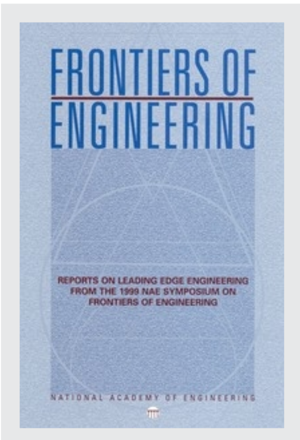

Frontiers of Engineering: Reports on Leading Edge Engineering from the 1999 NAE Symposium on Frontiers of Engineering

DETAILS

138 pages | 6 x 9 | HARDBACK
ISBN 978-0-309-38239-7 | DOI 10.17226/9774

AUTHORS

National Academy of Engineering

FIFTH ANNUAL SYMPOSIUM ON FRONTIERS OF ENGINEERING

NATIONAL ACADEMY OF ENGINEERING

NATIONAL ACADEMY PRESS
Washington, D.C.

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

Funding for the activity that led to this publication was provided by the Alcoa Foundation, Defense Advanced Research Projects Agency, Department of Defense—DDR&E-Research, National Aeronautics and Space Administration, Parsons Brinckerhoff, Science Applications International Corporation, and the National Academy of Engineering Fund.

International Standard Book Number 0-309-06933-5

Additional copies of this report are available from National Academy Press, 2101 Constitution Avenue, N.W., Lockbox 285, Washington, D.C. 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu

Printed in the United States of America

THE NATIONAL ACADEMIES

National Academy of Sciences
National Academy of Engineering
Institute of Medicine
National Research Council

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

ORGANIZING COMMITTEE

ROBERT H. WAGONER (Chair), Distinguished Professor of Engineering, Department of Materials Science and Engineering, Ohio State University
FRANCES ARNOLD, Professor, Division of Chemistry and Chemical Engineering, California Institute of Technology
MICHAEL L. CORRADINI, Associate Dean, Academic Affairs; Professor, Nuclear Engineering and Engineering Physics, University of Wisconsin
KENNETH Y. GOLDBERG, Associate Professor, Department of Industrial Engineering and Operations Research, University of California, Berkeley
CONNIE L. GUTOWSKI, Chief Program Engineer for F-150, Expedition, and Navigator, Product Development Center, Ford Motor Company
ERIK HAGERSTEN, Chair, Computer Architecture, Department of Computer Systems, Uppsala University
DAVID MCLAUGHLIN, Fellow Engineer, Westinghouse Science and Technology Center
JULIO M. OTTINO, Walter P. Murphy Professor and Chair, Department of Chemical Engineering, Northwestern University
CLIFFORD L. RENSCHLER, Manager, Electronic and Optical Materials Department, Sandia National Laboratories
EVA M. SEVICK-MURACA, Professor, Chemical Engineering Department, Texas A&M University
JIAN-GANG ZHU, Associate Professor, Department of Electrical and Computer Engineering, Carnegie Mellon University

Staff

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

Preface

This is the fifth book highlighting the presentations of the National Academy of Engineering's (NAE) annual symposium series, Frontiers of Engineering. The 1999 NAE Symposium on Frontiers of Engineering was held October 14–16, at the Academies' Beckman Center in Irvine, California. The 101 emerging engineering leaders (ages 30–45) from industry, academia, and federal laboratories who attended the meeting heard presentations and discussed cutting-edge research and technical work in four engineering fields. Symposium speakers were asked to prepare extended summaries of their presentations, and it is those papers that are contained here. The intent of this book, and of the four that precede it in the series, is to describe the content and underpinning philosophy of this unique meeting and to highlight some of the exciting developments in engineering 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 organizations 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 opportunity 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 1999 SYMPOSIUM

Magnetic data storage technology, applications of DNA array technology, renewable energy technologies, and optical applications of microelectromechanical systems (MEMS) were just a few of the topics covered at the 1999 symposium. The four broad areas covered were information technology, advances in bioengineering and chemical engineering, energy and the environment, and optics. Utilizing the theme *Drowning in Data,* presenters in the information technology session talked about managing the enormous amounts of on-line data being generated. Topics covered included the data storage technology of magnetic recording, the role of large multiprocessor servers in today's computing infrastructure, network survivability and information warfare, and the future of Web search. The bio/chemical engineering session began with a presentation on the generation of miniaturized, high-density arrays of oligonucleotide probes that have a variety of applications, including the analysis of genetic mutations and as a tool for genomic mapping. This was followed by two talks that highlighted developments in the engineering of novel structures: colloidal scale engineering—which has potential applications in optical information processing and storage, advanced coatings, and catalysis—and design strategies for creating synthetic macromolecular systems able to perform functions with biomimetic specificity. The third session of the meeting was titled *Energy for the Future and Its Environmental Impact.* Speakers in this session described the engineering challenges presented by deregulation in the electric power industry; the near, intermediate, and long-term role of nuclear energy; and the current status and future impact of renewable energy technologies such as biomass power, geothermal energy, wind, and solar power. In the optics session, one presentation covered the manufacture and commercialization of vertical-cavity surface emitting lasers (VCELS), which will enable applications in optical communications, biomedicine, consumer electronics, and optical computing. A second talk described the recent advances in MEMS that have opened up new opportunities for optical and optoelectronic systems, including optical systems on a chip and monolithic integration of a large number of optomechanical devices. (See Appendixes for complete program.)

As in past years, the varied backgrounds of the participants set the stage for lively question-and-answer periods and discussions both during the formal sessions and at breaks and mealtimes. The topics covered during these times ranged from technical questions to broader public policy issues related to engineering. In response to suggestions from past participants, two break-out sessions and a summary session on the last day were worked into the schedule. At those times, attendees met in smaller groups to discuss more "generic" engineering issues such as engineering education and public understanding of engineering. The outcomes of those sessions can be found in the Appendix.

As has been done in previous years, a distinguished engineer was invited to address the Frontiers of Engineering participants at dinner on the first evening of the symposium. At the 1999 meeting, Kent Kresa, chairman, president, and CEO of Northrop Grumman, talked about the rapidity of change and its impact on the engineering profession. Citing examples from the automotive and aerospace industries, he described how the ability to meet seemingly insurmountable challenges not only promotes innovation but can remake entire industries. Mr. Kresa urged the Frontiers participants to be cognizant of the impact of rapid change on their careers, to persevere in pursuing their visions, and to have fun in their jobs. Mr. Kresa's remarks are contained in this volume.

As part of an ongoing process 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 with which they were not as familiar 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. Suggestions for the future included having smaller break-out groups with more focused topics, poster sessions, photos of all participants in the agenda book, and an alumni newsletter with "reunions" at several-year intervals.

Funding for the Fifth Annual Symposium on Frontiers of Engineering was provided by the Alcoa Foundation, the Defense Advanced Research Projects Agency, the U.S. Department of Defense (DDR&E-Research), the National Aeronautics and Space Administration, Parsons Brinckerhoff, Science Applications International Corporation, and the NAE Fund. The National Academy of Engineering would like to express its appreciation to these sponsors for supporting the symposium as well as to the members of the Symposium Organizing Committee (see p. *iv*) for their work in planning and organizing the event. A special expression of gratitude is due Robert H. Wagoner, Distinguished Professor, Department of Materials Science and Engineering, Ohio State University, who contributed greatly to this activity by serving as chairman of the organizing committee for the 1997–1999 Frontiers of Engineering symposia.

Contents

OPTICS

DINNER SPEECH

APPENDIXES

Drowning in Data

Magnetic Recording: Winner of the Data Storage Technology Race

THOMAS R. ALBRECHT
IBM Almaden Research Center
San Jose, California

A recent article in The Washington Post (Burgess, 1999) contains a number of interesting statements about hard disk drives (HDDs), including:

> Its insides would warm the heart of a Swiss watchmaker—tiny, finely crafted components whirring, reaching, spinning in choreographed precision, doing their duty year after year with razor-sharp accuracy and reliability. . . . And the technology keeps getting better and cheaper, despite perennial predictions of its ultimate demise. . . . The hard-disk industry claims to do better than the famous "Moore's Law" of microchip progress. That law says that chips are meant to double in performance roughly every 18 months; diskmakers say they manage that about every 12. The strange thing is that despite such accomplishments, the hard disk gets so little respect.

While the public is well aware of the great strides made in recent decades in the areas of microprocessor and memory performance, HDD technology has been advancing at an even faster pace in some respects (Figure 1). Microprocessors have shown a 30 percent increase in clock speed and a 45 percent increase in MIPS (million instructions per second) per year over two decades. Dynamic random access memory chips deliver per dollar 40 percent more capacity per year. Yet HDDs have offered 60 percent more storage per disk (and per dollar) each year over the last decade, and in the last three years, this pace has accelerated to 100 percent per year. Progress is so rapid that today the industry suffers from a drive capacity glut. While unit shipments of HDDs continue to increase, the number of platters per drive is falling, indicating that many customers no longer choose to purchase the largest available capacities. Another factor in this

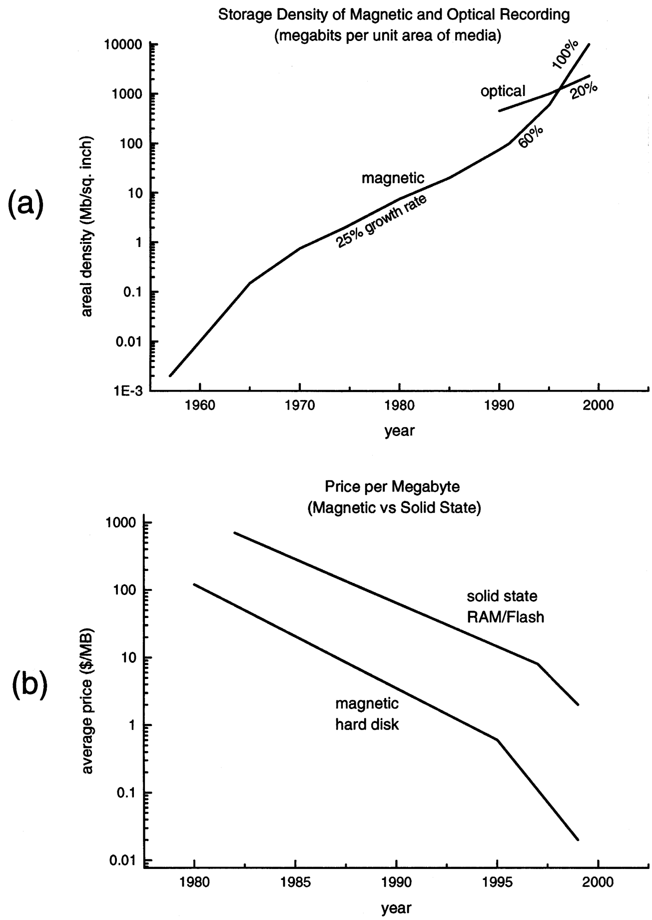

FIGURE 1 Comparison of magnetic data storage and competing technologies. (a) Until the 1990s, density of optical recording exceeded that of magnetic. Traditionally, magnetic densities have grown at 25 percent per year, increasing to 60 percent in the early 1990s with the introduction of anisotropic magnetoresistant (AMR) heads, and finally to 100 percent with giant magnetoresistant (GMR) heads in the last few years. Optical densities are growing at only 20 percent per year. (b) Although the rate of price reduction has accelerated for both magnetic and solid state technology in recent years, the price difference continues to widen. SOURCE: IBM.

trend is the replacement of large drives with arrays of smaller ones, which offers better performance, lower cost, and capability for fault tolerance without data loss.

The premature demise of magnetic recording as the leading storage technology has been predicted for decades. Projections from the past (Eschenfelder, 1970; Wildmann, 1974) typify the widely held perception during the 1970s and 1980s that engineering challenges would limit the achievable areal density of magnetic recording to levels much lower than could be achieved with optical recording. In fact, during the 1990s magnetic recording densities surpassed those of optical recording (Figure 1) and continue to improve at a faster pace than that of optical recording. After 50 years as the data storage medium of choice, magnetic recording, predominantly in the form of hard disk drives, remains the best choice for high capacity, high performance, and low cost. At the moment, more than 10^{19} bytes/year of HDD storage are being shipped. Yet even as magnetic recording is being pronounced the "winner" of the storage technology race, physical limitations that may slow or impede further progress are beginning to appear on the horizon.

HDDs are a combination of advanced technologies from several fields, including materials science, magnetics, signal processing, microfabrication, high-speed electronics, and mechanics. As the newspaper quote above mentions, HDDs are one of the few places in a computer where precision mechanics still plays a major role. The head-disk interface requires ultrasmooth disk surfaces (peak roughness less than 10 nm) with magnetic read/write elements riding on a slider (Figure 2) flying ~20 nm from the disk surface at velocities of up to 45 m/sec. Minimizing head-disk spacing is essential to maximizing areal density, while at the same time, head-disk physical contact must be avoided to prevent interface failure. Tight flying-height control is accomplished through the use of photolithographically defined multilevel air bearings that use combined positive and negative pressure regions to minimize sensitivity to interface velocity, atmospheric pressure, skew angle, and manufacturing tolerances. Although flying heights have decreased steadily over the years from ~20 μm to ~20 nm, further progress may require bringing the heads and disks fully into contact, which will require major technology breakthroughs to deal with friction and wear issues. Ever increasing areal density requires decreasing track widths, which are ~1 μm in today's products. Conventional rotary actuator and track-following servo technology appears to be approaching its limit. In the near future, HDDs are expected to incorporate two-stage actuators, with a conventional rotary actuator serving as a coarse head positioner, and a micromechanical actuator serving as a fine positioner. An example of a slider mounted on an electrostatic micromachined actuator (Fan et al., 1999) is shown in Figure 3.

Magnetic read/write elements have evolved from wire-wound ferrite inductive heads to microfabricated thin-film structures. Today, separate read and write elements are used, with a thin-film inductive element for writing and magnetoresistive element for reading. The quest for ever-increasing read head sensi-

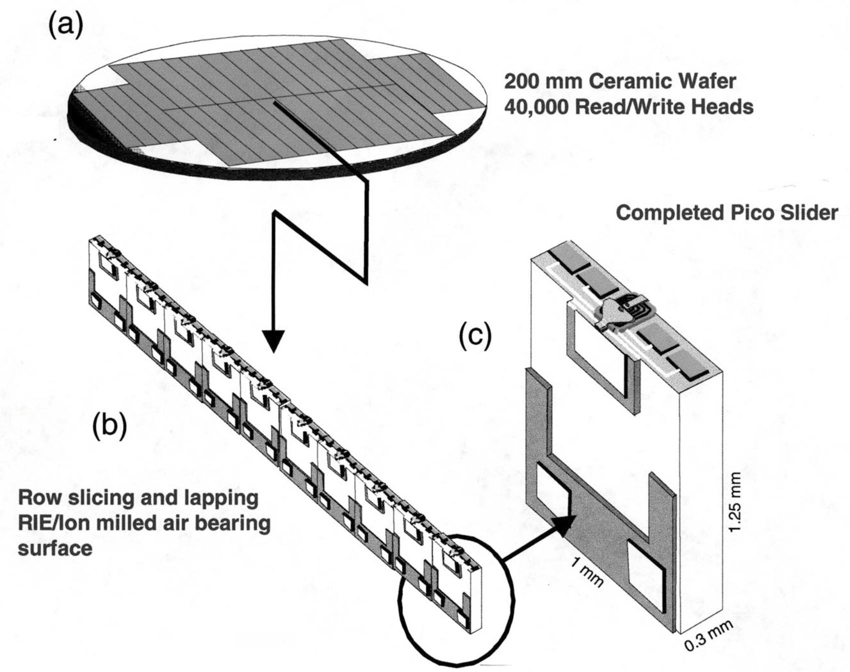

FIGURE 2 Fabrication of heads for HDDs. (a) Thin film inductive and GMR heads are batch fabricated on ceramic wafers. (b) Wafers are sliced into rows. At the row level, the air bearing surface is first lapped to tight flatness specifications, and then lithographically patterned and etched to create air bearing features. Air bearing features are typically 0.1–2 μm in height. (c) Finally, rows are parted into individual sliders 1 x 1.25 x 0.3 mm in size. SOURCE: IBM.

tivity to provide adequate signal-to-noise for reduced bit size has led to the adoption of read heads using the recently discovered giant magnetoresistance (GMR) effect, which has moved from discovery of a new physical effect (Baibich et al., 1988) to high-volume mass production of heads in just under 10 years. GMR heads incorporate a multilayer structure of magnetic and nonmagnetic thin films (Figure 4). The magnetization of a "pinned" layer is constrained by antiferromagnetic coupling to another layer. A thin nonmagnetic layer (e.g., 5 nm of Cu) separates the pinned layer from a second magnetic layer, whose magnetization is free to rotate under the influence of an external field (in this case, the fringing fields of magnetic bits written on the disk). Changes in the relative orientation of the magnetization in these two films result in changes in resistance on the order of 5–20 percent, an effect much larger than is achieved by the older anisotropic magnetoresistive (AMR) head technology. Although read elements have been the focus of recent advances in head technology, inductive write heads

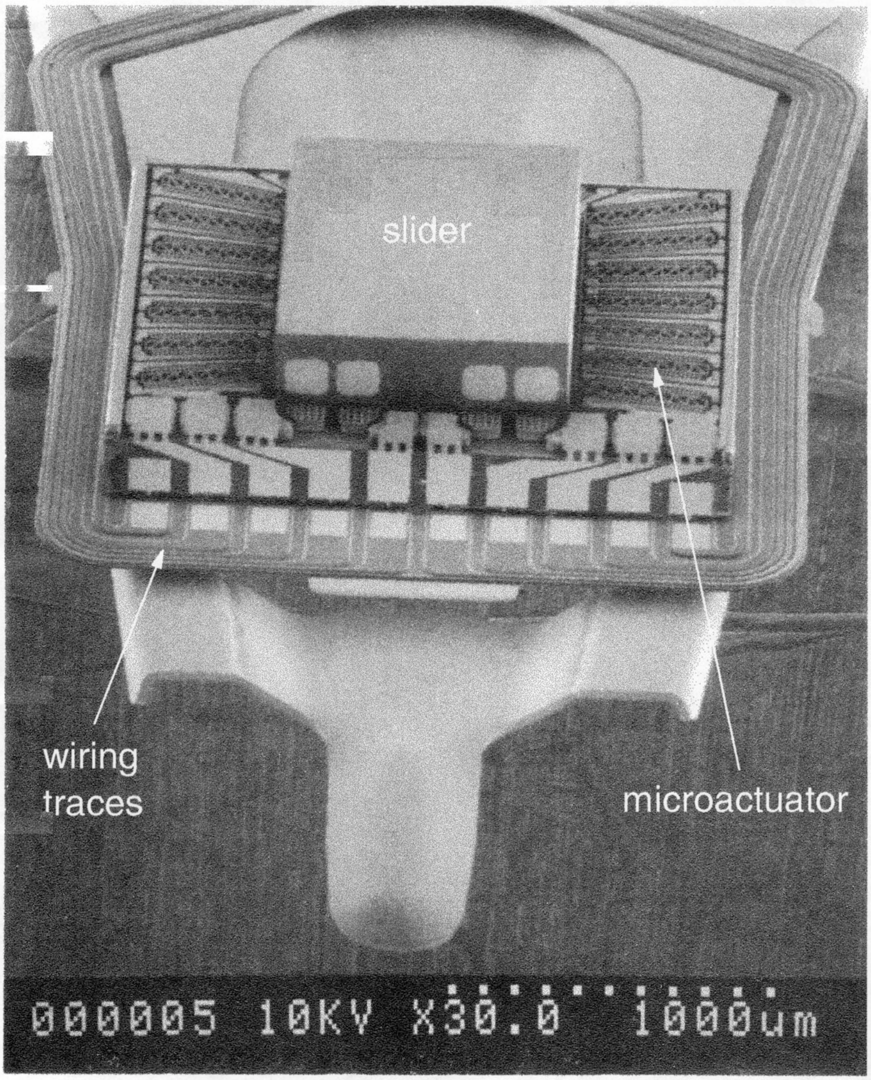

FIGURE 3 Slider (air bearing side up) mounted on a micromachined electrostatic secondary actuator on the end of a suspension. The microactuator serves as the fine positioning element for the track-following servo system of the HDD. SOURCE: IBM.

are also seeing a wave of innovation, with the implementation of higher moment materials, more compact coils, and better dimensional control to provide increased write field at higher speed in narrower tracks.

The greatest technical challenges facing the industry today are in the area of

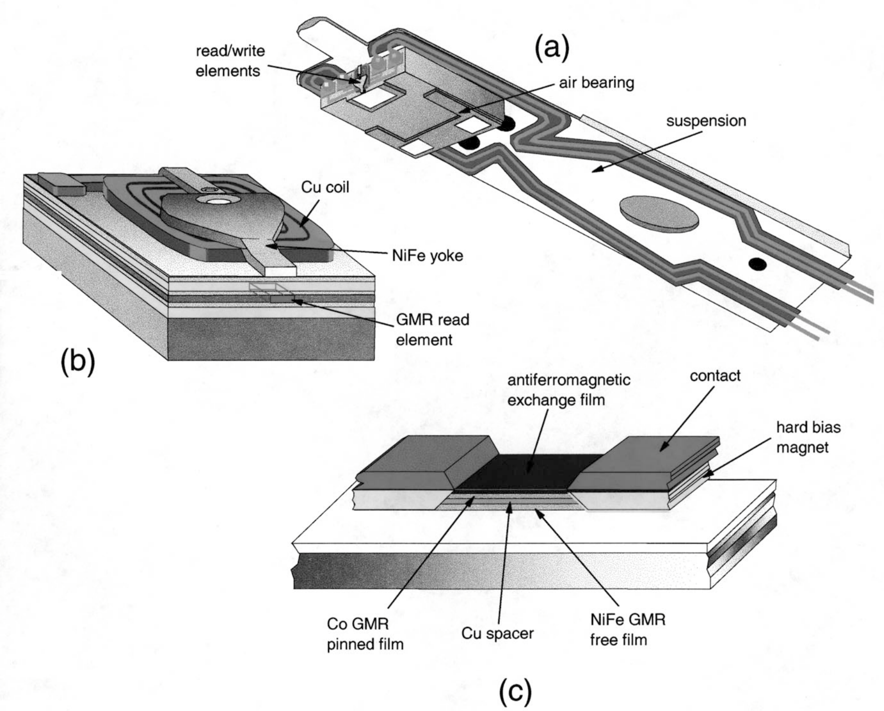

FIGURE 4 Magnetic read/write head. (a) The slider, with read/write elements on its rear surface facing downward toward the disk, is mounted on a spring-like suspension that holds the slider against the disk surface. An air bearing between the disk and slider surfaces prevents head-disk contact, while maintaining a ~20 nm separation between the head and disk. (b) The inductive write element is composed of Cu coils and a soft magnetic NiFe yoke, with poles facing the disk. (c) The GMR read element is a multilayer thin film structure, with pinned and free magnetic layers separated by a thin nonmagnetic Cu spacer layer. SOURCE: IBM.

disk materials and magnetics. Specifically, the size of recorded bits is approaching the superparamagnetic limit, where magnetization becomes thermally unstable. Disk media use granular cobalt alloy sputtered thin films, on the order of 15 nm thickness. Spontaneous changes in magnetization of individual grains can occur if the energy product K_uV (where K_u is the anisotropy energy of the material and V is the grain volume) is insufficiently large compared to the Boltzman energy k_BT (Weller and Moser, 1999). However, shrinking bit size requires shrinking grain size to maintain adequate signal-to-noise (Figure 5). Choosing materials with higher K_u poses problems in generating sufficiently large write fields (today's disk materials already require fields that cause magnetic saturation of write head poles). Thus, simple scaling of previous design points results in thermal

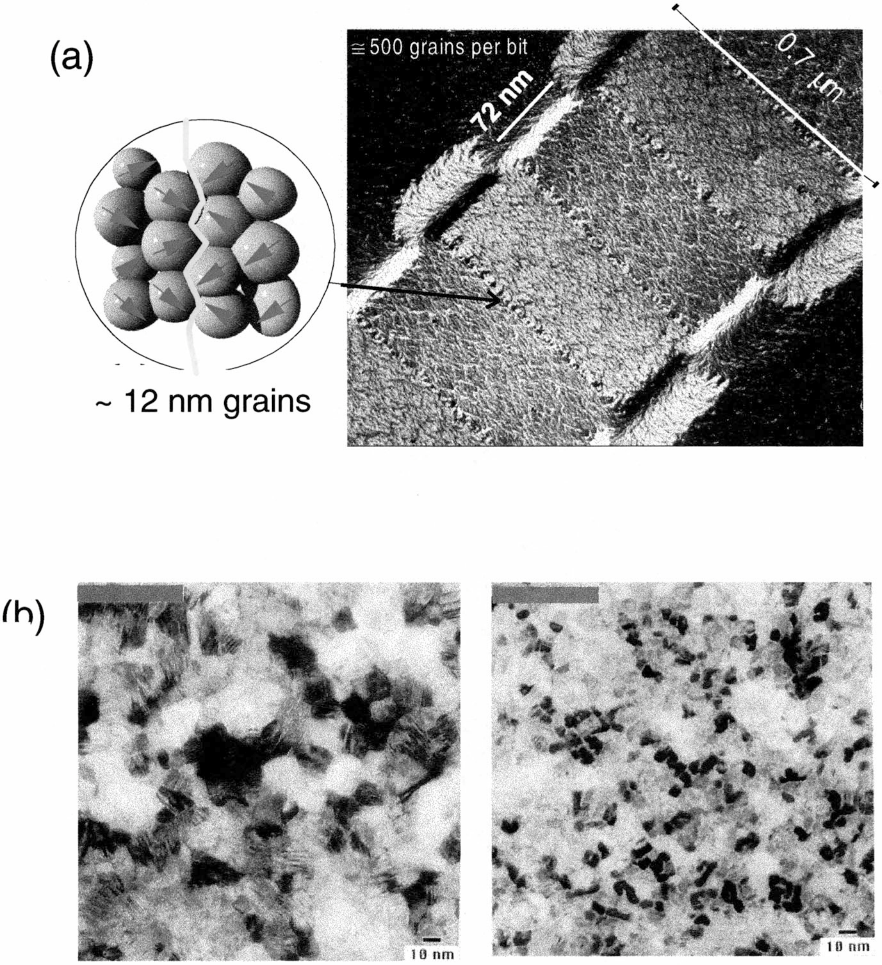

FIGURE 5 (a) Magnetic tracks are recorded on a granular thin-film magnetic medium on the disk. The finite size of grains in the medium results in "noisy" (not straight) boundaries between recorded bits. Achieving suitable disk signal-to-noise requires that grain size shrink as density is increased. Dimensions shown are for 10 Gbit/sq. inch density. The image of the track is a Lorentz microscopy image of a track recorded at lower density (1 Gb/sq. inch). (b) Choice of materials, underlayers, and sputtering conditions determines grain size. The medium on the right, with its smaller grains, can support a higher recording density; however, its thermal stability may be lower than the medium on the left. SOURCE: IBM.

instability (and lost data) when densities of 50-100 Gbit/sq. inch are reached, which is a design point only a few years away at the current rate of progress. To continue increasing density while maintaining thermal stability, some combination of the following will be used: perpendicular media and heads, reduced bit aspect ratio, thermally assisted writing, and enhanced error correction. Use of patterned media (one bit per pre-defined grain) offers the ultimate route to extremely high densities, although economical manufacturing strategies have not yet been shown.

Although there are many technical challenges, magnetic data storage is expected to remain a dominant player for years to come. As magnetic data storage reaches its 50th anniversary, it is branching out in new directions. Disk drives, once restricted to the domain of data storage for computers, are entering new markets. One example is the "microdrive," a tiny drive with a 1-inch disk for digital cameras and other handheld devices. Another example is the HDD video recorder, which may potentially revolutionize the television industry. Even the realm of solid-state memory may someday yield to magnetics: magnetic random access memory (MRAM) shows promise as a future replacement for solid-state flash memory (Parkin et al., 1999).

REFERENCES

Baibich, M. N., J. M. Broto, A. Fert, F. Nguyen Van Dau, and F. Petroff. 1988. Giant magnetoresistance of (001) Fe / (001) Cr magnetic superlattices. Physical Review Letters 61:2472–2474.

Burgess, J. 1999. Hard disks, which just keep improving, deserve more respect. The Washington Post (July 26):F23.

Eschenfelder, A. H. 1970. Promise of magneto-optic storage systems compared to conventional magnetic technology. Journal of Applied Physics 41:1372–1376.

Fan, L.-S., T. Hirano, J. Hong, P. R. Webb, W. H. Juan, W. Y. Lee, S. Chan, T. Semba, W. Imaino, T. S. Pan, S. Pattanaik, F. C. Lee, I. McFadyen, S. Arya, and R. Wood. 1999. Electrostatic microactuator and design considerations for HDD applications. IEEE Transactions on Magnetics 35:1000–1005.

