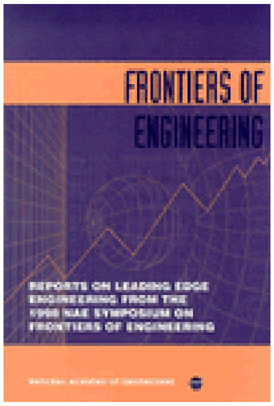


# Frontiers of Engineering: Reports on Leading Edge Engineering From the 1998 NAE Symposium on Frontiers of Engineering

National Academy of Engineering

ISBN: 0-309-52480-6, 160 pages, 6 x 9, (1999)

**This free PDF was downloaded from:**
**http://www.nap.edu/catalog/6411.html**

Visit the National Academies Press online, the authoritative source for all books from the National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine, and the National Research Council:

- Download hundreds of free books in PDF
- Read thousands of books online for free
- Purchase printed books and PDF files
- Explore our innovative research tools – try the Research Dashboard now
- Sign up to be notified when new books are published

Thank you for downloading this free PDF. If you have comments, questions or want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, visit us online, or send an email to comments@nap.edu.

This book plus thousands more are available at www.nap.edu.

Copyright © National Academy of Sciences. All rights reserved.
Unless otherwise indicated, all materials in this PDF file are copyrighted by the National Academy of Sciences. Distribution or copying is strictly prohibited without permission of the National Academies Press <http://www.nap.edu/permissions/>. Permission is granted for this material to be posted on a secure password-protected Web site. The content may not be posted on a public Web site.

THE NATIONAL ACADEMIES
*Advisers to the Nation on Science, Engineering, and Medicine*

# FOURTH ANNUAL SYMPOSIUM ON FRONTIERS OF ENGINEERING

NATIONAL ACADEMY OF ENGINEERING

NATIONAL ACADEMY PRESS
Washington, D.C.  1999

**NATIONAL ACADEMY PRESS - 2101 Constitution Ave., N.W. - Washington, D.C. 20418**

NOTICE: The National Academy of Engineering (NAE), a private nonprofit institution, was established in 1964 under the charter of the National Academy of Sciences (NAS) as a parallel organization of distinguished engineers. It is autonomous in its administration and in its selection of members, sharing with the NAS the responsibility of advising the federal government on scientific and technical matters; this mandate is carried out through joint supervision of the National Research Council. The NAE also sponsors an independent study program aimed at meeting national needs, encourages engineering education and research, and recognizes the superior achievement of engineers. Wm. A. Wulf is president of the NAE.

This publication has been reviewed according to procedures approved by a National Academy of Engineering report review process. Publication of signed work signifies that it is judged a competent and useful contribution worthy of public consideration, but it does not imply endorsement of conclusions or recommendations by the NAE. The interpretations and conclusions in such publications are those of the authors and do not purport to represent the views of the council, officers, or staff of the National Academy of Engineering.

Funding for the activity that led to this publication was provided by the National Science Foundation, U.S. Department of Defense, and National Academy of Engineering Fund.

International Standard Book Number 0-309-06287-X

Library of Congress Catalog Card Number 99-60544

Printed in the United States of America

# ORGANIZING COMMITTEE

ROBERT H. WAGONER (Chair), Professor, Department of Materials Science and Engineering, Ohio State University
FRANCES H. ARNOLD, Professor, Division of Chemistry and Chemical Engineering, California Institute of Technology
SUSAN CORWIN, Director, Computing Enhancement Architecture Lab, Intel Corporation
KENNETH Y. GOLDBERG, Associate Professor, Department of Industrial Engineering and Operations Research, University of California, Berkeley
CONNIE L. GUTOWSKI, Windstar Chassis Manager, Ford Motor Company
ROBERT D. HOWE, Associate Professor, Division of Engineering and Applied Sciences, Harvard University
PARVIZ MOIN, Professor, Department of Mechanical Engineering, Stanford University
JUN NI, Professor, Department of Mechanical Engineering and Applied Mechanics, Director, Wu Manufacturing Research Center, and Director, NSF-I/UCR Center
RICHARD S. PARNAS, Group Leader, Polymer Composites, National Institute of Standards and Technology
ELSA REICHMANIS, Head, Polymer and Organic Materials Research Department, Bell Laboratories, Lucent Technologies
JAMES G. SCHROTH, Staff Research Engineer, General Motors R&D Center
EVA M. SEVICK-MURACA, Associate Professor, School of Chemical Engineering, Purdue University

## Staff

JANET R. HUNZIKER, Program Officer
MARY W. L. KUTRUFF, Administrative Assistant

# Preface

In 1995 the National Academy of Engineering (NAE) initiated the Frontiers of Engineering Symposium program, which brings together 100 of the nation's future engineering leaders to learn about cutting-edge research and technical work in different engineering fields. On September 17-19, 1998, the NAE hosted its fourth Frontiers of Engineering Symposium. This book contains summary papers of the presentations given at that meeting. The intent of this book, and of the three that precede it in the series, is to describe the content and underpinning philosophy of this unique meeting and to highlight the kinds of pioneering research and technical work being done today.

## GOALS OF FRONTIERS OF ENGINEERING

The practice of engineering is changing. Not only must engineers be able to thrive in an environment of rapid technological change and globalization, but engineering is becoming more interdisciplinary. The frontiers of engineering are frequently occurring at the intersections of engineering disciplines, which compels researchers and practitioners alike to be aware of developments and challenges in areas other than their own.

At the three-day Frontiers of Engineering symposium, 100 of this country's best and brightest engineers, ages 30 to 45, learn from their peers about what is happening at the leading edge of engineering. This has great value for the participants in a couple of ways. First, it broadens their knowledge of current developments in other fields of engineering, leading to insights that may be applicable to the furthering of their own work. Second, because the engineers come from a variety of institutions in academia, industry, and government and from many different engineering disciplines, it allows them to make contacts with and learn from individuals whom they would not ordinarily meet in their usual round of professional meetings. This networking, it is hoped, will lead to collaborative work, facilitating the transfer of new techniques and approaches across fields.

The number of participants at each meeting is kept at 100 to maximize the opportunities for interaction and exchange among the attendees, who are invited to attend after a competitive nomination and selection process. The choice of topics and speakers for each year's meeting is carried out by an organizing committee composed of engineers in the same 30- to 45-year-old cohort as the participants. Each year different topics are covered and, with few exceptions, different individuals participate.

The speakers at the Frontiers of Engineering symposium have a unique challenge—to make the excitement of their field accessible to a technically sophisticated but nonspecialist audience. To achieve the objectives of the meeting, speakers are asked to provide a brief overview of their fields and to address such questions as: What are the frontiers in your field? What experiments, prototypes, and design studies are completed and in progress? What new tools and methodologies are being used? What are the current limitations on advances? What are the controversies? What is the theoretical, commercial, societal, and long-term significance of the work? Many elements of these topics are captured in the papers as well.

## CONTENT OF THE FOURTH ANNUAL SYMPOSIUM

Biomimetic robotic locomotion, tissue engineering, computational materials science, and simulation in manufacturing aluminum parts were just a few of the topics covered at the 1998 symposium. The four broad areas covered were biomaterials and optical engineering for biomedicine, advanced materials, simulation in manufacturing, and robotics. Presenters in the bio session talked about the challenges of creating living multidimensional tissues and organs, such as a liver; engineering macromolecular materials that combine the different virtues of natural and synthetic polymers; the benefits of optical coherence tomography, a type of optical biopsy that provides images of tissue in situ and in real time; and development of a confocal scanning laser microscope for the diagnosis of skin cancer without biopsy. The advanced materials session began with a presentation on current computational techniques being used to investigate the structure, dynamics, and properties of materials. This was followed by presentations on developments in advanced materials for applications in the integrated circuit and steel industries. Speakers in the session on simulation in manufacturing covered techniques used in the design of next generation engines for military aircraft, for analyzing surface roughness in formed aluminum parts, and for monitoring machine performance at remote sites in order to provide better service. In the robotics session, robot algorithms were discussed as well as applications such as snake-like robots that could assist in urban search-and-rescue operations following earthquakes, Mars rovers, and cobots—collaborative robots that interact with a human operator in a shared workspace. (See Appendixes for complete program.)

As in past years, there were lively discussions both during the question-and-answer sessions that followed each presentation and at breaks and mealtimes. The topics covered during these times ranged from technical questions to broader public policy issues related to engineering. For the second year in a row, the program included a tour of the Beckman Laser Institute.

William J. Perry, former Secretary of Defense and Berberian Professor of Engineering-Economic Systems and Operations Research at Stanford University, provided an insightful after-dinner address on the first evening of the symposium. His presentation, which is included in this volume, addressed the topic of technology innovation and American leadership. In it he cited examples of how innovations have played a critical role in the U.S. economy and national security and also described how technology is used not just for making weapons but also in national security decision-making. In conclusion, Dr. Perry quoted C. P. Snow, who wrote: "Technology is a queer thing. It brings you great gifts with one hand, and it stabs you in the back with the other." Dr. Perry noted that it is the emerging role of systems engineers to ensure that technology fulfills the former and guards against the latter.

As part of the ongoing process of seeking ways to make these meetings even more useful to participants, the attendees were asked to evaluate the Frontiers symposium. This feedback once again confirmed the value of the event. Attendees found that being informed about engineering areas they were not as familiar with was very useful and had the potential to affect their research and technical work. Others noted that the opportunity to interact with engineers from other sectors and disciplines was broadening and inspiring. It is clear that the symposium assessments by participants are very positive, and a survey is planned to solicit information on longer-term impacts of attending the meeting (e.g., collaborative work that has resulted from a contact there).

Funding for the Fourth Annual Symposium on Frontiers of Engineering was provided by the National Science Foundation (NSF), the U.S. Department of Defense (DOD), and NAE funds. The National Academy of Engineering would like to express its appreciation to NSF and DOD for supporting the symposium as well as to the members of the Symposium Organizing Committee (see p. *iii*) for their work in planning and organizing the event.

# Contents

## BIOMATERIALS AND OPTICAL IMAGING FOR BIOMEDICINE

## ADVANCED MATERIALS

## SIMULATION IN MANUFACTURING

## ROBOTICS

## DINNER SPEECH

## APPENDIXES

# FOURTH ANNUAL SYMPOSIUM ON FRONTIERS OF ENGINEERING

# BIOMATERIALS AND OPTICAL IMAGING FOR BIOMEDICINE

# Integration of Molecular and Macroscopic Scales in Tissue Engineering

LINDA G. GRIFFITH
*Department of Chemical Engineering*
*and*
*Division of Bioengineering and Environmental Health*
*Massachusetts Institute of Technology*
*Cambridge, Massachusetts*

## INTRODUCTION

The field of tissue engineering has emerged over the past decade, driven by a diverse range of clinical needs for replacement of diseased or damaged tissue and for delivering genetically engineered cells to patients (Langer and Vacanti, 1993). Although many artificial prosthetic devices are available to replace connective tissues such as joints, heart valves, blood vessels, and breasts, few synthetic devices are able to perform adequately over the lifetime of a patient, and devices vary greatly in their abilities to completely replace all of the functions of native tissue. No long-term replacements are available for some connective tissues, including heart, small-diameter blood vessels, and skin. Clinical needs are even more dire for such organs as livers, which can now be replaced only by organ transplantation.

The goal of tissue engineers is to meet these clinical needs by creating living three-dimensional tissues and organs using cells obtained from readily available sources such as biopsy of a patient's own (autologous) tissue or foreign tissue that would be discarded after surgical procedures such as circumcision. In every case the approach is to break the donor material down to the level of individual cells and then coax the isolated cells into forming a tissue structure of the appropriate size and/or shape by using a physical "scaffold" to organize cells on a macrosopic scale and providing molecular cues to stimulate appropriate cell growth, migration, and differentiation.

The scientific foundation for this new field lies in molecular cell biology, which has enabled the identification of hundreds of molecules involved in controlling cellular behavior from the external environment and development of new methods for assessing cellular responses. The engineering challenges in building on this science to meet clinical needs are at least twofold. The first challenge is to understand quantitatively how cells respond to these molecular signals and integrate multiple inputs to generate a given response—a significant challenge considering that the number of molecules identified so far represents only a fraction of the total that exist in the normal tissue environment. The second challenge is then to arrange cells in an appropriate three-dimensional configuration and present molecular signals in an appropriate spatial and temporal fashion so that the individual cells will grow and form the desired tissue structures—and do so in a way that can be carried out reproducibly, economically, and on a large scale. The specific examples described here are derived from work at the Massachusetts Institute of Technology and represent applications ranging from near term (less than one year) to very long term (more than 10 years).

## NEAR TERM—CONNECTIVE TISSUES

Tissues that have received the greatest attention from the commercial side—and thus are presumed to be feasible with current technology for near-term application in the clinic—include skin and cartilage. A common feature of these tissues is their relative avascularity and acellularity over dimensions important for tissue function, the ability to regenerate functional tissue from a single donor cell type, and the relative lack of cell-cell interactions in the normal tissue structure. Further, the cells in cartilage and skin grow readily in culture, allowing the number of cells from a single piece of donor tissue to be increased by several orders of magnitude. These factors make cartilage and skin highly attractive products for the emerging tissue engineering industry.

Among the many functions that cartilage serves in the human body, providing shape to such features as the outer ear and the nose is the least demanding from a tissue engineering perspective. Cartilage cells (chondrocytes) can readily be obtained by enzymatically digesting donor material to free cells from the extracellular matrix (ECM). Isolated chondrocytes placed in culture under appropriate conditions (e.g., in an agarose gel) exhibit a striking intrinsic ability to secrete ECM and form a stiff tissue similar to native cartilage. This property becomes useful for tissue engineering when the appropriate scaffold is used to direct tissue formation.

Our lab and many others have focused on synthetic bioresorbable polyesters in the polylactide/polyglycolide family as materials for scaffold construction in tissue engineering, as these materials have good mechanical properties, a long and favorable clinical record, are processable by solvent or thermal techniques, and break down by hydrolysis in body fluids to yield natural metabolites. In a collaboration initiated by a plastic surgeon at Boston Children's Hospital, we demonstrated that cartilage-like tissue in the shape of a human outer ear could be formed either in culture or by implanting beneath the skin a porous ear-shaped scaffold seeded with chondrocytes (Vacanti et al., 1992). The polyesters were chosen so that the degradation in mechanical strength of the scaffold was commensurate with the gain in mechanical strength of the tissue being formed by chondrocyte-secreted ECM, allowing the implant to retain its shape over the two-month period required for cartilage to form.

This sort of approach is being commercialized, notably for replacement of dermis of skin in diabetic ulcers and burn victims (Cooper et al., 1991). These tissues lead the way in determining what kinds of manufacturing and regulatory procedures will be needed for more complex applications.

## LONG-TERM—VASCULARIZED TISSUES

With few exceptions, tissues in the body are permeated by vascular networks to supply essential nutrients and regulatory factors. The distance between capillaries generally ranges from 20 to 200 microns, or about one to 10 cell diameters. The need for vascularity at almost the cellular scale is a major impediment to most cell-based approaches to tissue regeneration. Thus, tissues for which the microvascular network contributes strongly to overall tissue function, such as liver, have been more difficult to engineer and are correspondingly farther from clinical application. Highly vascularized tissues also tend to comprise several cell types arranged in a hierarchical structure, further complicating their reproduction by tissue engineering approaches. Cells derived from such tissues, such as hepatocytes from liver, often lose tissue-related functions when placed in culture; presumably, the hierarchy of structure also conveys a hierarchy of molecular control of cell behavior.

A place to start with such tissues is, then, understanding of the molecular signals that govern cell behavior. Virtually every aspect of cell behavior is governed at some level by interactions of transmembrane receptor molecules on the cell surface with ligands in the extracellular environment, and these complicated interactions are now being elucidated with the aid of engineering insights and analysis (Lauffenburger and Linderman, 1993). One example of a molecular control system is epidermal growth factor (EGF), which binds to the epidermal growth factor receptor (EGFR) to stimulate a diverse array of cell behaviors, including cell division. EGF was discovered in the early 1960s but has not yet entered clinical application despite many efforts to develop it and the wide range of effects it exerts under well-controlled in vitro conditions. It is inherently difficult to control local concentration of the peptide in vivo because of diffusive spread, cell uptake, and degradation of EGF. Further, when cells bind EGF, the EGF-EGFR complex is internalized and often degraded, leading to down regulation of receptors and attenuation of cellular responses. EGF is normally present in a soluble form, and some literature suggests that aspects of EGFR signaling can occur inside the cell after internalization. Nonetheless, we have shown that EGF retains biological activity when covalently tethered to the culture substrate where it has mobility but is physically prevented from entering the cell or diffusing away; further, cell response can be tuned by the density of tethered EGF ligand presented (Kuhl and Griffith-Cima, 1996). In addition to applications in tissue engineering, the ability to manipulate ligand presentation by purely physical means is now becoming a powerful tool for fundamental understanding and control of receptor function, complementing tools derived from molecular cell biology. The use of molecularly designed polymers is becoming a new tool for probing cell behaviors.

Moving to the macroscopic level, issues of how groups of cells can self-organize is important. It is becoming more apparent that the molecular signals from ECM may at least partially govern cell behavior by measurable biophysical outcomes such as the relative magnitude of adhesive bonds (Lauffenburger and Horwitz, 1996). For example, the morphology of aggregates of hepatocytes in culture—spread or spheriodal—can be predicted on the basis of the relative magnitudes of cell-substrate adhesion strength and cell contractile forces, and the morphology of more complex structures obtained from mixed hepatocyte/endothelial cultures can also be predicted based on relative cell-cell and cell-substrate adhesion strengths (Powers et al., 1997). We are exploiting this self-organizing ability of cells to generate three-dimensional tissue with a perfused microvascular structure in vitro starting from dispersed cells and a scaffold of precisely defined architecture and surface chemistry. Our initial focus is on trying to re-create the smallest functional unit in a tissue, the capillary bed, which has dimensions of 0.1 to 0.8 mm depending on the tissue and is ~0.8 mm for liver. We have developed and implemented new materials processing techniques to create scaffolds for this purpose, a technique that will allow integration of molecular cues, such as tethered EGF, with macroscopic signals (Griffith et al., 1997).

On the in vivo therapeutic side, our ultimate aim is to create a liver that can be transplanted directly into the portal vein of a patient beginning with a complex hierarchical scaffold seeded with cells. To create scaffolds with this high degree of complexity, we have implemented a solid free-form fabrication (SFF) technique, the 3 Dimensional Printing (3DP™) process. In SFF techniques, devices are built as a series of thin sequential layers. We have adapted the original 3DP™ process, developed for fabrication of ceramics and metals, to a range of polymeric and composite organic/inorganic materials important in tissue engineering and drug delivery. This technology is currently being commercialized.

In addition to the huge clinical needs for vascularized tissues, in vitro models of vascularized human tissues are desperately needed in the pharmaceuticals industry for determining the effects of drugs on humans and for basic research into disease states and development. Generation of tissue that can be controllably perfused in vitro and maintained in culture for extended periods will enhance virtually every effort to study tissue functions in vitro. We are implementing silicon microfabrication technology to create a miniature "liver on a chip" to allow rapid screening of drugs and other compounds.

## REFERENCES

Cooper, M. L., J. F. Hansbrough, R. L. Spielvogel, R. Cohen, R. L. Bartel, and G. Naughton. 1991. In vivo optimization of a living dermal substitute employing cultured human fibroblasts on a biodegradable polyglycolic acid or polyglactin mesh. Biomaterials 12(2):243–248.

Griffith, L., B. Wu, M. J. Cima, B. Chaignaud, and J. P. Vacanti. 1997. In vitro organogenesis of liver tissue. Annals of the New York Academy of Sciences 831(December 31):382–397.

Kuhl, P. R., and L. G. Griffith-Cima. 1996. Tethered epidermal growth factor as a paradigm for growth factor-induced stimulation from the solid phase. Nature Medicine 2(9):1022–1027.

Lauffenburger, D. A., and A. F. Horwitz. 1996. Cell migration: A physically integrated molecular process. Cell 84(3):359–369.

Lauffenburger, D. A., and J. L. Linderman. 1993. Receptors: Models for Binding, Trafficking, and Signaling. New York: Oxford University Press.

Langer, R., and J. P. Vacanti. 1993. Tissue engineering. Science 260(5110):920–926.

Powers, M. J., R. E. Rodriguez, and L. G. Griffith. 1997. Cell-substratum adhesion strength as a determinant of hepatocyte aggregate morphology. Biotechnology and Bioengineering 53(4): 415–426.

Vacanti, C. A., L. G. Cima, D. Rodkowski, and J. Upton. 1992. Tissue engineered growth of new cartilage in the shape of a human ear using synthetic polymers seeded with chondrocytes. Pp. 323–330 in Tissue-Inducing Biomaterials, Materials Research Society Symposium Proceedings, Vol. 252, L. G. Cima and E. S. Ron, eds. Pittsburgh, Pa.: Materials Research Society.

# Artificial Proteins: Bridging the Gap Between Natural and Synthetic Macromolecular Materials

DAVID A. TIRRELL
*Division of Chemistry and Chemical Engineering*
*California Institute of Technology*
*Pasadena, California*

How should one approach the engineering of macromolecular materials that combine the very different virtues of natural and synthetic polymers? Natural polymers, especially proteins and nucleic acids, serve as selective catalysts and as efficient information storage devices, while their synthetic counterparts (polyethylene, polypropylene, and so on) dominate modern materials technology because of their excellent mechanical, barrier, and processing behaviors (Rodriguez, 1996). The molecular architectural features that underlie these disparate patterns of behavior are strikingly different: proteins and nucleic acids are characterized by precisely defined chain lengths and sequences, whereas synthetic polymeric materials consist of complex mixtures of chain molecules in which length, sequence, and stereochemistry (molecular shape) vary widely from chain to chain.

## TEMPLATE POLYMERIZATION

The architectural differences between natural and synthetic polymers arise from differences in the nature of the polymerization processes that lead to each class of materials. Proteins and nucleic acids are made via "template polymerizations" in which each copy of the polymer chain is assembled on a nucleic acid template that dictates the length and sequence of each product molecule. Synthetic polymers, on the other hand, arise from statistical polymerization processes in which the relative rates of competing initiation, propagation, and termination steps determine the average chain structure as well as the distribution of structures formed. It is not surprising that these classes of macromolecules behave so differently; it is difficult to imagine, for example, how one could store genetic information in a population of chains in which product structure cannot be traced back to the structure of a precursor molecule.

The power of template polymerizations has been recognized by polymer materials scientists for decades, and the literature reports many attempts to use templates to control polymer structure and polymerization rate (Tan, 1989). These attempts have met with near total failure. To be sure, the addition of a template to a polymerizing mixture is often accompanied by changes in product structure or reaction rate, but the observed changes are rarely predictable and the causes of change rarely clear. These results highlight the difficulty of de novo design of molecular templates for control of chemical synthesis.

## NUCLEIC ACID TEMPLATES

An alternative approach to this problem acknowledges the difficulty of template design and exploits the known success of nucleic acids as templates for protein synthesis. Recombinant DNA methods now allow cells to be outfitted with genes that encode any desired sequence of amino acids, even if that sequence bears no relation to any natural protein. It has also been demonstrated that the protein biosynthetic apparatus can accommodate monomers other than the 20 amino acids normally used to build cellular proteins. The materials engineer is thus presented with an important new opportunity—that of designing polymeric materials of precisely controlled architectures without sacrificing the versatility characteristic of synthetic polymers. Thus, it may indeed be possible to combine the virtues of natural and synthetic macromolecular materials.

## ARTIFICIAL PROTEINS

Two complementary approaches to this problem have been explored over the past several years. The first uses recombinant DNA methods (or in some cases organic chemical synthesis) to prepare variants of natural structural proteins such as silk, elastin, or marine adhesives. Here biology serves very directly as a guide to new materials science in that the investigator makes relatively minor changes in polymer structure in the hope that the useful properties of the natural material can be preserved in a simplified variant that is more readily prepared. The most encouraging successes of this approach have been reported for analogs of elastin, a protein that—as its name implies—contributes to the elastic properties of various tissues. Simple elastin analogs have been shown to exhibit excellent mechanical properties and have been engineered to change shape in response to a variety of chemical and physical signals (Urry, 1993). Efforts to develop such materials for surgical and pharmaceutical applications are under way.

The second approach starts from the perspective of the polymer materials scientist and involves de novo design of artificial proteins. Protein structure and function continue to enter into the design process, and in some cases natural structural motifs are exploited, but the designs also embrace ideas drawn from the science and technology of synthetic polymeric materials. Because the materials of interest often differ substantially from any known natural proteins, this method tests the capacity of the biosynthetic apparatus to make novel and diverse macromolecular structures. The messenger RNA template that guides protein synthesis in this approach is derived not from any natural gene but rather from an artificial gene that specifies the sequence of amino acids dictated by the materials design process. The template polymerization that nature uses to build cellular proteins is co-opted by the materials engineer for his or her own purposes.

This approach has been used to prepare several new families of artificial proteins of interest in materials science and engineering. Early experiments addressed the concept of "encoded self-assembly," in which the artificial gene is designed to control not only the covalent structure of the chain but also the manner in which polymer chains assemble into larger-scale aggregates such as crystals or liquid crystals. By designing periodic polypeptide sequences, for example, one can control crystal dimensions and surface chemistry in ways that are not available to conventional polymeric materials (Krejchi et al., 1994). In similar fashion, by making populations of uniform rodlike polymers, one gains access to novel liquid crystal phases in polymer solutions and films (Yu et al., 1997).

## NEW BUILDING BLOCKS FOR ARTIFICIAL PROTEINS

Perhaps the most ambitious use of artificial genes to direct polymer synthesis involves the incorporation of amino acids other than the 20 normally used by cells to build proteins. It has been known for many years that analogs of some of the "canonical" amino acids can be activated and charged to transfer RNAs for incorporation into the growing protein chain, and recent experiments have shown that substitution by analogs can effect important changes in materials properties. Replacement of hydrogen by fluorine is particularly useful in this regard; hydrogen and fluorine differ little in size, making substitution straightforward in most instances. At the same time, fluoropolymers such as Teflon exhibit valuable surface properties that are markedly different from those of their hydrocarbon relatives. Other successful examples include amino acids bearing thienyl, alkenyl, and alkynyl side chains (van Hest and Tirrell, 1998).

## CHIMERIC ARTIFICIAL PROTEINS

Finally, one can combine the two approaches described above to create "chimeric" or "hybrid" proteins that include both natural and artificial polypeptide sequences. This may in fact be the simplest answer to the question posed at the beginning of this paper; the natural domain of the chimera performs its normal function while the artificial domain confers the needed materials properties such as strength, toughness, or adhesion to surfaces. In a recent example, so-called leucine-zipper peptide domains were attached to the ends of a water-soluble polypeptide to create triblock chimeric proteins in which the leucine zippers control polymer-polymer interactions in a manner that leads to reversible gelation of the protein solution in response to changes in pH or temperature (Petka et al., 1998).

## REFERENCES

Krejchi, M. T., E. D. T. Atkins, A. J. Waddon, M. J. Fournier, T. L. Mason, and D. A. Tirrell. 1994. Chemical sequence control of ß-sheet assembly in macromolecular crystals of periodic polypeptides. Science 265:1427–1432.

Petka, W. A., J. L. Harden, K. P. McGrath, D. Wirtz, and D. A. Tirrell. 1998. Reversible hydrogels from self-assembling artificial proteins. Science 281:389–392.

Rodriguez, F. 1996. Principles of Polymer Systems, 4th ed. Washington, D.C.: Taylor & Francis.

Tan, Y. Y. 1989. Pp. 245–259 in Comprehensive Polymer Science: Template Polymerization. Oxford, England: Pergamon Press.

Urry, D. W. 1993. Molecular machines: How motion and other functions of living organisms can result from reversible chemical changes. Angewandte Chemie. International Edition in English 32:819–841.

van Hest, J. C. M., and D. A. Tirrell. 1998. Efficient introduction of alkene functionality into proteins in vivo. FEBS Letters 428:68–70.

Yu, S. M., V. Conticello, C. Kayser, M. J. Fournier, T. L. Mason, and D. A. Tirrell. 1997. Smectic ordering in solutions and films of a rod-like polymer owing to monodispersity of chain length. Nature 389:167–170.

# Biomedical Imaging Using Optical Coherence Tomography

JAMES G. FUJIMOTO
*Department of Electrical Engineering and Computer Science*
*and*
*Research Laboratory of Electronics*
*Massachusetts Institute of Technology*
*Cambridge, Massachusetts*

## ABSTRACT

*Optical coherence tomography (OCT) is a new technology for performing high-resolution cross-sectional imaging. OCT functions as a type of optical biopsy that provides cross-sectional images of tissue structure on the micron scale. In combination with catheters and endoscopes, OCT can perform internal body imaging. OCT is a powerful imaging technology for medical diagnostics because, unlike conventional histopathology, which requires removal of a tissue specimen and processing for microscopic examination, OCT can provide images of tissue in situ and in real time.*

## INTRODUCTION

OCT is a recently developed optical imaging technique that performs high-resolution, cross-sectional imaging of microstructures in biological systems (Huang et al., 1991). OCT is analogous to ultrasound B mode imaging except that it uses light instead of sound. OCT performs imaging by using low-coherence interferometry to measure the optical backscattering of tissue as a function of echo delay and transverse position. The resulting two-dimensional dataset can be displayed as a gray scale or false color image.

OCT was originally developed and applied by our group for tomographic diagnostics in ophthalmology. OCT can provide images of the retina with resolutions of 10 μm, one order of magnitude higher than conventional ultrasound (Hee et al., 1995). Working in collaboration with the New England Eye Center and MIT's Lincoln Laboratory, we developed a clinical prototype OCT instrument for ophthalmic diagnosis. Several thousand patients have been examined to date (Hee et al., 1995; Puliafito et al., 1995a,b). The technology has been transferred to industry, and a commercial product was introduced into the ophthalmic market in 1996. More recently, advances in OCT imaging have enabled imaging to be performed in nontransparent tissues, thus enabling its application in a wide range of possible medical specialties (Brezinski et al., 1996; Fujimoto et al., 1995; Schmitt et al., 1994, 1995). Imaging depth is limited by optical attenuation due to scattering and absorption. However, in most tissues, imaging 2 to 3 mm deep can be achieved. OCT has been applied in vitro to image arterial pathology where it can differentiate plaque morphology with superior resolution to ultrasound (Brezinski et al., 1996). Imaging studies have also been performed in vitro to investigate applications in gastroenterology, urology, gynecology, surgery, and neurosurgery (Brezinski et al., 1997a,b; Tearney et al., 1997b,c). OCT has been applied in vivo to image developing biological specimens (African frog, leopard frog, and zebrafish tadpoles and embryos). For applications in developmental biology, OCT can permit the repeated imaging of developing morphology without the need to sacrifice specimens (Boppart et al., 1996).

OCT is a promising and powerful medical imaging technology because it can permit real-time in situ visualization of tissue microstructure without the need to excisionally remove and process a specimen, as in conventional biopsy and histopathology. The concept of “nonexcisional optical biopsy” provided by OCT and the ability to visualize tissue morphology in real time under operator guidance can be used both for diagnostic imaging and to guide surgical intervention. Coupled with catheter, endoscopic, or laparoscopic delivery, OCT holds the promise of having a widespread impact on medicine, ranging from improving the screening and diagnosis of cancer to enabling new microsurgical and minimally invasive surgical procedures.

## PRINCIPLES OF OPERATION AND TECHNOLOGY

OCT is analogous to ultrasound imaging but is based on optical ranging and the high-resolution, high-dynamic-range detection of backscattered light. In contrast to ultrasound, because the velocity of light is extremely high, the echo time delay of reflected light cannot be measured directly. One method for measuring the time delay of light is to use low-coherence interferometry or optical coherence domain reflectometry. Low-coherence interferometry was first developed for measuring reflections in fiber optics and optoelectronic devices and was first demonstrated in ophthalmology for measurements of axial eye length and corneal thickness (Fercher et al., 1988; Hitzenberger, 1991; Takada et al., 1987). Low-coherence interferometry uses heterodyne detection of light backscattered from the sample. Interference of the light reflected from the sample arm and reference arm of a Michelson interferometer (see Figure 1) can occur only when the optical path lengths of the two arms match to within the coherence length of

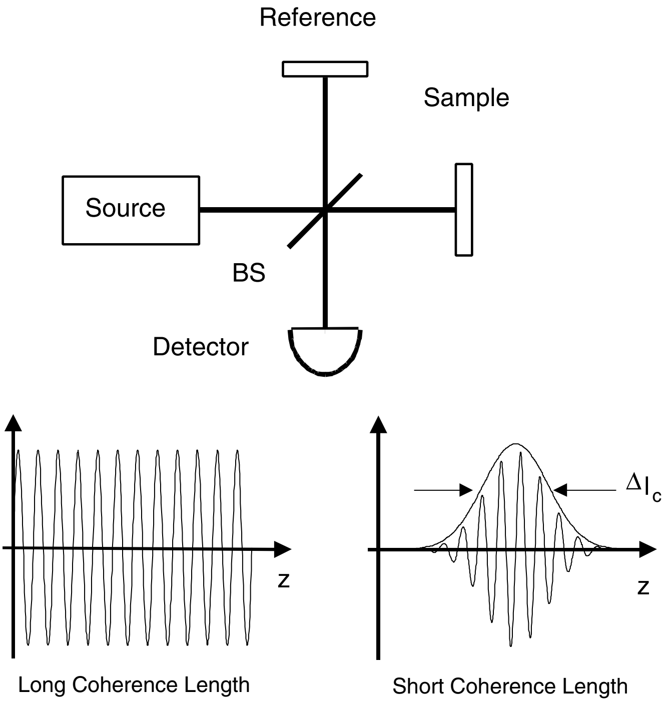


**FIGURE 1** OCT measures the echo time delay of reflected light by using low-coherence interferometry. The system is based on a Michelson-type interferometer. Reflections or backscattering from the object being imaged are correlated with light, which traverses a reference path.

the optical source. As the reference arm optical path length is scanned, backscattering sites within the sample arm are localized.

Figure 2 is a schematic illustrating how OCT performs cross-sectional imaging. The optical beam is focused into the object being imaged, and the echo time delay and intensity of the backscattered light are measured to yield an axial backscattering profile. The incident beam is then scanned in the transverse direction, and the axial backscattering is measured at several transverse positions to yield a two-dimensional dataset. This dataset represents the backscattering or back reflection through a cross-section of the object being imaged and can be displayed as a gray scale or false color image.

The axial resolution in OCT images is determined by the coherence length of the light source. The interference signal detected at the output port of the interferometer is the electric field autocorrelation of the source. The coherence length is the spatial width of the field autocorrelation. The envelope of the field autocorrelation is equivalent to the Fourier transform of its power spectrum.

