable x, y, or t, respectively. A line integral in xyz:

$$\int_A^B [P(x,y,z)dx + Q(x,y,z)dy + R(x,y,z)dz]$$

is similarly defined. The work done by a constant force F, acting on a particle moving a distance s along a straight line inclined at angle θ to the force, is $W = Fs \cos \theta$. If the path is defined as a curve C and the force is variable, the differential of work is $dW = F \cos \theta \, ds$, where ds is the differential of path. Then:

$$W = \int dW = \int_C F \cos \theta \, ds = \int_C (X \, dx + Y \, dy)$$

where X and Y are the x and y components of F, as shown in Fig. 7.

Surface Areas by Integration. If a plane curve is rotated about an axis in its plane, the area of the surface of revolution is given by:

$$S = 2\pi \int_a^b y \, ds$$

where y is the distance from axis to curve, and ds is the differential of length of arc.

For example, the surface area generated by rotating the hypocycloid $x^{2/3} + y^{2/3} = a^{2/3}$ about the x-axis is found by computing the area from $x = 0$ to $x = a$ and multiplying by 2:

$$y = (a^{2/3} - x^{2/3})^{3/2} \quad y' = -\frac{y^{1/3}}{x^{1/3}}$$

$$1 + y'^2 = 1 + \frac{y^{2/3}}{x^{2/3}} = \frac{a^{2/3}}{x^{2/3}}$$

$$ds = (1 + y'^2)^{1/2} \quad dx = \frac{a^{1/3}}{x^{1/3}} dx$$

$$S = 4\pi \int_0^a (a^{2/3} - x^{2/3})^{3/2} \left(\frac{a^{1/3}}{x^{1/3}}\right) dx$$

To integrate, let $z^2 = a^{2/3} - x^{2/3}$. Then $2z \, dz = -2/3x^{-1/3} \, dx$ and $dx = -3zx^{1/3} \, dz$. When $x = 0$, $z = a^{1/3}$; when $x = a$, $z = 0$:
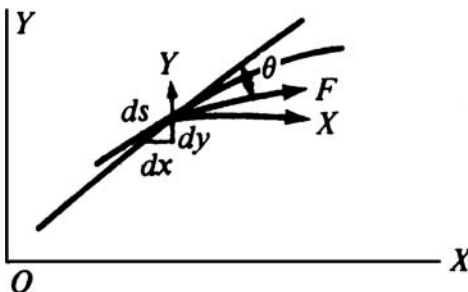
Fig. 7 Components of a force acting on a body moving along curve C

$$S = 4\pi a^{1/3} \int_{a^{1/3}}^0 z^3 x^{-1/3}(-3zx^{1/3} dz)$$

$$= 12\pi a^{1/3} \int_0^{a^{1/3}} z^4 dz$$

$$= 12\pi a^{1/3} [1/5z^5]_0^{a^{1/3}}$$

$$= \frac{12\pi a^2}{5}$$

The area of any curved surface in rectangular coordinates is given by:

$$A = \iint_S \left[1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2\right]^{1/2} dy \, dx$$

$$= \iint_S \left[1 + \left(\frac{\partial y}{\partial x}\right)^2 + \left(\frac{\partial y}{\partial z}\right)^2\right]^{1/2} dz \, dx$$

$$= \iint_S \left[1 + \left(\frac{\partial x}{\partial y}\right)^2 + \left(\frac{\partial x}{\partial z}\right)^2\right]^{1/2} dz \, dy$$

For example, the area of the surface of the sphere $x^2 + y^2 + z^2 = r^2$ is formed by computing the area in the first octant, then multiplying by 8:

$$\frac{\partial z}{\partial z} = -\frac{x}{z}$$

$$\frac{\partial z}{\partial y} = -\frac{y}{z}$$

$$1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 = 1 + \frac{x^2}{z^2} + \frac{y^2}{z^2}$$

$$= \frac{z^2 + x^2 + y^2}{z^2} = \frac{r^2}{r^2 - x^2 - y^2}$$

Integrate first with respect to y, holding x constant, to find the area of an increment dx wide and extending from the xz plane to the xy plane. Then integrate with respect to x, to add the increments from $x = 0$ to $x = r$:

$$A = 8 \int_0^r \int_0^{\sqrt{r^2 - x^2}} \left[\frac{r^2}{r^2 - x^2 - y^2}\right]^{1/2} dy \, dx$$

$$= 8r \int_0^r \left[\arcsin \frac{y}{\sqrt{r^2 - x^2}}\right]_0^{\sqrt{r^2 - x^2}} dx$$

$$= 8r \int_0^r \frac{\pi}{2} dx = 4\pi r[x]_0^r = 4\pi r^2$$

Volumes by Integration. When a plane curve is rotated about an axis in its plane, the volume of the solid generated is:

$$V = \pi \int_a^b y^2 \, dx$$

where y is the distance from axis to curve, and dx is the differential length along the axis. The integration is equivalent to adding circular cross sections of thickness dx.

A volume of revolution also is given by:

$$V = 2\pi \int_{r_1}^r rh \, dr$$

where r is the distance from axis to curve, and h is the length of element of volume normal

to r. The integration is equivalent to adding cylinders of thickness dr.

In polar coordinates, a volume of revolution is given by:

$$V = 2\pi \int_\alpha^\beta \int_c^d r^2 \sin \theta \, dr \, d\theta$$

The volume of a solid in rectangular coordinates generally is given by:

$$V = \int_a^b \int_m^n f(x,y)dy \, dx = \int_c^d \int_p^q f(x,y)dx \, dy$$

where m and n denote functions of x, and p and q denote functions of y.

In cylindrical coordinates, volume is given by:

$$V = \int_\alpha^\beta \int_a^b f(r,\theta)r \, dr \, d\theta = \int_c^d \int_\gamma^\delta f(r,\theta)r \, d\theta \, dr$$

where a and b denote functions of r, and γ and δ denote functions of θ.

Volume can also be calculated by triple integration:

$$V = \iiint_V dV = \int_a^b \int_m^n \int_u^v dz \, dy \, dx$$

where u and v denote functions of y, and m and n denote functions of x.

Green's Theorem. If $P(x,y)$ and $Q(x,y)$ are continuous functions with continuous partial derivatives $\partial P/\partial y$ and $\partial Q/\partial z$ in a simply connected region R bounded by a simple closed curve C, then:

$$\iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dx \, dy = \int_C (P \, dx + Q \, dy)$$

A simply connected region is defined as one in which any closed curve in the region can be reduced to a point without passing outside the region.

Stokes' Theorem. If $P(x, y, z)$, $Q(x, y, z)$, $R(x,y,z)$ are continuous functions having continuous first partial derivatives, and if S is a region that is bounded by a simple closed curve C and part of a surface $z = f(x, y)$, continuous with continuous first partial derivatives, then:

$$\iint_S \left[\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)dy \, dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)dz \, dx\right.$$

$$\left. + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dx \, dy\right] = \int_C (P \, dx + Q \, dy + R \, dz)$$

The signs are determined by a viewpoint in the direction of the normal resulting in the integration around C taken in the positive direction.

Applications of Integration

Typical applications of single integration are given in this section. Applications of double and triple integration are summarized in Tables 3 and 4.

Table 3 Applications of double integrals

Notation	Formula	In Cartesian coordinates	In polar coordinates
Area of a plane figure S	$\int_s dS$	$\iint dydx$	$\iint \rho\, d\rho d\varphi$
Area of a surface Σ	$\int_s \dfrac{dS}{\cos\gamma}$	$\iint \sqrt{1+\left(\dfrac{\partial z}{\partial x}\right)^2+\left(\dfrac{\partial z}{\partial y}\right)^2}\,dydx$	$\iint \sqrt{\rho^2+\rho^2\left(\dfrac{\partial z}{\partial \rho}\right)^2+\left(\dfrac{\partial z}{\partial \varphi}\right)^2}\,d\rho d\varphi$
Volume of a cylinder V	$\int_s z dS$	$\iint z\, dydx$	$\iint z\rho\, d\rho d\varphi$
Moment of inertia(a) I_x	$\int_s y^2 dS$	$\iint y^2 dydx$	$\iint \rho^3 \sin^2\varphi d\rho d\varphi$
Moment of inertia(b) I_o	$\int_s \rho^3 dS$	$\iint (x^2+y^2)dydx$	$\iint \rho^3 d\rho d\varphi$
Mass of a plane figure(c) M	$\int_s \delta dS$	$\iint \delta\, dydx$	$\iint \delta\rho\, d\rho d\varphi$
x_c	$\dfrac{\int_s x\, dS}{S}$	$\dfrac{\iint x\, dydx}{\iint dydx}$	$\dfrac{\iint \rho^2\cos\varphi\, d\rho d\varphi}{\iint \rho\, d\rho d\varphi}$
Coordinates of the center of gravity(d) y_c	$\dfrac{\int_s y\, dS}{S}$	$\dfrac{\iint y\, dydx}{\iint dydx}$	$\dfrac{\iint \rho^2\sin\varphi\, d\rho d\varphi}{\iint \rho\, d\rho d\varphi}$

(a) Of a plane figure with respect to the x-axis. (b) Of a plane figure with respect to a pole O. (c) With surface density δ given as a function of a point. (d) Of a plane homogeneous figure

Table 4 Applications of triple integrals

Notation	Formula	In Cartesian coordinates	In cylindrical coordinates	In spherical coordinates
Volume of a solid V	$\int_v dV$	$\iiint dzdydx$	$\iiint \rho dzd\rho d\varphi$	$\iiint r^3\sin\theta\, drd\theta\, d\varphi$
Moment of inertia(a) I_z	$\int_v \rho^2 dV$	$\iiint (x^2+y^2)dzdydx$	$\iiint \rho^3 dzd\rho d\varphi$	$\iiint r^4\sin^3\theta\, drd\theta\, d\varphi$
Mass of a solid(b) M	$\int_v \delta\, dV$	$\iiint \delta\, dzdydx$	$\iiint \delta\rho\, dzd\rho d\varphi$	$\iiint \delta r^3\sin\theta\, drd\theta\, d\varphi$
x_C	$\dfrac{\int_v y\, dV}{V}$	$\dfrac{\iiint y\, dzdydx}{\iiint dzdydx}$		
Coordinates of the center of gravity(c) y_C	$\dfrac{\int_v y\, dV}{V}$	$\dfrac{\iiint y\, dzdydx}{\iiint dzdydx}$		
z_C	$\dfrac{\int_v z\, dV}{V}$	$\dfrac{\iiint z\, dzdydx}{\iiint dzdydx}$		

(a) Of a solid with respect to the z-axis. (b) With density δ given as a function of a point. (c) Of a homogeneous solid

Fig. 8 Integration of surfaces formed by revolution of a curve. (a) Revolution about the x-axis. (b) Revolution about the y-axis

Length of Arc. For an arc of a plane curve $y = f(x)$, the length s from the point (a, b) to the point (c, d) is expressed as:

$$s = \int_a^c \sqrt{1+\left(\frac{dy}{dx}\right)^2}\; dx = \int_b^d \sqrt{1+\left(\frac{dx}{dy}\right)^2}\; dy$$

In polar coordinates, with $r = f(\theta)$, the length of the arc from the point (r_1, θ_1) to the point (r_2, θ_2) becomes:

$$s = \int_{\theta_1}^{\theta_2} \sqrt{r^2+\left(\frac{dr}{d\theta}\right)^2}\; d\theta = \int_{r_1}^{r_2} \sqrt{1+r^2\left(\frac{d\theta}{dr}\right)^2}\; dr$$

In three dimensions, with the curve represented by the equations $y = f_1(x)$ and $z = f_2(x)$, the length of arc from $x_1 = a$ to $x_2 = b$ is:

$$s = \int_a^b \sqrt{1+\left(\frac{dy}{dx}\right)^2+\left(\frac{dz}{dx}\right)^2}\, dx$$

Areas of Plane Figures. For the curve $y = f(x)$, the area bounded by $f(x)$, the x-axis, and the ordinates $x = a$ and $x = b$ is given by:

$$A = \int_a^b y\; dx$$

The area above the x-axis is represented as positive, that below the x-axis is negative. In polar coordinates, the area under a curve is:

$$A = \frac{1}{2}\int_\alpha^\beta \rho^2 d\theta$$

In rectangular coordinates, the area under a curve is also determined by the double integral:

$$A = \int_A \int dA = \iint dxdy$$

In polar coordinates, the double integral becomes:

$$A = \iint \rho\, d\rho\, d\theta$$

The area bounded by a closed curve C also can be determined by the line integral:

$$A = \frac{1}{2}\int_C (-y\, dx + x\, dy)$$

Area of a Surface. The formula for area of a surface formed by revolution of a curve $y = f(x)$ about the x-axis, as shown in Fig. 8(a), is determined by:
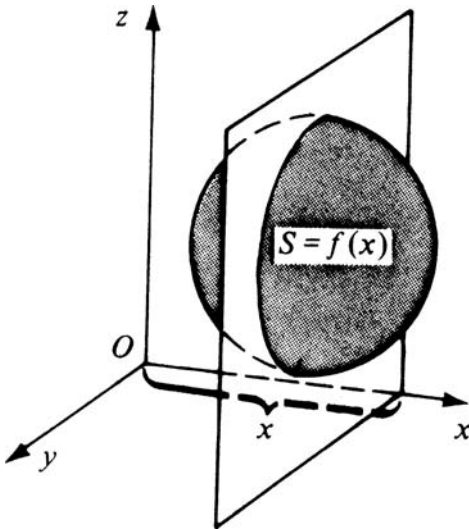
Fig. 9 Function for the surface area (*S*) perpendicular to the *x*-axis
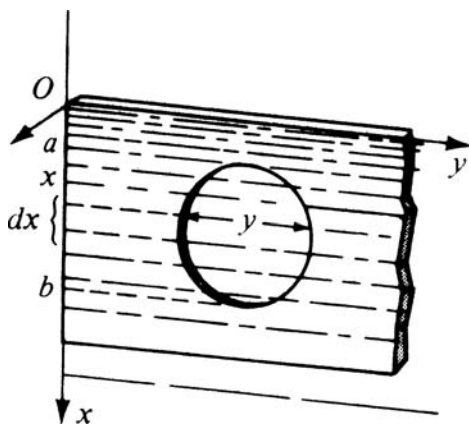
Fig. 10 Integration example for calculation of fluid pressure (see text)
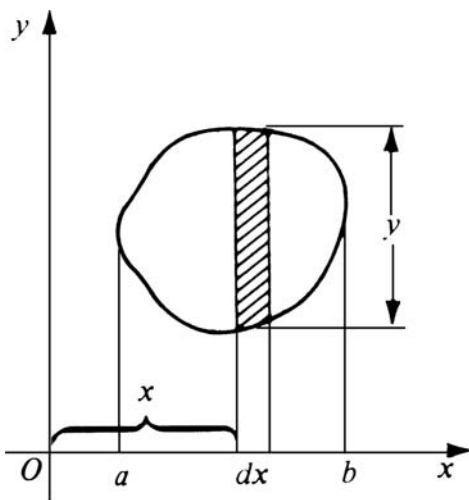
Fig. 11 Plane figure illustrating integration of moment of inertia
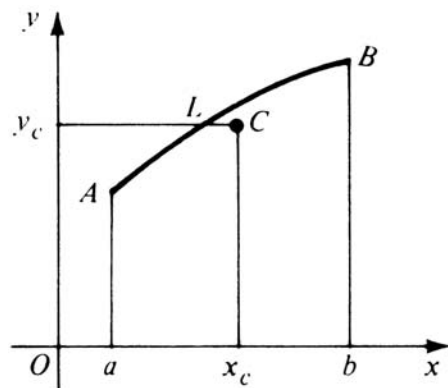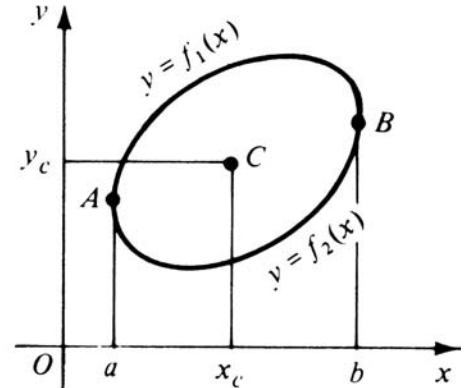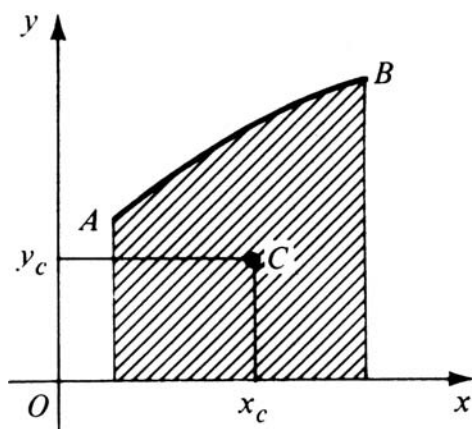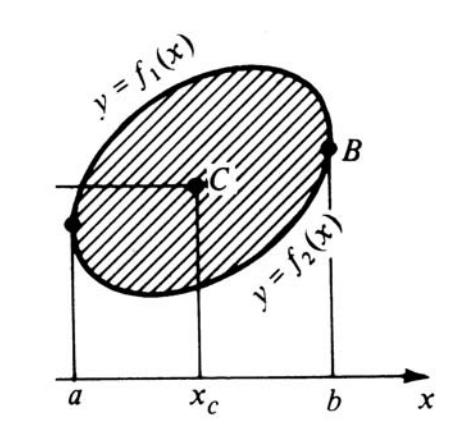
Fig. 12 Integration examples for center of gravity. (a) Curve. (b) Closed curve. (c) Curvilinear trapezoid. (d) Arbitrary plane figure

$$S = 2\pi \int_a^b y \; dl = 2\pi \int_a^b y \sqrt{1 + \left(\frac{dy}{dx}\right)^2} \, dx$$

The area of a surface formed by revolution of a curve $x = g(y)$ about the *y*-axis, as shown in Fig. 8(b), is determined by:

$$S = 2\pi \int_\alpha^\beta x \; dl = 2\pi \int_\alpha^\beta x \sqrt{\left(\frac{dx}{dy}\right)^2 + 1} \, dy$$

Volume. The formula for the interior volume of a surface of revolution, about the *x*-axis, as shown in Fig. 8(a), is determined by:

$$V = \pi \int_a^b y^2 dx$$

The volume of a surface of revolution about the *y*-axis, as shown in Fig. 8(b) is given by:

$$V = \pi \int_\alpha^\beta x^2 dy$$

For the volume of a solid, when the area of its plane section perpendicular to the *x*-axis is given as a function $S = f(x)$, as shown in Fig. 9, the formula is:

$$V = \int_a^b f(x) dx$$

Work. The work represented by the force *F* acting from the point $x = a$ to $x = b$ of the *x*-axis, when the force is a function $F = f(x)$, is given by the formula:

$$A = \int_a^b F \; dx$$

Fluid Pressure. For a fluid with specific gravity γ, the pressure on one side of a vertical plate immersed in the fluid can be determined by the following formula, when the distance *x* of the points of the plate from the level of the fluid varies from *a* to *b*, and where *y* is the length of a horizontal section of the plate ($y = f(x)$), as shown in Fig. 10:

$$P = \int_a^b \gamma xy \; dx$$

Center of Pressure. The depth $\bar{y}$ of the center of pressure against a surface vertical to the liquid surface between the depths a and b is:

$$\bar{y} = \frac{\int_{y=a}^{y=b} \gamma y^2 dA}{\int_{y=a}^{y=b} \gamma y dA}$$

Moment of Inertia. The moment of inertia of an arc of a homogeneous curve $y = f(x)$, with $a \leq x \leq b$, with respect to the y-axis is given by the formula:

$$I_y = \delta \int_a^b x^2 dl = \delta \int_a^b x^2 \sqrt{1 + (y')^2} dx$$

where δ is the linear density of the arc.

The moment of inertia of a plane figure, as shown in Fig. 11, with respect to the y-axis is given by:

$$I_y = \int_a^b x^2 y \; dx$$

where y is the length of a section parallel to the y-axis, and δ is the surface density of the figure.

Center of Gravity. The center of gravity C of an arc of a homogeneous plane curve $y = f(x)$, with $a \leq x \leq b$, as shown in Fig. 12(a), has the coordinates:

$$x_C = \frac{\int_a^b x\sqrt{1 + y'^2} dx}{L}$$

$$y_C = \frac{\int_a^b y\sqrt{1 + y'^2} dx}{L}$$

where L is the length of the arc. The center of gravity of a closed curve, as shown in Fig. 12 (b), has coordinates:

$$x_C = \frac{\int_a^b x\left(\sqrt{1 + |y_1'|^2} + \sqrt{1 + |y_2'|^2}\right) dx}{L}$$

$$y_C = \frac{\int_a^b \left(y_1\sqrt{1 + |y_1'|^2} + y_2\sqrt{1 + |y_2'|^2}\right) dx}{L}$$

where $y_1 = f_1(x)$ and $y_2 = f_2(x)$ are the equations of the upper and lower part of the bounding curve, and L is the length of the entire curve.

For a homogeneous curvilinear trapezoid, as shown in Fig. 12(c) the center of gravity C has the coordinates:

$$x_C = \frac{\int_a^b xy \; dx}{S}$$

$$y_C = \frac{\frac{1}{2}\int_a^b y^2 \; dx}{S}$$

where S is the area of the trapezoid, and $y = f(x)$ is the equation of the curve AB. For an arbitrary plane figure, as shown in Fig. 12(d), the center of gravity has the coordinates:

$$x_C = \frac{\int_a^b x(y_1 - y_2) dx}{S}$$

$$y_C = \frac{\frac{1}{2}\int_a^b (y_1^2 - y_2^2) dx}{S}$$

where $y_1 = f_1(x)$ and $y_2 = f_2(x)$ are the equations of the upper and lower portions of the bounding curve, and S is the area.

Laplace Transformations

THE METHODS OF LAPLACE AND FOURIER TRANSFORMATION and the Heaviside operational calculus are essentially aspects of the same method. This method simplifies solutions of such problems as ordinary differential equations with constant coefficients, linear differential equations with variable coefficients, integral equations of the convolution, or Faltung, type, and partial differential equations. Several Laplace transforms are listed in Table 1.

Fundamental Transformation Rules

Direct Laplace Transformation. The Laplace transform $\mathcal{L}(F)$ of a function $F(t)$ may be expressed as:

$$\mathcal{L}(F) = f(s) = \int_0^\infty e^{-st} F(t) dt \qquad \text{(Eq 1)}$$

with t a real variable; s a complex variable; $F(t)$ a real function of t, which equals zero for $t < 0$; $f(s)$ a function of s; and e the base of the natural logarithms. If the function $F(t)$ is known and the integral:

$$f(s) = \int_0^\infty e^{-st} F(t) dt \qquad \text{(Eq 2)}$$

can be computed, the function $f(s)$ may be determined and the direct Laplace transform:

$$f(s) = \mathcal{L}F(t) \qquad \text{(Eq 3)}$$

obtained. If the function $f(s)$ is known, the integral:

$$F(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} e^{st} f(s) ds \qquad \text{(Eq 4)}$$

must be used to determine the function $F(t)$. If this integral can be evaluated, the inverse transform:

$$F(t) = \mathcal{L}^{-1} f(s) \qquad \text{(Eq 5)}$$

may be obtained. The computation of several direct transforms may be evaluated by the integration of Eq 2.

Inverse Transformation. The inverse transform of a function $f(s)$ is computed by the use of the equation:

$$F(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} f(s) e^{st} ds \qquad \text{(Eq 6)}$$

The line integral for $F(t)$ can be evaluated by transforming it into a closed contour (Fig. 1). Defining the closed contour Γ as the straight line parallel to the axis of imaginaries at a distance c to the right of the axis and the large semicircle s_0 with center at (c, 0), then:

$$\oint_\Gamma e^{st} f(s) ds = \int_{c+jR}^{c-jR} e^{st} f(s) ds + \int_{s_0} e^{st} f(s) ds \qquad \text{(Eq 7)}$$

The evaluation of the contour integral along Γ is simplified by designating $\phi(s)$ as an integrable function of the complex variable s:

$$\lim_{|s|\to\infty} |\phi(s)| = 0 \qquad \text{(Eq 8)}$$

Then:

$$\lim_{R\to\infty} \left| \int_{s_0} e^{st} \phi(s) ds \right| = 0 \quad t > 0 \quad \text{Re } s \overset{\leq}{=} 0 \qquad \text{(Eq 9)}$$

The function $\phi(s) = f(s)$ frequently has such properties that the integral around the large semicircle defined by Eq 7 vanishes as $R \to \infty$. Thus:

$$F(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} f(s) e^{st} ds$$
$$= \lim_{R\to\infty} \frac{1}{2\pi j} \oint_s e^{st} f(s) ds \qquad \text{(Eq 10)}$$

and:

$$\oint_\Gamma e^{st} f(s) ds = 2\pi j \sum \text{Res } e^{st} f(s) \text{ inside } \Gamma \qquad \text{(Eq 11)}$$

Hence, from Eq 10:

$$F(t) = \sum \text{Res } e^{st} f(s) \text{ inside } \Gamma \qquad \text{(Eq 12)}$$

if R is sufficiently large to include all singularities.

Transformation of nth Derivative. If $\mathcal{L}[F(t)] = f(s)$, then:

$$\mathcal{L}\left[\frac{d^n F(t)}{dt^n} \right] = s^n f(s) - \sum_{k=0}^{n-1} F^{(k)}(0+) \cdot s^{n-1-k} \qquad \text{(Eq 13)}$$

where $F^{(2)}(0+)$ means $\frac{d^2 F(t)}{dt^2}$ evaluated for $t \to 0$, and $F^{(0)}(0+)$ means $F(0+)$, and $n = 1, 2, 3, \cdots n$.

Transformation of nth Integral. If $\mathcal{L}[F(t)] = f(s)$, then:

$$\mathcal{L}\left[\int\int\cdots\int F(t) dt \right] = s^{-n} f(s) + \sum_{k=-1}^{-n} F^{(k)}(0+) \cdot s^{-n-1-k} \qquad \text{(Eq 14)}$$

where $F^{(-2)}(0+)$ means $\int\int F(t) \, dt \, dt$ evaluated for $t \to 0$, and $n = 1, 2, 3, \cdots n$.

Inverse Transformation of Product. If:

$$\mathcal{L}^{-1}[f_1(s)] = F_1(t) \qquad \mathcal{L}^{-1}[f_2(s)] = F_2(t) \qquad \text{(Eq 15)}$$

then:

$$\mathcal{L}^{-1}[f_1(s) \cdot f_2(s)] = \int_0^t F_1(t-\lambda) \cdot F_2(\lambda) d\lambda \qquad \text{(Eq 16)}$$
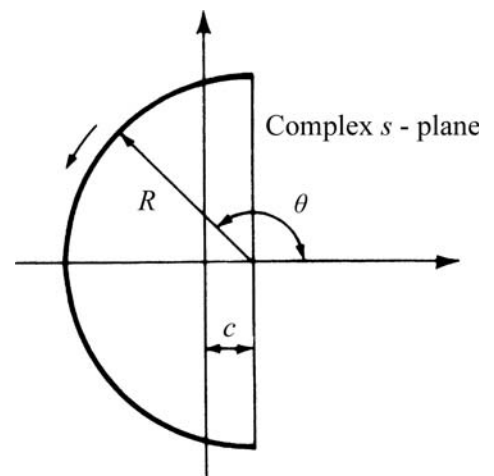
Fig. 1 Line integral transformed into a closed contour

Linear Transformations $\mathcal{L}$ and $\mathcal{L}^{-1}$. Designating k_1, k_2 as real constants, then:

$$\mathcal{L}[k_1 F_1(t) + k_2 F_2(t)] = k_1 \mathcal{L}[F_1(t)] + k_2 \mathcal{L}[F_2(t)] \tag{Eq 17}$$

and:

$$\mathcal{L}^{-1}[k_1 f_1(s) + k_2 f_2(s)] = k_1 \mathcal{L}^{-1}[f_1(s)] + k_2 \mathcal{L}^{-1}[f_2(s)] \tag{Eq 18}$$

Theorems

Some general theorems concerning operations on transforms are given as follows.