Parkin, S. S. P., K. P. Roche, M. G. Samant, P. M. Rice, R. B. Beyers, R. E. Scheuerlein, E. J. O'Sullivan, S. L. Brown, J. Bucchigano, D. W. Abraham, Yu Lu, M. Rooks, P. L. Trouilloud, R. A. Wanner, and W. J. Gallagher. 1999. Exchange-biased magnetic tunnel junctions and application to nonvolatile magnetic random access memory. Journal of Applied Physics 85(8): 5828–5833.

Weller, D., and A. Moser. 1999. Thermal effect limits in ultrahigh-density magnetic recording. IEEE Transactions on Magnetics 35(6):4423–4439.

Wildmann, M. 1974. Mechanical limitations in magnetic recording. IEEE Transactions on Magnetics 10:509–514.

Evolution of Large Multiprocessor Servers

KOUROSH GHARACHORLOO
Western Research Laboratory
Compaq Computer Corporation
Palo Alto, California

ABSTRACT

A decade ago, large multiprocessors were part of a niche market and were primarily used to solve difficult scientific and engineering problems. Today, these multiprocessors are enjoying a phenomenal growth, and commercial databases and World Wide Web servers constitute the largest and fastest growing segment of the market. This evolution is continuing with the increasing popularity of the Web. For example, many experts believe that there will be a push toward large centralized servers used as permanent data repositories, providing users with easy access to their data from a wide range of devices connected to the Internet. These trends have made multiprocessor servers a key component in today's computing infrastructure.

Multiprocessor server design has gone through a corresponding evolution during this period. The most significant change has been a shift from message passing to cache-coherent shared-memory systems. The key advantage of a shared-memory system is that it allows efficient and transparent sharing of resources (e.g., memory, disk, network) among all processors. This feature greatly simplifies the task of application programmers by reducing (and sometimes eliminating) the need for resource partitioning. In addition, shared-memory naturally lends itself to a single system image (from an operating system perspective), which greatly simplifies the task of system management. However, preserving these benefits for larger scale systems presents a new set of technical challenges. One key challenge is to achieve software and operating system scalability while maintaining a single system image. Another challenge is in satisfying the reliability, availability, and serviceability requirements of commercial applications

(such as databases) in an environment where resources are inherently shared among all processors. Similarly, the different behavior of commercial applications relative to scientific/engineering applications is beginning to have a more significant influence on design trade-offs. Finally, the ever-increasing gap between memory latency and processor speed presents its own set of challenges.

This paper provides an overview of the evolution of multiprocessor servers during the last decade. In addition, there will be a discussion of the new technical challenges along with an overview of the types of solutions being pursued.

BRIEF HISTORY

A decade ago, shared-memory multiprocessors were confined to a few processors connected by a common bus, and most people considered message passing multiprocessors as the only viable platform for scaling to larger numbers of processors. In contrast to shared-memory systems, communication has to be specified explicitly in message-passing systems, and data placement is required for correctness as opposed to only affecting performance. Motivated by the inherent benefits of the shared-memory programming model, a large amount of academic research was aimed at improving the scalability of shared-memory systems during the late 1980s and early 1990s.

The Stanford DASH (Lenoski et al., 1992) is one of the seminal research projects that paved the path for larger shared-memory multiprocessors. The most important of the many contributions of the project was to demonstrate an efficient mechanism for maintaining cache coherence in larger shared-memory systems. The cache coherence problem arises because multiple processors are allowed to keep cached copies of the same memory location for efficient access. Instead of using broadcast to keep these copies up-to-date (as in bus-based shared-memory), DASH opted for a more scalable directory-based scheme that maintains the identity of individual sharers and allows point-to-point messages to be sent to each sharer. Research projects such as DASH have influenced a large number of commercially available systems that have made large-scale shared-memory multiprocessing a reality.

Ironically, while shared-memory research was primarily directed toward scientific and engineering applications, the commercial success of shared-memory systems has been mainly spurred by workloads such as databases and Web servers. Compared to the high-performance technical computing (HPTC) market, which is a small niche market with little growth, the commercial market is large and enjoys high growth due to the increasing popularity of the Web and the push toward centralized information servers. Meanwhile, the rate of acceptance of the shared-memory programming model in the HPTC market has been slower than anticipated. Reasons for this include the lack of commercially available shared-memory systems with thousands of processors (due to the practical considerations described), availability of a large body of message-passing math library codes, and the need for portability to non-shared-memory systems.

The dramatic change in the targeted market for shared-memory servers has naturally led system vendors to cater more to commercial applications. For example, system vendors have begun to evaluate trade-offs that arise in the design of processors and servers with commercial applications in mind. Another consequence is that hardware vendors are no longer pushing for scalability to several thousand processors, even though it is technically feasible. This is primarily because achieving application and system software scalability is a more difficult task for commercial workloads. Therefore, the bulk of the market is in small-scale servers (4–8 processors), with fewer customers requiring large-scale servers (64–128 processors).

LARGE MULTIPROCESSORS VERSUS CLUSTERS

There are a number of large applications that lend themselves to the use of clusters. For example, most Web applications are "embarrassingly parallel." Therefore, it is just as easy, and usually much cheaper, to scale a service by installing multiple independent systems. The most compelling benefits for multiprocessor servers arise from operational concerns; management overheads typically increase with the number of individual systems. At the same time, using more systems in a loosely clustered environment increases availability and allows for incremental capacity expansion. Hence, for some applications, there is a happy medium with a loosely coupled cluster of smaller multiprocessor servers reducing management overheads, and yet maintaining reliability and lower system cost (less premium relative to larger multiprocessor servers).

RECURRING THEMES

Even though computer systems have a relatively short history, one can identify a number of recurring themes. Consider the use of centralized services during the mainframe era when users were connected to computing resources through simple terminals. The move to microcomputers changed this trend to the extent that users became responsible for tasks such as managing, upgrading, and backing up their systems, that were previously transparent. In reaction to this, many organizations moved back to more centralized servers for data storage and resource-intensive computations. Furthermore, the emergence of the Web has led to numerous centralized information services, with Web browsers being the analog to the simple terminals of the mainframe era. With these developments, issues such as reliability, availability, and maintainability that were the backbone of mainframes have once again come to the forefront. Similarly, commercial workloads that once ran on mainframes have risen once more to become the primary target of high performance computing.

Such recurring themes lead some to believe that there is nothing new in computer architecture, with new designs primarily being an adaptation of existing ideas to current technology trends and market needs.

CURRENT CHALLENGES

Operating System and Application Scalability

One of the myths in computer system design is that software is easier to change than hardware. In reality, commercial software, such as operating systems and databases, is quite complex (consisting of millions of lines of code), making it a daunting task to change it in a major way. In comparison, hardware systems are much less complex and are designed to a simpler and more precise specification. This allows hardware to be redesigned every 18 months to 2 years, and making major changes is quite feasible as long as architectural compatibility is maintained.

Compared to scientific and engineering applications, commercial applications make it challenging for system software to scale to a large number of processors, due to their frequent use of operating system services and input/output devices (I/O). The process of scaling software to a larger number of processors typically involves reducing the amount of inherent synchronization and communication by employing new data structures and algorithms. These types of changes are difficult to make in operating systems and applications such as databases because of their inherent size and complexity.

Major strides have been made during the last few years with some database and operating system software currently scaling up to around 64 processors. Nevertheless, there is still much work to be done.

Reliability, Availability, and Serviceability

A large number of mission-critical commercial applications require guarantees of little or no downtime. Isolating and protecting against hardware and software faults is especially difficult when resources are transparently shared, since the faults can quickly propagate and corrupt other parts of the system. Furthermore, a number of applications require incremental upgrades or replacement of various hardware and software components while the system is running. Finally, a major fraction of the total cost of ownership in commercial servers is in management and maintenance costs, and system designs and software that can tackle these costs are highly desirable.

Hardware System Design

The ever increasing gap between processor and memory speeds also introduces several challenges in multiprocessor designs, since memory system performance becomes a more dominant factor in overall performance. Comparing the Stanford DASH (Lenoski et al., 1992) with aggressive next-generation multiprocessor systems, processor speeds will be over 30 times higher with memory latencies improving by only 10 times. Fortunately, memory bandwidths have improved faster than processor speeds during this period. Therefore, in addition to designing for low latency, it is important to exploit concurrency (supported by higher memory bandwidths) to tolerate the remaining latency.

Techniques such as out-of-order instruction execution that are used in part to help tolerate higher memory latencies and provide more instruction-level parallelism have led to major increases in design time, number of designers, and number of verification engineers. This problem is exacerbated by the higher levels of integration that are making it feasible for a single chip to have several hundred million transistors. Hence, there will likely need to be a major shift in design methodology to deal with the increasing complexities of future chip designs.

Commercial Applications

Commercial workloads have been shown to have dramatically different behavior compared to scientific and engineering workloads (Barroso et al., 1998; Keeton et al., 1998). Due to the fraction of time spent in the memory system and lack of instruction-level parallelism, there is a relatively small gain from improving integer processor performance (also, there is no floating-point computation). Commercial applications also make frequent use of operating system services and I/O, making the performance of system software more important.

The size and complexity of commercial applications, in addition to their frequent interaction with the operating system, also make it difficult to study and simulate the behavior of these workloads for designing next-generation systems. Nevertheless, there has been progress in the areas of scaling down commercial workloads to make them amenable to simulation (Barroso et al., 1998) and complete system simulation environments that include the behavior of the operating system (Rosenblum et al., 1995).

INTERESTING DEVELOPMENTS

Hardware Trends

With increasing chip densities, future microprocessor designs will be able to integrate many of the traditional system-level modules onto the same chip as the processor. For example, the next-generation Alpha 21364 plans to exploit aggressively such integration by including a 1GHz 21264 core, separate 64KByte instruction and data caches, a 1.5MByte second-level cache, memory controllers, coherence hardware, and network routers all on a single die. Such integration translates into a lower latency and higher bandwidth cache hierarchy and memory system. Another key benefit is lower system component counts, which leads to more reliable and lower cost systems that are easier to manufacture and maintain. Finally, this approach lends itself to modular designs with incremental scalability.

As mentioned, exploiting concurrency to tolerate memory latencies is also becoming increasingly important. There are two promising techniques in this area that exploit the inherent and explicit parallelism in workloads such as commercial applications for this purpose. The first technique, called simultaneous multithreading (SMT) (Tullsen et al., 1995; Lo et al., 1998), is applicable to wide-issue out-of-order processors. SMT involves issuing instructions from multiple threads in the same cycle to better use the many functional units in a wide-issue processor. The second technique, called chip multiprocessing (CMP) (Hammond et al., 1997), is motivated by higher integration densities and involves incorporating multiple (simpler) processor cores on the same chip. SMT and CMP are alternative mechanisms for exploiting explicit application-level parallelism to better use a given chip area.

Hardware and Software Faults

As multiprocessor servers scale to a larger number of processors, it becomes exceedingly important to isolate or hide both hardware and software failures. There are a number of options for handling faults with varying degrees of complexity and performance impact. At one extreme, the goal of fault tolerance is to completely hide the occurrence of failures by using redundancy. An alternative is to confine or isolate the effect of the fault to a small portion of the system. In the latter case, the chance of failure for a task should depend only on the amount of resources it uses and not on the size of the system. Therefore, while failures will be visible, they will be limited to tasks that used the failing resource. Compared to fault tolerance, fault containment can typically be achieved with much less cost and complexity and with minimal effect on performance.

Fault containment is well understood in distributed systems where communication occurs only through explicit messages, and incoming messages can be checked for consistency. However, the efficient resource sharing enabled by shared-memory servers allows the effect of faults to spread quickly and makes techniques used in distributed systems too expensive given the low latency of communication. To provide fault isolation, most current shared-memory multiprocessors depend on statically dividing the machine into "hard" partitions that cannot communicate, thus removing the benefits of resource sharing across the entire machine. More recent work on fault containment in the context of the Stanford FLASH multiprocessor allows for "soft" partitions with support for resource sharing across partitions (Teodosiu et al., 1997; Chapin et al., 1995). On the hardware side, this approach requires a set of features that limit the impact of faults and allow the system to run despite missing hardware resources. In addition, there is a recovery algorithm that restores normal operation after a hardware fault. For software faults, the approach involves logically partitioning the system into cells that act as "failure units" and using various firewall mechanisms to protect cells from each other. This type of solution has yet to appear in commercially available systems, partly because of the lack of sufficient hardware support in current designs and partly because restructuring a commodity operating system for fault containment is a challenge.

Achieving Software Scalability and Reliability Through Virtual Machine Monitors

A virtual machine monitor is an extra layer of software introduced between the hardware and the operating system. The monitor exports and virtualizes the hardware interfaces to multiple operating systems, referred to as virtual machines. Virtual machine monitors were prevalent in mainframes in the 1970s and were used to multiplex the expensive hardware among several operating systems and users. However, the use of this technique eventually faded away.

The Disco project at Stanford (Bugnion et al., 1997) has proposed to rejuvenate virtual machine monitors as an alternative to dealing with scalability and reliability requirements for large-scale shared-memory systems. As mentioned, restructuring a commodity operating system with millions of lines of code to achieve scalability and reliability is a difficult task. Fortunately, while operating systems and application programs continue to grow in size and complexity, the machine-level interface has remained fairly simple. Hence, the virtual machine monitor can be a small and highly optimized piece of software written with scalability and fault containment in mind. The Disco prototype consists of only 13,000 lines and uses several techniques to reduce the overheads (measured to be 5–15 percent) from running the extra layer of software.

Disco allows multiple, possibly different, commodity operating systems to be running at the same time as separate virtual machines. To achieve scalability, multiple instances of the same operating system can be launched. For handling applications whose resource needs exceed the scalability of the commodity operating system, Disco provides the ability to explicitly share memory regions across virtual machines. For example, a parallel database server can share its buffer cache among multiple virtual machines. Disco also provides the alternative of developing specialized operating systems for resource-intensive applications that do not need the full functionality of commodity operating systems. Such specialized operating systems can coexist with commodity operating systems as separate virtual machines.

Virtual machines also serve as the unit of fault containment. Hardware or software faults only affect the virtual machines that actually used the faulty resource. Disco also handles memory management issues that arise from nonuniform memory access by transparently doing page replication and migration. Again, changing commodity operating systems to do this would be more difficult. Finally, Disco inherits advantages of traditional virtual machine monitors: (a) older versions of system software can be kept around to provide a stable platform for running legacy applications, and (b) newer versions of operating systems can be staged in carefully with critical applications residing on older operating systems until newer versions have proven themselves.

SUMMARY

The emergence of the information age and the World Wide Web has had a major impact on the market for and design of high performance computers. Commercial workloads such as databases and Web servers have become the primary target of multiprocessor servers. Furthermore, the reliance on centralized information services has created demand for designs that provide high availability and incremental scalability. Hardware designs have reacted to these needs relatively quickly, with many of the current challenges in scalability and reliability residing on the software side.

REFERENCES

Barroso, L. A., K. Gharachorloo, and E. D. Bugnion. 1998. Memory system characterization of commercial workloads. Pp. 3–14 in Proceedings of the 25th International Symposium on Computer Architecture, Barcelona, Spain, June 27–July 1, 1998. Washington, D.C.: IEEE Computer Society Press.

Bugnion E., S. Devine, K. Govil, and M. Rosenblum. 1997. Disco: Running commodity operating systems on scalable multiprocessors. ACM Transactions on Computer Systems 15(4):412–447.

Chapin, J., M. Rosenblum, S. Devine, T. Lahiri, D. Teodosiu, and A. Gupta. 1995. Hive: Fault containment for shared-memory multiprocessors. Pp. 12–25 in Proceedings of the 15th ACM Symposium on Operating Systems Principles, Copper Mountain Resort, Colo. New York: Association for Computing Machinery.

Hammond, L., B. A. Nayfeh, and K. Olukotun. 1997. A single-chip multiprocessor. Computer 30(9):79–85.

Keeton, K., D. A. Patterson, Y. Q. He, R. C. Raphael, and W. E. Baker. 1998. Performance characterization of the quad Pentium Pro SMP using OLTP workloads. Pp. 15–26 in Proceedings of the 25th International Symposium on Computer Architecture, Barcelona, Spain, June 27–July 1, 1998. Washington, D.C.: IEEE Computer Society Press.

Lenoski, D., J. Laudon, K. Gharachorloo, W.-D. Weber, A. Gupta, J. Hennessy, M. Horowitz, and M. S. Lam. 1992. The Stanford DASH multiprocessor. Computer 25(3):63–79.

Lo, J. L., L. A. Barroso, S. J. Eggers, K. Gharachorloo, H. M. Levy, and S. S. Parekh. 1998. An analysis of database workload performance on simultaneous multithreaded processors. Pp. 39–50 in Proceedings of the 25th International Symposium on Computer Architecture, Barcelona, Spain, June 27–July 1, 1998. Washington, D.C.: IEEE Computer Society Press.

Rosenblum, M., S. A. Herrod, E. Witchel, and A. Gupta. 1995. Complete computer system simulation: The SimOS approach. IEEE Parallel and Distributed Technology 3(4):34–43.

Teodosiu, D., J. Baxter, K. Govil, J. Chapin, M. Rosenblum, and M. Horowitz. 1997. Hardware fault containment in scalable shared-memory multiprocessors. Pp. 73–84 in Proceedings of the 24th International Symposium on Computer Architecture, Denver, Colo., June 2–4, 1997. Washington, D.C.: IEEE Computer Society Press.

Tullsen, D., S. Eggers, and H. Levy. 1995. Simultaneous multithreading: Maximizing on-chip parallelism. Pp. 392–403 in Proceedings of the 22nd International Symposium on Computer Architecture, Santa Margherita Ligure, Italy, June 18–24, 1995. Washington, D.C.: IEEE Computer Society Press.

Network Survivability and Information Warfare

MICHAEL K. REITER
Bell Laboratories, Lucent Technologies
Murray Hill, New Jersey

Perhaps the primary reason that we are "drowning in data" is the explosion of network connectivity in the last five years. Networked information systems (NISs) make accessible an unprecedented wealth of information and possibilities for communication, and do so in a very cost-efficient manner. We seem capable of deploying NISs with ever increasing capacities for transmitting, storing, and processing data, and the undeniable advantages of "getting connected" are drawing a broad array of applications—some of them life critical or financially critical—to be conducted or managed with the help of large NISs.

At the same time we seem embarrassingly incapable of deploying NISs that can withstand adversity and protect the data and services they manage. This was cogently displayed in a 1997 operation named "Eligible Receiver," in which a National Security Agency team demonstrated how to break into U.S. Department of Defense and electric power grid systems from the Internet; generate a series of rolling power outages and 911 overloads in Washington, D.C., and other cities; break into unclassified systems at four regional military commands and the National Military Command Center; and gain supervisory-level access to 36 networks, enabling e-mail and telephone service disruptions (NRC, 1999). Over a period of a year during the Gulf War, five hackers from the Netherlands penetrated computer systems at 34 military sites, gaining the ability to manipulate military supply systems and obtaining information about the exact locations of U.S. troops, the types of weapons they had, the capabilities of the Patriot missile, and the movement of U.S. warships in the Gulf region (Denning, 1999). During 1999, the United States military has been under a sophisticated and largely successful network attack originating from Russia, in which even top secret systems have been compromised (Campbell, 1999). The very NISs that are drowning us in data are also leaving us vulnerable to attacks that can take place from a distance, that can be conducted anonymously, and that can disable or degrade critical systems and national infrastructures. This is cause for deep concern in information warfare circles, as numerous countries and terrorist organizations are reported to be developing ways to exploit these vulnerabilities offensively.

The present state of affairs has led to increased research in areas such as intrusion detection and response, and "penetration tolerant" or "survivable" NIS technologies. The breadth and complexity of the problems that must be addressed renders it impossible to cover them here; the interested reader is referred to a recent National Research Council (NRC) study that proposes research directions to address some of the issues involved (NRC, 1999). For the purposes of the present discussion, however, we call attention to one research direction advocated in this study, namely the construction of trustworthy (survivable or penetration-tolerant) systems from untrustworthy components. The basic idea is to construct an NIS using replication, redundancy, and diversity, so that even if some of its components are successfully attacked, the NIS will nevertheless continue to provide critical service correctly. This notion has analogs in a broad range of engineering disciplines. In avionics, for example, a technique known as Triple Modular Redundancy (TMR) is commonly used to mask the failure of a processor by having three processors redundantly perform a computation and "vote" on the result. In this way, the result that occurs in a majority "wins," thereby hiding any effects of a single faulty processor. While a well-understood technique in the confines of an embedded system, these ideas are largely untested in larger, open settings in order to mask the misbehavior of data servers (e.g., file servers, name servers), network routers, or other NIS components.

Most work on applying these principles to build survivable data services has attempted to duplicate the TMR approach in networked settings, where it has come to be known as "state machine replication." These systems consist of an ensemble of closely coupled computers that all respond to each client request. Again, the correct answer of the ensemble of servers is determined by voting. Even this seemingly simple extension of TMR to network settings introduces complexities, however. First, concurrent requests from different clients, if processed in different orders at different servers, can cause the states of correct servers to diverge, so that even correct servers return different answers to each request. Most systems thus implement protocols to ensure that all correct servers process the same requests in the same order—a task that is theoretically impossible to achieve and simultaneously guarantee that the system makes progress (Fischer et al., 1985). Therefore, systems that adopt this approach employ request delivery protocols that impose additional timing assumptions on the network (e.g., Pittelli and Garcia-Molina, 1989; Shrivastava et al., 1992; Cristian et al., 1995) or are guaranteed to make progress only conditionally (e.g., Reiter, 1996; Kihlstrom et al., 1998; Castro and Liskov, 1999). Second, the use of every server to process every request precludes dividing server load among different servers. That is, adding more servers does not yield better performance; on the contrary, it tends to degrade performance, due to the additional obligations it imposes on the request ordering and output voting protocols. These properties, and the expense of the request ordering protocols themselves, thus seem to limit this approach to moderate replication of servers on local area networks only.

A more scalable approach presently being explored is the adaptation of quorum systems to the task of providing survivable services tolerant of server penetrations (Malkhi and Reiter, 1998a). In this approach, clients perform data operations only at a subset, called a quorum, of the servers. Quorums are constructed so that, for example, any two quorums intersect in $2t + 1$ servers, where t is a presumed maximum number of server penetrations that may occur. Using this property, it is possible to build access protocols that emulate a wide range of data objects despite the corruption of up to t data servers (Malkhi and Reiter, 1998b). An advantage of this approach over state machine replication is that quorums can be surprisingly small, e.g., $O(\sqrt{tn})$ servers out of a total of n servers. Consequently, this approach offers improved scaling and load balancing, so much so that it was recently employed in the design of a survivable nationwide electronic voting trial for the country of Costa Rica that was required to scale to hundreds of electronic polling stations spread across the country. There are limitations even to this approach, however. First, it complicates the task of detecting penetrated servers, and thus demands new mechanisms for doing so (Alvisi et al., 1999). Second, it is subject to a fundamental trade-off between availability and an ability to balance load among the servers; research is investigating techniques to break this trade-off by allowing a small and controlled probability of error in data operations (Malkhi et al., 1997). Third, in general, a new data access protocol must be designed to emulate each different data object, but on the other hand, many useful objects can be emulated in this environment without the limitations imposed by the impossibility result of Fischer et al. (1985).

Finally, there remain basic limitations to what can be achieved with any attempt to build highly resilient data services from untrustworthy components. While effective against attacks that an attacker cannot readily duplicate at all servers (e.g., attacks exploiting configuration errors or platform-specific vulnerabilities in particular servers, or physical capture or server administrator corruption), these approaches provide little protection against attacks that can simultaneously penetrate servers with little extra incremental cost per penetration to the attacker. Research in adding artificial diversity to servers, in the hopes of eliminating common vulnerabilities among servers, is underway, but it is in a fledgling state.

REFERENCES

Alvisi, L., D. Malkhi, L. Pierce, and M. K. Reiter. 1999. Fault detection for Byzantine quorum systems. Pp. 357–372 in Proceedings of the 7th IFIP Working Conference on Dependable Computing for Critical Applications, San Jose, Calif., January 6–8, 1999. Los Alamitos, Calif: IEEE Computer Society.

Campbell, M. 1999. U.S. losing cyber war to Russian hackers. The Australian, July 26.

Castro, M., and B. Liskov. 1999. Practical Byzantine fault tolerance. Paper presented at the 3rd Symposium on Operating Systems Design and Implementation, New Orleans, La., February 22–25, 1999.

Cristian, F., H. Aghili, R. Strong, and D. Dolev. 1995. Atomic broadcast: From simple message diffusion to Byzantine agreement. Information and Computation 18(1):158–179.

Denning, D. E. 1999. Information Warfare and Security. New York: ACM Press.

Fischer, M. J., N. A. Lynch, and M. S. Paterson. 1985. Impossibility of distributed consensus with one faulty process. Journal of the ACM 32(2):374–382.

Kihlstrom, K. P., L. E. Moser, and P. M. Melliar-Smith. 1998. The SecureRing protocols for securing group communication. Pp. 317–326 in Proceedings of the 31st Annual Hawaii International Conference on System Sciences, Vol. III, R. H. Sprague, Jr., ed. North Hollywood, Calif.: Western Periodicals Co.

Malkhi, D., and M. K. Reiter. 1998a. Byzantine quorum systems. Distributed Computing 11(4):203–213.

Malkhi, D., and M. K. Reiter. 1998b. Secure and scalable replication in Phalanx. Pp. 51–58 in Proceedings of the 17th IEEE Symposium on Reliable Distributed Systems, West Lafayette, Ind., October 20–23, 1998. Los Alamitos, Calif.: IEEE Computer Society.

Malkhi, D., M. K. Reiter, and R. Wright. 1997. Probabilistic quorum systems. Pp. 267–273 in Proceedings of the 16th ACM Symposium on Principles of Distributed Computing, Santa Barbara, Calif., August 21–24, 1997. New York: Association for Computing Machinery.

NRC (National Research Council). 1999. Trust in Cyberspace, F. B. Schneider, ed. Washington, D.C.: National Academy Press.

Pittelli, F. M., and H. Garcia-Molina. 1989. Reliable scheduling in a TMR database system. ACM Transactions on Computer Systems 7(1):25–60.

Reiter, M. K. 1996. Distributing trust with the Rampart toolkit. Communications of the ACM 39(4):71–74.

Shrivastava, S. K., P. D. Ezhilchelvan, N. A. Speirs, S. Tao, and A. Tully. 1992. Principal features of the VOLTAN family of reliable node architectures for distributed systems. IEEE Transactions on Computers 41(5):542–549.

Moving up the Information Food Chain: The Future of Web Search

OREN ETZIONI
Go2net, Inc., Seattle
and
Department of Computer Science and Engineering
University of Washington, Seattle

The World Wide Web is at the very bottom of the Information Food Chain. The Yahoos and Alta Vistas of the world are information herbivores, which graze on Web pages and regurgitate them as directories and searchable collections. The talk focuses on information carnivores, which hunt and feast on Web herbivores. Meta-search engines, shopbots, and the future of Web search are also considered.

URLs to visit include:

Meta-search:	www.metacrawler.com
Shopbot:	jango.excite.com
Concierge for the web:	www.askjeeves.com
Clustering search:	www.cs.washington.edu/research/clustering

Making Sense of the Human Genome

Genes, Chips, and the Human Genome

Stephen P. A. Fodor
Affymetrix, Inc.
Santa Clara, California

Photolithography and combinatorial chemical synthesis have been used to generate miniaturized, high-density arrays of oligonucleotide probes. These probe arrays are then used for parallel nucleic acid hybridization analysis, directly yielding high-information-content sequence data. Implementation of the DNA array technology has required the integration of multiple technical disciplines resulting in the development of fabrication methods for the probe arrays, assays for the detection of target hybridization, algorithms to analyze the hybridization data, and platform instruments to support the technology. Applications of the arrays include the analysis of genetic mutations, the simultaneous expression profiling of thousands of genes, a new method to quickly discover polymorphisms of the human genome, and a new tool for genomic mapping. The talk reviews specific oligonucleotide probe array designs and paradigm experiments.

ENGINEERING NOVEL STRUCTURES

Colloidal-Scale Engineering

Eric W. Kaler
Center for Molecular and Engineering Thermodynamics
Department of Chemical Engineering
University of Delaware
Newark, Delaware

Recent advances have made it possible to apply engineering principles to the control of processes on size scales ranging from nano- to micrometers. These are now the scales where the modern technologies of biology and electronics operate, but they have long been the domain ruled by the old discipline of colloid science. In this range the interfacial properties of a fluid or substrate become important, fluid flows are always laminar, and mass and heat transfer can become controlled by surface tension forces. Here processes of mixing and chemical reaction depend on "moving nanoliters nanometers," and the successful control of these processes results in devices of revolutionary size, speed, and selectivity. In this regime grinding and milling to make small particles no longer operate reliably, and much effort must be focused instead on self-assembly to synthesize new materials with spatially ordered features on the nanometer scale. These materials have present and potential applications in optical information processing and storage, advanced coatings, and catalysis. The intellectual driving force for this process results in part from efforts to mimic hard natural materials, such as bone and shell, and soft high-strength materials like spider silk; but, current efforts fall far short of producing the structural complexity of nature, and utterly fail to mimic its full function.