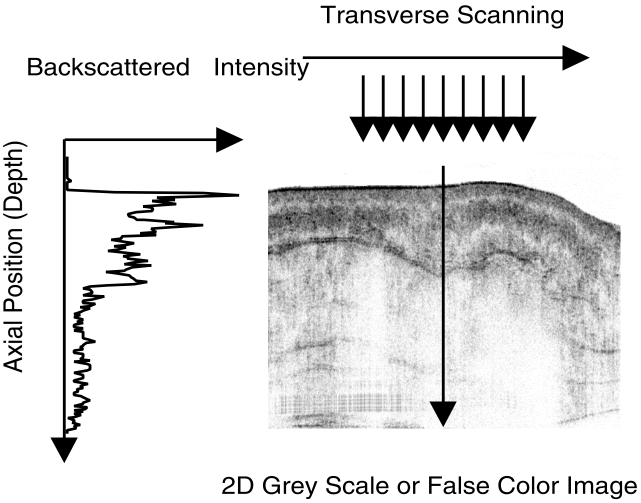


**FIGURE 2** Cross-sectional images are constructed by performing measurements of the echo time delay of light at different transverse positions. The result is a two-dimensional dataset that represents the backscattering in a cross-sectional plane of the tissue.

Thus, the width of the autocorrelation function, or the axial resolution, is inversely proportional to the width of the power spectrum. For a source with a Gaussian spectral distribution, the axial resolution $\Delta z$ is given by: $\Delta z = (2\ln 2/\pi)(\lambda^2/\Delta\lambda)$, where $\Delta z$ and $\Delta\lambda$ are the full widths at half maximum of the autocorrelation function and power spectrum, respectively, and $\lambda$ is the source central wavelength. Thus, broad-bandwidth optical sources are required to achieve high axial resolution. The transverse resolution achieved with an OCT imaging system is determined by the focused spot size in analogy with conventional microscopy. The transverse resolution is given by: $\Delta x = (4\lambda/\pi)(f/d)$, where $d$ is the spot size on the objective lens and $f$ is its focal length. High-transverse resolution can be obtained by using a large numerical aperture and focusing the beam to a small spot size.

OCT can be implemented using fiber optic technology. Figure 3 shows a schematic of an OCT system that uses a fiber optic Michelson-type interferometer. A low coherence light source is coupled into the interferometer and the interference at the output is detected with a photodiode. One arm of the interferometer emits a beam that is directed and scanned on the object being imaged, while the other arm of the interferometer is a reference arm with a scanning delay line. Because OCT uses fiber optics, it can easily be interfaced to a wide range of optical instruments.

For research applications, short-pulse lasers are used as light sources for

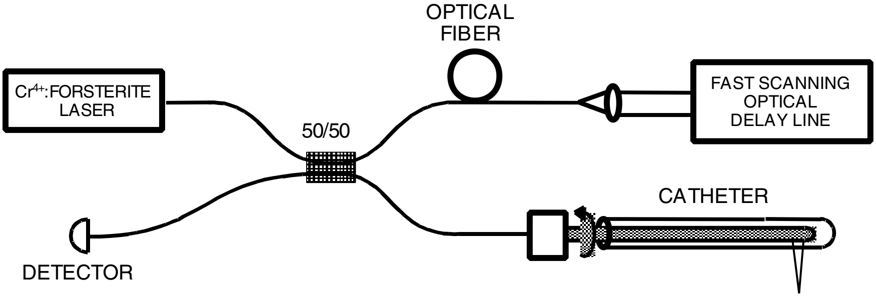


**FIGURE 3** Schematic of OCT instrument based on fiber optic implementation of a Michelson interferometer. One arm of the interferometer is interfaced to the measurement instrument, and the other arm has a scanning delay line. The system shown is configured for high-speed catheter-endoscope-based imaging.

OCT imaging because they have extremely short coherence lengths and high-output powers, thereby enabling high-resolution, high-speed imaging. For clinical applications, compact superluminescent diodes or semiconductor-based light sources can be used. The laser source for many of our studies was a short-pulse $Cr^{4+}$:Forsterite laser, which operates near 1,300 nm and achieves an axial resolution of 5 to 10 μm with a signal-to-noise ratio of 110 dB (Boppart et al., 1998). A rapidly scanning optical delay line based on a grating phase control device, similar to that used in ultrafast optics, is used for delay scanning (Tearney et al., 1997a). The grating-phase control device is attractive because it permits the phase and group velocity of the scan to be independently controlled and achieves extremely high scan speed. Images of 250 to 500 transverse pixels can be produced at four to eight frames per second. Future systems will operate at video rates.

## APPLICATIONS

OCT is a promising and powerful medical imaging technique because it can permit in situ and real-time visualization of tissue microstructure with resolutions that are one to two orders of magnitude higher than ultrasound. OCT was initially applied for imaging in the eye, and to date, has had the broadest clinical impact in ophthalmology. Figure 4 shows an example of an OCT image of the retina of a human subject (Hee et al., 1995). This image was taken at a 10-μm resolution and allows the detailed structure of the retina to be differentiated. The retinal thickness can be easily measured as well as the retinal nerve fiber layer, which is visible as a highly backscattered layer. Clinical studies in ophthalmology show that OCT is especially promising for the diagnosis and monitoring of

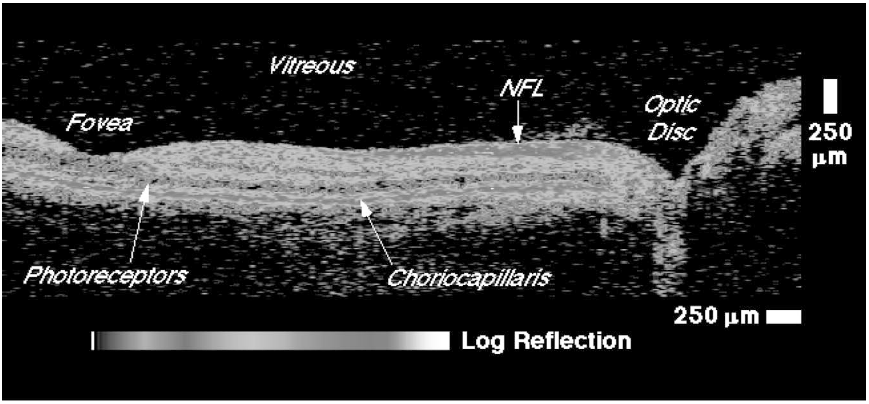


**FIGURE 4** OCT image of the human retina papillary-macular axis illustrating the ability to discriminate structural morphology in vivo. The highly backscattered retinal nerve fiber layer and choriocapillaris appear red. The optic disk as well as several of the retinal layers are observed. SOURCE: Reprinted with permission from the American Medical Association (Hee et al., 1995).

such diseases as glaucoma and macular edema associated with diabetic retinopathy, where it provides quantitative information on disease progression (Puliafito et al., 1995b). In many cases OCT has the ability to detect and diagnose early stages of disease before physical symptoms and loss of vision occur.

OCT can also be used for a variety of applications in internal medicine and internal body imaging of scattering tissues. Although the image penetration depth is limited to a few millimeters, this scale is comparable to the depth over which many biopsies are performed and many diagnostically important changes in tissue morphology occur near tissue or organ surfaces. OCT can resolve changes in architectural morphology that are important for diagnosis of diseases such as early neoplastic changes. Figure 5 shows an example of OCT images of human gastrointestinal tissues in vitro. These images were performed at 15-µm resolution using a 1,300-nm wavelength. The top image shows the structure of the normal esophageal mucosal tissue, which is characterized by a horizontally organized squamous epithelial structure. The middle image shows the structure of the normal intestinal mucosal tissue, which has a vertically organized columnar epithelial structure. Even at modest resolutions of 15 µm, the differences between the architectural morphology of these tissue types is evident. The bottom image shows an ampullary carcinoma. The carcinoma is evident in the left one-third of the image and is characterized by a loss of the glandular organization of the tissue. Normal bowel is seen at the right one-third of the image and is characterized by vertically organized columnar epithelial structure. The center part of the image shows a progressive disorganization and loss of structure.

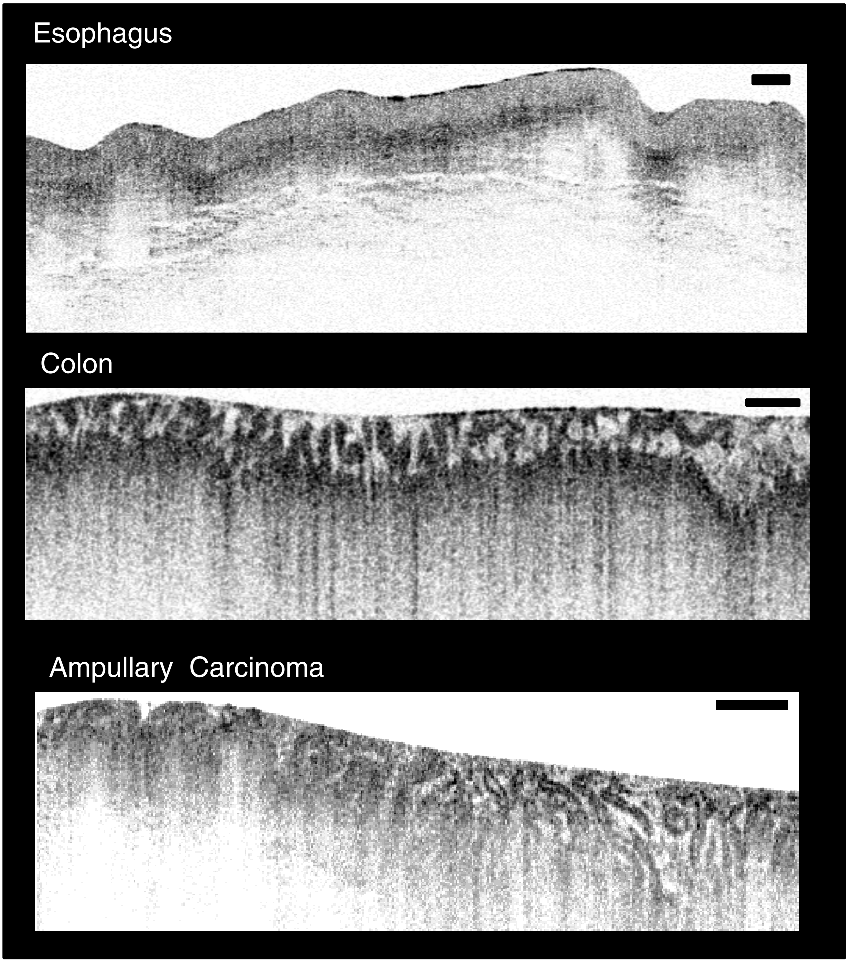


**FIGURE 5** In vitro images of gastrointestinal tissues and pathology, including normal human esophagus, colon, and ampullary carcinoma. These images illustrate the ability of OCT to discriminate architectural morphology.

Changes in architectural morphology such as these can be used for the screening and diagnosis of early neoplastic changes. Conventional excisional biopsy often suffers from high false-negative rates because the biopsy process relies on sampling tissue and the diseased tissues can easily be missed. OCT might be used to identify suspect lesions and guide excisional biopsy in order to reduce the false-negative rates. In future applications, when sufficient clinical data are available, OCT might be used directly for diagnosis.

For OCT imaging in internal organ systems, it is necessary to develop optical delivery technologies (Boppart et al., 1997; Tearney et al., 1996). Using fiber optics, a catheter-endoscope with an outer diameter of 2.9 French or 1.0 mm has been constructed. Figure 6 shows a schematic of an OCT catheter-endoscope. A single-mode optical fiber runs the length of the catheter, and the distal end consists of a GRIN lens and a microprism to direct the OCT beam radially. The fiber and distal optics are rotated so that the OCT beam scans an angular, radar-like, pattern and image cross-sectionally through internal organs.

The catheter-endoscope OCT system enables the acquisition of in vivo images of internal organ systems. Figure 7 shows an example of a catheter-endoscope OCT image of the esophagus of a rabbit in vivo. In vivo imaging of the pulmonary tract, gastrointestinal tract, and urinary tract as well as arterial imaging have been performed. These studies demonstrate the feasibility of performing OCT imaging of internal organ systems and suggest the possibility of its application clinically (Tearney et al., 1997a). Other research groups as well as ours are currently beginning OCT imaging studies in patients, and we expect results to be published shortly.

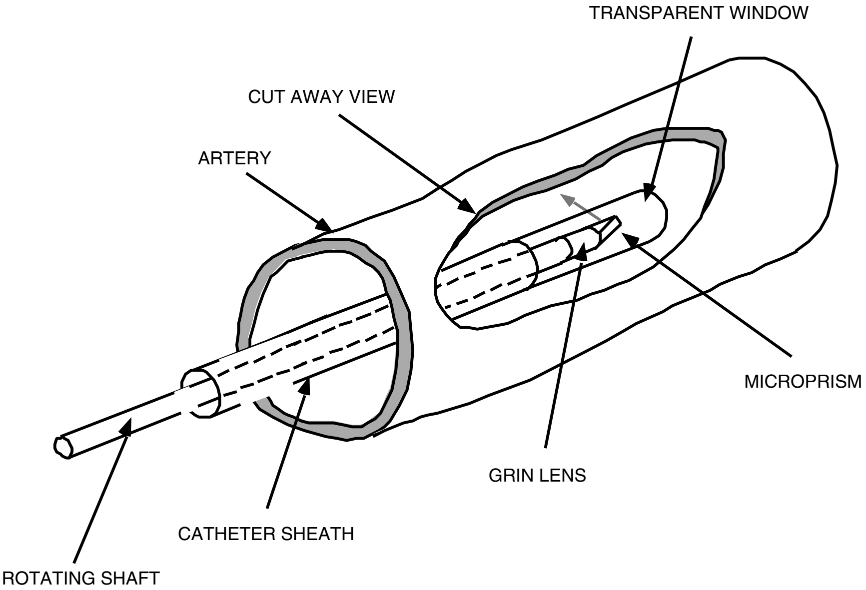


**FIGURE 6** OCT catheter for transverse intraluminal imaging. A single-mode fiber lies within a rotating flexible speedometer cable enclosed in a protective plastic sheath. The distal end focuses the beam at 90 degrees from the axis of the catheter. SOURCE: Reprinted with permission from Optical Society of America (Tearney et al., 1996).

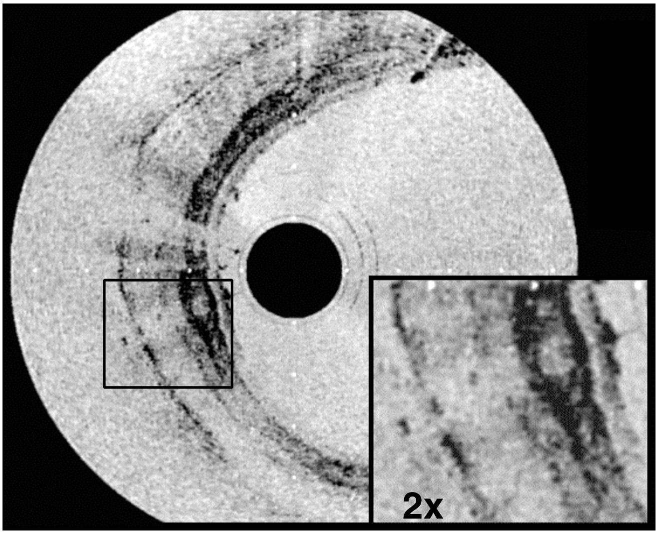


**FIGURE 7** OCT image of esophageal structures of New Zealand White rabbit in vivo performed by using prototype transverse scanning OCT catheter and high-speed imaging system. Image acquisition time was 250 ms per image for 300-pixel radial image.

The development of high-resolution OCT is also an important area of active research. Increasing resolutions to the cellular and subcellular levels are important for many applications, including the diagnosis of early neoplasias. One of the keys to achieving high resolution is the use of short-pulsed lasers to obtain short coherence length. High-resolution OCT imaging has been demonstrated in vivo in developmental biology specimens. Figure 8 shows an example of high-resolution OCT images of a *Xenopus laevis* (African frog) tadpole (Boppart et al., 1998). The OCT beam was focused to a 9-μm diameter spot (100-μm confocal parameter). A short-pulse $Cr^{4+}$:Forsterite laser that operates near 1,300 nm was used as the light source for these measurements. The free-space axial resolution was 5.1 μm. Assuming an average index of 1.35 for these specimens, the in vivo axial resolution was ~ 3.8 μm. In developmental biology the ability to image subcellular structure can be an important tool for studying mitotic activity and cell migration, which occur during development. The extension of these results to humans has important implications for the diagnosis of early neoplasias. Many of these diseases are manifest by changes that occur on cellular and subcellular levels.

The development of high-resolution high-speed OCT technology as well as OCT-compatible catheter-endoscopes represents enabling steps for many OCT imaging applications, including future endoscopic clinical applications. OCT is

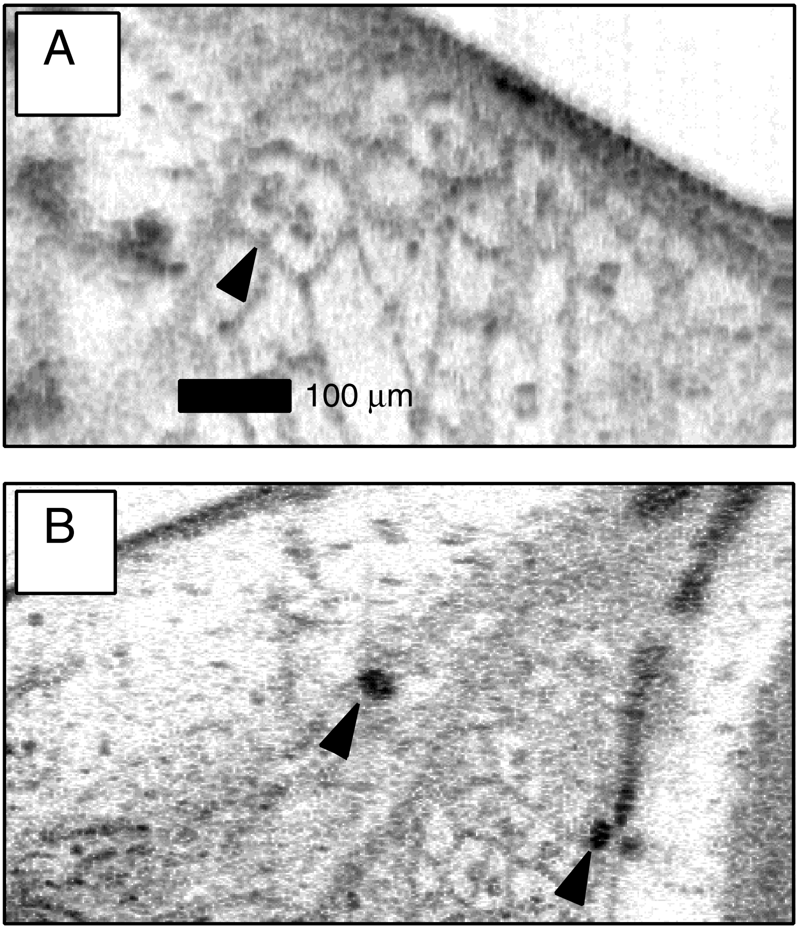


**FIGURE 8** High-resolution OCT images of a *Xenopus laevis* (African frog) tadpole in vivo. (A) Mesenchymal cells immediately following cell division with two daughter cells (arrow). Cell membranes and individual cell nuclei are apparent. (B) Melanin-laden neural crest cells, which migrate during development.

a powerful technique for optical biopsy because it can perform micron-scale imaging of cellular and architectural morphology in situ and in real time. Imaging information is available in real time without the need for excision and histological processing of a specimen. The capability to perform rapid in situ imaging can be used in a variety of clinical scenarios, including (1) to guide conventional biopsy and reduce false-negative rates due to sampling errors, (2) to perform imaging of tissue microstructure in situations where conventional excisional biopsy would be hazardous or impossible, and (3) to guide surgical or microsurgical intervention. More research remains to be done, and numerous clinical studies must be performed to determine in which clinical situations OCT can play a decisive role. The unique capabilities of OCT imaging suggest that it has the potential to have a significant impact on the diagnosis and clinical management of many diseases.

## ACKNOWLEDGMENTS

This work was performed in collaboration with Mark Brezinski of the Harvard Medical School and Massachusetts General Hospital. The contributions of Steve Boppart, Brett Bouma, Gary Tearney, Juergen Herrmann, Wolfgang Drexler, Xingde Li, and Costas Pitris are gratefully acknowledged. This research is supported in part by National Institutes of Health contracts NIH-9-RO1-CA75289-01 and NIH-9-RO1-EY11289-13 and the Medical Free Electron Laser Program, Office of Naval Research contract N000014-97-1-1066.

## REFERENCES

Boppart, S. A., M. E. Brezinski, B. Bouma, G. J. Tearney, and J. G. Fujimoto. 1996. Investigation of developing embryonic morphology using optical coherence tomography. Developmental Biology 177(1):54–63.

Boppart, S. A., B. E. Bouma, C. Pitris, G. J. Tearney, J. G. Fujimoto, and M. E. Brezinski. 1997. Forward-imaging instruments for optical coherence tomography. Optics Letters 22(21):1618–1620.

Boppart, S. A., B. E. Bouma, C. Pitris, J. F. Southern, M. E. Brezinski, and J. G. Fujimoto. 1998. In vivo cellular optical coherence tomography imaging. Nature Medicine 4(7):861–865.

Brezinski, M. E., G. J. Tearney, B. E. Bouma, J. A. Izatt, M. R. Hee, E. A. Swanson, J. F. Southern, and J. G. Fujimoto. 1996. Optical coherence tomography for optical biopsy: Properties and demonstration of vascular pathology. Circulation 93(6):1206–1213.

Brezinski, M. E., G. J. Tearney, S. A. Boppart, E. A. Swanson, J. F. Southern, and J. G. Fujimoto. 1997a. Optical biopsy with optical coherence tomography: Feasibility for surgical diagnostics. Journal of Surgical Research 71(1):32–40.

Brezinski, M. E., G. J. Tearney, N. J. Weissman, S. A. Boppart, B. E. Bouma, M. R. Hee, A. E. Weyman, E. A. Swanson, J. F. Southern, and J. G. Fujimoto. 1997b. Assessing atherosclerotic plaque morphology: Comparison of optical coherence tomography and high frequency intravascular ultrasound. Heart 77(5):397–403.

Fercher, A. F., K. Mengedoht, and W. Werner. 1988. Eye-length measurement by interferometry with partially coherent light. Optics Letters 13(3):186–188.

Fujimoto, J. G., M. E. Brezinski, G. J. Tearney, S. A. Boppart, B. E. Bouma, M. R. Hee, J. F. Southern, and E. A. Swanson. 1995. Biomedical imaging and optical biopsy using optical coherence tomography. Nature Medicine 1(9):970–972.

Hee, M. R., J. A. Izatt, E. A. Swanson, D. Huang, C. P. Lin, J. S. Schuman, C. A. Puliafito, and J. G. Fujimoto. 1995. Optical coherence tomography of the human retina. Archives of Ophthalmology 113(3):325–332.

Hitzenberger, C. K. 1991. Measurement of the axial eye length by laser Doppler interferometry. Investigative Ophthalmology and Visual Science 32(3):616–624.

Huang, D., E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto. 1991. Optical coherence tomography. Science 254(5035):1178–1181.

Puliafito, C. A., M. R. Hee, C. P. Lin, E. Reichel, J. S. Schuman, J. S. Duker, J. A. Izatt, E. A. Swanson, and J. G. Fujimoto. 1995a. Imaging of macular disease with optical coherence tomography. Ophthalmology 102(2):217–229.

Puliafito, C.A., M. R. Hee, J. S. Schuman, and J. G. Fujimoto. 1995b. Optical Coherence Tomography of Ocular Diseases. Thorofare, N.J.: Slack, Inc.

Schmitt, J. M., A. Knuttel, M. Yadlowsky, and A. A. Eckhaus. 1994. Optical coherence tomography of a dense tissue: Statistics of attenuation and backscattering. Physics in Medicine and Biology 39(10):1705–1720.

Schmitt, J. M., M. J. Yadlowsky, and R. F. Bonner. 1995. Subsurface imaging of living skin with optical coherence microscopy. Dermatology 191(2):93–98.

Takada, K., I. Yokohama, K. Chida, and J. Noda. 1987. New measurement system for fault location in optical waveguide devices based on an interferometric technique. Applied Optics 26(9):1603–1606.

Tearney, G. J., S. A. Boppart, B. Bouma, M. E. Brezinski, N. J. Weissman, J. F. Southern, and J. G. Fujimoto. 1996. Scanning single mode catheter/endoscope for optical coherence tomography. Optics Letters 21(7):543–545.

Tearney, G. J., M. E. Brezinski, B. E. Bouma, S. A. Boppart, C. Pitris, J. F. Southern, and J. G. Fujimoto. 1997a. In vivo endoscopic optical biopsy with optical coherence tomography. Science 276(5321):2037–2039.

Tearney, G. J., M. E. Brezinski, J. F. Southern, B. E. Bouma, S. A. Boppart, and J. G. Fujimoto. 1997b. Optical biopsy in human gastrointestinal tissue using optical coherence tomography. American Journal of Gastroenterology 92(10):1800–1804.

Tearney, G. J., M. E. Brezinski, J. F. Southern, B. E. Bouma, S. A. Boppart, and J. G. Fujimoto. 1997c. Optical biopsy in human urologic tissue using optical coherence tomography. Journal of Urology 157(5):1915–1919.

# Confocal Reflectance Microscopy: Diagnosis of Skin Cancer Without Biopsy?

MILIND RAJADHYAKSHA
*Lucid, Inc., Henrietta, New York*
*and*
*Massachusetts General Hospital, Boston*

## INTRODUCTION

X-ray tomography, magnetic resonance imaging, and ultrasound are noninvasive biomedical imaging modalities commonly used in the clinic. These imaging modalities have resolutions of 10 to 1,000 μm, which allows assessment of the gross (macro) structure of living tissue but not its detailed cellular and nuclear microstructures. Clinical assessment of cellular and nuclear microstructure (histology) requires a resolution of 0.1 to 10 μm and is performed by conventional optical microscopy. Conventional microscopy is invasive: one must remove (biopsy) the tissue, fix or freeze, excise into thin sections (typically 5-μm slices), and stain with dyes to enhance contrast. Biopsies destroy the site being investigated and prevent subsequent imaging of dynamic events. Tissue processing introduces artifacts and is expensive and time consuming. An alternative technique that potentially avoids biopsies or tissue processing is confocal reflectance microscopy. A confocal microscope can noninvasively image cellular and nuclear microstructures in thin sections within living tissue with high resolution and contrast (Pawley, 1995; Webb, 1996).

Confocal reflectance microscopy offers a noninvasive window into living human tissue for basic and clinical research. No biopsy, processing, or staining of tissue is necessary. Imaging is based on the detection of backscattered light with contrast due to naturally occurring refractive index variations of tissue microstructures. Between 1991 and 1996, scientists at L'Oreal in France demonstrated confocal imaging of living human skin with a white-light tandem scanning microscope (Bertrand and Corcuff, 1994; Corcuff et al., 1993, 1996; Corcuff and Leveque, 1993; New et al., 1991). In 1995 we developed a confocal scanning laser microscope for real-time imaging of human tissues (Rajadhyaksha et al., 1995). Cellular and nuclear microstructures in normal human skin and skin cancers, and dynamic events such as circulating blood flow, the response of skin to ultraviolet light, and wound healing were investigated (Rajadhyaksha and Zavislan, 1998).

## MOTIVATION

Dermatologists spend a large portion of their time diagnosing skin lesions or "moles" (see Figure 1). These lesions are of various shapes, sizes, and colors. Clinical screening is initially performed with either the naked eye or a low-power microscope (i.e., magnifying glass). Often the initial screening is not reliable because different cancers may look alike on the skin surface, and clinical pictures (such as Figure 1) do not reveal the subsurface tissue cellular and nuclear microstructures. Consequently, the accuracy of clinical screening of skin cancers is low; for example, the success rate for detecting melanomas (the most serious and potentially fatal skin cancer) is only 60 to 90 percent, depending on the dermatologist's expertise. Biopsies are almost always required for an accurate diagnosis. Of the approximately 3 million biopsies performed annually in the United States, 60 to 80 percent turn out to be noncancerous and therefore could have been avoided. Confocal reflectance microscopy offers dermatologists a noninvasive real-time imaging diagnostic tool that may potentially be useful for either screening lesions prior to biopsy or for diagnosis without biopsy.

## PRINCIPLES OF CONFOCAL MICROSCOPY

Confocal microscopy is one of those wonderful ideas that seem obvious but only after someone else (Minsky, 1957) has intuitively figured it out. A confocal

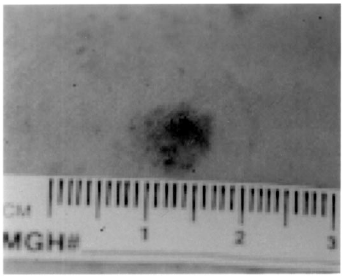


**FIGURE 1** Clinical appearance of a melanoma, which is the most serious and potentially fatal skin cancer. Clinical screening is based on this low-magnification, low-resolution photograph, which does not reveal subsurface tissue cellular and nuclear microstructures. By comparison, the microstructure of this melanoma can be noninvasively visualized with confocal microscopy (Figure 6b); thus, confocal imaging can potentially improve the accuracy of clinical screening.

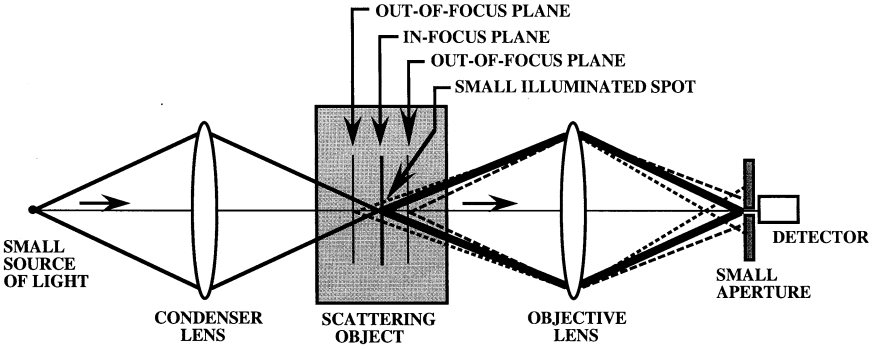


**FIGURE 2** A confocal microscope can noninvasively image a thin plane (section) that is in focus in a scattering object. This is called optical sectioning. The small aperture in front of the detector collects only the light that is in focus (solid lines) while spatially filtering light that is out of focus (dotted lines). Although the light is shown to penetrate the object, imaging of living tissue is based on the detection of backscattered light (Figure 3). SOURCE: Reprinted with permission from Mediscript Ltd. (Rajadhyaksha and Zavislan, 1998).

microscope (see Figure 2) consists of a "point" or small source of light that illuminates a "point" or small spot within an object, and the illuminated spot is then imaged onto a detector through a "point" or small aperture. The source, illuminated spot, and detector aperture are placed in optically <u>con</u>jugate <u>focal</u> planes, so we say they are "confocal" to each other. The detector aperture size is matched to the illuminated spot size through the intermediate optics. Because we illuminate a small spot and detect through a small aperture, we image only the plane that is in focus within the object. Light originating in planes that are out of focus is spatially filtered from entering the detector. A confocal microscope thus allows noninvasive imaging of a thin plane (section) within a scattering object with high axial (and also lateral) resolution; because we reject all the light that is not in focus, the image has high contrast. This is known as "optical sectioning."

The arrangement in Figure 2 shows that only a single spot may be illuminated and imaged at a time. Imaging a single point is often not useful in medicine. To view the whole object, one must then scan the illuminated spot over the desired field of view. We illuminate the object point by point in a two-dimensional raster and then create the image correspondingly point by point. Scanning may be done by either moving the specimen relative to a stationary illumination beam (object scanning) or by moving the beam relative to a stationary specimen (beam scanning). For imaging of human beings, beam scanning is obviously easier than object scanning. Although the configuration in Figure 2 shows the light transmitting through the object, imaging of living human tissue is based on

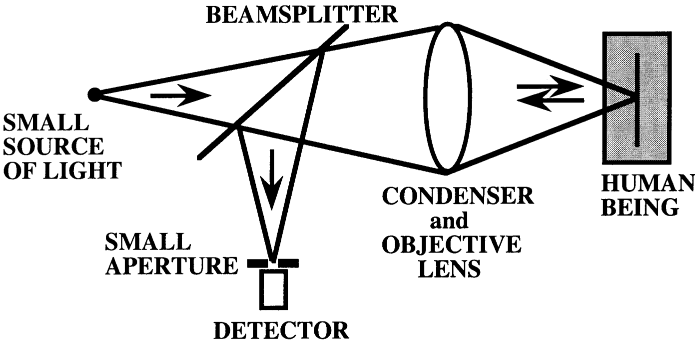


**FIGURE 3** Confocal imaging is based on the detection of backscattered light, with the illumination source and detector being on the same side of the human being. SOURCE: Reprinted with permission from Mediscript Ltd. (Rajadhyaksha and Zavislan, 1998).

the detection of backscattered light (see Figure 3) such that the illumination source and the detector are on the same side of the human being.

## DEVELOPMENT OF CONFOCAL SCANNING LASER MICROSCOPES

### Laboratory Prototype

At Wellman Laboratories, Massachusetts General Hospital (MGH), we built a video-rate confocal scanning laser microscope (CSLM) for noninvasive imaging of living human tissue (Rajadhyaksha et al., 1995). Figure 4 shows the optical design of our prototype CSLM. Any laser and wavelength may be used for illumination; we use near-infrared 800- to 1,064-nm wavelengths. Near-infrared wavelengths are preferred to visible wavelengths because of reduced scattering and absorption and hence deeper penetration into tissue (Anderson and Parrish, 1981). The collimated laser beam is scanned in the fast (X) direction with a rotating polygon mirror and in the slow (Y) direction with an oscillating galvanometric mirror. The scanning is at standard video rates of 15.734 kHz along X and 60 Hz along Y, so that the images can be displayed in real time on a television monitor. Two fields at 60 Hz are interlaced to produce a frame rate of 30 Hz, which is the standard for television in the United States. The X and Y scanning produces a raster of laser beam spots in the back focal plane of the objective lens, which, when demagnified by the objective lens, defines the field of view at the tissue. For 20× to 100× objective lenses that we usually use, the field of view is 800 to 160 μm. The raster illuminates an XY plane within the tissue through a standard microscope objective lens. The XY plane is a horizontal plane parallel to the surface of the tissue, and the Z axis (optical axis of the

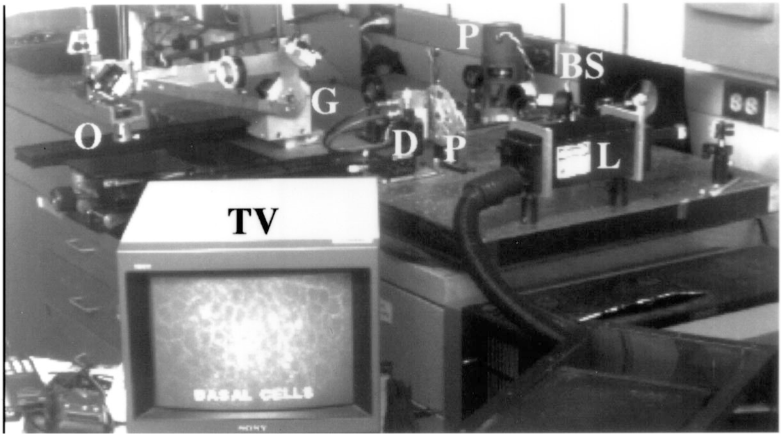


**FIGURE 4** Laboratory-prototype confocal scanning laser microscope for imaging living human skin. The illumination source is the collimated output beam from a near-infrared 800- to 1,064-nm laser (L). The collimated laser beam is scanned in the fast direction with a rotating polygon mirror (P) and in the slow direction with an oscillating galvanometric mirror (G). The scanning produces a two-dimensional raster of laser beam spots that illuminate the skin through an objective lens (O). Light that is backscattered from the tissue retraces its path through the objective lens and the two scanners. The returned light is descanned at the two scanners and then separates from the illumination path at the beamsplitter (BS). Beyond the beamsplitter, the backscattered light is detected through a pinhole (P) with a silicon avalanche photodiode (D). The detector output is sent to a television monitor (TV). The control and video-timing electronics is not shown.