Theorem 1. The Laplace transform of a constant k is that constant divided by s:

$$\mathcal{L}(k) = \frac{k}{s} \tag{Eq 19}$$

From the fundamental definition of the direct Laplace transform:

$$\mathcal{L}(k) = \int_0^\infty e^{-st} k \, dt = k \left[-\frac{e^{-st}}{s} \right]_0^\infty = \frac{k}{s} \tag{Eq 20}$$

and the integral vanishes at the upper limit.

Theorem 2. The equation:

$$\mathcal{L}k\phi(t) = k\mathcal{L}\phi(t) \tag{Eq 21}$$

where k is a constant, may be proved in the following manner:

$$\mathcal{L}k\phi(t) = \int_0^\infty e^{-st} k\phi(t)dt = k \int_0^\infty e^{-st} \phi(t)dt$$
$$= k\mathcal{L}\phi(t) \tag{Eq 22}$$

Theorem 3. If F is a continuous differentiable function and if F and dF/dt can be transformed, then:

$$\mathcal{L}\frac{dF}{dt} = s\mathcal{L}F(t) - F(0) \tag{Eq 23}$$

which is a useful equation for solving differential equations with constant coefficients. The equation is developed by:

$$\mathcal{L}\frac{dF}{dt} = \int_0^\infty e^{-st}\frac{dF}{dt}dt = Fe^{-st}\Big|_0^\infty + s\int_0^\infty e^{-st}F\,dt$$
$$= s\mathcal{L}F - F(0) \tag{Eq 24}$$

with integration performed by parts.

Theorem 4. If F is continuous with derivatives of orders $1, 2, \cdots, n$ which can be transformed, then:

$$\mathcal{L}\frac{d^n F}{dt^n} = s^n \mathcal{L}F - \sum_{k=0}^{n-1} F^{(k)}(0)s^{n-k-1} \tag{Eq 25}$$

where:

$$F^{(k)}(0) = \frac{d^k F}{dt^k}$$

is evaluated at $t = 0$. As an extension of theorem 3:

$$\mathcal{L}\frac{d^n F}{dt^n} = s\mathcal{L}\frac{d^{n-1}F}{dt^{n-1}} - \left(\frac{d^{n-1}F}{dt^{n-1}}\right)_{t=0} \tag{Eq 26}$$

By repeated applications, Eq 25 is obtained. Further, if:

$$F_r = \frac{d^r F}{dt^r} \tag{Eq 27}$$

is evaluated at $t = 0$, the transforms of the first four derivatives are:

$$\mathcal{L}\frac{dF}{dt} = s\mathcal{L}F - F_0 \tag{Eq 28}$$

$$\mathcal{L}\frac{d^2 F}{dt^2} = s^2 \mathcal{L}F - sF_0 - F_1 \tag{Eq 29}$$

$$\mathcal{L}\frac{d^3 F}{dt^3} = s^3 \mathcal{L}F - s^2 F_0 - sF_1 - F_2 \tag{Eq 30}$$

or:

$$\mathcal{L}\frac{d^4 F}{dt^4} = s^4 \mathcal{L}F - s^3 F_0 - s^2 F_1 - sF_2 - F_3 \tag{Eq 31}$$

which are useful in transforming differential equations.

Theorem 5. The Faltung or convolution theorem is developed as follows. It is sometimes called the superposition theorem. Let:

$$\mathcal{L}F_1(t) = f_1(s) \tag{Eq 32}$$

and:

$$\mathcal{L}F_2(t) = f_2(s) \tag{Eq 33}$$

Then, according to the theorem:

$$\mathcal{L}\int_0^t F_1(y)F_2(t-y)dy = \mathcal{L}\int_0^t F_2(y)F_1(t-y)dy$$
$$= f_1 f_2 \tag{Eq 34}$$

In developing this theorem, let:

$$\mathcal{L}F_3(t) = f_1(s)f_2(s) \tag{Eq 35}$$

then:

$$F_3(t) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} f_1(s)f_2(s)e^{st}ds \tag{Eq 36}$$

However:

$$f_2(s) = \int_0^\infty e^{-sy}F_2(y)dy \tag{Eq 37}$$

therefore:

$$F_3(t) = \frac{1}{2\pi j}\int_0^\infty F_2(y)dy \int_{c-j\infty}^{c+j\infty} f_1(s)e^{s(t-y)}ds \tag{Eq 38}$$

if the order of integration is reversed. However:

$$\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} f_1(s)e^{2(t-y)}ds = F_1(t-y) \tag{Eq 39}$$

hence:

$$F_3(t) = \int_0^\infty F_2(y)F_1(t-y)dy \tag{Eq 40}$$

Now, $F_1(t) = 0$, if t is less than 0, and:

$$F_1(t-y) = 0, \qquad \text{for } y > t \tag{Eq 41}$$

Because the infinite limit of integration may be replaced by the limit t, Eq 40 may be written as:

$$F_3(t) = \int_0^t F_2(y)F_1(t-y)dy \tag{Eq 42}$$

$$F_3(t) = \int_0^t F_1(y)F_2(t-y)dy \tag{Eq 43}$$

Applications of Laplace Transforms

Real Indefinite Integration. If $\mathcal{L}\,F(t) = f(s)$ and $\int F(t)dt$ can be transformed, then:

$$\mathcal{L}\int F(t)dt = \frac{1}{s}f(s) + \frac{1}{s}\int^0 F(t)dt \tag{Eq 44}$$

By the definition of the Laplace transform:

$$\mathcal{L}\int F(t)dt = \int_0^\infty e^{-st}\left[\int F(t)dt\right]dt \tag{Eq 45}$$

Integrating by parts produces:

$$\mathcal{L}\int F(t)dt = -\frac{1}{s}e^{-st}\int F(t)dt\Big|_0^\infty + \frac{1}{s}\int_0^\infty e^{-st}F(t)dt \tag{Eq 46}$$

Evaluating the limits and using $\mathcal{L}\,F(t) = f(s)$ gives:

$$\mathcal{L}\int F(t)dt = \frac{1}{s}f(s) + \frac{1}{s}\int^0 F(t)dt \tag{Eq 47}$$

for which $\int^0 F(t)dt$ indicates the initial value of the integral.

Real Definite Integration. If $\mathcal{L}\,F(t) = f(s)$ and the expression $\int_0^t F(t')dt'$ can be transformed, then:

$$\mathcal{L}\int_0^t F(t')dt' = \frac{1}{s}f(s) \tag{Eq 48}$$

Integrating by parts yields:

$$\mathcal{L}\int_0^t F(t')dt' = -\frac{1}{s}e^{-st}\int_0^t F(t')dt'\Big|_0^\infty + \frac{1}{s}f(s) \tag{Eq 49}$$

The first term vanishes at the upper limit because it contains an exponential function and at the lower limit because of the definite

integral. Hence, only the second term remains, as expressed in Eq 48. Following the same steps produces:

$$\mathcal{L} \int_0^t \int_0^t F(t'')dt''dt' = \frac{1}{s^2}f(s) \qquad (Eq\ 50)$$

where $\mathcal{L}\ F(t) = f(s)$.

Complex Differentiation. If $\mathcal{L}\ F(t) = f(s)$, then:

$$\mathcal{L}\ tF(t) = -\frac{df(s)}{ds} \qquad (Eq\ 51)$$

According to the definition of $f(s)$ as indicated by $\mathcal{L}F(t) = f(s)$:

$$f(s) = \int_0^t e^{-st}F(t)dt \qquad (Eq\ 52)$$

By differentiating both sides of Eq 52 with respect to s:

$$\frac{df(s)}{ds} = \frac{d}{ds}\int_0^t e^{-st}F(t)dt \qquad (Eq\ 53)$$

Because the integration involves t only, the order of integration and differentiation may be interchanged for Eq 53. Thus:

$$\frac{df(s)}{ds} = -\int_0^t e^{-st}tF(t)dt = -\mathcal{L}tF(t) \qquad (Eq\ 54)$$

Differentiation with a Second Independent Variable. For a function of two independent variables $F = F\ (x,\ t)$ if $\mathcal{L}F(x,t) = f(x,s)$, then:

$$\mathcal{L}\frac{\partial F(x,t)}{\partial x} = \frac{\partial f(x,s)}{\partial x} \qquad (Eq\ 55)$$

From the definition of Laplace transformation:

$$\mathcal{L}\frac{\partial F(x,t)}{\partial x} = \int_0^\infty e^{-st}\frac{\partial F(x,t)}{\partial x}dt \qquad (Eq\ 56)$$

Because x is not a variable of integration, the order of differentiation and integration may be interchanged. Thus:

$$\begin{aligned}\mathcal{L}\frac{\partial F(x,t)}{\partial x} &= \frac{\partial}{\partial x}\int_0^t e^{-st}F(x,t)dt \\ &= \frac{\partial f(x,s)}{\partial x}\end{aligned} \qquad (Eq\ 57)$$

Integration with a Second Independent Variable. For a function of two independent variables $F = F(x,\ t)$ if $\mathcal{L}F(x,t) = f(x,s)$, then:

$$\mathcal{L}\int_{x_1}^{x_2} F(x',t)dx' = \int_{x_1}^{x_2} f(x',s)dx' \qquad (Eq\ 58)$$

The method for developing Eq 58 is identical to the development of Eq 55 by using Eq 56 and 57.

Periodic Functions. If $F(t)$ is a periodic function with fundamental period T, $F(t + T) = F(t)$

where $t > 0$, and if $F(t)$ is sectionally continuous over a period $0 < t < T$, then its direct transformation is expressed as:

$$\begin{aligned}\mathcal{L}F(t) &= \int_0^\infty e^{-st}F(t)dt \\ &= \sum_{n=0}^\infty \int_{nT}^{(n+1)T} e^{-st}F(t)dt\end{aligned} \qquad (Eq\ 59)$$

Further, if

$$u = t - nT \qquad (Eq\ 60)$$

and, as a result of the periodicity of the function $F(t)$:

$$F(u + nT) = F(u) \qquad (Eq\ 61)$$

then Eq 59 may be written:

$$\mathcal{L}\ F(t) = \sum_{n=0}^\infty e^{-nTs}\int_0^T e^{-su}F(u)du \qquad (Eq\ 62)$$

Additionally, using:

$$\sum_{n=0}^{n=x} e^{-nTs} = \frac{1}{1 - e^{-Ts}} \qquad (Eq\ 63)$$

Equation 62 may be written in the form:

$$\mathcal{L}\ F(t) = \frac{1}{1 - e^{-Ts}}\int_0^T e^{-st}F(t)dt \qquad (Eq\ 64)$$

This equation can be used to obtain the transform of the function represented in Fig. 2. Substituting the expression:

$$\begin{aligned}\int_0^T F(t)e^{-st}dt &= \int_0^a e^{-st}dt - \int_a^{2a} e^{-st}dt \\ &= \frac{1 - e^{-sa} - e^{-sa} + e^{-2sa}}{s} \\ &= \frac{1 - 2e^{-sa} + e^{-2sa}}{s} = \frac{(1 - e^{-sa})^2}{s}\end{aligned} \qquad (Eq\ 65)$$

into Eq 64 yields

$$\begin{aligned}\mathcal{L}\ F(t) &= \frac{(1 - e^{-sa})^2}{s(1 - e^{-2sa})} = \frac{1 - e^{-sa}}{s(1 + e^{-sa})} \\ &= \frac{e^{sa/2} - e^{-sa/2}}{s(e^{sa/2} + e^{-sa/2})} = \frac{\sinh(sa/2)}{s\ \cosh\ (sa/2)} \\ &= \frac{\tanh(as/2)}{s}\end{aligned} \qquad (Eq\ 66)$$

Operational Method. The operational method for solving differential equations is essentially the same as that known as Heaviside's operational calculus. The relationship between a function $f(s)$ and another function $h(t)$ may be expressed in the form:

$$f(s) = \int_0^\infty e^{-st}h(t)dt \quad \text{Re } s > 0 \qquad (Eq\ 67)$$

where s is a complex number with its real part greater than zero, and $h(t)$ is a function for which the infinite integral of Eq 67 converges and is a function that satisfies the condition h
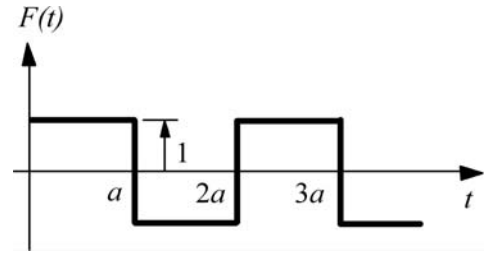
Fig. 2 Example function in transform of periodic function (see text)

$(t) = 0$ for $t < 0$. The functional relationship between $f(s)$ and $h(t)$ is written as:

$$f(s) = \mathcal{L}\ h(t) \qquad (Eq\ 68)$$

with the symbol $\mathcal{L}$ denoting the "Laplace transform of." The relationship is also written in the form:

$$h(t) = \mathcal{L}^{-1}f(s) \qquad (Eq\ 69)$$

with $h(t)$ the inverse Laplace transform of $f(s)$. For example, the functional relationship:

$$y(s) = \int_0^\infty e^{-st}x(t)dt \qquad (Eq\ 70)$$

is written symbolically as:

$$y(s) = \mathcal{L}\ x(t) \qquad (Eq\ 71)$$

To determine $\mathcal{L}(dx/dt)$ in terms of $y(s)$, integrate:

$$\mathcal{L}\left(\frac{dx}{dt}\right) = \int_0^\infty e^{-st}\left(\frac{dx}{dt}\right)dt \qquad (Eq\ 72)$$

by parts to obtain:

$$\int_0^\infty e^{-st}\left(\frac{dx}{dt}\right)dt = e^{-st}x\Big|_0^\infty + s\int_0^\infty e^{-st}x\ dt \qquad (Eq\ 73)$$

Using the assumptions that:

$$\lim_{t\to\infty}(e^{-st}x) = 0 \text{ and } \int_0^\infty e^{-st}x\ dt$$

exists for s greater than some fixed positive number, then Eq 73 becomes:

$$\mathcal{L}\left(\frac{dx}{dt}\right) = -x_0 + sy \qquad (Eq\ 74)$$

where $x_0 = x(0)$. To compute the Laplace transform of $d^2\ x/dt^2$, let:

$$u = \frac{dx}{dt} \qquad (Eq\ 75)$$

From Eq 74:

$$\begin{aligned}\mathcal{L}\left(\frac{d^2x}{dt^2}\right) = \mathcal{L}\left(\frac{du}{dt}\right) &= -u_0 + s\ \mathcal{L}\left(\frac{dx}{dt}\right) \\ &= -x_1 - sx_0 + s^2y\end{aligned} \qquad (Eq\ 76)$$

where x_1 is the value of dx/dt at $t = 0$. Continuing the process produces:

$$\mathcal{L}\frac{d^3x}{dt^3} = s^3y - s^2x_0 - sx_1 - x_2 \qquad \text{(Eq 77)}$$

$$\mathcal{L}\frac{d^4x}{dt^4} = s^4y - s^3x_0 - s^2x_1 - sx_2 - x_3 \qquad \text{(Eq 78)}$$

and:

$$\mathcal{L}\frac{d^nx}{dt^n} = s^ny - (s^{n-1}x_0 + s^{n-2}x_1 + s^{n-3}x_2 + \cdots + x_{n-1}) \qquad \text{(Eq 79)}$$

where:

$$x_n = \frac{d^nx}{dt^n} \qquad \text{(Eq 80)}$$

is evaluated at $x = 0$.

Table 1 Laplace transforms

Entries $f(s)$ give values of the Laplace transform $\mathcal{L}[F(t)] \equiv f(s)$ of the function $F(t)$. The entries in column $F(t)$ define $F(t)$ for $t \geq 0$; a, b, c, k are constants; $f(s) = \int_0^\infty e^{-st}F(t)dt$

$\mathbf{f(s)} \equiv \mathcal{L}[\mathbf{F(t)}]$	$\mathbf{F(t)}$	$\mathbf{f(s)} \equiv \mathcal{L}[\mathbf{F(t)}]$	$\mathbf{F(t)}$
$1/s$	1	$1/[s(s^2 + a^2)]$	$(1 - \cos at)/a^2$
$1/s^2$	t	$1/[s^2(s^2 + a^2)]$	$(at - \sin at)/a^3$
$1/s^n$	$t^{n-1}/(n-1)!$	$1/(s^2 + a^2)^2$	$(\sin at - at\cos at)/[2a^3]$
$1/(s - a)$	e^{at}	$s/(s^2 + a^2)^2$	$(t \sin at)/[2a]$
$1/(s - a)^2$	te^{at}	$s^2/(s^2 + a^2)^2$	$(\sin at + at\cos at)/[2a]$
$1/(s - a)^n$ $\quad n = 1, 2, \ldots$	$t^{n-1}e^{at}/(n-1)!$	$s/[(s^2 + a^2)(s^2 + b^2)]$ $\quad a^2 \neq b^2$	$(\cos at - \cos bt)/(b^2 - a^2)$
$a/(s^2 + a^2)$	$\sin at$	$1/[(s - a)^2 + b^2]$	$(e^{at}\sin bt)/b$
$s/(s^2 + a^2)$	$\cos at$	$(s - a)/[(s - a)^2 + b^2]$	$e^{at}\cos bt$
$a/(s^2 - a^2)$	$\sinh at$	$3a^2/(s^3 + a^3)$	$\begin{cases} e^{-at} - e^{at/2} \cdot A \\ \quad \text{where } A = \cos(\sqrt{3}at/2) - \\ \quad \sqrt{3}\sin(\sqrt{3}at/2) \end{cases}$
$s/(s^2 - a^2)$	$\cosh at$		
$(s\sin b + a\cos b)/(s^2 + a^2)$	$\sin(at + b)$	$3as/(s^3 + a^3)$	$\begin{cases} e^{-at} + e^{at/2} \cdot A \\ \quad \text{where } A = \cos(\sqrt{3}at/2) + \\ \quad \sqrt{3}\sin(\sqrt{3}at/2) \end{cases}$
$(s\cos b - a\sin b)/(s^2 + a^2)$	$\cos(at + b)$		
$2as/(s^2 + a^2)^2$	$t\sin at$	$4a^3/(s^4 + 4a^4)$	$\sin at\cosh at - \cos at\sinh at$
$(s^2 - a^2)/(s^2 + a^2)^2$	$t\cos at$	$s/(s^4 + 4a^4)$	$[1/(2a^2)]\sin at\sinh at$
$1/\sqrt{s}$	$1/\sqrt{\pi t}$	$1/(s^4 - a^4)$	$(\sinh at - \sin at)/[2a^3]$
$s^{-3/2}$	$2\sqrt{t/\pi}$	$s/(s^4 - a^4)$	$(\cosh at - \cos at)/[2a^2]$
$s^{-[n + (1/2)]}$ $\quad n = 1, 2, \ldots$	$2^n t^{n-(1/2)}/[1 \cdot 3 \cdot 5 \ldots (2n-1)\sqrt{\pi}]$	$8a^3 s^2/(s^2 + a^2)^3$	$(1 + a^2 t^2)\sin at - at\cos at$
$\Gamma(k)/s^k$	$k > 0 \; t^{k-1}$	$(1/s)[(s-1)/s]^n$(b)	$L_n(t) = (e^t/n!)[d^n(t^n e^{-t})/dt^n]$
$\Gamma(k)/(s - a)^k$	$k > 0 \; t^{k-1}e^{at}$	$\sqrt{s - a} - \sqrt{s - b}$	$(e^{bt} - e^{at})/(2\sqrt{\pi t^3})$
$1/[(s - a)(s - b)]$(a)	$(e^{at} - e^{bt})/(a - b)$	$s/(s - a)^{3/2}$	$e^{at}(1 + 2at)/\sqrt{\pi t}$
$s/[(s - a)(s - b)]$(a)	$(ae^{at} - be^{bt})/(a - b)$	$1/(\sqrt{s} + a)$(c)	$(1/\sqrt{\pi t}) - ae^{a^2 t}\text{erfc}(a\sqrt{t})$
$1/[(s - a)(a - b)(s - c)]$(a)	$\begin{cases} [(b-c)e^{at} + (c-a)e^{bt} + (a-b)e^{ct}]/A \\ \quad \text{where } A = -(a-b)(b-c)(c-a) \end{cases}$	$\sqrt{s}/(s + a^2)$	$(1/\sqrt{\pi t}) - (2ae^{-a^2 t}/\sqrt{\pi})\int_0^{a\sqrt{t}} e^{\lambda^2}d\lambda$

(continued)

Table 1   (continued)

$\mathbf{f(s)} \equiv \mathcal{L}[\mathbf{F(t)}]$		$F(t)$	$\mathbf{f(s)} \equiv \mathcal{L}[\mathbf{F(t)}]$		$F(t)$
$\sqrt{s}/(s-a^2)$(c)		$(1/\sqrt{\pi t}) + ae^{a^2 t}\mathrm{erf}(a\sqrt{t})$	$s^{-3/2}\,e^{-k/s}$		$(1/\sqrt{\pi k})\sin(2\sqrt{kt})$
$1/[\sqrt{s}(s+a^2)]$		$[2/(a\sqrt{\pi})]e^{-a^2 t}\int_0^{a\sqrt{t}} e^{\lambda^2}\,d\lambda$	$s^{-3/2}\,e^{k/s}$		$(1/\sqrt{\pi k})\sinh(2\sqrt{kt})$
$1/[\sqrt{s}(s-a^2)]$		$(e^{a^2 t}/a)\,\mathrm{erf}(a\sqrt{t})$	$s^{-\mu}\,e^{-k/s}$	$\mu > 0$	$(t/k)^{(\mu-1)/2}J_{\mu-1}(2\sqrt{kt})$
$1/[\sqrt{s}(\sqrt{s}+a)]$		$e^{a^2 t}\,\mathrm{erfc}(a\sqrt{t})$	$s^{-\mu}\,e^{k/s}$	$\mu > 0$	$(t/k)^{(\mu-1)/2}I_{\mu-J}(2\sqrt{kt})$
$(b^2-a^2)/[(s-a^2)$ $(\sqrt{s}+b)]$		$e^{a^2 t}[b - a\,\mathrm{erf}(a\sqrt{t})] - be^{b^2 t}\mathrm{erfc}(b\sqrt{t})$	$e^{-k\sqrt{s}}$	$k > 0$	$[k/2\sqrt{\pi t^3}]\exp[-k^2/(4t)]$
$(b^2-a^2)/[\sqrt{s}(s-a^2)$ $(\sqrt{s}+b)]$		$e^{a^2 t}[(b/a)\,\mathrm{erf}(a\sqrt{t}) - 1] + e^{b^2 t}\mathrm{erfc}(b\sqrt{t})$	$(1/\sqrt{s})e^{-k\sqrt{s}}$	$k \geqq 0$	$(1/\sqrt{\pi t})\exp[-k^2/(4t)]$
$1/[(s+a)\sqrt{s+b}]$		$(1/\sqrt{b-a})c^{-at}\,\mathrm{erf}(\sqrt{b-a}\sqrt{t})$	$s^{-1}e^{-k\sqrt{s}}$	$k \geqq 0$	$\mathrm{erfc}[k/(2\sqrt{t})]$
$1/[\sqrt{s+a}\sqrt{s+b}]$(d)		$e^{-(a+b)t/2}I_0[(a-b)t/2]$	$s^{-3/2}e^{-k\sqrt{s}}$	$k \geqq 0$	$\begin{cases} 2\sqrt{t/\pi}\exp[-k^2/(4t)]- \\ k\,\mathrm{erfc}[k/(2\sqrt{t})] \end{cases}$
$1/[(s+a)^{1/2}(s+b)^{3/2}]$		$\begin{cases} te^{-(a+b)t/2}\cdot B \\ \text{where } B \equiv I_0[(a-b)t/2]+ \\ I_1[(a-b)t/2] \end{cases}$	$ae^{-k\sqrt{s}}/[s(a+\sqrt{s})]$	$k \geqq 0$	$\begin{cases} -e^{ak}e^{a^2 t}\,\mathrm{erfc}\,A + \mathrm{erfc}[k/(2\sqrt{t})] \\ \text{where } A = a\sqrt{t}+k/(2\sqrt{t}) \end{cases}$
$\Gamma(k)/[(s+a)^k(s+b)^k]$	$k > 0$	$\begin{cases} \sqrt{\pi}C[t/(a-b)]^{k-(1/2)}e^{-(a+b)t/2} \\ \text{where } C \equiv I_{k-1/2}[(a-b)t/2] \end{cases}$	$e^{-k\sqrt{s}}/[\sqrt{s}(a+\sqrt{s})]$	$k \geqq 0$	$e^{ak}e^{a^2 t}\,\mathrm{erfc}\,A$
$(\sqrt{s+2a}-\sqrt{s})/$ $(\sqrt{s+2a}+\sqrt{s})$		$(1/t)\,e^{-at}\,I_1(at)$	$e^{-k\sqrt{s(s+a)}}/\sqrt{s(s+a)}$		$\begin{cases} 0 & \text{when } 0 < t < k \\ e^{-at/2}I_0\left(\tfrac{1}{2}a\sqrt{t^2-k^2}\right) & \text{when } t > k \end{cases}$
$(1-s)^n/s^{[n+(1/2)]}$(e)		$n!H_{2n+1}(\sqrt{t})/[(2n)!\sqrt{\pi t}]$	$e^{-k\sqrt{s^2+a^2}}/\sqrt{s^2+a^2}$		$\begin{cases} 0 & \text{when } 0 < t < k \\ J_0(a\sqrt{t^2-k^2}) & \text{when } t > k \end{cases}$
$(1-s)^n/s^{[n+(3/2)]}$		$-n!H_{2n+1}(\sqrt{t})/[\sqrt{\pi}(2n+1)!]$	$e^{-k\sqrt{s^2-a^2}}/\sqrt{s^2-a^2}$		$\begin{cases} 0 & \text{when } 0 < t < k \\ I_0(a\sqrt{t^2-k^2}) & \text{when } t > k \end{cases}$
$(\sqrt{s+2a}/\sqrt{s})-1$		$ae^{-at}[I_1(at)+I_0(at)]$	$e^{-k(\sqrt{s^2+a^2}-s)}/\sqrt{s^2+a^2}$	$k\geqq0$	$J_0(a\sqrt{t^2+2kt})$
$(\sqrt{s}+a+\sqrt{s})^{-2v}/$ $(\sqrt{s}\sqrt{s+a})$	$v > -1$	$(1/a^v)\,e^{-at/2}\,I_v(at/2)$	$e^{-ks}-e^{-k\sqrt{s^2+a^2}}$		$\begin{cases} 0 & \text{when } 0 < t < k \\ (ak/\sqrt{t^2-k^2})J_1(a\sqrt{t^2-k^2}) & \text{when } t > k \end{cases}$
$(a-b)^k/(\sqrt{s+a}$ $+\sqrt{s+b})^{2k}$	$k > 0$	$(k/t)\,e^{-(a+b)t/2}I_k[(a-b)t/2]$	$e^{-k\sqrt{s^2-a^2}}-e^{-ks}$		$\begin{cases} 0 & \text{when } 0 < t < k \\ (ak/\sqrt{t^2-k^2})/I_1(a\sqrt{t^2-k^2}) & \text{when } t > k \end{cases}$
$(s^2+a^2)^{-1/2}$		$J_0(at)$	$[a^v\exp(-k\sqrt{s^2+a^2})]/$ $[\sqrt{s^2+a^2}(\sqrt{s^2+a^2}+s)^v]$	$v > -1$	$\begin{cases} 0 & \text{when } 0 < t < k \\ [(t-k)/(t+k)]^{v/2}J_v(a\sqrt{t^2-k^2}) & \text{t} > \text{k} \end{cases}$
$(s^2+a^2)^{-k}$	$k > 0$	$[\sqrt{\pi}/\Gamma(k)][t/2a]^{k-(1/2)}J_{k-1/2}(at)$	$(1/s)\log s$(f)		$\Gamma'(1) - \log t,\ \Gamma'(1) = -0.5772157$
$(s^2-a^2)^{-k}$	$k > 0$	$[\sqrt{\pi}/\Gamma(k)][t/2a]^{k-(1/2)}I_{k-1/2}(at)$	$s^{-k}\log s$(f)	$k > 0$	$t^{k-1}[\Gamma'(k) - \Gamma(k)\log t][\Gamma(k)]^{-2}$
$(\sqrt{s^2+a^2}-s)^v/$ $\sqrt{s^2+a^2}$	$v > -1$	$a^v J_v(at)$	$(s-a)^{-1}\log s$ (f)	$a > 0$	$e^{at}[\log a - \mathrm{Ei}(-at)]$
$(\sqrt{s^2+a^2}-s)^k$	$k > 0$	$(ka^k/t)J_k(at)$	$(s^2+1)^{-1}\log s$(f)		$\cos t\,\mathrm{Si}\,t - \sin t\,\mathrm{Ci}\,t \equiv H(t)$
$(s-\sqrt{s^2-a^2})^v/$ $\sqrt{s^2-a^2}$	$v > -1$	$a^v I_v(at)$	$s(s^2+1)^{-1}\log s$(f)		$-\sin t\,\mathrm{Si}\,t - \cos t\,\mathrm{Ci}\,t$
$s^{-1}e^{-k/s}$		$J_0(2\sqrt{kt})$	$s^{-1}\log(1+ks)$(f)	$k > 0$	$-\mathrm{Ei}(-t/k)$
$s^{-1/2}e^{-k/s}$		$(1/\sqrt{\pi t})\cos(2\sqrt{kt})$	$\log[(s-a)/(s-b)]$		$t^{-1}(e^{bt}-e^{at})$
$s^{-1/2}e^{k/s}$		$(1/\sqrt{\pi t})\cosh(2\sqrt{kt})$	$\log[(s+a)/(s-a)]$		$2t^{-1}\sinh at\qquad \mathrm{Re}\,s > \mathrm{Re}\,a$
			$s^{-1}\log(1+k^2 s^2)$		$-2\,\mathrm{Ci}\,(t/k)$