A common approach to the production of nano-scale structures relies on templates that assemble into desired patterns and then can be replicated by a harder, more durable material. Surfactant molecules spontaneously assemble into a variety of useful structures and are commonly used as the template for materials with feature dimensions of a few nanometers, such as zeolites (Kresge et al., 1992). Larger templates are needed for materials with larger dimensions, and polymer beads have recently been used as templates. Surfactant molecules again play an important role in this process, because they can modify and control the growth of polymeric materials in an aqueous environment, such as in the process of emulsion polymerization. The possibility of using self-assembled surfactant structures to prepare polymers with novel supramolecular architectures has long tempted researchers, and polymerization in these complex fluids offers the potential to prepare polymers with morphologies precisely controlled on the length scale of nanometers. The variety of polymer morphologies potentially accessible by this approach include spherical latex particles in size ranges inaccessible to traditional emulsion polymerization techniques (radii as low as 7 nm have been reported), as well as highly porous materials with connected open channels some tens of nanometers in diameter. A roadblock to production of designed polymer structures has been the lack of a clear mechanistic picture of the microscopic events governing the formation of polymer, but recently the basic principles have been outlined (Morgan and Kaler, 1998). Larger (50–1000 nm) polymer beads are articles of commerce, and their surfaces can be functionalized with a variety of bioactive moieties.

Once these colloidal particles are in hand, the next challenge is to assemble them into larger scale superstructures. A promising class of such materials obtained by self-assembly is the colloidal crystals, ordered arrays of particles in the nanometer- to micrometer-size range. However, because the materials obtained after colloidal crystals are dried are brittle and can be redispersed in water, replication is necessary. In addition, the size and shape of the colloidal crystal needs to be controlled on a macroscopic scale. Several ways have been developed to control the organization of colloidal assemblies by using flow, surface tension forces, or electrical fields.

In the first example colloidal crystals are used to template a unique new material, nanostructured porous gold (Velev et al., 1997, 1999). The metallic structure is assembled on the surface of a filter membrane from nanometer-sized gold particles that are templated by colloidal crystals of larger latex microspheres that have been formed by flow over the filter. When the templates are removed chemically or by calcination they leave behind a three-dimensional metallic nanostructure with long-ranged ordering of the pores. The pore size is precisely controlled in the sub-micrometer range by the diameter of the latex microspheres.

The above procedure is shown in Figure 1. The templates are assembled from monodisperse, negatively charged polystyrene latex microspheres ranging in diameter from 300 to 1000 nm, with the colloidal crystals of interest formed by concentrating the particles in the vicinity of a membrane surface by filtration. Next, 15–25 nm colloidal gold particles are deposited in the cavities of the latex crystals. The colloidal gold fills the interstices in the latex arrays, forming a gold structure that is by itself porous on a smaller mesoscopic scale. This porosity allows the solvent to flow through the deposit until the pores are completely filled by the deposited gold colloid. After removal of the latex beads the gold superstructure appears as in Figure 2. The lace-like structure shown in Figure 2

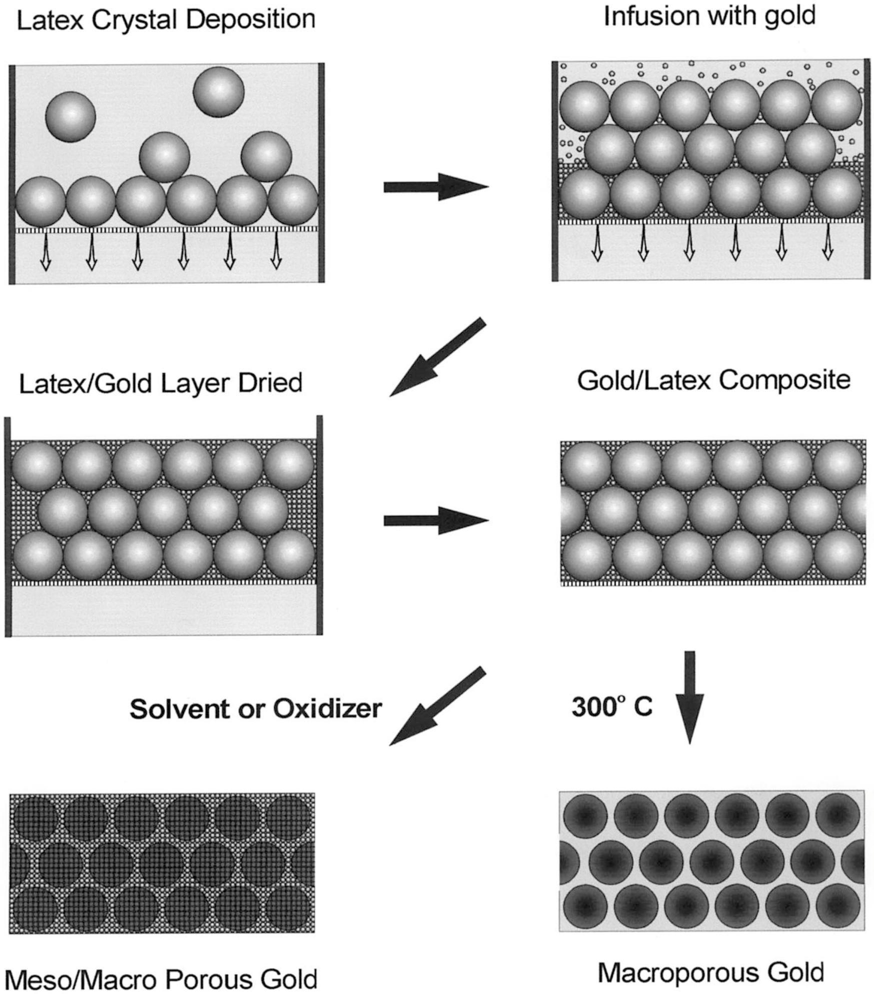

FIGURE 1 Schematic presentation of the method of assembly of polymer-templated gold.

is an almost exact replica of the three-dimensional wire-mesh photonic crystals described by Sievenpiper et al. (1996) but *scaled down by a factor of 20,000 to the sub-micrometer region*. The wire mesh photonic structure has been demonstrated to have an interesting combination of forbidden bands in the gigahertz region. As the photonic properties scale with array size, miniaturizing the structure should shift these properties to the infrared-visible region, which is of significant interest for optoelectronic devices.

Templating, therefore, guides the synthesis of new materials, but as yet

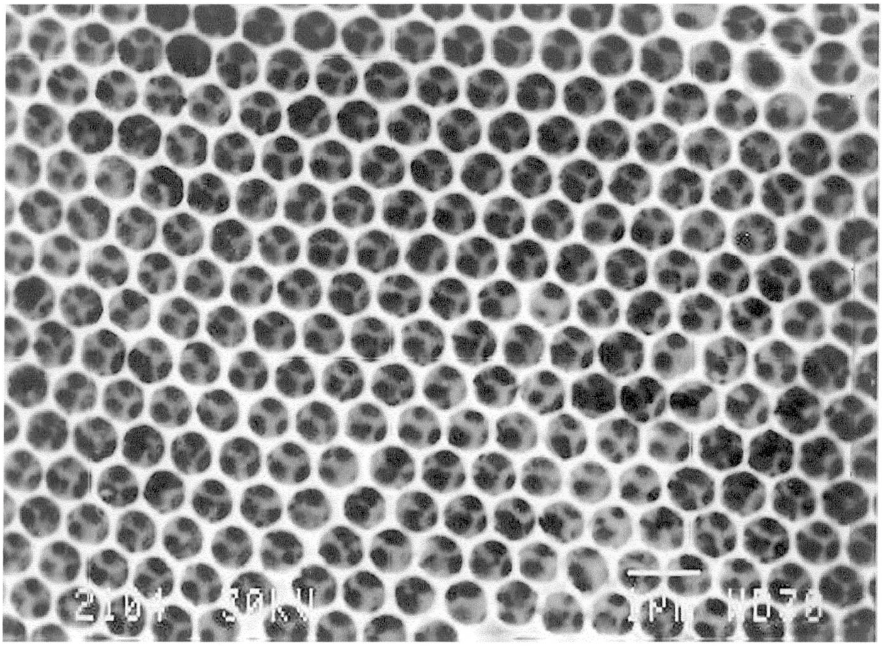

FIGURE 2 A representative photograph of a gold sample after dissolution of the latex. The highly ordered, lace-like gold structure is an analogue of a 3D wire-mesh photonic crystal scaled down to the sub-micrometer region. SOURCE: Adapted from Velev et al. (1999). Reprinted with permission from *Nature*.

there is a lack of simple and controllable methods for extraction of colloidal crystals from their aqueous environment and mounting or shaping them into usable solid objects. A step towards producing and processing colloidal crystalline materials is assembly of spherical or globular crystalline structures (microballs) in aqueous droplets suspended on the surface of a heavier liquid immiscible with water. Here, surface tension forces can be controlled to guide the formation of the particles. The colloidal particles in the suspended template droplets are gradually concentrated by drying to the point of transition to an ordered state. The tangential mobility of the droplet surfaces provides a "no-stick" substrate, where the particles are free to move, rotate, and rearrange into defect-free lattices. After complete evaporation of the water, highly ordered and symmetric composite particles with smooth surfaces are obtained. Control of the interplay between gravity and interfacial tensions allows control of the shape of the template droplets and their assemblies.

Colloidal self-assembly also is useful in the synthesis of devices in which electronic microchips are created out of, and interface with, fluid-borne colloidal and biological systems. Lithographically patterned substrates provide a well-

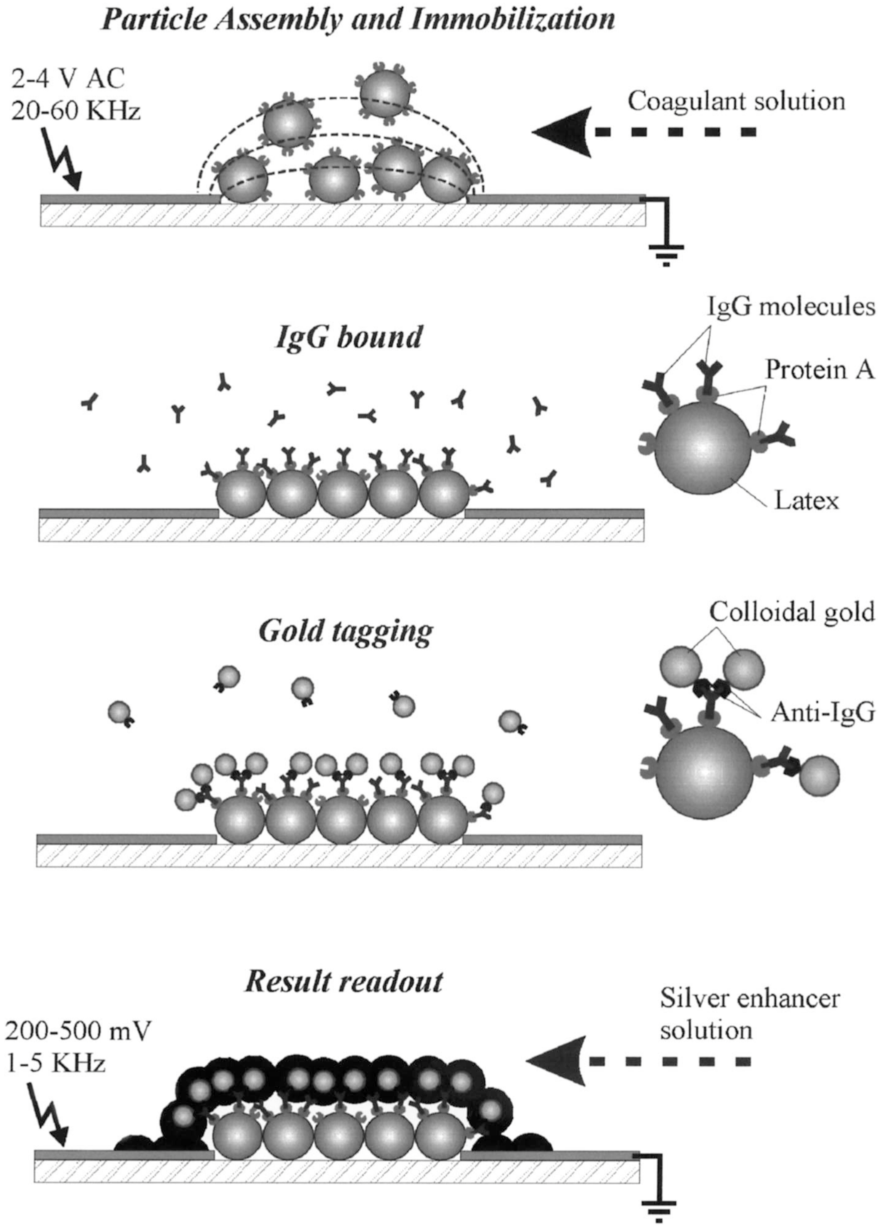

FIGURE 3 Schematics of the main stages of the assembly and operation of the sensor. The procedure is illustrated by an immunoglobulin test. SOURCE: Reprinted with permission from Velev and Kaler (1999). Copyright 1999 American Chemical Society.

defined, organized template for colloidal assembly and electrical manipulation of colloidal systems and living cells on a micron scale. Using such a substrate, we have developed a method for assembling microscopic sensor patches in situ from the same latex particles used in traditional agglutination assays (Velev and Kaler, 1999). The glass substrates for our sensors carry photolithographically fabricated gold patterns that form addressable electrodes of micron size with small gaps between them. To generate the active area of the sensors, the particles are collected in the gaps by dielectrophoresis and then stabilized. The assembly procedure can be repeated many times by exchanging the particle suspension and addressing different gaps so that different sensors can be assembled on the same "chip."

A schematic of the main stages involved in the preparation and read-out of the biosensors is shown in Figure 3. The limit of detection of the sensors is about $5x10^{-14}$ moles, which is comparable to the better IgG agglutination assays and immunosensors currently available. We believe this level can be reduced dramatically.

ACKNOWLEDGMENTS

This work is the result of a collaboration with O. D. Velev and A. M. Lenhoff.

REFERENCES

Kresge, C. T., M. E. Leonowicz, W. J. Roth, J. C. Vartuli, and J. S. Roth. 1992. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. Nature 359:710–712.

Morgan, J. D., and E. W. Kaler. 1998. Particle size and monomer partitioning in microemulsion polymerization. 1. Calculation of the particle size distribution. Macromolecules 31(10):3197–3202.

Sievenpiper, D. F., M. E. Sickmiller, and E. Yablonovitch. 1996. 3D wire mesh photonic crystals. Physical Review Letters 76(14):2480–2483.

Velev, O. D., T. A. Jede, R. F. Lobo, and A. M. Lenhoff. 1997. Porous silica via colloidal crystallization. Nature 389:447–448.

Velev, O. D., P. M. Tessier, A. M. Lenhoff, and E. W. Kaler. 1999. Nanostructured photonic metal synthesized via colloidal crystal templates. Nature 401:548.

Velev, O. D., and E. W. Kaler. 1999. In situ assembly of colloidal particles into miniaturized biosensors. Langmuir 15(11):3693–3698.

Design of Biomimetic Polymeric Materials

ARUP K. CHAKRABORTY
Departments of Chemical Engineering and Chemistry
University of California
Berkeley, California

INTRODUCTION

Recent technological advances are demanding materials that can carry out functions with high specificity, but how can we design materials from the molecular scale up so that they are able to perform these functions? One class of materials that offers potential for use in such advanced applications is polymers, large macromolecules made up of a connected set of monomers. Monomers can be of different types, and the sequence in which they are connected as well as the nature of the connections (i.e., the molecular architecture) can vary. Polymers are strong candidates for use in applications where the ability of a material to perform functions with specificity is crucial because we know that nature uses polymers (e.g., proteins and nucleic acids) the same way. Evolution has allowed nature to devise schemes that allow the design of macromolecular building blocks that can self-assemble into functionally interesting structures. Thus, one way to devise synthetic systems would be to take lessons from nature. This does not mean that the detailed chemistries of nature should be copied. Rather, we should search for underlying universalities in the schemes that natural systems employ in order to carry out a class of functions, and then should explore whether the universal schemes (if they exist) can affect biomimetic behavior. The interest in generic schemes has developed because they may be easier to implement in abiotic applications. Elucidating the universal schemes also provides insight into the underlying phenomena that effect self-assembly processes, and thus allows a specific class of functions to occur.

It is clear that to engineer macromolecules to function in specific ways a synergistic effort combining synthesis, physical experimentation, and theory is required. In this presentation, we will focus on how sophisticated theoretical and computational methods can be employed with experimental work to create synthetic macromolecular systems that can perform functions with biomimetic specificity. Two examples that illustrate the general ideas are: 1) biomimetic recognition between polymers and surfaces and 2) molecular architecture that can control surfactant properties of macromolecules with high sensitivity.

BIOMIMETIC RECOGNITION BETWEEN POLYMERS AND SURFACES

Many biological processes such as transmembrane signaling and pathogen-host interactions are initiated by a protein when it recognizes a specific pattern of binding sites on part of a membrane or cell surface. Recognition means that the polymer quickly finds and then adsorbs strongly on the pattern-matched region and not on others. The development of synthetic systems that can mimic such recognition between polymers and surfaces could have a significant impact on such advanced applications as the development of sensors, molecular scale separation processes, and synthetic viral inhibition agents (Sigal et al., 1996; Todd et al., 1994). Attempting to affect recognition in synthetic systems by copying nature's detailed chemistries does not seem practical for most applications. This leads to the question of whether there are any universal strategies that can affect recognition between polymers and surfaces. To deduce candidates for such strategies, some generalized observations regarding biological systems were made. Proteins are composed of many types of monomers, and have sequences that are not periodically repeating (Chan and Dill, 1991; Dewey, 1997; Irbäck et al., 1996; Pande et al., 1994). Similarly, the pattern of different binding sites on cell surfaces is not periodically repeating. These observations lead to the question of whether competing interactions (due to preferential interactions between different types of monomers and surface sites) and disorder are essential ingredients for biomimetic recognition in synthetic systems.

Disordered heteropolymers (DHPs) are synthetic macromolecules with chemically different monomer units distributed along the backbone in a disordered sequence that is described statistically.[1] Like proteins, they are composed of different monomers, with sequences that are aperiodic and quenched (the sequences cannot change in response to the environment). Unlike proteins, which carry a specific pattern encoded in their sequence distribution, DHP sequence distributions correspond to statistical patterns. (The meaning of statistical patterns is discussed below.) Imagine a DHP chain interacting with a surface bearing many types of sites that are distributed in a statistically described manner.

[1] If a number of DHPs made up of the same monomers are synthesized in a reactor, each DHP chain will have a different sequence. However, every such sequence belongs to the same statistical distribution determined by the reactivity ratios (Odian, 1991)

Such surfaces resemble cell surfaces wherein the pattern of binding sites is not periodically ordered. They are different from cell surfaces in that they bear statistical rather than specific patterns. Let the DHP segments exhibit preferential interactions with the surface sites (i.e., a particular type of segment prefers to interact with a certain type of surface site, and the other types of segments exhibit a different preference). Studying such a system allows an exploration of whether competing interactions and quenched disorder are sufficient ingredients for biomimetic recognition.

Our recent theoretical and computational studies (Bratko et al., 1997; Golumbfskie et al., 1999; Srebnik et al., 1996) suggest the occurrence of a phenomenon akin to recognition when the statistics characterizing the DHP sequence and that of the surface site distribution are related in a special way (i.e., matched). Specifically, these studies suggest that frustration (due to the competing segment-surface interactions and quenched disorder) and statistical pattern matching lead to one of the hallmarks of recognition, a sharp discrimination between regions of a surface to which a given type of DHP sequence binds strongly and those onto which it does not adsorb. In other words, statistical pattern matching is sufficient for biomimetic recognition to occur, provided the statistical patterns are designed properly. Our results show how a mixture of DHPs bearing different statistical patterns can be separated on a surface by discriminatory adsorption onto regions that bear complementary statistical patterns. The results also show dramatic differences in chain dynamics in the "wrong" and pattern-matched regions of the surface; these differences shed light on the kinetic behavior of frustrated systems, or systems that include biological macromolecules.

Some specific results can be illustrated by considering a surface comprised of four quarters, each of which bears a distribution of sites of two types (red and yellow) on a neutral background. Such an arrangement is shown in Figure 1. The top right quarter is such that in a certain correlation length there is a high probability of finding sites of opposite types adjacent to each other (statistically alternating). The bottom left quarter illustrates that in a certain correlation length there is a high probability of finding sites of the same type opposite to each other (statistically blocky). All quarters of the surface are characterized by an average total loading of 20 percent, and the correlation length is ~ 1.4 for the statistically patterned regions.

Consider a mixture of DHPs with statistically blocky and alternating type sequences in solution interacting with such a surface. DHP segments of type A prefer to interact with red sites on the surface, and those of type B prefer the yellow surface sites (Figure 1). Statistically blocky (alternating) DHPs are statistically better pattern matched with the statistically patchy (alternating) part of the surface. If the surface shown in Figure 1 is exposed to a fluid solution containing a mixture of statistically blocky and statistically alternating DHPs, will the chain molecules selectively adsorb on those regions of the surface with which they are statistically pattern matched? Such recognition due to statistical

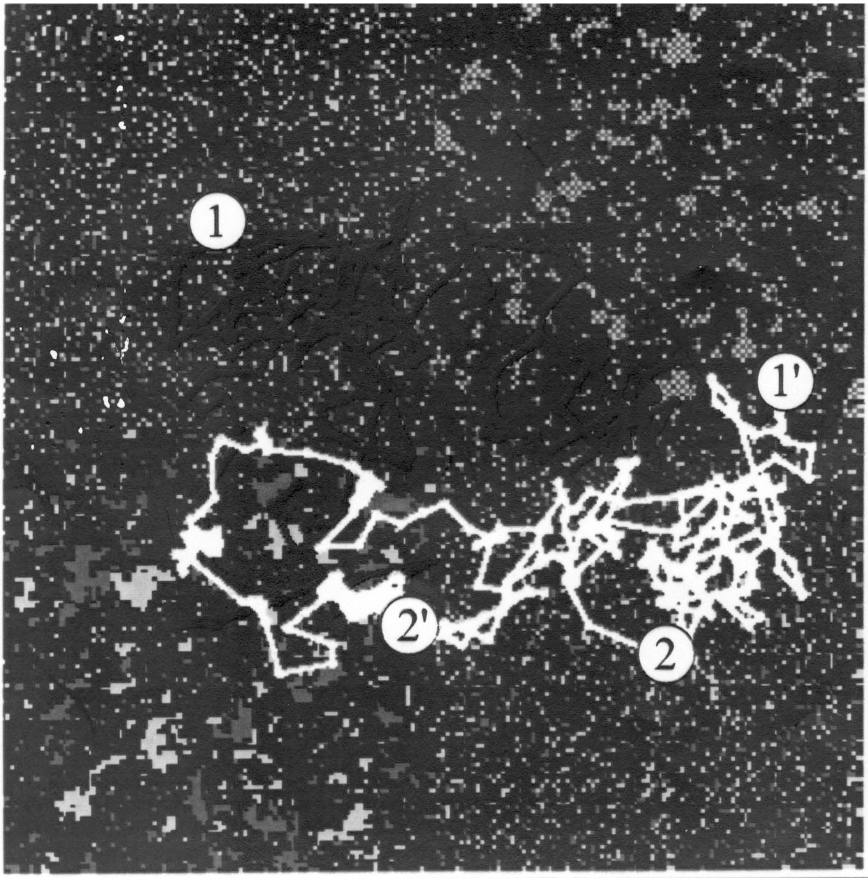

FIGURE 1 Surface bearing different statistical patterns. The two different types of sites are colored red and yellow, with the neutral background being blue. The top right (bottom left) quarter is statistically alternating (blocky); the other quarters are statistically random. The black and white trajectories are 2-D projections of the center of mass positions of statistically alternating and blocky DHPs, respectively. 1 (2) and 1' (2') are starting and ending positions of the statistically blocky (alternating) DHP. Figure can be viewed in color online at <www.nap.edu>. SOURCE: Reprinted with permission from Golumbfskie et al. (1999). Copyright (1999) National Academy of Sciences, U.S.A.

pattern matching requires not only that the "correct" patch be strongly favored for binding thermodynamically but also that the "wrong" regions of the surface not serve as kinetic traps.

Figure 1 depicts typical trajectories at $T/T_{ref} = 0.6$. All points on these trajectories do not correspond to adsorbed states. Both the statistically alternating and blocky DHPs begin on randomly patterned parts of the surface and ultimately find their way to the region of the surface that is statistically pattern matched with its sequence statistics. These results show that biomimetic recognition between polymers and surfaces is possible due to statistical pattern matching. It seems possible to exploit this notion to design inexpensive devices that can separate a large library of macromolecules into groups of statistically similar sequences. Sensor applications are also suggested.

The reason why statistical pattern matching allows biomimetic recognition to occur is made clear by the free-energy landscape shown in Figure 2. The "wrong" regions of the surface correspond to local free-energy minima that are

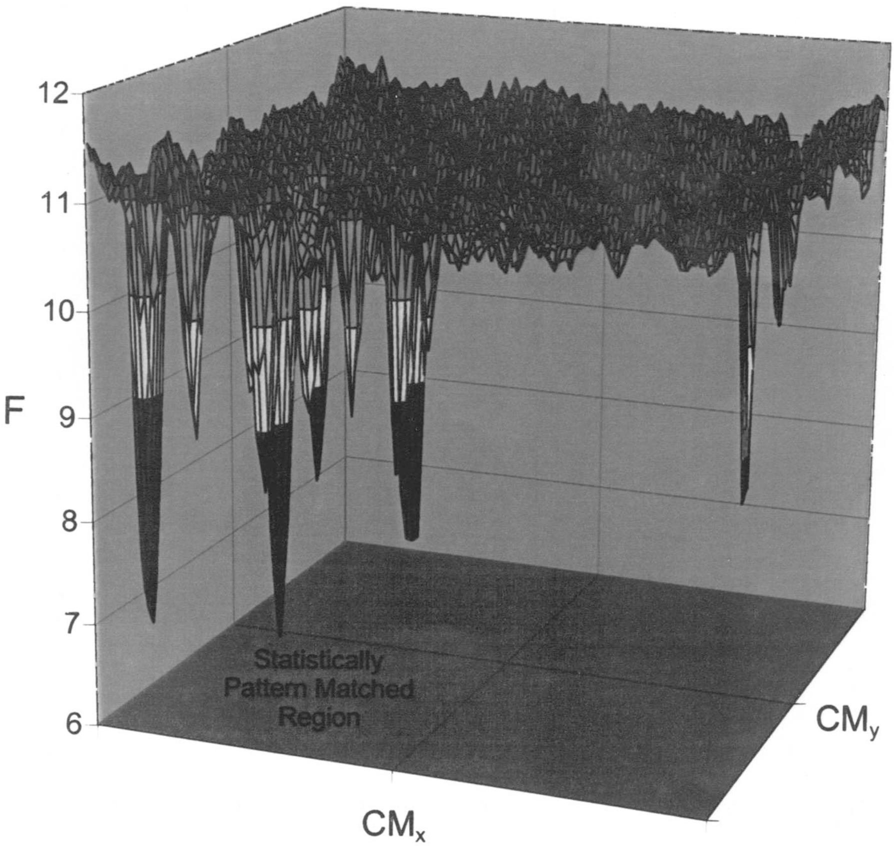

FIGURE 2 Free energy (in arbitrary units) versus center of mass position for a statistically alternating DHP interacting with the surface in Figure 1. Blue corresponds to deep free-energy minima, red patches are shallower free-energy minima, and yellow regions are free-energy barriers. The two deep free-energy minima along the right edge of the surface (wrong region) are due to periodic boundary conditions. Figure can be viewed in color online at <www.nap.edu>. SOURCE: Reprinted with permission from Golumbfskie et al. (1999). Copyright (1999) National Academy of Sciences, U.S.A.

separated by relatively small barriers from each other. In contrast, in the statistically pattern matched region there exist a few deep global minima; each of these minima, in turn, is very rugged.

The nature of this free-energy landscape leads to many interesting kinetic phenomena. One that is particularly stimulating is the shape of the macromolecule that evolves with time. When recognition occurs due to statistical pattern matching, the macromolecules adopt a small class of shapes.

Adsorbed macromolecules are characterized by loops. A loop is a length of contiguous unadsorbed chain segments between two consecutive adsorbed segments. The distribution of these loops fluctuates in time, and hence loop fluctuations are dynamic modes. Prior to considering the dynamics of these modes, let us examine the distribution of these loops as trajectories evolve, starting from wrong parts of the surface. All trajectories show the qualitative features depicted in Figure 3 where we plot the probability distribution P_n for an arbitrary segment to be part of a loop of length n for a statistically alternating DHP. Each panel shows P_n averaged over different time (Monte-Carlo step) windows. The first panel shows that, when the center of mass of the chain is in the wrong part of the surface, P_n is essentially structureless. This implies that all possible macromolecular shapes are being adopted as the chain samples the wrong parts of the surface. Remarkably, the other panels in Figure 3 demonstrate that, as the chain center of mass enters the statistically pattern-matched region, P_n begins to show structure. Ultimately, as shown in the panel labeled b, it exhibits a spectrum of peaks that corresponds to preferred loop lengths and hence macromolecular shape in the adsorbed state. This observation of adsorption in preferred shapes due to statistical pattern matching is very similar to recognition in biology. Results also suggest an intriguing connection between the experimentally realizable system studied and some provocative ideas concerning the competition between self-organization and selection in evolution (Kauffman, 1993).