CSLM) is perpendicular to it. The intermediate optics consist of folding mirrors and achromatic lenses. Light that is backscattered from the tissue retraces its path through the objective lens and the two scanners. The returned light is descanned at the two scanners and then separates from the illumination path at the beamsplitter. Beyond the beamsplitter, the backscattered light is detected through a pinhole by a silicon avalanche photodiode. The detector output is sent to a television monitor and storage devices such as a super-VHS videotape recorder and an 8-bits/pixel frame grabber. We built the control and video-timing electronics using well-known designs from the confocal scanning laser ophthalmoscope (Webb and Hughes, 1981; Webb et al., 1987).

Confocal imaging of living tissue is most useful if the resolution is similar to that of conventional microscopy (histology), so that cellular and nuclear microstructures can be seen. Histology involves viewing of typically 5-μm thin sections with lateral resolution of 1 μm. We optimized the CSLM design and operating parameters to achieve lateral resolution of 0.5 to 1 μm and axial resolution (section thickness) of 2 to 5 µm. Confocal microscopy provides the highest resolution yet of any noninvasive imaging modality, and this resolution compares very well to that of histology.

## Commercial Product

Our research at MGH-Wellman Laboratories with the prototype CSLM demonstrated that video-rate confocal imaging of cellular and nuclear microstructures with near-infrared light is possible noninvasively in living human skin (Rajadhyaksha et al., 1995; Rajadhyaksha and Zavislan, 1997, 1998). In 1997, Lucid, Inc. (Henrietta, N.Y.) entered into a five-year partnership with MGH-Wellman Labs to commercialize this technology for basic and clinical skin research. Engineers at Lucid, in collaboration with MGH scientists, reengineered the CSLM prototype into a turnkey user-friendly confocal imaging system called the VivaScope™, with flexible user-controlled operating parameters (see Figure 5). The VivaScope™ is much smaller than the prototype and portable, so we can move it easily between different labs and clinics. The optics and their mounts are robust, so alignment is not necessary, and setup time is 15 minutes when we move it to a new location. It is supported on a stand that can be raised or lowered relative to the skin site on the human subject to be imaged. We have built an arm and rotatable head that allow easy interfacing to different sites such as arms, legs, back, torso, face, and scalp. Stable imaging at different sites on the body is possible when using a specially designed skin-to-CSLM contact device.

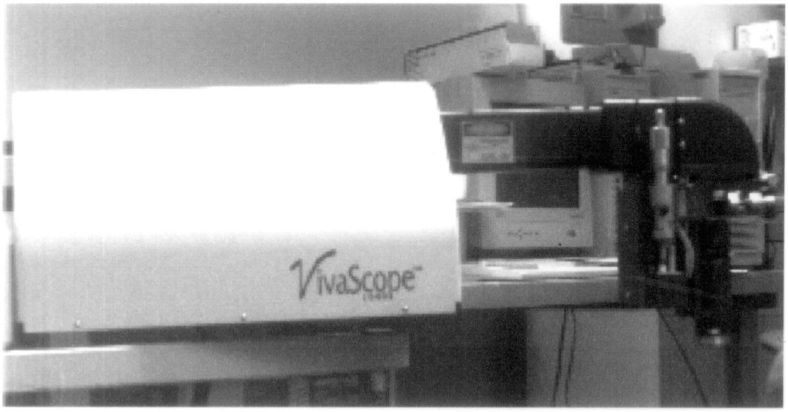


**FIGURE 5** Commercial confocal scanning laser microscope (VivaScope™) for imaging living human skin.

## IMAGING OF LIVING HUMAN SKIN

### Clinical Applications: Normal Skin Versus Skin Cancers

Confocal microscopy of normal human skin reveals cellular and nuclear microstructures in the stratum corneum and epidermis, and collagen and blood flow in the underlying dermis (Rajadhyaksha et al., 1995). We can image the superficial 100-μm thin epidermis and to a maximum depth of 350 μm in the dermis. Epidermal and dermal features in the confocal images were qualitatively and quantitatively analyzed; these correlated well to the corresponding histology. The epidermal features were cellular and nuclear shape and size, internucleus spacing, nuclear/cytoplasm ratio, cellular density, and spatial distribution of melanin (pigment that gives color to our skin). The shape, size, and density of circulating blood cells in the superficial dermal capillaries were measured. Morphological features included thickness of stratum corneum and epidermis and modulation depth of the epidermal-dermal junction. A group at L'Oreal has demonstrated similar imaging in normal human skin with their white light tandem scanning confocal microscope (Bertrand and Corcuff, 1994; Corcuff and Leveque, 1993; Corcuff et al., 1993, 1996; New et al., 1991). Two other groups have more recently reported images of skin (Masters et al., 1997) and skin structures such as nails (Kaufman et al., 1995).

Skin cancers appear different from normal skin (Figure 6). They have pathological differences such as atypical cells and nuclei that have abnormal shapes and sizes, lateral and vertical spreading of abnormal cells and formation of clusters of these cells, increase in melanin content and changes in its spatial distribution, formation of star-shaped projections (dendrites), loss of epidermal structure, elongation in the epidermal-dermal junction, and increased number of blood vessels and blood flow. In early clinical studies we have characterized melanomas and basal and squamous cell cancers. The control images were those of normal skin adjacent to the lesions. Preliminary analysis showed reasonably good qualitative correlation between the confocal images and the histology. Clinical research is currently progressing in the characterization of skin cancers as well as benign (noncancerous) lesions and other types of disorders, including detection of margins between lesions and surrounding normal tissue.

### Basic Research Applications

Confocal microscopy is an excellent imaging modality for basic skin (and other tissue) research. We can study living skin in its native state, without the artifacts of biopsy and histological processing. Because it is noninvasive, dynamic changes can be imaged over hours, days, weeks, and months. For example, we have investigated the response of the skin to ultraviolet-B (UVB) irradiation from sunlight, the process of wound healing, inflammatory responses to allergic or irritant agents, and delivery of drugs through blood vessels. Other

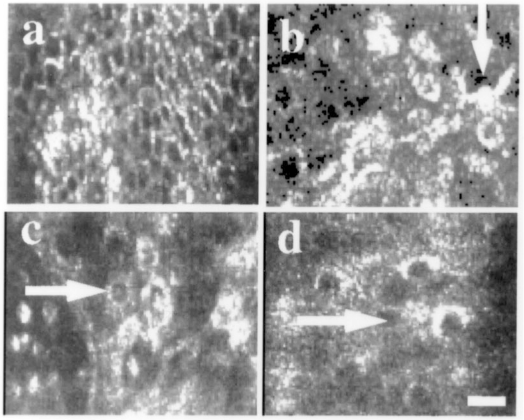


**FIGURE 6** Pathological differences in cellular and nuclear microstructures between normal human skin (a) and skin cancers (b-d) can be noninvasively visualized with confocal microscopy. The skin cancers shown are melanoma (b), basal cell carcinoma (c), and squamous cell carcinoma (d). Normal skin epidermis contains a regular chicken wire-mesh network of circular, oval, or polygonal cells. By comparison, the skin cancers show distinct differences: (b) melanoma cells develop star-shaped projections called dendrites (arrow), (c) basal cell carcinomas show increases in size and migration of basal cells from the superficial epidermis into the underlying dermis (arrow), (d) squamous cell carcinomas show changes in the shape and size of squamous cells in the superficial epidermis (arrow). Scale bar = 25 μm.

exciting applications may include cell-to-cell interactions, microcirculation, drug delivery through the skin, photoaging, and artificial tissue.

## FUTURE

Confocal imaging of living tissue is a new imaging modality. Our research effort will increase understanding of the morphology of normal versus abnormal skin. Other tissue types to which we have applied confocal microscopy are oral (lip and tongue) mucosa and bladder. When we compare confocal images to histology slides (i.e., the gold standard), two limitations are obvious: (1) confocal microscopy can probe only the upper 0.5 mm of tissue over fields of view limited to less than 1 mm, whereas histology probes down to depths of 2 to 3 mm over large fields of view, typically 5 mm; and (2) the use of dyes to stain specific cell types enhances tissue contrast, so that critical information necessary for diagnosis can be easily read in histology, whereas confocal microscopy relies on the natural (low) reflectance contrast of tissue without the advantages of stains. At present we are not certain whether diagnosis of skin cancers would be possible with confocal reflectance microscopy. However, results from our clinical studies strongly suggest that confocal microscopy may potentially be useful for screening skin cancers versus benign (noncancerous) lesions and normal skin, including detection of margins between the lesions and surrounding normal skin. Confocal images may thus provide a useful adjunct to clinical screening and histology by helping a dermatologist make decisions such as whether, where, or when to biopsy a lesion.

Further development of confocal scanning microscopes to make them highly effective for basic research and clinical applications includes several scientific and instrumentation challenges. The main scientific challenge is to understand light-tissue interaction and mechanisms of contrast and to determine techniques to enhance tissue contrast by staining cells types with reflectance microparticles. The instrumentation challenges are to make confocal imaging as similar to histology as possible through vertical sectioning, increasing the depth of imaging, and enlarging the field of view. As with any new imaging modality, we must learn to interpret, analyze, and extract meaningful information from the confocal images. This will involve an extensive correlation of confocal images to conventional (gold standard) histology.

The future will see the development of inexpensive handheld confocal microscopes that will become commonplace in the clinic and linked to each other in a telemedicine network. Although this is being initially applied to easily accessible tissues (skin, oral), the combination of confocal microscopy with other techniques such as laparoscopy should enable imaging of internal organs. Ultimately, optical imaging must be combined with nonoptical imaging modalities to create a noninvasive diagnostic tool kit for the medical profession.

## ACKNOWLEDGMENTS

Instrumentation development was funded in part by the U. S. Department of Energy and a Whitaker Foundation grant to M. Rajadhyaksha when he was at MGH-Wellman Labs. Clinical and basic research in skin, oral mucosa, and bladder was funded by Lucid, Inc. Credit for research goes to Richard Langley and Melanie Grossman (pigmented skin lesions), Steve Ugent (UVB effects), Salvador Gonzalez (tumor margins, proliferative and inflammatory disorders), Frank Koenig (bladder), and W. Matthew White (oral tissue) at MGH-Wellman Labs.

## REFERENCES

Anderson, R. R., and J. A. Parrish. 1981. The optics of human skin. Journal of Investigative Dermatology 77(1):13–19.

Bertrand, C., and P. Corcuff. 1994. In vivo spatio-temporal visualization of the human skin by real-time confocal microscopy. Scanning 16(3):150–154.

Corcuff, P., and. J. L. Leveque. 1993. In vivo vision of the human skin with the tandem scanning microscope. Dermatology 186(1):50–54.

Corcuff, P., C. Bertrand, and J. L. Leveque. 1993. Morphometry of human epidermis in vivo by real-time confocal microscopy. Archives of Dermatology Research 285(8):475–481.

Corcuff, P., G. Gonnord, G. E. Pierard, and J. L. Leveque. 1996. In vivo confocal microscopy of human skin: A new design for cosmetology and dermatology. Scanning 18(5):351–355.

Kaufman, S. C., R. W. Beuerman, and D. L. Greer. 1995. Confocal microscopy: A new tool for the study of the nail unit. Journal of the American Academy of Dermatology 32(4):668–670.

Masters, B. R., G. Gonnord, and P. Corcuff. 1997. Three-dimensional microscopic biopsy of in vivo human skin: A new technique based on a flexible confocal microscope. Journal of Microscopy 185(Part 3):329–338.

Minsky, M. November 7, 1957. Microscopy Apparatus. U.S. Patent No. 3013467.

New, K. C., W. M. Petroll, A. Boyde, L. Martin, P. Corcuff, J. L. Leveque, M. A. Lemp, H. D. Cavanagh, and J. V. Jester. 1991. In vivo imaging of human teeth and skin using real-time confocal microscopy. Scanning 13(5):369–372.

Pawley, J. B., ed. 1995. Handbook of Biological Confocal Microscopy, 2nd ed. New York: Plenum Press.

Rajadhyaksha, M., and J. M. Zavislan. 1997. Confocal laser microscope images tissue in vivo. Laser Focus World 33(2):119–127.

Rajadhyaksha, M., and J. M. Zavislan. 1998. Confocal reflectance microscopy of unstained tissue in vivo. Retinoids & Lipid-Soluble Vitamins in Clinical Practice 14(1):26–30.

Rajadhyaksha, M., M. Grossman, D. Esterowitz, R. H. Webb, and R. R. Anderson. 1995. In vivo confocal scanning laser microscopy of human skin: Melanin provides strong contrast. Journal of Investigative Dermatology 104(6):946–952.

Webb, R. H. 1996. Confocal optical microscopy. Reports on Progress in Physics 59(3):427–471.

Webb, R. H., and G. W. Hughes. 1981. Scanning laser ophthalmoscope. IEEE Transactions on Biomedical Engineering 28(7):488–492.

Webb, R. H., G. W. Hughes, and F. C. Delori. 1987. Confocal scanning laser ophthalmoscope. Applied Optics 26(8):1492–1499.

# ADVANCED MATERIALS

# Trends in Computational Materials Science for Materials Design and Processing

SHARON C. GLOTZER
*National Institute of Standards and Technology*
*Gaithersburg, Maryland*

Throughout the history of civilization, mankind has understood that the properties and usefulness of any material depend on its composition and processing. Early artisans and builders discovered secrets to transforming substances into functional materials often by happenstance, and modern-day experimentalists established trends through systematic study and categorization. Until recently, most material discoveries have resulted from exhaustive searches by trial and error.

In the 1960s, pioneering work in metallurgy revealed the possibility that general relationships linking macroscopic properties to atomic and molecular processes could be obtained. Comparing knowledge of metals, ceramics, polymers, and composites of these, researchers began to believe that the material behavior peculiar to each of these three major classes of materials is not so peculiar and could be subsumed under a general theory of materials built on first principles. This realization led to the hybrid discipline of materials science.

The ultimate mission of materials science and engineering is fast and accurate design of new materials and prediction of material properties from a fundamental knowledge of a material's constituents. Researchers in this field seek to understand the general relationships between structure and properties of materials with the ultimate goal of using these laws to tailor processes and properties to produce materials with specific behavior and functionality.

Computational materials science plays a pivotal role in this mission by allowing the investigation of structure, dynamics, and properties in ways not accessible to experiment. There are numerous computational techniques in use today, and it is useful to group them according to the length scale of the processes being studied. At the atomistic level, quantum molecular dynamics and quantum Monte

Carlo methods allow, for example, calculation of electronic band structures and prediction of chemical bond strength, which govern the electronic and mechanical properties of a material. At the next level, by devising formalisms that describe the effect of electrons implicitly, one can use classical molecular dynamics and classical Monte Carlo methods with effective pair and triplet potentials to simulate larger collections of atoms or molecules, on the scale of up to 10 nanometers, for times up to 100 nanoseconds on fast computers. Physical understanding of atomistic and molecular processes at this level can be used to construct atomistic-based continuum equations to study processes at the mesoscopic level of a few microns using finite difference and finite element solution methods. Finally, all of this knowledge can be coarse grained into constitutive laws and other continuum equations that, combined with finite element simulation techniques, can be used to model the bulk structural, electronic, thermal, and relaxational properties of materials. Typically, individual researchers are expert in one or two methods particular to either micro-, meso-, or macroscopic-length scales. The hierarchical coupling of length scales and methods is often called multiscale modeling of materials and is one of the most important trends in computational materials science today.

A major challenge to materials science is the prediction and control of microstructure. Historically, this activity was central to the field of metallurgy, e.g., in understanding and controlling the growth of dendritic structures during alloy solidification. Now it is also a major theme in ceramics and polymers. Often at the heart of microstructure formation is a thermodynamic instability. One example of such an instability occurs when a material is cooled below its freezing point and crystallization occurs. In a material of two different polymers (a polymer blend), a different thermodynamic instability can arise either at low or high temperature when the entropic gain from mixing no longer compensates for the repulsion between unlike species. Because polymers are long-chain molecules, the entropy of mixing is typically extremely small and thus blends tend toward “phase separation” either by nucleation or spinodal decomposition into two phases, each rich in one of the components (Glotzer, 1995).

When a polymer-blend phase separates by spinodal decomposition, bicontinuous interconnected patterns emerge and coarsen as the system phase separates (see Figure 1). However, the growth of these patterns is commonly found to be sensitive to perturbations that break the isotropy of the coarsening process. This suggests the possibility of producing diverse morphologies through the adjustment of molecular architecture and the type of perturbation. Indeed, artificially structured materials present endless opportunities to create new applications, and computational materials science is playing a key role in guiding experiments. In problems of this type, simulations have focused on investigating mesoscale phenomena, from a few manometers to tens or hundreds of microns, using methods that average details of the molecular architecture and specific interactions. Such methods are used because current computational and algo-

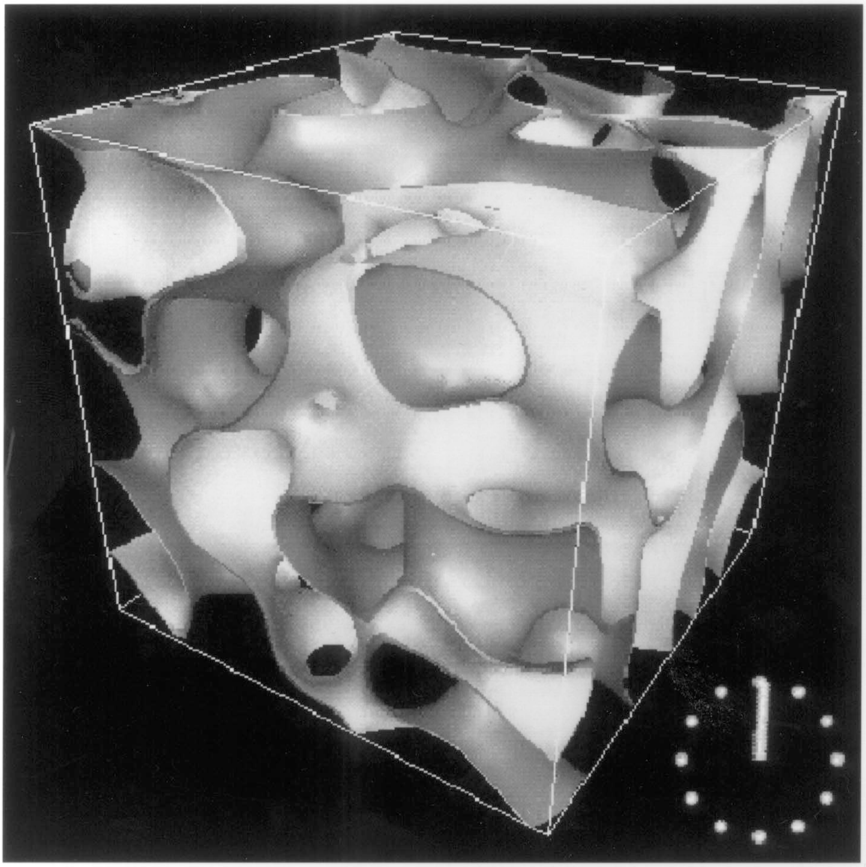


**FIGURE 1** Snapshot of computer simulation of a polymer blend undergoing phase separation by spinodal decomposition. Shown are isosurfaces of constant composition at the boundary between the two phases. The Cahn-Hilliard-Cook simulation technique (Glotzer, 1995) used here focuses on mesoscopic-length scales, where individual atomic details have been averaged into an "effective" polymer interaction. SOURCE: Reprinted from Computational Materials Science, Vol. 6, 1996, with permission from Elsevier Science.

rithmic limitations make the simulation of long-time processes such as phase separation in bulk polymeric materials with atomistic methods extremely time consuming. Three examples of using simulation to elucidate trends and guide experiments in pattern formation in polymer blends are as follows:

- *Phase separation of ultrathin polymer-blend films on patterned substrates.* Recently, researchers have studied the influence of modulated surface interactions on an ultrathin polymer-blend film spun cast onto a substrate (Böltau et al., 1998; Karim et al., 1998). Important insights into pattern-directed spinodal decomposition gained from computer simulations helped in the design of the experiments. The simulations were performed (Karim et al., 1998) by solving the Cahn-Hilliard-Cook coarse-grained model of phase separation (Glotzer, 1995) with a modulated boundary interaction on the substrate surface. The boundary interaction was designed such that the magnitude of the chemical potential favoring a particular component alternated on the substrate surface with some chosen periodicity. Cross sections of critical composition films showed that the surface interaction induced composition waves, both in the plane and transverse to the substrate, which led to a pattern-directed lateral-phase separation (stripes) in thin films and to a "checkerboard" morphology in thicker films (see Figure 2).

These results were verified experimentally using ultrathin deuterated polystyrene (dPS) and polybutadiene (PB)-blend films spun cast onto self-assembled monolayer (SAM) substrates. The local surface interactions of these SAM lay-

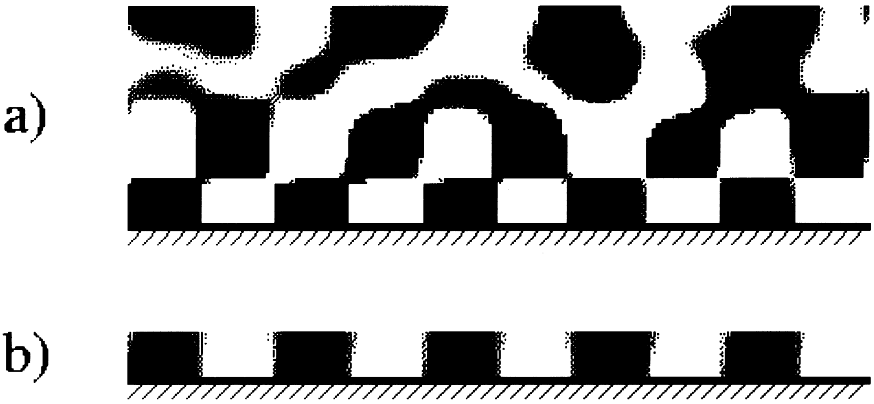


**FIGURE 2** Cross sections showing simulation of composition variations in phase-separating-blend films on a patterned substrate. The simulations used the Cahn-Hilliard method with a modulated boundary interaction. The surface pattern wavelength is about twice the maximally unstable phase separation scale ("spinodal wavelength") in these simulations. Variations in film height because of unequal surface and interfacial tensions, deformability of the polymer-air boundary, and hydrodynamic interactions were not treated in this simulation. Thermal noise was included. (a) "Checkerboard" pattern in thicker film cross sections arising from surface-directed phase separation in combination with pattern-directed lateral phase separation. (b) Thinner films exhibit only pattern-directed lateral phase separation. SOURCE: Karim et al., 1998.

ers were varied through microcontact printing of hydrophobic and hydrophilic end-group alkanethiols to provide a modulation of the surface interaction as in the simulations. Control of the local boundary interaction with these patterned surfaces allowed the selection of symmetry and scale of the resulting phase separation morphology through templates that direct the phase separation. The experiments confirmed that the spinodal decomposition process can be manipulated by using surface patterning of the solid substrate supporting the blends. The blend components tracked the surface pattern and produced stable and well-aligned stripes for a particular range of film thickness (see Figure 3). The strategy of creating surface structure by nanofabricating surface template patterns should find increasing application in engineering, especially as fabrication of "master" surface patterns becomes more routine.

- *Target patterns in filled polymer blends.* Filled polymers represent a substantial share of the world market in plastics, competing strongly with costly high-performance plastics because of the promise of improving material properties with cheap additives. A major limitation to the potential growth and development of this industry has been the lack of understanding of interactions between fillers and the polymer matrix. This limitation has forced the filled polymers industry to adopt empirical approaches with inevitably long product development cycles. Computer simulations have recently been used to study the effect of impurities and fillers on microstructure development in an immiscible blend (Lee et al., 1998). Generally, a filler particle will prefer one of the blend components, leading to preferential wetting of that component at the surface of the filler. This "encapsulation" of the filler particle by one of the components induces a depletion layer adjacent to the wetting layer. If the blend is thermodynamically unstable, the simulations showed that the pattern will continue in the form of spherical spinodal waves whose width and extent depend on the degree of instability (see Figure 4). These patterns have been observed experimentally using atomic force microscopy and phase contrast microscopy (see Figure 5).
- *Pattern formation in liquid crystal display materials.* Materials used in liquid crystal displays, like those in laptop computers, are typically composed of droplets of liquid crystal dispersed in a polymer matrix (so-called polymer-dispersed liquid crystals). The microstructure of these materials, e.g., droplet size and shape, is formed during phase separation of the liquid crystal and polymer (see Figure 6a) and affects the properties and performance of the final product (Figure 6b). Computer simulations (Figure 6c) using methods similar to those described above have been used to investigate microstructure formation in these materials and show that the resulting patterns depend crucially on composition, relative molecule size, and degree of immiscibility (Langer and Glotzer, 1997; Lapeña et al., 1998).

In summary, computational materials science is playing a key role in materials development and processing and in the design of new materials and material

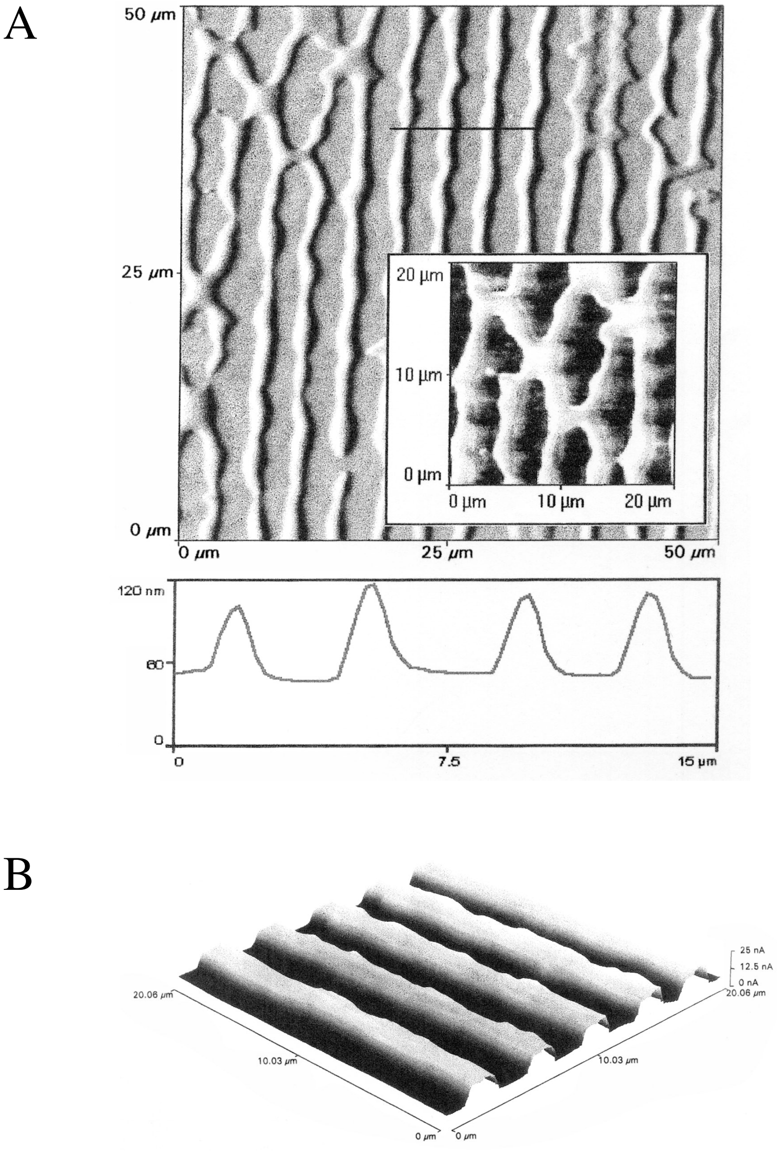
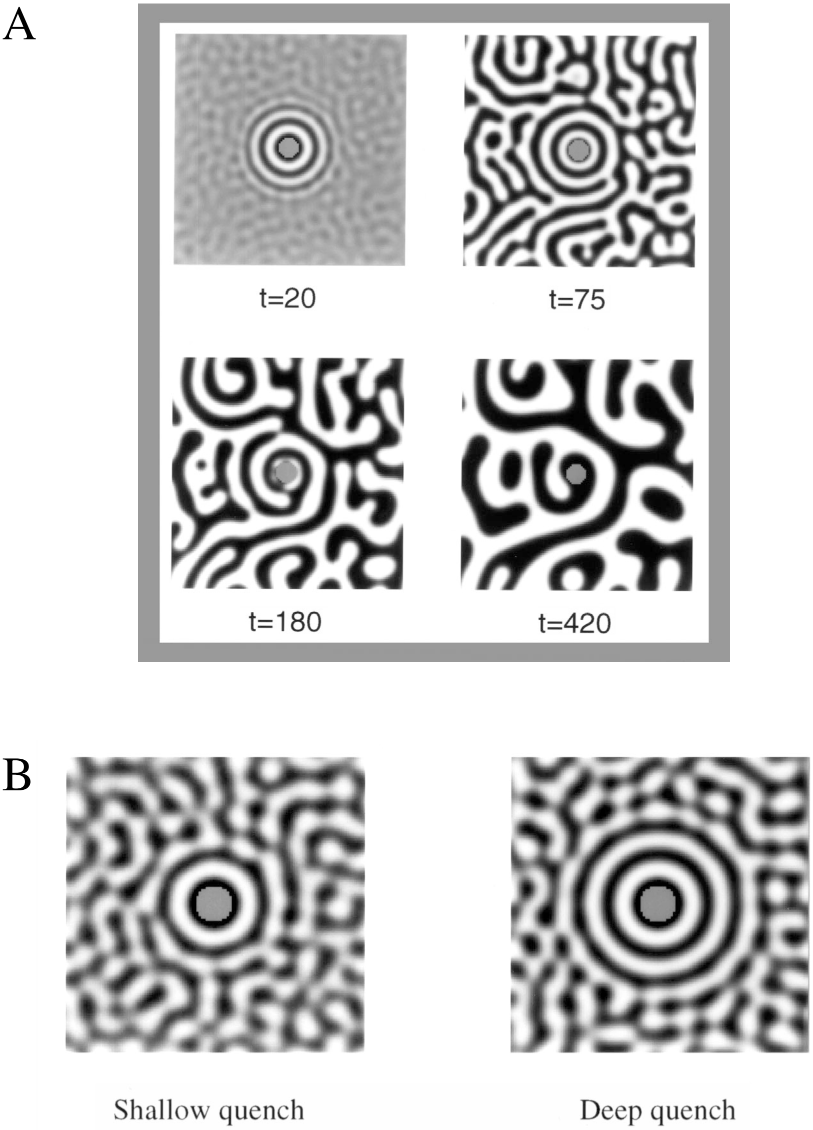


**FIGURE 3** (a) Intermediate stage of dPS/PB-blend film phase separating on a patterned substrate. AFM image of topography with accompanying line profile. Solid line denotes line profile position. Inset shows lateral force image (lighter PB-rich regions have higher friction) corresponding to the topography at the upper left-hand corner. (b) Lateral force AFM image in profile view late stage of same blend. Elevated regions correspond to PB-rich domains. The pattern remains unchanged after 24 hours. SOURCE: Karim et al., 1998.



Shallow quench

Deep quench

**FIGURE 4** (a) Computer simulation using the Cahn-Hilliard method of a polymer-blend phase separating in the presence of a filler particle, at various times *t* during a temperature quench. Composition waves radiate outward from the filler particle and join with the "normal" spinodal decomposition pattern in the bulk. Eventually, only a thin wetting layer remains. (b) The strength of the "target" pattern depends strongly on the degree of immiscibility of the two polymers. Left: shallow quench. Right: deep quench. SOURCE: Lee et al., 1998.

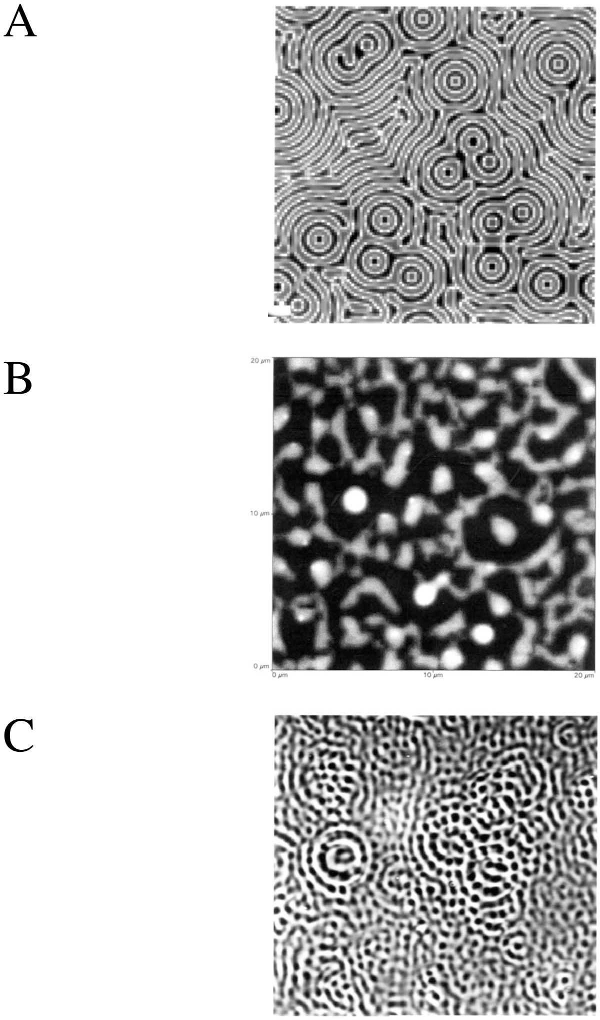


**FIGURE 5** (a) Computer simulation using the Cahn-Hilliard method of a polymer-blend phase separating in the presence of many filler particles. Composition waves interfere to produce a transient spiral pattern. SOURCE: Lee et al., 1998. (b) AFM image of phase separation in a thin polymer-blend film containing many filler particles. "Target" patterns form around filler (bright spots), reflecting the preferential wetting of one of the blend components. The width of the image is 20 microns. SOURCE: Image courtesy of A. Karim, National Institute of Standards and Technology. (c) Phase contrast image of target patterns observed in a PS/PVME (50/50)-blend phase separating during photo-crosslinking. Here the crosslink junctions are believed to act as filler particles. The scale of the image is 10 microns. SOURCE: Reprinted with permission from Marcel Dekker (Tran-Cong, 1998).