(continued)

Table 1 (continued)

$\mathbf{f(s)} \equiv \mathcal{L}[\mathbf{F(t)}]$		$F(t)$
$s^{-1} \log (s^2 + a^2)$	$a > 0$	$2 \log a - 2 \operatorname{Ci}(at)$
$s^{-2} \log (s^2 + a^2)$	$a > 0$	$2a^{-1}[at \log a + \sin at - at \operatorname{Ci} at]$
$\log[(s^2 + a^2) s^2]$		$2t^{-1}(1 - \cos at)$
$\log[(s^2 - a^2)/s^2]$		$2t^{-1}(1 - \cosh at)$
$\cot^{-1}[(s - b)/a]$		$t^{-1} e^{bt} \sin at$
$\cot^{-1}(s/k)$		$t^{-1} \sin kt$
$(1/s) \cot^{-1}(s/k)$		$\operatorname{Si} kt$
$e^{k^2 s^2} \operatorname{erfc}(ks)$	$k > 0$	$[1/(k\sqrt{\pi})] \exp[-t^2/(4k^2)]$
$s^{-1} e^{k^2 s^2} \operatorname{erfc}(ks)$	$k > 0$	$\operatorname{erf}[t/(2k)]$
$e^{ks} \operatorname{erfc} \sqrt{ks}$	$k > 0$	$\sqrt{k}/[\pi\sqrt{t}(t + k)]$
$s^{-1/2} \operatorname{erfc} \sqrt{ks}$		$\begin{cases} 0 & \text{when } 0 < t < k \\ (\pi t)^{-1/2} & \text{when } t > k \end{cases}$
$s^{-1/2} e^{k^2/s} \operatorname{erfc}(k/\sqrt{s})$	$k > 0$	$1/\sqrt{\pi(t + k)}$
$s^{-1/2} \operatorname{erfc} \sqrt{ks}$		$(1\sqrt{\pi t}) e^{-2k\sqrt{t}}$
$\operatorname{erf}(k/\sqrt{s})$		$[1/(\pi t)] \sin (2k\sqrt{t})$
$K_0(ks)$(g)		$\begin{cases} 0 & \text{when } 0 < t < k \\ (t^2 - k^2)^{-1/2} & \text{when } t > k \end{cases}$
$K_0(k\sqrt{s})$		$[1/(2t)] \exp[-k^2/(4t)]$
$s^{-1} e^{ks} K_1(ks)$		$(1/k)\sqrt{t(t + 2k)}$
$s^{-1/2} K_1(k\sqrt{s})$		$(1/k) \exp[-k^2/(4t)]$
$s^{-1/2} e^{k/s} K_0(k/s)$		$(2/\sqrt{\pi t}) K_0(2\sqrt{2kt})$
$\pi e^{-ks} I_0(ks)$		$\begin{cases} [t(2k - t)]^{-1/2} & \text{when } 0 < t < 2k \\ 0 & \text{when } t > 2k \end{cases}$
$e^{-ks} I_1(ks)$		$\begin{cases} (k - t)/[\pi k\sqrt{t(2k - t)}] & \text{when } 0 < t < 2k \\ 0 & \text{when } t > 2k \end{cases}$
$-e^{as} \operatorname{Ei}(-as)$		$(t + a)^{-1} \quad a > 0$
$a^{-1} + se^{as} \operatorname{Ei}(-as)$		$(t + a)^{-2} \quad a > 0$
$[(\pi/2) - \operatorname{Si} s] \cos s$		$(t^2 + 1)^{-1}$
$\quad + \operatorname{Ci} s \sin s$	$s > 0$	

Step functions

s^{-1}		$U(t) = \begin{cases} 0 & \text{when } t < 0 \\ 1 & \text{when } t \geqq 0 \end{cases}$

Heaviside unit function

$s^{-1} e^{-ks}$		$U(t - k) = \begin{cases} 0 & \text{when } t < k \\ 1 & \text{when } t \geqq k \end{cases}$
e^{-as}	$a \geqq 0$	$\delta(t - a)$

$\mathbf{f(s)} \equiv \mathcal{L}[\mathbf{F(t)}]$		$F(t)$
$s^k e^{-as}$	$a \geqq 0$	$\delta^{(k)}(t - a) \quad k = 1, 2, \ldots$
$s^{-2} e^{-ks}$		$S_k(t) = \begin{cases} 0 & \text{when } t \leq k \\ t - k & \text{when } t \geqq k \end{cases}$
$2s^{-3} e^{-ks}$		$\begin{cases} 0 & \text{when } t < k \\ (t - k)^2 & \text{when } t \geqq k \end{cases}$
$\Gamma(\mu) \cdot s^{-\mu} e^{-ks}$	$\mu > 0$	$\begin{cases} 0 & \text{when } t < k \\ (t - k)^{\mu - 1} & \text{when } t \geqq k \end{cases}$
$s^{-1}(1 - e^{-ks})$		$\begin{cases} 1 & \text{when } 0 < t < k \\ 0 & \text{when } t > k \end{cases}$
$s^{-1}(1 - e^{ks})^{-1} = (2s)^{-1}$ $\left(1 + \coth \dfrac{1}{2}ks\right)$(h)		$\begin{cases} 1 + [t/k] = n \\ \quad \text{when } (n - 1)k < t < nk, \\ \qquad n = 1, 2, \ldots \end{cases}$
$s^{-1} \tanh ks$		$\begin{cases} M(2k, t) = (-1)^{n-1} \\ \quad \text{when } 2k(n - 1) < t < 2kn, \\ \qquad n = 1, 2, \ldots \end{cases}$

Square-wave function

$s^{-1}(1 + e^{-ks})^{-1}$		$\begin{cases} [M(k, t) + 1]/2 = [1 - (-1)^n]/2 \\ \quad \text{when } (n - 1)k < t < nk, \\ \qquad n = 1, 2, \ldots \quad t > 0 \end{cases}$
$(s \sinh ks)^{-1}$		$\begin{cases} F(t) = 2(n - 1) \\ \quad \text{when } (2n - 3)k < t < (2n - 1)k \\ \qquad n = 1, 2, \ldots t > 0 \end{cases}$
$s^{-1} \coth ks$		$\begin{cases} F(t) = 2n - 1 \\ \quad \text{when } 2k(n - 1) < t < 2kn, \\ \qquad n = 1, 2, \ldots \end{cases}$
$(s \cosh ks)^{-1}$		$\begin{cases} M(2k, t + 3k) + 1 = 1 + (-1)^n \\ \quad \text{when } (2n - 3)k < t < (2n - 1)k, \\ \qquad n = 1, 2, \ldots \quad t > 0 \end{cases}$
$s^{-1}(e^{-as} - e^{-bs})$		$F(t) = \begin{cases} 0 & \text{for } 0 < t < a \\ 1 & \text{for } a < t < b \\ 0 & \text{for } t > b \end{cases}$
$(m/s^2) - (ma/2s)$ $[\coth(as/2) - 1]$		$\begin{cases} F(t) = m(t - na) \\ \quad \text{when } na < t < (n + 1)a, \\ \qquad n = 0, 1, 2, \ldots \end{cases}$
$s^{-2} \tanh(cs/2)$		$\begin{cases} H(c, t) = \begin{cases} t & \text{when } 0 < t < c \\ 2c - t & \text{when } c < t < 2c \end{cases} \\ H(c, t + 2nc) = H(c, t) \\ \qquad n = 1, 2, \ldots \end{cases}$

Triangular wave

$k(s^2 + k^2)^{-1} \coth[\pi s/(2k)]$	$|\sin ki|$

Full-wave rectification of sin kt

$[(s^2 + 1)(1 - e^{-\pi s})]^{-1}$	$\begin{cases} \sin t \text{ when } (2n - 2)\pi < t < (2n - 1)\pi \\ 0 \quad \text{when } (2n - 1)\pi < t < 2n\pi \\ \qquad n = 1, 2, \ldots \end{cases}$

(continued)

Table 1   (continued)

$\mathbf{f(s)} \equiv \mathcal{L}[\mathbf{F(t)}]$	$F(t)$	$\mathbf{f(s)} \equiv \mathcal{L}[\mathbf{F(t)}]$	$F(t)$
Half-wave rectification of sin t		$(m/s^2)(1 - 2e^{-as} + e^{-2as})$ $= (m/s^2)\,(1 - e^{-as})^2$	$F(t) = \begin{cases} mt & \text{for } 0 < t < a \\ -m(t-2a) & \text{for } a < t < 2a \\ 0 & \text{for } t > 2a \end{cases}$
$[(E/s) + (m/s^2)]e^{-as}$	$F(t) = \begin{cases} 0 & \text{for } 0 < t < a \\ E + m(t-a) & \text{for } t > a \end{cases}$	$(m/s^2)\,[1 - (1 + as)\,e^{-as}]$	$F(t) = \begin{cases} mt & \text{for } 0 < t < a \\ 0 & \text{for } t > a \end{cases}$
$(m/s^2)(1 - e^{-as})$	$F(t) = \begin{cases} mt & \text{for } 0 < t < a \\ ma & \text{for } t > a \end{cases}$		

(a) a, b, and c are distinct constants.
(b) $L_n\,(t)$ is the Laguerre polynomial of degree n.
(c) erf $y = \dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^{v} e^{-u^2}\,du$

  erfc $y = 1 - $ erf y.
(d) $I_n\,(x) \equiv i^{-n}\,J_n\,(ix)$, where J_n is Bessel's function of the first kind.
(e) $H_n(x) \equiv e^{x^2}[d^n(e^{-x^2})/dx^n]$ is the Hermite polynomial.
(f) $\log s \equiv \log_e s \equiv \ln s$.

  Ei$t = \displaystyle\int_{-x}^{t} r^{-1}e^{r}\,dr \quad t < 0 \quad$ (the exponential-integral function)

  $-$ Ei $(-t) = \displaystyle\int_1^{x} x^{-1}e^{-tx}\,dx \quad t > 0$

  Ci $t = -\displaystyle\int_t^{\infty} r^{-1}\cos r\,dr = -\int_1^{\infty} x^{-1}\cos tx\ dx \quad t > 0 \quad$ (the cosine-integral function)

  Si $t = \displaystyle\int_0^{t} r^{-1}\sin r\,dr \quad t > 0 \qquad$ (the sine-integral function)

  $H(t) = \cos\ t\ \text{Si}\ t - \sin\ t\ \text{Ci}\ t \quad t > 0$.
(g) $K_n(x)$ is Bessel's function of the second kind for the imaginary argument.
(h) When $t > 0$, $[t]$ denotes the greatest integer $(0, 1, 2, \ldots)$ that does not exceed the number t.

Officers and Trustees of ASM International (2008–2009)

Members of the ASM Handbook Committee (2008–2009)

Chairs of the ASM Handbook Committee

Authors and Contributors

John E. Allison
Ford Motor Company

D.-H. Bae
University of Michigan

Andrew L. Banka
Air Flow Sciences Corporation

L.A. Barrales Mora
Institut für Metallkunde und Metallphysik
Aachen, Germany

A. Beaudoin
University of Illinois at Urbana-Champaign

J.H. Beynon
Swinburne University of Technology,
Australia

H.K.D.H. Bhadeshia
University of Cambridge, UK

D.E. Boyce
Cornell University

Carelyn E. Campbell
National Institute of Standards and
Technology

H.-B. Cao
University of Wisconsin

W. Cao
CompuTherm, LLC

Y.A. Chang
University of Wisconsin

Indrajit Charit
University of Idaho

S.-L. Chen
CompuTherm, LLC

Y.B. Chun
Inha University, Incheon, Korea

Emmanuel Clouet
CEA, DEN, Service de Recherches
de Métallurgie Physique,
France

Peter C. Collins
The Ohio State University

Jonathan A. Dantzig
University of Illinois at Urbana-Champaign

C.H.J. Davies
Monash University, Australia

P.R. Dawson
Cornell University

Alain Dollet
Processes, Materials and Solar Energy
Laboratory (PROMES), CNRS,
France

David Dye
Imperial College, London, UK

Olaf Engler
Hydro Aluminium Deutschland GmbH,
Germany

Jeffrey L. Evans
University of Alabama in Huntsville

Hans J. Fecht
Institute for Micro and Nanotechnology,
University of Ulm, Germany

Hamish L. Fraser
The Ohio State University

D.U. Furrer
Rolls-Royce Corporation

Charles-Andre Gandin
Ecole Nationale Superieure
des Mines de Paris, France

A.K. Ghosh
University of Michigan

Michael G. Glavicic
Rolls-Royce Corporation

G. Gottstein
Institut für Metallkunde und Metallphysik
Aachen, Germany

Jürgen Hirsch
Hydro Aluminium Deutschland GmbH,
Germany

S.K. Hwang
Inha University, Incheon, Korea

O.M. Ivasishin
Kurdyumov Institute for Metal Physics,
Kyiv, Ukraine

John J. Jonas
McGill University, Canada

Herng-Jeng Jou
QuesTek Innovations LLC

Kai Karhausen
Hydro Aluminium Deutschland GmbH,
Germany

P.A. Kobryn
Air Force Research Laboratory,
Wright-Patterson Air Force Base

M. Krzyzanowski
The University of Sheffield,
United Kingdom

Charles Kuehmann
QuesTek Innovations LLC

Howard Kuhn
The Ex One Company

Mei Li
Ford Motor Company

Witold Lojkowski
Institute of High Pressure Physics, Polish
Academy of Sciences, Poland

D. Scott MacKenzie
Houghton International Inc.

Savko Malinov
Queen's University, Belfast, UK

David L. McDowell
Georgia Institute of Technology

H. Mecking
Technical University Hamburg-Harburg,
Germany

Fred Meisenkothen
Air Force Research Laboratory,
Wright-Patterson Air Force Base

Mark Miodownik
King's College London, UK

Rajiv S. Mishra
Missouri University of Science and
Technology

V. Mohles
Institut für Metallkunde und
Metallphysik Aachen, Germany

Frank Montheillet
Centre Science des Matériaux et des
Structures, Ecole Nationale Superieure
des Mines, de Saint-Etienne, France

John Morral
The Ohio State University

P.D. Nicolaou
Air Force Research Laboratory,
Wright-Patterson Air Force Base

G.B. Olson
Northwestern University

P.M. Pauskar
The Timken Company

Henry R. Piehler
Carnegie Mellon University

Andrew H. Rosenberger
Air Force Research Laboratory,
Wright-Patterson Air Force Base

Ashok Saxena
University of Arkansas

A. Saxena
Los Alamos National Laboratory

S.L. Semiatin
Air Force Research Laboratory,
Wright-Patterson Air Force Base

Wei Sha
Queen's University, Belfast, UK

Chen Shen
GE Global Research

S.V. Shevchenko
Kurdyumov Institute for Metal Physics,
Kyiv, Ukraine

R. Shivpuri
The Ohio State University

L.S. Shvindlerman
Institute of Solid State Physics,
Russian Academy of Sciences, Moscow,
Russia

I. Steinbach
Ruhr-University Bochum,
Germany

H.J. Stone
University of Cambridge, UK

XuMing Su
Ford Motor Company

Brian G. Thomas
University of Illinois at Urbana-Champaign

J.P. Thomas
ATI Allvac

Sammy Tin
Illinois Institute of Technology

Yunzhi Wang
The Ohio State University

Christopher Woodward
Air Force Research Laboratory,
Wright-Patterson Air Force Base

K. Wu
CompuTherm, LLC

Y. Yang
CompuTherm, LLC

F. Zhang
CompuTherm, LLC

Officers and Trustees of ASM International (2008–2009)

Members of the ASM Handbook Committee (2008–2009)

Chairs of the ASM Handbook Committee

Officers and Trustees of ASM International (2008–2009)

Members of the ASM Handbook Committee (2008–2009)

Chairs of the ASM Handbook Committee

Policy on Units of Measure

By a resolution of its Board of Trustees, ASM International has adopted the practice of publishing data in both metric and customary U.S. units of measure. In preparing this Handbook, the editors have attempted to present data in metric units based primarily on Système International d'Unités (SI), with secondary mention of the corresponding values in customary U.S. units. The decision to use SI as the primary system of units was based on the aforementioned resolution of the Board of Trustees and the widespread use of metric units throughout the world.

For the most part, numerical engineering data in the text and in tables are presented in SI-based units with the customary U.S. equivalents in parentheses (text) or adjoining columns (tables). For example, pressure, stress, and strength are shown both in SI units, which are pascals (Pa) with a suitable prefix, and in customary U.S. units, which are pounds per square inch (psi). To save space, large values of psi have been converted to kips per square inch (ksi), where 1 ksi = 1000 psi. The metric tonne (kg $\times$ 10^3) has sometimes been shown in megagrams (Mg). Some strictly scientific data are presented in SI units only.

To clarify some illustrations, only one set of units is presented on artwork. References in the accompanying text to data in the illustrations are presented in both SI-based and customary U.S. units. On graphs and charts, grids corresponding to SI-based units usually appear along the left and bottom edges. Where appropriate, corresponding customary U.S. units appear along the top and right edges.

Data pertaining to a specification published by a specification-writing group may be given in only the units used in that specification or in dual units, depending on the nature of the data. For example, the typical yield strength of steel sheet made to a specification written in customary U.S.

units would be presented in dual units, but the sheet thickness specified in that specification might be presented only in inches.

Data obtained according to standardized test methods for which the standard recommends a particular system of units are presented in the units of that system. Wherever feasible, equivalent units are also presented. Some statistical data may also be presented in only the original units used in the analysis.

Conversions and rounding have been done in accordance with IEEE/ASTM SI-10, with attention given to the number of significant digits in the original data. For example, an annealing temperature of 1570 °F contains three significant digits. In this case, the equivalent temperature would be given as 855 °C; the exact conversion to 854.44 °C would not be appropriate. For an invariant physical phenomenon that occurs at a precise temperature (such as the melting of pure silver), it would be appropriate to report the temperature as 961.93 °C or 1763.5 °F. In some instances (especially in tables and data compilations), temperature values in °C and °F are alternatives rather than conversions.

The policy of units of measure in this Handbook contains several exceptions to strict conformance to IEEE/ASTM SI-10; in each instance, the exception has been made in an effort to improve the clarity of the Handbook. The most notable exception is the use of g/cm^3 rather than kg/m^3 as the unit of measure for density (mass per unit volume).

SI practice requires that only one virgule (diagonal) appear in units formed by combination of several basic units. Therefore, all of the units preceding the virgule are in the numerator and all units following the virgule are in the denominator of the expression; no parentheses are required to prevent ambiguity.

Reviewers

Michel Bellet
CEMEF—Centre de Mise en Forme des Matériaux/Materials Processing Center, France

Thomas R. Bieler
Michigan State University

Hongbo Cao
General Electric Company

E. Buddy Damm
The Timken Company

Craig Darragh
The Timken Company (ret.)

Dennis M. Dimiduk
Air Force Research Laboratory, Wright-Patterson Air Force Base

Dave Furrer
Rolls-Royce Corporation

Zhanli Guo
Sente Software Ltd

Larry D. Hanke
Materials Evaluation and Engineering, Inc

Niels Hansen
Materials Research Division, Risø National Laboratory for Sustainable Energy, Technical University of Denmark

Jeffrey A. Hawk
National Energy Technology Laboratory

Edmond Ilia
Metaldyne

Kent L. Johnson
Harvard Engineering, Inc.

Michael Kassner
University of Southern California

Ursula R. Kattner
National Institute of Standards and Technology

David E. Laughlin
Carnegie Mellon University

Antoinette Maniatty
Rensselaer Polytechnic Institute

William Mankins
Metallurgical Services Incorporated

Claire Maurice
Ecole Nationale Superieure des Mines de Saint-Etienne, France

R. Craig McClung
Southwest Research Institute

David McDowell
Georgia Institute of Technology

Terry R. McNelley
Naval Postgraduate School

Jonathan Miller
Air Force Research Laboratory Wright-Patterson Air Force Base

Pierre Montmitonnet
CEMEF—Centre de Mise en Forme des Matériaux, Centre for Material Forming, France

John Morral
The Ohio State University

Toby Padfield
ZF NAO Technical Center

John H. Perepezko
University of Wisconsin-Madison

Elwin L. Rooy
Rooy and Associates

Kenneth Russell
Massachusetts Institute of Technology

Taku Sakai
The University of Electro-Communications, Tokyo

Ayman Salem
Air Force Research Laboratory, Wright-Patterson Air Force Base

Gordon Sargent
Air Force Research Laboratory, Wright-Patterson Air Force Base

Rainer Schmid-Fetzer
Clausthal University of Technology, Germany

S.L. Semiatin
Air Force Research Laboratory Wright-Patterson Air Force Base

Yongho Sohn
University of Central Florida

Leyun Wang
Michigan State University

Paul Wawrzynek
Fracture Analysis Consultants, Inc. and Cornell University

Dieter Wolf
Idaho National Laboratory

Yiyi Yang
Michigan State University

Glossary of Terms

A

ab initio. From the beginning (Latin); often used to refer to first-principles modeling approaches.

abnormal grain growth. Rapid, nonuniform, and usually undesirable growth of one or a small fraction of grains in a polycrystalline material during annealing. The phenomenon is most frequent in fine-grained materials in which a larger-than-average grain (or grains) consumes surrounding small grains whose growth is limited by particle pinning. Also known as *secondary recrystallization*.

acicular alpha. A product of nucleation and growth from β to the lower-temperature allotrope α-phase. It may have a needlelike appearance in a micrograph and may have needle, lenticular, or flattened bar morphology in three dimensions. This definition applies specifically to titanium alloys.

aging. A change in material property or properties with time. See also *quench aging* and *strain aging*.

alligatoring. The longitudinal splitting of flat slabs in a plane parallel to the rolled surface.

allotriomorphic crystal. A crystal having a normal lattice structure but an outward shape that is imperfect, because it is determined to some extent by the surroundings. The grains in a metallic aggregate are allotriomorphic crystals.

allotropy. The property by which certain elements may exist in more than one crystal structure.

allotropic transformation. The ability of a material to transform from one crystal structure to another. Closely synonymous with polymorphism.

alloying element. An element added to and remaining in a metal that changes structure and properties.

amorphous material. A material that lacks the long-range three-dimensional atomic periodicity that is characteristic of a crystalline solid.

angle of bite. In the rolling of metals, the location where all of the force is transmitted through the rolls; the maximum attainable angle between the roll radius at the first contact and the line of roll centers. Operating angles less than the angle of bite are termed contact angles or rolling angles.

angstrom (Å). A unit of linear measure equal to 10^{-10} m, or 0.1 nm. Although not an accepted SI unit, it is occasionally used for small distances, such as interatomic distances, and some wavelengths.

angularity. The conformity to, or deviation from, specified angular dimensions in the cross section of a shape or bar.

anion. A negatively charged ion that migrates toward the anode (positive electrode) under the influence of a potential gradient.

anisotropy. Variations in one or more physical or mechanical properties with direction with respect to a fixed reference system in the material.

annealing. A generic term denoting a treatment—heating to and holding at a suitable temperature followed by cooling at a suitable rate—used primarily to soften metallic materials but also to produce desired changes simultaneously in other properties or in microstructure. When applied only for the relief of stress, the process is called stress relieving or stress-relief annealing. In ferrous alloys, annealing is carried out above the upper critical temperature, but the time-temperature cycles vary widely in maximum temperature attained and cooling rate used, depending on composition, material condition, and desired results. In nonferrous alloys, annealing cycles are designed to remove part or all of the effects of cold working (recrystallization may or may be involved), cause complete coalescence of precipitates from the solid solution in relatively coarse from, or both, depending on composition and material condition.

asperities. Protrusions rising above the general surface contours that constitute the actual contact areas between touching surfaces.

austempering. A heat treatment for ferrous alloys in which a part is quenched from the austenitizing temperature at a rate fast enough to avoid formation of ferrite or pearlite and then held at a temperature just above the martensite start temperature until transformation to bainite is complete. Although designated as bainite in both austempered steel and austempered ductile iron, austempered steel consists of two-phase mixtures containing ferrite and carbide, while austempered ductile iron consists of two-phase mixtures containing ferrite and austenite.

austenite. A high-temperature form of iron. In steel heat treating, the steel is heated into the austenite region before rapidly cooling (quenching).

austenite stabilizer. An alloying element that, when added to iron, increases the region of the phase diagram in which austenite (face-centered cubic iron) is stable. Strong austenite stabilizers are carbon, nickel, and manganese.

average grain diameter. The mean diameter of an equiaxed grain section whose size represents all the grain sections in the aggregate being measured. See also *grain size*.

Avrami plot/(Avrami equation). Plot describing the kinetics of phase transformations in terms of the dependence of fraction X of microstructure that has transformed (e.g., recrystallized, decomposed, etc.) as a function of time (t) or strain (ε). Avrami plots usually consist of a graph of log $[\ln(1/(1 - X))]$ versus log t (or log ε) and are used to determine the so-called Avrami exponent n in the relation $X = 1 - \exp(-Bt^n)$.

axial ratio. The ratio of the length of one axis to that of another, for example, c/a, or the continued ratio of three axes, such as $a{:}b{:}c$.

axis (crystal). The edge of the unit cell of a space lattice. Any one axis of any one lattice is defined in length and direction relative to other axes of that lattice.

