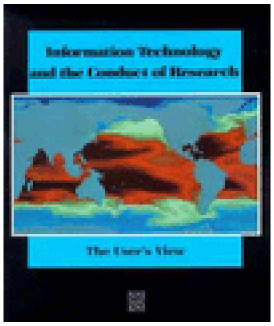
http://www.nap.edu/catalog/763.html

We ship printed books within 1 business day; personal PDFs are available immediately.



Information Technology and the Conduct of Research: The User's View

Panel on Information Technology and the Conduct of Research, National Academy of Sciences, National Academy of Engineering, Institute of Medicine

ISBN: 0-309-58199-0, 88 pages, 8.5 x 10, (1989)

This PDF is available from the National Academies Press at: http://www.nap.edu/catalog/763.html

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

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

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

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

Copyright © National Academy of Sciences. All rights reserved.
Unless otherwise indicated, all materials in this PDF File are copyrighted by the National Academy of Sciences. Distribution, posting, or copying is strictly prohibited without written permission of the National Academies Press. Request reprint permission for this book.

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

Information Technology and the Conduct of Research

The User's View

REPORT OF THE PANEL ON INFORMATION TECHNOLOGY AND THE CONDUCT OF RESEARCH
NATIONAL ACADEMY OF SCIENCES
NATIONAL ACADEMY OF ENGINEERING
INSTITUTE OF MEDICINE
COMMITTEE ON SCIENCE, ENGINEERING, AND PUBLIC POLICY

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. 1989

NATIONAL ACADEMY PRESS 2101 Constitution Avenue, NW Washington, DC 20418

The National Academy of Sciences (NAS) is a private, self-perpetuating society of distinguished scholars in scientific and engineering research, dedicated to the furtherance of science and technology and their use for the general welfare. Under the authority of its congressional charter of 1863, the Academy has a working mandate that calls upon it to advise the federal government on scientific and technical matters. The Academy carries out this mandate primarily through the National Research Council, which it jointly administers with the National Academy of Engineering and the Institute of Medicine. Dr. Frank Press is President of the NAS.

The National Academy of Engineering (NAE) was established in 1964, under the charter of the NAS, as a parallel organization of distinguished engineers, autonomous in its administration and in the selection of members, sharing with the NAS its responsibilities for advising the federal government. Dr. Robert M. White is President of the NAE.

The Institute of Medicine (IOM) was chartered in 1970 by the NAS to enlist distinguished members of appropriate professions in the examination of policy matters pertaining to the health of the public. In this, the Institute acts under both the Academy's 1863 congressional charter responsibility to be an adviser to the federal government and its own initiative in identifying issues of medical care, research, and education. Dr. Samuel O. Thier is President of the IOM.

The Committee on Science, Engineering, and Public Policy is a joint committee of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. It includes members of the councils of all three bodies.

This study received support from the Department of Energy, the National Aeronautics and Space Administration, the National Bureau of Standards and National Oceanic and Atmospheric Administration of the Department of Commerce, the National Library of Medicine, and the National Science Foundation. Opinions, findings, conclusions, or recommendations expressed in this publication are those of the Panel on Information Technology and the Conduct of Research of the Committee on Science, Engineering, and Public Policy, and do not necessarily reflect the views of the sponsors.

Library of Congress Cataloging-in-Publication Data
Panel on Information Technology and the Conduct of Research (U.S.)
Information technology and the conduct of research.
Bibliography: p.
Includes index.
1. Information technology—Scientific applications. 2. Research—Technological innovations. I. Title.
Q180.55.I45P36 1989 001.4'2 88-28903

ISBN 0-309-03888-X

Printed in the United States of America

Cover Photograph:
Temperature field in January at a depth of 225 meters, from a global model of the oceanic general circulation. Deep reds represent temperatures of 24°C and deep blues are -2°C. Picture courtesy of Michael Cox. Reproduced from "Computer modeling in physical oceanography from the global circulation to turbulence," William R. Holland and James C. McWilliams. Physics Today, vol. 40, no. 10, p. 52.