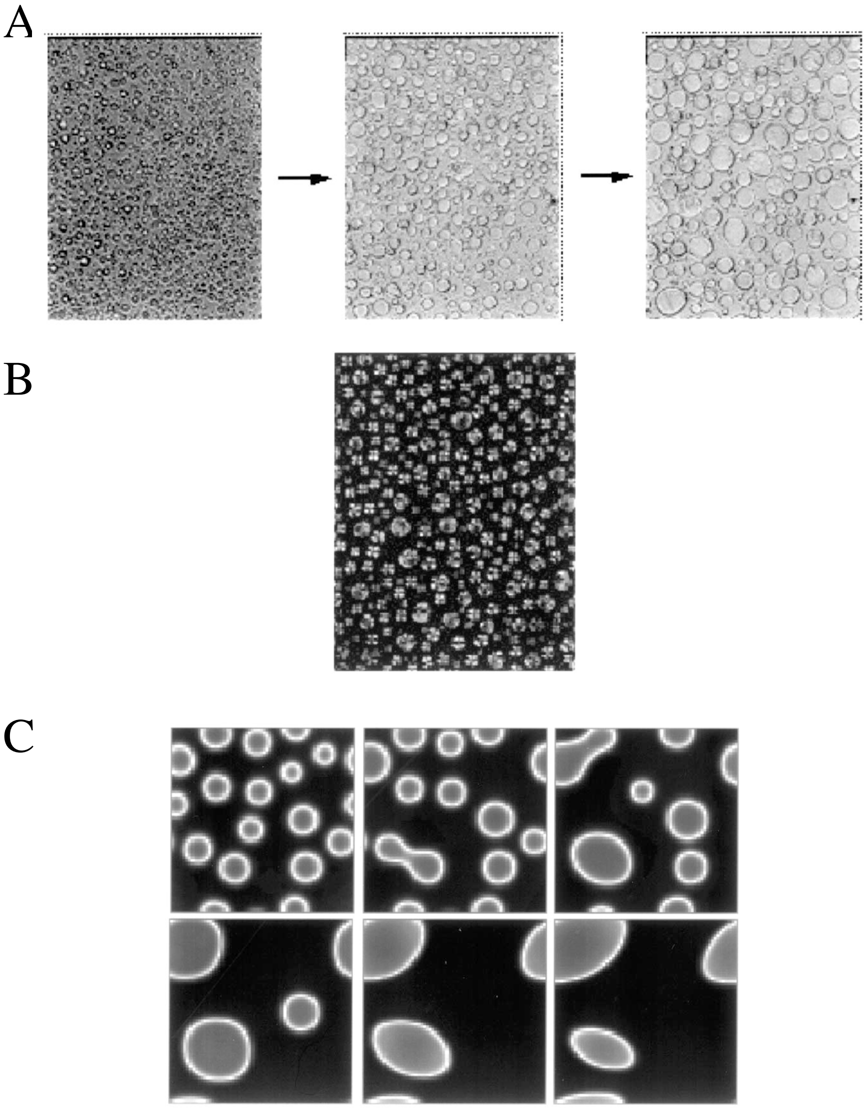


**FIGURE 6** (a) Polymer-dispersed liquid crystal undergoing phase separation. (b) Final morphology, shown between crossed polarizers. SOURCE: Both images courtesy of http://abalone.cwru.edu/tutorial/. (c) Computer simulation of a polymer-dispersed liquid crystal undergoing phase separation following a thermal quench from a mixed, isotropic phase to a de-mixed, nematic liquid crystal-rich phase coexisting with an isotropic polymer-rich phase. As spinodal decomposition proceeds (left to right, top to bottom), alignment of the liquid crystal causes elongation of the liquid crystal droplets in the polymer matrix. SOURCE: Glotzer, 1997.

applications. In the field of polymers, for example, simulations provide insight and predictions that allow experiments to control patterns at mesoscopic-length scales. As computer power continues to increase and algorithms continue to improve, the ability of simulation to model materials at all relevant length scales will grow and take us closer to first principles prediction and design of real materials.

## REFERENCES

Böltau, M., S. Walheim, J. Mlynek, G. Krausch, and U. Steiner. 1998. Surface-induced structure formation of polymer blends on patterned substrates. Nature 391(6670):877–878.

Glotzer, S. C. 1995. Computer simulations of spinodal decomposition in polymer blends. Pp. 1–46 in Annual Reviews of Computational Physics, Vol. 2, D. Stauffer, ed. River Edge, N. J.: World Scientific.

Glotzer, S. C. 1997. Unpublished research.

Karim, A., J. F. Douglas, B. P. Lee, S. C. Glotzer, J. A. Rogers, R. J. Jackman, E. J. Amis, and G. M. Whitesides. 1998. Phase separation of ultrathin polymer-blend films on patterned substrates. Physical Review E 57(6):R6273–R6276.

Langer, S. A., and S. C. Glotzer. 1997. Morphogenesis in nematic liquid crystal/polymer materials. Physica A 239:358–362.

Lapeña, A., S. C. Glotzer, S. A. Langer, and A. J. Liu. 1998. Effect of ordering on spinodal decomposition of liquid-crystal/polymer mixtures. Submitted to Physical Review Letters. [Online]. Available: http://xxx.lanl.gov/abs/cond-mat/9810166.

Lee, B. P., J. F. Douglas, and S. C. Glotzer. 1998. Filler-Induced Composition Wares in Phase-Separation Polymer Blends. Manuscript in preparation.

Tran-Cong, Q. 1998. Structure and Properties of Multiphase Polymeric Materials, T. Araki, Q. Tran-Cong, and M. Shibayama, eds. New York: Marcel Dekker.

# Design, Synthesis, Development, and Integration into Manufacturing of New Polymer Materials Architectures for Advanced Integrated Circuit Fabrication

OMKARAM NALAMASU
*Bell Laboratories, Lucent Technologies*
*Murray Hill, New Jersey*

Invention of the point contact transistor in 1947 heralded the dawn of the microelectronics era, which has had impacts on every aspect of our lives. Much of the tremendous progress in integrated circuit (IC) technology, performance, and functionality in the past 30 years has been fueled by advances in lithographic technology. The ability to pack an ever-increasing number of individual circuit elements into a device has enabled faster devices, higher densities, and lower-power dissipation in complementary metal-oxide-silicon (CMOS) circuits. Device complexity and functionality have increased, while minimum feature size has dramatically decreased (see Figure 1), resulting in unprecedented productivity gains in the integrated circuit industry (Moore, 1995). This very steep performance curve historically improved cost per function of integrated circuits by 30 percent per year over this period. Roughly half of the productivity gains are a direct result of fundamental and incremental advances in lithography technology.

Materials chemistry in general and organic and polymer chemistry in particular have enabled the advancements in microelectronics technology. Since the ability to shrink feature size is critically dependent on the technologies involved in delineation of the circuit pattern, advances in high-resolution imaging materials could be considered one of the key enablers that have contributed to the unprecedented growth in microelectronics technologies (Thompson et al., 1994).

The lithographic technology for printing integrated circuits consists of three main elements: the exposure tool, the mask technology, and the resist technology. In IC lithography an image of the mask (usually reduced four to five times) is projected onto a wafer substrate that has been coated with a photosensitive material (resist). The solubility of the resist is changed by exposure to light, so that a copy of the mask emerges upon development, much like a photograph.

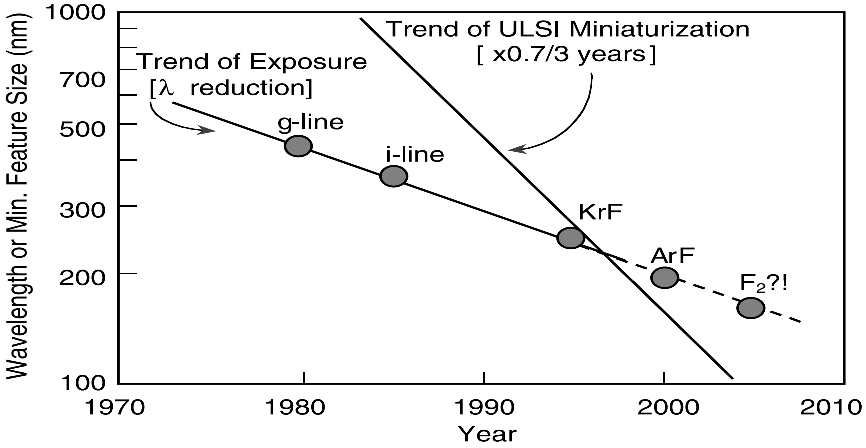


**FIGURE 1** Comparison of lithography wavelength trends with integrated circuit feature size trend. SOURCE: Bell Laboratories/Lucent Technologies. Courtesy: S. Okazaki, Hitachi.

The remaining resist pattern is then used for subsequent process steps such as etching or implantation doping. Thus, any lithography technology must have fully developed exposure tool, mask, and resist technologies for it to be successful.

An overwhelming preponderance of devices continue to be fabricated via "conventional photolithography" employing 350- to 450-nm light, referred to as G- (435-nm) and I-line (365-nm) lithographies that correspond to the emission bands of a mercury lamp. Incremental improvements in tool design and performance with concomitant refinements in the novolac/diazonaphthoquinone resist materials chemistry and processing have enabled this platform to meet the printing needs of 10-μm geometries at its insertion to today's 0.35-μm minimum geometry circuits (Dammel, 1993). The cost of introducing a new technology, which includes the costs associated with development and implementation of new hardware and resist materials, is a strong driving force to push photolithography to its absolute resolution limit and extend its commercial viability.

As device feature sizes approached 0.25 μm, the industry moved toward using 248-nm excimer laser ultraviolet (UV) light as the exposing wavelength for advanced lithographic applications. The conventional photoresists were not appropriate for use with this deep-UV (DUV) light source because of their opacity and low throughput owing to insufficient photospeeds because of low quantum yields and low-energy density at the wafer plane of the deep-UV exposure tools. The materials community met the challenge by delivering the first revolutionary change in resist materials chemistry by adopting chemically amplified resist schemes, in which one photoproduct catalyzes several hundred chemical events, leading to very sensitive high-resolution resist systems (Reichmanis et al., 1991).

The initial pioneering studies by Ito, Willson, and Frechet dealt with the catalytic deprotection of poly([tert-butoxycarbonyloxy]styrene) (TBS) in which the thermally stable acid-labile tert-butoxycarbonyl group is used to mask the hydroxyl functionality of poly(vinylphenol). Irradiation of TBS films containing small amounts of an onium salt, such as diphenyliodonium hexafluoroantimonate with UV light, liberates an acid species, which upon subsequent baking catalyzes cleavage of the protecting group to generate poly(p-hydroxystyrene) with catalytic chain lengths higher than 1,000. Thus, exposure and postexposure bake (PEB) result in a large polarity change in the exposed areas of the film. Whereas the substituted phenol polymer is a nonpolar material soluble in nonpolar lipophilic solvents, poly(vinylphenol) is soluble in aqueous base. This change in polarity allows formation of either positive or negative tone images, depending on the developer employed.

Since inception of the chemically amplified resist concept in 1980, numerous research groups have expanded on this revolutionary concept. The first commercial 248-nm resist was introduced in 1990 after almost 10 years of research and was based on the copolymer derived from tert-butoxycarbonyloxystyrene and sulfur dioxide used in conjunction with a nitrobenzyl tosylate photoacid generator (PAG) material. Several resist materials using a variety of thermally stable and acid-labile protective groups and base-soluble parent polymers have been marketed since then.

The problems with chemically amplified resist systems did not really become apparent until they were in full-scale manufacturing. The very high catalytic chain lengths in chemically amplified resists that result in high sensitivity (low photospeed) also turn out to be its biggest problem, as the deactivation of photoacid by airborne adventitious amines and basic functionalities on semiconductor substrate surfaces have resulted in “T-tops” and “scum,” respectively, reducing the process margin and reproducibility. Strong collaborative efforts between process engineering and resist chemistry communities have led to both a fundamental understanding of the process issues and new generations of resist materials that alleviated or eliminated the issues by innovative resist materials design. The invention and innovation of materials that use partially protected poly(vinyl)phenol or its copolymers in combination with PAGs that generate weaker photoacids and addition of bases to resist formulation have allowed the use of chemically amplified resists for optical projection printing at and below the conventional Rayleigh diffraction limit, contributing to the extension of optical lithography. The chemically amplified principles have also been applied to all other resist systems for photospeed, resolution, and process latitude enhancements.

The opacity of traditional UV and deep-UV organic matrix resins and photoresist components at 193 nm introduces yet another paradigm shift in the design of resist materials. This, in turn, offers opportunities for innovation and invention in the design of novel chemistries and process schemes. The constraints to this reengineering are, of course, to retain aqueous base solubility, etch resistance, resolution, photospeed, and a process latitude similar to that of phenolic-based systems but built into aliphatic polymer and dissolution inhibitor components. Moreover, regulatory constraints on the volatile organic emissions also provide opportunities to design revolutionary resist schemes.

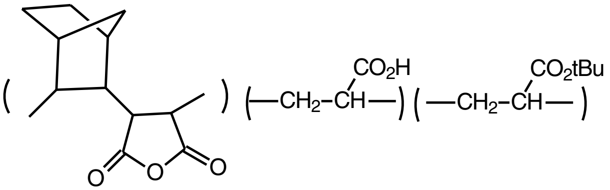
To summarize, the challenge to a resist chemist is to design a 193-nm resist system that is largely based on aliphatic polymers and dissolution inhibitors but functionally identical to DUV and I- or G-line resists with a process that is transparent to the process engineer. Base solubility in a majority of the matrix polymers used for lithography is due to the presence of either phenolic hydroxides or carboxylic (aliphatic or aromatic) acid moieties. Much of the effort in designing 193-nm resists has focused on derivatized acrylate and methacrylate copolymers. While resists based on these materials showed some promise, etch stability and incompatibility with conventional aqueous-base developers are issues that are still unresolved.

Bell Labs has pioneered cyclo-olefin-maleic anhydride alternating copolymers with acrylates as an attractive alternative to methacrylate-based matrix resins and other alternate “all alicyclic backbone” polymers (Houlihan et al., 1997). These polymers are prepared easily by standard radical polymerization techniques, and a large pool of cycloolefin feedstocks enabled us to engineer the desired imaging and etch properties. As a rule, cycloolefins with maleic anhydride yield alternating copolymers with high-Tg (glass transition temperature). Acrylic or methacrylic acids or protected analogs thereof could be incorporated in small percentages without disruption of the essentially alternating nature of the copolymer (see Figure 2).

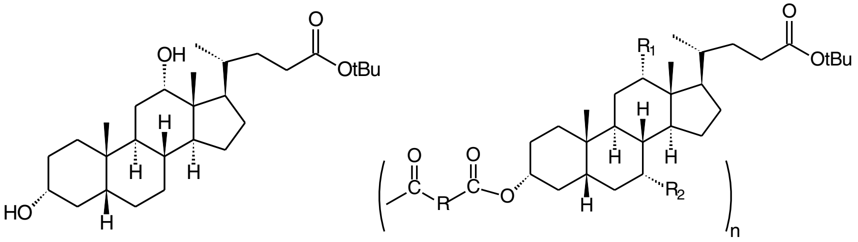
Aqueous-base solubility and imaging functionalities could be introduced into the polymer by either the addition of acrylic acid or acrylates into the polymer backbone or the use of appropriately substituted cycloolefins or maleic anhydride monomers. High structural carbon in the polymer allows for high etch stability and the presence of acrylate acidic moieties in the polymer makes the system compatible with aqueous-base developers.

Additionally, the use of multifunctional monomeric as well as oligomeric t-butyl esters of cholic, deoxycholic, and lithocholic acids as dissolution inhibitors (DI) dramatically enhances the dissolution selectivity by inhibiting dissolution in the unexposed regions and enhancing that in the exposed regions. The multifunctional nature of the DI allows for strong interaction between the polymer and dissolution inhibitor and imparts nonlinear dissolution behavior to the resist, contributing to high contrast and resolution.

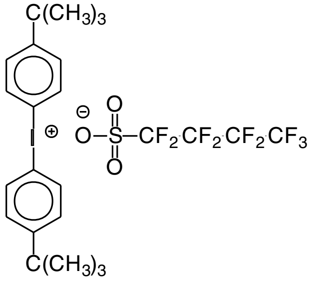
While the aromatic PAGs are highly absorptive at 193 nm, they are only needed in small quantities (typically <5 wt%) in resist formulations. Consequently, 193-nm resists could be designed using the same PAGs as those used



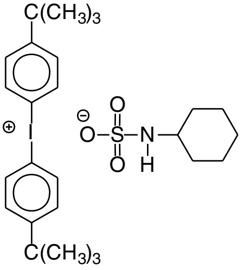
P(NB-alt-MA-co-TBA-co-AA)



Dissolution Inhibitors (DI)



bis(4-t-butylphenyl iodonium)
nonaflate
PAG



bis(4-t-butylphenyl)iodonium
cyclamate
PDB

**FIGURE 2** 193-nm resist components. Alternating copolymers of norbornene with maleic anhydride with acrylic acid and t-butyl acrylate “impurities,” monomeric and oligomeric dissolution inhibitors, photoacid generators (PAG), and photodefinable bases (PDB). SOURCE: Bell Laboratories/Lucent Technologies.

with 248 nm. In addition to absorption, the considerations in designing a PAG are solubility, volatility of both the PAG and its photoproducts, acid strength, cost, and toxicity. Depending on the choice of protecting group, PAGs that generate strong or weak acids can be employed. One typical example of a PAG is shown in Figure 2.

Base additives generally have been added to chemically amplified resist formulations to "buffer" the resist from the fluctuations in airborne adventitious amine (e.g., ammonia, N-methyl pyrrolidone) concentrations in clean room environments. We have introduced a new class of photodecomposable aminosulfonate moieties capable of photogenerating free aminosulfonic acids. Since these materials are bases that generate acid upon exposure to light, they are called photodefinable bases.

Optimized 193-nm resist formulations show resolution down to 0.12-μm features with binary masks and 60-nm resolution with phase-shifting mask technology. The resists also demonstrate excellent process margin and exhibit etch stability better than that of deep-UV resists. The materials will be inserted for production of 0.15- to 0.13-μm devices in the year 2000 to 2002 time frame.

## REFERENCES

Dammel, R. 1993. Diazonaphthoquinone-Based Photoresists. Bellingham, Wash.: SPIE Press.

Houlihan, F. M., T. I. Wallow, O. Nalamasu, and E. Reichmanis. 1997. Synthesis of cycloolefin-maleic anhydride alternating copolymers for 193 nm imaging. Macromolecules 30(21):6517–6524.

Moore, G. E. 1995. Lithography and the future of Moore's law. Proceedings of the International Society for Optical Engineering (SPIE) 2438:2–17.

Reichmanis, E., F. M. Houlihan, O. Nalamasu, and T. X. Neenan. 1991. Chemical amplification mechanisms for microlithography. Chemistry of Materials 3(3):394–407.

Thompson, L. F., C. G. Willson, and M. J. Bowden, eds. 1994. Introduction to Microlithography. Washington, D.C.: American Chemical Society.

# Novel Sheet Steel Developments

John G. Speer
*Advanced Steel Processing and Products Research Center*
*Colorado School of Mines*
*Golden, Colorado*

Steel is sometimes thought of by the general public and the scientific community as a mature "low-tech smokestack" industry. Such a perspective is grossly inaccurate, and thus it is gratifying and appropriate to see steel included in the Advanced Materials session of the National Academy of Engineering's Symposium on Frontiers of Engineering. This paper will provide an introduction to the importance of technology in the steel industry, an overview of current developments related to automotive applications of sheet steels, and specific highlights of recent accomplishments in the area of "bake-hardening" sheet steels.

## TECHNOLOGY IN THE RUSTBELT

Steel products are available in many forms and with varying chemical compositions and processing, which provide innumerable combinations of properties that are useful in different applications. Important properties may include stiffness, strength, fracture resistance, formability, weldability, corrosion resistance, and cryogenic or electromagnetic characteristics. This incredible versatility contributes to the extensive use of steels in diverse and widespread applications. A watchful eye will quickly identify applications of steels in all aspects of everyday life, from bridges and buildings to appliances and cars, furniture and ductwork, doors and roofs, food and aerosol cans, ships and armor, guard rails and belted tires, pipelines and pressure vessels, tools and fasteners, gears and bearings, poles and towers, turbines and transformers, blades and rails, shafts and springs, trucks and tractors, chains and cables, kitchen surfaces and implements, and on and on.

The contributions of recent steel developments to our economy, lifestyle, national security, and society often go unappreciated, however, since steel is usually a raw material rather than a finished product, and the improvements in steel products and production technology are often "invisible" to the naked eye. Nonetheless, new steel technologies often play an enabling or important role in the development of new or improved products or product designs in industries that are thought of as more high-tech. New processes, including vacuum degassing, ladle treatment, and enhanced computer controls, have stimulated product development opportunities that take advantage of these new capabilities. Furthermore, advancements in other areas such as casting technology and production of new feedstock materials (to complement scrap recycling) have rapidly altered the structure of the steel industry, including its geography, whereby facilities are becoming smaller and located closer to growing industrial centers.

The low cost of steel is a further attribute of great significance, and technological advancements in steel processing have contributed profoundly to the industry's cost reduction efforts. It is perhaps surprising to learn that the price of steel has not increased substantially during the working lifetime of most of the symposium attendees, and the per-pound cost of premium-quality steel products can be substantially lower than that of, for example, bottled drinking water! The scale of steel production is enormous, with about 200 billion pounds (valued at close to $50 billion) manufactured annually in the United States by about 170,000 steel industry employees (American Iron and Steel Institute, Steel Manufacturers Association, and Department of Energy, 1998). Finally, environmental awareness in the industry has increased substantially, and scientific and engineering advances have contributed to dramatic reductions in air and water pollution and solid waste. Steel recycling has increased, and steel is reported to be the most recycled of materials.

These introductory remarks are not intended to be an "advertisement" for the steel industry, although it is interesting to note that the North American steel community has embarked on an information campaign to correct misconceptions about the industry among the general public (Ritt, 1998). Instead, these remarks are intended for scientists and engineers, to provide background and insight into steel's importance, and the excitement and advancements that will continue to be associated with the steel technology arena.

## BROAD-BASED TECHNOLOGY NEEDS AND OPPORTUNITIES—A MULTIDISCIPLINARY PERSPECTIVE

The technological needs and opportunities in the steel industry remain extensive and widespread, and involve numerous technical disciplines. It is not possible to review these opportunities in depth here. Readers are referred to the *Steel Industry Technology Roadmap* published in March 1998, which was prepared by the domestic steel industry (member companies of the American Iron and Steel Institute and the Steel Manufacturers Association) in cooperation with the U.S. Department of Energy (Kavanagh, 1998; American Iron and Steel Institute, Steel Manufacturers Association, and Department of Energy, 1998). The *Roadmap* was prepared explicitly to identify and prioritize the trends and drivers, new and emerging technologies, and remaining technical challenges for the industry and covers a vast scope from raw materials (iron ore, coal, etc.) processing to product development and applications technology (i.e., user) needs. Consequently, the *Roadmap* offers a clear definition of technological development and transfer opportunities whereby existing capabilities or research interests can be effectively matched with the future needs of industry. A small number of generic opportunity areas are summarized in Box 1, which is intended only to identify examples that illustrate the breadth of technical disciplines involved; each example is a sophisticated area in itself. While the medical sciences and related disciplines also are of interest to the engineering community and are not specifically addressed in the *Roadmap*, it should be noted that advancements in these areas are highly relevant to the steel industry, as they influence health insurance costs and provide a sound basis for understanding environmental issues.

## ADVANCEMENTS RELATED TO AUTOMOTIVE SHEET STEELS

Automobile manufacturing is an important application for a variety of steels. Sheet steels are the predominant material in the body structure and closure panels (i.e., outer "skin") and provide (1) formability and weldability for ease of manufacturing; (2) requisite strength and stiffness for ride stability, occupant protection during collisions, etc.; and (3) dent resistance and surface finish for optimum appearance of the painted vehicle. Bar steels are applied at even higher strength levels for shafts, gears, and bearings in engine and drive system applications, where strength, toughness, fatigue, and wear resistance are needed for durability and reliability. Finally, stainless steels provide resistance to oxidation/corrosion in exhaust systems and other chemical environments and have recently been proposed for structural applications as well (Emmons and Douthett, 1996). Steels compete vigorously with other materials in many of these applications, but steel has generally remained the material of choice due to its favorable combination of property characteristics and economics. Automotive sheet steels will be used here to provide an example of some current development areas in steel technology, and a specific product development area also will be highlighted.

Improvements in technologies related to automotive sheet steel contribute to performance or to cost reductions in manufacturing and assembly. A few recent advancements in sheet products/manufacturing/applications include the following (e.g., Shah et al., 1997, 1998).

- *Corrosion Resistant Coatings.* A variety of zinc-based coatings have been developed and applied, which have dramatically improved resistance to both

### BOX 1

| **Discipline** | **Examples of Technology Need/ Opportunity** |
|---|---|
| Mechanical engineering | Processing equipment design and maintenance, rolling process improvements |
| Electrical engineering | Reduce energy consumption of electric arc furnace for melting scrap, etc. |
| Instrumentation and sensors | Advanced process/product monitoring |
| Chemical engineering | Optimization of chemical/physical processing, improve process fluids (e.g., lubricants for rolling, forming), etc. |
| Process metallurgy | Enhanced liquid steel processing (e.g., refining, casting), iron reduction, etc. |
| Solid-state metallurgy | Product development, solid processing, etc. |
| Materials science | High-performance refractory ceramics |
| Computer science/information systems | Improved production scheduling and logistics models |
| Biology/environmental science | New biological or chemical waste treatments |
| Geology | Characterization and processing of minerals used as raw materials, for reclamation of refractories, etc. |
| Physics | Improved defect detection and property characterization |
| Chemistry | Improved corrosion/stain-resistant coatings and electroplating processes and/or consumables |
| Applied mathematics | Artificial intelligence and improved process modeling and control |
| Civil engineering | Framing techniques for residential construction, including wind and seismic loading |

cosmetic and perforation corrosion. While the coatings are not visible to the consumer in the painted condition, this development is one that is perhaps easiest for individual consumers to recognize by considering their experiences with rusted vehicles in the past, in dramatic comparison to the longevity of current auto bodies.

- *High Strength Steels.* New families of sheet steels with higher strength levels have been developed and implemented to allow for thickness reductions that provide vehicle weight savings and associated improvements in fuel economy, while at the same time reducing material cost. Even higher strengths have been developed to provide energy absorption in safety-critical components such as bumper systems and door-impact beams. A related area of current activity and future interest involves understanding the influence of steel composition/ processing/ microstructure on deformation behavior at very high strain rates (applicable to crash situations) in order to develop improved steels for optimized occupant protection.
- *Uniformity of Properties.* Manufacturing with higher-strength steels is inherently somewhat more difficult from the standpoint of forming parts without breakage and forming dimensionally consistent parts (due to a change in shape, or "springback," that occurs when a formed part is released from the stamping die cavity). Uniformity of properties (i.e., product consistency) is critical to successful manufacturing, and continued improvements are being made through improved processes and process control technologies and better understanding of the sources and effects of metallurgical and process variations.
- *Improved Surface Finish.* The painted surface appearance of vehicles is an attribute of considerable importance to some consumers. Steel "surface texture" characteristics are created and controlled during the sheet rolling and coating processes, and recent developments have included understanding the evolution of surface finish during processing and the influence of the steel surface finish on the final appearance after painting, as well as developing new surface textures, new paint formulations, and new paint application technologies (e.g., Bastawros et al., 1993).
- *Improved Exhaust and Fuel Systems.* Specialty steels such as chromium-containing ferritic stainless steel have been incorporated into exhaust systems, providing a significant increase in service life in comparison to earlier products. Similarly, new metallic/organic coating systems have been developed for sheet steels used in fuel tank applications, thereby reducing the use of (earlier-generation) lead-containing coatings and providing improved corrosion performance in the presence of alternative fuel formulations.
- *Dent-Resistant Steels.* One consequence of the trend toward reduced thickness of sheet steels for weight reduction is increased sensitivity to denting of the outer panels, from shopping carts, doors opening/closing, hail, flying stones, etc. New steel grades with improved dent resistance have been developed for such applications, including "bake-hardenable" products, which are becoming extensively used and derive a portion of their strength from metallurgical changes that are designed to occur in the elevated temperature paint curing cycle during post-stamping assembly of the vehicle.
- *New Manufacturing Technologies.* Advancements in the technology for manufacturing steel and for manufacturing with steel are highly relevant to the

development and application of automotive sheet products but are too extensive to review here. A few examples of important application technologies may be worthwhile for illustration, however, and include such developments as robotic spot welding for vehicle assembly, improved stamping lubricants applied directly by the steel producer prior to shipment, and use of tailor-welded blanks and tube hydroforming. Tailor-welded blanks involve the use of combinations of steels having different properties, thicknesses, and/or coatings, which are welded together prior to stamping rather than during assembly. This technology allows greater local optimization of material characteristics and also provides cost reduction via a reduction in the numbers of individual parts that need to be produced and stamping dies needed to produce them. Tube hydroforming is another emerging technology whereby assemblies of stampings are replaced by a single component that is shaped by internal hydraulic pressure, conforming a cylindrical tube into a complex shape defined by the surrounding die cavity.

- *Improved Design Processes.* New computing and modeling capabilities have provided the opportunity to design better, faster, and at lower cost. Simulations of sheet steel behavior during stamping allow for improved stamping feasibility evaluations and improved die designs, while deflection and crash modeling provide opportunities for improved vehicle designs. Recently, for example, a collaborative design investigation directed by a consortium of steel manufacturers identified the potential for significant automobile weight and cost savings by integrating new steel products, applications technology, and design concepts. Prototype structures of this "Ultra-Light Steel Auto Body" design have been manufactured for evaluation (Bagsarian, 1998). Opportunity remains for continued advancement in the accuracy of simulations, understanding of material behavior inputs into the models, and so forth.

## ANATOMY OF A STEEL PRODUCT DEVELOPMENT

The overview thus far has discussed broad frontiers of steel development and some general areas of interest with respect to automotive sheet steels. It is appropriate to consider one particular metallurgical development in more depth to appreciate some of the excitement, interest, and opportunity that remains in the field of ferrous physical metallurgy. Ferrous physical metallurgy remains at the heart of steel product development, and we will consider specifically here the development of bake-hardenable sheet steels for dent-resistant auto body applications produced via a hot-dip coating process.

As mentioned earlier, bake-hardenable steels are strengthened during the thermal paint curing cycle and are associated with an inherently improved strength/formability balance because the baking mechanism provides a strengthening contribution after forming. (Strength and formability are often conflicting attributes.) The bake-hardening mechanism utilizes a phenomenon in metallurgy known as strain aging, where carbon atoms dissolved in the solid iron alloy diffuse (at the paint baking temperature in this instance!) to defects in the crystalline structure called dislocations. The motion of dislocations under stress provides the means by which permanent deformation of the solid is accomplished, as in forming a stamped part. The carbon atoms that diffuse and interact with the dislocations are found to restrict dislocation mobility under stress, thereby providing the strengthening contribution we associate with bake hardening.

Carbon atoms are present in steels in the form of carbides or dissolved in “solid solution.” From the standpoint of design and processing of bake-hardenable steels, careful control of the amount of dissolved (or solute) carbon is the critical factor. Too little solute carbon results in an inadequate hardening response, while too much results in an uncontrolled response, whereby the baking occurs at room temperature prior to stamping rather than during the paint baking step at elevated temperature after stamping.

Bake-hardenable steels are readily produced and have been used to an increasing degree over the past few years in applications where a zinc-based coating is applied to the steel substrate by electrodeposition from aqueous solution near ambient temperature. In these instances the control of solute carbon is accomplished by careful processing and is unaffected by the coating process. For vehicle manufacturers that prefer to use hot-dip coatings, the situation is much different. Hot-dip coatings are applied by immersion into a molten zinc-rich bath at a relatively high temperature, and exposure to such temperatures can influence the solute carbon concentration in some steels by dissolving iron carbides. In other steels, solute carbon is totally avoided by the addition of elements that form stable carbides that are unaffected by the coating process. Design of a bake-hardenable steel for processing via this route therefore requires a strategy for obtaining a controlled level of solute in the face of these different processing responses. Such strategies have been developed and employed on a limited basis and involve controlling solute levels by careful control of either (1) the total carbon concentration in the alloy, (2) the amount of stable carbides that are present, or (3) processing at higher temperatures to dissolve some of the “stable” carbides. Any of these options can prove difficult to accomplish consistently, however, and thus a new approach was sought.

A metallurgical concept was desired whereby dissolved carbon, and therefore bake-hardening characteristics, might be readily controlled in a fashion that was relatively insensitive to small variations in chemical composition or processing. Thermodynamic considerations of carbide stability, or solubility, in different steel alloys indicated that either very stable carbides (steels containing small additions of titanium or niobium) or relatively unstable carbides (in steels without specific carbide-forming additions) are typical. Identification of an alloying approach associated with “intermediate” carbide stability was considered to be promising, which eventually led to the selection of vanadium as a key component for subsequent laboratory investigation. Vanadium is not a common alloy addition in formable sheet steels. Steel processing simulations have demonstrated the feasibility of this new concept, and follow-up technical publications, patents, and commercial activity have suggested that it may offer one means for the steel industry to meet its customers' needs for hot-dip coated bake-hardenable sheet steels (Mitchell and Gladman, 1998; Taylor and Speer, 1996, 1998). Other options may also be identified and implemented, as this exciting area of study remains highly active at several laboratories.

## CONCLUSION

This overview has provided some general understanding and appreciation of steel as a vital industry with exciting challenges and opportunities. Fruitful areas for future research were shown to cover a broad range of technical disciplines, and one example highlighted a particular area of solid-state metallurgy. Steel technology has advanced at an accelerating pace over the past few thousand years but truly remains a frontier of engineering.

## ACKNOWLEDGMENTS

Support of the sponsors of the Advanced Steel Processing and Products Research Center, an NSF Industry/University Cooperative Research Center, is gratefully acknowledged. Contributions by Ed Case, Armco, Inc., and Lawrence Kavanagh, American Iron and Steel Institute, are appreciated. Special thanks are due James Schroth, General Motors Corp., and David Matlock, Colorado School of Mines, for helpful comments on the manuscript.

## REFERENCES

American Iron and Steel Institute, Steel Manufacturers Association, and Department of Energy. 1998. Steel Industry Technology Roadmap. Washington, D.C.: American Iron and Steel Institute.

Bagsarian, T. 1998. Automakers study Ulsab's body-in-white. Iron Age—New Steel 14(June):35.

Bastawros, A. F., J. G. Speer, G. Zerafa, and R. P. Krupitzer. 1993. Effects of steel surface texture on appearance after painting. SAE Paper 930032. Warrendale, Pa.: Society of Automotive Engineers.

Emmons, J. B., and J. Douthett. 1996. Automotive frames of stainless steel. Advanced Materials and Processes 150(August):23–25.

Kavanagh, L. W. New frontiers in steelmaking. 1998. Advanced Materials and Processes 153(January):29–31.

Mitchell, P., and T. Gladman. 1998. Vanadium in interstitial free steels. Pp. 37–48 in Mechanical Working and Steel Processing Conference Proceedings, 1997. Warrendale, Pa.: Iron and Steel Society.