B

bar. A section hot rolled from a *billet* to a form, such as round, hexagonal, octagonal, square, or rectangular, with sharp or rounded corners or edges and a cross-sectional area of less than 105 cm² (16 in.²). A solid section that is long in relationship to its cross-sectional dimensions, having a completely symmetrical cross section and a width or greatest distance between parallel faces of 9.5 mm (⅜ in.) or more.

barreling. Convexity of the surfaces of cylindrical or conical bodies, often produced unintentionally during upsetting or as a natural consequence during compression testing.

basal plane. That plane of a hexagonal or tetragonal crystal perpendicular to the axis of highest symmetry. Its Miller indices are (0001).

Bauschinger effect. A reduction in yield strength on straining a material in the opposite direction to the initial testing.

bendability. The ability of a material to be bent around a specified radius without fracture.

bending. The straining of material, usually flat sheet or strip metal, by moving it around a straight axis lying in the neutral plane. Metal flow takes place within the plastic range of the metal, so that the bent part retains a *permanent set* after removal of the applied stress. The cross section of the bend inward from the neutral plane is in compression; the rest of the bend is in tension.

bending and forming. The processes of bending, flanging, folding, twisting, offsetting, or otherwise shaping a portion of a blank or a whole blank, usually without materially changing the thickness of the metal.

bending moment. The moments (force times distance) that tend to bend a beam in the plane of the loads.

bending stress. A stress involving tensile and compressive forces, which are not uniformly distributed. Its maximum value depends on the amount of flexure that a given application can accommodate. Resistance to bending can be termed stiffness.

bending under tension. A forming operation in which a sheet is bent with the simultaneous application of a tensile stress perpendicular to the bend axis.

bend (longitudinal). A forming operation in which the axis is perpendicular to the rolling direction of the sheet.

bend or twist (defect). Distortion similar to warpage generally caused during forging or trimming operations. When the distortion is along the length of the part, it is termed bend; when across the width, it is termed twist. When bend or twist exceeds tolerances, it is considered a defect. Corrective action consists of hand straightening, machine straightening, or cold restriking.

bend radius. The radius measured on the inside of a bend that corresponds to the curvature of a bent specimen or the bent area in a formed part.

bend test. Evaluation of a sheet metal response to a bending operation, such as around a fixed radius tool.

bend (transverse). A forming operation in which the bend axis is parallel to the rolling direction of the sheet.

beta structure. Structurally analogous body-centered cubic phases (similar to β-brass) or electron compounds that have ratios of three valence electrons to two atoms.

beta transus. The minimum temperature above which equilibrium α does not exist. For β eutectoid additions, the β transus ordinarily is applied to hypoeutectoid compositions or those that lie to the left of the eutectoid composition. This definition applies to titanium alloys.

biaxial stretchability. The ability of sheet material to undergo deformation by loading in tension in two directions in the plane of the sheet.

billet. A semifinished section that is hot rolled from a metal ingot, with a rectangular cross section usually ranging from 105 to 230 cm^2 (16 to 36 in.^{2}), the width being less than twice the thickness. Where the cross section exceeds 230 cm^2 (36 in.^{2}), the term *bloom* is properly but not universally used. Sizes smaller than 105 cm^2 (16 in.^{2}) are usually termed bars. A solid semifinished round or square product that has been hot worked by forging, rolling, or extrusion. See also *bar*.

bite. Advance of material normal to the plane of deformation and relative to the dies prior to each deformation step.

blank. In forming, a piece of sheet material, produced in cutting dies, that is usually subjected to further press operations. A piece of stock from which a forging is made.

blanking. The operation of punching, cutting, or shearing a piece out of stock to a predetermined shape.

blister. A local protuberance in the surface of the sheet, often elongated, resulting from an internal separation due to the expansion of entrapped gas. The gas may be entrapped during casting, pickling, annealing, or electroplating in a previously existing subsurface defect.

block and finish. The forging operation in which a part to be forged is blocked and finished in one heat through the use of tooling having both a block impression and a finish impression in the same die block.

blocking. A forging operation often used to impart an intermediate shape to a forging, preparatory to forging of the final shape in the finishing impression of the dies. Blocking can ensure proper working of the material and can increase die life.

bloom. A semifinished hot rolled product, rectangular in cross section, produced on a blooming mill. See also *billet*. For steel, the width of a bloom is not more than twice the thickness, and the cross-sectional area is usually not less than approximately 230 cm^2 (36 in.^{2}). Steel blooms are sometimes made by forging.

board hammer. A type of forging hammer in which the upper die and ram are attached to "boards" that are raised to the striking position by power-driven rollers and let fall by gravity. See also *drop hammer*.

boss. A relatively short, often cylindrical protrusion or projection on the surface of a forging.

bottom draft. Slope or taper in the bottom of a forge depression that tends to assist metal flow toward the sides of depressed areas.

boundary condition. A requirement to be met by a solution to a set of differential equations on a specified set of values of the independent variables.

bow. The tendency of material to curl downward during shearing, particularly when shearing long, narrow strips.

Bravais lattices. The 14 possible three-dimensional arrays of atoms in crystals (see *space lattice*).

brick element. The element for three-dimensional finite-element modeling that is brick shaped (six faces) and has eight nodes.

Bridgman correction. Factor used to obtain the flow stress from the measured axial stress during tension testing of metals in which necking has occurred.

Brillouin zones. Energy states for the free electrons in a metal, as described by the use of the band theory (zone theory) of electron structure. Also called electron bands.

brittle fracture. A fracture that occurs without appreciable plastic deformation.

brittleness. A tendency to fracture without appreciable plastic deformation.

buckling. A bulge, bend, kink, or other wavy condition of the workpiece caused by compressive stresses.

bulge test. A test wherein the blank is clamped securely around the periphery and, by means of hydrostatic pressure, the blank is expanded. The blank is usually gridded so that the resulting strains can be measured. This test is usually performed on large blanks of 20 to 30 cm (8 to 12 in.) in diameter.

bulging. The process of increasing the diameter of a cylindrical shell (usually to a spherical shape) or of expanding the outer walls of any shell or box shape whose walls were previously straight.

bulk forming. Forming processes, such as extrusion, forging, rolling, and drawing, in which the input material is in billet, rod, or slab form and a considerable increase in surface-to-volume ratio in the formed part occurs under the action of largely compressive loading. Compare with *sheet forming*.

Burgers vector. The crystallographic direction along which a dislocation moves and the unit displacement of dislocations; the magnitude of the Burgers vector is the smallest unit distance of slip in the direction of shear due to the movement of one dislocation.

burnishing. The smoothing of one surface through frictional contact with another surface.

burr. A thin ridge or roughness left on forgings or sheet metal blanks by cutting operations such as slitting, shearing, trimming, blanking, or sawing.

C

CAD/CAM. An abbreviation for computer-aided design/computer-aided manufacturing.

camber. The tendency of material being sheared from sheet to bend away from the sheet in the same plane.

canned extrusion. A coextrusion process in which the billet consists of a clad material, or can, that is relatively ductile and nonreactive, and the core is a reactive, brittle powder or other material.

canning. A dished distortion in a flat or nearly flat sheet metal surface, sometimes referred to as oil canning. Distortion of a flat or nearly flat metal surface that can be deflected by finger pressure but will return to its original position when the pressure is removed. Enclosing a highly reactive metal within a relatively inert material for the purpose of hot working without undue oxidation of the active metal.

casting. (1) Metal object cast to the required shape by pouring of injecting liquid metal into a mold, as distinct from one shaped by a mechanical process. (2) Pouring molten metal into a mold to produce an object of desired shape. (3) Ceramic forming process in which a body slip is introduced into a porous mold, which absorbs sufficient water from the slip to produce a semirigid circle.

casting modulus. A simplified approach to determining solidification time. The time is proportional to the square of the section modulus (the ratio of volume to surface area), known as Chvorinov's rule.

cation. A positively charged ion that migrates through the electrolyte toward a cathode (negative electrode) under the influence of a potential gradient.

cavitation. The formation of microscopic cavities during the cold or hot deformation of metals, generally involving a component of tensile stress. Cavities may nucleate at second-phase particles lying within grains or at grain boundaries (with or without particles) as a result of slip intersection or grain-boundary sliding. Under severe conditions, cavities may grow and coalesce to give rise to fracture. In liquids, cavitation is the formation and instantaneous collapse of cavities or bubbles caused by rapid and intense pressure changes. Cavitation caused by untrasonic radiation is sometimes used to effect violent local agitation. Cavitation caused by severe turbulent flow often leads to damage of adjacent material surfaces.

cell. Micron-sized volume bounded by low-misorientation walls comprised of dislocation tangles.

center bursting. Internal cracking due to tensile stresses along the central axis of products being extruded or drawn.

chamfer. A beveled surface to eliminate an otherwise sharp corner. A relieved angular cutting edge at a tooth corner.

check. A crack in a die impression corner, generally due to forging strains or pressure, localized at some relatively sharp corner. Die blocks too hard for the depth of the die impression have a tendency to check or develop cracks in impression corners. One of a series of small cracks resulting from thermal fatigue of hot forging dies (often called a heat check or heat checking).

chord modulus. The slope of the chord drawn between any two specific points on a stress-strain curve. See also *modulus of elasticity*.

circle grid. A regular pattern of circles, often 2.5 mm (0.1 in.) in diameter, marked on a sheet metal blank.

circle-grid analysis. The analysis of deformed circles to determine the severity with which a sheet metal blank has been deformed.

clad. Outer layer of a coextruded or codrawn product. See also *sleeve*.

clamping pressure. Pressure applied to a limited area of the sheet surface, usually at the periphery, to control or limit metal flow during forming.

clearance. In punching and shearing dies, the gap between the die and the punch. In forming and drawing dies, the difference between this gap and metal thickness.

closed-die forging. The shaping of hot metal completely within the walls or cavities of two dies that come together to enclose the workpiece on all sides. The impression for the forging can be entirely in either die or divided between the top and bottom dies. Impression die forging, often used interchangeably with the term closed-die forging, refers to a closed-die operation in which the dies contain a provision for controlling the flow of excess material, or *flash*, that is generated. By contrast, in flashless forging, the material is deformed in a cavity that allows little or no escape of excess material.

closed dies. Forging or forming impression dies designed to restrict the flow of metal to the cavity within the die set, as opposed to open dies, in which there is little or no restriction to lateral flow.

closed pass. A pass of metal through rolls where the bottom roll has a groove deeper than the bar being rolled and the top roll has a collar fitting into the groove, thus producing the desired shape free from *flash* or *fin*.

close-tolerance forging. A forging held to unusually close dimensional tolerances so that little or no machining is required after forging. See also *precision forging*.

cluster mill. A rolling mill in which each of two small-diameter work rolls is supported by two or more backup rolls.

coarsening. The increase in the average size of second-phase particles, accompanied by the reduction in their number, during annealing, deformation, or high-temperature service exposure. Coarsening thus leads to a decrease in the total surface energy associated with the matrix-particle interfaces.

codrawing. The simultaneous drawing of two or more materials to form an integral product.

coefficient of friction. A measure of the ease with which one body will slide over another. It is obtained by dividing the tangential force resisting motion between the two bodies by the normal force pressing the two bodies together.

coefficient of thermal expansion (CTE). Change in unit of length (or volume) accompanying a unit change of temperature, at a specified temperature or temperature range.

coextrusion. The simultaneous extrusion of two or more materials to form an integral product.

cogging. The reducing operation in working an ingot into a billet with a forging hammer or a forging press.

coil breaks. Creases or ridges that appear as parallel lines transverse to the direction of rolling and extend across the width of the sheet. Coil breaks are caused by the deformation of local areas during coiling or uncoiling of annealed or insufficiently temper-rolled steel sheets.

coining. A closed-die squeezing operation in which all surfaces of a workpiece are confined or restrained, resulting in a well-defined imprint of the die on the work. A *restriking* operation used to sharpen or change an existing radius or profile. Coining can be done while forgings are hot or cold and is usually performed on surfaces parallel to the parting line of the forging.

cold forming. See *cold working*.

cold heading. Working metal at room temperature such that the cross-sectional area of a portion or all of the stock is increased. See also *heading* and *upsetting*.

cold lap. A flaw that results when a workpiece fails to fill the die cavity during the first forging. A seam is formed as subsequent dies force metal over this gap to leave a seam on the workpiece surface. See also *cold shut*.

cold rolled sheet. A mill product produced from a hot rolled pickled coil that has been given substantial cold reduction at room temperature. After annealing, the usual end product is characterized by improved surface, greater uniformity in thickness, increased tensile strength, and improved mechanical properties as compared with hot rolled sheet.

cold shut. A fissure or lap on a forging surface that has been closed without fusion during the forging operation. A folding back of metal onto its own surface during flow in the die cavity; a forging defect.

cold trimming. The removal of flash or excess metal from a forging at room temperature in a trimming press.

cold-worked structure. A microstructure resulting from plastic deformation of a metal or alloy below its recrystallization temperature.

cold working. The plastic deformation of metal under conditions of temperature and strain rate that induce *strain hardening*. Usually, but not necessarily, conducted at room temperature. Also referred to as cold forming or cold forging. Contrast with *hot working*.

columnar structure. A coarse structure of parallel, elongated grains formed by unidirectional growth that is most often observed in castings but sometimes seen in structures. This results from diffusional growth accompanied by a solid-state transformation.

compact (noun). The object produced by the compression of metal powder, generally while confined in a die.

compact (verb). The operation or process of producing a compact; sometimes called pressing.

compression test. A method for assessing the ability of a material to withstand compressive loads.

compressive strength. The maximum compressive stress a material is capable of developing. With a brittle material that fails in compression by fracturing, the compressive strength has a definite value. In the case of ductile, malleable, or semiviscous materials (which do not fail in compression by a shattering fracture), the value obtained for compressive strength is an arbitrary value dependent on the degree of distortion that is regarded as effective failure of the material.

compressive stress. A stress that causes an elastic or plastic body to deform (shorten) in the direction of the applied load. Contrast with *tensile stress*.

computational fluid dynamics (CFD). An area of computer-aided engineering devoted to the numerical solution and visualization of fluid-flow problems.

computer-aided design (CAD). Any design activity that involves the effective use of the computer to create or modify an engineering design. Often used synonymously with the more general term computer-aided engineering (CAE).

computer-aided materials selection system. A computerized database of materials properties operated on by an appropriate knowledge base of decision rules through an *expert system* to select the most appropriate materials for an application.

concurrent engineering. A style of product design and development that is done by concurrently using all of the relevant information in making each decision. It replaces a sequential approach to product development in which one type of information was predominant in making each sequential decision. Concurrent engineering is carried out by a multifunctional team that integrates the specialties or functions needed to solve the problem. Sometimes called simultaneous engineering.

constitutive equation. Equation expressing the relation between stress, strain, strain rate, and microstructural features (e.g., grain size). Constitutive equations are generally phenomenological (curve fits based on measured data) or mechanism-based (based on mechanistic model of deformation and appropriate measurements). Phenomenological constitutive equations are usually valid only within the processing regime in which they were measured, while mechanism-based relations can be extrapolated outside the regime of measurement, provided the deformation mechanism is unchanged. A mathematical relationship that describes the flow stress of a material in terms of the plastic strain, the strain rate, and temperature.

constraint modeling. A form of computer modeling in which constraints are used to create a set of rules that control how changes are made to a group of geometric elements (lines, arcs, form features, etc.). These rules are typically embodied in a set of equations. Constraint models are defined as either parametric or variational.

continuum mechanics. The science of mathematically describing the behavior of continuous media. The same basic approach can apply to descriptions of stress, heat, mass, and momentum transfer.

controlled rolling. Multistand plate or bar rolling process, typically for ferrous alloys, in which the reduction per pass, rolling speed, time between passes, and so on are carefully chosen to control recrystallization, precipitation, and phase transformation in order to develop a desired microstructure and set of properties.

conventional forging. Forging process in which the work material is hot and the dies are at room temperature or slightly elevated temperature. To minimize the effects of die chilling on metal flow and microstructure, conventional forging usually involves strain rates of the order of 0.05 s^{-1} or greater. Also known as nonisothermal forging.

core. (1) Inner material in a coextruded or codrawn product. (2) In casting, a specially formed material inserted in a mold to shape the interior part of the casting that cannot be shaped as easily by the pattern. (3) In ferrous alloys prepared for case hardening, that portion of the alloy that is not part of the case.

coring. (1) A central cavity at the butt end of a rod extrusion; sometimes called *extrusion pipe*. (2) A condition of variable composition between the center and surface of a unit of microstructure (such as a dendrite, grain, or carbide particle); results from nonequilibrium solidification, which occurs over a range of temperature.

corrugating. The forming of sheet metal into a series of straight, parallel alternate ridges and grooves with a rolling mill equipped with matched roller dies or a press brake equipped with specially shaped punch and die.

corrugations. Transverse ripples caused by a variation in strip shape during hot or cold reduction.

Coulomb friction. Interface friction condition for which the interface shear stress is proportional to the pressure normal to the interface. The proportionality constant is called the Coulomb coefficient of friction, μ, and takes on values between 0 (perfect lubrication) and $1/\sqrt{3}$ (sticking friction) during metalworking. See also *friction shear factor*.

cracked edge. A series of tears at the edge of the sheet resulting from prior processing.

crank. Forging shape generally in the form of a "U" with projections at more or less right angles to the upper terminals. Crank shapes are designated by the number of throws (for example, two-throw crank).

creep. Time-dependent strain occurring under stress.

creep forming. Forming, usually at elevated temperatures, where the material is deformed over time with a preload, usually weights placed on the parts during a stress-relief cycle.

crimping. The forming of relatively small *corrugations* in order to set down and lock a seam, to create an arc in a strip of metal, or to reduce an existing arc or diameter. See also *corrugating*.

cross breaks. Visually apparent line-type discontinuities more or less transverse to the coil rolling direction, resulting from bending the coil over too sharp a radius and thus kinking the metal.

crystal. A solid composed of atoms, ions, or molecules arranged in a pattern that is periodic in three dimensions.

crystal lattice. A regular array of points about which the atoms or ions of a crystal are centered.

crystalline. That form of a substance comprised predominantly of (one or more) crystals, as opposed to glassy or amphorous.

crystalline defects. The deviations from a perfect three-dimensional atomic packing that are responsible for much of the structure-sensitive properties of materials. Crystal defects can be point defects (vacancies), line defects (dislocations), or surface defects (grain boundaries).

crystal-plasticity modeling. Physics-based modeling techniques that treat the phenomena of deformation by way of slip and twinning in order to predict strength and the evolution of crystallographic texture during the deformation processing of polycrystalline materials. See also *deformation texture, slip, Schmid's law, Taylor factor*, and *twinning*.

crystal system. One of seven groups into which all crystals may be divided: triclinic, monoclinic, orthorhombic, hexagonal, rhombohedral, tetragonal, and cubic.

cube texture. A texture found in wrought metals in the cubic system in which nearly all the crystal grains have a plane of the type (100) parallel or nearly parallel to the plane of working and a direction of the type [001] parallel or nearly parallel to the direction of elongation.

cumulative distribution function (CDF). A frequency distribution arranged to give the number of observations that are less than given values. 100% of the observations will be less than the largest class interval of the observations.

cup. (1) A sheet metal part; the product of the first drawing operation. (2) Any cylindrical part or shell closed at one end.

cup fracture (cup-and-cone fracture). A mixed-mode fracture, often seen in tensile test

specimens of a ductile material, in which the central portion undergoes plane-strain fracture and the surrounding region undergoes plane-stress fracture. One of the mating fracture surfaces looks like a miniature cup; it has a central depressed flat-face region surrounded by a shear lip. The other fracture surface looks like a miniature truncated cone.

cupping. (1) The first step in *deep drawing*. (2) Fracture of severely worked rods or wire in which one end looks like a cup and the other a cone.

cupping test. A mechanical test used to determine the ductility and stretching properties of sheet metal. It consists of measuring the maximum part depth that can be formed before fracture. The test is typically carried out by stretching the testpiece clamped at its edges into a circular die using a punch with a hemispherical end. See also *cup fracture*, *Erichsen test*, and *Olsen ductility test*.

curling. Forming of an edge of circular cross section along a sheet or along the end of a shell or tube, either to the inside or outside, for example, the pinholes in sheet metal hinges and the curled edges on cans, pots, and pans.

D

damage. General term used to describe the development of defects such as cavities, cracks, shear bands, and so on that may culminate in gross fracture in severe cases. The evolution of damage is strongly dependent on material, microstructure, and processing conditions (strain, strain rate, temperature, and stress state).

DBTT. See *ductile-to-brittle transition temperature (DBTT)*.

dead-metal zone. Region of metal undergoing limited or no deformation during bulk forming of a workpiece, generally developed adjacent to the workpiece-tooling interface as a result of friction, die chilling, or deformation-zone geometry.

deep drawing. Forming operation characterized by the production of a parallel-wall cup from a flat blank of sheet metal. The blank may be circular, rectangular, or a more complex shape. The blank is drawn into the die cavity by the action of a punch. Deformation is restricted to the flange areas of the blank. No deformation occurs under the bottom of the punch—the area of the blank that was originally within the die opening. As the punch forms the cup, the amount of material in the flange decreases. Also called cup drawing or radial drawing.

deflection. The amount of deviation from a straight line or plane when a force is applied to a press member. Generally used to specify the allowable bending of the bed, slide, or frame at rated capacity with a load of predetermined distribution.

deformation (adiabatic) heating. Temperature increase that occurs in a workpiece due to the conversion of strain energy, imparted during metalworking, into heat.

deformation energy method. A metalforming analysis technique that takes into account only the energy required to deform the workpiece.

deformation limit. In *drawing*, the limit of deformation is reached when the load required to deform the flange becomes greater than the load-carrying capacity of the cup wall. The deformation limit (limiting drawing ratio) is defined as the ratio of the maximum blank diameter that can be drawn into a cup without failure, to the diameter of the punch.

deformation-mechanism map. Strain rate/temperature map that describes forming or service regimes under which deformation is controlled by micromechanical processes such as dislocation glide, dislocation climb, and diffusional flow limited by bulk or boundary diffusion.

deformation processing. A class of manufacturing operation that involves changing the shape of a workpiece by plastic deformation through the application of a compressive force. Often carried out at elevated temperature.

deformation texture. Preferred orientation of the crystals/grains comprising a polycrystalline aggregate that is developed during deformation processing as a result of slip and rotation within each crystal that comprises the aggregate.

Demarest process. A *fluid forming* process in which cylindrical and conical sheet metal parts are formed by a modified rubber bulging punch. The punch, equipped with a hydraulic cell, is placed inside the workpiece, which in turn is placed inside the die. Hydraulic pressure expands the punch.

density. Mass per unit volume. Weight per unit volume.

design of experiments. Methodology for choosing a small number of screening experiments to establish the important material and process variables in a complex manufacturing process.

developed blank. A sheet metal blank that yields a finished part without trimming or with the least amount of trimming.

deviatoric. The nonhydrostatic component of the state of stress on a body. It is the deviatoric component that causes shape change (plastic deformation).

die. (1) A tool, usually containing a cavity, that imparts shape to solid, molten, or powdered metal primarily because of the shape of the tool itself. Used in many press operations (including blanking, drawing, forging, and forming), in die casting, and in forming green powder metallurgy compacts. Die-casting and powder metallurgy dies are sometimes referred to as molds. See also *forging dies*. (2) A complete tool used in a press for any operation or series of operations, such as forming, impressing, piercing, and cutting.

The upper member or members are attached to the slide (or slides) of the press, and the lower member is clamped or bolted to the bed or bolster, with the die members being so shaped as to cut or form the material placed between them when the press makes a stroke. (3) The female part of a complete die assembly as described previously.

die assembly. The parts of a die stamp or press that hold the die and locate it for the punches.

die block. A block, often made of heat treated steel, into which desired impressions are machined or sunk and from which closed-die forgings or sheet metal stampings are produced using hammers or presses. In forging, die blocks are usually used in pairs, with part of the impression in one of the blocks and the rest of the impression in the other. In sheet metal forming, the female die is used in conjunction with a male punch. See also *closed-die forging*.

die button. An insert in a die that matches the punch and is used for punching and piercing operations. The die button is readily removable for sharpening or replacement as an individual part of a die.

die cavity. The machined recess that gives a forging or stamping its shape.

die check. A crack in a die impression due to forging and thermal strains at relatively sharp corners. Upon forging, these cracks become filled with metal, producing sharp, ragged edges on the part. Usual die wear is the gradual enlarging of the die impression due to erosion of the die material, generally occurring in areas subject to repeated high pressures during forging.

die chill. The temperature loss experienced by a billet or preform when it contacts dies that are maintained at a lower temperature.

die clearance. Clearance between a mated punch and die; commonly expressed as clearance per side. Also called clearance or punch-to-die clearance.

die coating. Hard metal incorporated into the working surface of a die to protect the working surface or to separate the sheet metal surface from direct contact with the basic die material. Hard-chromium plating is an example.

die forging. A forging that is formed to the required shape and size through working in machined impressions in specially prepared dies.

die forming. The shaping of solid or powdered metal by forcing it into or through the *die cavity*.

die impression. The portion of the die surface that shapes a forging or sheet metal part.

die insert. A relatively small die that contains part or all of the impression of a forging or sheet metal part and is fastened to the master *die block*.

die line. A line or scratch resulting from the use of a roughened tool or the drag of a foreign particle between tool and product.

die lock. A phenomenon in which the deformation is limited in a forging near the die face due to chilling of the workpiece and/or friction at the workpiece-die interface.

die lubricant. In forging or forming, a compound that is sprayed, swabbed, or otherwise applied on die surfaces or the workpiece during the forging or forming process to reduce friction. Lubricants also facilitate release of the part from the dies and provide thermal insulation. See also *lubricant*.

die match. The alignment of the upper (moving) and lower (stationary) dies in a hammer or press. An allowance for misalignment (or mismatch) is included in forging tolerances.

die radius. The radius on the exposed edge of a deep-drawing die, over which the sheet flows in forming drawn shells.

die section. A section of a cutting, forming, or flanging die that is fastened to other sections to make up the complete working surface. Also referred to as cutting section.

die set. (1) The assembly of the upper and lower die shoes (punch and die holders), usually including the guide pins, guide pin bushings, and heel blocks. This assembly takes many forms, shapes, and sizes and is frequently purchased as a commercially available unit. (2) Two (or, for a mechanical upsetter, three) machined dies used together during the production of a *die forging*.

die shift. The condition that occurs after the dies have been set up in a forging unit in which a portion of the impression of one die is not in perfect alignment with the corresponding portion of the other die. This results in a mismatch in the forging, a condition that must be held within the specified tolerance.

die stamping. The general term for a sheet metal part that is formed, shaped, or cut by a die in a press in one or more operations.

dimpling. (1) The stretching of a relatively small, shallow indentation into sheet metal. (2) In aircraft, the stretching of metal into a conical flange for a countersunk head rivet.

direct (forward) extrusion. See *extrusion*.

directional solidification. Controlled solidification of molten metal in a casting so as to provide feed metal to the solidifying front of the casting. Usually, this results in the metal solidifying in a preferred direction. In the limit, the solidification can be controlled to grow as a single grain (single-crystal casting).

discontinuous yielding. The nonuniform plastic flow of a metal exhibiting a yield point in which plastic deformation is inhomogeneously distributed along the gage length. Under some circumstances, it may occur in metals not exhibiting a distinct yield point, either at the onset of or during plastic flow.

dishing. The formation of a shallow concave surface in which the projected area is very large compared with the depth of the impression.