MOLECULAR ARCHITECTURE CAN CONTROL SURFACTANT PROPERTIES OF MACROMOLECULES

The ability of amphiphilic molecules to form organized assemblies in solution has important commercial and biological consequences (Discher et al., 1999; Gompper and Schick, 1994; Larson, 1999; Lasic, 1993; Safran and Clark, 1987; Won et al., 1999; Zana, 1986; Zhao et al., 1998). A variety of products, such as detergents, emulsifiers, catalysts, and vehicles for drug delivery, rely on this ability. Membranes in plant and animal cells are composed of self-assembled phospholipid bilayers. Self-assembly into these organized motifs is driven by the amphiphilic character of the molecular building blocks (i.e., different chemical groups in the molecules exhibit different solvent affinities). The simplest organized assembly is a micelle, which is formed to minimize unfavorable interactions between the medium and the poorly solvated moieties of the amphiphile.

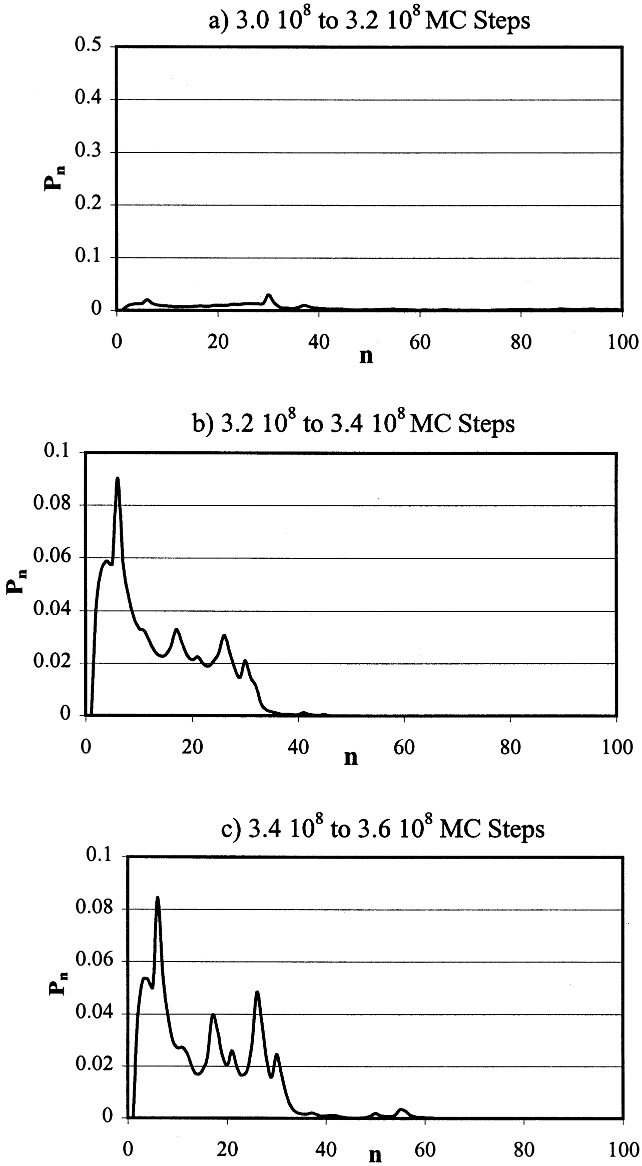

FIGURE 3 The probability distribution of loops (P_n). a) center of mass in the wrong parts of the surface; b) center of mass has entered the statistically pattern matched region; c) shape selective adsorption has occurred. SOURCE: Reprinted with permission from Golumbfskie et al. (1999). Copyright (1999) National Academy of Sciences, U.S.A.

Although a large majority of studies have been conducted using water as the solvent, the amphiphilic character of molecules can be expressed in a variety of media such as organic solvents and polymers (Balsara, 1998; DeSimone et al., 1994; Discher et al., 1999; Gompper and Schick, 1994; Johnston et al., 1996; Larson, 1999; Lasic, 1993; Safran and Clark, 1987; Won et al., 1999; Zana, 1986; Zhao et al., 1998). Extensive theoretical and experimental studies of micelles formed by single-tailed and double-tailed amphiphiles (Figure 4a), including those of biological importance, have been conducted (e.g., Balsara, 1998; DeSimone et al., 1994; Discher et al., 1999; Gawrisch et al., 1992; Gompper and Schick, 1994; Johnston et al., 1996; Karaborni and Smit, 1996; Karbaroni et al., 1994; Larson, 1999; Lasic, 1993; Lewis et al., 1994; Pochan et al., 1996; Safran and Clark, 1987; Tate et al., 1991; Won et al., 1999; Zana, 1986; Zhao et al., 1998). Recent studies have focused on synthetic double-tailed surfactants (Gemini surfactants) due to their superior properties (e.g., Danino et al., 1995; Zana, 1996). However, most efforts to control surfactant properties focus on proper choice of the chemical structure of the amphiphilic molecule.

The surfactant properties of macromolecules (and hence their ability to self-assemble into functionally interesting motifs) can be controlled with high sensitivity by manipulating molecular architecture without changing the chemical identity of the amphiphilic moieties. In addition to differences in surfactant properties between macromolecules in different architectural classes, subtle variations in an architectural class also lead to significant effects. This is due to the importance of conformational entropy for self-assembly processes of polymers. This notion of choosing the nature of the connections between the amphiphilic moieties to control surfactant properties may prove useful in applications where the choice of chemical structure is restricted (e.g., for concerns related to biocompatibility or toxicity).

We use light scattering experiments and a field-theoretic model to show that a class of branched macromolecules (see Figure 4c) is an extremely efficient surfactant. It is worth remarking that agrecans, one of the most effective biological surfactants, have a similar architecture (Alberts et al., 1994). For the branched macromolecules shown in Figure 4c, a slight affinity of the medium toward the backbone relative to the branches is sufficient for micelle formation. Such a small selectivity is insufficient to induce micelle formation in linearly connected polymeric amphiphiles such as diblock copolymers (Figure 4b). In addition, an intermediate branching density optimizes surfactant properties due to the interplay between molecular architecture and conformational entropy (i.e., a particular architecture in an architectural class works best).

ACKNOWLEDGMENTS

Financial support for the work described here was provided by the National Science Foundation, the U.S. Department of Energy (Basic Energy Sciences),

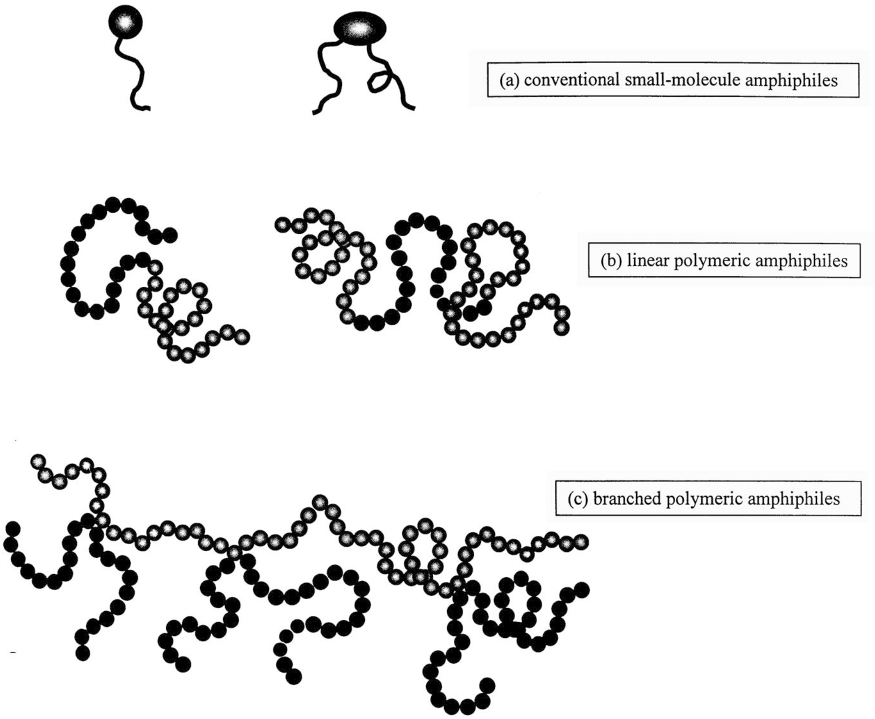

FIGURE 4 Amphiphilic molecules. (a) small molecule amphiphiles: single-tailed and double-tailed surfactants; (b) linear polymeric amphiphiles: diblock and triblock copolymers; (c) branched amphiphiles with a well-solvated backbone and poorly solvated branches.

and the Camille-Dreyfus Foundation. I would like to acknowledge collaborations with Prof. E. I. Shakhonovich, Prof. N. Balsara, Prof. V. Pande, Dr. S. Y. Qi, and Mr. A. Golumbfskie.

REFERENCES

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. 1994. Molecular Biology of the Cell. New York: Garland Publishing.

Balsara, N. P. 1998. Multicomponent polyolefin blends with ordered and disordered microstructures. Current Opinion in Solid State and Materials Science 3(6):589–595.

Bratko, D., A. K. Chakraborty, and E. I. Shakhnovich. 1997. Recognition between random heteropolymers and multifunctional disordered surfaces. Chemical Physics Letters 280(1–2):46–52.

Chan, H. S., and K. A. Dill. 1991. Polymer principles in protein structure and stability. Annual Review of Biophysics and Biophysical Chemistry 20:447–490.

Danino, D., Y. Talmon, and R. Zana. 1995. Branched threadlike micelles in an aqueous solution of a trimeric surfactant. Science 269(5229):1420.

DeSimon, J. M., T. J. Romack, and J. R. Combes. 1994. Dispersion polymerizations in supercritical carbon dioxide. Science 265(5170):356.

Dewey, G. 1997. Fractals in Molecular Biophysics. Oxford: Oxford University Press.

Discher, B. M., Y.-Y. Won, D. S. Ege, J. C.-M. Lee, F. S. Bates, D. E. Discher, and D. A. Hammer. 1999. Polymersomes: Tough vesicles made from diblock copolymers. Science 284(5417): 1143–1146.

Gawrisch, K., V. A. Parsegian, D. A. Hajduk, M. W. Tate, S. M. Gruner, N. L. Fuller, and R. P. Rand. 1992. Energetics of a hexagonal lamellar hexagonal-phase transition sequence in dioleoylphosphatidylethanolamine membranes. Biochemistry 31(11):2856–2864.

Golumbfskie, A. J., V. Pande, and A. K. Chakraborty. 1999. Simulation of biomimetic recognition between polymers and surfaces. Proceedings of the National Academy of Sciences U.S.A. 96(21):11707–11712.

Gompper, G., and M. Schick. 1994. Self-Assembling Amphiphilic Systems. San Diego, Calif.: Academic Press.

Irbäck, A., C. Peterson, and F. Potthast. 1996. Evidence for nonrandom hydrophobicity structures in protein chains. Proceedings of the National Academy of Sciences U.S.A. 93(18):9533–9538.

Johnston, K. P., K. L. Harrison, M. J. Clarke, S. M. Howdle, M. P. Heitz, F. V. Bright, C. Carlier, and T. W. Randolph. 1996. Water-in-carbon dioxide microemulsions: An environment for hydrophiles including proteins. Science 271(5249):624–626.

Karaborni, S., and B. Smit. 1996. Computer simulations of surfactant structures. Current Opinion in Colloid and Interface Science 1(3):411–415.

Karaborni, S., K. Esselink, and R. Os. Zana. 1994. Simulating the self-assembly of Gemini (dimeric) surfactants. Science 266(5183):254.

Kauffman, S. A. 1993. The Origins of Order: Self-Organization and Selection in Evolution. New York: Oxford University Press.

Larson, R. G. 1999. The Structure and Rheology of Complex Fluids. New York: Oxford University Press.

Lasic, D. D. 1993. Liposomes: From Physics to Applications. Amsterdam: Elsevier.

Lewis R. N., R. N. McElhaney, P. E. Harper, D. C. Turner, and S. M. Gruner. 1994. Studies of the thermotropic phase behavior of phosphatidylcholines containing 2-alkyl substituted fatty acyl chains: A new class of phosphatidylcholines forming inverted nonlamellar phases. Biophysical Journal 66(4):1088–1103.

Odian, G. G. 1991. Principles of Polymerization. New York: Wiley.

Pande, V. S., A. Y. Grosberg, and T. Tanaka. 1994. Thermodynamic procedure to synthesize heteropolymers that can renature to recognize a given target molecule. Proceedings of the National Academy of Sciences U.S.A. 91(26):12976–12979.

Pochan, D. J., S. P. Gido, S. Pispas, and J. W. Mays. 1996. Morphological transitions in an I2S simple graft block copolymer: From folded sheets to folded lace to randomly oriented worms at equilibrium. Macromolecules 29(15):5099–5105.

Safran, S. A., and N. A. Clark, eds. 1987. Physics of Complex and Supramolecular Fluids. New York: Wiley.

Sigal, G. B., M. Mammen, G. Dahmann, and G. M. Whitesides. 1996. Polyacrylamides bearing pendant-sialoside groups strongly inhibit agglutination of erythrocytes by influenza virus: The strong inhibition reflects enhanced binding through cooperative polyvalent interactions. Journal of the American Chemical Society 118(16):3789–3800.

Srebnik, S., A. K. Chakraborty, and E. I. Shakhnovich. 1996. Adsorption-freezing transition for random heteropolymers near disordered 2D manifolds due to “pattern matching.” Physical Review Letters 77(15):3157–3160.

Tate, M. W., E. F. Eikenberry, D. C. Turner, E. Shyamsunder, and S. M. Gruner. 1991. Nonbilayer phases of membrane lipids. Chemistry and Physics of Lipids 57(2–3):147–164.

Todd, R. J., R. D. Johnson, and F. H. Arnold. 1994. Multiple-site binding interactions in metal-affinity chromatography. I. Equilibrium binding of engineered histidine-containing cytochromes c. Journal of Chromatography A 662(1):13–26.

Won, Y.-Y., H. T. Davis, and F. S. Bates. 1999. Giant wormlike rubber micelles. Science 283(5404):960–963.

Zana, R, ed. 1986. Surfactant Solutions. New York: Marcel Dekker.

Zana, R. 1996. Gemini (dimeric) surfactants. Current Opinion in Colloid and Interface Science 1(5):556–571.

Zhao, D., J. Feng, Q. Huo, N. Melosh, G. H. Fredrickson, B. F. Chmelka, and G. D. Stucky. 1998. Triblock copolymer syntheses of mesoporous silica with periodic 50 to 300 angstrom pores. Science 279(5350):548–552.

Energy for the Future and Its Environmental Impact

Deregulating the Electric Grid: Engineering Challenges

THOMAS J. OVERBYE
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Urbana, Illinois

Most people seldom give much thought to electric power. And why should they? The electrical system was designed as the ultimate in plug-and-play convenience, and the regulated monopoly structure of the industry means they have had no choice but to buy their power from the local utility. Furthermore, our society's tremendous dependence on electric power has been transparent to most, except, of course, during the occasional blackout when it becomes all too apparent. However, there has been the issue of the monthly bill. In particular, the issue of why electric rates have been so different across the country. There had to be a better way! So began the process of deregulating the electrical grid, with no one quite sure where it would end up. This paper describes several of the engineering challenges associated with this deregulation.

BACKGROUND

Prior to deregulation electric utilities were vertically integrated "natural" monopolies serving captive markets governed by a regulatory compact. That is, in a particular service territory the electric utility did everything from owning and operating the generation, owning the transmission grid in that market, providing the wires that actually connected to the customer, to reading the customer's meter. It was a cost-plus business—the utility and regulators determined the allowable expenses, which were used to determine the customer rates, which the customers (previously known as ratepayers) had to pay. In exchange for this monopoly franchise, the utility accepted an obligation to serve all existing and future customers on a nondiscriminatory basis.

From an engineering standpoint this vertical monopoly provided a stable basis for building a reliable system. In an era of economies of scale, large power plants, and the high-voltage transmission system needed to move the power from these plants to the customers, could be engineered, built, and operated with the assurance that legitimate costs could be passed on to the ratepayers. Control was centralized, with the transmission grid shared by a relatively small group of utilities, more colleagues than competitors, all operating under the same paradigm of vertical monopolies.

Starting in the 1970s with the OPEC oil embargo, things began to change. Slowly the grid was opened to competition. Key recent events were the passing by the U.S. Congress of the Federal Power Act of 1992, which required opening the transmission system to competition, and in 1996 the issuance by the U.S. Federal Energy Regulatory Commission (FERC) of Orders 888 (Promoting Wholesale Competition Through Open Access, Nondiscriminatory Transmission Services by Public Utilities) and 889 (Open Access Same-Time Information Systems). The aim of these changes is simple: to provide nondiscriminatory access to the high-voltage transmission system so as to open the grid to true competition in the generation market, with the eventual goal of providing choice to the customers. However, the engineering challenges in doing this can be significant. This paper addresses three of these challenges: 1) bulk electricity market development, 2) market power assessment in electricity markets, and 3) power system data aggregation and visualization.

BULK ELECTRICITY MARKET DEVELOPMENT

FERC Orders 888 and 889 provided the broad guidelines for restructuring the U.S. power industry. How best to achieve this restructuring was left to individual state governments. Given the divergent political views of the states and their differences in average electric rates, it should not come as too much of a surprise that restructuring is progressing at vastly different rates across the country. Those states with the highest electric rates have been the first to restructure, while those with low rates are finding it difficult to see any advantage to changing the status quo. In this section I will examine some of the generic issues associated with an important step in restructuring, the development of a bulk electricity market.

The foundation for such a market is the high voltage electric transmission grid. The interconnected electric transmission grid in North America is one of the largest and most complex man-made objects ever created. This grid, which encompasses the entire North American continent, consists of four large 60-Hz ac synchronous subsystems. These subsystems are 1) the Eastern Interconnect, which supplies electric power to most users east of the Rocky Mountains; 2) the Western Interconnect, which supplies power to most users west of the Rockies and portions of Northern Mexico; 3) the Texas Interconnect, which supplies most of Texas; and 4) the Quebec Interconnect. These four subsystems are in turn connected to each other by dc transmission lines and back-to-back converters that allow for limited power transfers between the subsystems. Altogether the grid consists of billions of individual components, tens of millions of miles of wire, and thousands of individual generators.

The high degree of interconnection is beneficial, since it allows for economy and emergency transfers of electric power between regions, and hence the establishment of large electricity markets. A high degree of connectivity also has a detrimental side effect: failures in one location can propagate through the system at almost the speed of light. Large-scale blackouts can quickly affect tens of millions of people with losses reaching billions of dollars, as was illustrated by widespread outages of the Western Interconnect in 1994 and 1996 (Taylor, 1999). The engineering challenge of open access is to set up an open electricity market on what is really just one huge electrical circuit and simultaneously maintain reliability.

The transmission grid has several aspects that make setting up effective markets challenging. First, there is no mechanism to store electrical energy efficiently; total electrical generation must equal total load plus losses at all times. This continual matching must occur even as the load on the grid is constantly varying, with daily changes in demand of over 100% not uncommon. Second, with few exceptions, there are no mechanisms to directly control the flow of electricity on the tens of thousands of individual high-voltage transmission lines and transformers in the grid. Rather, the electric flow through the grid is dictated by the impedances of the transmission lines and the locations where electric power is injected by the generators and removed by the loads. Consequently, there are no analogous electrical elements to a gas industry control valve, a busy signal in the telecommunications industry, or a holding pattern in the airline industry. Thus, to transfer say 1,000 MW of power from Tennessee to Illinois the actual power flow would "loop" around on a large number of transmission lines in the Eastern Grid; this effect is known as "loop flow."

Third, the transmission grid has capabilities to transfer power that are finite but often difficult to quantify. These values have been defined by the North American Electric Reliability Council (NERC) as the available transfer capability (ATC) (NERC, 1996). ATC values are dependent on a number of constraints, including the need to avoid exceeding transmission/transformer thermal limits, voltage magnitudes limits, transient stability limits, and oscillatory stability limits. When an element is loaded to its limit it is said to be congested. Any additional power transfers that would increase the loading on that element are not allowed. However, because of loop flow when a single element is congested it can have a major systemwide impact. The real power losses associated with moving power in a market are nonlinear, varying as the square of the individual transmission line current. Thus, there is no unique way to assign the losses to particular market participants. FERC Orders 888 and 889 mandated the need for a functional unbundling of system services to allow separate pricing and hence competition for these so-called ancillary services. Ancillary services include power scheduling and dispatch, loading following, operating reserves, energy imbalance, real power loss replacement, and voltage (Hirst and Kirby, 1996). Previously these services had been bundled and supplied by the local utility. How much these services can really be unbundled and supplied by external markets remains an unanswered question.

While space limits a discussion of how these constraints are dealt with in individual markets, there is one consequence that merits special attention—extreme price volatility. Just about all electricity markets have experienced substantial price variation, with the price spike of June 1998 in the Midwest as a prime example (FERC, 1998). During this incident, which lasted for several days, spot market prices for electricity experienced almost unheard of volatility, soaring over two hundredfold from typical values of about $25 per megawatt hour (MWh) up to values of $7,500 per MWh. While the causes of this volatility are complex, they ultimately arise from several of the underlying characteristics of the electrical grid. During the June 1998 price spike, load levels were at or near record levels in the Midwest. As the load went up and with no way to store electricity, generation became an increasingly valuable commodity. Generation was in short supply, but was available elsewhere on the grid. Because of congestion due to thermal limits on just two elements, a transmission line in northwest Wisconsin and a transformer in southeast Ohio, no additional power could be transferred into the Midwest from either the West or the East. This situation allowed the remaining suppliers of power to rapidly raise prices to levels never before seen. Efficiently managing electricity markets where congestion on a single element can impact thousands of other elements, as well as power transfers, continues to be a challenge. This leads to a second key engineering challenge, developing markets in which market power is not abused by the market participants.

MARKET POWER ASSESSMENT IN ELECTRICITY MARKETS

One of the key goals of deregulation is obtaining lower prices through the advent of competition. However, as the grid is deregulated, there are significant concerns that the benefits gained from breaking up the vertical market power of a traditional utility may be lost through the establishment of horizontal market power, particularly in generation markets. Market power is the antithesis of competition. It is the ability of a particular seller or group of sellers to maintain prices profitably above competitive levels for a significant period of time. When an entity exercises market power, it ceases to be a price taker and becomes a price maker.

Market power analysis typically involves three steps: 1) identification of the products and services; 2) identification of the geographic market in which the product competes; and 3) evaluation of market concentration in that geographic market, using an index such as the Herfindahl-Hirschman index (Scherer, 1980).

For electricity markets step two is by far the most difficult since the geographic market depends on the actual loading of the transmission grid (Overbye et al., 1999a). Even in large electrical networks that contain a large number of market participants, congestion in the transmission grid can create "load pockets" in which there are only a small number of participants in the generation market, for example. Experimental evidence suggests that in such situations exploitation of market power by the participants could be expected (Zimmerman et al., 1999). This presence of load pockets should not be surprising since the transmission grid was originally designed to meet the needs of a vertically integrated utility moving power from its generators to its load. Market power is an issue that needs to be considered by regulators when examining utility mergers and third-party generation acquisition.

POWER SYSTEM DATA AGGREGATION AND VISUALIZATION

As the electricity industry becomes increasingly competitive, knowledge concerning the capacity and constraints of the electric system will become a commodity of great value. Understanding the rapidly changing electricity market before others do can give an important competitive advantage. The problem is that this knowledge is often contained in a tidal wave of data. The calculation of ATC for the Mid-America Interconnected Network (MAIN) is a prime example (Januzik et al., 1997). MAIN is one of the NERC regional reliability councils and covers most of Illinois, Wisconsin, the eastern part of Missouri, and the upper peninsula of Michigan. Thirty times each week MAIN calculates ATC values for over 200 buy/sell directions using a 14,000-bus power system containing about 20,000 transmission lines. For each direction about 1,300 contingencies must be considered. Thus, assuming one is interested in bus voltage magnitude and angle, and the real/reactive flow on each transmission line, the weekly data output of MAIN's ATC studies is about 1800 gigabytes of data. And MAIN is just one of ten regions, and ATC is just one value needed to participate in electricity markets.

The result has been that transmission providers and market participants are being overwhelmed by ATC data, but they have gained little insight into the mechanisms impacting their ability to obtain transmission service. Without this insight, it is nearly impossible for participants to make informed business decisions regarding the interaction between their desired transactions and the constraints imposed by the transmission system. For example, when electricity market players need transmission service they must determine whether to pay a higher premium and purchase nonrecallable transmission services or try to get by with less expensive recallable transmission services. Also, with either type of service they must develop a feel for the likelihood of curtailments. These situations require business decisions about managing the risk associated with the transmission system. The successful market participants will be the ones who understand the underlying transmission systems and can thus make fully informed decisions. Newer visualization techniques such as animation of power system flow values, contouring of transmission line flow values, data aggregation techniques, and interactive 3D visualization (Overbye, 1999b) have started to make a dent in this mountain of data, but more extensive methods are definitely needed.

REFERENCES

FERC (U.S. Federal Energy Regulatory Commission). 1998. Causes of wholesale electric pricing abnormalities in the Midwest. Staff report to Federal Energy Regulatory Commission, June 1998.

Hirst, E., and B. Kirby. 1996. Electric Power Ancillary Services. Oak Ridge, Tenn.: Oak Ridge National Laboratory.

Januzik, L. R., R. F. Paliza, R. P. Klump, and C. M. Marzinzik. 1997. MAIN regional ATC calculation effort. Proceedings of the American Power Conference, Chicago, April 1–3, 1997. Chicago: Illinois Institute of Technology.

NERC (North American Electric Reliability Council). 1996. Available Transfer Capability Definitions and Determination. Princeton, N.J.: NERC.

Overbye, T. J., G. Gross, P. W. Sauer, and M. J. Laufenberg. 1999a. Market power evaluation in power systems with congestion. Pp. 61–69 in Game Theory Applications in Electric Power Markets. New York: Institute of Electrical and Electronics Engineers.

Overbye, T. J., K. P. Klump, and J. D. Weber. 1999b. A virtual environment for interactive visualization of power system economic and security information. Paper presented at the IEEE Power Engineering Society 1999 Summer Meeting, Edmonton, Alberta, Canada, July 18–22, 1999.

Scherer, F. M. 1980. Industrial Market Structure and Economic Performance. Chicago: Rand McNally College Publishing Co.

Taylor, C. W. 1999. Improving grid behavior. IEEE Spectrum 36(6):40–45.

Zimmerman, R. D., J. C. Bernard, R. J. Thomas, and W. Schulze. 1999. Energy auctions and market power: An experimental examination. Paper presented at the 32nd Hawaii International Conference on System Sciences, Maui, January 5–8, 1999.

The Future of Nuclear Energy

PER F. PETERSON
Department of Nuclear Engineering
University of California
Berkeley, California

The discovery of nuclear fission in 1938 fundamentally altered the trajectory of the twentieth century. As this century closes, the future role of nuclear energy—both fission and fusion—remains unclear, but will range somewhere between limited specialty applications like isotope production and underwater propulsion, up to the large-scale commodity production of electricity and hydrogen. The next century should see nuclear energy's long-term role become defined, as fossil fuel dominance eventually erodes and as the environmental and economic costs of energy alternatives are explored in larger-scale deployments, and as research in advanced fission and fusion energy provides more attractive commercial products. Simultaneously, civilian nuclear infrastructure will play a critical role in managing nuclear weapons materials declared excess to military needs; in establishing increasingly rigorous norms for accounting, aggregating, protecting, and disposing of nuclear materials; and in seeing the global adoption of these stringent norms, particularly in the former Soviet Union.

NEAR TERM: THE NEXT DECADE

The near term will not see new nuclear power plants built in the United States, and only modest numbers constructed elsewhere. Most existing U.S. nuclear plants will continue to run, with a substantial fraction receiving license extensions. Deregulation will play an important role, opening the possibility of sale of nuclear power plants to dedicated operating companies (Joosten, 1999). The recent sale of the troubled Clinton power plant, which has been shut down for three years to address safety problems, suggests that plants that formerly would have shut down due to poor utility management can instead become attractive opportunities for purchase by operating companies with the technical and managerial capability to maintain high levels of reliability and safety, and thus maintain high capacity factors. With license extensions, the momentum provided by existing nuclear infrastructure—17 percent of current global electricity production—will maintain a foundation of substantial nuclear capability for three or more decades into the 21st century.

Nuclear energy research with the greatest near-term impact will focus on improving the economic performance of existing light water reactor (LWR) nuclear power plants. Interesting research opportunities will include studies to increase further fuel burn-up levels to extend the time between outages and reduce spent fuel generation, and to develop and license dry-cask storage systems, transportation systems, and high-level radioactive waste repositories. Coupled with increased fuel burn-up, intriguing research will also be directed toward high burn-up thorium and inert matrix-based LWR fuel designs that generate smaller amounts of plutonium.