Ritt, A. 1998. Fashioning a better image of steel. Iron Age—New Steel 14(March):50–54.

Shah, V. C., J. G. Speer, B. Zuidema, and G. D. Millar, eds. 1997. Steel in automotive applications. SAE Special Publication 1259. Warrendale, Pa.: Society of Automotive Engineers.

Shah, V. C., J. G. Speer, B. Zuidema, and G. D. Millar, eds. 1998. Steel sheet and steel bar products and processing for automotive applications. SAE Special Publication 1349. Warrendale, Pa.: Society of Automotive Engineers.

Taylor, K. A., and J. G. Speer. 1996. Bake-Hardenable Vanadium Containing Steel and Method of Making Thereof. United States Patent No. 5,556,485. September 17.

Taylor, K. A., and J. G. Speer. 1998. Development of vanadium-alloyed, bake-hardenable sheet steels for hot-dip coated applications. Pp. 49–61 in Mechanical Working and Steel Processing Conference Proceedings, 1997. Warrendale, Pa.: Iron and Steel Society.

# SIMULATION IN MANUFACTURING

# Role of Simulation in the Design of Next-Generation Engines for Military Aircraft

MARK N. GLAUSER
*Department of Mechanical and Aeronautical Engineering*
*Clarkson University*
*Potsdam, New York*

To stimulate and enable the emergence of new concepts for affordable advanced flight vehicles and propulsion systems, fast, adequate, and affordable models are necessary to support computational fluid dynamics (CFD)-based design in realistic complex flows. Complex turbomachinary flows, such as those used in current and future military aircraft engines, exhibit unsteady aerodynamics, forced blade response and high-cycle fatigue, complex confluent interacting flows with strong rotation, compressor surge and stall instabilities, and losses and extremely high temperatures in both the combustor and turbine components. In addition to the complexity of the systems themselves, the flows exiting these propulsion systems are complex three-dimensional shear flows characterized by large temperature, velocity, and density fluctuations. Such exit flows can be detrimental to other components of the aircraft such as control surfaces like flaps because of extremely high heat loading. They also impact the acoustic and thermal signatures of the flight vehicle. In this paper, a brief overview of current uses of CFD methods will be presented and elucidated by example. We will begin at the compressor, skip the combustor for now (we will discuss an exciting very recent development for simulating combustor flows later), move to the turbine, and finally discuss applications to exit flows. Problem areas as well as successes will be noted and future directions discussed.

Currently, CFD tools, as well as various experimental methods such as laser Doppler anemometry and particle image velocimetry, are being used by William W. Copenhaver at Wright Patterson Air Force Base to focus on the influence of stator wakes on transonic stage performance in compressors. Copenhaver (1998) is attempting to identify the influences that upstream wakes have on transonic blade row aerodynamic and aeromechanical responses. The simulations are based on APNASA, which is a National Aeronautics and Space Administration Lewis steady-state code with k-epsilon turbulence model. The wake information computed from the APNASA code is substantially different from experimental results at this point. The key problems with the simulations are the lack of inclusion of the unsteady nature of the flow and the inadequate turbulence model.

Richard Rivir, also of Wright Patterson Air Force Base, is focusing on the prediction of unsteady heat transfer at the leading edge of the high-pressure turbine to assess leading-edge cooling methods. Rivir (1998) is using an unsteady Navier Stokes code with limited success, sometimes having as much as a two to one variation between simulation and experiment. Clearly, more physics must be built into such simulations, in particular to properly account for turbulence.

A very exciting application that impacts the exit flow involves the use of CFD, coupled with careful analysis and experiment, to guide the development of active flow control of the hot jet plume for the reduction of the C-17 flap temperature. Currently, the hot plume impingement necessitates titanium plates on the C-17 flap system and susceptibility to infrared threat. CFD results have been used to guide the development of active control strategies for reduction of the plume temperature through enhanced mixing. This effort, termed active core exhaust control, has transitioned to a full-scale engine test. Results from the full-scale test have shown the feasibility of mixing out the plume so that by 10 diameters downstream the peak temperature has been reduced by a factor of two. The CFD played a key role in parameter space surveys, providing guidance on optimum injection flow rate and optimum injection frequency as well as placement of injectors.

As the above examples show, simulations play an increasingly important role in design methodologies for complex turbomachinary flows, although much fundamental work is still required to improve existing simulation tools. We can categorize the simulations of flows as follows. The direct numerical simulation (DNS) of turbulent flows involves solving the full Navier Stokes equations, resolved to all scales (Moin and Mahesh, 1998). DNS provides a powerful tool for model development and fundamental physical insight into flow physics, although it is limited to relatively low Reynolds numbers and simple geometries. Large eddy simulations (LESs) resolve the larger scales in the flow, and the smaller scales (subgrid scale) are modeled (Lesieur and Métais, 1996). Currently, LES applications have been limited to relatively simple geometries but at somewhat higher Reynolds numbers than DNS. Recently much progress has been made in applying LES to more complex flows, and an application to gas turbine combustors will be discussed below. Reynolds Averaged Navier Stokes (RANS) and unsteady RANS (called by some “very large eddy simulations”) are the current workhorses in industry (Speziale, 1991). These methods involve solving evolution equations for the mean flow as well as various levels of the turbulent statistics. Finally, new methods, termed low-dimensional modeling (LDM), are being explored to help in the guidance and implementation of active flow control strategies (Lumley and Blossey, 1998). The philosophy of LDM is to develop very simple physics-based simulation models that contain sufficient information for implementing flow control and insight into the dynamics of the large eddies (coherent structures). Except for DNS, some form of modeling is required for all of the above.

In terms of future directions for simulation of complex flows, for turbulence modeling, unsteady effects must be included as well as some component of the turbulent structure. One example of an approach to include structure involves including the structure of the turbulence in single-point turbulence models (Reynolds and Kassinos, 1995). The structure of turbulence often plays a key role in the transport of turbulence stresses. The absence of this type of information from practically all currently used one-point turbulence models (Reynolds stress transport and eddy viscosity) limits their performance. The structure-based model of Reynolds and Kassinos (1995) is now approaching a level of maturity that will allow testing in more realistic complex flows. For implementation of LESs in practical flows, methods must be developed to deal with complicated boundary conditions and improve subgrid-scale modeling. As an example, Pierce and Moin (1998) have developed an LES-based prediction method for combustor flows with swirl. The effects of swirl are cleverly added by imposing a body force via swirling inflow conditions, resulting in a practical LES tool for studying realistic combustor flows. Finally, for implementation of low-dimensional models in practical flows, methods must be developed to deal with complicated boundary conditions and improve subgrid-scale modeling. A key need is to extend such models to flows with separation since their control is a focus of much current interest.

## REFERENCES

Copenhaver, W. 1998. Turbomachinary fluid mechanics. Pp. 79–81 in Proceedings of the Aerodynamics and Turbulence Conference. Washington, D.C.: Air Force Office of Scientific Research.

Lesieur, M., and O. Métais. 1996. New trends in large-eddy simulations of turbulence. Annual Review of Fluid Mechanics 28:4845–4882.

Lumley, J. L., and P. Blossey. 1998. Control of turbulence. Annual Review of Fluid Mechanics 30:311–327.

Moin, P., and K. Mahesh. 1998. Direct numerical simulation: A tool in turbulence research. Annual Review of Fluid Mechanics 30:539–578.

Pierce, C. D., and P. Moin. 1998. Method for generating equilibrium swirling inflow conditions. AIAA Journal 36(7):1325–1327.

Reynolds, W. C., and S. C. Kassinos. 1995. One-point modeling of rapidly deformed homogeneous turbulence. Proceedings of the Royal Society of London A 451:87–104.

Rivir, R. B. 1998. Turbine aerothermal research at the Air Force Research Laboratory. Pp. 345–347 in Proceedings of the Aerodynamics and Turbulence Conference. Washington, D.C.: Air Force Office of Scientific Research.

Speziale, C. G. 1991. Analytical methods for the development of Reynolds-stress closures in turbulence. Annual Review of Fluid Mechanics 23:107–157.

# Role of Simulation in Understanding Surface Roughness in Formed Aluminum Parts

ARMAND J. BEAUDOIN
*Department of Mechanical and Industrial Engineering*
*University of Illinois at Urbana-Champaign*
*Urbana, Illinois*

## INTRODUCTION

One does not have to look far to observe metals as an engineering material with widespread application. Metals offer the designer a class of materials with a broad range of behaviors. Further, metals are accompanied by a long history of experimental characterization and continuous refinement of techniques for analysis. The ability to describe metal deformation is critical to both the design and the manufacture of metal products. Accurate analysis is realized through the wedding of three engineering disciplines: materials science, mechanical engineering, and computer science. It is the intent of this paper to demonstrate how fruits from these branches of engineering science may be brought together to reveal insight into fundamental aspects of metal deformation—on a scale that bears relevance to industrial application.

In the following, a general framework that incorporates a detailed and physically based material description—crystal plasticity—into finite element procedures is outlined. The success of crystal plasticity theory in predicting behavior in metal forming is built on several things: initialization of a mathematical description for a complex microstructure using experimental data, effective use of computational resources, and visualization of results in a manner that reflects experiment. It is argued that such detail offers both improved predictive capability and effective utilization of computing resources. In closing it is proposed that continued development of physically motivated models will enable the simulation of more general thermomechanical processes for metal production.

## APPROACH

Metals are composed of crystals (see Figure 1). Taken individually, the crystals are anisotropic. That is to say that their properties are dependent on the direction; the yielding of a crystal depends on the orientation of the applied load. Typically, a metal product intended for an engineering application is a "bulk" material—there are lots of crystals. Provided a material point[1] in the bulk workpiece contains a sufficient number of crystals, the properties of that point may be treated as a polycrystal (Kocks, 1970). The distribution of crystal orientations in the polycrystal is commonly referred to as the texture. Because the distribution is rarely a random one, the texture contains preferred orientations and thus imparts directional properties to the polycrystal.

The field of materials science provides a wealth of theory for describing the deformation response at the scale of the crystal (Bishop and Hill, 1951; Taylor, 1938) and collecting the responses of underlying crystals to model the plasticity of the polycrystal. Accompanying the development of theory are allied experi-

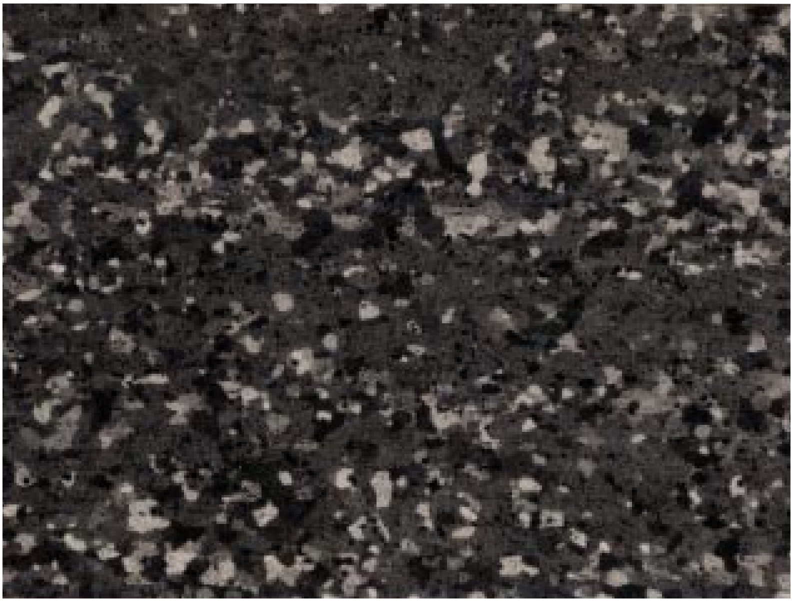


**FIGURE 1** Metal microstructure composed of crystals. SOURCE: Reprinted with permission from The Minerals, Metals and Materials Society and ASM International (Beaudoin et al., 1998).

---

[1] A material point has a spatial dimension over which behavior may be characterized in some average sense. This is in contrast to the usual mathematical concept of a point.

mental techniques for measuring the texture of a metal. The distribution of normal vectors to crystal planes may be collected using x-ray diffraction. The techniques of quantitative texture analysis enable the development of distribution functions for crystal orientation from the x-ray data. As the name suggests, the result is a quantitative description of the texture. Alternatively, backscatter Kikuchi diffraction (BKD) patterns may be collected through electron microscopy (Adams et al., 1993; Wright and Adams, 1992). This technique provides not only the crystallographic orientation but also the spatial location of the orientation measurement. As will be pointed out below, parallel computation enables the treatment of large datasets generated through material characterization. The combination of experimental methodology and crystal plasticity theory provide for both the initialization and evolution of texture in a computer simulation.

At the heart of finite element methods is the capability to treat inhomogeneity in the field variables of interest. With regard to plastic deformation of metals, these variables are forces and displacements. The finite element formulation may be melded with the constitutive description provided by crystal plasticity theory described above. In this way, deformation of workpieces may be examined with gradients in the field variables accommodated through finite elements and concomitant response at material points within elements achieved through the crystal plasticity description. In general, there are two approaches to combining finite elements with crystal plasticity theory. A polycrystal may be contained within each finite element (see Figure 2), or a finite element mesh may be used to discretize a polycrystal (see Figure 3).

For the analysis of a bulk processing operation, an aggregate of crystals is

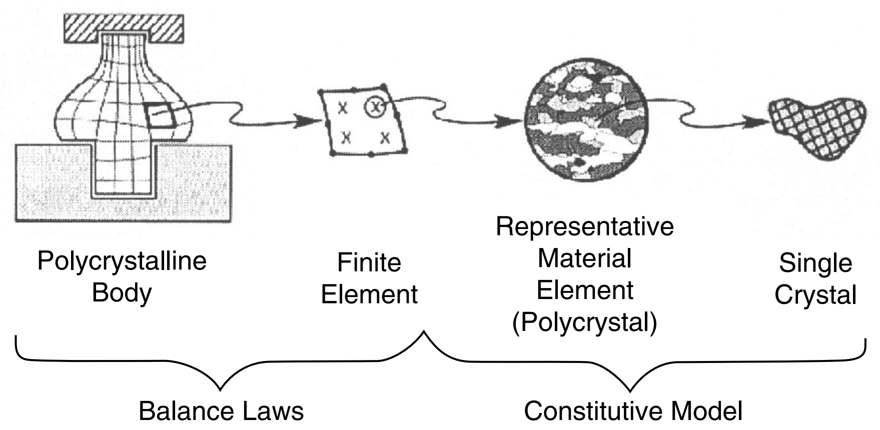


**FIGURE 2** Finite element analysis of bulk forming operation. SOURCE: Reprinted with permission from Cambridge University Press (Kocks et al., 1998).

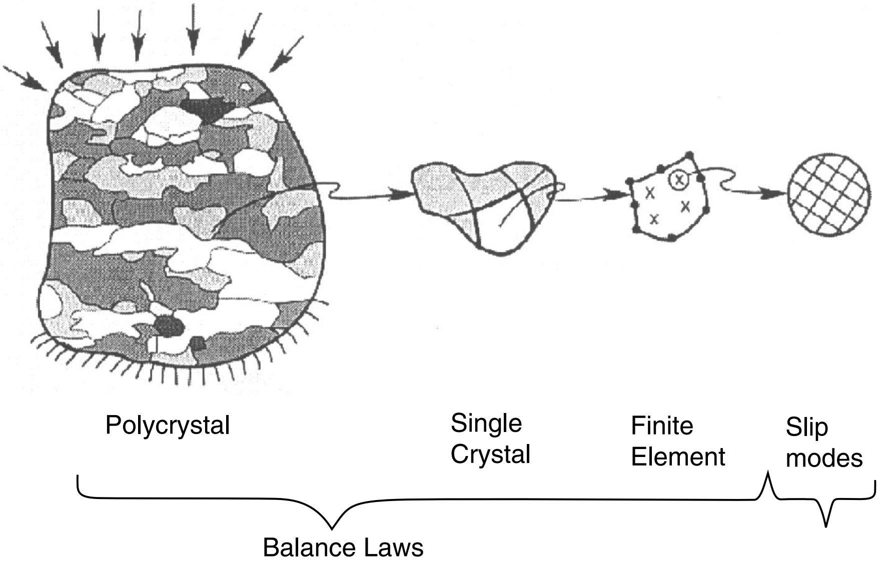


**FIGURE 3** Finite element analysis of metal microstructure. SOURCE: Reprinted with permission from Cambridge University Press (Kocks et al., 1998).

placed at integration points within the finite element. The element response follows from the aggregate response of a contained polycrystal. The spatial dimension of a finite element should be orders of magnitude larger than the respective dimension of an individual (representative) crystal (Figure 2). Within the aggregate each crystal deforms in accordance with the deformation of the overlying finite element. This approach is useful for treating deformation inhomogeneity that arises from interaction between boundary conditions (typically owing to frictional contact between the metal workpiece and the tooling) and overall material anisotropy. Examples of operations that have been treated by using this approach are drawing (Balasubramanian and Anand, 1996; Smelser and Becker, 1991), rolling (Mathur and Dawson, 1989), hydroforming (Beaudoin et al., 1994), and autobody sheet forming (Bryant et al., 1994). The deformed shape of the workpiece arising from the simulation may differ significantly from that derived using isotropic material properties—a situation more in line with industrial practice.

Finite elements may also be applied to examine the deformations within a polycrystal. The emphasis is on inhomogeneities that arise from interactions between crystals. The dimension of a finite element is typically orders of magnitude less than the polycrystal, and boundary conditions are constructed to render a homogeneous deformation (Figure 3); that is, the entire finite element mesh represents a material point. In contrast to the polycrystal fully contained within a single element, crystals gain additional degrees of freedom and conform with neighbors both in the sense of equilibrium and compatible deformation. This technique has been used to study grain interactions, origin of recrystallization nuclei (Beaudoin et al., 1996), and surface roughening (Becker, 1998). Such simulations provide considerable insight into the metal physics.

## APPLICATION

Surface roughening developed during sheet forming operations may take a variety of forms. If the roughening has a random character, the surface may develop the character of an orange peel. Another possibility is that the roughness may take on a more directional character, with striations aligned in one direction. This defect is referred to as “ridging” or “roping” and is observed in aluminum alloys and ferritic stainless steels. Collectively, the ridges impart a corrugated pattern to the part surface. Alignment of the ridges is generally in the direction of prior manufacturing operations for rolling of the sheet. Such an effect is undesirable from the standpoint of visual appearance. In severe manifestations of the defect, grooves that formed between ridges act to limit formability of the product.

An idealized model was developed to study potential surface deformations arising from bands of grains with a similar crystallographic orientation. Automated collection of BKD measurements from samples of aluminum sheet was carried out by using the orientation imaging technique mentioned above. A mesh was formed by using 14,640 eight-node bricks, and a crystal orientation was assigned based on the BKD data. The surface deformation predicted for forming in biaxial stretch is shown in Figure 4. For plotting purposes the coordinate normal to the sheet was magnified so as to emphasize the roughening tendency. Shading is used to indicate orientations of crystals forming the microstructure. The cube orientation, generally associated with recrystallization following thermal treatment, is shaded in dark gray. Crystallographic orientations that have origins in the cold working of face-centered cubic metals (labeled S, Cu, and Br as a matter of historical convention) are marked with lighter shades of gray. In general, the cube grains tend to yield regions of greater thinning, forming a groove. Adjacent to the band of cube orientations is a ridge lying along the midline of the mesh.

## FUTURE DIRECTIONS

The strength of crystal plasticity in describing deformation is founded principally on the ability of the theory to detail the *average* behavior within a crystal. Indeed, it is this average response that dictates behavior in deformation. Yet in many manufacturing processes deformation is closely allied with thermal treat-

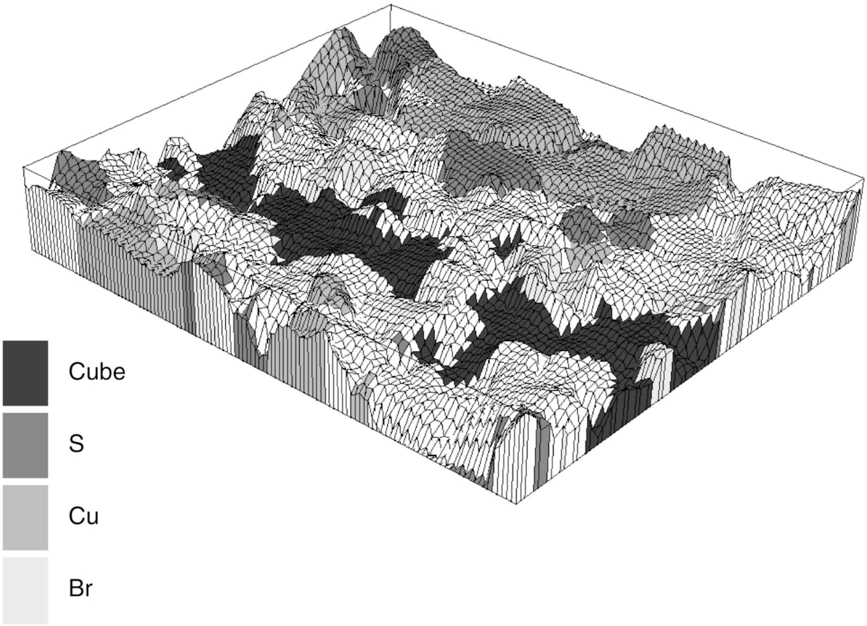


**FIGURE 4** Simulation of surface roughening. SOURCE: Reprinted with permission from The Minerals, Metals and Materials Society and ASM International (Beaudoin et al., 1998).

ment. Thermomechanical processing elicits the nucleation and growth of crystals within a microstructure, often dramatically changing the metal anisotropy. This nucleation and growth process is dependent on what happens “at the edges” (Doherty et al., 1997); in these processes the tails of the orientation distribution may bear more influence than the mean. The development of substructure in polycrystals and the origins of recrystallization are active areas of research that will lead to a more complete description of metal behavior. As an example, while present technologies reveal insight into the mechanisms of surface roughening (as presented herein), the correction of the defect through hot rolling and heat treatment is not—at present—a tractable problem for simulation.

The results presented follow from theory that has been developed since the early 1900s. The success of the crystal plasticity approach follows from maintaining fidelity in the description of metal physics at the expense of detail. However, parallel computation and sophisticated numerical methods are available, and the approach outlined herein provides opportunity for leveraging such resources. The challenge lies in continued development of physically motivated, quantitative descriptions of metal behavior. The extension to more general thermomechanical processing mandates the development of accurate models with variables of state that are initializable and may be evolved to characterize properties. The continuous improvement of existing production processes and the development of novel processes will be enhanced by ongoing development of theories for metal behavior.

## REFERENCES

Adams, B. L., S. I. Wright, and K. Kunze. 1993. Orientation imaging: The emergence of a new microscopy. Metallurgical Transactions A 24A(4):819–831.

Balasubramanian, S., and L. Anand. 1996. Single crystal and polycrystal elasto-viscoplasticity: Application to earing in cup drawing of F.C.C. Materials. Computational Mechanics 17(4):209–225.

Beaudoin, A. J., P. R. Dawson, K. K. Mathur, U. F. Kocks, and D. A. Korzekwa. 1994. Application of polycrystal plasticity to sheet forming. Computer Methods in Applied Mechanics and Engineering 117(1–2):49–70.

Beaudoin, A. J., H. Mecking, and U. F. Kocks. 1996. Development of localized orientation gradients in FCC polycrystals. Philosophical Magazine A 73(6):1503–1517.

Beaudoin, A. J., J. D. Bryant, and D. A. Korzekwa. 1998. Analysis of ridging in aluminum auto body sheet metal. Metallurgical and Materials Transactions A 29A(9):2323–2331.

Becker, R. 1998. Effects of strain localization on surface roughening during sheet forming. Acta Materialia 46(4):1385–1401.

Bryant, J. D., A. J. Beaudoin, and R. T. VanDyke. 1994. The effect of crystallographic texture on the formability of AA 2036 autobody sheet. SAE Transactions Sect. 5 103:49–57.

Bishop, J. F. W., and R. Hill. 1951. A theory of the plastic distortion of a polycrystalline aggregate under combined stresses. Philosophical Magazine 42:414–427.

Doherty, R. D., D. A. Hughes, F. J. Humphreys, J. J. Jonas, D. Juul-Jensen, M. E. Kassner, W. E. King, T. R. McNelley, H. J. McQueen, and A. D. Rollett. 1997. Current issues in recrystallization: A review. Materials Science and Engineering A238(2):219–274.

Kocks, U. F. 1970. The relation between polycrystal deformation and single crystal deformation. Metallurgical Transactions 1(5):1121–1143.

Kocks, U. F., C. N. Tomé, and H.-R. Wenk, eds. 1998. Texture and Anisotropy: Preferred Orientations in Polycrystals and Their Effect on Materials Properties. Cambridge: Cambridge University Press.

Mathur, K. K., and P. R. Dawson. 1989. On modeling the development of crystallographic texture in bulk forming processes. International Journal of Plasticity 5(1):67–94.

Taylor, G. I. 1938. Plastic strain in metals. Journal of the Institute of Metals 62:307–324.

Smelser, R. E., and R. Becker. 1991. Earing in Cup Drawing of a (100) Single Crystal. Pp. 457–471 in ABAQUS User's Conference Proceedings. Providence, R.I.: Hibbitt, Karlsson & Sorensen.

Wright, S. I., and B. L. Adams. 1992. Automatic analysis of electron backscatter diffraction patterns. Metallurgical Transactions A 23A(3):759–767.

# Machine Performance Assessment Methodology and Advanced Service Technologies

JAY LEE
*United Technologies Research Center*
*East Hartford, Connecticut*

## ABSTRACT

*This paper introduces an assessment methodology for machine performance degradation through correlated neural behavior reasoning. This methodology provides a behavior-based model for machine performance monitoring and failure prevention. In addition, research challenges and opportunities in digital service technology for the life-cycle support of products and engineering systems in a global enterprise are introduced.*

## INTRODUCTION

Today's industries are facing serious structural problems brought about by their rapid development of overseas activities within a globally integrated manufacturing enterprise. Service and maintenance are becoming extremely important for companies to maintain their manufacturing productivity and customers' satisfaction in foreign regions. The recent rush to integrate highly sophisticated manufacturing equipment with locally supported design and manufacturing engineering practices has further increased the use of relatively unknown and untested technologies. Frequent breakdowns of machinery have been a serious concern. Difficulty in identifying the causes of system failures has been attributed to several factors, including system complexity, uncertainties, and lack of adequate troubleshooting tools (National Research Council, 1995). A new paradigm should focus on quality of service in the products we design and manufacture. Factories in different regions need to be coordinated through use of state-of-the-art information technologies to ensure consistent manufacturing and product quality. For example, the performance of a machine could be monitored and assessed from anywhere in the world. In addition, information on productivity, diagnostics, and service evaluation of manufacturing systems could be shared among different locations and partners.

Digital service technology is an emerging field that addresses service issues for manufacturers and customers. With growing manufacturing globalization activities, companies are looking for ways to assess the performance of their manufacturing operations and products in remote sites. Digital maintenance diagnostics and maintenance tools such as "watchdog"-type information mechatronics with integrated media will improve the effectiveness of maintenance activities (Goncharenko et al., 1997; Lee, 1992, 1995a,b, 1996c; Lee and Kramer, 1993; Shi et al., 1997). A typical digital service technology-based teleservice engineering system is illustrated in Figure 1.

There are three critical emerging technologies involved in the teleservice engineering system (Goncharenko et al., 1997; Lee, 1996b; Shi et al., 1997):

*Behavior Assessment and Performance Degradation Evaluation Agent (Watchdog Neural Chip).* Assessment of a machine's performance information to operators in remote sites requires an integration of many different sensory devices and reasoning agents. Generally speaking, the operational performance of components, machines, and processes can be divided into four states: normal operation state, degraded state, maintenance state, and failure state (Lee, 1992, 1995b). Figure 2 shows the typical performance states of a machine. The degraded states can be defined at a gross level or a detailed level. At a gross

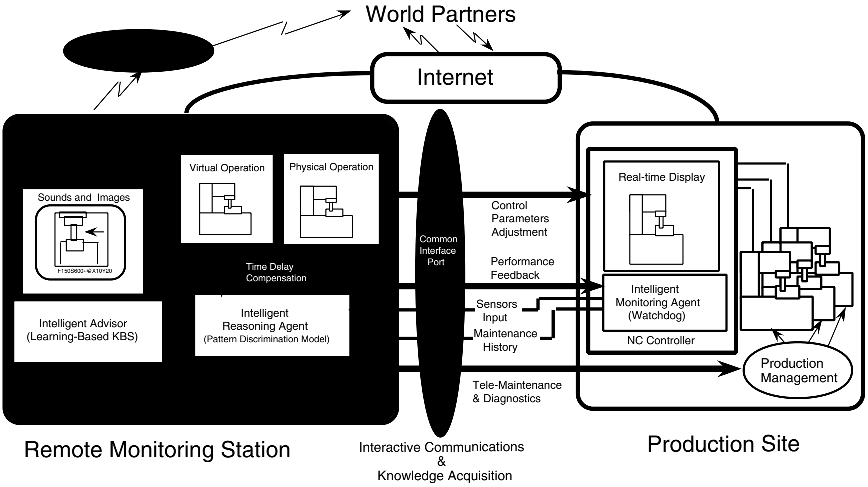


**FIGURE 1** A typical teleservice engineering system. SOURCE: Lee, 1996a.

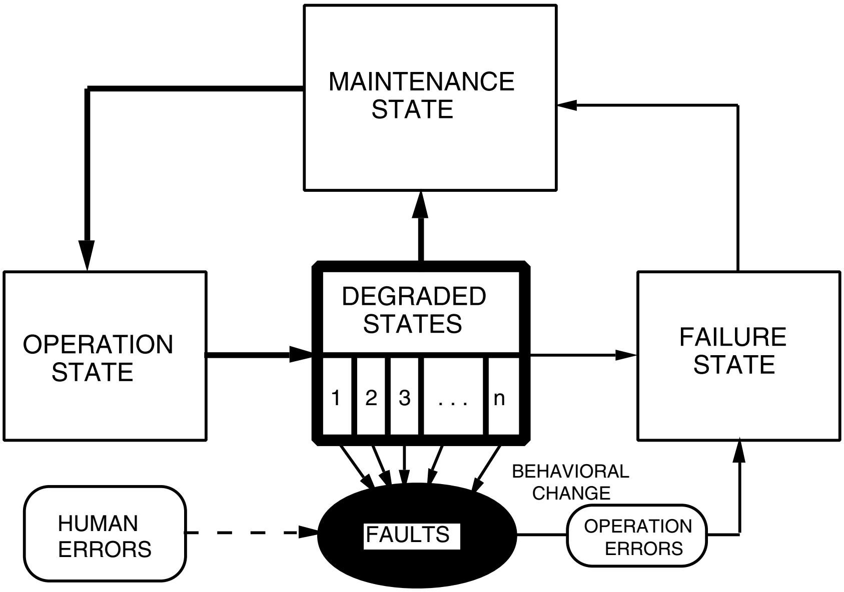


**FIGURE 2** Machine's performance states. SOURCE: Lee, 1992.

level, a component is described as degraded whenever a deterioration occurs but does not cause loss of its function. For example, component degradation occurs whenever corrective maintenance is required, such as a loose belt, a worn brush of a motor, or dust on a photosensor, but the components have not failed. Detailed degradation is associated with a given range of characteristics of the components, or performance of the components, such as detailed degraded states for a car battery indicator and temperature indicator. The advantage of defining more detailed degradation states is that we can accurately predict impacts on the failure of a component. When aging occurs, the component and machine generally progress through a series of degradation states before failure occurs. If a degradation condition can be measured and detected, proactive corrective maintenance activities can be performed before further degradation or failure occurs.

To effectively measure the degradation of a machine, its behavior and associated information from operators and its working environment need to be assessed adaptively. A watchdog agent, a neural computing algorithm, has been developed by the author to provide on-line composition and reasoning. In addition, this agent could be connected with a telephone jack so that machine behavior and its performance information could be accessed and evaluated from a remote site. The watchdog agent and its working principles are shown in Figure 3.

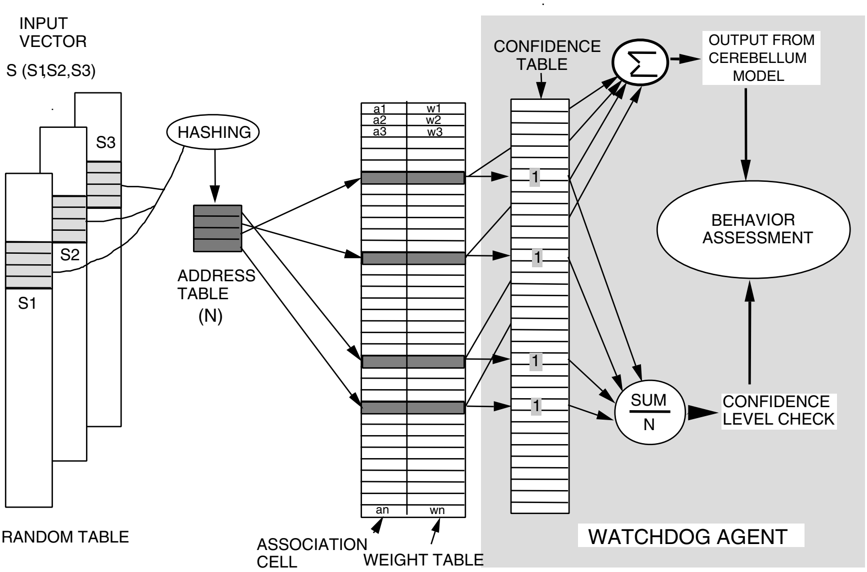


**FIGURE 3** Watchdog agent for machine behavior assessment. SOURCE: Lee, 1992.

Briefly speaking, the watchdog agent can be described as a computing device that accepts an input vector $\mathbf{S} = (\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_n)$ and produces an output vector $\mathbf{P} = \mathbf{F}(\mathbf{S})$. To compute the output vector **P** for a given input state **S**, pair mapping is performed, namely

$$\mathbf{f}: \quad \mathbf{S} \rightarrow \mathbf{A}$$
$$\mathbf{g}: \quad \mathbf{A} \rightarrow \mathbf{P}$$

where **A** is called the association cell vector, which is actually a large table of memory addresses. Given an input vector $\mathbf{S} = (\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \ldots, \mathbf{S}_n)$, the mapping function **f** points to some memory addresses (location in the **A** table); these locations are among the association cells **A** and are referred to as the active association cells. The number of active association cells for any given input is a fixed parameter of the watchdog module selected by the user. Weight(s), or tabulated values that contribute to the output response, are attached to each association cell. One weight if output vector **P** is one dimensional, and **n** weights if **P** is **n** dimensional, is attached to each cell. Mapping function **g** uses the weights associated with the active association cells and generates an output response. Function **g** may be a simple summation if the weights are attached to active association cells, or it may vary during the training and storage process. Any input vector is therefore a set of address pointers, and the output vector, in its simplest form, is the sum of weights attached to these address pointers. Thus, each element of the output $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n)$ is computed by a separate watchdog from the formula

$$\mathbf{p(k) = a_1(k) \times w_1(k) + a_2(k) \times w_2(k) + \ldots + a_n(k) \times w_n(k)}$$

where $\mathbf{A(k)} = \mathbf{(a_1(k), a_2(k), \ldots, a_n(k))}$ is the association cell vector of the $\mathbf{k^{th}}$ watchdog and $\mathbf{W(k)} = \mathbf{(w_1(k), w_2(k), \ldots, w_n(k))}$ is the weight vector of the $\mathbf{k^{th}}$ watchdog. The value of the output for a given input can be changed by changing the weights attached to the active association cells.