dislocation. A linear imperfection in a crystalline array of atoms. Two basic types include: (1) an edge dislocation corresponds to the row of mismatched atoms along the edge formed by an extra, partial plane of atoms within the body of a crystal; and (2) a screw dislocation corresponds to the axis of a spiral structure in a crystal, characterized by a distortion that joins normally parallel planes together to form a continuous helical ramp (with a pitch of one interplanar distance) winding about the dislocation. Most prevalent is the so-called mixed dislocation, which is any combination of an edge dislocation and a screw dislocation.

dislocation density. The total length of dislocation lines per unit volume, or the number of dislocation lines that cut through a unit cross-sectional area.

dispersion strengthening. The strengthening of a metal or alloy by incorporating chemically stable submicron-sized particles of a nonmetallic phase that impede dislocation movement at elevated temperature.

double-cone test. Simulative bulk forming test consisting of the compression of a sample shaped like a flying saucer between flat dies. The variation of strain and stress state developed across the sample is used to obtain a large quantity of data on microstructure evolution and failure in a single experiment.

draft. The amount of taper on the sides of the forging and on projections to facilitate removal from the dies; also, the corresponding taper on the sidewalls of the die impressions. In *open-die forging*, draft is the amount of relative movement of the dies toward each other through the metal in one application of power.

drawability. A measure of the *formability* of a sheet metal subject to a drawing process. The term is usually used to indicate the ability of a metal to be deep drawn. See also *drawing* and *deep drawing*.

draw bead. An insert or riblike projection on the draw ring or hold-down surfaces that aids in controlling the rate of metal flow during deep-drawing operations. Draw beads are especially useful in controlling the rate of metal flow in irregularly shaped stampings.

draw forming. A method of curving bars, tubes, or rolled or extruded sections in which the stock is bent around a rotating form block. Stock is bent by clamping it to the form block, then rotating the form block while the stock is pressed between the form block and a pressure die held against the periphery of the form block.

drawing. A term used for a variety of forming operations, such as *deep drawing* a sheet metal blank; *redrawing* a tubular part; and drawing rod, wire, and tube. The usual drawing process with regard to sheet metal working in a press is a method for producing a cuplike form from a sheet metal disk by holding it firmly between blankholding surfaces to prevent the formation of wrinkles while the punch travel produces the required shape.

drawing die. A type of die designed to produce nonflat parts such as boxes, pans, and so on. Whenever practical, the die should be designed and built to finish the part in one stroke of the press, but if the part is deep in proportion to its diameter, redrawing operations are necessary.

drawing ratio. The ratio of the blank diameter to the punch diameter.

draw radius. The radius at the edge of a die or punch over which sheet metal is drawn.

draw stock. The forging operation in which the length of a metal mass (stock) is increased at the expense of its cross section; no *upset* is involved. The operation includes converting ingot to pressed bar using "V," round, or flat dies.

drop forging. The forging obtained by hammering metal in a pair of closed dies to produce the form in the finishing impression under a *drop hammer*; forging method requiring special dies for each shape.

drop hammer. A term generally applied to forging hammers in which energy for forging is provided by gravity, steam, or compressed air.

drop hammer forming. A process for producing shapes by the progressive deformation of sheet metal in matched dies under the repetitive blows of a gravity-drop or power-drop hammer. The process is restricted to relatively shallow parts and thin sheet from approximately 0.6 to 1.6 mm (0.024 to 0.064 in.).

drop through. The type of ejection where the part or scrap drops through an opening in the die.

dry-film lubricant. A type of lubricant applied by spraying or painting on coils or sheets prior to blanking, drawing, or stamping. The lubricant can have a wax base and be sprayed hot onto the sheet surface and solidify on cooling, or be a water-based polymer and be roll coated onto the surface (one or both sides) and be heated to cure and dry. Such lubricants have uniform thickness, low coefficients of friction, and offer protection from corrosion in transit and storage.

ductile fracture. Failure of metals as a result of cavity nucleation, growth, and coalescence. Ductile fracture may occur during metal forming at both cold and hot working temperatures.

ductile-to-brittle transition temperature (DBTT). A temperature or range of temperatures over which a material reaction to impact (high strain rate) loads changes from ductile, high-energy-absorbing to brittle, low-energy-absorbing behavior. The DBTT determinations are often done with Charpy or Izod test specimens measuring absorbed energy at various temperatures.

ductility. A measure of the amount of deformation that a material can withstand without breaking.

dynamic friction. The friction forces between two surfaces in relative motion. See also *static friction*.

dynamic material modeling. A methodology by which macroscopic measurements of flow stress as a function of temperature and strain rate are used with continuum criteria of instability to identify regions of temperature and strain rate in which voids, cracks, shear bands, and flow localization are likely to occur.

dynamic recovery. Recovery process that occurs during cold or hot working of metals, typically resulting in the formation of low-energy dislocation substructures/subgrains within the deformed original grains. Dynamic recovery reduces the observed level of strain hardening due to dislocation multiplication during deformation.

dynamic recrystallization. The formation of strain-free recrystallized grains during hot working. It results in a decrease in flow stress and formation of equiaxed grains, as opposed to dynamic recovery in which the elongated grains remain.

dynamic strain aging. A behavior in metals in which solute atoms are sufficiently mobile to move toward and interact with dislocations during deformation. This results in strengthening over a specific range of elevated temperature and strain rate.

E

earing. The formation of ears or scalloped edges around the top of a drawn shell, resulting from directional differences in the plastic-working properties of rolled metal with, across, and at angles to the direction of rolling.

edge dislocation. A line imperfection that corresponds to the row of mismatched atoms along the edge formed by an extra, partial plane of atoms within the body of a crystal.

edger (edging impression). The portion of a die impression that distributes metal during forging into areas where it is most needed in order to facilitate filling the cavities of subsequent impressions to be used in the forging sequence.

edge strain. Repetitive areas of local deformation extending inwardly from the edge of temper-rolled sheet.

edging. (1) In sheet metal forming, reducing the flange radius by retracting the forming punch a small amount after the stroke but before release of the pressure. (2) In rolling, the working of metal in which the axis of the roll is parallel to the thickness dimension. Also called edge rolling. The result is changing a rounded edge to a square edge. (3) The forging operation of working a bar between contoured dies while turning it 90° between blows to produce a varying rectangular cross section.

effective draw. The maximum limits of forming depth that can be achieved with a multiple-action press; sometimes called maximum draw or maximum depth of draw.

effective strain. The (scalar) strain conjugate to effective stress defined in such a manner that the product of the effective stress and the effective strain increment is equal to the increment in imposed work during a deformation process.

effective stress. A mathematical way to express a two- or three-dimensional stress state by a single number.

elastic deformation. A change in dimensions that is directly proportional to and in phase with an increase or decrease in applied force; deformation that is recoverable when the applied force is removed.

elasticity. The property of a material by which the deformation caused by stress disappears upon removal of the stress. A perfectly elastic body completely recovers its original shape and dimensions after the release of stress.

elastic limit. The maximum stress a material can sustain without any permanent strain (deformation) remaining upon complete release of the stress.

elastic modulus. See *Young's modulus*.

elastohydrodynamic lubrication. A condition of lubrication in which the friction and film thickness between two bodies in relative motion is determined by the elastic properties of the bodies in combination with the viscous properties of the lubricant.

electrical resistivity. The electrical resistance offered by a material to the flow of current, times the cross-sectional area of current flow and per unit length of current path; the reciprocal of the conductivity. Also called resistivity or specific resistance.

electric-discharge machining (EDM). Metal-removal (machining) process based on the electric discharge/spark erosion resulting from current flowing between an electrode and workpiece placed in close proximity to each other in a dielectric fluid. The electrode may be a wire (as in wire EDM) or a contoured shape (so-called plunge EDM); the latter technique is used for making metalworking dies.

electromagnetic forming. A process for forming metal by the direct application of an intense, transient magnetic field. The workpiece is formed without mechanical contact by the passage of a pulse of electric current through a forming coil. Also known as magnetic pulse forming.

electron backscatter diffraction (EBSD). Materials characterization technique conducted in a scanning electron microscope (and sometimes a transmission electron microscope) used to establish the crystallographic orientation of individual (micronsized) regions of material through analysis of Kikuchi patterns formed by backscattered electrons. Automated EBSD systems can

thus be used to determine the texture over small-to-moderate-sized total volumes of material.

elongation. (1) A term used in mechanical testing to describe the amount of extension of a testpiece when stressed. (2) In tensile testing, the increase in the gage length, measured after fracture of the specimen within the gage length, usually expressed as a percentage of the original gage length.

elongation, percent. The extension of a uniform section of a specimen expressed as a percentage of the original gage length:

$$\text{Elongation}, \% = \frac{(L_x - L_0)}{L_0} \times 100$$

where L_0 is the original gage length, and L_x is the final gage length.

embossing. A process for producing raised or sunken designs in sheet material by means of male and female dies, theoretically with minimal change in metal thickness. Examples are letters, ornamental pictures, and ribs for stiffening. Heavy embossing and *coining* are similar operations.

engineering strain. A term sometimes used for average linear strain or nominal strain in order to distinguish it from true strain. In tensile testing, it is calculated by dividing the change in the gage length by the original gage length.

engineering stress. A term sometimes used for average linear stress or nominal stress in order to differentiate it from true stess. In tension testing, it is calculated by dividing the load applied to the specimen by its original cross-sectional area.

Erichsen test. A *cupping test* used to assess the ductility of sheet metal. The method consists of forcing a conical or hemispherical-ended plunger into the specimen and measuring the depth of the impression at fracture.

error function. The function that often results as a solution to a partial differential equation. The error function is defined as:

$$\text{erf(x)} = \frac{2}{\pi} \int_0^x e^{-x^2} dx$$

The error function is also called the probability integral.

erosion resistance. The ability of a die material to resist sliding wear and thus maintain its original dimension.

etching. Production of designs, including grids, on a metal surface by a corrosive reagent or electrolytic action.

Euler angles. Set of three angular rotations used to specify unambiguously the spatial orientation of crystallites relative to a fixed reference frame.

expert system. A computer-based system that captures the knowledge of experts through the integration of databases and knowledge

bases using search and logic deduction algorithms.

extruded hole. A hole formed by a punch that first cleanly cuts a hole and then is pushed farther through to form a flange with an enlargement of the original hole. This may be a two-step operation.

extrusion. The conversion of an ingot or billet into lengths of uniform cross section by forcing metal to flow plastically through a die orifice. In forward (direct) extrusion, the die and ram are at opposite ends of the extrusion stock, and the product and ram travel in the same direction. Also, there is relative motion between the extrusion stock and the die. In backward (indirect) extrusion, the die is at the ram end of the stock, and the product travels in the direction opposite that of the ram, either around the ram (as in the impact extrusion of cylinders, such as cases for dry cell batteries) or up through the center of a hollow ram.

extrusion billet. A metal slug used as *extrusion stock*.

extrusion forging. Forcing metal into or through a die opening by restricting flow in other directions. A part made by the operation.

extrusion pipe. A central oxide-lined discontinuity that occasionally occurs in the last 10 to 20% of an extruded bar. It is caused by the oxidized outer surface of the billet flowing around the end of the billet and into the center of the bar during the final stages of extrusion. Also called *coring*.

extrusion stock. A rod, bar, or other section used to make extrusions.

eyeleting. The displacing of material about an opening in sheet or plate so that a lip protruding above the surface is formed.

F

fatigue. The phenomenon leading to fracture under repeated or fluctuating stresses having a maximum value less than the ultimate tensile strength of the material. Fatigue failure generally occurs at loads that, applied statically, would produce little perceptible effect. Fatigue fractures are progressive, beginning as minute cracks that grow under the action of the fluctuating stress.

fatigue limit. The maximum cyclic stress that a material can withstand for an infinitely large number of stress cycles. Also called endurance limit.

fatigue-strength reduction factor. The ratio of the fatigue strength of a member or specimen with no stress concentration to the fatigue strength with stress concentration. This factor has no meaning unless the stress range and the shape, size, and material of the member or specimen are stated.

fiber texture. Crystallographic texture in which all or a large fraction of the crystals in a polycrystalline aggregate are oriented such that a specific direction in each crystal is parallel to a specific sample direction, such as the axis of symmetry of a cylindrical object. Often found in wrought products such as wire and round extrusions that have been subjected to large axisymmetric deformation.

fillet. The concave intersection of two surfaces. In forging, the desired radius at the concave intersection of two surfaces is usually specified.

film strength. The ability of a surface film to resist rupture by the penetration of asperities during sliding or rolling of two surfaces over each other.

fin. The thin projection formed on a forging by trimming or when metal is forced under pressure into hairline cracks or die interfaces.

finishing dies. The die set used in the last forging step.

finish trim. Flash removal from a forging; usually performed by trimming but sometimes by band sawing or similar techniques.

finite-element analysis (FEA). A computer-based technique used to solve simultaneous equations that is used to predict the response of structures to applied loads and temperature. The FEA is a tool used to model deformation and heat treating processes.

finite-element modeling (FEM). A numerical technique in which the analysis of a complex part is represented by a mesh of elements interconnected at node points. The coordinates of the nodes are combined with the elastic properties of the material to produce a stiffness matrix, and this matrix is combined with the applied loads to determine the deflections at the nodes and hence the stresses. All of the aforementioned is done with special FEM software. The FEM approach also may be used to solve other field problems in heat transfer, fluid flow, acoustics, and so on. Also known as finite-element analysis (FEA).

fixture. A tool or device for holding and accurately positioning a piece or part on a machine tool or other processing machine.

flame hardening. A heat treating method for surface hardening steel of the proper specifications in which an oxyacetylene flame heats the surface to a temperature at which subsequent cooling, usually with water or air, will give the required surface hardness.

flame straightening. The correction of distortion in metal structures by localized heating with a gas flame.

flange. A projecting rim or edge of a part; usually narrow and of approximately constant width for stiffening or fastening.

flanging. A bending operation in which a narrow strip at the edge of a sheet is bent down (up) along a straight or curved line. It is used for edge strengthening, appearance, rigidity, and the removal of sheared edges. A flange is often used as a fastening surface.

flaring. The forming of an outward acute-angle *flange* on a tubular part.

flash. Metal in excess of that required to completely fill the blocking or finishing forging impression of a set of dies. Flash extends out from the body of the forging as a thin plate at the line where the dies meet and is subsequently removed by trimming. Because it cools faster than the body of the component during forging, flash can serve to restrict metal flow at the line where dies meet, thus ensuring complete filling of the impression.

flash extension. That portion of flash remaining on a forged part after trimming; usually included in the normal forging tolerances.

flash line. The line left on a forging after the flash has been trimmed off.

flash pan. The machined-out portion of a forging die that permits the flow through of excess metal.

flattening. (1) A preliminary operation performed on forging stock to position the metal for a subsequent forging operation. (2) The removal of irregularities or distortion in sheets or plates by a method such as *roller leveling or stretcher leveling*. (3) For wire, rolling round wire to a flattened condition.

flattening dies. Dies used to flatten sheet metal hems; that is, dies that can flatten a bend by closing it. These dies consist of a top and bottom die with a flat surface that can close one section (flange) to another (hem, seam).

fleck scale. A fine pattern of scale marks on the sheet surface that can be either dark or light. The dark pattern originates from scale and other impurities embedded in the strip during hot rolling that are not removed during pickling. The light pattern originates from a scale pattern imprinted on the work rolls in the finishing stands in the hot mill being printed onto the strip.

flex roll. A movable roll designed to push up against a sheet as it passes through a roller leveler. The flex roll can be adjusted to deflect the sheet any amount up to the roll diameter.

flex rolling. Passing sheets through a *flex roll* unit to minimize yield point elongation in order to reduce the tendency for *stretcher* strains to appear during forming.

flexural strength. A property of solid material that indicates its ability to withstand a flexural or transverse load.

flow curve. A curve of true stress versus true strain that shows the stress required to produce plastic deformation. A graphical representation of the relationship between load and deformation during plastic deformation.

flow lines. (1) Texture showing the direction of metal flow during hot or cold working. Flow lines can often be revealed by etching the surface or a section of a metal part. (2) In mechanical metallurgy, paths followed by minute volumes of metal during deformation.

flow localization. A situation where material deformation is localized to a narrow zone. Such zones often are sites of failure. Flow

localization results from poor lubrication, temperature gradients, or flow softening resulting from adiabatic heating, generation of softer crystallographic texture, grain coarsening, or spheroidization of second phases.

flow softening. Stress-strain behavior observed under constant strain-rate conditions characterized by decreasing flow stress with increasing strain. Flow softening may result from deformation heating as well as a number of microstructural sources, such as the generation of a softer crystallographic texture and the spheroidization of a lamellar phase.

flow stress. The uniaxial true stress required to cause plastic deformation at a particular value of strain, strain rate, and temperature.

flow through. A forging defect caused by metal flow past the base of a rib with resulting rupture of the grain structure.

fluid-cell process. A modification of the Guerin process for forming sheet metal, the fluid-cell process uses higher pressure and is primarily designed for forming slightly deeper parts, using a rubber pad as either the die or punch. A flexible hydraulic fluid cell forces an auxiliary rubber pad to follow the contour of the form block and exert a nearly uniform pressure at all points on the workpiece. See also *fluid forming* and *rubber-pad forming*.

fluid forming. A modification of the Guerin process, fluid forming differs from the fluid-cell process in that the die cavity, called a pressure dome, is not completely filled with rubber but with hydraulic fluid retained by a cup-shaped rubber diaphragm.

fluting. A series of sharp parallel kinks or creases that can occur when sheet steel is formed clylindrically. Fluting is caused by inhomogeneous yielding of these sheets.

foil. Metal in sheet form less than 0.15 mm (0.006 in.) thick.

fold. A forging defect caused by folding metal back onto its own surface during its flow in the die cavity.

forgeability. Term used to describe the relative ability of material to deform without fracture. Also describes the resistance to flow from deformation.

forging. The process of working metal to a desired shape by impact or pressure in hammers, forging machines (upsetters), presses, rolls, and related forming equipment. Forging hammers, counterblow equipment, and high-energy-rate forging machines apply impact to the workpiece, while most other types of forging equipment apply squeeze pressure in shaping the stock. Some metals can be forged at room temperature, but most are made more plastic for forging by heating.

forging dies. Forms for making forgings; they generally consist of a top and bottom die. The simplest will form a completed forging in a single impression; the most complex, consisting of several die inserts, may have a number of impressions for the progressive working of complicated shapes. Forging dies are usually in pairs, with part of the impression in one of the blocks and the rest of the impression in the other block.

forging plane. In forging, the plane that includes the principal die face and is perpendicular to the direction of ram travel. When the parting surfaces of the dies are flat, the forging plane coincides with the parting line. Contrast with *parting plane*.

forging quality. Term used to describe stock of sufficient quality to make it suitable for commercially satisfactory forgings.

forging stock. A wrought rod, bar, or other section suitable for subsequent change in cross section by forging.

formability. The ease with which a metal can be shaped through plastic deformation. Evaluation of the formability of a metal involves measurement of strength, ductility, and the amount of deformation required to cause fracture. The term *workability* is used interchangeably with formability; however, formability refers to the shaping of sheet metal, while workability refers to shaping materials by *bulk forming*. See also *forgeability*.

formability-limit diagram. An empirical curve showing the biaxial strain levels beyond which failure may occur in sheet metal forming. The strains are given in terms of major and minor strains measured from deformed circles, previously printed onto the sheet.

formability parameters. Material parameters that can be used to predict the ability of sheet metal to be formed into a useful shape.

forming. The plastic deformation of a billet or a blanked sheet between tools (dies) to obtain the final configuration. Metalforming processes are typically classified as *bulk forming* and *sheet forming*. Also referred to as metalworking.

forming-limit diagram (FLD) or forming-limit curve (FLC). An empirical curve in which the major strains at the onset of necking in sheet metal are plotted vertically and the corresponding minor strains are plotted horizontally. The onset-of-failure line divides all possible strain combinations into two zones: the safe zone (in which failure during forming is not expected) and the failure zone (in which failure during forming is expected).

form rolling. Hot rolling to produce bars having contoured cross sections; not to be confused with the *roll forming* of sheet metal or with *roll forging*.

forward extrusion. Same as direct extrusion. See also *extrusion*.

fracture. The irregular surface produced when a piece of metal is broken.

fracture criterion. A mathematical relationship among stresses, strains, or a combination of stresses and strains that predicts the occurrence of ductile fracture. Should not be confused with fracture mechanics equations,

which deal with more brittle types of fracture.

fracture-limit line. An experimental method for predicting surface fracture in plastically deformed solids. Is related to the forming-limit diagram used to predict failures in sheet forming.

fracture load. The load at which splitting occurs.

fracture-mechanism map. Strain rate/temperature map that describes regimes under which different damage and failure mechanisms are operative under either forming or service conditions.

fracture strain (ε_f). The true strain at fracture defined by the relationship:

$$\varepsilon_f = \ln\left[\frac{\text{Initial cross} - \text{sectional area}}{\text{Final cross} - \text{sectional area}}\right]$$

fracture strength. The engineering stress at fracture, defined as the load at fracture divided by the original cross-sectional area. The fracture strength is synonymous with the breaking strength.

fracture stress. The true stress at fracture, which is the load for fracture divided by the final cross-sectional area.

fracture toughness. A generic term for measures of resistance to extension of a crack. The term is sometimes restricted to results of fracture mechanics tests, which are directly applicable in fracture control. However, the term commonly includes results from simple tests of notched or precracked specimens not based on fracture mechanics analysis. Results from tests of the latter type are often useful for fracture control, based on either service experience or empirical correlations with fracture mechanics tests. See also *stress-intensity factor*.

free bending. A bending operation in which the sheet metal is clamped at one end and wrapped around a radius pin. No tensile force is exerted on the ends of the sheet.

friction. The resisting force tangential to the common boundary between two bodies when, under the action of an external force, one body moves or tends to move relative to the surface of the other.

friction coextrusion. A solid core along with a tube made of a cladding material is friction extruded.

friction extrusion. A rotating round bar is pressed against a die to produce sufficient frictional heating to allow softened material to extrude through the die.

friction hill. Shape of the normal pressure-position plot that pertains to the axisymmetric and plane-strain forging of simple and complex shapes. The pressure is approximately equal to the flow stress at the edge of the forging and increases toward the center, thus producing the characteristic hilllike shape. The exact magnitude of the increase in pressure is a function of interface friction

and the diameter-to-thickness or width-to-thickness ratio of the forging.

friction shear factor. Interface friction coefficient for which the interface shear stress is taken to be proportional to the flow stress divided by $\sqrt{3}$. The proportionality constant is called the friction shear factor (or interface friction factor) and is usually denoted as m. The friction shear factor takes on values between 0 (perfect lubrication) and 1 (sticking friction) during metalworking. See also *Coulomb friction*.

Fukui cup test. A cupping test combining stretchability and drawability in which a round-nosed punch draws a circular blank into a conical-shaped die until fracture occurs at the nose. Various parameters from the test are used as the criterion of formability.

fuzzy logic. The use of fuzzy sets in the representation and manipulation of vague information for the purpose of making decisions or taking actions. Fuzzy logic enables computers to make decisions based on information that is not clearly defined.

G

gage. (1) The thickness of sheet or the diameter of wire. The various standards are arbitrary and differ with regard to ferrous and nonferrous products as well as sheet and wire. (2) An aid for visual inspection that enables an inspector to determine more reliably whether the size or contour of a formed part meets dimensional requirements.

gage length. The original length of that part of a test specimen over which strain or other characteristics are measured.

galling. A condition whereby excessive friction between high spots results in localized welding with subsequent spalling and further roughening of the rubbing surface(s) of one or both of two mating parts.

grain. An individual crystal in a polycrystalline metal or alloy.

grain boundary. A narrow zone in a metal or ceramic corresponding to the transition from one crystallographic orientation to another, thus separating one grain from another; the atoms in each grain are arranged in an orderly pattern.

grain-boundary sliding. The sliding of grains past each other that occurs at high temperature. Grain-boundary sliding is common under creep conditions in service, thus leading to internal damage (e.g., cavities) or total failure, and during superplastic forming, in which undesirable cavitation may also occur if diffusional or deformation processes cannot accommodate the sliding at a sufficient rate.

grain growth. The increase in the average size of grains in a crystalline aggregate during annealing (static conditions) or deformation (dynamic conditions). The driving force for grain growth is the reduction in total grain-boundary

area and its associated surface energy.

grain size. A measure of the area or volume of grains in a polycrystalline material, usually expressed as an average when the individual sizes are fairly uniform.

gridding. Imprinting an array of repetitive geometrical patterns on a sheet prior to forming for subsequent determination of deformation. Imprinting techniques include: (1) Electrochemical marking (also called electrochemical or electrolytic etching)—a grid-imprinting technique using electrical current, an electrolyte, and an electrical stencil to etch the grid pattern into the blank surface. A contrasting oxide usually is redeposited simultaneously into the grid. (2) Photoprint—a technique in which a photosensitive emulsion is applied to the blank surface, a negative of the grid pattern is placed in contact with the blank, and the pattern is transferred to the sheet by a standard photographic printing practice. (3) Ink stamping. (4) Lithographing.

gripper dies. The lateral or clamping dies used in a forging machine or mechanical upsetter.

Guinier-Preston (G-P) zone. A small precipitation domain in a supersaturated metallic solid solution. A G-P zone has no well-defined crystalline structure of its own and contains an abnormally high concentration of solute atoms. The formation of G-P zones constitutes the first stage of precipitation and is usually accompanied by a change in properties of the solid solution in which they occur.

H

Hall-Petch dependence. A reflection of the effect of grain size on the yield strength of a metal. It states that the yield strength is inversely proportional to the square root of the grain size.

Hall-Petch relationship. A general relationship for metals that shows that the yield strength is linearly related to the reciprocal of the square root of the grain diameter.

hammering. The working of metal sheet into a desired shape over a form or on a high-speed hammer and a similar anvil to produce the required dishing or thinning.

hand forge (smith forge). A forging operation in which forming is accomplished on dies that are generally flat. The piece is shaped roughly to the required contour with little or no lateral confinement; operations involving mandrels are included. The term hand forge refers to the operation performed, while hand forging applies to the part produced.

hardness test. A test to measure the resistance to indentation of a material. Tests for sheet metal include Rockwell, Rockwell Superficial, Tukon, and Vickers.

Hartmann lines. See *Lüders lines*.

heading. The *upsetting* of wire, rod, or bar stock in dies to form parts that usually

contain portions that are greater in cross-sectional area than the original wire, rod, or bar.

healed-over scratch. A scratch that occurred during previous processing and was partially obliterated in subsequent rolling.

hemming. A bend of 180° made in two steps. First, a sharp-angle bend is made; next, the bend is closed using a flat punch and a die.

high-angle boundary. Boundary separating adjacent grains whose misorientation is at least 15°.

high-energy-rate forging. The production of forgings at extremely high ram velocities resulting from the sudden release of a compressed gas against a free piston. Forging is usually completed in one blow. Also known as HERF processing, high-velocity forging, and high-speed forging.

high-energy-rate forming. A group of forming processes that applies a high rate of strain to the material being formed through the application of high rates of energy transfer. See also *high-energy-rate forging* and *electromagnetic forming*.

hodograph. A curve traced in the course of time by the tip of a vector representing some physical quantity. In particular, the path traced by the velocity vector of a given particle.

hole expansion test. A formability test in which a tapered punch is forced through a punched or a drilled and reamed hole, forcing the metal in the periphery of the hole to expand in a stretching mode until fracture occurs.

hole flanging. The forming of an integral collar around the periphery of a previously formed hole in a sheet metal part.

homogenization. Heat treatment used to reduce or eliminate nonuniform chemical composition that develops on a microscopic scale (microsegregation) during the solidification processing of ingots and castings. Homogenization is commonly used for aluminum alloys and nickel-base superalloys.