Panel on Information Technology and the Conduct of Research

DONALD N. LANGENBERG (*Chair*), Chancellor, University of Illinois at Chicago
W. RICHARDS ADRION, Chair, Computer and Information Science Department, University of Massachusetts, Amherst
JOSEPH BALLAM, Professor, Department of Physics, Stanford Linear Accelerator Center, Stanford, California
BRUCE G. BUCHANAN, Professor and Co-Director, Center for Parallel, Distributed, and Intelligent Systems, University of Pittsburgh, Pittsburgh, Pennsylvania
WILLIAM J. EMERY, Professor, Department of Aerospace Engineering Science, University of Colorado, Boulder
DAVID A. HODGES, Professor, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley
DAVID A. HOFFMAN, Professor, Department of Mathematics and Statistics, University of Massachusetts, Amherst
F. THOMAS JUSTER, Professor of Economics, University of Michigan, Ann Arbor
SARA B. KIESLER, Professor, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania
KENNETH M. KING, President, EDUCOM, Princeton, New Jersey
ROBERT LANGRIDGE, Professor, Schools of Pharmacy and Medicine, University of California, San Francisco
NINA W. MATHESON, Director, William H. Welch Medical Library, The Johns Hopkins University, Baltimore, Maryland
DAVID A. PENSAK, Corporate Advisor, Computing Technology, E. I. du Pont de Nemours and Co., Wilmington, Delaware
ALLAN H. WEIS, Vice President, Data Systems Division, IBM Enterprise Systems, White Plains, New York

Staff

JOHN R. B. CLEMENT, *Study Director*
AUDREY PENDERGAST, *Staff Officer*
ANN K. FINKBEINER, *Writer*
NISHA GOVINDANI, *Secretary*
JOAN ROOD, *Secretary*

Committee on Science, Engineering, and Public Policy

CORNELIUS J. PINGS (Chairman), Provost, University of Southern California, Los Angeles, California
GILBERT S. OMENN, Dean, School of Public Health and Community Medicine, University of Washington, Seattle, Washington (*former Chairman**)

H. NORMAN ABRAMSON, Executive Vice-President, Southwest Research Institute, San Antonio, Texas
ALBERT M. CLOGSTON, Member, Center for Materials Science, Los Alamos National Laboratory, Los Alamos, New Mexico
PHILIP M. CONDIT, Executive Vice-President, Boeing Commercial Airplane Company, Seattle, Washington
EMILIO Q. DADDARIO, Washington, D.C.
GERALD P. DINNEEN,* Vice President, Science and Technology, Honeywell Incorporated, Minneapolis, Minnesota
ALFRED P. FISHMAN, William Maul Measey Professor of Medicine, and Director, Cardiovascular-Pulmonary Division, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania
RALPH E. GOMORY, Senior Vice-President for Science and Technology, IBM Corporation, Armonk, New York
ZVI GRILICHES,* Nathaniel Ropes Professor of Political Economy, Department of Economics, Harvard University, Cambridge, Massachusetts
ARTHUR KELMAN, Wisconsin Alumni Research Foundation Senior Research Professor of Plant Pathology and Bacteriology, Department of Plant Pathology, University of Wisconsin, Madison, Wisconsin
FRANCIS E. LOW, Institute Professor, Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts

---

* Term expired June 30, 1988.