INTERMEDIATE TERM: 10–40 YEARS

During this period increasing oil and gas prices, driven by depletion and potentially by carbon taxes, will result in growing fractions of electrical power generation coming from non-fossil sources, if the world rejects the environmental costs of further increases of coal combustion without carbon sequestration. In the absence of major technical breakthroughs, LWRs will remain the lowest-cost nuclear power option. Uranium prices are projected to remain low during this period, so that once-through fuel use will remain the most economically attractive option. The regional competitiveness of new nuclear power generation capacity will depend on resource availability. The initiation of new plant construction in the United States will be impeded by higher interest rates due to uncertainty about the performance of new regulatory systems and risk of construction delays, and by the economies of scale that make LWRs most attractive at sizes greater than 1 GW and thus make the capital investment for the first new LWR quite high. Following construction of a first plant, particularly if construction occurs successfully in under five years as recently achieved in Japan, interest rates for subsequent plants could be substantially lower.

Recent boiling water reactor (BWR) designs with internal pumps or natural circulation, which eliminate bulky external reactor equipment like jet pumps and steam generators, permit extremely compact containment buildings. Because modular construction technology can be readily applied to these BWR containments, the quantities of material and construction times are reduced substantially. In the absence of subsidies, BWRs can be expected to increasingly dominate the economic competition in the LWR market.

Research activities with the greatest potential to accelerate new reactor construction will focus on further improvements to advanced LWRs. In particular, research on LWR passive-containment cooling systems for accident response (e.g., Peterson et al., 1998) will have large leverage due to the potential to reduce the plant cost by eliminating active cooling systems and diesel-generator power supplies, to reduce the safety envelope volume, to simplify operations and maintenance activities, and to improve reliability and safety. At the back end of the fuel cycle, progress on siting, design, and construction of geologic repositories will be important regardless of growth or decline in fission energy production. The evolution of international infrastructure and standards for managing nuclear materials will have major long-term nonproliferation implications, substantially negative if the evolution continues toward increasingly dispersed long-term surface storage rather than toward aggregation and disposal in regional or multinational facilities (Peterson, 1996).

Improved understanding of the health effects of low levels of radiation, where currently standards are set by simple linear extrapolation of health effects observed at large doses, may potentially result in reassessment, either up or down, in the safety of nuclear power and the consequences of nuclear accidents. The demonstration of a threshold for radiation effects, postulated by some researchers, would result in a major decrease in the calculated consequences of severe accidents, and would also affect design requirements for radioactive waste disposal.

Markets for nuclear energy will also depend strongly on the evolution of other energy technologies. Particularly favorable developments for nuclear energy would include substantial improvements in energy storage technology and in time-of-day pricing and demand-side management. Storage potentially opens new electricity markets for mobile applications, and also displaces fossil-fired peaking capacity and increases the value of nighttime sales from capital-intensive base-load sources. More effective demand management also reduces the market share for expensive fossil-peaking capability and shifts demand toward off-peak periods.

LONG TERM: BEYOND 40 YEARS

In the long term, if coal's environmental costs are rejected, traditional fossil energy will play a substantially diminished role in energy production, with a corresponding increase occurring in the contributions of non-carbon-emitting energy sources. In this time range modest or potentially substantial contributions to energy production may come from both fission and fusion energy sources. For growing contributions from fission energy, most economic analyses suggest that uranium prices will become sufficiently high after some five decades to economically warrant the recycle of spent fuel to capture its additional energy content, and a transition to fast spectrum reactors capable of operating at breeding ratios slightly above one.

Proliferation resistance will be an important goal for future changes to the nuclear fuel cycle. Issues related to proliferation resistance are better understood now, particularly because it is now widely known that all isotopes of plutonium must be treated as potentially weapons-usable, as well as the transuranic elements neptunium and americium. Likewise, the United Nation's International Atomic Energy Agency has now determined that geologic repositories for spent fuel will require permanent safeguards monitoring and institutional control (Linsley and Fattah, 1994). While several decades of global spent fuel production can be placed in a small number of multinational repositories to reduce the long-term risks and burdens of the safeguards monitoring, long-term commitments to nuclear fission will require gradual transition to technologies that generate waste streams qualifying unambiguously for permanent safeguards termination. Here the most promising direction for research may be toward lead and lead/bismuth coolants for reactors.

Early U.S. fast reactor research chose to focus on liquid sodium as a coolant, due to lower corrosivity as well as low density, which enables high velocity flow and higher power density, achieving more rapid breeding of new plutonium for the startup of new fast reactors. Currently, high power density is no longer considered a virtue, and sufficient plutonium exists for the startup of large numbers of fast reactors if it is ever desirable. Furthermore, the end of the Cold War brought information that the Russians have solved the corrosion issues associated with lead coolants, and use lead/bismuth cooled reactors in their existing submarine fleet (Chekunov et al., 1997).

Lead, with its high atomic weight, extracts essentially no energy in elastic collisions with neutrons, and thus lead-cooled reactors can have an extremely hard neutron spectrum. This has several beneficial effects. Little spectrum hardening occurs due to void generation, while neutron leakage increases substantially, making it much easier to obtain negative void reactivity coefficients, a task that is difficult with sodium. The exceptionally hard spectrum also allows effective burning of the transuranic elements neptunium and americium that are of concern from a waste safeguards termination perspective. Conversion ratios above one are possible without blankets, so that the fuel self-breeds enough plutonium internally to maintain constant reactivity, and the only limitation on core life comes from radiation damage, potentially allowing core lifetimes approaching 15 to 20 years, three or four times greater than current LWRs. Because the cores can be homogeneous, multiple recycle of the fuel can be accomplished by high-temperature volatility-based methods that remove only fission products and are incapable of separating uranium and plutonium, making them highly proliferation resistant and capable of consuming LWR spent fuel inventories.

Because lead is chemically inert with the water required to generate steam for power production, and because the vapor pressure of lead is extremely low, lead-cooled reactors do not require the massive containment structures of LWRs and thus have the potential to be less expensive than current LWR technology, particularly if refueling occurs at greater than 10-year intervals. In whatever form, improvements in fast-spectrum reactor technology and the ability to produce waste streams qualifying unambiguously for permanent safeguards termination will be important if fission makes long-term contributions to global energy production.

Fusion faces important scientific and engineering hurdles to becoming an economical energy source. However, because the fuel supply for fusion is effectively infinite, the waste generation and accident consequences are far smaller than for fission plants, and the land resources and environmental impact required by fusion could be much smaller than for renewables. Therefore, fusion energy deserves substantial attention and continued development. For magnetic fusion energy (MFE), confinement concepts can be categorized on a scale leading from plasmas externally controlled by large magnets to simpler self-organized plasma configurations that create confining magnetic fields primarily by internal plasma currents. The diverse range of potential plasma configurations, and the risks and benefits associated with each, provides the primary dilemma for prioritizing MFE research. Recent reviews of the U.S. fusion program (SEAB, 1999) have recognized this dilemma and have recommended a balanced research portfolio that continues efforts with externally controlled plasmas (e.g., tokamaks) where the path to success is well understood but where development costs are high, and self-organized plasmas where greater technical uncertainty exists but where the development costs, and potential power plant costs, may be very low.

For inertial fusion energy (IFE), major progress has been made with the ongoing construction of the National Ignition Facility, a large laser system that is anticipated to ignite fusion targets around 2006, and with design of heavy-ion accelerator and laser driver systems that could operate at the 5-Hz rate required for fusion power plants. National development efforts for MFE and IFE fusion power sources include extensive technology development components. Very interesting are developments toward the use of high temperature liquids to shield fusion chamber structures and remove fusion energy (Moir, 1995). By minimizing activation and waste generation and providing high power density, liquid protection would have strong, positive economic benefits. Combined with a simple self-organized plasma configuration, or inertial fusion with an innovative, inexpensive driver system, liquid protection would offer the potential for fusion energy to be less expensive than fission with greater innovation, and to be competitive with current natural gas prices. Such success would have enormous implications for future human welfare.

The careful management and use of nuclear technologies and materials will remain a permanent responsibility for this and future generations, with large long-term leverage on global economic, environmental, and security conditions. Research, coupled with continued strengthening of the international regime for the management of nuclear materials, can influence this important leverage, under the assumption that bright and motivated people continue to enter the field and work to maximize the benefits that nuclear technologies can potentially bring.

REFERENCES

Chekunov, V. V., D. V. Pankratov, Y. G. Pashkin, G. I. Toshinsky, B. F. Gromov, B. A. Shmatko, Y. S. Belomitcev, V. S. Stepanov, E. I. Yefimov, M. P. Leonchuk, P. N. Martinov, Y. I. Orlov. 1997. Use of lead-bismuth coolant in nuclear reactors and accelerator-driven systems. Nuclear Engineering and Design 173(1–3):207–217.

Joosten, J. 1999. U.S. electric market restructuring: Implications for nuclear plant operation and safety. Nuclear News 42(6):41–48.

Linsley, G., and A. Fattah. 1994. The interface between nuclear safeguards and radioactive waste disposal: Emerging issues. IAEA Bulletin 36(2):22–26.

Moir, R. W. 1995. The logic behind thick, liquid-walled, fusion concepts. Fusion Engineering and Design 29:34–42.

Peterson, P. F. 1996. Long-term safeguards for plutonium in geologic repositories. Science and Global Security 6(1):1–29.

Peterson, P. F., V. E. Schrock, and R. Greif. 1998. Scaling for integral simulation of mixing in large, stratified volumes. Nuclear Engineering and Design 186(1–2):213–224.

SEAB (Secretary of Energy Advisory Board). 1999. Realizing the Promise of Fusion Energy. Final Report of the Task Force on Fusion Energy. Washington, D.C.: U.S. Department of Energy. [Online]. Available: http://www.hr.doe.gov/seab [December 21, 1999].

Renewable Energy Technologies: Today and Tomorrow

JAMES M. CHAVEZ
Sandia National Laboratories
Albuquerque, New Mexico

JANE H. DAVIDSON
Department of Mechanical Engineering
University of Minnesota
Minneapolis, Minnesota

ABSTRACT

Renewable energy sources have served humankind for thousands of years, and they will continue to be vital resources in the future. Because they are sustainable, solar, wind, biomass, and geothermal systems can meet many of our energy needs with minimal impact on the environment. Yet to date, they generate less than 1 percent of the worldwide energy needs and only 2 percent of the electricity generation in the United States. Experts believe that these technologies have not become major sources of energy because conventional sources cost less and have perceived advantages over renewables. However, they are enjoying renewed appeal and opportunities because of issues such as global climate change, carbon emissions, environmental concerns, utility restructuring, and growth in energy demands in developing nations. In fact, some studies (e.g., by the World Energy Council, Shell Corp., and the United Nations) project that renewable energy technologies are going to contribute to the world's energy supplies at the 20 percent to 50 percent level by the year 2040. Such rapid growth in the use of renewable energy requires that the technologies become more cost-effective. Improvements in manufacturing processes, efficiency of operation and maintenance, and technical advances in materials, processes, and storage will be needed. In this paper, we discuss renewable energy technologies, their current status, and their potential to have a major impact in the future.

INTRODUCTION

Renewable energy technologies have improved dramatically in the last 25 years. Efficiencies are higher, reliability is better, and costs are lower. However, because the cost of power from conventional energy sources continues to decline, these technologies have not made a significant impact on the marketplace to date. Renewable energy (excluding hydro) currently generates less than 1 percent of the energy needs worldwide and accounts for only 2 percent of the electricity generated in the United States, as shown in Figures 1 and 2. However, the combination of continuing growth of energy consumption (more than 3 percent) annually and the threat of increased emissions of greenhouse gases (see Figure 3 for carbon emissions in the world), demands that we consider a more sustainable energy mix.

No "silver bullet" exists to solve the world's energy needs and mitigate the environmental impact of energy production. Instead, a portfolio of energy sources, including the renewables—biomass, geothermal, wind, and solar—will be needed. All have cost and technological issues that must be addressed if they are to be widely commercialized without a substantial increase in the cost of electricity and without government subsidies. Even the relatively mature renewable technologies require technological advances (efficiency and reliability improvements) that will drive cost down. In this paper, emerging renewable technologies will be addressed from the perspective of electricity generation (hydropower is considered a mature technology). The following overview includes a brief introduction to the technologies and current areas of research to improve them.

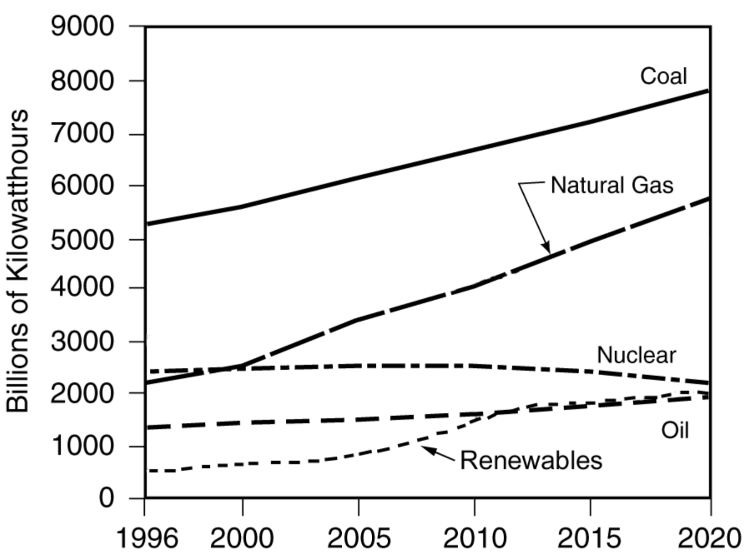

FIGURE 1 World consumption for energy generation. SOURCE: Data courtesy of the Energy Information Administration (1999).

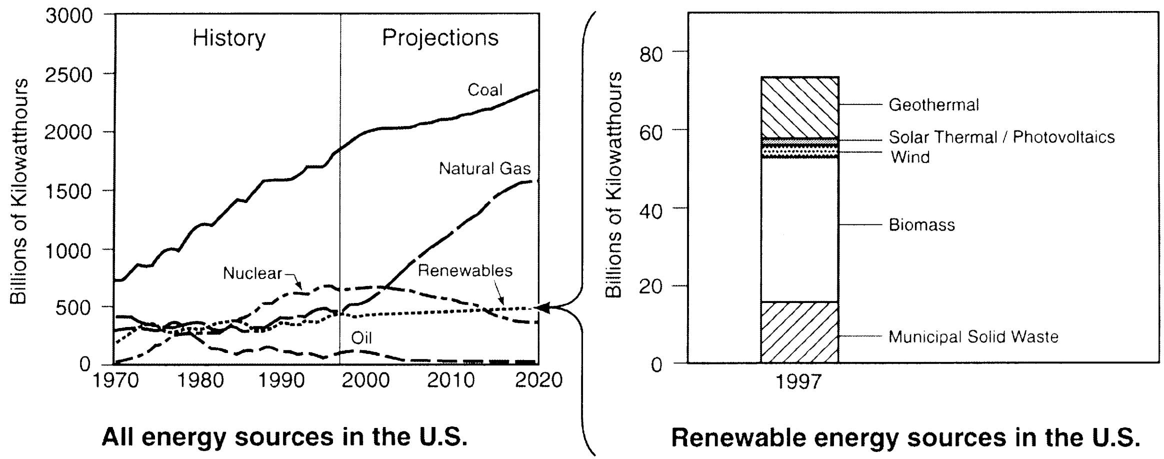

FIGURE 2 Electricity generation in the United States by source and from renewable energy sources for 1997. SOURCE: Data courtesy of the Energy Information Administration (1998).

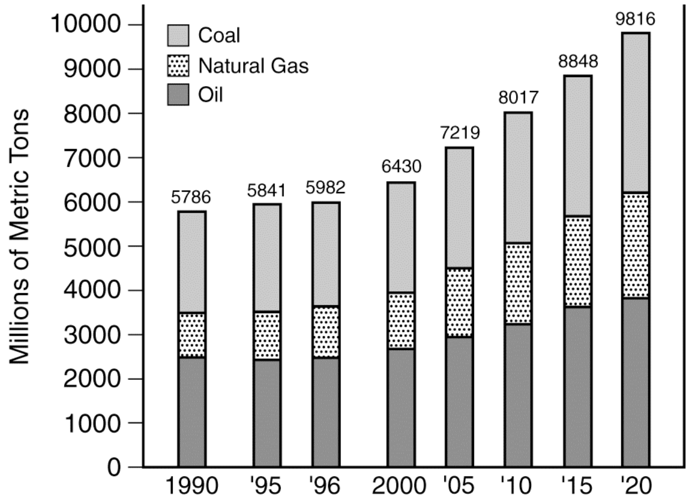

FIGURE 3 World carbon emissions by source. SOURCE: Data courtesy of the Energy Information Administration (1999).

OVERVIEW OF RENEWABLE ENERGY TECHNOLOGIES

Biomass power is combustion of plant-derived material, wood and agricultural residues, cultivated trees or herbaceous plants such as sorghum or sugar cane, solid municipal waste, and, in the less developed nations, dung. The energy density of biomass is approximately two-thirds that of bituminous coals, but it offers other advantages. It has lower sulfur and ash content than coal, is easier to gasify, and is not considered to impact greenhouse gases. The carbon dioxide emissions from burning are offset by absorption during photosynthesis. However, potential environmental impacts are land and water use.

Biomass power accounted for approximately 10,500 MW or 1.2 percent of U.S. electric generating capacity in the years 1993 to 1997. Direct combustion of municipal solid waste, which is not truly a renewable resource, accounts for 3,400 MW. U.S. demonstrations of biomass gasifiers capable of connecting to gas turbines are in Hawaii and Vermont. Currently, six U.S. power plants are co-firing coal and wood residue products on a commercial basis. According to the U.S. Department of Energy, domestic biomass generation capacity could reach 20–30 GW by the year 2020.

The emerging biomass technologies are production of biofuels, including ethanol, methanol and biodiesel from vegetable oils, production of gasified biomass for use in gas turbines, and co-firing of biomass with coal to reduce sulfur emissions. Technology needs include development of high-growth-rate biomass crops, lower cost generation of biofuels and combined-cycle gasifier plants running on biomass.

Geothermal energy is obtained from the vast heat resources beneath the Earth's surface. The low-temperature heat contained in shallow ground can be used to heat or cool homes and commercial buildings either directly or with ground-coupled heat pumps. In active geothermal zones, the temperature gradient is about four times the nominal value of 30°C per kilometer, and high-pressure steam or water is used for electric power generation. Geothermal power plants release on average only 5 percent of the carbon dioxide emitted by a fossil fuel plant. However the environmental impact of exploiting geothermal resources is controversial. Most of the U.S. hydrothermal systems with obvious surface resources have been explored.

Twenty-one countries generate 8,000 MW of electricity from geothermal resources, and 11,300 thermal megawatts are used for applications such as aquaculture, greenhouse operations, and industrial processing. Domestic power plants generate 2,800 MW at 5 to 7.5 cents per kilowatt-hour.

Power production is possible from hydrothermal reservoirs, geopressurized reservoirs that contain methane under high-pressure, hot dry rock, and magma, but currently only hydrothermal reservoirs are used. The needs of this technology are to develop models and instrumentation for exploration and characterization of new reservoirs, for cost-effective deep well drilling, to improve cost-effective binary conversion cycle efficiency, and to develop high-temperature corrosion-resistant materials and coatings. Exploitation of hot dry rock, magma, and geopressured aquifers remain formidable challenges to this technology.

Wind is the most cost competitive of the renewable energy technologies. Advanced turbines, primarily horizontal axis designs, have replaced the sails and blades of traditional farm windmills. The focus is on large power plants with individual turbines as large as 1,000 kW. Since the 1980s, cost has dropped, efficiency has increased, and reliability has improved. Modern wind turbines produce electricity for 5 to 6 cents/kWh. Bird kill, visual impact, and noise are perceived as environmental issues. The intermittent and site-specific nature of the wind poses challenges.

The world's installed capacity exceeds 10 GW, with large installations in Europe, the United States and India. Forecasts indicate a growth rate of over 2 GW per year. In the United States, 15,000 wind turbines in California have supplied more than 1 percent of the electricity for that state since 1980. In the last 2 years an additional 900 MW was installed in the United States, bringing the national total to about 2.5 GW. In fact, one estimate suggests wind could provide 20 percent of U.S. electricity needs with current technology. The new "Wind Powering America" initiative has goals of 5 GW by 2005, 10 GW by 2010, and 40 GW by 2020 (about 5 percent of the U.S. electricity consumption).

Technological advances to provide cheap, reliable power focus on power electronics, materials, structure design and manufacturing, and transmission with a continuously fluctuating input power. Sandia National Laboratories engineer Paul Veers summarizes the challenges this way, "The airfoils must work in a dynamic stall environment. The structure must withstand billions of stress cycles with low cost materials. . . . A control system should smooth all power fluctuations. . . . It must be accomplished in an environmentally non-intrusive way" (Veers, 1996).

Solar technologies are simply divided into thermal and direct-conversion technologies. The thermal technologies include systems used for domestic heating and cooling, process heat, water and air treatment, thermochemical processing and power generation. This discussion addresses only electricity production.

Photovoltaics uses semiconductor technology to convert sunlight directly into electricity with no heat engine. The solar cell or photovoltaic cell was discovered in 1954 by Bell Telephone researchers examining the sensitivity of a silicon wafer to sunlight. Photovoltaic products are commercially available for lighting, communications, remote site electrification, traffic signs, water pumping, vehicle battery charging, as well as for grid-interactive electricity generation.

In 1995 global shipments of photovoltaics were nearly 90 MW, electric grid-connected systems making up 4 percent of the total. Most commercial sales are small 1- to 2-kWpeak standalone modules for lighting and remote sites. High cost limits their use in grid-connected applications; the estimate is 50 cents/kWh, 10 times the cost of power from a fossil fuel plant. Costs must be reduced to about $3/peak watt for photovoltaics to gain significant market share. Efforts to reduce cost focus on increased efficiencies and advanced manufacturing techniques.

Theoretical conversion efficiencies of photovoltaic systems depend on the semiconductor materials used in the cells and on the ambient temperature. The materials currently used to make photovoltaic cells can be grouped into three broad categories: 1) expensive, efficient monocrystalline silicon, 2) less efficient but much lower cost polycrystalline silicon, and 3) the lowest cost and poorest performer, amorphous silicon material. Conversion efficiencies of commercial polycrystalline silicon cells are 10 to 15 percent. Now the primary development areas are in how to use monocrystalline silicon with solar concentrators and making thin-film cells by depositing a 5- to 20-micron film of silicon onto an inexpensive substrate, because the estimated efficiency of these cells is above 20 percent. Work is ongoing with other materials, including amorphous silicon (a-Si), copper indium diselenide ($CuInSe_2$ or CIS) and related materials, and cadmium telluride (CdTe).

Concentrating solar power systems use mirrors to concentrate sunlight to produce temperatures high enough to drive modern, efficient heat engines and to produce electrical power. All concentrating solar power technologies rely on four basic systems: collector (mirror), receiver (absorber), transport-storage, and power conversion. Concentrating solar power stations need to concentrate large amounts of sunlight from the collector into a small area, the receiver, to produce high-temperature heat, which in turn can be converted into electricity in a conventional heat engine. Three types of concentrating solar systems have been developed, characterized by the shape of the mirrored surface on which sunlight is collected and concentrated: parabolic troughs, power towers, and dish/engine systems. All concentrating solar power systems can be put in series with a fossil-fuel-driven heat source that can either heat the working fluid, charge storage, or drive the power conversion system during periods of low sunlight.

Depending on the system, concentrating solar power generation can be designed to produce from tens to hundreds of megawatts of electricity. The power from it can be dispatchable, because these systems have cost-effective thermal storage and they can be hybridized (coupled with a conventional power plant). Consequently, these plants can produce power before or after sunrise (or 24-hour operation if desired). Concentrating solar power stations are best suited to be either peak-load or intermediate-load power stations.

All concentrating solar technologies have been validated and demonstrated, with trough-electric systems being the most mature of the three types. In fact, nine systems were installed in California in the 1980s, and they are still operating reliably. Power tower technologies have been demonstrated at the 10-MW Solar One and Solar Two pilot plants in California, and dish/Stirling systems have and are being demonstrated in the Middle East, the U.S. desert southwest, and the Mediterranean region. Although these technologies have been under development for the last 20 years, additional technology development is needed to continue to reduce costs and ensure their reliability. The major areas of technology development are manufacturing of the collectors and improving system reliability.

CONCLUSIONS

Perhaps the most dramatic effect on the renewable energy industry is the push toward cleaner technologies, as global climate change has become the focus of international attention. Forecasts indicate that in the near future, we can expect more industry support and acceleration of research and development into renewable energy sources. For renewable energy to take its place in the power picture of the new millennium, we need 1) research and development to improve the efficiency and reliability of these systems, 2) policy changes that address the barriers to use of renewable energy, and 3) acceptance by the government and the public that renewable energy is needed to help mitigate environmental problems and contribute to the world's energy supply. Biomass, geothermal, wind, and solar resources have been used for thousands of years. With changes in the environment and improvements in technology, they will become a larger part of power supplied in the future.

REFERENCES

Energy Information Administration. 1998. Annual Energy Outlook, 1999 with Projections to 2020. DOE/EIA-0383(99). Washington, D.C.: U.S. Department of Energy.

Energy Information Administration. 1999. International Energy Outlook 1999 with Projections to 2020. DOE/EIA-0484(99). Washington, D.C.: U.S. Department of Energy.

Veers, P. S. 1996. Foreword to special issue on wind energy. ASME Journal of Solar Energy Engineering 118(4):197.

ADDITIONAL RECOMMENDED READINGS

Arvizu, D. E., and T. E. Drennen. 1997. Technology Progress for Sustainable Development. Report SAND097-0555C. Albuquerque, N.M.: Sandia National Laboratories.

Dorf, R. C., ed. 1996. The Engineering Handbook. New York: CRC Press.

Eliasson, B. 1998. Renewable Energy: Hydro, Wind, Solar, Biomass, Geothermal, Ocean: Status and Prospects. Baden-Dattwil, Switzerland: ABB Corporate Research Ltd.

EUREC Agency. 1996. The Future for Renewable Energy: Prospects and Directions. London: James and James Ltd.

Kreith, F., and R. E West, eds. 1997. CRC Handbook of Energy Efficiency. New York: CRC Press.

Lay, K. L. 1998. Changes and innovation: the evolving energy industry. Presented at the 17th Congress of the World Energy Council, Houston, Tex., September 14, 1998.

Shell International Limited. 1996. The Evolution of the World's Energy Systems. London: Shell International.

Energy Information Administration. 1998. Renewable Energy Annual 1998 with Data for 1997. DOE/EIA-0603(98)/1. Washington, D.C.: U.S. Department of Energy.

Zweibel, K. 1995. Thin films: Past, present, future. Progress in Photovoltaics: Research and Applications 3(5):279–293.

OPTICS

Issues Associated with the Volume Manufacturing of Vertical-Cavity Surface-Emitting Lasers

THOMAS M. BRENNAN AND H. Q. HOU
EMCORE Photovoltaics
Albuquerque, New Mexico

ABSTRACT

Vertical-cavity surface-emitting lasers (VCSELs) have been grown by metal-organic chemical vapor deposition. Excellent uniformity of Fabry-Perot cavity wavelength for VCSEL materials of ±0.2 percent across a 3" diameter wafer was achieved. This results in excellent uniformity of the lasing wavelength and threshold current of VCSEL devices. Employing pregrowth calibrations on growth rates periodically with an in situ reflectometer, we obtained a run-to-run wavelength reproducibility for 770- and 850-nm VCSELs of ±0.3 percent over the course of more than a hundred runs. These developments have brought VCSEL devices out of the laboratory and into the commercial arena, where they are now available for many advanced technical applications, such as optical communications, biomedical applications, consumer electronics, and optical computing. The ability to manufacture very exquisite light sources, such as the VCSEL, theoretically for the cost of a light-emitting diode, provides for the first time the opportunity to economically process multitudes of information with light.

INTRODUCTION

Vertical-cavity surface-emitting lasers (VCSELs) are promising for a variety of applications such as optical interconnects, optical communications, optical recording, and remote sensing (see, for example, Lee, 1995). Growth of VCSEL materials requires a wide range of composition and doping-level variation. Metalorganic chemical vapor deposition (MOCVD) technology is increasingly recognized as a superior epitaxy platform because of its high throughput, continuous compositional grading control, flexibility for materials and dopant choices, and the ability to achieve low surface defect density (Hibbs-Brenner et al., 1996; Sun et al., 1995). However, the material uniformity across the entire wafer is often limited by the complex fluid and thermodynamics in a reactor. Another important concern lies in the run-to-run reproducibility. It is essential that thickness and composition control for a VCSEL structure be accurate to align the distributed Bragg reflector (DBR) stopband, Fabry-Perot mode, and the active-layer gain spectrum to a specific wavelength range. Slight day-to-day variation can drift the VCSEL wavelength and change the laser characteristics significantly. Recently, it was shown that good reproducibility and uniformity can be achieved from MOCVD-grown VCSELs (Hibbs-Brenner et al., 1996; Sun et al., 1995). However, the traditional geometry of the MOCVD reactor prohibits the use of in situ tools, such as pyrometric interferometry (Houng et al., 1994) and ellipsometry (Aspnes et al., 1990), which are commonly used in a molecular beam epitaxy (MBE) system for growth monitoring. The MOCVD growth conditions are usually determined in advance through an often tedious set of calibration runs. In this paper we demonstrate excellent reproducibility for VCSEL growth employing pregrowth calibrations with an in situ reflectance setup (Breiland and Killeen, 1995; Hou et al., 1996; Kawai et al., 1987) and high material uniformity by optimizing the growth conditions.