Hooke's law. A generalization applicable to all solid material, which states that stress is directly proportional to strain and is expressed as:

$$\text{Stress/strain} = \sigma/\varepsilon = \text{Constant} = E$$

where E is the modulus of elasticity (Young's modulus). The constant relationship between stress and strain applies only below the proportional limit.

hot forming. See *hot working*. Similar to hot sizing, however, the forming is done at temperatures above the annealing temperature, and deformation is usually larger.

hot isostatic pressing (HIP). A process for simultaneously heating and forming a powder metallurgy compact in which metal powder, contained in a sealed flexible mold, is subjected to equal pressure from all direc-

tions at a temperature high enough for full consolidation to take place. Hot isostatic pressing is also frequently used to seal residual porosity in castings and to consolidate metal-matrix composites. A process that subjects a component (casting, powder forgings, etc.) to both elevated temperature and isostatic gas pressure in an autoclave. The most widely used pressurizing gas is argon. When castings are hot isostatically pressed, the simultaneous application of heat and pressure virtually eliminates internal voids and microporosity through a combination of plastic deformation, creep, and diffusion.

hot rolled sheet. Steel sheet reduced to required thickness at a temperature above the point of scaling and therefore carrying hot mill oxide. The sheet may be flattened by cold rolling without appreciable reduction in thickness or by roller leveling, or both. Depending on the requirements, hot rolled sheet can be pickled to remove hot mill oxide and is so produced when specified.

hot shortness. A tendency for some alloys to separate along grain boundaries when stressed or deformed at temperatures near the melting point. Hot shortness is caused by a low-melting constituent, often present only in minute amounts, that is segregated at grain boundaries.

hot size. A process where a preformed part is placed into a hot die above the annealing temperature to set the shape and remove springback tendencies.

hot strip or pickle line scratch. Scratches that are superficially similar to slivers or skin laminations but originate from mechanical scoring of the strip in the hot mill, pickle line, or slitter.

hot upset forging. A *bulk forming* process for enlarging and reshaping some of the cross-sectional area of a bar, tube, or other product form of uniform (usually round) section. It is accomplished by holding the heated forging stock between grooved dies and applying pressure to the end of the stock, in the direction of its axis, by the use of a heading tool, which spreads (upsets) the end by metal displacement. Also called hot heading or hot upsetting. See also *heading* and *upsetting*.

hot working. The plastic deformation of metal at such a temperature and strain rate that recrystallization or a high degree of recovery takes place simultaneously with the deformation, thus avoiding any *strain hardening*. Also referred to as hot forging and hot forming. Contrast with *cold working*.

hub. A *boss* that is in the center of a forging and forms a part of the body of the forging.

hubbing. The production of die cavities by pressing a male master plug, known as a *hub*, into a block of metal.

hydrodynamic lubrication. A system of lubrication in which the shape and relative motion of the sliding surfaces causes the formation of a liquid film having sufficient pressure to separate the surfaces. See also *elastohydro-dynamic lubrication.*

hydrostatic extrusion. A method of extruding a *billet* through a die by pressurized fluid instead of the ram used in conventional *extrusion*.

hydrostatic stress. The average value of the three normal stresses. The hydrostatic stress is a quantity that is invariant relative to the orientation of the coordinate system in which the stress state is defined.

I

IACS. See *percent IACS (%IACS)*.

impact extrusion. The process (or resultant product) in which a punch strikes a slug (usually unheated) in a confining die. The metal flow may be either between punch and die or through another opening. The impact extrusion of unheated slugs is often called cold extrusion.

impact line. A blemish on a drawn sheet metal part caused by a slight change in metal thickness. The mark is called an impact line when it results from the impact of the punch on the blank; it is called a recoil line when it results from transfer of the blank from the die to the punch during forming, or from a reaction to the blank being pulled sharply through the draw ring.

impact strength. A measure of the resiliency or toughness of a solid. The maximum force or energy of a blow (given by a fixed procedure) that can be withstood without fracture, as opposed to fracture strength under a steady applied force.

impression. A cavity machined into a forging die to produce a desired configuration in the workpiece during forging.

inclusion. A physical and mechanical discontinuity occurring within a material or part, usually consisting of solid, encapsulated foreign material. Inclusions are often capable of transmitting some structural stresses and energy fields, but to a noticeably different degree than from the parent material.

increase in area. An indicator of sheet metal forming severity based on percentage increase in surface area measured after forming.

ingot. A casting intended for subsequent rolling, forging, or extrusion.

ingot conversion. A primary metalworking process that transforms a cast ingot into a wrought mill product.

ingot metallurgy. A processing route consisting of casting an ingot that is subsequently converted into mill products via deformation processes.

inoculant. Materials that, when added to molten metal, modify the structure and thus change the physical and mechanical properties to a degree not explained on the basis of the change in composition resulting from their use. Ferrosilicon-base alloys are commonly used to inoculate gray irons and ductile irons.

intellectual property. Knowledge-based property, usually represented by patents, copyrights, trademarks, or trade secrets.

interface heat-transfer coefficient (IHTC). Coefficient defined as the ratio of the heat flux across an interface to the difference in temperature of material points lying on either side of the interface. In bulk forming, the IHTC is usually a function of the die and workpiece surface conditions, lubrication, interface pressure, amount of relative sliding, and so on.

intermetallic alloy. A metallic alloy usually based on an ordered, stoichiometric compound (e.g., Fe_3Al, Ni_3Al, $TiAl$) and often possessing exceptional strength and environmental resistance at high temperatures, unlike conventional (less highly alloyed) disordered metallic materials.

interstitial-free steels. Steels where carbon and nitrogen are removed in the steelmaking process to very low levels, and any remaining interstitial carbon and nitrogen is tied up with small amounts of alloying elements that form carbides and nitrides, that is, titanium and niobium. Although these steels have low strength, they exhibit exceptional formability.

ionic bond. (1) A type of chemical bonding in which one or more electrons are transferred completely from one atom to another, thus converting the neutral atoms into electrically charged ions. These ions are approximately spherical and attract each other because of their opposite charges. (2) A primary bond arising from the electrostatic attraction between two oppositely charged ions.

ironing. An operation used to increase the length of a tube or cup through reduction of wall thickness and outside diameter, the inner diameter remaining unchanged.

isostatic pressing. A process for forming a powder metallurgy compact/metal-matrix composite or for sealing casting porosity by applying pressure equally from all directions. See also *hot isostatic pressing (HIP)*.

isothermal forging. A hot forging process in which a constant and uniform temperature is maintained in the workpiece during forging by heating the dies to the same temperature as the workpiece.

isothermal transformation (IT) diagram. A diagram that shows the isothermal time required for transformation of austenite to begin and to finish as a function of temperature.

isotropy. The condition in which the properties are independent of the direction in which they are measured.

J

J-integral. A mathematical expression involving a line or surface integral that encloses the crack front from one crack surface to the other, used to characterize the *fracture*

toughness of a material having appreciable plasticity before fracture. The *J*-integral eliminates the need to describe the behavior of the material near the crack tip by considering the local stress-strain field around the crack front; J_{Ic} is the critical value of the *J*-integral required to initiate growth of a preexisting crack.

Joffe effect. Change in mechanical properties, especially the fracture strength, resulting from testing in an environment that modifies the surface characteristics of the material.

K

Keeler-Goodwin diagram. The *forming-limit diagram* for low-carbon steel commonly used for sheet metal forming.

kinetics. Term describing the rate at which a metallurgical process (e.g., recovery, recrystallization, grain growth, phase transformation) occurs as a function of time or, if during deformation, of strain.

kinks. Sharp bends or buckles caused by localized plastic deformation of a sheet.

klink. An internal crack caused by too rapid heating of a large workpiece.

knockout. A mechanism for releasing workpieces from a die.

knockout mark. A small protrusion, such as a button or ring of flash, resulting from depression of the *knockout pin* from the forging pressure or the entrance of metal between the knockout pin and the die.

knockout pin. A power-operated plunger installed in a die to aid removal of the finished forging.

Kronecker symbol. A second-order tensor, δ_{ij}. $\delta_{ij} = 1$ for $i = j$; $\delta_{ij} = 0$ for $i \neq j$.

L

lancing. Cutting along a line in the workpiece without detaching a slug from the blank.

laser cutting. A cutting process that severs material with the heat obtained by directing a laser beam against a metal surface. The process can be used with or without an externally supplied shielding gas.

lateral extrusion. An operation in which the product is extruded sideways through an orifice in the container wall.

lattice. A regular geometrical arrangement of points in space.

lattice constants. See *lattice parameter*.

lattice parameter. The length of any side of a unit cell of a given crystal structure. The term is also used for the fractional coordinates x, y, z of lattice points when these are variable.

leveling. The flattening of rolled sheet, strip, or plate by reducing or eliminating distortions.

limiting dome height (LDH) test. A mechanical test, usually performed unlubricated on sheet metal, that simulates the fracture conditions in a practical press-forming opera-

tion. The results are dependent on the sheet thickness.

limiting drawing ratio (LDR). See *deformation limit*.

linear elastic fracture mechanics. A method of fracture analysis that can determine the stress (or load) required to induce fracture instability in a structure containing a crack-like flaw of known size and shape. See also *stress-intensity factor*.

liners. Thin strips of metal inserted between the dies and the units into which the dies are fastened.

loose metal. A defect in an area of a stamping where very little contour is present. The metal in the area has not been stretched, resulting in a shape with no stiffness. The area may have waves in it or may sag so that there is a dish in an area that is intended to be flat or nearly flat. This defect differs from oil canning in that the metal cannot be snapped back into the desired shape when a load is removed or reversed on the area.

low-angle boundary. Boundary separating adjacent grains whose misorientation is less than 15°. See also *subgrain*.

lubricant. A material applied to dies, molds, plungers, or workpieces that promotes the flow of metal, reduces friction and wear, and aids in the release of the finished part.

lubricant residue. The carbonaceous residue resulting from lubricant that is burned onto the surface of a hot forged part.

Lüders lines. Elongated surface markings or depressions, often visible with the unaided eye, that form along the length of a round or sheet metal tension specimen at an angle of approximately 55° to the loading axis. Caused by localized plastic deformation, they result from discontinuous (inhomogeneous) yielding. Also known as Lüders bands, Hartmann lines, Piobert lines, or stretcher strains.

lumped-parameter model. A mathematical model in which the distributed properties of physical quantities are replaced with their lumped equivalents. When a problem can be analyzed in terms of a finite number of discrete elements, it can be expressed by ordinary differential equations. To describe the more realistic case of distributed parameters having many values spread over a field in space requires the use of partial differential equations.

M

machinability. The relative ease with which material is removed from a solid by controlled chip-forming in a machining process.

major strain. Largest principal strain in the sheet surface. Often measured from the major axis of the ellipse resulting from deformation of a circular grid.

malleability. The characteristic of metals that permits <M> plastic deformation in compression without fracture.

mandrel. (1) A blunt-ended tool or rod used to retain the cavity in a hollow metal product during working. (2) A metal bar around which other metal can be cast, bent, formed, or shaped. (3) A shaft or bar for holding work to be machined.

mandrel forging. The process of rolling or forging a hollow blank over a mandrel to produce a weldless, seamless ring or tube.

Mannesmann process. A process for piercing tube billets in making seamless tubing. The billet is rotated between two heavy rolls mounted at an angle and is forced over a fixed mandrel.

Marforming process. A *rubber-pad forming* process developed to form wrinkle-free shrink flanges and deep-drawn shells. It differs from the Guerin process in that the sheet metal blank is clamped between the rubber pad and the blankholder before forming begins.

martensite. A generic term for microstructures formed by diffusionless phase transformation in which the parent and product phases have a specific crystallographic relationship. Martensite is characterized by an acicular pattern in the microstructure in both ferrous and nonferrous alloys. In alloys where the solute atoms occupy interstitial positions in the martensitic lattice (such as carbon in iron), the structure is hard and highly strained; but where the solute atoms occupy substitutional positions (such as nickel in iron), the martensite is soft and ductile. The amount of high-temperature phase that transforms to martensite on cooling depends to a large extent on the lower temperature attained, there being a rather distinct beginning temperature (M_s) and a temperature at which the transformation is essentially complete (M_f).

mass-conserving process. A manufacturing process in which the mass of the starting material is approximately equal to the mass of the final product or part. Examples are casting, precision forming, and powder processes.

match. A condition in which a point in one die half is aligned properly with the corresponding point in the opposite die half, within specified tolerance.

matched edges (match lines). Two edges of the die face that are machined exactly at 90° to each other and from which all dimensions are taken in laying out the die impression and aligning the dies in the forging equipment.

material heat. The pedigree of the starting stock or billet used to make a forging.

matrix phase. The continuous (interconnected) phase in an alloy with two or more phases. In cast or wrought materials, the matrix phase is often comprised of the first phase to solidify.

mechanical texture. Directionality in the shape and orientation of microstructural features such as inclusions, grains, and so on.

mechanical working. The subjecting of material to pressure exerted by rolls, hammers, or

presses in order to change the shape or physical properties of the material.

mechanistic modeling. An approach that requires a fundamental understanding of the physics and chemistry governing the process. These laws and principles are used to describe the process and its parameters. The results are then validated against controlled test results.

mesh. (1) The number of screen openings per linear inch of screen; also called mesh size. (2) The screen number on the finest screen of a specified standard screen scale through which almost all of the particles of a powder sample will pass.

metadata. Descriptive data about the material for which data are reported. Metadata include a complete description of the material (producer, heat number, grade, temper, etc.), a complete description of the test method, and information about the test plan.

metal. An opaque lustrous elemental chemical substance that is a good conductor of heat and electricity and, when polished, a good reflector of light. Most elemental metals are malleable and ductile and are, in general, denser than other elemental substances. As to structure, metals may be distinguished from nonmetals by their atomic binding and electron availability. Metallic atoms tend to lose electrons from the outer shells, the positive ions thus formed being held together by the electron gas produced by the separation. The ability of these "free electrons" to carry an electric current, and the fact that this ability decreases as temperature increases, establish the prime distinctions of a metallic solid.

microalloyed steel. A low-to-medium-carbon steel usually containing small alloying additions of niobium, vanadium, nitrogen, and so on whose thermomechanical processing is controlled to obtain a specific microstructure and thus a suite of properties.

microhardness test. An indentation test using diamond indentors at very low loads, usually in the range of 1 to 1000 g.

microstructure. The structure of polished and etched metals as revealed by a microscope.

mill finish. A nonstandard (and typically nonuniform) surface finish on mill products that are delivered without being subjected to a special surface treatment (other than a corrosion-preventive treatment) after the final working or heat treating step.

mill product. Any commercial product of a mill.

mill scale. The heavy oxide layer that forms during the hot fabrication or heat treatment of metals.

minimum bend radius. The smallest radius about which a metal can be bent without exhibiting fracture. It is often described in terms of multiples of sheet thickness.

minor strain. The principal strain in the sheet surface in a direction perpendicular to the major strain. Often measured from the minor axis of the ellipse resulting from deformation of a circular grid.

mischmetal. From the German *mischmetall*, with roots mischen (to mix) and metall (metal), it is a natural mixture of rare earth metals containing approximately 50 wt% Ce, 25% La, 15% Nd, and 10% other rare earth metals, iron, and silicon. It is commonly used to make rare earth additions to alloys (e.g., magnesium alloys), rather than using more expensive pure forms of the rare earth metals.

mismatch. The misalignment or error in register of a pair of forging dies; also applied to the condition of the resulting forging. The acceptable amount of this displacement is governed by blueprint or specification tolerances. Within tolerances, mismatch is a condition; in excess of tolerance, it is a serious defect. Defective forgings can be salvaged by hot reforging operations.

misorientation. Angular difference between the orientations of two grains adjacent to a grain boundary, between a twin and its parent matrix, and so on.

mixed dislocation. Any combination of a *screw dislocation* and an *edge dislocation*.

model. A physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process.

modeling. Application of a standard, rigorous, structured methodology to create and validate a physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process.

modulus of elasticity (E). The measure of rigidity or stiffness of a metal; the ratio of stress, below the proportional limit, to the corresponding strain. In terms of the stress-strain diagram, the modulus of elasticity is the slope of the *stress-strain curve* in the range of linear proportionality of stress to strain. Also known as *Young's modulus*. For materials that do not conform to *Hooke's law* throughout the elastic range, the slope of either the tangent to the stress-strain curve at the origin or at low stress, the secant drawn from the origin to any specified point on the stress-strain curve, or the chord connecting any two specific points on the stress-strain curve is usually taken to be the modulus of elasticity. In these cases, the modulus is referred to as the *tangent modulus, secant modulus, or chord modulus*, respectively.

modulus of rigidity. See *shear modulus*.

Monte Carlo modeling. Numerical modeling technique, based on statistical mechanics, that can be used to describe the migration of grain boundaries in polycrystalline aggregates during annealing or deformation processes and thus is applied to describe recrystallization, grain growth, and the accompanying evolution of texture. Also referred to as the Potts technique.

m-value. See *strain-rate sensitivity*.

N

near-net shape forging. A forging produced with a very small finish allowance over the final part dimensions and requiring some machining prior to use.

necking. (1) The reduction of the cross-sectional area of metal in a localized area by uniaxial tension or by stretching. (2) The reduction of the diameter of a portion of the length of a cylindrical shell or tube.

necklace recrystallization. Partial static or dynamic recrystallization that nucleates heterogeneously on grain boundaries in various steels, nickel-base superalloys, and so on. A microstructure of fine (necklace-like) grains lying on the original grain boundaries is thus produced.

net shape forging. A forging produced to finished part dimensions that requires little or no further machining prior to use.

neural network. Nonlinear regression-type methodology for establishing the correlation between input and output variables in a physical system. For example, neural networks can be used to correlate processing variables to microstructural features or microstructural features to mechanical properties.

neuron. A node in a neural-network system that can be considered as an internal variable and whose value is a function of the neurons in the previous layer.

Newtonian fluid. A fluid exhibiting Newtonian viscosity wherein the shear stress is proportional to the rate of shear.

no-draft (draftless) forging. A forging with extremely close tolerances and little or no draft that requires minimal machining to produce the final part. Mechanical properties can be enhanced by closer control of grain flow and by retention of surface material in the final component.

nominal strain. The unit elongation given by the change in length divided by the original length. Also called *engineering strain*.

nominal stress. The unit force obtained when the applied load is divided by the original cross-sectional area. Also called *engineering stress*.

nonfill (underfill). A forging condition that occurs when the die impression is not completely filled with metal.

normal anisotropy. A condition in which a property or properties in the sheet thickness direction differ in magnitude from the same property or properties in the plane of the sheet.

normal distribution. The *probability density function* used to describe the various properties of materials and the distribution of most random variables encountered in engineering design.

notching. An unbalanced shearing or blanking operation in which cutting is done around only three sides (usually) of a punch.

n-value. A term commonly referred to as the strain-hardening exponent derived from the relationship between true stress, σ, with true strain, ε, given by $\sigma = K\varepsilon^n$. Preferably called the *strain-hardening exponent.*

O

objective function. Mathematical function describing a desired material or process characteristic whose optimization is the goal of process design. In bulk forming, typical objective functions may include forging weight (minimum usually is best), die fill (minimum underfill is best), and uniformity of strain or strain rate (maximum uniformity is best). In design optimization, it is the grouping of design parameters that is attempted to be maximized or minimized, subject to the problem constraints. Also known as criterion function.

offal. Sheet metal section trimmed or removed from the sheet during the production of shaped blanks or the formed part. Offal is frequently used as stock for the production of small parts.

offset. The distance along the strain coordinate between the initial portion of a stress-strain curve and a parallel line that intersects the stress-strain curve at a value of stress (commonly 0.2%) that is used as a measure of the *yield strength.* Used for materials that have no obvious *yield point.*

offset yield strength. The stress at which the strain exceeds by a specified amount (the *offset*) an extension of the initial proportional portion of the stress-strain curve; expressed in force per unit area.

Olsen ductility test. A *cupping test* in which a piece of sheet metal, restrained except at the center, is deformed by a standard steel ball until fracture occurs. The height of the cup at the time of fracture is a measure of the ductility.

open-die forging. The hot mechanical forming of metals between flat or shaped dies in which metal flow is not completely restricted. Also known as hand or smith forging. See also *hand forge (smith forge).*

optimization. The process of searching for the best combination of design parameters. Design optimization suggests that, for a given set of possible designs and design criteria, there exists a single design that is best or optimal.

orange peel. In metals, a surface roughening in the form of a pebble-grained pattern that occurs when a metal of unusually coarse grain size is stressed beyond its elastic limit.

orientation-distribution function (ODF). Mathematical function describing the normalized probability of finding grains of given crystallographic orientations/Euler angles. Because crystallographic orientations are in terms of Euler angles, the description of texture using ODFs is unambiguous, unlike pole figures. See also *texture, preferred orientation*, and *pole figure.*

Ostwald ripening. The increase in the average size of second-phase particles, accompanied by the reduction in their number, during annealing, deformation, or high-temperature service exposure. Ostwald ripening leads to a decrease in the total surface energy associated with matrix-particle interfaces. Also known as *coarsening.*

outliers. Observed values much lower or higher than most other observations in a data set. Values identified as outliers should be investigated. Data should be screened for outliers, because they have substantial influence on statistical analysis.

oxidation. A reaction where there is an increase in valence resulting from a loss of electrons. Such a reaction occurs when most metals or alloys are exposed to atmosphere and the reaction rate increases as temperature increases.

oxide-dispersion-strengthened (ODS) alloys. A class of materials in which fine oxide particles are incorporated in metal powders, compacted, and then fabricated into finished forms by deformation processing. The resulting material has improved thermal softening resistance with excellent thermal and electrical conductivity. Examples are ODS copper alloys and sintered aluminum powder.

oxidized surface. A tightly adhering oxide surface layer that results in modified surface color and reduced reflectivity. It is often accompanied by surface penetration of oxide that causes brittleness.

P

pack rolling. Hot, flat rolling process in which the workpiece (or a stack of workpieces) in the form of plate, sheet, or foil is encased in a sacrificial can to reduce/eliminate contamination (e.g., oxygen pickup) or poor workability due to roll chill.

pancake forging. A rough forged shape, usually flat, that can be obtained quickly with minimal tooling. Usually made by upsetting a cylindrical billet to a large height reduction in flat dies. Considerable machining is usually required to attain the finish size.

parting. A shearing operation used to produce two or more parts from a stamping.

parting line. The line along the surface of a forging where the dies meet, usually at the largest cross section of the part. *Flash* is formed at the parting line.

parting plane. The plane that includes the principal die face and is perpendicular to the direction of ram travel. When parting surfaces of the dies are flat, the parting plane coincides with the parting line. Also referred to as the forging plane.

pass. (1) A single transfer of metal through a stand of rolls. (2) The open space between two grooved rolls through which metal is processed.

peak count. In surface measurements, the number of asperities above a given (defined) height cut-off level and within a given width cut-off. Frequency is taken at 50 μin./in.

peak density. The average number of peaks within the specified cut-off levels.

peak height. Peak-to-valley magnitude as measured by a suitable stylus instrument. Peak height is related to roughness height, depending on uniformity of surface irregularities.

peen forming. A dieless, flexible-manufacturing technique used primarily in the aerospace industry for forming sheet metals by way of the deformation imparted by the controlled-velocity impact of balls.

percent IACS (%IACS). In 1913, values of electrical conductivity were established and expressed as a percent of a standard. The standard chosen was an annealed copper wire with a density of 8.89 g/cm³, a length of 1 m, a weight of 12 g, with a resistance of 0.1532 Ω at 20 °C (70 °F). The 100% IACS (International Annealed Copper Standard) value was assigned with a corresponding resistivity of 0.017241 Ωmm²/m. The percent IACS for any material can be calculated by % IACS = 0.017241 Ωmm²/m × 100/volume resistivity.

perforating. The punching of many holes, usually identical and arranged in a regular pattern, in a sheet, workpiece blank, or previously formed part. The holes are usually round but may be any shape. The operation is also called multiple punching. See also *piercing.*

permanent set. The deformation or strain remaining in a previously stressed body after release of the load.

permeability (magnetic). A general term used to express various relationships between magnetic induction and magnetizing force. These relationships are either "absolute permeability," which is a change in magnetic induction divided by the corresponding change in magnetizing force, or "specific (relative) permeability," the ratio of the absolute permeability to the permeability of free space.

phenomenological model. Empirical or data-based modeling that relies on collecting data from observations, specifying the correlation structure between variables, using numerical techniques to find parameters for the structure such that the correlation between the data is maximized, and then validating the model against a test data set.

phonon (wave). An organized movement of atoms or molecules, such as a sound wave.

pickup. Small particles of oxidized metal adhering to the surface of a *mill* product.

piercing. The general term for cutting (shearing or punching) openings, such as holes and slots, in sheet material, plate, or parts. This

operation is similar to blanking; the difference is that the slug or piece produced by piercing is scrap, while the blank produced by *blanking* is the useful part.

pinchers. Surface defects having the appearance of ripples or of elongated areas of variable surface texture extending at an acute angle to the sheet rolling direction and often branching. Pinchers are usually caused by poor hot rolled strip shape or improper drafting or incorrect crown on the cold reduction mill. They are sometimes referred to as feather pattern.

pinning. The retardation or complete cessation of grain growth during annealing or deformation by second-phase particles acting on grain boundaries.

Piobert lines. See *Lüders lines*.

pit. A small, clean depression in a sheet surface caused by the rolling-in of foreign particles such as sand, steel, and so on that subsequently fall out.

planar anisotropy. A term indicating variation in one or more physical or mechanical properties with direction in the plane of the sheet. The planar variation in plastic strain ratio is commonly designated as Δr, given by $\Delta r = (r_0 + r_{90} - 2r_{45})/2$. The earing tendency of a sheet is related to Δr. As δr increases, so does the tendency to form ears.

plane strain. Deformation in which the normal and shear components associated with one of the three coordinate directions are equal to zero. Bulk forming operations that approximate plane-strain conditions include sheet rolling and sheet drawing.

plane stress. Stress state in which the normal and shear components of stress associated with one of the three coordinate directions are equal to zero. Most sheetforming operations are performed under conditions approximating plane stress.

planishing. Smoothing a metal surface by rolling, forging, or hammering; usually the last pass or passes of a shaping operation.

plastic deformation. The permanent (inelastic) distortion of metals under applied stresses that strain the material beyond its *elastic limit*. The ability of metals to flow in a plastic manner without fracture is the fundamental basis for all metalforming processes.

plastic flow. The phenomenon that takes place when metals or other substances are stretched or compressed permanently without rupture.

plastic instability. The deformation stage during which plastic flow is nonuniform and necking occurs.

plasticity. The property of a material that allows it to be repeatedly deformed without rupture when acted upon by a force sufficient to cause deformation and that allows it to retain its shape after the applied force has been removed.

plastic-strain ratio (*r*-value). A measure of normal plastic anisotropy is defined by the ratio of the true width strain to the true thickness strain in a tensile test. The average plastic strain ratio, r_m, is determined from tensile samples taken in at least three directions from the sheet rolling direction, usually at 0, 45, and 90°. The r_m is calculated as $r_m = (r_0 + 2r_{45} + r_{90})/4$. The ratio of the true width strain to the true thickness strain in a sheet tensile test is $r = \varepsilon_w/\varepsilon_t$. A formability parameter that relates to drawing, it is also known as the anisotropy factor. A high r-value indicates a material with good drawing properties.

ploughing. Plastic deformation of the surface of the softer component of a friction pair.

point lattice. A set of points in space located so that each point has identical surroundings. There are 14 ways of so arranging points in space, corresponding to the 14 *Bravais lattices*.