A procedure for entering a function in a watchdog agent is as follows:

1. Assume **F** is the function we want the watchdog agent to compute. Then $\mathbf{P} = \mathbf{F(S)}$ is the desired value of the output vector of each point in the input space.
2. Select a point **S** in input space where **P** is to be stored. Compare the current value of the function at that point $\mathbf{P^*} = \mathbf{F(S)}$.
3. For every element in $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n)$ and $\mathbf{P^* = (p^*_1, p^*_2, \ldots, p^*_n)}$, if $\mathbf{l\ p_I} - \mathbf{p^*_i\ l} < \mathbf{e_i}$, where $\mathbf{e_i}$ is an acceptable error, do nothing; the desired value is already stored. However, if $\mathbf{l\ p_I} - \mathbf{p^*_i\ l} > \mathbf{e_i}$, add to every weight that contributed to $\mathbf{p^*}$ the quantity $\mathbf{M_i} = (\mathbf{p_I} - \mathbf{p^*_i}) \,/\, \mathbf{N},$ where **N** is the number of weights that contributed to $\mathbf{p^*_i}$.

The values of a set of inputs act as pointers to a table of random numbers. These numbers are hashed to form a set of addresses in a weight table. The values in the weight table pointed to by these addresses are summed to yield an output. Training consists of adjusting the values in the weight table based on the error between their present summation value and the desired output for the particular values of the input that resulted in these addresses.

*Knowledge Learning and System Failure Recovery Agent.* Knowledge-intensive intelligent tools are required for the acquisition and organization of data in a machine and its working environment to track the behavior of the machine at any given time. A watchdog chip will serve as the blackbox of a machine and is able to keep the signatures of major components. In case of failure, operators can access the blackbox and obtain the last several minutes of information about the behavior of the machine. As a result, faults can be located quickly, and the system can be recovered rapidly. Knowledge-based information can also be shared with other user sites.

*Digital Service Technologies for Collaborative Maintenance.* Multimedia-based tools are required to support remote users for maintenance assistance. Interactive and collaborative tools will enable technical personnel to perform diagnostics from a remote distance. Digital maintenance diagnostics and maintenance tools such as a smart helmet with integrated media will improve the effectiveness of the production equipment through collaborative maintenance and diagnostics. A smart glove would enable operators to perform machine maintenance and performance adjustment collaboratively.

## RESEARCH ISSUES AND OPPORTUNITIES IN DIGITAL SERVICE TECHNOLOGY

To achieve the digital service enterprise in today's globalized manufacturing industries, issues and technical challenges need to be addressed. The fundamental research issue that prevents us from resolving these problems is an inadequate understanding of the behavior of manufacturing machines and equipment on a daily basis. We simply do not know how to measure the performance degradation of components and machines. We lack validated predictive models that could tell us what would happen when the process parameters take on specified values. The research challenge, therefore, will be to construct models and agents that can be used to assess a machine's behavior remotely. These models must be interoperable and responsive to agents, so that remote users do not always have to communicate with them through a human intermediary.

Some research challenges and opportunities in selected technical areas have been discussed by Goncharenko et al. (1997) and Shi et al. (1997). Below is a summary of these challenges and research opportunities:

- *Sensing System Standardization.* In-process sensing and in-machine sensing are the foundation for teleservice systems. Because of the complexity of a machine and its associated environment, different types of sensors may be used, requiring different data acquisition protocols and systems. As teleservice practice expands, more machines/processes will be linked within the remote diagnosis system. Standards should include the selection of sensor type for typical signals, sensor signal output ranges, protocols, and so forth. There are several industry standards currently available. Research efforts are needed to develop synthesis tools for sensory system design and implementation.
- *Adaptive Sensor Fusion and Affordability.* One physical machine fault (e.g., an unbalanced shaft) may generate different symptoms (e.g., vibration, temperature changes, motor load variation) and can be measured by different sensors (e.g., an accelerometer, thermal couples, motor current). Similarly, one sensor may sense different types of machine faults that are occurring simultaneously, and the sensor's sensitivity may vary as the operating condition changes. Thus, adaptive sensor fusion, which will improve the reliability of the diagnosis, should be emphasized. However, affordability should also be emphasized. Issues of affordability will lead to cheaper sensing techniques, fewer demands on sensing data accuracy, and high requirements on the noise rejection capability of the

algorithms. In addition, optimal objective-oriented sensor placement strategy needs to be investigated. Remote diagnosis provides a better opportunity to develop the aforementioned techniques by implementing knowledge, experience-sharing, and self-learning.

- *Data Compression, Features Extraction, and Task Allocation.* Sensing data will be transferred to a central server through the Internet. If transferring the raw data directly, long time delays due to heavy traffic may be experienced. In addition, more storage space may be required in the central server. One area that needs to be addressed is data preprocessing for data compression and feature extraction. It should be noted that the data compression task would be different from the conventional approach in image analysis and signal processing, even though there are many similarities and some techniques can be borrowed. The research emphasis here is on how to combine the engineering knowledge and diagnosis requirements in the data compression and preprocessing stages. Engineering feature-based data compression should be developed, which considers the important features in diagnostic analysis and system performance assessment. Examples in this category include (1) identifying engineering model parameters from the data and transferring only the model coefficients to the server, (2) extracting features from the original sensing signature and adjusting the threshold based on the interested signal information, and (3) transferring data continuously to the server from low- to high-wavelet coefficients and developing stopping criteria based on the decision-making strategy. Task allocation is another research topic closely related to teleservice engineering systems. According to the requirements for response time to faults, the monitoring and diagnosis tasks can be classified as immediate response (e.g., tool breakage, collision), intermediate response (e.g., tool wear, temperature compensation), and slow response (e.g., machine wear and degradation, environment changes). According to the information required and the complexity involved in decision making, the tasks can be classified as single-variable process change detection, multivariate analysis, and integrated decision making. A study needs to be conducted to classify all tasks into various categories. As a result, the watchdog agent could perform simple on-line and real-time process change detection for tasks requiring immediate response. In the future, teleservice engineering systems should be directed toward more complex diagnostic tasks that necessitate correlated degradation assessment for multivariate data.
- *Collaborative Maintenance and Diagnostics.* A major advantage of implementing teleservice engineering systems is the opportunity for collaborative maintenance and diagnostics. These can be achieved through (1) fault condition data collection, in which the remote diagnostics system provides an opportunity to accumulate more machine/process fault conditions and thus a better diagnosis algorithm can be developed from the fault conditions in various remote sites, and (2) fault diagnosis, in which the information is available in the server and can be accessed by experts at various locations. Thus, knowledge distributed over var-

ious sites can be integrated to perform more complex collaborative diagnoses. However, significant research on how to manage the information and distributed decision-making components is required. Such topics as distributed artificial intelligence, competitive decision making, and risk management should be researched in the context of collaborative diagnosis.

- *Self-learning and Supervised Learning for Smart Service Agents.* Even though the topics of supervised learning and self-learning have been studied by various researchers, they become more critical in a remote diagnosis environment. Because of the nature of remote diagnosis, process data from many different locations can be accessed. Updated knowledge will be acquired much faster than with traditional diagnosis techniques. Thus, supervised learning (or self-learning) is more critical. Furthermore, it is anticipated that fewer well-trained supervisors may perform the task due to the availability of new information. Another challenge is how to assess new information before using it for the purpose of learning. The information assessment stage can be integrated as part of the supervised/self-learning research.
- *Integrated Performance Assessment.* The teleservice system will provide several categories of information together, such as on-line process/machine sensing data, historical fault/degradation data, and machine design information. All information should be integrated when conducting a performance evaluation. The author has proposed the concept of machine physiology, which focuses on behavior-based computation rather than model-based computation for machine performance degradation assessment without using machine fault condition information. In addition, knowledge of machine performance can be learned and modeled and eventually used in machine performance compensation. Thus, the performance of a typical machine may not necessarily degrade over time but instead be improved by using learning-modeling-compensation techniques.
- *Self-maintenance and Dependability.* A backup strategy should be considered for overall system reliability improvement and for determining what to do if the Internet system malfunctions. Under various conditions, the system should be able to operate and perform basic functions, albeit with deteriorated performance. For example, if the Internet is not available for data transfer and remote diagnosis, local data processing should be executed to conduct essential tasks that are nominally carried at the remote site. How to design a redundant system with minimum cost and high performance is a challenge. This is also related to task allocation and analysis, simplified diagnosis algorithms, and decision making under incomplete information.
- *Reconfigurablity and Transferability.* The remote system is a complex system involving intensive hardware and software development. The systems developed for different applications should share common modules. Furthermore, computer/Internet technology and diagnosis methodologies are advancing rapidly, and the developed system should be able to incorporate advancements

easily without major modifications. Thus, reconfigurability and transferability are important in all aspects of system development.

## CONCLUSION

Tomorrow's manufacturing industries must cope with the life-cycle support of the products they produce in a globalized enterprise. Information technology will play an indispensable role in supporting and enabling the complex practices of design and manufacturing by providing the mechanisms to facilitate and manage the integrated system discipline. The need for improved understanding of a product's performance for after-market service will lead to digital service technology for reasons other than troubleshooting. Manufacturers and users will benefit from the increased equipment and process reliability that digital service technology offers. In addition, manufacturing suppliers will benefit from the opportunity to sell process solutions, not just equipment, to their customers.

## REFERENCES

Goncharenko, I., K. Mori, and J. Lee. 1997. Real-time monitoring of machining processes via the internet: VR modeling approach. Proceedings of the International Conference on Manufacturing Milestones toward the 21st Century. Tokyo: Japan Society of Mechanical Engineers.

Lee, J. 1992. In-Process Machine Degradation Monitoring and Fault Detection Using a Neural Networks Approach. Ph.D. dissertation. Washington, D.C.: George Washington University.

Lee, J. 1995a. Machine performance monitoring and proactive maintenance in computer integrated manufacturing: Review and perspective. International Journal for Computer Integrated Manufacturing 8(5):370–380.

Lee, J. 1995b. Modern computer-aided maintenance of manufacturing equipment and systems: Review and perspective. Computers and Industrial Engineering 28(4):793–811.

Lee, J. 1996a. Manufacturing Study in Japan. NSF Report. Arlington, Va.: National Science Foundation.

Lee, J. 1996b. Measurement of machine performance degradation using a neural networks model. Computers in Industry 30(3):193–209.

Lee, J. 1996c. Strategy on globalization support remote manufacturing system. Paper presented at the International Conference on Advances in Production Management Systems (APMS '96), Kyoto, Japan, November 4–6, 1996.

Lee, J., and B. M. Kramer. 1993. Analysis of machine degradation using a neural network based pattern discrimination model. Journal of Manufacturing Systems 12(5):379–387.

National Research Council. 1995. Information Technology for Manufacturing: A Research Agenda. Computer Science and Telecommunications Board and Manufacturing Studies Board. Washington, D.C.: National Academy Press.

Shi, J., J. Ni, and J. Lee. 1997. Research challenges and opportunities in remote diagnostics and system performance assessment. Intelligent Manufacturing Systems (IMS '97): A Proceedings Volume from the 4th IFAC Workshop, J. Kim, ed. New York: Pergamon.

# ROBOTICS

# A Brief History of Robotics

KENNETH Y. GOLDBERG
*Department of Industrial Engineering and Operations Research*
*University of California*
*Berkeley, California*

Almost all cultures have myths about artificial creatures: the Homunculus, Golem, Sorcerer's Apprentice, and, of course, Frankenstein. One theme that runs through these myths is that things generally turn out rather badly for the mortals who attempt to play creator.

In 1923 the playwright Karel Capek coined the word "robot"—"worker" in Czech. Industrialization was in full swing; Capek's extrapolation of this trend, combined with the development of the electronic computer in the 1940s, fired the imagination of many science fiction writers. Robots were the focus of Asimov's *I Robot* series and were featured in many popular films (*2001*) and television shows (*Lost in Space*).

What is a robot? The definition is surprisingly controversial, even among roboticists. At one extreme are humanoids, friendly or unfriendly, with anthropomorphic features. At the other extreme are the repetition-loving mechanical arms of industrial automation. The former are highly flexible, the latter highly efficient. Roboticists such as Whitney (1986) suggest that there is an inherent design tradeoff between flexibility and efficiency: a humanoid household robot would not be nearly as efficient at the standard home dishwashing machine. Neither is it obvious that anthropomorphism is necessary: machines with flapping wings were far less successful at flying than fixed-wing aircraft.

The International Standards Organization defines an industrial robot as "an automatically controlled, reprogrammable, multipurpose, manipulator with three or more axes" (ISO 8373), a reasonable definition that excludes dishwashers and most talking dolls. The first U.S. patent for a robot that falls under this definition was granted to George Devol in 1956. In the 1960s, research and development on numerical control (NC) machines for milling and lathing grew into a large commercial industry.

In the early 1980s, around the time the film *Star Wars* came out, robots became America's media darlings. They were suddenly friendly and desirable. Intoxicating predictions were made about robots leading to the end of human labor. Many companies invested in robots but faced the challenge of making them work reliably. Unrealistic expectations led to disillusionment and the failure of many robot companies at the end of that decade.

Today, there are almost 700,000 robots at work in industry. Approximately 80,000 robots were sold throughout the world in 1996, accounting for gross sales of over $5 billion. Almost half of those robots were installed in Japan, about 10,000 were installed in the United States and Germany, respectively, and the remaining 20,000 were installed in Korea, Italy, France, and other countries. By far the largest application areas are welding and painting, followed by machining and assembly. The largest customers are the automotive industry, followed by electronics, food, and pharmaceuticals (United Nations/Economic Commission for Europe and International Federation of Robotics, 1997).

Throughout this turbulence robotics continues to be an active and thriving area of research. Some of its subareas are kinematics (the study of positions and velocities), dynamics (the study of forces), and motion planning (how to get an object from here to there while avoiding obstacles). The so-called "piano movers" problem was solved in the 1980s when a breakthrough showed that it could be reduced to a well-known problem of deciding the truth value of algebraic sentences (Latombe, 1991). Other areas include grasping, locomotion, actuator design (the direct-drive robot arm was another breakthrough in the 1980s), and sensor design (a reliable tactile sensor is still being pursued).

Robotics is highly interdisciplinary, including specialists from fields such as mechanical engineering, computer science, electrical engineering, and industrial engineering. Research is sponsored in this country by the National Science Foundation, the Defense Advanced Research Projects Agency, the U.S. Department of Energy, the National Aeronautics and Space Administration, and industry. Similar sponsors are found for research in Japan, Australia, and Europe. Almost all universities have research groups working in robotics. The largest international research organization is the Institute of Electrical and Electronics Engineers Society of Robotics and Automation, founded in 1984.

There are many frontiers of robotics. Theories from areas such as nonlinear control, Lie algebra, computational geometry, and computational algebra are being applied to such topics as medical and surgical robots, microscale robots, Internet robots, modular robots, and robot toys for education and entertainment. It is impossible to cover all of these frontiers in one session. For the Frontiers symposium I worked closely with Susan Corwin of Intel and Rob Howe of Harvard to select the following four roboticists to represent a cross section of the best new work in our field.

## REFERENCES

Latombe, J.-C. 1991. Robot Motion Planning. Boston: Kluver Academic Publishers.

United Nations/Economic Commission for Europe and International Federation of Robotics. 1997. World Industrial Robots 1997— Statistics, Analysis, and Forecasts to 2000. New York: United Nations.

Whitney, D. E. 1986. Real robots don't need jigs. Proceedings of the IEEE International Conference on Robotics and Automation. Washington, D.C.: IEEE Computer Society Press.

# Algorithms in Robotics: The Motion Planning Perspective

LYDIA E. KAVRAKI
*Department of Computer Science*
*Rice University*
*Houston, Texas*

The development of robotics depends not only on the advancement of robotic technology and the building of novel robotic systems but also on improved understanding and design of robot algorithms. Like computer algorithms, robot algorithms are abstract descriptions of processes in the physical world whose development and analysis are, to some extent, independent of a particular implementation technology. But robot algorithms also differ in significant ways from computer algorithms. While the latter have full control over the data to which they apply, robot algorithms deal with physical objects that they attempt to control despite the fact that these objects are subject to the independent and imperfectly modeled laws of nature. This leads robot algorithms to blend in a unique way basic control issues (controllability and observability) and computational issues (calculability and complexity). They pose in turn fundamental questions that require a new set of analytical concepts and tools for the evaluation of their performance (Latombe, 1994).

Among robot algorithms, those that deal with the planning of continuous motion are central to robotics. Robots accomplish tasks by moving objects (including themselves) in the real world, and they need to reason about the continuous geometry and physics of their environment. The generation of motion clearly poses both control and computational problems. A number of key results have influenced recent work. In particular, the computational complexity of several motion planning problems was established by using theoretical computer science methodology (Reif, 1979). The concept of the configuration space permitted discussion of motion in a unified framework (Lozano-Pérez, 1983). Under this framework, the problem of reasoning about a robot in its original environment is transformed to the problem of reasoning about a point in a new space that is called the configuration space and whose dimension equals the number of degrees of freedom of the robot. Critically based decomposition techniques first addressed the problem of computing with continuous geometric data by partitioning the space into regions such that some pertinent property remains invariant over each region (Schwartz and Sharir, 1983). Moving away from purely geometric issues, the subtle relations between state reachability and state recognizability were studied (Lozano-Pérez et al., 1984), and the role of task mechanics in the generation of motion was emphasized (Mason, 1986). As with robot algorithms in general, the amount of physical resources required (number of hands, beacons, etc.) defines the physical complexity of the algorithm. Understanding the intrinsic physical complexity of motion tasks has aided the design of reliable algorithms that induce minimal engineering costs (Canny and Goldberg, 1994; Donald, 1993).

The following focuses on the computational issues of motion planning by examining the basic problem of planning a continuous geometric path between two free configurations of a robot. The geometry of the robot, the kinematics of the robot, and the geometry of the environment are completely known. The path planning problem is a central research area in robotics and several variants of the problem have been explored (Latombe, 1991). These include parts manipulation, assembly sequencing, multirobot coordination, on-line execution, controllability, planning with incomplete knowledge or uncertainty in control and sensing, and others. Currently, the application of planning techniques extends well beyond the realm of traditional robotics to applications in medicine (robot-assisted surgery), computational biology and computational chemistry (computer-assisted pharmaceutical drug design), and graphics animation (digital actors).

Complete algorithms are known for the path planning problem. However, the fastest among them is exponential in the number of degrees of freedom of the robot (Canny, 1988), and it is unrealistic, from a computational point of view, to apply such an algorithm to robots with more than two or three degrees of freedom. Nevertheless, remarkable progress has been made in practical path planning over the past few years. Planners with weaker performance guarantees (probabilistically complete planners) have been developed for robots with many degrees of freedom with increasingly improved performance. An important milestone was the introduction of randomization techniques for solving high-dimensional problems (Barraquand and Latombe, 1991). Recently, a unified randomized framework has dealt successfully with 5 to 25 degrees-of-freedom robots (Kavraki and Latombe, 1994). The analysis of this framework, called the probabilistic roadmap approach to path planning, demonstrates how the geometric properties of the underlying space (other than the number of vertices, edges, faces, etc., of the objects in the environment) can be exploited to yield a planner with performance guarantees.

The probabilistic roadmap approach proceeds as follows. During a preprocessing phase a roadmap is constructed. This is done by first generating random free configurations of the robot and then connecting these configurations using a simple, but very fast, motion planner. The roadmap thus formed in the free configuration space of the robot is stored as an undirected graph. The configurations are the nodes of the graph, and the paths computed by the local planner are the graph edges. A subsequent processing of the graph attempts to increase its connectivity again by using probabilistic techniques. Following the preprocessing phase, multiple path planning queries can be answered. A query asks for a path between two free configurations of the robot. To answer a query, the planner first attempts to find a path from the start and goal configurations to two nodes of the roadmap. Next, a graph search is performed to find a sequence of edges connecting these nodes in the roadmap. Concatenation of the successive path segments transforms the sequence found into a feasible path for the robot.

Importantly, the performance of the method can be related to basic geometric properties of the space in which the problem is solved (Kavraki et al., 1995). The evaluation shows that in spaces that are ε-good (in these spaces each point "sees" an ε fraction of the whole free space), a roadmap with polynomial in ε number of nodes captures the connectivity of the free space with high probability. This allows path planning queries to be answered efficiently and reliably once such a roadmap has been constructed. The analysis explains the good experimental performance of the planner in realistic settings and introduces a new methodology for evaluating the performance of planning algorithms.

Recent work is extending the probabilistic framework above to planning for multiple robots, planning for nonholonomic robots (robots with no-slip constraints, such as carllike robots), planning in environments with moving objects, and planning for flexible objects. Introducing physical properties in geometric settings (as done in the case of flexible objects; Kavraki et al., 1998) gives rise to even more difficult discretization and sampling issues. Furthermore, the probabilistic framework is used to plan paths for drug molecules to their docking sites in proteins (A. Singh, J.-C. Latombe, and D. Brutlag, Stanford University, personal communication, 1998). Interestingly, work in robotics is finding many applications in the area of computer-assisted pharmaceutical drug design (Kavraki, 1997).

Understanding of the intrinsic geometric properties of motion tasks and the development of efficient techniques that rely on these properties is still at an early stage. It is now becoming necessary to study these issues in depth and develop a computational framework that can directly relate the difficulty of tasks to their geometric and physical characteristics. The goal is to allow, with minimal effort, the tailoring of robot planning algorithms to a variety of tasks—from specialized operations in extremely constrained environments to everyday life activities.

## REFERENCES

Barraquand, J., and J.-C. Latombe. 1991. Robot motion planning: A distributed representation approach. International Journal of Robotics Research 10(6):628–649.

Canny, J. 1988. The Complexity of Robot Motion Planning. Cambridge, Mass.: MIT Press.

Canny, J., and K. Y. Goldberg. 1994. RISC for industrial robotics: Recent results and open problems. Pp. 1951–1958 in Proceedings of the IEEE International Conference on Robotics and Automation. Washington, D.C.: IEEE Computer Society Press.

Donald, B. 1993. Information invariants in robotics. Pp. 276–283 in Proceedings of the IEEE International Conference on Robotics and Automation. Washington, D.C.: IEEE Computer Society Press.

Kavraki, L. 1997. Geometry and the discovery of new ligands. Pp. 435–448 in Algorithms for Robotic Motion and Manipulation, J.-P. Laumond and M. Overmars, eds. Wellesley, Mass.: AK Peters.

Kavraki, L., and J.-C. Latombe. 1994. Randomized preprocessing of configuration space for fast path planning. Pp. 2138–2145 in Proceedings of the IEEE International Conference on Robotics and Automation. Washington, D.C.: IEEE Computer Society Press.

Kavraki, L., J.-C. Latombe, R. Motwani, and P. Raghavan. 1995. Randomized query processing in robot motion planning. Pp. 353–362 in Proceedings of the 27th Annual ACM Symposium on the Theory of Computing. New York, N.Y.: Association for Computing Machinery.

Kavraki, L., F. Lamiraux, and C. Holleman. 1998. Towards planning for elastic objects. Pp. 313–325 in Robotics: The Algorithmic Perspective, P. Agarwal, L. Kavraki, and M. Mason, eds. Wellesley, Mass.: A K Peters.

Latombe, J.-C. 1991. Robot Motion Planning. Boston, Mass.: Kluwer.

Latombe, J.-C. 1994. Robot algorithms. Pp. 1–8 in Algorithmic Foundations of Robotics, K. Y. Goldberg, D. Halperin, J.-C. Latombe, and R. H. Wilson, eds. Wellesley, Mass.: A K Peters.

Lozano-Pérez, T. 1983. Spatial planning: A configuration space approach. IEEE Transactions on Computers C-32(2):108–120.

Lozano-Pérez, T., M. Mason, and R. Taylor. 1984. Automatic synthesis of fine-motion strategies for robots. International Journal of Robotics Research 3(1):3–24.

Mason, M. 1986. Mechanics and planning of manipulator pushing operations. International Journal of Robotics Research 5(3):53–71.

Reif, J. 1979. Complexity of the mover's problem and generalizations. Pp. 421–427 in Proceedings of the 20th Annual Symposium on Foundations of Computer Science. New York, N.Y.: Institute of Electrical and Electronics Engineers.

Schwartz, J., and M. Sharir. 1983. On the "piano movers" problem. II. General techniques for computing topological properties of real algebraic manifolds. Advances in Applied Mathematics 4(3):298–351.

# Mechanics, Control, and Applications of Biomimetic Robotic Locomotion

JOEL W. BURDICK
*Department of Mechanical Engineering*
*California Institute of Technology*
*Pasadena, California*

## INTRODUCTION

Since mobility can be an essential requirement for the operation of many autonomous systems, robotic locomotion has been actively studied for over three decades. Most mobile robots use wheels, since they provide the simplest means for mobility. However, some terrains are inaccessible to wheeled vehicles, and wheels are undesirable for a number of applications. Biomimetic locomotion refers to the movement of robotic mechanisms in ways that are analogous to the patterns of movement found in nature. Biomimetic robotic locomotors do not rely on wheels, tracks, jets, thrusters, or propellers for their propulsion.

Despite more than three decades of research effort, biomimetic robotic locomotors have largely remained laboratory curiosities. However, promising applications of biomimetic mobility still motivate research in this area. The potential for robots to assist the elderly in their homes has recently motivated the Honda Motor Co. (1998) to invest heavily in bipedal walking research, since anthropomorphic robots can adapt better to preexisting environments that were designed for humans. In the 1980s, the Defense Advanced Research Projects Agency provided significant support for Ohio State University's Adaptive Suspension Vehicle, a six-legged, 7,000-pound locomotor that was intended to support combat operations in complex terrain (Song and Waldron, 1989). Nonlegged locomotory machines might find significant applications as well. For example, efforts are under way at Caltech to develop a “snakelike” robotic endoscope that would enable minimally invasive access to the human small bowel system, which is currently inaccessible by conventional endoscopes. Snakelike robots have also been investigated for use in urban search-and-rescue operations following earthquakes or other natural disasters. Biomimetic fluid propulsion based on changes in the mechanism's shape offers an alternative to traditional underwater vehicle propulsion based on propellers and control surfaces. Such fluid propulsion may be very maneuverable at many size scales and is free from the motor noise, vibrations, and propeller cavitation associated with propellers.

## CURRENT STATUS

Quasi-static legged locomotion (where the locomotor's center of mass is always supported by at least three legs in ground contact) has been the most extensively studied type of biomimetic locomotion, and many four- and six-legged robots have been successfully demonstrated (Song and Waldron, 1989). Beginning with Raibert (1986), legged hopping robots have received considerable experimental and analytical attention (M'Closkey and Burdick, 1993). Bipedal walking and running also has been an active area of study (McGeer, 1990; Honda Motor Co., 1998). "Snakelike" robots can potentially enter environments that are inaccessible to legged or wheeled vehicles. Significant work in snakelike locomotion was initiated in the 1970s by Hirose and Umetani (1976) and more recently by Chirikjian and Burdick (1995). Realistic efforts to develop fishlike robots have emerged only recently (Barrett, 1996; Kelly et al., 1998).

## RESEARCH NEEDS AND OBJECTIVES

There are many issues that limit widespread deployment of biomimetic locomotors, including limitations in actuation technology, onboard power-carrying capability, and sensing. While all of these issues merit serious attention, this paper briefly discusses the associated limitations in theory. In an attempt to derive useful results for specific examples, prior biomimetic locomotion studies have generally focused on a particular robot morphology (such as a biped or quadruped). Unfortunately, results derived for one morphology typically do not extend to other morphologies. To enable future widespread deployment of cheap and robust robotic locomotion platforms, we must ultimately seek a more unifying and comprehensive framework for biomimetic robotic locomotion engineering. This framework should have the following properties:

- the analysis, design, and control methodologies can be uniformly applied to a broad class of locomotory problems;
- significant aspects of the framework can be encoded in automated software tools; and
- the underlying methods are sufficiently rigorous to predict and enable robust system performance.

Realization of such a framework would enable more widespread deployment of effective biomimetic locomotors. One strategy being pursued at Caltech to realize this framework is to (1) establish general forms for the equations of motion of locomotion systems, (2) develop a control theory for this class of nonlinear equations, (3) abstract motion planning and feedback control algorithms from the control theory, and (4) develop paradigms (perhaps rules of thumb) for designing systems for specific applications. Caltech's research program, whose underlying concepts are briefly sketched in the next section, has reached a good understanding of the underlying mechanics principles. The associated control theory and algorithms are currently in development.

## PRINCIPLES OF BIOMIMETIC PROPULSION

Biomimetic propulsion is typically generated by a coupling of periodic mechanism deformations to external constraints (i.e., mechanical interactions with the environment). The forces generated by these constraint interactions (e.g., pushing, rolling, sliding) induce net robot movement. The creeping, sidewinding, and undulatory gaits of snakes rely on no-slip, or nonholonomic, constraints. Slug and snail movement depends on the viscous fluid constraint of slime trails, while amoebae and paramecia move via a constraint between their surfaces and the surrounding fluid. Fish use a variety of fluid mechanical constraint principles. Surprisingly, common principles underlie the mechanics and control of these seemingly different systems (Ostrowski and Burdick, 1998).

The language of geometric mechanics has proven to be a useful way to precisely phrase these intuitive notions. Two simple concepts motivated by geometric thinking have proven useful in developing a comprehensive basis for the mechanics and control of biomimetic locomotion. The first key observation is that it is always possible to divide a locomoting robot's configuration variables into two classes. The first class of variables describes the position of the robot—that is, the displacement of a robot fixed coordinate frame with respect to a fixed reference frame. The set of frame displacements is $SE(m)$, $m \leq 3$, or one of its subgroups—that is, a Lie group. The second set of variables defines the mechanism's internal configuration or shape. The set of all possible shapes (the "shape space") is a manifold, $M$. The Lie group, $G$, together with the shape space, $M$, form the total configuration space of the system, denoted by $Q = G \times M$. The configuration space of both terrestrial and aquatic biomimetic locomotors is a trivial principal fiber bundle.

The importance of the principal fiber bundle structure of the configuration space of locomoting systems is related to the following facts. The shape and position variables are coupled by the constraints acting on the robot. By making changes in the shape variables, it is possible to effect changes in the position variables through the constraints. A central goal of locomotion analysis is the systematic derivation of an expression that answers the question: If I wiggle the body, how far does the mechanism locomote? Formally, this all-important relationship between shape changes and position changes can be described via a connection, which is an intrinsic geometric feature of principal fiber bundles. Recent Caltech efforts have shown that there is a systematic way to derive the connection for a very large class of locomotion problems. The connection provides not only a unified way of thinking about mechanics but also one about motion planning and control (Radford and Burdick, 1998).

The second key idea is the use of symmetries in locomotion analysis. Because a locomotor is a mechanical system, we can assume there exists a Lagrangian and a set of constraint equations that describe the interaction principle underlying the given locomotion scheme. A symmetry corresponds to a group of transformations that leave the Lagrangian (and possibly the constraints) invariant. In the absence of constraints, these symmetries correspond to the well-known principles of momentum conservation. Unfortunately, conservation laws are not necessarily preserved in the presence of most constraints, which are essential to locomotion. However, it is possible to extend classical theory (see Bloch et al., 1996, for the case of nonholonomic constraints) to develop a generalized momentum equation that describes the evolution of the momenta due to the interaction constraints. The mixture of symmetry and interaction constraints can give rise to the ability to increase or control momentum via the action of internal forces. This is an extremely important effect in generating biomimetic locomotion. The aforementioned connection and the use of invariance principles (and their associated momenta equations) yield a comprehensive framework for the fundamental analysis of biomimetic locomotion mechanics and control.

## REFERENCES

Barrett, D. S. 1996. Propulsive Efficiency of a Flexible Hull Undersea Vehicle. Ph.D. dissertation. Massachusetts Institute of Technology, Cambridge.

Bloch, M., P. S. Krishnaprasad, J. E. Marsden, and R. M. Murray. 1996. Nonholonomic mechanical systems with symmetry. Archive for Rational Mechanics and Analysis 136(l):21-99.

Chirikjian, G. S., and J. W. Burdick. 1995. The kinematics of hyper-redundant robot locomotion. IEEE Transactions on Robotics and Automation 11(6):781-793.

Hirose, S., and Y. Umetani. 1976. Kinematic control of active cord mechanism with tactile sensors. Pp. 241–252 in Proceedings of the Second International CISM-IFT Symposium on Theory and Practice of Robots and Manipulators, A. Morecki and K. Kñedzior, eds. New York: Elsevier Scientific Publishers.

Honda Motor Co., Ltd. 1998. Humanoid robot. [Online]. Available: http://www.honda.co.jp/home/hpr/e_news/robot/index.html.

Kelly, S., R. Mason, C. Anhalt, R. Murray, and J. Burdick. 1998. Modeling and experimental investigation of carangiform locomotion for control. Paper presented at the American Control Conference, Philadelphia, Pa., June 24–26, 1998.

McGeer, T. 1990. Passive dynamic walking. International Journal of Robotics Research 9(2):62–82.

M'Closkey, R. T., and J. W. Burdick. 1993. On the periodic motions of a hopping robot with vertical and forward motion. International Journal of Robotics Research 12(3):197-218.

Ostrowski, J. P., and J. W. Burdick. 1998. The mechanics of undulatory robotic locomotion. International Journal of Robotics Research 17(7):683-701.

Radford, J., and J. Burdick. 1998. Local motion planning for nonholonomic control systems evolving on principal bundles. Paper presented at the International Symposium on Mathematical Theory of Networks and Systems (MTNS '98), Padova, Italy, July 6–11, 1998.

Raibert, M. H. 1986. Legged Robots that Balance. Cambridge, Mass.: MIT Press.

Song, S. M., and K. J. Waldron. 1989. Machines that Walk: The Adaptive Suspension Vehicle. Cambridge, Mass.: MIT Press.

# Robotic Perception for Autonomous Navigation of Mars Rovers

LARRY H. MATTHIES
*Jet Propulsion Laboratory*
*Pasadena, California*

The U.S. government has been working on autonomous navigation of robotic vehicles since the 1960s, when the Jet Propulsion Laboratory (JPL) developed a prototype lunar rover for the Surveyor program. The earliest U.S. and Russian rovers were essentially teleoperated, which was acceptable for the few seconds of communications delay between earth and the moon. In the 1970s, JPL began working on rovers for Mars, where communication latency of up to 40 minutes, owing to roundtrip light time, required a higher degree of autonomy on the rover if exploration was to be efficient (O'Handley, 1973). Key technical barriers were relatively poor three-dimensional (3-D) sensing of the environment, lack of accurate means to keep track of the position of the rover as it traveled, and the low performance of onboard computers. This paper surveys key progress in sensors, algorithms, and processors that has alleviated these specific barriers, thereby enabling practical autonomous navigation.