MOCVD GROWTH AND IN SITU REFLECTOMETRY

Epitaxial growth was performed in an EMCORE GS3200 MOCVD rotating disk reactor. GaAs and $Al_xGa_{1-x}As$ were grown at 750°C by using trimethylgallium (TMG), trimethylaluminium (TMA), and 100 percent arsine (AsH_3). Dopants for *n*- and *p*-type materials were from disilane (Si_2H_6) and carbon-tetrachloride (CCl_4), respectively. The alkyl and doping precursors were mixed in an injection block, and were carried by high-purity H_2 to the EMCORE vertical reactor through three different injection zones distributed along the radial direction on the top flow-injection flange. The AsH_3 was injected through two other zones. The reactants were isolated from the stainless steel chamber wall with a high hydrogen flow along the shroud. Therefore, there is little upstream contamination or carryover to a substrate. The substrate was rotated at 1,000 rpm. The reactor pressure was 60 torr.

The normal-incidence reflectance setup consists of a 7-watt W-halogen lamp as the light source and a silicon detector with a 10 nm bandwidth interference filter at 633 nm to detect the reflectance signal. The whole assembly was mounted directly on the top miniflange window of the reactor. A virtual interface model (Breiland and Killeen, 1995) extracts the growth rate and optical constants from the absolute reflectance of only the topmost layer by referencing the starting signal to the known reflectance of the substrate before epitaxy begins. It does not require any knowledge of epitaxial materials, interface position, thickness, or composition of underlying epitaxial layers. This provides a simple,

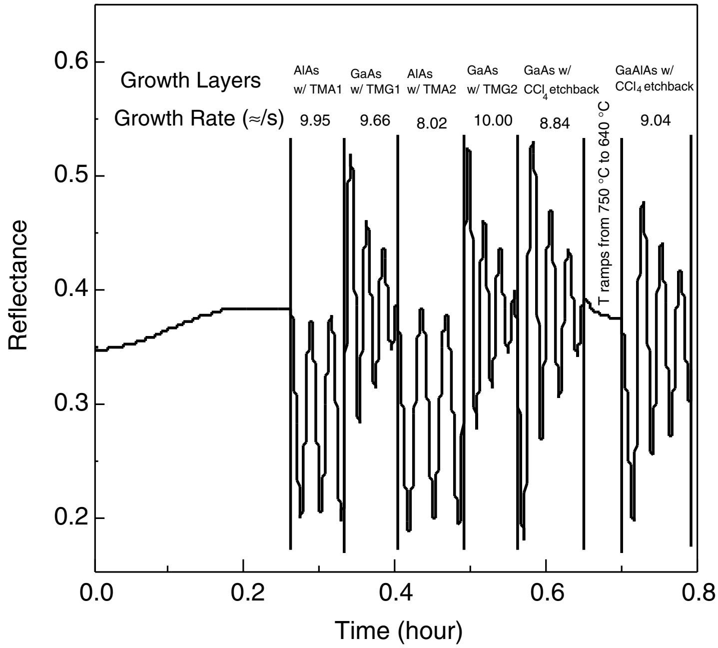

FIGURE 1 Temporal reflectance of a typical calibration run for the growth rate and etchback rate during AlGaAs growth. Within an hour, growth and etchback rates are obtained for the various growth conditions.

robust, and accurate measurement of the growth rate. Shown in Figure 1 is the temporal reflectance from a typical calibration run. The growth rates achieved by using various alkyl sources were determined from fitting the reflectance waveform with the virtual interface model. The reduction of the growth rate due to an etchback effect of AlGaAs by CCl_4 with different C doping levels at different temperatures is also extracted from this single calibration run of less than an hour (Hou et al., 1996). The small deviation of the growth rate from the expected value was corrected afterwards.

MATERIAL AND DEVICE UNIFORMITY

The sample uniformity was optimized by changing the gas-flow partition for alkyl sources among the three injection zones with hydride flow partition of 80 percent:20 percent between the inner and outer injection zones and a total reactor flow of 32.6 slm. Mirror structures of 20-period of AlGaAs/AlAs DBRs were grown under different conditions. The center wavelength of the DBR reflectivity stopband was measured for different radial positions of a 3-inch diameter

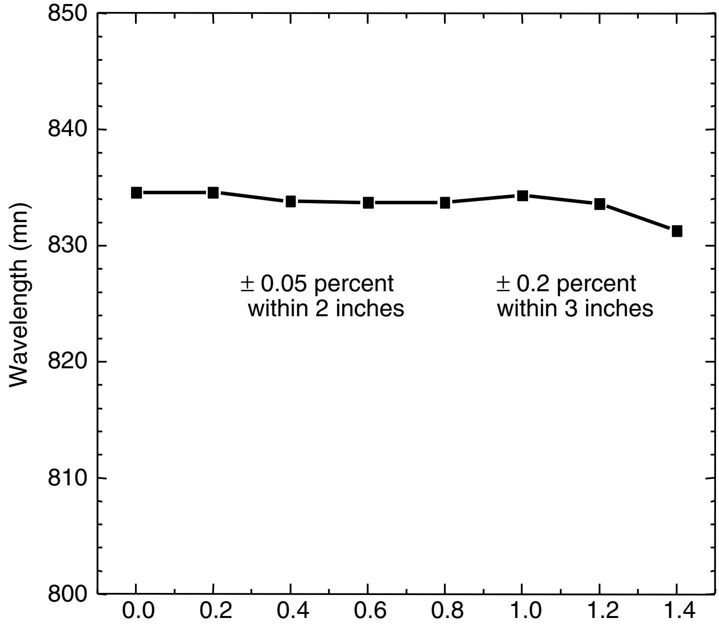

FIGURE 2 Uniformity of the center cavity wavelength for an 850nm VCSEL sample. Excellent uniformity of ±0.05 percent is achieved in the center 2-inch area, and ±0.2 percent in the whole 3-inch area.

wafer. It was found that equal alkyl distribution among the three injection zones on the top flange did not give a uniform growth rate on the substrate. Instead, a setting of the gas-flow distribution of 7 percent:78 percent:15 percent between the inner, middle, and outer injection zones yielded the most uniform thickness. The uniformity, which is most sensitive to the partition setting, was further optimized with the AsH_3 flow rate, or the V/III ratio. We found that the uniformity could be tuned slightly with AsH_3 flow rates from 230 to 270 sccm, or the equivalent V/III ratio from 36 to 42.3, and the best uniformity across an entire 3" wafer was achieved with an AsH_3 flow of 248.5 sccm, or a V/III ratio of approximately 39.

Figure 2 shows the wavelength of the Fabry-Perot cavity mode for an 850-nm VCSEL structure grown with the above-optimized conditions as a function of the distance from the wafer center. The uniformity was found to be rotation symmetric except for a little disturbance near the major flat of the wafer. The wavelength for the area ~3 mm from the wafer edge (the last data point in Figure 2) can be slightly influenced by the thickness of material coated on the susceptor. The wavelength variation ($\Delta\lambda$) was measured to be 0.8 nm in the center 2 inches and 3.3 nm across the entire 3-inch wafer. This corresponds to a wavelength and thickness uniformity of ±0.05 percent in the center 2-inch area of the 3-inch wafer, and ±0.2 percent over the entire 3-inch wafer. Since the alkyl gases are mixed uniformly before being injected into the reactor, and the incorporation ratio of Al/Ga is fairly insensitive to a small temperature variation around 750°C, the composition variation across a wafer should be negligible. This uniformity result is believed to be among the best achievable for III/V semiconductor epitaxial technology for an entire 3-inch wafer area.

VCSELs were fabricated from the material discussed above with oxide current apertures on both sides of the cavity to provide electrical and optical confinements (Choquette et al., 1995). The structure consists of 26 periods of top and 38 periods of bottom $Al_{0.16}Ga_{0.84}As/Al_{0.92}Ga_{0.08}As$ DBRs. The composition was parabolically graded across the interfaces of high and low Al containing layers to reduce the series resistance. There is one period of $Al_{0.16}Ga_{0.84}As/Al_{0.98}Ga_{0.02}As$ DBR on each side adjacent to the cavity. The $Al_{0.98}Ga_{0.02}As$ layer can be selectively oxidized laterally in a steam ambient to form an Al-oxide current aperture since the oxidation rate of $Al_{0.98}Ga_{0.02}As$ is 3.5 times higher than $Al_{0.92}Ga_{0.08}As$ (Choquette et al., 1995). The lasing wavelength and threshold current of VCSEL devices with 9 x 9 μm^2 oxide apertures on both sides of the cavity are plotted in Figure 3 as a function of the distance from the

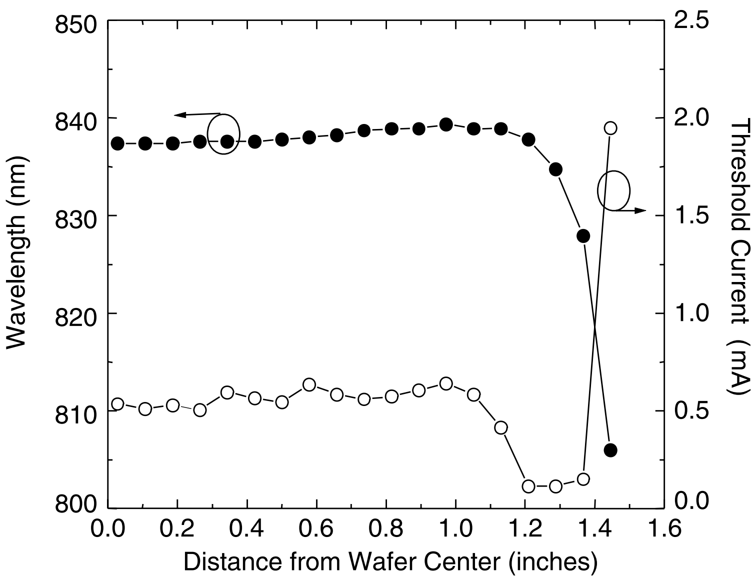

FIGURE 3 Uniformity of the lasing wavelength and threshold current of 850-nm VCSELs with 9 x 9 μm^2 oxide apertures.

center of the wafer. Even with fabrication nonuniformity, the lasing wavelength uniformity is still ±0.3 percent (or variation of ~5 nm) over nearly the whole wafer. The threshold current variation is less than ±10 percent in the center 2-inch area of the 3-inch wafer. The drop of the threshold current in the outer 1-inch area is due to the laser self pulsation (Choquette et al., 1996) resulting from a high series resistance of the *p*-DBRs. The C-doping efficiency drops when the substrate temperature increases (Hou et al., submitted) in the outer area of the 3-inch diameter platen where the temperature is estimated to be 20~30°C higher than the center. This results in an increase of series resistance by a factor of 2 in the outer area. Therefore, slightly over-doped *p*-DBRs should be used to ensure a low series resistance for more uniform device performance to compensate for the inherent effect caused by temperature nonuniformity of the substrate.

RUN-TO-RUN REPRODUCIBILITY

The MOCVD growth runs are also highly reproducible when a pregrowth calibration with the in situ reflectance is performed about once every 20~30 runs. Figure 4 shows the cavity wavelengths for a number of 770- and 850-nm VCSELs plotted as a function of the run number. The cavity was not changed from run to run for the 770- and 850-nm VCSELs, but the mirror design (number of periods and doping concentration) was varied for a structure-design optimization process. Twelve 850-nm VCSELs (run number 1068 to 1073, 1114, and 1163 to 1167) and seven 770-nm VCSELs (run number 1112 to 1113 and 1115 to 1119) were grown at different times over the course of about 100 runs. Other materials and device structures were grown in between. To ensure reproducibility, a calibration run as shown in Figure 1 was carried out before a series of VCSELs were grown. Slight changes in the flow rate calibration were made accordingly. As shown in Figure 4, we obtained wavelength run-to-run reproducibility of ±0.3 percent for both 770- and 850-nm VCSELs. Since the calibration included a wide range of composition and doping levels (Hou et al., 1996), reproducible results were achieved in spite of the varying device structures.

CONCLUSION

In summary, we have demonstrated an optimized MOCVD technique for highly uniform and reproducible VCSEL growth. In situ reflectance is used to monitor the growth routinely. With a pregrowth calibration of the growth rate using a test structure periodically, ±0.3 percent run-to-run cavity wavelength reproducibility was achieved over a course of about 100 growth runs. The thickness and composition uniformity is as good as ±0.2 percent, and the lasing wavelength uniformity is as good as ±0.3 percent in most of a 3-inch wafer. This ultrahigh uniformity of the epitaxial material enables uniform VCSEL performance of a large number of VCSEL arrays. Our results show the potential of

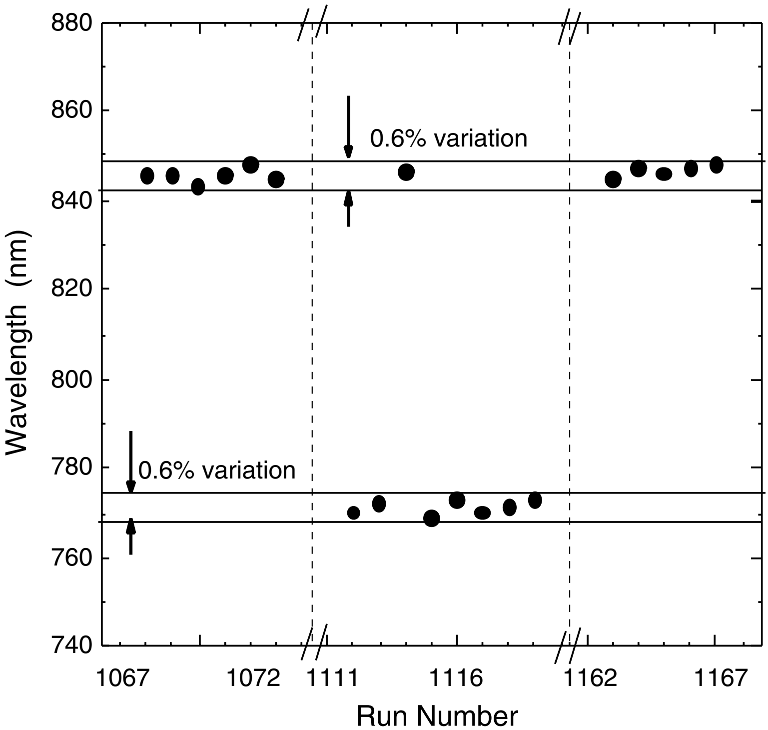

FIGURE 4 Reproducibility of the cavity wavelength of 770- and 850-nm VCSEL structures versus the growth run number. Pregrowth calibrations with the in situ reflectometer were made with Runs 1067, 1111, and 1162.

MOCVD as a very stable, reproducible, and uniform growth platform for VCSEL manufacturing.

REFERENCES

Aspnes, D. E., W. E. Quinn, and S. Gregory. 1990. Application of ellipsometry to crystal growth by organometallic molecular beam epitaxy. Applied Physics Letters 56(25):2569–2571.

Breiland, W. G., and K. P. Killeen. 1995. A virtual interface method for extracting growth rates and high temperature optical constants from thin semiconductor films using *in situ* normal incidence reflectance. Journal of Applied Physics 78(11):6726–6736.

Choquette, K. D., H. Q. Hou, K. L. Lear, H. C. Chui, K. M. Geib, B. E. Hammons, and A. Mar. 1996. Self-pulsing oxide-confined vertical-cavity lasers with ultralow operating current. Electronics Letters 32(5):459–460.

Choquette, K. D., K. L. Lear, R. P. Schneider Jr., K. M. Geib, J. J. Figiel, and R. Hull. 1995. Fabrication and performance of selectively oxidized vertical-cavity lasers. IEEE Photonics Technology Letters 7(11):1237–1239.

Hibbs-Brenner, M. K., R. A. Morgan, R. A. Walterson, J. A. Lehman, E. L. Kalweit, S. Bounnak, T. Marta, and R. Gieske. 1996. Performance, uniformity, and yield of 850nm VCSEL's deposited by MOCVD. IEEE Photonics Technology Letter 8(1):7–9.

Hou, H. Q., W. G. Breiland, B. E. Hammons, and H. C. Chui. 1996. *In situ* growth rate measurements by normal-incidence reflectance during MOCVD growth. Electrochemical Society Proceedings 96(2):27–35.

Hou, H. Q., B. E. Hammons, and H. C. Chui. Submitted. Doping and *in situ* selective etching of AlGaAs with CCl_4 in metalorganic vapor phase epitaxy. Applied Physics Letters.

Houng, Y. M., M. R. T. Tan, B. W. Liang, S. Y. Wang, and D. E. Mars. 1994. *In situ* thickness monitoring and control for highly reproducible growth of distributed Bragg reflectors. Journal of Vacuum Science & Technology 12(2):1221–1224.

Kawai, H., S. Imanaga, K. Kaneko, and N. Watanabe. 1987. Complex refractive indices of AlGaAs at high temperature measured by *in situ* reflectometry during growth by metalorganic chemical vapor deposition. Journal of Applied Physics 61:328–332.

Lee, T. P., ed. 1995. Current Trends in Vertical Cavity Surface Emitting Lasers. Singapore: World Scientific.

Sun, S., E. Armour, K. Zheng, and S. D. Hersee. 1995. Ultra-Uniform MOCVD Growth of GaAs/AlGaAs for VCSELs. Pp. 123–127 in Proceedings of the 7^{th} Biennial Workshop on Organometallic Vapor Phase Epitaxy. Fort Meyers, Fla., Apr. 2–6, 1995.

Optical Applications of Microelectromechanical Systems

MING C. WU
Electrical Engineering Department
University of California
Los Angeles, California

ABSTRACT

Microelectromechanical systems (MEMS) technology has opened up many new opportunities for optics. For the first time, reliable microactuators and three-dimensional optomechanical structures can be monolithically integrated with micro-optical elements. This new technology will impact many applications including display, scanning, and telecommunications. In this paper, we will discuss two MEMS applications: optical systems on a chip and monolithic integration of a large array of optomechanical devices.

INTRODUCTION

Integrated optical systems have been a dream of many optics researchers for decades. Though some progress has been achieved in semiconductor optoelectronics, most of the other optical systems are still made of discrete components and assembled precisely on optical breadboards. As a result, these optical systems are expensive and bulky. Recent advances in MEMS have made it possible to produce compact optomechanical structures and microactuators at low cost, using batch-processing techniques. These advances have opened up many new possibilities for optical and optoelectronic systems, including optical systems on a chip, and monolithic integration of a large number of optomechanical devices. MEMS technology provides a paradigm shift for optics from labor-intensive manual assembly to batch fabrication using integrated circuit-type microfabrication. Compared with conventional optical systems, MEMS optical systems (also called micro-optoelectromechanical systems or MOEMS) offer many advantages (Muller and Lau, 1998; Wu, 1997) by being more compact, lighter, faster, more rugged, and more scalable. Their applications include projection and head-mounted displays, optical data storage, printing, optical scanners, switches, modulators, sensors, and optoelectronic components packaging.

This paper describes two types of applications for optical MEMS: (1) optical systems on a chip, and (2) monolithic integration of a large array of optomechanical devices. Specific examples will be given for each type of application.

OPTICAL SYSTEMS ON A CHIP

MEMS technology has made it possible, for the first time, to integrate an entire optical table onto a single silicon chip. Optical elements such as lenses, mirrors, and gratings are batch fabricated along with the XYZ stages (micropositioning stages with three independent degrees of freedom) and the microactuators. (See Figure 1 for an example of a single-chip MEMS optical system.) The discrete lenses and bulk optomechanical supports are now batch fabricated and monolithically integrated on a single chip. As a result, the size of the system can be dramatically reduced by one to two orders of magnitude.

The main challenge to implementing single-chip optical systems is the ability to make different optical components using the same fabrication process. Typical optical components for free-space optical systems include (1) optical elements such as lenses, mirrors, and refractive and diffractive optical elements; (2) three-dimensional optomechanical support, such as lens mounts; and (3) adjustable structures and actuators, such as XYZ micropositioners. In the last several years, our research group at the University of California, Los Angeles (UCLA) has been developing a MEMS optical bench technology that can simultaneously fabricate these three types of components using the same fabrication process. Our fabrication process is based on standard polysilicon surface-micromachining processes. Such processes are now available through commercial foundries (e.g., the multi-user MEMS processes, or MUMPs, offered by Cronos Microsystems). MUMPs is a three-polysilicon surface-micromachining process in which the two structural polysilicon layers can be selectively patterned and released to form free-standing structures. Details of the MUMPs process can be found at http://mems.mcnc.org/mumps.html.

One of the most commonly used optical components in free-space optical systems is XYZ stage. Building micro-XYZ stages using MEMS technology is a challenging task. The out-of-plane translation is particularly difficult since most of the surface-micromachined actuators move in the in-plane directions. We have overcome this difficulty by using a novel micro-elevator-by-self-assembly (MESA) structure (Fan and Wu, 1997). MESA consists of five polysilicon plates joined by microhinges. When the two outer plates are moved towards each other, the center plate will buckle up and rise above the substrate. By integrating MESA with the two sliding structures a full micro-XYZ stage with three independent degrees of freedom is achieved. The scanning electron

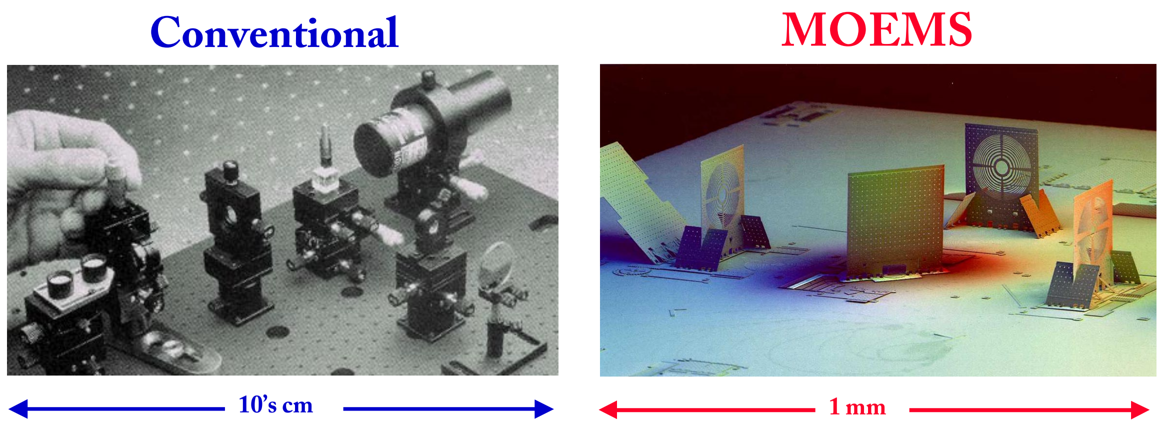

FIGURE 1 (Left) Conventional optical systems on a breadboard. (Right) MEMS optical systems on a single chip of silicon. SOURCE: (Left) Reprinted with permission from Melles Griot. (Right) UCLA. Photograph by David Scharf.

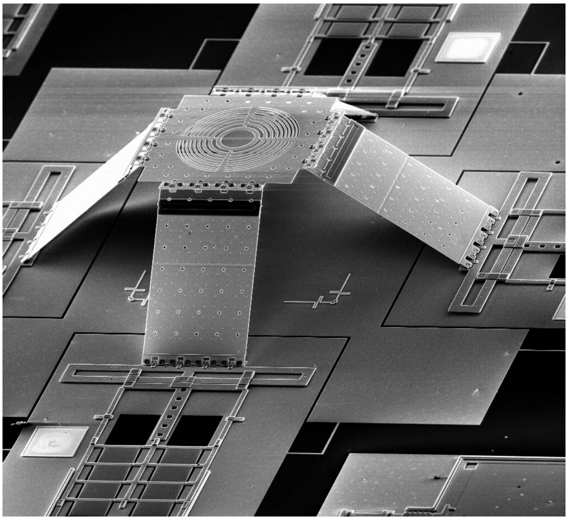

FIGURE 2 Scanning electron micrograph of a micro-Fresnel lens integrated on a micro-XYZ stage. The micro-XYZ stage has three independent degrees of freedom, with the travel distance greater than 100 μm in all three directions. SOURCE: UCLA.

micrograph of the micro-XYZ stage is shown in Figure 2. It is integrated with scratch drive actuator arrays. The scratch drive actuator is basically a stepper motor with an extremely fine step size (Akiyama et al., 1997). It consists of a polysilicon plate with a vertical bushing. Upon application of an electrical pulse, the electrostatic force will deform the polysilicon plate. The downward motion pushes the front bushing forward by a small step. Upon release of the bias, the friction at the front bushing, which is larger than at the backside of the plate, pulls the entire scratch drive actuator forward. The average step size is extremely fine, 20~30 nm, which is ideal for positioning optical elements requiring sub-0.1 μm accuracy.

Micro-optical elements can be monolithically integrated with the actuated micromechanical structures. One of the unique advantages of MEMS optical bench technology is that all three-dimensional structures are assembled after the fabrications are finished. Before the micromechanical structures are released, the surface of wafers remains planar. Therefore, we can combine MEMS with the well established planar micro-optics technology to build three-dimensional optical components. For example, a refractive microlens is monolithically integrated with a three-dimensional lens mount by using reflow technique. A cylinder of polymer (e.g., photoresist) is patterned on the polysilicon plate. It is then heated up at a high temperature, causing the polymer cylinder to reflow. The surface tension will pull the surface into a spherical shape. This is illustrated in Figure 3. Microlenses with a wide range of f—number and aperture can be

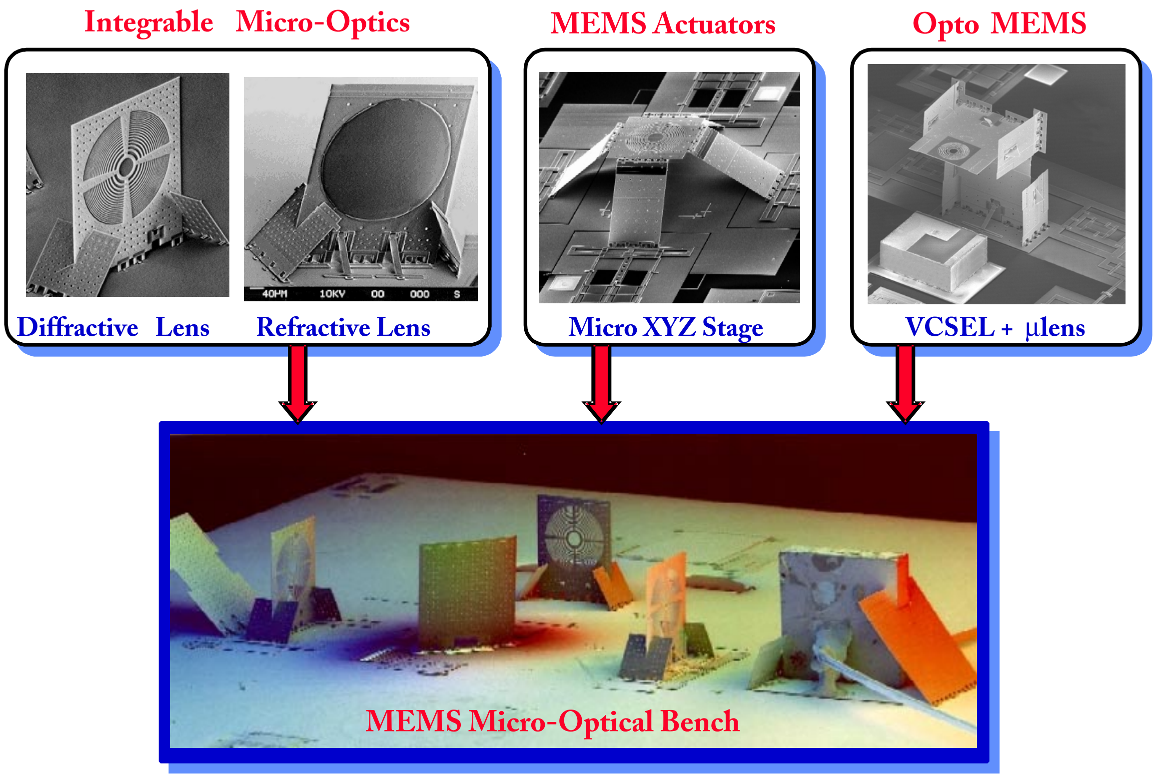

FIGURE 3 The basic building blocks of MEMS micro-optical bench. SOURCE: UCLA. Lower photograph by David Scharf.

made by this technique. Diffractive optical elements can also be readily incorporated into the MEMS structures. We have also successfully used the microfabricated XYZ stage to position a bulk microlens with 300 micrometer diameter. Hybrid integration with active optoelectronic source and detectors are also of interest. We have successfully integrated a vertical surface-emitting laser with a suspended microlens, as shown in Figure 3.