Poisson's ratio (*ν*). The absolute value of the ratio of transverse (lateral) strain to the corresponding axial strain resulting from uniformly distributed axial stress below the *proportional limit* of the material in a tensile test.

pole figure. Description of crystallographic texture based on a stereographic-projection representation of the times-random probability of finding a specific crystallographic pole with a specific orientation relative to sample reference directions. For axisymmetric components, the sample reference directions are usually the axis and two radial directions; for a sheet material, the rolling, transverse, and sheet-normal directions are used. Because pole figures provide information only with regard to the orientation of one crystallographic pole, several pole figures or an orientation-distribution function (derivable from pole-figure measurements) are needed to fully describe crystallographic texture. See also *orientation-distribution function*.

polycrystalline aggregate. The collection of grains/crystals that form a metallic material.

polygonization. A recovery-type process during the annealing of a worked material in which excess dislocations of a given sign rearrange themselves into low-energy, low-angle tilt boundaries.

population. A statistical concept describing the total set of objects or observations under consideration.

porosity. Fine holes or voids within a solid; the amount of these pores is expressed as a percentage of the total volume of the solid.

postforming. Any treatment after the part has been formed, such as annealing, trimming, finishing, and so on.

powder forging. The plastic deformation of a powder metallurgy *compact or preform* into a fully dense finished shape by using compressive force; usually done hot and within closed dies.

powder metallurgy (PM). The technology and art of producing metal powders and using metal powders for production of mass materials and shaped objects.

precipitation hardening. Hardening in metals caused by the precipitation of a constituent from a supersaturated solid solution.

precision. In testing, a measure of the variability that can be expected among test results. The precision of an instrument indicates its ability to reproduce a certain reading. Precision is the inverse of standard deviation. A decrease in the scatter of test results is represented by a smaller standard deviation, leading directly to an increase in precision.

precision forging. A forging produced to closer tolerances than normally considered standard by the industry.

preferred orientation. Nonrandom distribution of the crystallographic orientations of the grains comprising a polycrystalline aggregate.

preform. (1) The forging operation in which stock is preformed or shaped to a predetermined size and contour prior to subsequent die forging operations. When a preform operation is required, it will precede a forging operation and will be performed in conjunction with the forging operation and in the same heat. (2) The initially pressed powder metallurgy *compact* to be subjected to *repressing*.

prelubed sheet. A sheet or coil that has had a lubricant applied during mill processing to serve as a forming lubricant in the fabrication plant.

press forging. The forging of metal between dies by mechanical or hydraulic pressure; usually accomplished with a single work stroke of the press for each die station.

press forming. Any sheet metal forming operation performed with tooling by means of a mechanical or hydraulic press.

pressing. The product or process of shallow drawing sheet or plate.

press load. The amount of force exerted in a given forging or forming operation.

principal strain. The normal strain on any of three mutually perpendicular planes on which no shear strains are present.

principal strain direction. The direction of action of the normal strains.

principal stress. One of the three normal stresses in the coordinate system in which all of the shear stresses are equal to zero.

prior particle boundary (PPB). An apparent boundary between the pre-existing powder metal particles that is still evident within the microstructure of consolidated powder metallurgy products because of the presence of carbide or other phases that form at these boundaries.

probability density function (PDF). A mathematical function that, when integrated between two limits, gives the probability that a random variable assumes a value between these limits.

processing map. A map of strain rate versus temperature that delineates the regions that should be avoided in processing to prevent

the formation of poor microstructures or voids or cracks. These maps are generally created by the dynamic material modeling method or by mapping extensive results of processing experience.

process model. A mathematical description of the physical behavior underlying a manufacturing process that is used to predict performance of the process in terms of operating parameters. Most often, process models are reduced to software and are manipulated with computers.

process modeling. Computer simulation of deformation, heat treating, and machining processes for the purpose of improving process yield and material properties.

profile (contour) rolling. In *ring rolling*, a process used to produce seamless rolled rings with a predesigned shape on the outside or the inside diameter, requiring less volume of material and less machining to produce finished parts.

progressive die. A die planned to accomplish a sequence of operations as the strip or sheet of material is advanced from station to station, manually or mechanically.

progressive forming. Sequential forming at consecutive stations with a single die or separate dies.

projection welding. Electrical resistance welding in which the welds are localized at embossments or other raised portions of the sheet surface.

proof load. A predetermined load, generally some multiple of the service load, to which a specimen or structure is submitted before acceptance for use.

proof stress. The stress that will cause a specified small permanent set in a material. A specified stress to be applied to a member or structure to indicate its ability to withstand service loads.

proportional limit. The greatest stress a material is capable of developing without a deviation from straight-line proportionality between stress and strain. See also *elastic limit* and *Hooke's law*.

punch. (1) The male part of a die—as distinguished from the female part, which is called the die. The punch is usually the upper member of the complete die assembly and is mounted on the *slide* or in a *die set* for alignment (except in the inverted die). (2) In double-action draw dies, the punch is the inner portion of the upper die, which is mounted on the plunger (inner slide) and does the drawing. (3) The act of piercing or *punching* a hole. Also referred to as punching.

punching. The die shearing of a closed contour in which the sheared-out sheet metal part is scrap.

punch nose radius. The shape of the punch end, contacting the material being formed to allow proper material flow or movement.

punch press. (1) In general, any mechanical press. (2) In particular, an endwheel gap-frame press with a fixed bed, used in piercing.

Q

quarter hard. A temper of nonferrous alloys and some ferrous alloys characterized by tensile strength approximately midway between that of dead soft and half-hard tempers.

quench aging. Hardening by precipitation that results after the rapid cooling from solid solution to a temperature below which the elements of a second phase become supersaturated. Precipitation occurs after the application of higher temperatures and/or times and causes increases in yield strength, tensile strength, and hardness.

quenching. Rapid cooling of metals from a suitable elevated temperature, generally accomplished by immersion in water, oil, polymer solution, or salt, although forced air is sometimes used.

R

rabbit ear. Recess in the corner of a metal-forming die to allow for wrinkling or folding of the blank.

radial draw forming. The forming of sheet metals by the simultaneous application of tangential stretch and radial compression forces. The operation is done gradually by tangential contact with the die member. This type of forming is characterized by very close dimensional control.

radial forging. A process using two or more moving anvils or dies for producing shafts with constant or varying diameters along their length or tubes with internal or external variations in diameter. Often incorrectly referred to as *rotary forging*.

radial roll (main roll, king roll). The primary driven roll of the rolling mill for rolling rings in the radial pass. The roll is supported at both ends.

ram. The moving or falling part of a drop hammer or press to which one of the dies is attached; sometimes applied to the upper flat die of a steam hammer. Also referred to as the *slide*.

***R*-curve.** In fracture mechanics, a plot of crack-extension resistance as a function of stable crack extension, which is the difference between either the physical crack size or the effective crack size and the original crack size. *R*-curves normally depend on specimen thickness and, for some materials, on temperature and strain rate.

recovery. Process occurring during annealing following cold or hot working of metals in which defects such as dislocations are eliminated or rearranged by way of mechanisms such as dipole annihilation, the formation of subgrains, and subgrain growth. Recovery usually leads to a reduction in stored energy, softening, reduction or elimination of residual stresses, and, in some instances, changes in physical properties. Recovery may also serve as a precursor to

static recrystallization at sufficient levels of prior cold or hot work. See also dynamic *recovery*.

recrystallization. A process of nucleation and growth of new strain-free grains or crystals in a material. This process occurs upon heating above the recrystallization temperature (approximately 40% of the metal absolute melting temperature) during/after hot working or during annealing after cold working. Recrystallization can be dynamic (occurring during straining), static (occurring following deformation, typically during heat treatment), or metadynamic (occurring immediately after deformation due to the presence of recrystallization nuclei formed during deformation).

recrystallization texture. Crystallographic texture formed during static or dynamic recrystallization. The specific texture components that are formed are dependent on the nature of the stored work driving recrystallization and the nucleation and growth mechanisms that underlie recrystallization.

redrawing. The second and successive deep-drawing operations in which cuplike shells are deepened and reduced in cross-sectional dimensions.

reduction. (1) In cupping and deep drawing, a measure of the percentage of decrease from blank diameter to cup diameter, or of the diameter reduction in redrawing. (2) In forging, extrusion, rolling, and drawing, either the ratio of the original to the final cross-sectional area or the percentage of decrease in cross-sectional area.

reduction in area. The difference between the original cross-sectional area and the smallest area at the point of rupture in a tensile test; usually stated as a percentage of the original area.

redundant work. Energy in addition to that required for uniform flow expended during processing due to inhomogeneous deformation.

relative density. Ratio of density to pore-free density.

relief. Clearance obtained by removing material, either behind or beyond the cutting edge of a punch or die.

repressing. The application of pressure to a sintered compact; usually done to improve a physical or mechanical property or for dimensional accuracy.

rerolling quality. Rolled billets from which the surface defects have not been removed or completely removed.

residual stress. An elastic stress that exists in a solid body without an imposed external force. Residual stresses often result from forming or thermal processing and are caused by such factors as cold working, phase changes, temperature gradients, or rapid cooling.

response surface modeling. A statistical, mathematical, or graphical model that describes the variation of the response variable in terms of the parameters of the problem.

restriking. (1) The striking of a trimmed but slightly misaligned or otherwise faulty forging with one or more blows to improve alignment, improve surface condition, maintain close tolerances, increase hardness, or effect other improvements. (2) A sizing operation in which coining or stretching is used to correct or alter profiles and to counteract distortion. (3) A salvage operation following a primary forging operation in which the parts involved are rehit in the same forging die in which the pieces were last forged.

retained austenite. An amount of the high-temperature face-centered cubic phase of iron (austenite) that does not transform to martensite (is retained) when quenched to room temperature.

reverse drawing. *Redrawing* of a sheet metal part in a direction opposite to that of the original drawing.

reverse flange. A sheet metal flange made by shrinking, as opposed to one formed by stretching.

reverse redrawing. An operation after the first drawing operation in which the part is turned inside out by inverting and redrawing, usually in another die, to a smaller diameter.

rheology. The science of deformation and the flow of matter.

ring compression test. A workability test that uses the expansion or contraction of the hole in a thin compressed ring to measure the frictional conditions. The test can also be used to determine the flow stress.

ring rolling. The process of shaping weldless rings from pierced disks or shaping thick-walled ring-shaped blanks between rolls that control wall thickness, ring diameter, height, and contour.

rod. A solid round section 9.5 mm (3/8 in.) or greater in diameter whose length is great in relation to its diameter.

roll. Tooling used in the rolling process to deform material stock.

roll bending. The curving of sheets, bars, and sections by means of rolls.

rolled-in scale. Localized areas of heavy oxide not removed by the hot mill descaling sprays and rolled out to elongated streaks during further processing.

roller leveling. *Leveling* by passing flat sheet metal stock through a machine having a series of small-diameter staggered rolls that are adjusted to produce repeated reverse bending.

roll feed. A mechanism for feeding strip or sheet stock to a press or other machine. The stock passes between two revolving rolls mounted one above the other, which feed it under the dies a predetermined length at each stroke of the press. Two common types of drive are the oscillating-lever type and the rack-and-pinion type. The single-roll feed may be used to either push or pull the stock to or from the press. The double-roll feed is commonly used with wider presses (left to right) or in other cases where a single-roll feed is impractical.

roll flattening. The flattening of sheets that have been rolled in packs by passing them separately through a two-high cold mill with virtually no deformation. Not to be confused with roller leveling.

roll forging. A process of shaping stock between two driven rolls that rotate in opposite directions and have one or more matching sets of grooves in the rolls; used to produce finished parts of preforms for subsequent forging operations.

roll former. A device with three or more rolls positioned to progressively plastically form sheet or strip metal into curved or linear shapes.

roll forming. A process in which coil sheet or strip metal is formed by a series of shaped rolls into the desired configuration. Metal-forming through the use of power-driven rolls whose contour determines the shape of the product; sometimes used to denote power *spinning*.

rolling. The reduction of the cross-sectional area of metal stock, or the general shaping of metal products, through the use of rotating rolls.

rolling mandrel. In ring rolling, a vertical roll of sufficient diameter to accept various sizes of ring blanks and to exert rolling force on an axis parallel to the main roll.

rolling mills. Machines used to decrease the cross-sectional area of metal stock and to produce certain desired shapes as the metal passes between rotating rolls mounted in a framework comprising a basic unit called a stand. Cylindrical rolls produce flat shapes; grooved rolls produce rounds, squares, and structural shapes.

roll mark. A mark in light relief on the sheet surface produced by an indentation in the cold reduction mill work roll surface.

roll straightening. The straightening of metal stock of various shapes by passing it through a series of staggered rolls (the rolls usually being in horizontal and vertical planes) or by reeling in two-roll straightening machines.

roll threading. The production of threads by rolling the piece between two grooved die plates, one of which is in motion, or between rotating grooved circular rolls.

roping. A surface defect consisting of a series of generally parallel markings or ripples on areas of rolled formed sheet parts that have undergone substantial strain. The ripples are always parallel to rolling direction.

rotary forging. A process in which the work-piece is pressed between a flat anvil and a swiveling (rocking) die with a conical working face; the platens move toward each other during forging. Also called orbital forging. Compare with *radial forging*.

rotary shear. A sheet metal cutting machine with two rotating-disk cutters mounted on parallel shafts driven in unison.

rotary swager. A swaging machine consisting of a power-driven ring that revolves at high speed, causing rollers to engage cam surfaces and force the dies to deliver hammerlike blows on the work at high frequency. Both straight and tapered sections can be produced.

rotary swaging. A *bulk forming* process for reducing the cross-sectional area or otherwise changing the shape of bars, tubes, or wires by repeated radial blows with one or more pairs of opposed dies.

rough blank. A blank for a forming or drawing operation, usually of irregular outline, with necessary stock allowance for process metal, which is trimmed after forming or drawing to the desired size.

roughing stand. The first stand (or several stands) of rolls through which a reheated billet or slab passes in front of the finishing stands. See also *rolling mills.*

roughness cut-off level. Terms used in the measurement of surface roughness. (a) Width cut-off: the greatest spacing of repetitive surface irregularities used in the measurement of roughness, usually 0.030 in. (b) Height cut-off: the minimum surface irregularity in peak count determinations, usually 50 μin.

roughness height. The average height of surface irregularities with reference to a mean or nominal surface as determined by height and width cut-offs. It may be expressed as the deviation from the nominal surface, as arithmetic average, or as root mean square.

rubber forming. A sheet metal forming process in which rubber is used as a functional die part.

rubber-pad forming. A sheet metal forming operation for shallow parts in which a confined, pliable rubber pad attached to the press slide (ram) is forced by hydraulic pressure to become a mating die for a punch or group of punches placed on the press bed or baseplate. Developed in the aircraft industry for the limited production of a large number of diversified parts, the process is limited to the forming of relatively shallow parts, normally not exceeding 40 mm (1.5 in.) deep. Also known as the Guerin process. Variations of the Guerin process include the *Marforming process*, the *fluid-cell process*, and *fluid forming*.

run. The quantity produced in one setup.

r-value. The ratio of true width strain to true thickness strain. Often called *plastic-strain ratio*.

S

saddening. The process of lightly working an ingot in the initial forging operation to break up and refine the coarse, as-cast structure at the surface.

scale pattern. A transverse surface pattern on cold rolled sheet caused by intermittent

removal of the scale in the scale-breaker operation prior to pickling. The result is a pattern of overpickled areas that are not eliminated in cold reduction.

Schmid factor. In a uniaxial tension test, the geometric factor that corresponds to the product of the cosine of the angle between the tension axis and the slip-plane normal and the cosine of the angle between the tension axis and the slip direction. Often denoted as *m*.

Schmid's law. Criterion that slip in metallic crystals is controlled by a critical resolved shear stress that depends on specific material, strain rate, and test temperature but is independent of the stress normal to the slip plane.

scoring. (1) The marring or scratching of any formed part by metal pickup on the punch or die. (2) The reduction in thickness of a material along a line to weaken it intentionally along that line.

scratches. Lines generally caused by sliding of the sheet surface over sharp edges of processing equipment or over other sheets.

scratch resistance. The ability of a material to resist scratching. It is a function of the material hardness, although the lubricity of the surface will also play a part.

screw dislocation. A line imperfection that corresponds to the axis of a spiral structure in a crystal and is characterized by a distortion joining normally parallel lines together to form a continuous helical ramp (with a pitch of one interplanar distance) winding about the dislocation.

screw press. A high-speed press in which the ram is activated by a large screw assembly powered by a drive mechanism.

scuffing. Localized damage caused by the occurrence of solid-phase welding between sliding surfaces. No local surface melting occurs.

seam. A surface defect appearing as thin lines in the rolling direction of sheet metals due to voids elongated during rolling.

seaming. The process of joining sheet metal parts by interlocking bends.

secant modulus. The slope of the secant drawn from the origin to any specified point on the stress-strain curve. See also *modulus of elasticity*.

secondary recrystallization. See *abnormal grain growth*.

secondary sheet. A shearing action that occurs between soft work metal and a cutting edge as a result of insufficient clearance.

secondary tensile stress. Tensile stress that develops during a bulk deformation process conducted under nominally compressive loading due to nonuniform metal flow resulting from geometry, friction, or die-chilling effects. Secondary tensile stresses are most prevalent in open-die forging operations.

segregation. A nonuniform distribution of alloying elements, impurities, or microphases.

seizure. The stopping of relative motion between two bodies as the result of severe interfacial friction. Seizure may be accompanied by gross surface welding.

semifinisher. An impression in a series of forging dies that only approximates the finish dimensions of the forging. Semifinishers are often used to extend die life or the finishing impression, to ensure proper control of grain flow during forging, and to assist in obtaining desired tolerances.

seminotching. A process similar to notching except that the cutting operation is a partial one only, permitting the cut shape to remain with the blank or part.

set. The shape remaining in a stamped or press-formed part after the punch force is removed. See also *permanent set*.

severe plastic deformation. Processes of plastic deformation with accumulated natural logarithmic strains more than 4 that are usually used to change material structure and properties.

shank. The portion of a die or tool by which it is held in position in a forging unit or press.

shape distortion. A dimensional change due to warping or bending resulting mainly from thermal treatment.

shape fixability. The ability of a material to retain the shape given to it by a forming operation.

shaving. Backflow of the clad or sleeve material during hydrostatic coextrusion.

shear. (1) A machine or tool for cutting metal and other material by the closing motion of two sharp, closely adjoining edges, for example, squaring shear and circular shear. (2) An inclination between two cutting edges, such as between two straight knife blades or between the punch cutting edge and the die cutting edge, so that a reduced area will be cut each time. This lessens the necessary force but increases the required length of the working stroke. This method is referred to as angular shear. (3) The act of cutting by shearing dies or blades, as in a squaring shear. (4) The type of force that causes or tends to cause two contiguous parts of the same body to slide relative to each other in a direction parallel to their plane of contact.

shear band. Region of highly localized shear deformation developed during bulk forming (and sometimes during sheet forming) as a result of material properties (such as a high flow-softening rate and low rate sensitivity of the flow stress), metal flow geometry, friction, chilling, and so on.

shear burr. A raised edge resulting from metal flow induced by blanking, cutting, or punching.

shearing. A cutting operation in which the work metal is placed between a stationary lower blade and movable upper blade and severed by bringing the blades together. Cutting occurs by a combination of metal shearing and actual fracture of the metal.

shear modulus (*G*). The ratio of shear stress to the corresponding shear strain for shear stresses below the proportional limit of the material. Values of shear modulus are usually determined by torsion testing. Also known as modulus of rigidity.

shear strength. The maximum shear stress a material can sustain. Shear strength is calculated from the maximum load during a shear or torsion test and is based on the original dimensions of the cross section of the specimen.

shear stress. (1) A stress that exists when parallel planes in metal crystals slide across each other. (2) The stress component tangential to the plane on which the forces act.

sheet. Any material or piece of uniform thickness and of considerable length and width as compared to its thickness. With regard to metal, such pieces under 6.5 mm (¼ in.) thick are called sheets, and those 6.5 mm (¼ in.) thick and over are called plates. Occasionally, the limiting thickness for steel to be designated as sheet steel is No. 10 Manufacturer's Standard Gage for sheet steel, which is 3.42 mm (0.1345 in.) thick.

sheet forming. The plastic deformation of a piece of sheet metal by tensile loads into a three-dimensional shape, often without significant changes in sheet thickness or surface characteristics. Compare with *bulk forming*.

shell four-ball test. A lubricant test in which three balls are clamped in contact, as in an equilateral triangle. The fourth ball is held in a rotating chuck and touches each of the stationary balls. Load is applied through a lever arm system that pushes the stationary balls upward against the rotating ball.

shim. A thin piece of material used between two surfaces to obtain a proper fit, adjustment, or alignment.

shot peening. A method of cold working metals in which compressive stresses are induced in the exposed surface layers of parts by impingement of a stream of shot (small spherical particles), directed at the metal surface at high velocity under controlled conditions. It differs from blast cleaning in primary purpose and in the extent to which it is controlled to yield accurate and reproducible results. Although shot peening cleans the surface being peened, this function is incidental. The major purpose of shot peening is to increase fatigue strength. Shot for peening is made of iron, steel, or glass.

shrinkage. The contraction of metal during cooling after hot forging. Die impressions are made oversized according to precise shrinkage scales to allow the forgings to shrink to design dimensions and tolerances.

shrink flanging. The reduction of the length of the free edge after the flanging process.

shuttle die. A multiple-station die in which the separated workpieces are fed from station to station by bars that are positioned in the die proper. Also known as a transfer die.

sidepressing. A deformation process in which a cylinder is laid on its side and deformed in compression. It is a good test to evaluate the tendency for fracture at the center of a billet, or for evaluating the tendency to form shear bands.

side thrust. The lateral force exerted between the dies by reaction of a forged piece on the die impressions.

single-stand mill. A rolling mill designed such that the product contacts only two rolls at a given moment.

sinking. The operation of machining the impression of a desired forging into die blocks.

sintering. The densification and bonding of adjacent particles in a powder mass or compact by heating to a temperature below the melting point of the main constituent.

size effect. The behavior in which the dimensions of the test specimen affect the value of the mechanical property measured. Most prominent for fatigue properties and strength of brittle materials, where strength is lower for large section size.

sizing. (1) Secondary forming or squeezing operations needed to square up, set down, flatten, or otherwise correct surfaces to produce specified dimensions and tolerances. (2) Some burnishing, broaching, drawing, and shaving operations are also called sizing. (3) A finishing operation for correcting ovality in tubing. (4) Final pressing of a sintered powder metallurgy part.

skin lamination. A subsurface separation that can result in surface rupture during forming.

slab. A flat-shaped semifinished rolled metal ingot with a width not less than 250 mm (10 in.) and a cross-sectional area not less than 105 cm² (16 in.²).

slabbing. The hot working of an ingot to a flat rectangular shape.

sleeve. Outer layer of a coextruded or codrawn product. See also *clad*.

slide. The main reciprocating member of a press, guided in the press frame, to which the punch or upper die is fastened; sometimes called the *ram*. The inner slide of a double-action press is called the plunger or punch-holder slide; the outer slide is called the blankholder slide. The third slide of a triple-action press is called the lower slide, and the slide of a hydraulic press is often called the platen.

slide adjustment. The distance that a press slide position can be altered to change the shut height of the die space. The adjustment can be made by hand or by power mechanism.

sliding friction test—flat dies. A test in which a sheet steel sample is placed between two flat, hardened die faces and pulled through the dies under conditions that permit recording of the load applied to the dies and the force required to pull the strip.

sliding friction test—wedge dies. Similar to the flat die test assembly except that one or both of the flat dies has a wedge configuration to confine the edges of the specimen. This permits development of unit loadings in excess of compressive strength of the specimen, and cold reduction of drawn strip is readily accomplished.

slip. Crystallographic shear process associated with dislocation glide that underlies the large plastic deformation of crystalline metals and alloys. Slip is usually observed on close-packed planes along close-packed directions, in which case it is referred to as restricted slip. In body-centered cubic materials, such as alpha iron, slip occurs along any plane containing a close-packed direction and is referred to as pencil glide.

slip-line field. Graphical technique used to estimate the deformation and stresses involved in plane-strain metalforming processes.

slitting. Cutting or shearing along single lines to cut strips from a sheet or to cut along lines of a given length or contour in a sheet or workpiece.

sliver. A surface defect consisting of an elongated thin layer of partially attached metal.

slotting. A stamping operation in which elongated or rectangular holes are cut in a blank or part.

slug. (1) The metal removed when punching a hole in a forging; also termed punchout. (2) The forging stock for one workpiece cut to length.

smith forging. See *hand forge (smith forge)*.

smudge. A dark-appearing surface contamination on annealed sheet generally resulting from cold-reduction oil residues or carbon deposited from annealing gas with an unfavorable CO/CO_2 ratio. Smudge may adversely affect painting or plating but may be beneficial in the prevention of galling.

smut. A contaminant consisting of fine, dark-colored particles on the surface of pickled sheet products. This usually results from heavy oxidation of the steel surface during hot rolling.

snap through. Shock in a die due to the sudden beginning and completion of fractures in cutting dies, causing the compressed punch to elastically snap into tension.

solid modeling. A form of computer modeling in which the three-dimensional features of the part or object are represented. With solid modeling, a cut through the model reveals interior details. The method also permits accurate calculation of mass properties (e.g., mass and moment of inertia), and, with full color, shading, and shadowing, it creates realistic displays. Solid models may be integrated with motion analysis software to create realistic simulations. Solid models may also be linked with finite-element models.

sow block. A block of heat treated steel placed between the anvil of the hammer and the forging die to prevent undue wear to the anvil. Sow blocks are occasionally used to hold insert dies. Also called anvil cap.

space lattice. A set of equal and adjoining parallelepipeds formed by dividing space by three sets of parallel planes, the planes in any one set being equally spaced. There are seven ways of so dividing space, corresponding to the seven *crystal system* structures. The unit parallelepiped is usually chosen as the unit cell of the system. Due to geometrical considerations, atoms can only have one of 14 possible arrangements, known as *Bravais lattices*.

spalling. The removal of small pieces of metal from the working face of the die, usually as a result of severe heat checking. Spalling is most likely in hard materials with low ductility.

spank. A press operation used to reform parts that have already had their major contour formed or drawn in the conventional manner. The spank operation is often used where it is not possible to produce the final contour, such as sharp creases or corners, in a single forming operation. It is also used at the end of a production line where large sheet metal parts have become distorted due to previous operations, such as trimming, punching, forming, and flanging. Spanking is used to bring the panels back to the desired contour. See also *restriking*.

special boundary. A grain boundary between two grains whose crystallographic lattices have a certain fraction ($1/N$, in which N is an integer) of coincident lattice points. Such boundaries, denoted using the notation ΣN, may have low mobility and surface energy.

specific heat. Amount of heat required to change the temperature of one unit weight of a material by one degree.

specific modulus. The material elastic modulus divided by the material density.

specific strength. The material strength divided by the material density.

spheroidization. Process of converting a lamellar, basketweave, or acicular second phase into an equiaxed morphology via deformation, annealing, or a combination of deformation followed by annealing.

spinning. The forming of a seamless hollow metal part by forcing a rotating blank to conform to a shaped mandrel that rotates concentrically with the blank. In the typical application, a flat-rolled metal blank is forced against the mandrel by a blunt, rounded tool; however, other stock (notably, welded or seamless tubing) can be formed. A roller is sometimes used as the working end of the tool.

spinoidal hardening. Strengthening caused by the formation of a periodic array of coherent face-centered cubic solid-solution phases on a submicrostructural size level.

springback. The elastic recovery of metal after stressing. The extent to which metal tends to

return to its original shape or contour after undergoing a forming operation. This is compensated for by overbending or by a secondary operation of restriking.

stacking-fault energy (SFE). The energy associated with the planar fault formed by dissociated dislocations in crystalline materials. Low-SFE materials typically have wide stacking faults, and high-SFE materials very narrow or no stacking faults. The SFE affects a number of material properties, such as work-hardening rate and recrystallization. Materials with low SFE undergo rapid dislocation multiplication and hence show high work-hardening rates and relative ease of dynamic recrystallization because of the difficulty of dynamic recovery. Materials with high SFE energies usually exhibit low work-hardening rates because of the ease of dynamic recovery and are difficult to recrystallize.