## BACKGROUND

Research on robotic vehicles for terrestrial applications accelerated in the late 1970s and throughout the 1980s. The Autonomous Land Vehicle (ALV) Program, funded by the Defense Advanced Research Projects Agency (DARPA) from 1984 to 1988, developed a vehicle the size of a large bread truck that used a scanning laser rangefinder to locate and avoid obstacles while driving cross-country at speeds on the order of 2 mph (Olin and Tseng, 1991). Progress in this area was retarded by the size, cost, and power consumption of the laser rangefinder and by the size and cost of the onboard computers. Meanwhile, in the more structured environment of autonomous lane-following on highways, researchers in Germany were able to achieve driving speeds of up to 60 miles an hour for a Mercedes van that used onboard cameras and computers to track lane markings (Dickmanns and Mysliwetz, 1992). This application was helped by the strong geometric constraints available from knowledge of highway design and by the relatively low amount of computation required to track lane markings in imagery. Nevertheless, this effort made a significant contribution to the field by demonstrating effective application of system dynamics models and Kalman filtering in a computer vision system for autonomous navigation.

## REAL-TIME 3-D PERCEPTION FOR MARS ROVERS

A key breakthrough in 3-D perception for autonomous cross-country navigation came in 1990, when JPL developed efficient and reliable algorithms for estimating elevation maps in real time from stereo image pairs using compact commercial processors onboard a prototype Mars rover (Matthies, 1992). Three-dimensional sensing with stereo cameras has advantages over laser scanners for Mars rover applications because it is easier to make the sensor mechanically robust enough to survive the rigors of launch and landing. This stereo vision approach computes range by triangulation, given matching features in two images. The prevailing approach to stereo matching in the computer vision community at that time, popularized by David Marr (1982) at the Massachusetts Institute of Technology, was to first extract edges from both images and then match the edges found in the left and right images. The edge detection process reduces the amount of information to be processed in the matching stage, which is what made this approach appear attractive; however, by definition it produces "sparse" range data because it measures range only where prominent edges are found in the imagery. For Mars rover applications this would mean that range data would only be found around the outlines of high-contrast rocks. However, to do effective traverse planning, it was preferable to sample the elevation of the terrain more densely, so as to be guaranteed of finding small obstacles that might trap a wheel.

The heart of JPL's innovation was to develop an efficient stereo matching algorithm based on area correlation, which was able to produce reliable range measurements at almost every pixel in the image for applications like Mars rovers. Most components of this algorithm had been developed previously for different problems; hence, the key contribution was to recognize how to put them together to produce a fast and reliable solution to this problem. The first step in this algorithm is a process called rectification, which resamples both images in such a way that corresponding image features lie on corresponding scanlines in the two images. This reduces the search for matching features to a 1-D search along the corresponding scanlines. The second step applies a band-pass filter to each image to compensate for overall differences in brightness between images from the two cameras. The third step finds corresponding features by evaluating a least squares similarity criterion for a small image patch from the left image at several trial match positions along the scanline in the right image. This evaluation is performed for each pixel in the left image to produce a range estimate at each pixel; the large number of correlations to perform makes this the bottleneck step in the whole algorithm. In straightforward implementations the number of arithmetic operations needed to evaluate the least squares criterion is $3N$, not including addressing operations, where $N$ is the number of pixels in the image patch used for matching; $N$ is typically around 50 (i.e., a 7×7 patch). However, there exists an incremental technique that maintains intermediate results to reduce the cost to six operations per trial match position, independent of the size of the image patch used for matching. This was key to making the entire process practical for real-time implementation onboard a robotic vehicle.

The first incarnation of a stereo vision system using this algorithm required nearly 10 processor boards occupying about 1 cubic foot of space and using 100 to 200 watts of power. This system produced about 1,000 range measurements per second; this enabled the JPL Mars rover prototype "Robby" to drive autonomously over a 100-meter cross-country course in four hours in September 1990 (Matthies, 1992)—a significant "first" for a robotic vehicle.

## FURTHER DEVELOPMENT OF REAL-TIME 3-D PERCEPTION

This technology was picked up by DARPA and the U.S. Army for use in research programs aimed at developing unmanned ground vehicles for military reconnaissance applications (Mettala, 1992). By 1996 the speed of stereo vision systems had increased to about 30,000 range measurements per second, with faster computers that occupied slightly less space. This enabled semiautonomous High Mobility Multipurpose Wheeled Vehicles (HMMWVs) to execute rudimentary reconnaissance missions covering about 2 miles of open terrain at speeds up to 5 mph (Matthies et al., 1996).

Since 1996, the advent of general-purpose microprocessors with limited vector processing capability has enabled substantial speed improvements and size reductions for stereo vision systems. The MMX feature in the Intel Pentium is the best-known example of such a capability; it allows up to four 16-bit integer operations to be performed in parallel. Stereo algorithms implemented on these processors can now perform about 700,000 range measurements per second (Konolige, 1998). Lower-performance systems based on the same algorithm but different CPUs have been built on circuit cards that fit in the palm of the hand, including the CPU and both cameras. This highlights another recent development that is enabling compact, low-cost computer vision systems for robotic vehicles: the introduction of low-power complementary metal-oxide silicon (CMOS) imagers with clocking, control, and analog-to-digital conversion functions fully integrated onchip. Forthcoming advances in vector processing for general-purpose CPUs will accelerate these trends; for example, Motorola has announced a vector-processing extension to the PowerPC architecture that will allow 16 byte-wide arithmetic operations to proceed in parallel. Within a year this will enable stereo vision systems to produce on the order of 2 million range measurements per second with a single microprocessor—that is, 256×256 range measurements per frame at full video rate (30 frames per second). Compared to 1990, this is a speed increase of three orders of magnitude with a simultaneous reduction in size and power dissipation of one order of magnitude.

These advances are enabling a suite of new applications of robotic vehicles. In 1999 a robotic vehicle carrying stereo cameras is to enter Chernobyl to attempt to create a 3-D model of the interior to facilitate further cleanup efforts. JPL's stereo algorithm will help enable the next U.S. Mars rover, currently scheduled for launch in 2003, to explore several kilometers, in comparison with the 100 meters or so covered by the Sojourner rover in the summer of 1997. DARPA and the U.S. Army are also continuing to use this technology for further development of military robotic vehicles. For example, the Demo III program managed by the Army Research Laboratory aims to enable autonomous cross-country navigation at 20 mph by 2001, for a robotic vehicle the size of a large desk. The DARPA Tactical Mobile Robot Program is currently funding development of robotic vehicles the size of a large briefcase for reconnaissance applications in urban warfare. Both of these programs will employ stereo vision among their sensor suites for autonomous navigation and will depend on the aforementioned advances in algorithms, low-power CMOS imagers, and high-performance embedded CPUs to provide the increased speed and smaller size required.

## LIMITATIONS AND APPROACHES TO SOLUTIONS

These advances have produced a viable solution to real-time 3-D perception for robotic vehicles operating during the day in barren or semiarid terrain. Limitations that arise as we push for broader applicability include the following:

- For military applications, operability at night is essential. It appears that stereo vision with thermal infrared imagery works quite well, although thermal cameras are currently very expensive. Two-axis scanning laser rangefinders work well at night but are also still large and expensive.
- Stereo vision fails in textureless environments, such as painted walls in indoor mobile robot applications. This can be solved by adding low-cost, low-resolution active sensors (e.g., sonar), or compact single-axis scanning laser rangefinders to sense the floorplan of a room.
- For terrestrial applications, robotic vehicles need to perceive both the 3-D geometry and the composition of the terrain (e.g., to discriminate traversable vegetation from nontraversable rocks). For some basic discriminations, viable solutions are in hand; in particular, live vegetation is easily distinguished from

soil and rocks using visible and near-infrared imagery (Matthies et al., 1996). Other discriminations are still poorly solved, especially for real-time applications, such as distinguishing dead vegetation from soil. For some of these cases we are studying the use of image texture for terrain classification.

- In addition to obstacle detection, position estimation is a major part of the autonomous navigation problem. Although the global positioning system (GPS) can largely solve this problem outdoors on Earth, it is still an important problem for indoor robot applications and in planetary exploration. Visual feature tracking with stereo cameras has been employed successfully for robot motion estimation and for terminal guidance to human-designated objectives (Matthies and Shafer, 1987; Wettergreen et al., 1997). A number of methods are under development that use maps together with images and other sensors for various forms of landmark recognition (Cozman and Krotkov, 1996; Lu and Milios, 1997; Matthies et al., 1997). Some of these methods are fairly mature for Earth applications now; methods suitable for Mars rovers will likely come to maturity and be deployed over the next five to seven years.
- In terrestrial applications, autonomous navigation among other moving objects requires a significant extension of perception, planning, and local world modeling capabilities beyond that addressed above. Much work is in progress on this problem, using sonar, scanning-laser rangefinders, and imagery.

## PREDICTIONS

Robotic perception systems for Mars rovers should enable autonomous navigation up to a kilometer or more from the lander by the year 2003. Laser rangefinders and image-based feature tracking and landmark recognition algorithms are expected to be used for autonomous precision landing on a comet in less than 10 years from now. Within 10 years it is also possible that unmanned ground vehicles will be sufficiently mature to proceed with fielding them for selected military applications. The cost of such systems may also be low enough, and the capability high enough, to support commercialization for some civil applications. Potential commercial applications of such technology include autonomous material transportation and collision avoidance sensors for smart passenger vehicles. Applications for computer vision technology also exist outside autonomous navigation, such as in PC-based camera systems that use 3-D shape and motion-sensing capabilities to enhance video conferencing and human-computer interfaces. In fact, commercial imaging applications are part of what is driving the rapid progress in low-power CMOS imagers, low-power embedded processors, and vector processing extensions to general-purpose microprocessor architectures. These advances will lead to computer vision systems in the next 10 years that will make robotic vehicles and the vision systems themselves cost-effective for new applications. Finding and exploiting these markets will be an exciting opportunity for engineers in the next decade.

## ACKNOWLEDGMENTS

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, with support from the National Aeronautics and Space Administration, Defense Advanced Research Projects Agency, and the Joint Robotics Program of the Office of the Secretary of Defense.

## REFERENCES

Cozman, F., and E. Krotkov. 1996. Position estimation from outdoor visual landmarks for teleoperation of lunar rovers. Proceedings of the Third IEEE Workshop on Applications of Computer Vision. Los Alamitos, Calif.: IEEE Computer Society Press.

Dickmanns, E. D., and B. Mysliwetz. 1992. Recursive 3-D road and relative ego-state recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2):199–213.

Konolige, K. 1998. Small vision systems: Hardware and implementation. Proceedings of the Eighth International Symposium on Robotics Research, Y. Shohira and S. Hirose, eds. New York: Springer-Verlag.

Lu, F., and E. Milios. 1997. Robot pose estimation in unknown environments by matching 2D range scans. Journal of Intelligent and Robotic Systems 18:249–275.

Marr, D. 1982. Vision. San Francisco, Calif.: W. H. Freeman.

Matthies, L. H. 1992. Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation. International Journal of Computer Vision 8(1):71–91.

Matthies, L. H., and S. A. Shafer. 1987. Error modeling in stereo navigation. IEEE Journal of Robotics and Automation RA-3(3):239–248.

Matthies, L., A. Kelly, T. Litwin, and G. Tharp. 1996. Obstacle detection for unmanned ground vehicles: A progress report. Proceedings of the Seventh International Symposium on Robotics Research, G. Giralt and G. Hirzinger, eds. New York: Springer-Verlag.

Matthies, L., C. Olson, G. Tharp, and S. Laubach. 1997. Visual localization methods for Mars rovers using lander, rover, and descent imagery. Proceedings of the International Symposium on Artificial Intelligence, Robotics, and Automation in Space (i-SAIRAS). Washington, D.C.: National Aeronautics and Space Administration.

Mettala, E. G. 1992. The OSD tactical unmanned ground vehicle program. Pp. 159–172 in Proceedings of the DARPA Image Understanding Workshop. San Mateo, Calif.: Morgan Kaufmann Publishers.

O'Handley, D. A. 1973. Scene analysis in support of a Mars rover. Computer Graphics and Image Processing 2:281–297.

Olin, K. E., and D. Y. Tseng. 1991. Autonomous cross-country navigation. IEEE Expert 6(4):16–32.

Wettergreen, D., H. Thomas, and M. Bualat. 1997. Initial results from vision-based control of the Ames Marsokhod rover. Proceedings of the International Conference on Intelligent Robots and Systems (IROS). New York: Institute of Electrical and Electronics Engineers.

# Cobots

MICHAEL A. PESHKIN
*Mechanical Engineering Department*
*Northwestern University*
*Evanston, Illinois*

It is often assumed that the benefit of robots is their strength, speed, or accuracy—qualities that a human operator's "help" can only diminish and that indeed may pose a hazard to him or her. Less recognized is that it may be the robots' interface to computers and information systems that is their primary benefit. Furthermore, even in heavily mechanized environments, people have an important continuing role because of their sensing and dexterity, which cannot be matched or replaced by robots.

Collaborative robots—"cobots"—are a new type of robotic device, intended for direct interaction with a human operator in a shared workspace. Cobots allow a true sharing of control between human and computer. The human operator supplies motive power and exerts forces directly on the payload, while the mechanism of the cobot serves to redirect or "steer" the motion of the payload under computer control. The computer monitors the force (direction as well as magnitude) applied by the operator to the payload. In real time these operator forces can be compared with programmed guiding surfaces, and motion in the direction that the operator pushes can be allowed, disallowed, or redirected. The human operator may be allowed complete freedom of motion of the payload, or in the opposite extreme the payload may be restricted to a single curve through space. Thus, the full versatility of software is available for the production of virtual surfaces and other haptic effects.

At Northwestern, our first application area has been automotive assembly, with the help of General Motors. Unlike body welding and painting, automobile assembly is a worker-intensive process because of the need for human dexterity. Ergonomics issues have become a major concern: even payloads well within human strength limits for lifting can cause significant problems. In response, the materials handling industry has developed a great variety of so-called "assist devices" that provide support against gravity and sometimes guide motion but do so at the expense of much greater inertia, anisotropic response to the operator's applied forces, restriction of motion to few dimensions, and greater possibilities of jamming in assembly operations. Thus, conventional assist devices usually reduce productivity. Their clunkiness frustrates operators, often leading to disuse in practice. Damaging collisions may occur, reducing productivity.

Virtual surfaces, whether enforced by a cobot or a conventionally actuated robot, can solve many of the problems of assist devices, if they can be implemented on an appropriately large scale of workspace size and strength. However, the force magnitudes needed are almost by definition at least those of humans. The speeds, if we are to increase rather than decrease productivity, must be at least those of humans. This would seem to imply a need for large motors with greater-than-human power. Workers are understandably leery of such motors in the context of a general-purpose manipulator, if they are intended to work within its workspace.

Cobots, in contrast, rely on the worker to provide motive power or can give some small powered assistance that requires only small motors. The much greater need for force is that required for changes of direction, sometimes called "inertia management." In cobots this is accomplished by the physical mechanism of the cobot rather than by motors, with a consequent improvement in both safety and smoothness of operation.

Space permits only the briefest explanation of the mechanism of cobots. The simplest cobot has a two-dimensional (planar) workspace and a mechanical heart that is a single wheel (see Figure 1). Most interest lies in extension of the cobot idea to many dimensions of motion and to revolute joints. The latter uses a continuously variable transmission in place of a wheel. Interested readers are referred to http://cobot.com for further information.

The motor in Figure 1 simply steers the wheel. No amount of malevolent steering by the control computer can cause the cobot to move on its own. Only the operator can cause it to move, by applying forces to the handle. A force sensor (top) monitors these user forces.

The unicycle cobot displays two essential behaviors: free mode and virtual surface. Free mode is invoked when the cobot's position in its planar workspace is away from all defined constraint surfaces. The cobot should therefore permit any motion the user attempts to impart. To do this, the steering angle of the wheel is servocontrolled such that user forces perpendicular to the wheel's rolling direction are nulled. The behavior is similar to that of a caster wheel on a rolling item of furniture, although there is no physical caster at all.

When the user brings the cobot's position in the plane to a place where a constraint surface is defined, control of the steering angle changes over to virtual surface mode. The wheel is steered such that its rolling direction becomes tangent to the constraint surface, and this tangency is maintained as the user moves

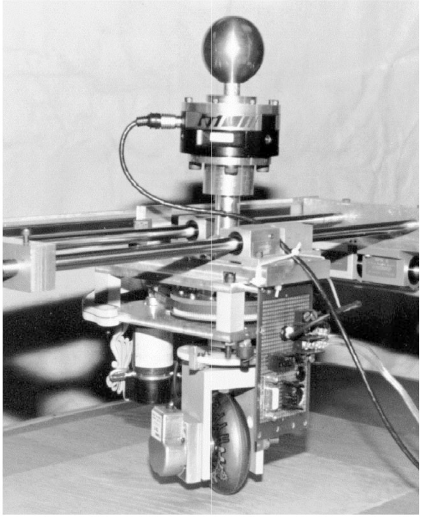


**FIGURE 1** The unicycle cobot is the simplest possible cobot. Constraint surfaces, defined in software, delimit excluded areas of the plane. The Cartesian rails serve only to keep the unicycle upright and are not needed in practical cobots of higher dimensionality. The cobot mechanism consists of a free-rolling wheel in contact with a working surface. The wheel's rolling velocity is monitored by an encoder but is not driven by a motor.

the cobot in "virtual contact" with the constraint surface. The user perceives contact with a hard frictionless constraint surface. In practice the illusion is convincing. Virtual surface mode is ended when the measured user forces are found to be directed away from the constraint surface, at which point free mode resumes.

The unicycle cobot can be generalized to higher dimensions as well as to the revolute architecture characteristic of most industrial robots. The latter rely on a distinct kinematic element in place of the wheel—a continuously variable transmission. Many other haptic effects are also possible: virtual paths and attractive surfaces.

In materials handling applications such as automobile assembly, even the simplest haptic effect—free mode—can be very useful. In free mode the cobot gives the operator the perception that the payload is responding in an unconstrained and natural way to his applied forces. This is actually a simulated lack of constraint, and the existence of a computer in the loop gives an opportunity for many improvements over the natural behavior of the payload—virtual haptic effects. For instance, the lack of isotropy of the underlying kinematic mechanism (e.g., an overhead rail system or an articulated arm) can be masked by the cobot in free mode, so that the payload responds in a more predictable way to the operator's intentions. A prototype railcobot at Ford Motor Company includes this masking effect. Or the inertia of the payload—its reluctance to change its direction of motion—can be masked so that it is perceived as lighter and more maneuverable than it actually is.

Virtual surfaces are another haptic effect. To be useful they must be hard (abrupt), strong (large forces sustained without penetration), and smooth (no friction as a payload is pushed along a virtual surface). Since a cobot's virtual surfaces rely on a physical mechanism rather than on actuators, these desirable qualities are innate. Virtual surfaces are useful for productivity because an operator can push a payload against a virtual surface and "swoop" around a corner quickly and pleasantly.

Virtual surfaces can also be used to prevent collisions in close quarters or assembly operations. A cobot now under test at General Motors assists in removing doors from car bodies after painting and prior to assembly. It confines the door's motion (rotation as well as translation) to a well-chosen curved escape path over a few critical inches as it is removed from the car body, preventing collision of finished surfaces.

While materials handling and automotive assembly in particular have been the first applications area, many others exist as well.

- *Image-guided surgery.* In this area safety is essential, and a cobot's ability to guide motion without possessing a corresponding ability to move on its own can totally remove concern about some failure modes. Perhaps more importantly, the quality of a virtual surface enforced by a cobot originates in its physical mechanism, rather than in servo-controlled actuators, thus yielding harder and smoother surfaces than can be achieved by a robot. Preserving the critical sense of touch in surgery requires high-quality "shared control" between surgeon and robot, for which smoothness of motion is essential.
- *Haptic display.* Computer-aided design (CAD) models of contoured objects (e.g., beverage bottles, car bodies) can be displayed visually, but the feel of these objects cannot be experienced prior to building them. Cobots can display hard, smooth virtual surfaces from CAD models.
- *Rehabilitation and exercise.* Popular weight training equipment, originally designed for rehabilitation, uses shaped cams and other mechanical components to confine a user's limb or body motion to a particular trajectory. While these trajectories are somewhat adjustable, far greater versatility could be achieved if the motion trajectories were encoded in software rather than frozen into the mechanical design of the equipment. Cobots can enforce virtual trajectories with the smoothness, hardness, and safety required for this application.

Many fascinating research areas have been exposed in building and controlling cobots. Many of the research topics that have been explored in robotics suggest new and different questions in the context of cobots. A sampling includes the following:

- *Path planning.* Creating the appropriate motion trajectory for a given task is a current issue in robotics. In cobots the corresponding problem is to create the virtual surfaces that bound and guide the motion of a payload controlled by a human operator, in support of a given task.
- *Haptic effects.* Free mode and virtual surface mode are but two poles of an unlimited range of haptic effects that can be invented. For instance, a virtual surface may have a "penetration strength" beyond which it gives way, or it may have a simulated attractive potential field, or it may yield compliantly to operator pressure against it.
- *High dimensions.* For cobots with workspace dimension greater than two, virtual surfaces can exist with a variety of dimensionalities ("surface" remains the generic term). Describing these surfaces efficiently and usefully is nontrivial.
- *Control.* Novel control issues are created by the essential role of the human operator in the motion of a cobot. For instance, a robot trajectory is a path through space parameterized by time. In cobot control, progress along a path may be entirely at the discretion of the human operator, who may stop or even reverse direction along a path. The utility of time as a parameter is thus greatly reduced, yet control software must maintain the cobot on the path.

## ACKNOWLEDGMENTS

The invention of cobots, by the author and colleague J. Edward Colgate, came about as part of a research project on intelligent assist devices funded by the General Motors Foundation. Further support by the National Science Foundation is gratefully acknowledged. For more information and bibliography references, please begin at http://cobot.com. Intellectual property rights to the cobot are held by Northwestern University.

# DINNER SPEECH

# Technology Innovation and American Leadership

WILLIAM J. PERRY
*Former U.S. Secretary of Defense*
*Professor, Stanford University*
*Stanford, California*

Last year, America's information technology industry earned more than $700 billion—more than the gross domestic product of most nations! The world is truly in the midst of an information technology revolution that is bigger and moving faster than the Industrial Revolution, and the United States is the undisputed leader and major beneficiary of this revolution. As a result, the U.S. economy today is the wonder and envy of the world.

Last year I experienced a powerful example of this economy at work when, after four years in Washington, I returned to Silicon Valley and set out to buy a house. I discovered to my horror that, to buy a house equivalent to the one I had sold four years earlier, I would have to pay 50 percent more. I was thus able once more to practice my time-proven economic principle: "sell low, buy high!"

Earlier this year our local newspaper reported on the economic factors underlying this staggering rise in house prices. During the four years I was gone, Santa Clara County added more than 100,000 new jobs. During the period I was house hunting, jobs were increasing at a rate of 1,000 a week. In the year before I returned, 70 companies in Silicon Valley had their initial public offerings, creating each week four or five new millionaires—each of them wanting to buy a $2 million house! So this is one measure of the extent of technological innovation and its economic impact.

While the entrepreneurs in Silicon Valley were applying their technology to building our nation's wealth, I was back in Washington applying that same technology to building our nation's security. For decades our security strategy has depended on the full use of our technological leadership, especially our leadership in information and aerospace technologies. Today I will talk about the role of technological innovation in two different but complementary tasks: building our nation's wealth and building our nation's security. Both endeavors depend on America sustaining its world leadership in technology, a condition that many Americans take for granted. But this leadership is not a divine right; indeed, it has not always been true, and it is not certain that it will be true in the future. So it is worth asking: How did we achieve this position of leadership—and how can we sustain it?

During the nineteenth century and the early part of this century, America was an also-ran in the field of science. In the 1930s many of the leading scientists in the world came to the United States to escape the scourge of fascism, and most stayed here after the war. They trained a whole new generation of native-born scientists and engineers, including hundreds of thousands of veterans whose graduate education was provided by the GI bill. Then when the Cold War started, our government, which had provided strong support for technology during World War II, decided to continue that support. So the new engineers and scientists, having just completed their education under the GI bill, were put to work advancing the state of the art in technology.

As a result, the Department of Defense (DOD) developed and was the first to use supercomputers, communications satellites, integrated circuits, high-order software languages, and the Internet (in fact, the predecessor to the Internet was the ARPANet). All of these developments were part of a strategy to maintain technological superiority as a way of offsetting the numerical superiority of Soviet ground forces. Fortunately, we never had to put this offset strategy to a test with the Red Army. But the high-tech weapons systems we developed for that purpose were put to a test in Desert Storm. There we were fighting against a foe equipped with Soviet weapons but in about equal numbers to the allied forces.

American forces were equipped with the new weapons systems developed during the 1970s, which used information technologies to locate enemy targets on the battlefield, embedded computers to guide weapons precisely to those targets, and stealth technology to evade enemy weapons. As a consequence, the allied forces won quickly, decisively, and with remarkably few casualties. That is, when our technological advantage is not needed to offset superior numbers, it can be used to achieve battlefield dominance over a foe with equal numbers.

Our military leaders, having seen the results of battlefield dominance in Desert Storm, decided that they liked it and that they wanted to keep it. And so today our military strategy calls for maintaining battlefield dominance over any regional power with whom we might be engaged in conflict, and to do that through our leadership in technology. This is the same strategy of technological superiority we had during the Cold War but now for a different reason. Not only has the reason changed, but so has the way of applying this technological superiority.

During the 1950s, 1960s, and 1970s, DOD was the principal supporter of research and development for the computer, communications, and semiconductor industries. In effect, our nation's commercial industry was riding on the shoulders of the DOD. Today that has all changed. The technological explosion in computers, communications, and semiconductors has led to an amazing new set of products for industry, businesses, schools, and homes. Indeed, all of these different users are being tied together today by the World Wide Web in a way few could have predicted a decade ago. Commercial applications of computers are leading military applications, and for computer companies commercial revenues dwarf DOD revenues.

Today DOD is obliged to ride on the shoulders of our commercial information technology industry, and this poses a difficult problem for military planners. Our major weapons systems take 10 to 15 years to develop and then are in the inventory for 20 to 40 years. But the computer technology that most influences their competitive advantage changes every two or three years.

When I became Secretary of Defense, I was determined to do something about this problem. I knew that I needed a systems strategy that would keep major weapons systems in the field for several decades but update them every few years with new information technology. So I approved the creation of a large-scale experimental program to do just that; the Army calls it Force 21—the digitized battlefield. Force 21 holds promise to dramatically improve the way the United States adopts and adapts computer technology to military uses.

The concept behind the experiment was simple. The Army inserted appliques of commercially available digital subsystems into their current weapons systems—tanks, artillery, attack helicopters—thereby giving them a quantum increase in capability. In aggregate these appliques form a "system of systems," an integrated network of powerful computers and high-speed communications—basically, an Internet on the battlefield. This system of systems will transform the way commanders and troops see and communicate on the battlefield.

How does this work in battle? When a tank commander spots enemy forces, he will have a choice: he could engage the enemy with the weapons on his tanks or he could call in nearby attack helicopters, artillery, strike aircraft, or naval gunfire. Because of digital technology and the constant flow of battlefield information to all combatants, these other units will see exactly what the tank commander sees. And any one of them—or any combination of them—will be able to respond with equal precision in attacking the targets. As combat is under way, the supporting logistics unit will be monitoring the ammo usage. So it will be able to conduct resupply at the time and with the amount needed, thereby reducing the huge logistics tail needed to support combat operations. This system of systems is a brilliant application of information technology to achieve battlefield dominance without designing all new weapons platforms. This is the army of the future, and it is not just on viewgraphs. The Army has already outfitted the 4th Division at Fort Hood with this new equipment and is testing it in simulated combat.

This example makes it clear that our national security strategy depends on

U.S. leadership in information technologies. But can we count on our commercial industry to sustain that leadership? To get some insight into this question, consider how the United States achieved world leadership in integrated circuits many decades ago.

After the semiconductor had been discovered, engineers all over the world were rushing to develop applications for this new device. During those heady times, a conference was held of the engineers working on these applications, and a seminal paper was presented at that conference by the British engineer, G. W. A. Dummer. In his paper he said: "With the advent of the transistor (and the work in semiconductors in general), it seems now possible to envisage electronic equipment in a solid block with no layers of insulating, conducting, rectifying, and amplifying materials, the electrical functions being connected directly by cutting out areas of the various layers." Well, you don't have to be an electronics engineer to understand that he was describing what later came to be called the integrated circuit, and he was racing to be its inventor.

But the integrated circuit was not invented by a British engineer named Dummer; it was invented by two American engineers named Noyce and Kilby. At a similar technical seminar held a few years after the integrated circuit was announced, Dummer presented another paper in which he mused about the reasons he and other European engineers came in second. "It is worth remembering," he said, "that the giant American electronics companies were formed since the war by a relatively few enterprising electronics engineers, setting up with either their own capital or risk capital from a bank. Often a government contract would start them off. Hard work was necessary, and the large home market was a great asset, but the climate of innovation was such that any advanced technical product could be sold. Successful businesses are almost always dependent on a few people who are innovative and enthusiastic." That story encapsulates how the United States gained technical leadership first in the semiconductor industry and later in the information technology industry. In essence, Dummer was saying that, although military contracts were helpful, America won primarily because of its innovators.

But the world has changed in many ways since then. In particular, DOD support is much less consequential, and without that support Japan has risen as a world-class competitor in the information technology field and has challenged U.S. leadership. So what is the likelihood that our commercial industry can sustain its world leadership without the major support from DOD that it had during the Cold War? Let me relate "A Tale of Two Countries" because I believe this tale, which is about competitive technology programs in the United States and Japan, illustrates how critically important innovation is to a country's ability to maintain technological leadership.

Just a decade ago the United States was falling behind Japan in its ability to compete in world markets, and all the trends seemed to be negative. The United States had a very high rate of consumption and a low rate of savings, especially relative to Japan. Japanese auto companies outcompeted American auto companies, even in domestic U.S. markets, with a profound negative effect on our balance of trade. Japanese companies dominated the world market in consumer electronics. And Japan's leading electronics companies were embarked on an intensive and concentrated effort to overtake the lead of American companies in information technology. It appeared that the United States was about to lose one more major market to Japan and this in a field of traditional American market dominance.

How do things look after a decade of this predicted demise of U.S. competitiveness? The U.S. consumption rate is still high and its savings rate is still low, and in my opinion that is not likely to change. Japan's savings rate is even higher than it was, resulting in low consumption, and this is believed to be one of the factors in its current economic slump. That is ironic because a decade ago economists cited Japan's high savings rate as a major driver in its economic growth. U.S. auto companies, by adopting Japanese production techniques and an emphasis on quality, have gained back some of their lost market. While the United States has not gained back any significant part of the market for consumer electronics, Japan has lost market to increasingly difficult competition from the "Tiger" countries. But most importantly, Japan has failed in its bold attempt to seize leadership in the important market for information technologies.

What happened? The Japanese strategy, orchestrated by the Ministry of International Trade and Industry, focused on gaining leadership in three products involving leading-edge technologies: memory chips, fifth-generation (artificial intelligence) computers, and high-density television. They reasoned that these were important products in and of themselves but, more importantly, that leadership in these products would lead to compelling competitive advantages in all other products in the field of information technology. Many industry leaders in the United States agreed with this assessment and became increasingly alarmed as Japanese companies gained an increasingly larger share of the market for memory chips. Today the situation looks very different—indeed, it appears that Japanese companies "bet on the wrong horses." Memory chips have become a commodity product, with increasingly lower margins prevailing after Korea and the other "Tigers" entered the competitive fray. Artificial intelligence applications for computers have not matured, as envisioned a decade ago. And the market for high-density television has been very slow to materialize. In the meantime, information technology companies in the United States have proceeded in very different directions, concentrating their efforts in three areas: microprocessors, which were used in increasingly capable workstations and increasingly versatile desktop computers; telecommunications networks, which created the exploding market for Internet products; and software, which provided a competitive advantage across the board in information technology products. How did these two different countries, each with very capable technologists and managers, come to such different outcomes?

I believe three primary factors affected this outcome: (1) The Japanese strategy was driven to a large degree by its government, while the U.S. strategy was driven by a large number of individual entrepreneurs. When a government puts its support behind a product, it can be a powerful force, but if it guesses wrong the corrective forces found in the marketplace may be rendered ineffective. (2) American entrepreneurs initiating new products or new companies found an abundant supply of risk capital, both from venture capitalists and in the dynamic market for public offerings for high-tech companies. There are no comparable markets for risk capital in Japan. (3) The technological skills required for success in these new markets were in abundant supply at the leading technical universities in the United States. This resulted in new product ideas from university labs and, even more importantly, a fresh wave of scientists and engineers trained in information technology. The training at America's technical universities was relevant and at the cutting edge because of the unique bonding between America's technical universities and its high-tech companies. In sum, the success of American companies stemmed from three great assets: our entrepreneurial spirit, our innovative markets for risk capital, and our great technical universities.

These advantages are quite fundamental and are likely to allow us to sustain our leadership in any product where innovation is the key to success. But if we want to sustain these advantages into the twenty-first century, we must value them and nurture them. I believe that our entrepreneurial spirit will continue to thrive for the indefinite future. It is a basic cultural advantage we have, and we should recognize and cherish it. Similarly, I believe that we will continue to lead the world in our innovative markets for risk capital, but to ensure that we should strive to educate the public and the Congress about the importance that risk capital has to our overall economy. By so doing we should be able to help lawmakers resist the temptation to overregulate these markets, thereby killing the goose that is laying these golden eggs. I am concerned, though, about our ability to maintain the position of world leadership of our great technical universities.

For more than four decades DOD provided the majority of funding for America's technology base, which was underlying all of the remarkable technical products developed during that period. With the end of the Cold War, the DOD budget has decreased, in real terms, by 40 percent; and the funding for the technology base has decreased proportionately. Moreover, DOD's production contracts have decreased about 70 percent, and thus the defense contractor's independent R&D, which is proportional to overall sales, has decreased proportionately. And while there has been some increase in the rest of the federal R&D budget, the increase has not been proportional, and it has mostly been in health-related technologies.

On the positive side, our information technology companies have had dramatic increases in revenues and profits this past decade and have made corresponding increases in their expenditures for R&D. But these expenditures have been almost exclusively for product development and have not served to replenish our technology base. So this is a serious problem for the future and will be solved only by an increase in federal funds for technology-based programs and by industry consortia that pool some of their R&D funds to support R&D base programs at our universities.

I have given some examples here of how technological innovations have played a critical role in our economy and our national security. I would like to conclude by describing how technology is used not just in the making of our weapons systems but in decisionmaking in national security.

Government officials in national security find decisions in this field particularly challenging for three reasons: First, the stakes are so high—the price of a wrong decision can be thousands of lives. Second, the problems are incredibly complex—quantitative decision techniques can generally be applied only to a segment of the problem. Third, the decisionmaker rarely has either sufficient data or sufficient time to make an analysis of all important factors bearing on the decision. This last dilemma was captured brilliantly by C. P. Snow (1962), the British scientist who worked on technically complex defense projects during World War II. In his book *Science and Government,* he wrote: “One of the most bizarre features of any advanced industrial society in our time is that the cardinal choices have to be made by a handful of men who cannot have a first-hand knowledge of what those choices depend upon or what their results may be. . . . When I say the ‘cardinal choices,’ I mean those which determine in the crudest sense whether we live or die.”