Using the building blocks described above, we have built several optical systems on MEMS micro-optical benches. A femtosecond autocorrelator has been constructed to measure the pulse width of extremely short optical pulses (120 femtoseconds has been demonstrated experimentally). A single-chip optical disk pickup head has also been realized.

LARGE ARRAY OF OPTICAL MEMS DEVICES

Like integrated circuits, a large number of optical devices can be integrated on the same chip using MEMS technology. This is similar to the random access memory of electronic integrated circuits. The most well-known example of this type of optical MEMS device is perhaps the digital micromirror device of Texas Instruments, which consists of more than one million torsion mirrors. The digital micromirror device (DMD™) has been used in commercial projection display systems, and is particularly attractive for compact portable projection systems.

In addition to display, MEMS optical devices are very attractive for telecommunication applications, including optical switches, optical crossconnect, wavelength division add/drop multiplexers, and tunable lasers/detectors. MEMS optical switches offer many advantages over other types of switches, including low insertion loss and low crosstalk. They are transparent to the optical data and are independent of polarization, wavelength, bit rate, and modulation format. Moreover, MEMS technology allows a large number of switches to be integrated on a single substrate. This is very attractive for optical crossconnect applications. The basic building blocks of a MEMS optical crossconnect is illustrated in Figure 4.

Under Defense Advance Research Project Agency support, UCLA has been investigating manufacturable, low-cost fiber optic switches using MEMS technology. Figure 5 shows the scanning electron micrograph of our first-generation MEMS fiber optic switch (Lee et al., 1997). It consists of a moveable micromirror, four fiber guides, a mechanical restoring spring, and microactuators. Depending on the mirror position, light emitted from the input fiber is either transmitted to the opposite fiber or reflected to the orthogonal fiber. The scratch drive actuator and a restoring spring are used to move the micromirror. The restoring spring allows the switch to be used as a bypass switch for the fiber distribution data interface (FDDI) ring networks. During power failure, the spring always pulls the mirror back to the bypass state to maintain the continuity of the fiber ring network. The switch has been successfully demonstrated. Low

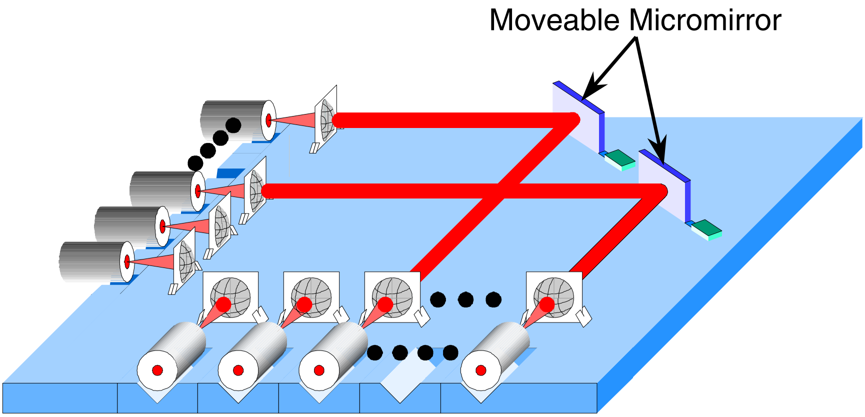

FIGURE 4 Schematic of MEMS optical crossconnect. SOURCE: UCLA.

insertion loss was obtained, however the switching time is on the order of 10 msec. Though this is sufficient for FDDI applications, sub-millisecond switching times and low operating voltages are desired for many other applications.

For optical crossconnect applications large mirror areas and low operating voltages are desired. There is, however, a trade-off between these two parameters. A large mirror area requires a large displacement, which translates into high operating voltages for typical electrostatic actuators. Recently, we devel-

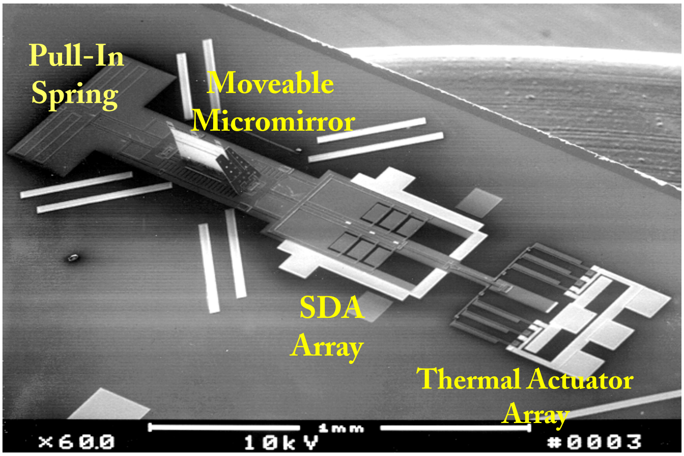

FIGURE 5 Schematic of the first-generation MEMS fiber optic switch made at UCLA. SOURCE: UCLA.

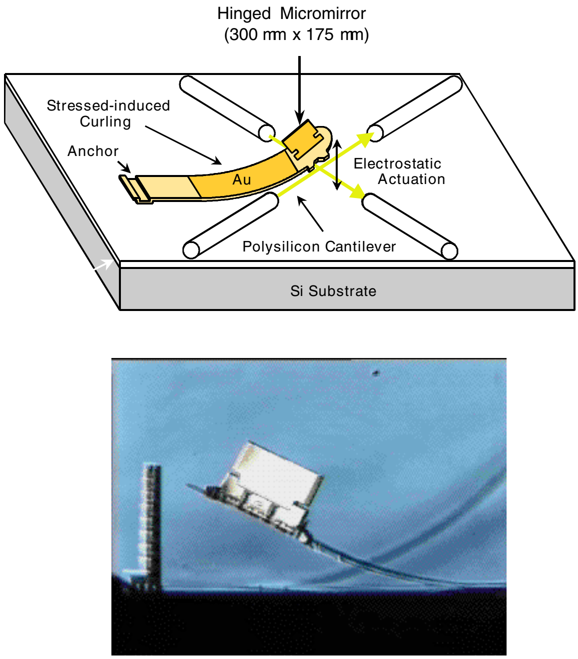

FIGURE 6 Schematic and photograph of the curled cantilever MEMS optical switches. SOURCE: UCLA.

oped a novel curled cantilever switch (Chen et al., 1999) that can overcome this trade-off. Figure 6 shows the schematic drawing and photograph of the switch. The micromirror is attached to the end of the curled cantilever beam. The actuation voltage is determined by the initial small gap, while the mirror displacement is determined by the motion of the tip. We have achieved a operating voltage of 20 volts for a mirror displacement of 300 micron. Low insertion loss (0.7 dB) and low crosstalk (< -80 dB) have also been achieved. The switching time is less than 1 msec.

One unique advantage of the MEMS optical switch is its ability to scale up to large switch arrays. With the miniaturization of MEMS, a large number of the optical switches can be integrated monolithically on the silicon substrate. This technology is ideally suited for optical crossconnect applications. Optical crossconnects with a large number of input-output fibers and a very low insertion loss and crosstalk are key components for reconfigurable dense wavelength-division-multiplexed optical fiber networks. Recently, an 8x8 MEMS optical crossconnect was demonstrated (Lin et al., 1998). Continued development in optical MEMS technology could lead to even larger switches in the near future.

SUMMARY

The ability to integrate optical and actuated mechanical components on the same chip using MEMS technology opens up many new opportunities. An optical switch with small insertion loss and low crosstalk is just one example. Eventually, an entire optical system, such as the large-scale fiber optic switch matrix could be integrated monolithically on a single chip. A new family of optical and optoelectronic devices and systems could be built at low cost with batch-processing techniques. These devices will be smaller, lighter, faster, more rugged, and more functional than their bulk optomechanical counterparts. This new technology will have a significant impact on optical switching, telecommunication, display, printing, scanning, and optical data storage applications in the near future.

ACKNOWLEDGMENTS

The author would like to thank his former and current graduate students at UCLA for their contributions to this work. The work at UCLA is supported in part by the Defense Advanced Research Project Agency and the Packard Foundation.

REFERENCES

Akiyama, T., D. Collard, and H. Fujita. 1997. Scratch drive actuator with mechanical links for self-assembly of three-dimensional MEMS. Journal of Microelectromechanical Systems 6(1):10–17.

Chen, R., H. Nguyen, and M. C. Wu. 1999. A low-voltage micromachined optical switch by stress-induced bending. Pp. 424–428 in Proceedings of the IEEE 12th International Workshop on Micro Electro Mechanical Systems (MEMS '99), Orlando, Fla., January 17–21, 1999. New York: Institute of Electrical and Electronics Engineers.

Fan, L., M. C. Wu, K. D. Choquette, and M. H. Crawford. 1997. Self-assembled microactuated XYZ stages for optical scanning and alignment. Pp. 319–322 in Proceedings of the 1997 International Conference on Solid-State Sensors and Actuators (Transducers 97), Chicago, Ill., June 16–19, 1997. New York: Institute of Electrical and Electronics Engineers.

Lee, S. S., E. Motamedi, and M. C. Wu. 1997. Surface-micromachined free-space fiber optic switches with integrated microactuators for optical fiber communication systems. Pp. 85–88 in Proceedings of the 1997 International Conference on Solid-State Sensors and Actuators (Transducers 97), Chicago, Ill., June 16–19, 1997. New York: Institute of Electrical and Electronics Engineers.

Lin, L. Y., E. L. Goldstein, and R. W. Tkach. 1998. Free-space micromachined optical switches with submillisecond switching time for large-scale optical crossconnects. IEEE Photonics Technology Letters 10:525–527.

Muller, R. S., and K. Y. Lau. 1998. Surface-micromachined microoptical elements and systems. Proceedings of the IEEE 86:1705–1720.

Wu, M. C. 1997. Micromachining for optical and optoelectronic systems. Proceedings of the IEEE 85:1833–1856.

Dinner Speech

Career Flexibility in Rapidly Changing Times

KENT KRESA
Northrop Grumman Corporation
Los Angeles, California

It is an honor to be here with such a select group of the country's best and brightest. First, I congratulate all of you for being chosen for this symposium. This year's symposium addresses several exciting areas, including information technology, chemical and biological engineering, optics, and energy. You are all certainly getting your minds stretched. As you heard, I am in the defense business, and a couple of these topics are extremely important to my industry and me.

Information technology, for example, reminds me, as a designer and builder of defense electronics and other battlefield management systems, that information warfare is right behind it. The interesting part of information technology is that while the "good guys" evolve the technology to create and protect the systems of tomorrow, the "bad guys" are trying to bring down these same systems. The more our society becomes dependent on these systems, the more susceptible we will be to sabotage and information warfare. Cyberspace is the new dimension that will tie the world together and, as you can imagine, it is a key concern of the defense industry.

Certainly chemical and biological weapons, another focus of this year's symposium, are also a great concern for my industry. My past experiences in the 1970s with these issues suggest that we will make little headway unless we get better intelligence. And, quite honestly, I am not confident that we will be able to handle this threat of bugs and gas, given our overall lack of success in drug interdiction. But that is the nature of today's world. Small and rogue states may learn how to create these weapons a lot easier than ballistic missiles, which take a larger infrastructure.

As you heard in the introduction, I have engineering degrees, but I haven't practiced engineering for 25 years. Some of you, like me over the past two and a half decades, are also focusing on management issues, and I am sure that, over time, more of you will spend a greater percentage of your day in that management capacity. I found that my engineering background has been a tremendous benefit to me as a manager in two ways. First, engineering and its black-and-white, hard-lined perspective allows you to look at problems in a relatively dispassionate way, hopefully leading to quicker and more productive solutions. As an engineer, using raw data and hard analysis, I try to understand the conditions and how I can solve the issue, whether it is a technical problem or a human relations problem.

The second benefit of having an engineering background has been the adoption of an enormous "BS filter." Proposals from my organization and from the outside are thrown at me every day. Applying an engineering perspective to issues has allowed me to quickly turn on that filter, even as people are explaining these proposals. So don't be afraid if you've got to go into the management side of things, because, believe it or not, your engineering background is really going to help.

As I thought about my talk tonight, I realized that I certainly can't tell you anything new about engineering disciplines. But what I might be able to offer is a little bit about how I see the world. The first thing I'd like to say is that change is accelerating, and if I remember my engineering terms correctly, that's a third derivative. Change, including in my business, is occurring at such an intense rate—far faster than at any time that I have ever witnessed. This change will undoubtedly continue to affect your own careers and actually your career structure. It used to be that an engineer could go through a whole career while focusing on a specific engineering discipline. Today, engineers probably maintain a specific discipline for half of their career, at most.

As a result, you are going to have to become quite adaptable, innovative, and agile in order to have a full and prosperous career. Product cycle times are coming down dramatically. Today they are measured in months, not years. And it wasn't so long ago that American business cycle times were operating in decades—in fact I was involved in one that took more than a decade.

The real reasons change is happening are computers and communications and the ability to automate processes in various ways. Projects that I spent quite a bit of time on as a young engineer are done trivially today on computers. As a matter a fact, I almost wasn't an engineer. At MIT, I was in a co-op program and then went off to Boeing as a junior to learn about engineering. I was placed in a room with 1,200 people with Frieden calculators (mechanical calculators to calculate the stresses on an aircraft wing). This process amounted to a human computer—the information would come in one end of the room and go out the other. I quickly realized that this was not what I wanted to do for the rest of my life.

Now that was engineering in the 1950s, at least for the aircraft industry. When I went back to school, professors told me that a "new world" was coming involving space and other great advancements in aviation. These comments excited me, and so I remained an engineer. Change really began to occur in the late 1950s and early 1960s, and change has been taking place ever since.

Today, change is so intense that I am not clear how businesses are going to operate in the next century. The barriers to market penetration and the ability to guess wrong, and therefore be out of business in a very short period of time, are something new. Size used to matter quite a bit in business, because a company could weather an occasional storm. This is not true anymore. Just look at what is happening with marketing on the Internet and the effect this is having on people's buying habits.

Blockbuster, for example, is roughly only a decade old, and has experienced an extremely impressive growth spurt with the evolution of videotapes. Now, just 10 short years later, the company faces extinction. Management understands this dilemma and is working hard to transform Blockbuster's business. But what Blockbuster is struggling with after a decade of growth is occurring in just a few years at many other companies.

I was deeply involved in a significant time of change while on the board of Chrysler Corporation. During the 1980s, the American auto industry was under tremendous pressure. The Japanese automobile manufacturers went from a modest presence in the United States to becoming a significant force. They were producing great high-quality cars at a much lower cost than General Motors, Chrysler, or Ford.

Management was initially concerned, but it discounted the severity of the problem and decided that Japan's lower labor and capital costs were the sole reasons for that country's success. But when the Japanese began to open factories on U.S. soil with the same impressive results, management at the Big Three automakers took notice. The American managers realized that the Japanese system focused essentially on cycle time and value stream analysis—novel concepts back then. In those days, teaming was not even considered. A stylist designed the car, then threw it over the fence to the engineers. The engineers turned over the process to the production group and finally a car came off the production line. It was the Japanese model that dramatically changed this process. Partnership agreements were entered into to take advantage of Japanese production methods while retaining American strengths in automation.

As a result of this dramatic shift in operating philosophy, both U.S. and Japanese auto industries are more robust, innovative, and cost efficient today than ever before. The real winner is the consumer, because the quality of today's cars is so much better than a decade ago. This is true of virtually any product these days. I can remember when consumers always purchased the warranty when buying a television, because they would last only six months or so. Today, warranties are not nearly as popular because the quality of the products is so good.

One very big project that required a change of thinking for our company engineers was the B-2 stealth bomber. When we started this project in 1987, we went beyond state-of-the-art in every conceivable direction. In the decade that followed, there were about 3,000 patents that were produced to make this aircraft a reality. It was the first aircraft to be designed, developed, and manufactured on a computer screen. We networked plants all over the country to conduct this computer-aided design in a highly secure fashion. As you can imagine, it was an enormous undertaking. The network program was definitely way ahead of its time, but just a decade later it's an extremely common approach. Everybody's got CAD for whatever purpose they want, and it's probably on their PCs.

This time lapse will continue to shrink. I know some of you are in businesses that have six-month cycle times to get a product out to the market. I would imagine that pretty soon cycle time would be measured in months for most businesses. This approach will certainly have a tremendous impact on your careers.

Another important quality of a good engineer is vision. Vision may be a scholastic process. I think that everybody has a vision, but most people fail at those visions. Every one of us, I'm sure, has examples of key visionaries. There are two that come to mind for me. The first one is of a couple of guys at ARPA in the late 1960s who pushed the concept of packet switching. They pushed because they needed a cheap way to couple computers at various large research universities in order to perform complex nuclear bomb computations without having to build a massive computational facility. So they developed a device—I think it's called an IMP—and went around selling the concept to universities. Over time, people at universities figured out how to send data on this equipment, and every now and then, they added a message to their colleagues along with the data. This was the start of the ARPANET, which has evolved from ".edu" to ".mil" to ".com."

The other group of visionaries, in the early 1970s, came from government. They developed the global positioning satellite. The GPS was not in vogue at that time, in fact, the military services did not want it. At the time, the Navy and Air Force each had its own systems, and there were other navigational schemes going on. A few visionaries pushed extremely hard for a unified system and, over time, others determined that this unified approach could open tremendous possibilities for military and nonmilitary functions, including precision weapons and navigation.

For every successful visionary story, I am sure there are dozens of failures. If you truly believe in your product or idea, you must maintain your drive to make it happen. Again, if you have a vision, pursue it with all of your might. And if you fail, get back up and try something new. The system will accept failure as long as people are trying to push great new ideas.

The final piece of advice I will share with you is to have fun with your job. Many people take their jobs very seriously, and that is important and admirable. But we spend more than the traditional 40 hours a week, including Saturdays and

Sundays, worrying about our jobs and careers. My point is, enjoy the ride. You are all in enviable positions as leaders in your professions. You are part of exciting and rapidly changing times. Enjoy it! And maybe, when you get to be as old as I am, you can stand before a group of bright young engineers and discuss some great visions that you contributed to. I wish you continued success in your careers.

APPENDIXES

Break-out Session Outcomes

INTRODUCTION

At the 1999 Frontiers of Engineering meeting, a new element—break-out sessions—was added to the program. On each of the first two days of the symposium, one-hour break-out sessions were held, and on Saturday there was a plenary session to discuss selected break-out session topics. The goal of the break-out sessions was primarily to give an opportunity for participants to talk in smaller groups about engineering issues of importance to them. However, after the first day, it became evident that the results of the sessions also could provide input to the format for next year's sessions, perhaps by providing a narrower focus to the topics discussed, or could guide the selection of the dinner speaker. The outcomes of the break-out sessions are summarized here.*

DAY 1

Participants were assigned to groups of approximately 15 members. Attention was paid to ensure that each group was diverse in terms of the kinds of work institutions and engineering fields represented. Participants were asked to answer one of these questions:

- **What are the three most significant issues facing engineers today?**
- **What will be the three most significant issues facing engineers in the next 10 years?**

* Recommendations and comments are those of the Frontiers participants, and publication in this volume does not indicate an endorsement by the National Academy of Engineering.

Responses were as follows and are not necessarily in priority order:

Q1: What are the three most significant issues facing engineers today?

Red Group:

1. Funding
   - The need for better industry and academic partnerships especially relating to commercialization of intellectual property
   - The need to extend the payback period on research investment beyond the short term

2. Interdisciplinary nature of engineering research
   - The multidisciplinary education required by engineers today results in greater need of interdisciplinary communication

3. Communication
   - The communication gap between engineers and the general public results in poor policy based on uninformed or irrational information
   - The public does not appreciate engineering and often makes uninformed decisions
   - Engineers need to communicate better with policymakers

Dark Blue Group:

1. An increasingly broad skill set is required to avoid commodification (work being outsourced). This means education of engineers needs to be changed to supply students with these skills, and practicing engineers need to maintain and broaden their skills.

2. Fostering two-way communication between the engineering community and the "rest of the world" in order to increase respect, improve public perception, and foster responsibility to the community, with emphasis on ethical issues.

3. Instilling in students and new engineers solid engineering judgment, understanding of fundamentals, and value-added proposition in the context of changing technology. Defining the technical challenges of the knowledge age. Developing the tools and technologies needed in the new knowledge age.

Green Group:

1. Lack of influence by engineering on public policy
2. Keeping up with current information
3. The gap between theory and practice

Pink Group:

1. Professional/continuing education – keeping pace with technology, functioning on multidisciplinary projects

2. Business aspects of engineering/globalization
3. Education – quality of science and engineering education in the United States (secondary and college)
4. Supply and demand of engineers/women in engineering

Q2: What will be the three most significant issues facing engineers in the next 10 years?

Brown Group:

1. Engineering policy
   - Advocacy for the profession
   - Continuing support for long-term/basic research
   - Public education and awareness, improved visibility

2. Professional renewal
   - Continuing education
   - Maintaining interdisciplinary knowledge, individually and in groups (to counteract overspecialization)
   - Teaching critical thinking rather than just handling facts and formulas

3. Handling data
   - Management of information sources
   - Interpretation/utilization of existing knowledge bases
   - Processing of data. Information → knowledge

Yellow Group:

1. Improved education in the face of rapid change
   - Effective teaching of scientific fundamentals
   - Effective teaching of breakthroughs at the interface of traditional disciplines
   - Effective continuous life-long education
   - Effective scientific/engineering immersion at K-12 by impacting teacher/guidance counselor training including continuous involvement with the engineering community

2. Ethical issues
   - Global environment
   - Integrity
   - Human/technological interface
   - Protection of privacy

3. Increasing complexity
   - Managing the information explosion

- Verification of complex systems
- Interdependence of technologies
- Security issues

Light Blue Group:

1. Education
   - Addressing the need for interdisciplinary knowledge while retaining expertise
   - Clarifying the mission of engineers to society as a whole → increasing the draw of American citizens to engineering education
   - Increase visibility of engineering role models

2. Communication
   - Being able to effectively communicate to peers, superiors, public, etc.

3. Ethics
   - Engineers' ability to address ethical issues related to technology advancement, e.g., privacy issues in computer software/communication and bioengineering

4. Interfacing multiple systems and managing complexity and unintended consequences

5. Public policy
   - Implications of technology and society
   - Engineering talent from foreign countries
   - Technology investment

Gray Group:

1. Predicting and coping with the ethical and social impacts of new and expanding technologies
2. Pressures on the engineer to keep up with changes in the workforce and rapidly expanding technology
3. Understanding and managing engineering innovations as applied to complex systems, e.g., environmental, public health, economic, and sustainability issues

Not surprisingly, there was no discernible difference between the responses given to the first question (challenges facing engineers today) and responses given to the second question (challenges in the next 10 years).

DAY 2

For this day's break-out session, the previous day's responses were evaluated. It was determined that the responses tended to coalesce around three significant issues, each with several subpoints. These issues were:

Issue 1: Engineering Education in the 21st Century
- Lifelong learning
- Ethics
- Increasingly interdisciplinary nature of engineering
- Preparation at the K-12 level for engineering careers
- Teaching and learning engineering fundamentals and critical thinking

Issue 2: Engineering and Public Policy
- Communication between engineers and the public and engineers and public policymakers
- Advocacy by the engineering community
- Support for R&D (short-term vs. long-term objectives)
- Ways to avoid uninformed public policy decisions

Issue 3: Managing Complexity/Information Explosion
- Implications for education
- Need for a systems perspective/multidisciplinary approach in engineering
- Effect of globalization
- Industry-university partnerships

Groups were given one of the issues with at least two of its related subpoints and asked to discuss concrete steps that could be taken to address them.

Issue 1: Engineering Education in the 21st Century

1.1 Lifelong learning
- Need formal accreditation or other reward mechanisms for short courses
- Companies should freely support continuing education

1.2 Increasingly interdisciplinary nature of engineering
- Students should be involved in interdisciplinary teams in introductory engineering courses, senior-level project courses, graduate-level projects/courses
- Use case studies of post-hoc analysis of failures/disasters to illustrate key issues: There is a need for many disciplines (within and outside engineering) in order to understand a problem. Illustrate ethical issues in decisions affecting the outcomes of the disaster, e.g., cost vs. safety, adherence to procedures

- Support collaboration among universities, professional societies, industry, etc. to develop modules or case studies for integration into curricula
- Internships/co-op programs: While generally valuable to students and employers, the quality of the experience varies. Require students to have at least two different experiences
- Create internships with faculty members, involve undergraduates in research
- Faculty could serve as role models by doing interdisciplinary work
- University curriculum issues: Require/reward interdisciplinary teamwork. Offer more broad-based core courses across disciplines. (But what to eliminate? 5-yr program?)
- The burden should not all be on the university and formal education
- Require one major interdisciplinary team project as undergrad
- Encourage teamwork even within discipline
- Change the tenure value system, which discourages interdisciplinary work
- In industry, expose young employees to several areas of the business early
- Promote continued learning on the job

1.3 Preparation at the K-12 level for engineering careers

- Problems
  - Many science teachers do not know science well enough to teach it adequately—perhaps doing more harm than good
  - Guidance counselors are inadequately informed about engineering careers
  - US education system is oriented towards common denominator, passing everyone (unlike some in Europe, for example)
  - Media lacks positive (any?) portrayal of engineers or engineering
  - Teamwork and communication skills are essential to learn at the high school level
- Proposals
  - Encourage the media to show engineers and engineering positively—TV, cartoons, web
  - Distribute informational packages to teachers
  - Educate guidance counselors about engineering careers
  - Institute more stringent requirements for students
  - Encourage volunteer activities (industry, academic, and personal initiatives)—help at schools, talk to classes, talk to guidance counselors
  - Encourage retiring engineers/professors to teach in high schools
  - Start before college to bring new people into our profession. Answer the following for high school students: What is engineering about? What preparation is needed in order to be an engineer? What is a typical day in the life of an engineer? Include examples that touch upon ethics and

consequences of engineering activities and upon multidisciplinary consideration.

- Develop a framework for interaction, e.g., institutional relationships

1.4 Teaching and learning engineering fundamentals and critical thinking

- Educational Needs Assessment

From an industry perspective, the graduate needs a strong background in a fundamental area, appreciation of other fields, ability to be open and communicate, job-specific skills, "drive." At the graduate level, expect person to be able to step in more readily as a "design leader."

In general, graduates need to acknowledge/appreciate what others contribute, including those in other disciplines, as well as technicians, machinists, etc.

- Length and Sequencing of Educational Activities

Can only do so much in a B.S. degree program, therefore education must be an ongoing, lifelong process. Many practicing engineers will get master's degrees and many companies provide training.

When to specialize vs. generalize? Different models may work for different people, e.g., "specialize" in traditional disciplinary program at B.S. level and expand on disciplinary grounding at graduate level or begin with a broad degree (B.A.) and then specialize at graduate level. Marketability . . . will companies hire a nontraditional engineer?

- Role of Master's Level Education

It may be easier to modify master's vs. B.S. degree programs. Students already have some discipline-specific fundamentals. Convey a sense of how things fit together and interact. Have multidisciplinary teams tackle real-world projects, e.g., analyze failures.

1.5 Other Considerations

- Money talks. Funding is needed to encourage interdisciplinary education, e.g., from NSF
- Use Fundamentals of Engineering exam as a motivator for interdisciplinary study within engineering
- Bring in outside experts to address ethics and interdisciplinary issues

Issue 2: Engineering and Public Policy

2.1 Communication between engineers and the public and engineers and public policymakers

- What would we say to each constituency?
  - To the public: notice and respect us
  - To policymakers: need more money and a greater voice

- Proposals
  - Link professional societies
  - Improve public image
  - Outreach

2.2 Advocacy for the engineering community

- No single "silver bullet"
- Encourage the emergence of an engineering spokesperson at the national political level, e.g., the next Vannevar Bush, a technology adviser to the president, a technology laureate
- Identify and emulate best practices in influencing policy, e.g., the medical community, bio community
- Expose people to positive aspects of engineering, e.g., the human side, require engineering for nonengineers in college, inform students at a young age, get an engineer on Sesame Street
- Heavily publicize national engineering awards
- Engage the popular press—editorials, etc.

2.3 Support for R&D (short-term vs. long-term objectives)

- Problems
  - The time necessary to generate support for R&D funding is greater than election cycles
  - A crisis is needed to galvanize support for R&D funding, and there are few of these
- Proposals
  - Articulate the importance of R&D
  - Advertise previous successes
  - Create a crisis???