stamping. A general term to denote all press-working. In a more specific sense, stamping is used to imprint letters, numerals, and trademarks in sheet metal, machined parts, forgings, and castings. A tool called a stamp, with the letter or number raised on its surface, is hammered or forced into the metal, leaving a depression on the surface in the form of the letter or number.

standard deviation. A measure of the dispersion of observed values or results from the average, expressed as the positive square root of the variance.

static friction. The force tangential to the interface that is just sufficient to initiate relative motion between two bodies under load.

sticker breaks. Repetitive transverse lines, often curved, caused by localized welding of coil wraps during annealing and subsequent separation of these welded areas during an uncoiling operation. Also referred to as sticker marks.

stiffness. Resistance to elastic deformation.

stochastic search methods. A large group of optimization techniques that uses probabilistic methods. Two common methods are genetic algorithms and simulated annealing.

stock. A general term used to refer to a supply of metal in any form or shape and also to an individual piece of metal that is formed, forged, or machined to make parts.

stoichiometric. Having the precise weight relation of the elements in a chemical compound, or quantities of reacting elements or compounds being in the same weight relation as the theoretical combining weight of the elements involved.

straightening. A finishing operation for correcting misalignment in a forging or between various sections of a forging.

strain. The unit of change in the size or shape of a body due to force, in reference to its original size or shape.

strain aging. The changes in ductility, hardness, yield point, and tensile strength that

occur when a metal or alloy that has been cold worked is stored for some time. In steel, strain aging is characterized by a loss of ductility and a corresponding increase in hardness, yield point, and tensile strength.

strain hardening. An increase in hardness and strength caused by plastic deformation at temperatures below the recrystallization range. Also known as work hardening.

strain-hardening coefficient or exponent. The value n in the relationship $\sigma = K\varepsilon^n$, where σ is the true stress; ε is the true strain; and K, which is called the strength coefficient, is equal to the true stress at a true strain of 1.0. The strain-hardening exponent, also called n-value, is equal to the slope of the true stress/true strain curve up to maximum load, when plotted on log-log coordinates. The n-value relates to the ability of a sheet material to be stretched in metalworking operations. The higher the n-value, the better the formability (stretchability). Also called work-hardening exponent.

strain lines. Surface defects in the form of shallow line-type depressions appearing in sheet metals after stretching the surface a few percent of unit area or length. See also *Lüders lines*.

strain rate. The time rate of deformation (strain) during a metalforming process.

strain-rate sensitivity. The degree to which mechanical properties are affected by changes in deformation rate. Quantified by the slope of a log-log plot of flow stress (at fixed strain and temperature) versus strain rate. Also known as the *m*-value.

strength. The ability of a material to withstand an applied force.

strength coefficient (K). A constant related to the tensile strength used in the power-law equation $\sigma = K\varepsilon^n$. In mechanical engineering nomenclature, it is called so, and the power-law equation is given as $\sigma = \sigma_o\varepsilon^n$. See also *n*-value.

stress. The intensity of the internally distributed forces or components of forces that resist a change in the volume or shape of a material that is or has been subjected to external forces. Stress is expressed in force per unit area. Stress can be normal (tension or compression) or shear.

stress concentration. On a macromechanical level, the magnification of the level of an applied stress in the region of a notch, void, hole, or inclusion.

stress-concentration factor (K_t). A multiplying factor for applied stress that allows for the presence of a structural discontinuity such as a notch or hole; K_t equals the ratio of the greatest stress in the region of the discontinuity to the nominal stress for the entire section. Also called theoretical stress-concentration factor.

stress-intensity factor. A scaling factor, usually denoted by the symbol K, used in linear elastic fracture mechanics to describe the

intensification of applied stress at the tip of a crack of known size and shape. At the onset of rapid crack propagation in any structure containing a crack, the factor is called the critical stress-intensity factor, or the fracture toughness. Various subscripts are used to denote different loading conditions or fracture toughnesses.

stress raisers. Design features (such as sharp corners) or mechanical defects (such as notches) that act to intensify the stress at these locations.

stress relaxation. Drop in stress with time when material is maintained at a constant strain. The drop in stress is a result of plastic accommodation processes.

stress relief. The removal or reduction of residual stress by thermal treatment, mechanical treatment (shot peening, surface rolling, stretching, bending, and straightening), or vibratory stress relief.

stress-strain curve. A graph in which corresponding values of stress and strain from a tension, compression, or torsion test are plotted against each other. Values of stress are usually plotted vertically (ordinates or y-axis) and values of strain horizontally (abscissas or *x*-axis). Also known as stress-strain diagram.

stretchability. The ability of a material to undergo stretch-type deformation.

stretcher leveling. The leveling of a piece of sheet metal (that is, removing warp and distortion) by gripping it at both ends and subjecting it to a stress higher than its yield strength.

stretcher straightening. A process for straightening rod, tubing, and shapes by the application of tension at the ends of the stock. The products are elongated a definite amount to remove warpage.

stretcher strains. Elongated markings that appear on the surface of some sheet materials when deformed just past the yield point. These markings lie approximately parallel to the direction of maximum shear stress and are the result of localized yielding. See also *Lüders lines*.

stretch flanging. The stretching of the length of the free edge after the flanging process.

stretch forming. The shaping of a sheet or part, usually of uniform cross section, by first applying suitable tension or stretch and then wrapping it around a die of the desired shape.

stretching. The mode of deformation in which a positive strain is generated on the sheet surface by the application of a tensile stress. In stretching, the flange of the flat blank is securely clamped. Deformation is restricted to the area initially within the die. The stretching limit is the onset of metal failure.

striking surface. Those areas on the faces of a set of dies that are designed to meet when the upper die and lower die are brought together. The striking surface helps protect impres-

sions from impact shock and aids in maintaining longer die life.

strip. A flat-rolled metal product of some maximum thickness and width, arbitrarily dependent on the type of metal; narrower than *sheet*.

stripping. The removal of the metal strip from the punch after a cutting operation. Also a term referring to the removal of a part adhering to the punch on the upstroke after forming.

subcritical crack growth (SCG). A failure process in which a crack initiates at a preexisting flaw and grows until it attains a critical length. At that point, the crack grows in an unstable fashion, leading to catastrophic failure. Typical examples of SCG processes are fatigue failure and stress corrosion.

subgrain. Micron-sized volume bounded by well-defined dislocation walls. The misorientations across the walls are low angle in nature, that is, $<15°$.

subpress die. A die that is closed by the press ram but opened by springs or other means because the upper shoe is not attached to the ram.

subsow block (die holder). A block used as an adapter in order to permit the use of forging dies that otherwise would not have sufficient height to be used in the particular unit or to permit the use of dies in a unit with different *shank* sizes.

superplastic forming. Forming using the superplasticity properties of material at elevated temperatures.

superplastic forming and diffusion bonding. The process of combining the diffusion bonding cycle into the superplastic forming.

superplasticity. The ability of certain metals to develop extremely high tensile elongations at elevated temperatures and under controlled rates of deformation. Materials that show high strain-rate sensitivity (≥ 0.5) at deformation temperatures often exhibit superplasticity. The phenomenon is often developed through a mechanism of grain-boundary sliding in very fine-grained, two-phase alloys.

support plate. A plate that supports a draw ring or draw plate. It also serves as a spacer.

surface finish. The classification of a surface in terms of roughness and peak density.

surface hardening. A localized heat treating process that produces a hard-quenched surface in steel without introducing additional alloying elements. Surface hardening can be produced by flame, induction, or laser or electron beam thermal treatments.

surface hardness. The hardness of that portion of the material very near the surface, as measured by microhardness or superficial hardness testers.

surface oxidation. Development of an oxide film or layer on the surface of metals in oxidizing environments. Oxidation at high temperatures is occasionally referred to as sealing.

surface roughness. The fine irregularities in the surface texture that result from the production process. Considered as vertical deviations from the nominal or average plane of the surface.

surface texture. Repetitive or random deviations from the nominal surface that form the pattern of the surface. Includes roughness, waviness, and flaws.

surface topography. The fine-scale features of a surface as defined by the size and distribution of asperities. Surface topography is measured by surface roughness and the direction of surface features (lay).

swage. (1) The operation of reducing or changing the cross-sectional area of stock by the fast impact of revolving dies. (2) The tapering of bar, rod, wire, or tubing by forging, hammering, or squeezing; reducing a section by progressively tapering lengthwise until the entire section attains the smaller dimension of the taper.

sweep device. A single or double arm (rod) attached to the upper die or slide of a press and intended to move the operator's hands to a safe position as the dies close, if the operator's hands are inadvertently within the point of operation.

Swift cup test. A simulative test in which circular blanks of various diameter are clamped in a die ring and deep drawn into a cup by a flat-bottomed cylindrical punch. The ratio of the largest blank diameter that can be drawn successfully to the cup diameter is known as the *limiting drawing ratio or deformation limit*.

T

Taguchi method. A technique for designing and performing experiments to investigate processes in which the output depends on many factors (e.g., material properties, process parameters) without having to tediously and uneconomically run the process using all possible combinations of values of those variables. By systematically choosing certain combinations of variables, it is possible to separate their individual effects.

tailor-welded blank. Blank for sheet forming typically consisting of steels of different thickness, grades/strengths, and sometimes coatings that are welded together prior to forming. Tailor-welded blanks are used to make finished parts with a desirable variation in properties such as strength, corrosion resistance, and so on.

tangent bending. The forming of one or more identical bends having parallel axes by wiping sheet metal around one or more radius dies in a single operation. The sheet, which may have side flanges, is clamped against the radius die and then made to conform to the radius die by pressure from a rocker-plate die that moves along the periphery of the radius die. See also *wiper forming (wiping)*.

tangent modulus. The slope of the stress-strain curve at any specified stress or strain. See also *modulus of elasticity*.

Taylor factor. The ratio of the required stress for deformation under a specified strain state to the critical resolved shear stress for slip (or twinning) within the crystals comprising a *polycrystalline aggregate*. The determination of the Taylor factor assumes uniform and identical strain within each crystal in the aggregate and provides an upper bound on the required stresses. The Taylor factor averaged over all crystals in a polycrystalline aggregate ($= \overline{M}$) provides an estimate of the effect of texture on strength.

tearing. Failure and localized separation of a sheet metal.

temper. In nonferrous alloys and in some ferrous alloys (steels that cannot be hardened by heat treatment), the hardness and strength produced by mechanical or thermal treatment, or both, and characterized by a certain structure, mechanical properties, or reduction in area during cold working.

temperature-compensated strain rate. Parameter used to describe the interdependence of temperature and strain rate in the description of thermally activated (diffusion-like) deformation processes. It is defined as $\dot{\varepsilon}$ exp (Q/RT), in which $\dot{\varepsilon}$ denotes the strain rate, Q is an apparent activation energy characterizing the micromechanism of deformation, R is the gas constant, and T is the absolute temperature. Flow stress, dynamic recrystallization, and so on at various strain rates and temperatures are frequently interpreted in terms of the temperature-compensated strain rate. Also known as the *Zener-Hollomon parameter (Z)*.

tempering. (1) In heat treatment, reheating hardened steel to some temperature below the eutectoid temperature to decrease hardness and/or increase toughness. (2) The process of rapidly cooling glass from near its softening point to induce compressive stresses on the surface balanced by interior tension, thereby imparting increased strength.

template (templet). A gage or pattern made in a die department, usually from sheet steel; used to check dimensions on forgings and as an aid in sinking die impressions in order to correct dimensions.

tensile ratio. The ratio of the tensile strength to yield strength. It is the inverse of the yield ratio.

tensile strength. In tensile testing, the ratio of maximum load to original cross-sectional area. Also known as ultimate strength. Compare with *yield strength*.

tensile stress. A stress that causes two parts of an elastic body, on either side of a typical stress plane, to pull apart. Contrast with *compressive stress*.

tension. The force or load that produces elongation.

tensor order. A measure of the number of directional dimensions associated with a

quantity. A scalar, for example, has a tensor order of 0, indicating that it has no directionality associated with it. A vector, having a single direction, is a quantity with a tensor order of 1. The Reynolds stress tensor, the product of two vectors, has a tesnor order of 2. Higher-order tensors exist with a number of directions equal to their tensor order.

tetrahedral element. The element for three-dimensional finite-element modeling that is tetrahedron shaped (four faces) and has four nodes; also called the tet element.

texture. The description of the relative probability of finding the crystals comprising a polycrystalline aggregate in various orientations.

thermal conductivity. Ability of a material to conduct heat. The rate of heat flow, under steady conditions, through unit area, per unit temperature gradient in the direction perpendicular to the area. It is given in SI units as watts per meter Kelvin (W/m K); in customary units as (Btu/ft^2 °F).

thermal fatigue. Fracture resulting from the presence of temperature gradients that vary with time in such a manner as to produce cyclic stresses in a structure.

thermocouple. A device for measuring temperature, consisting of two dissimilar metals that produce an electromotive force roughly proportional to the temperature difference between their hot and cold junction ends.

thermomechanical processing (TMP). A general term covering a variety of processes combining controlled thermal and deformation treatments to obtain synergistic effects, such as improvement in strength without loss of toughness. Same as thermal-mechanical treatment.

thick-film lubrication. A condition of lubrication in which the film thickness of the lubricant is appreciably greater than that required to cover the surface asperities when subjected to the operating load. See also *thin-film lubrication*.

thin-film lubrication. A condition of lubrication in which the film thickness of the lubricant is such that the friction and wear between the surfaces is determined by the properties of the surfaces as well as the characteristics of the lubricant.

three-point bending. The bending of a piece of metal or a structural member in which the object is placed across two supports and force is applied between and in opposition to them.

tilt boundary. Grain boundary for which the crystal lattices of the grains on either side of the boundary are related by a rotation about an axis that lies in the plane of the boundary.

torsion. A twisting deformation of a solid or tubular body about an axis in which lines that were initially parallel to the axis become helices.

torsional stress. *The shear stress* on a transverse cross section resulting from a twisting action.

total elongation. The total amount of permanent extension of a testpiece broken in a tensile test; usually expressed as a percentage over a fixed gage length. See also *elongation, percent*.

toughness. The ability of a material to resist an impact load (high strain rate) or to deform under such a load in a ductile manner, absorbing a large amount of the impact energy and deforming plastically before fracturing. Such impact toughness is frequently evaluated with Charpy or Izod notched impact specimens. Impact toughness is measured in terms of the energy absorbed during fracture. Fracture toughness is a measure of the ability of a material to withstand fracture in the presence of flaws under static or dynamic loading of various types (tensile, shear, etc.). An indicator of damage tolerance, fracture toughness is measured in terms of $Mpa\sqrt{m}$ or $ksi\sqrt{in}$.

transformation-induced plasticity (TRIP). A phenomenon occurring chiefly in certain highly alloyed steels that have been heat treated to produce metastable austenite or metastable austenite plus martensite, whereby, on subsequent deformation, part of the austenite undergoes strain-induced transformation to martensite. Steels capable of transforming in this manner, commonly referred to as TRIP steels, are highly plastic after heat treatment but exhibit a very high rate of strain hardening and thus have high tensile and yield strengths after plastic deformation at temperatures between 20 and 500 °C (70 and 930 °F). Cooling to –195 °C (–320 °F) may or may not be required to complete the transformation to martensite. Tempering usually is done following transformation.

transformation temperature. The temperature at which a change in phase occurs. This term is sometimes used to denote the limiting temperature of a transformation range.

transition temperature. (1) An arbitrarily defined temperature that lies within the temperature range in which metal fracture characteristics (as usually determined by tests of notched specimens) change rapidly, such as the ductile-to-brittle transition temperature (DBTT). The DBTT can be assessed in several ways, the most common being the temperature for 50% lowest temperature at which the fracture is 100% ductile (100% fibrous criterion). The DBTT is commonly associated with temper embrittlement and radiation damage (neutron irradiation) of low-alloy steels. (2) Sometimes used to denote an arbitrarily defined temperature within a range in which the ductility changes rapidly with temperature.

Tresca yield criterion. Prediction of yielding in ductile materials when the maximum shear stress on any plane reaches a critical value, $\tau = \tau_c$.

triaxiality. The ratio of the hydrostatic (mean) stress to the flow (effective) stress. Triaxiality provides a measure of the tendency for cavities to grow during deformation processing.

trimming. The mechanical shearing of flash or excess material from a forging with a trimmer in a trim press; can be done hot or cold.

triple junction/triple point. Point at which three grains meet in a polycrystalline aggregate. Also, region in which high stress concentrations may develop during hot working or elevated-temperature service, thus nucleating wedge cracking.

true strain. (1) The ratio of the change in dimension, resulting from a given load increment, to the magnitude of the dimension immediately prior to applying the load increment. (2) In a body subjected to axial force, the natural logarithm of the ratio of the gage length at the moment of observation to the original gage length. Also known as natural strain.

true stress. The value obtained by dividing the load applied to a member at a given instant by the cross-sectional area over which it acts.

tube stock. A semifinished tube suitable for subsequent reduction and finishing.

twinning. Also called deformation or mechanical twinning, it is a deformation mechanism, similar to dislocation slip, in which small (often plate- or lens-shaped) regions of a crystal or grain reorient crystallographically to adopt a twin relationship to the parent crystal. It is particularly common in noncubic metals (e.g., alpha-titanium and tetragonal tin) and in many body-centered cubic metals deformed at high rates and/or low temperatures. Twinning is often accompanied by an audible crackling sound, from which "crying tin" gets its name.

twist boundary. Grain boundary for which the crystal lattices of the grains on either side of the boundary are related by a rotation about an axis that lies perpendicular to the plane of the boundary.

TZM. A high-creep-strength titanium, zirconium, and molybdenum alloy used to make dies for the isothermal forging process.

U

ultimate strength. The maximum stress (tensile, compressive, or shear) a material can sustain without fracture; determined by dividing maximum load by the original cross-sectional area of the specimen. Also known as nominal strength or maximum strength.

ultrasonic inspection. The use of high-frequency acoustical signals for the purpose of nondestructively locating flaws within raw material or finished parts.

underfill. A portion of a forging that has insufficient metal to give it the true shape of the impression.

uniform elongation (e_u). The elongation that occurs at maximum load and immediately preceding the onset of necking in a tensile test.

upset. The localized increase in cross-sectional area of a workpiece or weldment resulting from the application of pressure during mechanical fabrication or welding.

upset forging. A forging obtained by *upset* of a suitable length of bar, billet, or bloom.

upsetter. A horizontal mechanical press used to make parts from bar stock or tubing by *upset forging*, piercing, bending, or otherwise forming in dies. Also known as a header.

upsetting. The working of metal so that the cross-sectional area of a portion or all of the stock is increased. See also *heading*.

V

vacuum forming. Sheetforming process most commonly used for titanium in which a blank is placed into a chamber that has a heated die, and a vacuum is applied to creep form the part onto the die. The part is usually covered with an insulating material, and the bag is outside this material.

validation. The process of substantiating that material property test data have been generated according to standard methods and practices, or other indices of quality, reliability, and precision. The validation process is the first step toward ratification or confirmation of the data, making them legally effective and binding in some specified application.

variability of data. The degree to which random variables deviate from a central value or mean. In statistical terms, this is measured by the sample standard deviation or sample variance.

vent. A small hole in a punch or die for admitting air to avoid suction holding or for relieving pockets of trapped air that would prevent die closure or action.

vent mark. A small protrusion resulting from the entrance of metal into die vent holes.

viscoelasticity. A property involving a combination of elastic and viscous behavior that makes deformation dependent on both temperature and strain rate. A material having this property is considered to combine the features of a perfectly elastic solid and a perfect fluid.

viscosity. Bulk property of a fluid or semifluid that causes it to resist flow.

visioplasticity. A physical-modeling technique in which an inexpensive, easy-to-deform material (e.g., clay, wax, lead) is gridded and deformed in subscale tooling to establish the effects of die design, lubrication, and so forth on metal flow and defect formation by way of postdeformation examination of grid distortions.

W

warm working. Deformation at elevated temperatures below the recrystallization temperature. The flow stress and rate of strain hardening are reduced with increasing temperature; therefore, lower forces are required than in cold working. See also *cold working and hot working*.

wear plates. Replaceable elements used to face wearing surfaces on a hydraulic press.

wear resistance. Resistance of a sheet metal to surface abrasion. See also *erosion resistance*.

web. A relatively flat, thin portion of a forging that effects an interconnection between ribs and bosses; a panel or wall that is generally parallel to the forging plane.

wedge compression test. A simple workability test in which a wedge-shaped specimen is compressed to a certain thickness. This gives a gradient specimen in which material has been subjected to a range of plastic strains.

Widmanstätten structure. Characteristic structure produced when preferred planes and directions in the parent phase are favored for growth of a second phase, resulting in the precipitated second phase appearing as plates, needles, or rods within a matrix.

wiper forming (wiping). Method of curving sheet metal sections or tubing over a form block or die in which this form block is rotated relative to a wiper block or slide block.

wire. A thin, flexible, continuous length of metal, usually of circular cross section and usually produced by drawing through a die.

wire drawing. Reducing the cross section of wire by pulling it through a die.

wire drawing test. A test in which a cylindrical draw die is used to reduce the diameter of wire. The drawing force is measured and reflects lubricant effectiveness.

wire rod. Hot rolled coiled stock that is to be cold drawn into wire.

workability. See also *formability*, which is a term more often applied to sheet materials. The ease with which a material can be shaped through plastic deformation in bulk forming processes. It involves both the measurement of the resistance to deformation (the flow properties) and the extent of possible plastic deformation before fracture occurs (ductility).

work hardening. See strain hardening.

work-hardening exponent. See strain-hardening exponent.

workpiece. General term for the work material in a metalforming operation.

wrap forming. See *stretch forming*.

wrinkling. A wavy condition obtained in deep drawing of sheet metal in the area of the metal between the edge of the flange and the draw radius. Wrinkling may also occur in other forming operations when unbalanced compressive forces are set up.

wrought material. Material that is processed by plastic deformation, typically to produce a recrystallized microstructure. Cast and wrought materials are produced by ingot casting and deformation processes to produce final mill products.

Y

yield. Evidence of plastic deformation in structural materials. Also known as plastic flow or creep.

yield point. The first stress in a material, usually less than the maximum attainable stress, at which an increase in strain occurs without an increase in stress. Only certain metals—those that exhibit a localized, heterogeneous type of transition from elastic to plastic deformation—produce a yield point. If there is a decrease in stress after yielding, a distinction can be made between upper and lower yield points. The load at which a sudden drop in the flow curve occurs is called the upper yield point. The constant load shown on the flow curve is the lower yield point.

yield point elongation. The extension associated with discontinuous yielding that occurs at approximately constant load following the onset of plastic flow. It is associated with the propagation of Lüder's lines or bands.

yield ratio. The ratio of the yield strength to the tensile strength. It is the inverse of the tensile ratio.

yield strength. The stress at which a material exhibits a specified deviation from proportionality of stress and strain. An offset of 0.2% is used for many metals. Compare with *tensile strength*.

yield stress. A stress at which a steel exhibits the first measurable permanent plastic deformation.

Young's modulus. A measure of the rigidity of a metal. It is the ratio of stress, within the proportional limit, to corresponding strain. Young's modulus specifically is the modulus obtained in tension or compression.

Z

Zener-Hollomon parameter (Z). See *temperature-compensated strain rate*.

SELECTED REFERENCES

- J.R. Davis, Ed., A*SM Materials Engineering Dictionary*, ASM International, 1992
- *DoD Modeling and Simulation (M&S) Glossary*, U.S. Department of Defense, (DoD 5000.59-M), Jan 1998
- *Introduction to Finite-Difference Methods for Numerical Fluid Dynamics,* Los Alamos

National Laboratories, released to OSTI, 1996

- *Dictionary of Scientific and Technical Terms*, 5th ed., McGraw-Hill

- *Metallogragphy and Microstructures,* Vol 9, *ASM Handbook,* ASM International, 2004
- *Metalworking: Sheet Forming*, Vol 14B, ASM Handbook, ASM International, 2006

- *Materials Selection and Design*, Vol 20, *ASM Handbook,* ASM International, 1997

INDEX

This page has been reformatted by Knovel to provide easier navigation.

www.iran-mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

www.iran-mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

www.iran-mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

www.iran-mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

www.iran−mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

www.iran-mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

Index Terms	Links

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

Index Terms	Links	

This page has been reformatted by Knovel to provide easier navigation.

www.iran-mavad.com

مرجع مهندسی مواد و متالورژی

Index Terms	Links

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

www.iran‑mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

www.iran‑mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

www.iran—mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

www.iran-mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

www.iran–mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

www.iran‑mavad.com

مرجع مهندسی مواد و متالورژی

Index Terms	Links		

This page has been reformatted by Knovel to provide easier navigation.

www.iran-mavad.com

مرجع مهندسی مواد و متالورژی

Index Terms	Links		
Lubrication			
emulsions	135		
overview	135		
regimes	136		
rolling aluminum with	135		
Lüders lines	709		
Lumped-parameter methods	34		
Lumped-parameter model	34	36	709

M

This page has been reformatted by Knovel to provide easier navigation.

www.iran–mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

www.iran–mavad.com

مرجع مهندسی مواد و متالورژی

Index Terms	Links

This page has been reformatted by Knovel to provide easier navigation.

www.iran-mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

www.iran–mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

www.iran-mavad.com

مرجع مهندسی مواد و متالورژی

Index Terms	Links			

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

www.iran‒mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

Index Terms	Links			
Stochastic search methods	717			
Stock	130	717		
Stoichiometric	55	717		
Stoichiometric phases	176			
Stowell Model	328			
Straightening	108	717		
Strain	23	717		
Strain aging	717			
Strain hardening	4	717		
Strain lines	355	717		

Strain rate	11	717		
Strain-hardening coefficient	388	462	513	717
Strain-hardening exponent	35	523		
Strain-rate sensitivity	35	717		
Strength	10	717		
Strength coefficient (K)	717			
Strengthening mechanisms				
age/precipitation hardening				
legacy understanding	380			
overview	379			
precipitation strengthening,				
modeling	381			
cold work				
legacy understanding	386			
modeling strengthening by	387			
overview	385			
dispersion strengthening				
modeling	383			
overview	383			
grain-size strengthening/hardening				
legacy understanding	382			
modeling	384			
overview	383			
slip transmission, strengthening from				
interfacial barriers to	388			

This page has been reformatted by Knovel to provide easier navigation.

www.iran‒mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

This page has been reformatted by Knovel to provide easier navigation.

www.iran-mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

www.iran-mavad.com

مرجع مهندسی مواد و متالورژی

This page has been reformatted by Knovel to provide easier navigation.

Index Terms	Links				
Viscoelasticity	720				
Viscoplastic self-consistent (VPSC)					
model	240	243	339	517	545
	546				
Viscosity	57	720			
Visioplasticity	347	355	356	720	
Voronoi tessellation	283				

W

Warm working	542	720			
Wear plates	720				
Wear resistance	155	720			
Web	33	720			
Wedge compression test	720				
Widmanstätten structure	392	394	528	531	532
	720				
WIMV approach	240				
Wiper forming (wiping)	720				
Wire	25	720			
Wire drawing	127	348	349	720	
Wire drawing test	720				
Wire rod	720				
Work hardening	23				
See also Strain hardening					
Workability	35	720			
See also Formability					
Work-hardening exponent	720				
Workpiece	23	720			
Wrap forming. *See* **Stretch forming**..					
Wrinkling	720				
Wrought material	543	720			

Y

Yield	35	720			
Yield point	465	720			
Yield point elongation	720				
Yield ratio	720				
Yield strength	5	720			

This page has been reformatted by Knovel to provide easier navigation.

www.iran−mavad.com

مرجع مهندسی مواد و متالورژی