JOHN D. ROBERTS,* Institute Professor of Chemistry, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California

KENNETH J. RYAN, Kate Macy Ladd Professor of Obstetrics and Gynecology, Harvard Medical School; and Chairman, Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, Massachusetts

HERBERT A. SIMON, Richard King Mellon University Professor, Department of Computer Science and Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania

Ex Officio

FRANK PRESS, President, National Academy of Sciences
ROBERT M. WHITE, President, National Academy of Engineering
SAMUEL O. THIER, President, Institute of Medicine

Staff

ALLAN R. HOFFMAN, *Executive Director*
MYRON F. UMAN, *Associate Executive Director*
BARBARA A. CANDLAND, *Administrative Coordinator*
CATHY D. WILLIAMS, *Administrative Secretary*

Preface

The ever-present urge to categorize our fellow humans leads in this computer age to the categories “computer literate” and “computer illiterate.” It might seem obvious that scientists and engineers, particularly those engaged in research, must all be computer literate. After all, such people work with numbers and data, and isn't that what computers are all about? Yet the most superficial survey of researchers will reveal a wide range of capabilities in the use of information technology (computers plus telecommunications) in research. It will also reveal endemic frustration and dissatisfaction.

Why? Is not the work of those researchers whose subject is information technology itself yielding a steady stream of new capabilities for their colleagues in other fields? Yes it is, and some of the new capabilities can truly be called revolutionary. Are not researchers in many fields continually finding new ways to apply information technology to do old things faster, better, and cheaper, and to do new things which just yesterday were beyond the realm of possibility? Yes, that is so. So much so, in fact, that there are many who believe that the pervasive use of information technology in the conduct of research is changing profoundly the very meaning of the word “research.” Are not our institutions, agencies, and companies, our policymakers, managers, and vendors finding ways to place the new instruments of information technology in the hands of more and more researchers? Yes, despite the usual fiscal constraints, they are.

Then what's the problem? Indeed, is there one? It was the suspicion that there is a problem (many, actually), that there are serious impediments to the wider and more effective use of information technology in research, that led the Committee on Science, Engineering, and Public Policy (COSEPUP) to form the Panel on Information Technology and the Conduct of Research and to charge it to explore the situation and to report its findings, conclusions, and recommendations. I agreed to chair the Panel because, as a scientist turned university administrator and federal official, and a computer illiterate, I was excited by the prospect of learning something of an issue I sense is of paramount importance to the future of the global research enterprise. This report is the result of the Panel's deliberations. Our subject is broad, its literature is scattered, and some of its facets are still more art than science. We cannot claim to have produced the definitive picture of the issue. We hope we have made a case for the importance of understanding and addressing it, and perhaps shed some light on a creature that reminds some of us of the elephant once investigated by an earlier panel.

If this report has value, it is due to the salient characteristic of the Panel reflected in the report's subtitle, “The User's View.” Most of the Panel members are researchers active in disciplines not encompassed by the term “information technology.” They are expert but skeptical users of information technology in their own research, in possession of exciting visions of what this technology might bring to their fields, and of experienced views of what it has brought, and at what cost. From my youth I remember an ad for an automobile, which urged the reader to “Ask the man who owns one!” There's wisdom in that slogan; in the absence of a considerable body of established knowledge, our Panel focused on asking the men and women “who own one.” The result is a report that should speak to researchers experienced in the application of information technology, as well as those who would like to gain more experience, and of course to those engaged in supporting research. We believe it is worth the reader's attention.

The conception and early stages of the study owe much to the Committee on Science, Engineering, and Public Policy. In particular, two former members of the Committee deserve mention: Floyd E. Bloom, whose term with the committee ended soon after this study's initiation; and Gilbert S. Omenn. Dr. Bloom conceived the topic and played a central role in its birth. Dr. Omenn chaired COSEPUP during the inception and most of the execution of the study. We also must thank Norman Metzger of the National Research Council, who developed the study's charge and initially served as study director.

Support for the study was provided by several federal agencies: the Department of Energy; the National Aeronautics and Space Administration; the National Bureau of Standards and National Oceanic and Atmospheric Administration of the Department of Commerce; the National Library of Medicine; and the National Science Foundation. Sun Microsystems, Inc., donated an advanced workstation for the purpose of report production.