2.4 Ways to avoid uninformed public policy decisions

- Identify points of contact in congressional offices and with congressmen/senators with interests in science and technology
- Prepare fast response to hot issues, e.g., cloning
- Write to representatives
- Be proactive with influential politicians
- Take a unified stand on issues among professional societies; project the voice of engineering
- Sponsor the Super Bowl—use public service messages aimed at a general, nontechnical audience

Issue 3: Managing Complexity/Information Explosion

3.1 Implications for education

- Combine engineering departments
- Students follow tracks that gradually diverge
- Students take mandatory class on complex problem solving
- All students take capstone design course together
- Introduce complex simulation software packages; learn how to perform sensitivity analyses with these
- Emphasize iterative problem solving, e.g., design, test, redesign
- Real problems introduced at all levels
- Teach people to access information: Identify erroneous information and filter irrelevant information

3.2 Need for a systems perspective/multidisciplinary approach in engineering

- Cross list more courses
- Require multidisciplinary "capstone" design projects
- Return to an interdisciplinary set of core courses
- Encourage faculty collaboration across disciplines and joint appointments
- Eliminate departments
- Improve industry/university post-hiring education (courses, leadership programs)
- Require senior-level multidisciplinary design project
- Influence ABET policy (include a complexity component)

3.3 Effect of globalization

- Take advantage of communication and data sharing capabilities to work on problems with multinational teams 24 hours/day
- Teach second language and communication skills to engineers
- Use immersion programs to learn other cultures
- Design for global markets, but use mass customization to access all markets
- Promote diversity
- Increase machine translation of journals
- Utilize international/distributed design teams
- Consider global/environmental issues in design projects
- Understand human/technology interface
- Cultivate relationships for research exchange on a global level

3.4 Industry-university partnerships

- Address intellectual property issues: What is fair? Security of information. Up-front agreements
- Work towards longer-term relationships
- Create "centers" and consortia

- Increase government role in funding/mandating cooperation
- Increase the value of industrial research in the tenure process
- Get more industrial representation on academic advisory boards (and vice versa)
- Encourage industrial and academic sabbaticals

Contributors

THOMAS R. ALBRECHT is a research staff member and manager at IBM Almaden Research Center in San Jose, California. Dr. Albrecht received a B.A. in physics from Carleton College in Northfield, Minnesota and a Ph.D. in applied physics from Stanford University. His thesis research at Stanford focused on Atomic Force Microscopy (AFM), and included the first demonstration of micro-fabricated force-sensing cantilever probes, which are used routinely in AFM today. After graduate school, he worked briefly for Park Scientific Instruments, helping launch their first AFM product. At IBM, Dr. Albrecht's work has centered primarily on technologies for tape and disk drives. In tape technology, he contributed to a novel track-following servo technology, referred to as Timing-Based Servo, which allows higher track density and higher reliability. This technology, along with various mechanical designs of his, has been implemented in IBM's 3570 Magstar MP family of tape products, and has become the standard servo technology for the newly-launched Linear Tape Open (LTO) tape platform adopted by Hewlett-Packard, IBM, and Seagate. For disk drives, Dr. Albrecht spearheaded the development of ramp load/unload technology for IBM's Travelstar mobile disk drives. His most recent project has been mechanical design work for IBM's Microdrive 1-inch disk drive, the world's smallest, lightest disk drive intended for new markets such as digital cameras and handheld computers. Dr. Albrecht has 30 issued U.S. patents and 28 publications. He has been elected to the IBM Academy of Technology.

THOMAS M. BRENNAN is vice president and general manager of EMCORE PhotoVoltaics in Albuquerque, New Mexico. Prior to this assignment, he was the co-founder and co-president of MicroOptical Devices (MODE), an Albuquerque-based high technology company, which was purchased by EMCORE in December, 1997. MODE manufactures vertical-cavity surface-emitting laser (VCSEL) components and integrated modules in volume. VCSELs are revolutionary compound semiconductor microlaser diodes that emit light vertically from the surface of the fabricated wafer. Prior to founding MODE, Mr. Brennan was a senior member of the technical staff at Sandia National Labs from 1986 to 1996, where he focused his efforts on the material growth of III-V compound semiconductors, reactor design, in situ reactor diagnostics, and material characterization. His activities included growth of some of the first VCSEL material at Sandia and in the U.S. and development of new and unique manufacturing techniques for VCSEL material growth. Prior to joining Sandia, Mr. Brennan was a member of the technical staff at Bell Communications Research and was an associate member of the technical staff at AT&T Bell Labs. At both facilities, he focused his efforts on epitaxial materials growth and characterization and epitaxial reactor design. Mr. Brennan holds an A.S. in chemistry from the Florida Institute of Technology, a B.S. in biochemistry from Monmouth College, and an M.S. in solid state physics from Stevens Institute of Technology. He has published more than 100 articles and holds 12 patents issued and pending.

ARUP K. CHAKRABORTY is a professor of chemical engineering and professor of chemistry at the University of California, Berkeley. His research interests are in polymeric materials science and engineering, interfacial phenomena, zeolite catalysis and separations, quantum mechanics, and statistical mechanics. Dr. Chakraborty received his Ph.D. in chemical engineering from the University of Delaware and was a postdoctoral fellow in chemical engineering and materials science at the University of Minnesota. His awards include: research professor, Miller Institute for Basic Science, Berkeley; the Allan P. Colburn Award from AIChE, the Camille Dreyfus Teacher-Scholar Award, and the NSF Young Investigator Award.

JAMES M. CHAVEZ is technical manager of the Solar Thermal Test Department at Sandia National Laboratories and also manages Sandia policy and governmental interactions on energy issues. He has been involved in the research, development, and demonstration of renewable technologies for utility applications since 1986. During this time he worked on both large and small solar power generation technologies. From 1994 to1999, Mr. Chavez managed Sandia National Laboratories' activities for the Department of Energy's Concentrating Solar Power Program, Biomass Power Program, and Solar Buildings Program. He also managed the National Solar Thermal Test Facility at Sandia National Labs, where he was responsible for the technical staff and for all testing at the largest U.S. solar test facility. From 1991 to 1995, Mr. Chavez was the technical advisor and chair of the Solar Two Technical Advisory Committee to successfully implement the Solar Two 10 MW_e molten salt solar power plant in

Barstow, California. He is an active member of ASME and was a recipient of the 1997 Hispanic Professional Engineer Award, for professional achievement in solar energy development. Mr. Chavez received his B.S. in mechanical engineering from New Mexico State University and M.S. in mechanical engineering from the University of California, Berkeley.

JANE H. DAVIDSON is a professor of mechanical engineering at the University of Minnesota where she conducts fundamental and applied research in solar energy and air pollution control. Current emphases are in electrohydrodynamics, mixed convection heat transfer in thermosyphon heat exchangers, development of polymer heat exchangers for liquid-to-liquid heat transfer, reduction of ozone generation rates in electrostatic air cleaners, chemical vapor deposition in gaseous electrical discharges, and control of indoor air pollution. Dr. Davidson has held positions at Union Carbide Nuclear Division in Oak Ridge, Tenn.; the International Agency for Research on Cancer in Lyon, France; Research Triangle Institute; Duke University; and Colorado State University. She is a Fellow of the ASME and recipient of numerous ASME Best Paper Awards. Dr. Davidson received her bachelor's and master's degrees from the University of Tennessee and Ph.D. from Duke University.

OREN ETZIONI is currently on leave from his position as an associate professor in the Department of Computer Science and Engineering at the University of Washington to serve as chief technical officer of Go2net, Inc., based in Seattle, Washington. He received his bachelor's degree in computer science from Harvard University and his Ph.D. from Carnegie Mellon. He joined the University of Washington as an assistant professor of computer science and engineering in 1991 and was promoted to associate professor in 1996. In Fall 1991, Dr. Etzioni launched the Internet Softbots project. In 1993, he received an NSF Young Investigator Award. In 1995, his Rodney Softbot was chosen as one of five finalists in the National DISCOVER Awards for Technological Innovation in Computer Software. Dr. Etzioni has published more than 50 technical papers in the fields of web search, intelligent agents, data mining, user interfaces, and artificial intelligence. He is the co-founder of Netbot, Inc., which was acquired by Excite in 1997. Dr. Etzioni's research has yielded two award-winning search tools, which have been transferred to industry. His group's work on software agents has been featured in *The New York Times*, *The Economist*, *Business Week*, *Discover Magazine*, *Forbes Magazine*, and *The New Scientist*. His research interests include: software agents, web navigation and search technology, and human-computer interaction.

STEPHEN P. A. FODOR is chairman and CEO of Affymetrix in Santa Clara, California. He received his Ph.D. in chemistry at Princeton, and was an NIH post-doctoral fellow at the University of California, Berkeley. In 1989, his group at Affymax led the development of new technologies merging photolithography with combinatorial solid-phase chemistry to create highly dense arrays of biological compounds. In 1993, Dr. Fodor was a founding member of Affymetrix as scientific director, and in 1997, he was appointed president and CEO. Affymetrix has extended the array technology to nucleic acids and is using and commercializing oliogonucleotide arrays to study a variety of nucleic acid molecular recognition processes. Dr. Fodor is the recipient of numerous awards, including most recently in 1998, the Jacob Heskel Gabbay Award in Biotechnology and Medicine and the Association for Laboratory Automation Achievement Award.

KOUROSH GHARACHORLOO is a research scientist in the Western Research Laboratory at Compaq Computer Corporation (formerly Digital Equipment Corporation). His research interests are parallel computer architecture and software, including hardware and software distributed shared memory systems, and the study of database and webserver applications. Before joining Digital, Dr. Gharachorloo was a key contributor in the Stanford Dash and Flash multiprocessor projects. In addition, his thesis research on memory consistency models has influenced academic research and impacted implementations of several commercial microprocessors. Recently, Dr. Gharachorloo has been involved in the development of Shasta (a software distributed shared memory system) and in the study of commercial applications such as Oracle and Alta Vista. In addition, he has contributed substantially to the design and specification of Alpha processors and server platforms (both 21264- and 21364-based systems) and has authored a chapter of the Alpha Architecture Reference Manual. Dr. Gharachorloo received a B.S. in electrical engineering, a B.A. in economics, an M.S. in electrical engineering, and a Ph.D. in electrical engineering and computer science, all from Stanford University. He has authored over 35 technical conference and journal papers and has filed over 10 patents. He also taught several short courses at Digital, lectured numerous times at Stanford, and served on various invited panels and program committees.

ERIC W. KALER is Elizabeth Inez Kelley Professor of Chemical Engineering and chair of the Chemical Engineering Department at the University of Delaware. He received his B.S. from the California Institute of Technology and his Ph.D. from the University of Minnesota. From 1982 until 1989, Dr. Kaler was on the faculty at the University of Washington as an assistant, then associate professor. He joined the faculty at the University of Delaware in 1989. He has received numerous awards, including the ACS Delaware Section Award, the ACS Award in Colloid or Surface Chemistry, the ASEE Curtis W. McGraw Award, and the Presidential Young Investigator Award. Dr. Kaler serves on editorial boards of several journals and is a member of several advisory boards. He holds six patents.

KENT KRESA is chairman of the board, president, and chief executive officer of the Northrop Grumman Corporation. Mr. Kresa joined Northrop Grumman in 1975 as vice president and manager of the company's Research and Technology Center, developing new proprietary processes and products for the company. From 1976–1982 he served as corporate vice president and general manager of the Ventura Division, a leader in the production of unmanned aeronautical vehicles. In 1982 he was appointed group vice president of the company's Aircraft Group and in 1986 was named senior vice president for technology development and planning. Before joining Northrop Grumman, Mr. Kresa served with the Defense Advanced Research Projects Agency, where he was responsible for broad applied research and development programs in the tactical and strategic defense arena. From 1961–1968 he was associated with the Lincoln Laboratory at the Massachusetts Institute of Technology (MIT), where he worked on ballistic missile defense research and reentry technology. Mr. Kresa is the recipient of many awards, including the California Industrialist of the Year, the Bob Hope Distinguished Citizen Award, the Navy League of New York's Admiral John J. Bergen Leadership Award, and 1994 Executive of the Year by the Los Angeles Business Journal. In 1998, he was elected an honorary fellow of the American Institute of Aeronautics and Astronautics. He is a member of the National Academy of Engineering, the past chairman of the Board of Governors of the Aerospace Industries Association, and chairman of the Defense Policy Advisory Committee on Trade. He serves on the boards of directors of numerous corporations and foundations, and on the Board of Trustees of the California Institute of Technology. Mr. Kresa is a graduate of MIT and received a B.S. in 1959, an M.S. in 1961 and an E.A.A. in 1966, all in aeronautics and astronautics.

THOMAS J. OVERBYE is an associate professor of electrical and computer engineering at the University of Illinois at Urbana-Champaign. He received his B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison. He was employed with Madison Gas and Electric Company from 1983 to 1991, where he helped develop their energy management system. In 1993 Dr. Overbye was the recipient of the IEEE Power Engineering Society Walter Fee Outstanding Young Engineer Award. Additionally, he is the author of a number of papers in the area of power system analysis, simulation, and restructuring, and is the principal developer of the PowerWorld Simulator software package.

PER F. PETERSON is a professor in the Department of Nuclear Engineering at the University of California, Berkeley, where he teaches and conducts research in heat and mass transfer, multi-phase/multi-component flows, thermal hydraulics, and nuclear materials management. He is also chair of the Energy Resources Group, an interdisciplinary academic unit of U.C., Berkeley with programs that treat issues of energy, resources, development and international security as the intersection of technological, economic, environmental and sociopolitical components. Dr. Peterson received his bachelor's degree in mechanical engineering from the University of Nevada, Reno, and his master's degree and Ph.D. from the University of California, Berkeley. Prior to joining the faculty at Berkeley as an assistant professor in 1990, he was a Japan Society for the Promotion of Science Fellow at the Tokyo Institute of Technology. He was also an engineer with Bechtel National, Inc. for three years. Dr. Peterson was a recipient of the NSF Presidential Young Investigator Award and two NURETH Best Paper Awards.

MICHAEL K. REITER is department head of the Secure Systems Research Department at Bell Laboratories, Lucent Technologies. He received a B.S. degree in mathematical sciences from the University of North Carolina and M.S. and Ph.D. degrees in computer science from Cornell University. He joined AT&T Bell Labs in 1993 and became a founding member of AT&T Labs Research when NCR and Lucent Technologies (including Bell Labs) were split from AT&T in 1996. He returned to Bell Labs in 1998 to take his current position. During 1998–2000, Dr. Reiter will serve as program chair of the flagship computer security conferences of both the Association for Computing Machinery and the Institute of Electrical and Electronics Engineers. He serves on numerous other conference program committees in the areas of computer security and distributed computing. Dr. Reiter is a member of the INFOSEC Science and Technology Study Group, which was chartered by the INFOSEC Research Council to advise government agencies on funding priorities for computer security research. His research interests include all areas of computer and communications security, electronic commerce, and distributed computing. With respect to the topic of his talk, his work on system survivability includes the Rampart system, which served as the foundation for AT&T's Omega Cryptographic key management service, and the Phalanx system, which was a central component of an electronic voting system targeted for trial in the 1998 Costa Rican presidential elections.

MING C. WU is a professor in the Department of Electrical Engineering and the director of the MURI Center on RF Photonic Material Devices, sponsored by the Office of Naval Research, at the University of California, Los Angeles. Previously, he was a member of the technical staff at AT&T Bell Laboratories (now Lucent Technologies), where he conducted research in high-speed semiconductor lasers and optoelectronics. His current research interests include optical MEMS (microelectromechanical systems) or MOEMS, high-speed optoelectronics, and microwave photonics. Dr. Wu has served as chair or on the program committee for numerous conferences in his field. He received a Packard Foundation Fellowship in 1992. He has published over 100 journal papers, 150 conference papers, contributed one book chapter, and holds 8 U.S. patents. Dr. Wu received his M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley.

Program

NATIONAL ACADEMY OF ENGINEERING

Fifth Annual Symposium
on Frontiers of Engineering
October 14–16, 1999

DROWNING IN DATA
Organizers: Kenneth Goldberg, Erik Hagersten, and Jian-Gang Zhu

Magnetic Recording: Winner of the Data Storage Technology Race
Thomas R. Albrecht, IBM Almaden Research Center

Evolution of Large Multiprocessor Servers
Kourosh Gharachorloo, Compaq Computer Corporation

Network Survivability and Information Warfare
Michael K. Reiter, Bell Laboratories, Lucent Technologies

Moving up the Information Food Chain: The Future of Web Search
Oren Etzioni, University of Washington

* * *

MAKING SENSE OF THE HUMAN GENOME
Organizers: Frances Arnold and Julio M. Ottino

Genes, Chips, and the Human Genome
Stephen P. A. Fodor, Affymetrix, Inc.

ENGINEERING NOVEL STRUCTURES

Colloidal Scale Engineering
Eric W. Kaler, University of Delaware

Design of Biomimetic Polymeric Materials
Arup K. Chakraborty, University of California, Berkeley

* * *

ENERGY FOR THE FUTURE AND ITS ENVIRONMENTAL IMPACT

Organizers: Michael Corradini and David McLaughlin

Deregulating the Electric Grid: The Engineering Challenge
Thomas J. Overbye, University of Illinois

The Future of Nuclear Power
Per F. Peterson, University of California, Berkeley

Renewable Energy Technologies: Today and in the Future
James M. Chavez, Sandia National Laboratories
Jane H. Davidson, University of Minnesota

* * *

OPTICS

Organizers: Connie Gutowski, Clifford Renschler, and Eva Sevick-Muraca

Surface-Emitting Lasers: A Key Component for Global Communication
Thomas M. Brennan, EMCORE PhotoVoltaics

Optical Applications of Microelectromechanical Systems
Ming C. Wu, University of California, Los Angeles

* * *

DINNER SPEECH

Career Flexibility in Rapidly Changing Times
Kent Kresa, Northrop Grumman Corporation

Participants

NATIONAL ACADEMY OF ENGINEERING

FIFTH ANNUAL SYMPOSIUM ON FRONTIERS OF ENGINEERING
October 14–16, 1999

Thomas R. Albrecht
Research Staff Member and Manager
IBM Almaden Research Center

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

Daniel F. Baldwin
Assistant Professor
George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology

Annelise E. Barron
Assistant Professor
Department of Chemical Engineering
Northwestern University

John F. Brady
Professor of and Executive Officer for Chemical Engineering
California Institute of Technology

Thomas M. Brennan
Vice President and General Manager
EMCORE PhotoVoltaics

Carla E. Brodley
Assistant Professor
School of Electrical and Computer Engineering
Purdue University

Kevin L. Buescher
Technical Staff Member
Applied Theoretical and Computational Physics Division
Los Alamos National Laboratory

Arup K. Chakraborty
Professor
Chemical Engineering and Chemistry
University of California, Berkeley

Kavitha Chandra
Assistant Professor
Department of Electrical and Computer Engineering
University of Massachusetts

James Chavez
Technical Manager
Government Relations Department
Sandia National Laboratories

Catherine A. Chess
Manager, Research Staff Member
Research Division Technical Strategy & Worldwide Operations
IBM Research

Todd K. Citron
General Manager
System Products
Hughes Space & Communications

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

Francesco Costanzo
Assistant Professor
Engineering Science and Mechanics Department
Pennsylvania State University

George W. Coulston
Technology Manager
DuPont Company
DuPont Dacron® Intermediates

Peter T. Cummings
Distinguished Professor and UT/ORNL Distinguished Scientist
Department of Chemical Engineering
University of Tennessee, Knoxville

Robert P. Currier
Team Leader
Chemical Science and Technology Division
Los Alamos National Laboratory

Jane H. Davidson
Professor
Department of Mechanical Engineering
University of Minnesota

Brian Dealy
Group Chief Technologist
Science Applications International Corporation

Jorge R. Duyos
President
JRD & Associates, Inc.

John M. Dzenitis
Senior Research Engineer
Monsanto Company

Chang-Beom Eom
Associate Professor
Dept. of Mechanical Engineering
and Materials Science
Duke University

Oren Etzioni
Associate Professor
Department of Computer Science
and Engineering
University of Washington

Stephen P. A. Fodor
Chairman and CEO
Affymetrix, Inc.

Wm. Garth Frazier
Materials Research Engineer
Air Force Research Laboratory

H. Christopher Frey
Associate Professor
Dept. of Civil Engineering
North Carolina State University

Sarah Frisken-Gibson
Senior Research Scientist
MERL - A Mitsubishi Electric
Research Laboratory

Hannes P. Fuchs
Staff Research Engineer
General Motors R&D Center

Albin J. Gasiewski
Group Leader
National Oceanic and Atmospheric
Administration
Environmental Technology
Laboratory

Kourosh Gharachorloo
Research Scientist
Western Research Laboratory
Compaq Computer Corporation

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

Lori L. Graham
Assistant Professor
Department of Civil Engineering
University of Virginia

Dennis G. Grubb
Assistant Professor
School of Civil and Environmental
Engineering
Georgia Institute of Technology

Connie L. Gutowski
Chief Program Engineer for F-150,
Expedition, and Navigator
Product Development Center
Ford Motor Company

Rafael E. Guzmán
Reservoir Engineer
BP Amoco Colombia

Erik Hagersten
Chair, Computer Architecture
Department of Computer Systems
Uppsala University

G. Chris Han-Adebekun
Senior Research Engineer
Eastman Kodak Company

Patrick Hanrahan
CANON USA Professor
Computer Science and Electrical Engineering Departments
Stanford University

Sarah W. Harcum
Assistant Professor
Department of Chemical Engineering
New Mexico State University

Michael P. Harold
Research Manager
Central Research and Development
E.I. du Pont de Nemours & Co.

Kathleen C. Hinge
Member Research and Technology Staff
Xerox Corporation

Alex Q. Huang
Assistant Professor
Department of Electrical Engineering
Virginia Polytechnic Institute and State University

Charles M. Hura
Manager
Mechanical Engineering Excellence Center
Eastman Kodak Company

Rebecca H. Jordan
Member of the Technical Staff
Materials and Chemistry Research Dept.
Bell Laboratories, Lucent Technologies

Scott W. Jorgensen
Staff Research Engineer
Chemical and Environmental Sciences Lab
General Motors R&D Center

Jinder Jow
Development Scientist
Union Carbide Corporation

Eric W. Kaler
Professor and Chair
Department of Chemical Engineering
University of Delaware

Mary E. F. Kasarda
Assistant Professor
Department of Mechanical Engineering
Virginia Polytechnic Institute and State University

Steven N. Kempka
Technical Manager
Thermal and Fluid Sciences Department
Sandia National Laboratories

Agnes Kim
Technical Specialist
Advanced Engineering Center
Ford Motor Company

Kevin T. Kornegay
Assistant Professor
School of Electrical Engineering
Cornell University

Ekaterini Korovessi
Senior Research Engineer
Central Research & Development
E.I. du Pont de Nemours & Co.
Experimental Station, E1/104

Mikko H. Lipasti
Assistant Professor
Department of Electrical and Computer Engineering
University of Wisconsin

Patrick J. Loughlin
Associate Professor and Fulton C. Noss Faculty Fellow
Depts. of Electrical Engineering and Bioengineering
University of Pittsburgh

Narayan B. Mandayam
Assistant Professor, WINLAB
Dept. of Electrical & Computer Engineering
Rutgers University

Steven P. Marsh
Research Metallurgist
Physical Metallurgy Branch
Naval Research Laboratory

Charles J. Mart
Section Head, Gas Conversion Development
Gas Conversion Laboratory
Exxon Research and Development Laboratories

David F. McLaughlin
Fellow Engineer
Westinghouse Science and Technology Center

Paul J. McWhorter
Deputy Director
Microsystems Science, Technology and Component Center
Sandia National Laboratories

Deirdre R. Meldrum
Associate Professor
Dept. of Electrical Engineering
University of Washington

William C. Messner
Associate Professor
Department of Mechanical Engineering
Carnegie Mellon University

Matthew P. Miller
Assistant Professor
Sibley School of Mechanical and Aerospace Engineering
Cornell University

Barbara S. Minsker
Assistant Professor
Department of Civil and Environmental Engineering
University of Illinois at Urbana-Champaign

Chu Whan Moon
Project Manager
Otis Elevator Company

Purnendu K. Mozumder
Vice President
PDF Solutions, Inc.

Rakesh Nagi
Associate Professor
Department of Industrial Engineering
State University of New York at Buffalo

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

Thomas Overbye
Associate Professor
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

Girish Pathak
Vice President and Director
Operations Systems Laboratory
GTE Laboratories Inc.

Brian K. Paul
Assistant Professor
Industrial and Manufacturing Engineering
Oregon State University

Per F. Peterson
Professor
Department of Nuclear Engineering
University of California, Berkeley

John P. Pollinger
Engineering Manager
AlliedSignal Ceramic Components

Stephen R. Quake
Associate Professor
Dept. of Applied Physics
California Institute of Technology

Sudhendu Rai
Member Research and Technology Staff II
Xerox Wilson Center for Research and Technology

Rajeev J. Ram
Assistant Professor
Electrical Engineering and Computer Science Department
Massachusetts Institute of Technology

Richard M. Ratliff
Vice President
Sabre

Michael Reiter
Department Head
Secure Systems Research Department
Bell Laboratories, Lucent Technologies

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

Stephen R. Sammataro
Senior Design Engineer
Sikorsky Aircraft Corporation

Sunita Satyapal
Manager, Energy Systems
United Technologies Research Center

Christine E. Schmidt
Assistant Professor
Department of Chemical Engineering
University of Texas

William F. Schneider
Senior Technical Specialist
Ford Research Laboratory
Ford Motor Company

Erik R. Scott
Research Scientist
ALZA Corporation

Eva M. Sevick-Muraca
Professor
Chemical Engineering Department
Texas A&M University

Patrick M. Shanahan
767-400 ER Program Manager
Commercial Airplane Group
Boeing Company

Mrityunjay Singh
Chief Scientist
Dynacs Engineering Co.
NASA Glenn Research Center

Ganesh Skandan
Vice President R&D
Nanopowder Enterprises, Inc. - an SMT Company

Winston O. Soboyejo
Professor
Department of Mechanical & Aerospace Engineering and The Princeton Materials Institute
Princeton University

Olav Solgaard
Assistant Professor
Department of Electrical Engineering
Stanford University

Anna Stefanopoulou
Assistant Professor
Department of Mechanical and Environmental Engineering
University of California, Santa Barbara

Margaret M. Steinbugler
Research Scientist, Energy Systems, and Fuel Cell Program Technical Leader
United Technologies Research Center

Hung-Jue Sue
Associate Professor
Department of Mechanical Engineering
Texas A&M University

Yuri Suzuki
Assistant Professor
Dept. of Materials Science and Engineering
Cornell University

Marvin M. Theimer
Senior Researcher
Microsoft Research

Matthew Turk
Researcher
Microsoft Research

Paul R. Van Tassel
Assistant Professor
Department of Chemical Engineering and Materials Science
Wayne State University

Annette von Jouanne
Assistant Professor
Electrical and Computer Engineering Department
Oregon State University

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

Sean Wakayama
Senior Engineer/Scientist
Boeing Company

Scott H. Wasserman
Project Scientist
Union Carbide Corporation

James J. Watkins
Assistant Professor
Chemical Engineering Department
University of Massachusetts

Marc B. Wisnudel
Chemical Engineer
Polymer Materials Lab
GE Corporate Research and Development

Ann Wollrath
Senior Staff Engineer
Sun Microsystems, Inc.

Ming C. Wu
Professor
Electrical Engineering Department
University of California, Los Angeles

Bennet S. Yee
Assistant Professor
Dept. of Computer Science and Engineering
University of California, San Diego

John Yin
Associate Professor and Cargill Faculty Fellow
Dept. of Chemical Engineering
University of Wisconsin, Madison

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

Guest Speaker

Kent Kresa
Chairman, President and CEO
Northrop Grumman Corporation

U.S.-Japan Frontiers of Engineering

Jonathan Bray
Professor
Civil and Environmental Engineering
University of California, Berkeley

Yasutaka Iguchi (*co-chair*)
Professor
New Industry Creation Hatchery Center
Tohoku University

Masahiro Kumeta
Department of International Affairs
Japan Science and Technology Corporation (JST)

Hideaki Matsubara
Chief Management Researcher
Research and Development Laboratory
Japan Fine Ceramics Center

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

James H. Smith
Director
Growth and Technology
Allied Signal I&SS - Redmond

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

Kazuo Tani
Associate Professor
Department of Civil Engineering
Yokohama National University

David A. Tirrell
Professor
Department of Chemistry
California Institute of Technology

Robert H. Wagoner (*co-chair*)
Professor
Dept. of Materials Science and Engineering
Ohio State University

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

Sponsor Representatives

Brian Dealy
Group Chief Technologist
Science Applications International Corporation

Alan S. Rudolph
Program Manager
Defense Sciences Office
Defense Advanced Research Projects Agency

George A. Schlutermann
Senior Hydrogeologist
Parsons Brinckerhoff

Guests

Malcolm J. Abzug
Retired President
ACA Systems, Inc.

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

Douglas Bauer
Executive Director
Commission on Engineering and Technical Systems
National Research Council

Matthew Cottle
Development Office
The National Academies

Ed Furtek
Associate Vice Chancellor
Science and Technology Policy and Projects
University of California, San Diego

Gerald Nadler
IBM Chair Emeritus in Engineering Management and Director, Breakthrough Thinking Group
University of Southern California

Benedict R. Schwegler, Jr.
Executive Director
Walt Disney Imagineering R&D, Inc.

Robert H. Wertheim
Consultant
Science Applications International Corporation

Abe M. Zarem
Consultant

National Academy of Engineering

Penelope Gibbs
Administrative Assistant

Janet Hunziker
Program Officer

Mary Kutruff
Administrative Assistant

Barbara Lee Neff
Executive Assistant to the President

William A. Wulf
President