As Secretary of Defense I was faced every day with cardinal choices—choices that determined whether our soldiers would live or die. And I generally had to make my decision without enough data and without enough time to analyze even the data I had. I understood that I would generally not be able to apply analytical tools to the decisions I had to make. Nevertheless, I found that knowledge of those tools was invaluable. That knowledge allowed me to approach decisionmaking with an objective framework for assessing the validity and relevance of the data I had—and didn’t have—and for assessing alternative solutions. This analytical rigor served me well on many of the important decisions that I made.

Let me illustrate some of the important principles of decisionmaking by giving some examples from my tenure as Secretary of Defense. The first principle is that you never have enough data to solve the most important problems analytically, but that does not mean that analytical techniques cannot be used—it means instead that you have to separate the variables in the problem. It may then be possible to apply analytical tools to one or more of the separated parts. For example, the President did not use analytical tools when he decided to deploy U.S. troops to Bosnia in December 1995. But his decision depended on my assuring him of the feasibility and relative safety of the mission, and I based my assurance to him on some rather detailed analyses. Our simulated war games provided the basis for believing that we could carry out our part of the missions in the Dayton Agreement with 25,000 U.S. troops. Our detailed flow analysis showed that we could move these troops into Bosnia from Germany fast enough that there would not be a period when we would be so weak as to invite attack. And our detailed logistical analysis showed that we could supply these troops from Hungary, where we had been offered a support base. Some of the detailed analyses underlying my assurances to the President required our logisticians to employ rather sophisticated tools from linear programming and queuing theory.

A second important principle is that all important decisions involve uncertainties that cannot be resolved because they involve data not known at the time a decision must be made. Because of the pervasive nature of statistical uncertainty in nearly all of our decisions, our decision tools must be able to deal with random variables.

Let me now give a real example of statistical uncertainty in the defense field that occurred in the 1970s. This was during the height of the Cold War. The United States and the Soviet Union both practiced a doctrine known as mutual assured destruction, aptly nicknamed MAD, which had our two countries locked together in a balance of terror. At one stage we believed that we might be able to break this balance of terror with the deployment of antiballistic missiles (ABMs). But we knew that a deployment would be costly, both in dollars and in the possibility that an ABM system might provoke an attack while it was being built. Therefore, we wanted to be sure that our ABM system could not only shoot down missiles but also that in so doing it would protect the country from the threat we postulated. Calculation of the protection provided by an ABM system involved random variables—some of the relevant data (e.g., the yield of the Soviet warheads) was known with precision but not to us, so we treated warhead yield not as a fixed number but as a probability distribution. Other factors, such as the actual impact point of a missile aimed at a certain target, were unknowable, even to the side firing the missile, except as a probability distribution. Finally, we knew that we were protecting against a large-scale attack directed at diverse targets, but we did not know how specific warheads would be assigned to specific targets; therefore, that also had to be treated as a random variable.

After many false starts and misleading analyses, DOD finally developed the analytical tools that dealt adequately with this complex problem. All relevant intelligence data were assembled about the characteristics of Soviet weapons, the hardness of presumed U.S. targets, and the weapons effects against various targets, as determined by field measurements made on our own weapons. A half dozen of the most significant variables were assigned probability distributions that reflected the best information available about them, including the uncertainty about these data. Then a variety of attack scenarios were planned and run thousands of times on the computer, using Monte Carlo techniques to reflect the probability distributions.

What resulted from this extensive set of calculations was that the results of the missile attack were reflected in a probability distribution that in turn reflected the real uncertainty of the input data. Decisionmakers then were presented with the resulting calculated damage not as a single number but as an average and a deviation about that average. This extensive set of calculations was an important factor in the final decision—namely, not to depend on the deployment of an ABM system to defend the country but instead to negotiate an arms control treaty that limited the deployment of both missiles and ABMs.

This is an example of how a complex operational analysis was used to influence a major decision on a real-life security problem. I do not mean to imply that because such extensive analysis was performed that the answer is right or even that all analysts agreed with the decision—that is far from true. Nor do I mean to suggest that this analysis done on the missile threat in 1970s is relevant to decisions about how to deal with the missile threat in the twenty-first century. Instead, I want to suggest that technology by itself is not enough. It is also important that we make the best decisions about how to use technology to most effectively strengthen our national security.

Let me close with a final quote from C. P. Snow. He wrote: "Technology is a queer thing. It brings you great gifts with one hand, and it stabs you in the back with the other." The time-honored role of engineers is to bring technology to the public in the form of products. The emerging role of the systems engineer is to ensure that the system of systems created by the new technology does not stab us in the back; that technology really does, as we hope, bring us great gifts.

## REFERENCE

Snow, C. P. 1962. Science and Government. New York: New American Library.

# APPENDIXES

# Contributors

**ARMAND J. BEAUDOIN** is an associate professor in the Department of Mechanical and Industrial Engineering at the University of Illinois, Champaign-Urbana. His work experience includes 15 years with the metals industry. In the early stages of his career Dr. Beaudoin developed supervisory control systems for the nondestructive evaluation of aluminum aircraft plate. Here the disparity between the rudimentary state of metal-forming analysis and considerable artistry of the production mill operator became apparent. This led to further study of metal forming with a combined metallurgical and mechanical perspective. To this end, Dr. Beaudoin received a master's of materials science degree from the University of Virginia and a Ph.D. in mechanical engineering from Cornell University. Returning to industry, he pursued computer simulation of material behavior as a basis for process design. Such analysis resulted in the development of thermomechanical processes for production of aluminum alloys used in packaging, auto body panels, and aircraft plate. His interests have extended to other materials, such as steels used in the construction of ships and submarines.

**JOEL W. BURDICK** is an associate professor in the Department of Mechanical Engineering, California Institute of Technology. His research is in the areas of sensor-based robotic motion planning, mechanics and control of robotic and biological locomotion, applied nonlinear control, hyper-redundant robotic systems, multifingered robotic hand manipulation, industrial parts fixturing, and medical applications of robotics. While at CalTech, Dr. Burdick has been the recipient of the National Science Foundation Presidential Young Investigator Award, the Office of Naval Research Young Investigator Award, and the Feynman Fellowship. He has also received the Associated Students of the California

Institute of Technology award for excellence in undergraduate teaching and the Graduate Student Council award for excellence in graduate student education. Dr. Burdick received his undergraduate degrees in mechanical engineering and chemistry from Duke University and his M.S. and Ph.D. in mechanical engineering from Stanford University.

**JAMES G. FUJIMOTO** is a professor of electrical engineering at the Massachusetts Institute of Technology (MIT) and an adjunct professor of ophthalmology at Tufts University. He received his undergraduate, S.M., and Ph.D. degrees from MIT. Dr. Fujimoto's research interests include femtosecond laser generation and measurement techniques, studies of ultrafast processes in electronic and optoelectronic materials and devices, and laser medicine. He was the recipient of a National Science Foundation Young Investigator Award, an AT&T New Research Fund Award, an IEEE Lasers and Electro-Optic Society Traveling Lecturer Award, and the National Academy of Sciences Award for Initiatives in Research.

**MARK N. GLAUSER** is a professor in the Department of Mechanical and Aeronautical Engineering at Clarkson University in Potsdam, New York. His research has ranged from relatively simple mixing layers (2D and axisymmetric) to complicated 3-D transitioning boundary layers and separated flows. Dr. Glauser is also leading an effort at Clarkson to develop what have been termed interactive labs of the future—laboratory course modules that utilize a multimedia interactive environment (based on MAPLE and LabVIEW) to integrate experimental, computational, and analytical tools in a complementary and coherent fashion. In the academic year 1994–1995 he was a visiting senior research scientist at NASA Langley and a Fulbright Fellow at the Centre d'Etudes Aerodynamiques & Thermiques/Laboratoire d'Etudes Aerodynamiques in Poitiers, France. Dr. Glauser received his Ph.D. in fluid dynamics from the University at Buffalo, State University of New York.

**SHARON C. GLOTZER** is a computational physicist in the Polymers Division of the Materials Science and Engineering Laboratory at the National Institute of Standards and Technology (NIST) and serves as director of the NIST Center for Theoretical and Computational Materials Science (CTCMS), which she cofounded. Dr. Glotzer began her career at NIST as a National Research Council postdoctoral fellow. Her research interests include theoretical and computational investigations of phase transitions and pattern formation in polymer materials and dynamics of dense liquids and glasses. In 1997 she received the U.S. Department of Commerce Bronze Medal Award for her contributions to the creation of the CTCMS and in 1998 the Presidential Early Career Award for Scientists and Engineers. Dr. Glotzer received a B.S. degree in physics from the University of

California, Los Angeles, and a Ph.D. in theoretical condensed matter physics from Boston University.

**KENNETH Y. GOLDBERG** is an associate professor in the Department of Industrial Engineering and Operations Research at the University of California, Berkeley. He received his Ph.D. from the School of Computer Science at Carnegie Mellon University and his B.S.E. from the University of Pennsylvania. Dr. Goldberg and his students study geometric algorithms for feeding, sorting, and fixturing industrial parts. He was named a National Science Foundation (NSF) Young Investigator in 1994 and an NSF Presidential Faculty Fellow in 1995.

**LINDA G. GRIFFITH** is the Karl Van Tassel Career Development Associate Professor, Department of Chemical Engineering and Division of Bioengineering and Environmental Health, Massachusetts Institute of Technology. Dr. Griffith's research is in the rapidly emerging field of tissue engineering, where cells are manipulated using biochemical factors, synthetic materials, and mechanics to form multi-dimensional structures that carry out the functions of normal tissue in vitro or in vivo. Her work focuses on controlling the spatial and temporal presentation of molecular ligands and physical cues that are known to influence cell behavior. Her research combines molecular design and synthesis of surfaces that interact with cells via receptor-mediated phenomena as well as design and synthesis of macroscopic 3-D devices, with an emphasis on developing liver tissue in vitro. One of her inventions, "injectable cartilage," is currently in clinical trials for treatment of urological disorders, and a second provided the basis for a new company, Therics, devoted to creation of complex devices for tissue engineering and drug delivery. Dr. Griffith received her B.S. from the Georgia Institute of Technology and a Ph.D. from the University of California at Berkeley, both in chemical engineering. She was the recipient of an National Science Foundation Presidential Young Investigator Award in 1991.

**LYDIA E. KAVRAKI** is an assistant professor of computer science at Rice University. She received a B.A. degree at the University of Crete in Greece and her Ph.D. degree at Stanford University. Before joining the Rice University faculty in 1996, she worked as a postdoctoral fellow and as a research associate at Stanford. Dr. Kavraki's research investigates algorithms and system architectures for solving geometric problems arising in the physical world. She is particularly interested in problems in the areas of motion planning, assembly sequencing, manufacturing, and applications in computational chemistry (pharmaceutical drug design) and medicine (robot-assisted surgery). Dr. Kavraki was a cochair of the International Workshop on the Algorithmic Foundations of Robotics in 1998 and received the National Science Foundation Early Career Development Award in 1997.

**JAY LEE** is director for product development and manufacturing, United Technologies Research Center (UTRC), where he is responsible for strategic direction and research and development activities in the areas of product development processes, manufacturing systems, sustainable process development, machining systems, quality systems, green products and processes, and advanced service technologies. Prior to joining UTRC, Dr. Lee served as director of the Industry/University Cooperative Research Center Program, the Engineering Research Centers Program, and the Materials Processing and Manufacturing Program at the National Science Foundation. Previously, he held several engineering and management positions in the automotive, precision machinery, and service industries. Dr. Lee has master's degrees in mechanical engineering from the University of Wisconsin, Madison, and industrial management from the State University of New York, Stony Brook, as well as a Ph.D. from George Washington University. He is active with the American Society of Mechanical Engineers and the National Research Council and serves as an adviser to programs in Japan and the United Kingdom. Dr. Lee has received fellowship awards from the Japan Society for the Promotion of Science and the Japan Science and Technology Agency and received the Society of Manufacturing Engineers' Outstanding Young Manufacturing Engineer Award.

**LARRY H. MATTHIES** is supervisor of the Machine Vision and Tracking Sensors Group at the Jet Propulsion Lab (JPL) in Pasadena, California. Dr. Matthies joined JPL after obtaining a Ph.D. in computer science from Carnegie Mellon University in 1989. His research interests center on computer vision for autonomous navigation, and he has been instrumental in the development of several 3-D perception systems for mobile robots. Dr. Matthies was a member of the flight team for the Sojourner Mars rover. He is now a member of the editorial board of the journal *Autonomous Robots* and an adjunct professor in the Department of Computer Science at the University of Southern California. His interest in computer vision for mobile robots grew out of prior interests in computer graphics and artificial intelligence. His work provides a rewarding balance of theoretical development, prototype system building, field trials in interesting locations, and, most of all, application to exciting missions.

**OMKARAM NALAMASU** is manager of the optical lithography and imaging materials research programs at Bell Laboratories/Lucent Technologies. He has made critical contributions to the deep-ultraviolet resist materials chemistry and process development areas and has published over 90 papers, review articles, and book chapters on various aspects of photoresist materials design, synthesis, formulation, process development, and their implementation in integrated circuit fabrication. Dr. Nalamasu has chaired and organized conferences on lithography and photoresist materials and has taught courses on materials for electronics in the United States and Japan through the American Chemical Society and the State University of New York continuing education organizations. Dr. Nalamasu obtained his Ph.D. in chemistry from the University of British Columbia.

**WILLIAM J. PERRY** is a professor at Stanford University, with a joint appointment at the Department of Engineering-Economic Systems/Operations Research and the Institute for International Studies. From February 1994 to January 1997, Dr. Perry was Secretary of Defense of the United States. His previous government experience was as Deputy Secretary of Defense (1993-1994) and as Under Secretary of Defense for Research and Engineering (1977-1981). Dr. Perry's business experience includes serving as a laboratory director for General Telephone and Electronics (1954-1964); founding and serving as the president of ESL, Inc. (1964-1977); Executive Vice President of Hambrecht & Quist, Inc. (1981-1985); and founder and chairman of Technology Strategies & Alliances (1985-1993). He serves on the board of United Technologies, Hambrecht & Quist, and several emerging high-tech companies. Dr. Perry is a member of the National Academy of Engineering, a fellow of the American Academy of Arts and Sciences, and the recipient of numerous military and civilian awards. He received his B.S. and M.S. from Stanford University and his Ph.D. from Penn State, all in mathematics.

**MICHAEL A. PESHKIN** is an associate professor of mechanical engineering at Northwestern University. Dr. Peshkin's research interests are in the area of robotics: image-guided surgery, collaborative robots or "cobots," the use of compliance characteristics to ease manual or automated assembly operations, and the use of sliding motions in robotic manipulation. He holds patents for cobots and for a fluoroscope-based image-guided surgical system, both of which are licensed to fledgling companies. Dr. Peshkin received his bachelor's degree from the University of Chicago; master's degree from Cornell University; and a Ph.D. in physics from Carnegie Mellon University, where he also worked at the Robotics Institute.

**MILIND RAJADHYAKSHA** is an optical engineer at Lucid, Inc. in Henrietta, New York, jointly with Massachusetts General Hospital-Wellman Labs of Photomedicine in Boston. He is responsible for the design and development of confocal microscopes at Lucid and for imaging and clinical research at MGH-Wellman Labs. Prior to joining Lucid, Dr. Rajadhyaksha was a member of the research faculty at Harvard Medical School-Massachusetts General Hospital, where he worked mainly in confocal microscopy. He is also a visiting faculty member at Tufts University, where he teaches microscopy to graduate students. Dr. Rajadhyaksha holds a B.S. from the Indian Institute of Technology, Bombay, and M.S. and Ph.D. degrees from Purdue University. He did his postdoctoral training at Harvard Medical School-MGH, jointly with the Schepens Eye Research Institute in Boston.

**JOHN G. SPEER** is a professor of metallurgy at the Colorado School of Mines. He teaches graduate and undergraduate courses in physical metallurgy and participates in the research efforts of the Advanced Steel Processing and Products Research Center, a National Science Foundation Industry/University Cooperative Research Center. He joined Bethlehem Steel's Homer Research Laboratories as a research engineer in 1983 and initially worked in the areas of microalloying and plate product development, including development of new high-strength low-alloy grades for construction and Navy ship-building applications. Dr. Speer also participated in sheet steel product development and formability programs and became a research supervisor of Bethlehem's Cold Rolled Products Group in 1989, responsible for directing product development, operations support, and customer technical support activities related to cold rolled sheet, coated sheet formability, cold mill process fluids, and surface appearance. Dr. Speer received his undergraduate degree in metallurgy and materials engineering from Lehigh University and a Ph.D. in physical metallurgy from Oxford University.

**DAVID A. TIRRELL** is Ross McCollum-William H. Corcoran Professor in the Division of Chemistry and Chemical Engineering, California Institute of Technology, a position he assumed in 1998. After earning a B.S. in chemistry from the Massachusetts Institute of Technology, Dr. Tirrell enrolled in the newly created Department of Polymer Science and Engineering at the University of Massachusetts, receiving his Ph.D. in 1978. Following a postdoctoral fellowship at Kyoto University and an assistant professorship in the Department of Chemistry at Carnegie Mellon University, Dr. Tirrell returned to Amherst, where he was Barrett Professor of Polymer Science and Engineering and director of the Materials Research Laboratory. He has served as a visiting professor at universities in Australia and France and at the University of Wisconsin and currently holds an appointment as a Rotschild Fellow at the Institut Curie in Paris. Dr. Tirrell's most important contributions to chemistry and chemical engineering have come in the areas of radical copolymerization mechanism, biomimetic membrane chemistry, and development of molecular biological approaches to new polymeric materials.

# Program

**NATIONAL ACADEMY OF ENGINEERING**

**FOURTH ANNUAL SYMPOSIUM ON FRONTIERS OF ENGINEERING**
September 17-19, 1998

## BIOMATERIALS AND OPTICAL IMAGING FOR BIOMEDICINE

*Organizers*: Frances Arnold and Eva Sevick-Muraca

**Integration of Molecular and Macroscopic Scales in Tissue Engineering**
Linda G. Griffith, Massachusetts Institute of Technology

**Artificial Proteins: Bridging the Gap Between Natural and Synthetic Macromolecular Materials**
David A. Tirrell, California Institute of Technology

**Biomedical Imaging Using Optical Coherence Tomography**
James G. Fujimoto, Massachusetts Institute of Technology

**Confocal Reflectance Microscopy: Diagnosis of Skin Cancer Without Biopsy?**
Milind Rajadhyaksha, Lucid Inc. and
Massachusetts General Hospital, Wellman Labs

* * *

## ADVANCED MATERIALS

*Organizers*: Richard Parnas, Elsa Reichmanis, and Jim Schroth

### Trends in Computational Materials Research for Materials Design and Development

Sharon C. Glotzer, National Institute of Standards and Technology

### Design, Synthesis, Development and Integration into Manufacturing of New Polymer Materials Architectures for Advanced Integrated Circuit Fabrication

Omkaram Nalamasu, Bell Labs/Lucent Technologies

### Novel Sheet Steel Developments

John G. Speer, Colorado School of Mines

* * *

## SIMULATION IN MANUFACTURING

*Organizers*: Connie Gutowski, Parviz Moin, Jun Ni, and Jim Schroth

### Role of Simulation in the Design of the Next-Generation Engines for Military Aircraft

Mark N. Glauser, Clarkson University

### Role of Simulation in Understanding Surface Roughness in Formed Aluminum Parts

Armand Beaudoin, University of Illinois at Urbana-Champaign

### Performance Assessment Methodology and Advanced Service Technologies

Jay Lee, United Technologies Research Center

* * *

## ROBOTICS

*Organizers*: Susan Corwin, Ken Goldberg, and Robert Howe

### A Brief History of Robotics

Kenneth Y. Goldberg, University of California, Berkeley

**Algorithms in Robotics: The Motion Planning Perspective**
Lydia Kavraki, Rice University

**Mechanics, Control, and Applications of Biomimetic Robotic Locomotion**
Joel W. Burdick, California Institute of Technology

**Robotic Perception for Autonomous Navigation of Mars Rovers**
Larry H. Matthies, Jet Propulsion Laboratory

**Cobots**
Michael A. Peshkin, Northwestern University

* * *

## DINNER SPEECH

**Technology Innovation and American Leadership**
William J. Perry, Stanford University

# Participants

## NATIONAL ACADEMY OF ENGINEERING

## FOURTH ANNUAL SYMPOSIUM ON FRONTIERS OF ENGINEERING

September 17-19, 1998

Venkat Allada
Assistant Professor
Dept. of Engineering Management
University of Missouri-Rolla

Mark G. Allen
Manager
Aeropropulsion Technologies
Physical Sciences Inc.

Nancy M. Amato
Assistant Professor
Dept. of Computer Science
Texas A&M University

Kristi S. Anseth
Assistant Professor
Dept. of Chemical Engineering
University of Colorado

Frances H. Arnold
Professor
Division of Chemistry and Chemical Engineering
California Institute of Technology

Carol R. Ashe
Associate
Camp Dresser & McKee, Inc.

Beauford W. Atwater
Program Manager
Sensors, Systems, and Telemetry Group
Bell Communications Research

Thomas G. Bifano
Associate Professor
Dept. of Aerospace and Mechanical Engineering
Boston University

Thierry A. Blanchet
Assistant Professor
Dept. of Mechanical Engineering, Aeronautical Engineering, Mechanics
Rensselaer Polytechnic Institute

Reid A. Brennen
Member of Technical Staff
Hewlett-Packard Laboratories

Joel W. Burdick
Associate Professor
Dept. of Mechanical Engineering
California Institute of Technology

Andrea Califano
Program Director
IBM Computational Biology Center

Jacqueline Chen
Senior Member of Technical Staff
Combustion Research Facility
Sandia National Laboratories

Wilfred Chen
Assistant Professor
Dept. of Chemical & Environmental Engineering
University of California, Riverside

Charles P. Coleman
Manager, System Diagnostics
Joseph C. Wilson Center for Research and Technology
Xerox Corporation

Michael L. Corradini
Associate Dean for Academic Affairs
Professor of Engineering Physics
University of Wisconsin-Madison

Susan Corwin
Director
Computing Enhancement Architecture Lab
Intel Corporation

John P. Coulter
Associate Professor
Dept. of Mechanical Engineering and Mechanics
Lehigh University

David C. Dankworth
Staff Engineer
Exxon Research and Engineering Co.

Mark E. Dean
IBM Fellow and Director
IBM Austin Research Laboratory

Kay C. Dee
Assistant Professor of Biomedical Engineering
Director, Cell and Tissue Engineering Laboratory
Tulane University

William Dowling
Senior Technical Specialist
Manufacturing Systems Department
Ford Research Lab

Michael P. Ekart
Principal Research Engineer
Eastman Chemical Company

James G. Fujimoto
Professor
Dept. of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Henri P. Gavin
Assistant Professor
Dept. of Civil and Environmental Engineering
Duke University

Mark D. Gehlsen
Senior Process Development Engineer
3M Company

Apostolos T. Georgiou
Associate Engineer
Modeling & Advanced Control Group
Mobil Technology Company

Mark N. Glauser
Professor
Dept. of Mechanical and Aeronautical Engineering
Clarkson University

Sharon C. Glotzer
Acting Director
Center for Theoretical and Computational Materials Science
National Institute of Standards and Technology

Kenneth Y. Goldberg
Associate Professor
Industrial Engineering and Operations Research
University of California, Berkeley

Larry Gray
Mechanical Engineer
DEKA Research & Development

Linda G. Griffith
Associate Professor
Department of Chemical Engineering
Division of Bioengineering and Environmental Health
Massachusetts Institute of Technology

Connie L. Gutowski (*unable to attend*)
Windstar Chassis Manager
Vehicle Center #2
Ford Motor Company

Erik Hagersten
Chief Architect
High-End Servers
Sun Microsystems

Paula T. Hammond
Assistant Professor
Chemical Engineering Department
Massachusetts Institute of Technology

Brian E. Holtz
Senior Engineer
Computer Motion Inc.

Robert D. Howe
Associate Professor
Division of Engineering and Applied Sciences
Harvard University

James F. Huber
Senior Product Manager
Network Products
FANUC Robotics North America, Inc.

Michael B. Jones
Researcher
Microsoft Research
Microsoft Corporation

J. Woody Ju
Professor
Dept. of Civil and Environmental Engineering
University of California, Los Angeles

John Kallinderis
Associate Professor
Dept. of Aerospace Engineering
University of Texas at Austin

Lydia Kavraki
Assistant Professor
Dept. of Computer Science
Rice University

Robert G. Kelly
Associate Professor
Dept. of Materials Science and Engineering
University of Virginia

Kevin L. Kloker
Director
Corporate Systems Research Labs
Motorola, Inc.

Kathleen S. Krisch
Member of Technical Staff
Bell Laboratories, Lucent Technologies

David N. Ku
Professor
School of Mechanical Engineering
Georgia Institute of Technology

Dimitris C. Lagoudas
Professor
Department of Aerospace Engineering
Texas A&M University

Jay Lee
Director
Product Development and Manufacturing
United Technologies Research Center

Ming Li
Staff Scientist
Material Mechanics and Microstructure Center
Aluminum Company of America

John C. Lichauer
Senior Program Engineer
Westinghouse Electric Corporation

Don M. Lipkin
Materials Scientist
Physical Metallurgy Laboratory
GE Corporate Research and Development

Jon P. Longtin
Assistant Professor
Dept. of Mechanical Engineering
State University of New York at Stony Brook

Rick L. Martineau
Team Leader
Los Alamos National Laboratory

Larry H. Matthies
Supervisor
Machine Vision and Tracking Sensors Group
Jet Propulsion Laboratory

David F. McLaughlin
Fellow Engineer
Westinghouse Science and Technology Center

Alex Meduvsky
Manager
Heated Gas Inflator Program
TRW Vehicle Safety Systems

Parviz Moin
Professor
Department of Mechanical Engineering
Stanford University

George G. Muntean
Technical Specialist
Cummins Engine Co., Inc.

Omkaram Nalamasu
Technical Manager
Optical Lithography and Imaging Materials
Bell Laboratories, Lucent Technologies

Jun Ni
Professor, Mechanical Engineering and Applied Mechanics
Director, Wu Manufacturing Research Center, Director, NSF-I/UCR Center
University of Michigan

David P. Norton
Research Staff Member
Solid State Division
Oak Ridge National Laboratory

Linda K. Nozick
Associate Professor
School of Civil and Environmental Engineering
Cornell University

Victor G. Oancea
Senior Research Engineer
Hibbitt, Karlsson & Sorensen, Inc.

Julio M. Ottino
Walter P. Murphy Professor and Chair
Dept. of Chemical Engineering
Northwestern University

Richard S. Parnas
Group Leader
Polymer Composites, Polymers Division
National Institute of Standards and Technology

Douglas H. Pearson
Research Physicist
Naval Research Laboratory

Charles C. Peck
Manager, Advanced Distributed Architecture Research
Advanced Technology Laboratories
Lockheed Martin Corporation

Michael A. Peshkin
Associate Professor
Mechanical Engineering Department
Northwestern University

Melinda Piket-May
Assistant Professor
Dept. of Electrical and Computer Engineering
University of Colorado

Milind Rajadhyaksha
Optics Engineer
Massachusetts General Hospital - Wellman Labs

Elsa Reichmanis (*unable to attend*)
Head
Polymer and Organic Materials Research Department
Bell Laboratories, Lucent Technologies

Clifford L. Renschler
Manager
Electronic and Optical Materials Dept.
Sandia National Laboratories

Philip M. Rice
Senior Professional Associate
Central Artery/Tunnel Project
Parsons Brinckerhoff, Inc.

Steven D. Rizzi
Assistant Vice President and Division Manager
Advanced Information Technology Division
Science Applications International Corporation

William L. Roberts
Assistant Professor
Dept. of Mechanical and Aerospace Engineering
North Carolina State University

Krista B. Rossiter
Engineer
Kemper Engineering Lab
The Procter & Gamble Company

Sarah M. Ryan
Assistant Professor
Industrial and Management Systems Engineering
University of Nebraska-Lincoln

Meera Sampath
Member, Research and Technology Staff
Control and Diagnostics Laboratory
Joseph C. Wilson Center for Research and Technology
Xerox Corporation

John Saw
Director
Hardware Development
AT&T Wireless Services

James G. Schroth
Staff Research Engineer
General Motors R&D Center

Daniel T. Schwartz
Associate Professor
Dept. of Chemical Engineering
University of Washington

Eva M. Sevick-Muraca
Associate Professor
School of Chemical Engineering
Purdue University

Michael C. Shaw
Manager
Design and Reliability, Materials Science
Rockwell Science Center

Scott Singerman
Technology Manager
Compression Systems Component Center
Pratt & Whitney

Louis J. Soslowsky
Associate Professor and Director of Orthopaedic Research
Depts. of Bioengineering and Orthopaedic Surgery
University of Pennsylvania

John G. Speer
Professor
Dept. of Metallurgical and Materials Engineering
Advanced Steel Processing and Products Research Center
Colorado School of Mines

Yu-Chong Tai
Associate Professor of Electrical Engineering
California Institute of Technology

Robert D. Tilton
Associate Professor
Dept. of Chemical Engineering
Carnegie Mellon University

David A. Tirrell
Ross McCollum-William H. Corcoran Professor
Division of Chemistry and Chemical Engineering
California Institute of Technology

Scott Tychsen
Product Manager
Raytheon Aircraft Company

John H. van Zanten
Assistant Professor
Dept. of Chemical Engineering
Johns Hopkins University

Steven J. Visco
Principal Investigator
Materials Sciences Division
Lawrence Berkeley National Laboratory

Robert H. Wagoner
Professor
Dept. of Materials Science and Engineering
Ohio State University

Howard Waldman
Consulting Scientist
Raytheon Systems Company

Andrew M. Weiner
Professor
School of Electrical and Computer Engineering
Purdue University

Denise Wilson
Assistant Professor
Dept. of Electrical Engineering
University of Kentucky

Michael Winter
Director
Hamilton Standard Program
United Technologies Research Center

Fuquan (Frank) Zhao
Product Engineer
Chrysler Corporation

Jian-Gang (Jimmy) Zhu
Associate Professor
Dept. of Electrical and Computer Engineering
Carnegie Mellon University

Steven J. Zinkle
Research Staff Member
Metals and Ceramics Division
Oak Ridge National Laboratory

*Guest Speaker*

William J. Perry
Michael and Barbara Berberian Professor
Engineering-Economic Systems and Operations Research
Stanford University

*German-American Frontiers of Engineering*

Wolfgang Benecke
Managing Director
Institute for Microsensors, -actuators and -systems (IMSAS)
University of Bremen

*U.S.-Japan Frontiers of Engineering*

Tetsuo Miwa
Research Head
Medicinal Chemistry Laboratories
Takeda Chemical Industries, Ltd.

Yoshihiro Takiguchi
Researcher
Central Research Lab
Hamamatsu Photonics K.K.

Ken-ichi Yoshie
Senior Research Scientist
Yokohama Research Center
Mitsubishi Chemical Corporation

*Senior Engineers*

Malcolm J. Abzug
Retired President
ACA Systems, Inc.

Irving L. Ashkenas
Vice Chairman of the Board
Systems Technology, Inc.

William F. Ballhaus, Sr.
President
International Numatics, Inc.

George A. Bekey
Gordon Marshall Professor of Computer Science and
Director, Robotics Research Laboratory
University of Southern California

Robert H. Wertheim
Consultant
Science Applications International Corporation

## *Agency Representatives*

William S. Butcher
Senior Engineering Advisor
Directorate for Engineering
National Science Foundation

Clifford Lau
Program Officer
Corporate Programs Division
Office of Naval Research

## *National Research Council*

Mahadevan (Dev) Mani
Executive Office
National Research Council

# Glossary

**acellularity:** being without cells or composed of tissue not divided into separate cells.

**alkanethiols:** simple organosulfur compounds that are derivatives of hydrogen sulfide in the same way that alcohols are derivatives of water, in which an alkyl group (e.g., methane, ethane, propane) is joined to a mercapto group.

**alkenyl:** any univalent aliphatic hydrocarbon radical $C_nH_{2n-1}$ derived from an alkene by removal of one hydrogen atom.

**alkynyl:** a univalent aliphatic hydrocarbon radical containing a triple bond.

**avascularity:** having few or no blood vessels.

**catalytic deprotection:** in a catalytic reaction, cleavage of a protecting group from a molecule by a heat-stable species or substance.

**corneum:** horny layer; in the stratum corneum, this is the outer, more or less horny part of the epidermis.

**cytoplasm:** cell substance between the cell membrane and the nucleus, containing the cytosol, organelles, cytoskeleton, and various particles.

**derivatized:** a chemical substance that is so related structurally to another substance as to be theoretically inferred even when it is not obtainable in practice.

**deuterated:** containing the isotope of hydrogen, especially as the constituent of a chemical compound.

**elastin:** a protein, similar to collagen, that forms the chief constituent of elastic fibers and contributes to the elastic properties of various tissues.

**endothelial:** a type of epithelium composed of a single layer of smooth, thin cells that lines the heart, blood vessels, lymphatics, and serous cavities.

**epithelial:** any animal tissue that covers a surface, or lines a cavity or the like, and that, in addition, performs any of various secretory, transporting, or regulatory functions. It usually consists almost entirely of closely packed flat (squamous) or columnar (cuboidal or pyramidal) cells, with little intercellular material.

**haptic:** relating to, or based on, the sense of touch.

**hepatocytes:** epithelial parenchymatous cells of the liver.

**heterodyne:** noting or pertaining to a method of changing the frequency of an incoming radio signal by adding it to a signal generated within the receiver to produce fluctuations or beats of a frequency equal to the difference between the two signals.

**hydrophilic:** having a strong affinity for water.

**hydrophobic:** having a lack of affinity for water, or not readily wet by water.

**hydroxyl functionality:** OH group (one atom of bonded oxygen to hydrogen) bound to another atom on a molecule.

**laparoscopy:** examination of the abdominal cavity or performance of minor abdominal surgery using a flexible fiberoptic instrument passed through a small incision in the abdominal wall and equipped with biopsy forceps, an obturator, scissors or the like.

**leucine:** a white, crystalline, water-soluble amino acid essential in the nutrition of humans and animals, which is obtained by the decomposition of proteins and made synthetically.

**ligands:** molecules, such as antibodies, hormones, or drugs, that bind to a receptor.

**lipophilic:** promoting the solubilization or absorption of lipids.

**macular**: relating to an irregularly oval, yellow-pigmented area on the central retina, containing color-sensitive rods and the central point of sharpest vision.

**moieties:** parts or portions of a molecule, generally complex, having a characteristic chemical property.

**neoplastic:** relating to, or having the characteristics of, a new growth of tissue resembling the tissue from which it arises but having no physiologic function and being benign, potentially malignant, or malignant in character.

**nucleation:** In crystallization processes, the formation of new crystal nuclei in supersaturated solutions.

**retinopathy:** any diseased condition of the retina, the innermost coat of the posterior part of the eyeball that receives the image produced by the lens, especially a disease that is noninflammatory.

**spinodal decomposition:** an unmixing process in which crystals with bulk composition in the central region of the phase diagram separate when cooled.

**thienyl:** either of two univalent isomeric radicals derived from thiophene by removal of a hydrogen atom from either the alpha or 2-position or the beta or 3-position.