The study benefited from the opinions and reviews of many people: research users, experts in computing, computing applications, and communications, and policymakers. We received helpful advice and suggestions from too many persons to mention by name; but we acknowledge the vital part their input played. At the final stages of the report's preparation, four members of COSEPUP served as a review group: John D. Roberts, as chair, and Alfred P. Fishman, Francis E. Low, and Herbert A. Simon. We did not always take the advice offered; but we always profited from it. Of course, the report's statements, findings, and recommendations remain the sole responsibility of the Panel members.

The Panel is particularly grateful to its professional staff: John Clement, Audrey Pendergast, Ann Finkbeiner, and Nisha Govindani, who supported the study both intellectually and logistically, while exhibiting exemplary patience. The Panel also owes special thanks to Allan Hoffman, executive director of COSEPUP, for input and support from conception to final dissemination; and also to Barbara Candland and Cathy D. Williams of the COSEPUP staff. Without the efforts of all these people, the study truly would not have taken place.

One final, and important, point: I share with many researchers a strong belief that much of the power of science (whether practiced by scientists, engineers, or clinical researchers) derives from the steadfast commitment to free and unfettered communication of information and knowledge. This principle has been part of the ethos of the global research community for centuries, and has served it and the rest of humanity well. If asked to distill one key insight from my service on this panel, I would respond with the assertion that information technology is of truly enormous importance to the research community, and hence to all humanity, precisely because it has the potential to enhance communication of information and knowledge within that community by orders of magnitude. We can now only dimly perceive what the consequences of that fact may be. That there is a revolution occurring in the creation and dissemination of information, knowledge, and, ultimately, understanding is clear to me. It is also clear to me that it is critically important to maintain our commitment to free and unfettered communication as we explore the uses of information technology in the conduct of research. If my colleagues and I succeed through this report in conveying some sense of this to our readers, and of the necessity that many individuals and institutions be attentive to it, we will have discharged our duty.

DONALD N. LANGENBERG
CHAIR

Contents

Information Technology and the Conduct of Research

The User's View

Executive Summary

Information technology—the set of computer and telecommunications technologies that makes possible computation, communication, and the storage and retrieval of information—has changed the conduct of scientific, engineering, and clinical research. This report examines present trends, future potential, and impediments to the use of information technology in support of research. Written from the viewpoint of the researcher using information technology and including many examples, the report offers a number of recommendations directed to two principal audiences: policymakers and leaders of institutions responsible for the support and management of research, and researchers themselves.

The first programmable, electronic, digital computer was created nearly five decades ago. At first, computers simply substituted for other means of carrying out arithmetic calculations; they were large, expensive, often unreliable, and accessible only to a minority of scientists and engineers. With the advent of the integrated circuit (the semiconductor “chip”), computational speed and power increased dramatically, and computer use became widespread. Recently, computer technology has been joined with telecommunications technology to create a new entity: information technology, which has done much to remove the constraints of speed, cost, and distance from the researcher.

On the whole, information technology has led to improvements in research. New avenues for scientific exploration have opened. Researchers can collaborate more widely and efficiently. Much more data are available for analysis. Analytic capabilities have improved significantly, along with the capability to present results as visual images. New information technologies offer further opportunities to improve research. But widespread use of computers in research has not come about without problems. Some of these difficulties are technological, some financial. Underlying many of them are complex institutional and behavioral constraints.

The report examines three aspects of the research process: data collection and analysis, communication and collaboration among researchers, and information storage and retrieval.

In data collection and analysis, a number of trends are discussed, including

- Growth in the amount of information researchers can store and analyze;
- Creation of new families of computer-controlled instruments;
- Proliferation of computer networks dedicated to research; and
- Increasing availability of software “packages” supporting research activities.

Among the difficulties associated with data collection and analysis are uneven access to computing resources, problems in obtaining support for software development and maintenance, and unnecessary complexities of transmitting data over computer networks.

Communication and collaboration among researchers are changing. Not only can information be shared more and more quickly, but researchers are also developing new collaborative arrangements. Three technologies are involved: word processing, electronic mail, and computer communications networks. Word processing and electronic mail are arguably the most pervasive of all the routine uses of computers in research communication. Electronic mail—sending text from one computer user to another over the networks—is partially replacing written and telephone communication among many communities of scientists. Scientists increasingly use networks for conversation and for repeated exchanges of text and data files. Among the most important of the potential applications of information technology is the emergence of a truly national research network.

The principal difficulties with communicating via electronic mail and file transfer technologies involve incompatibility between different text and data processing systems and between network protocols. Also significant are network limitations: addressing conventions are cumbersome and unhelpful, locator services are nearly nonexistent, and overall network availability and reliability need improvement.

Electronic storage and retrieval of information hold enormous advantages: information can be stored economically, found quickly without going to another location, and moved easily. For all disciplines, both scientific data and reference database promise to be significant sources of knowledge for basic research. However, a number of problems need to be resolved. Researchers have difficulty getting access to data stored by other researchers. Even when researchers get access to colleagues' data, they have difficulty reading them. Finally, when researchers get access to and read each others' databases, they often lack information on the quality of the data.

The primary difficulty encountered with reference databases is in conducting searches. Most information searches at present are incomplete, cumbersome, inefficient, expensive, and executable only by specialists.

There is a pressing need for new, more compact, and more permanent forms of data storage. Stored data gradually become useless, either because the storage media decay or the storage technology itself becomes obsolete. Underlying difficulties in information storage and retrieval are significant problems in the institutional management of resources.

New computer-based technologies offer the prospect of new ways of finding, understanding, storing, and communicating information, and should increase both the capabilities and the productivity of researchers. Among these new technologies are simulations, new methods of presenting observational and computational results as visual images, the use of knowledge-based systems as “intelligent assistants,” and more flexible and intuitive ways for people to interact with, and control, computers.

The Panel has identified a number of problem areas in which institutional and behavioral impediments underline many difficulties in the use of information technology in research. These areas include

- Issues of costs and cost sharing: financial impediments are chronic. Although institutions will continue to do their best, information technology for research will continue to need more funds. The Panel believes that increased support of information technology in research deserves high priority.
- The problem of standards: simplified, consistent standards for operation of, and interconnection among, computer systems could have major impacts on research communications and productivity; however, such standards are largely absent, and their development is a slow and controversial process.
- Legal and ethical constraints: the need to safeguard and maintain confidentiality of data on human subjects is a major issue; also likely to become increasingly significant is the question of responsibility in computer-supported decision making in engineering, clinical practice, or research.
- Gaps in training and education: learning to use information technology presently requires significant initial investments of time and effort, and researchers who make these investments often receive insufficient help. Although the problem is likely to diminish with time, it affects current attitudes of many researchers toward the use of information technology.
- The perceived risks of organizational change: organizations and administrators can understandably be reluctant to make the substantial changes required to make use of information technology.
- Of fundamental importance, the lack of an infrastructure for the use of information technology in research: access to expertise, and support mechanisms to encourage such experts; tools for developing and managing software; systems for storing and retrieving information; and support services for communication and collaboration among researchers.

The report concludes with three major recommendations.

RECOMMENDATION I

The institutions supporting the nation's researchers must recognize and meet their responsibilities to develop and support policies, services, and standards that help researchers use information technology more widely and productively. Specifically

- *Universities* should provide accessible, expert help in learning and using information technology.
- *Universities departments,* and *scientific and professional groups,* should establish career ladders for scientific programming positions.
- *Funding agencies* should provide support for scientific programming and for help services in learning and using information technology systems for research.
- *Scientific associations* should establish disciplinary standards for the storage and indexing of scientific data.
- *University departments,* and *scientific and professional groups,* should implement mechanisms for the evaluation, merit (peer) review, and dissemination of software useful in the conduct of research.
- *Vendors,* in collaboration with *scientific groups,* should establish standards for simplified and consistent user-machine interfaces.
- *Network administrators* should provide simple user interfaces and addressing schemes, add gateways to other networks, improve system reliability and capacity, and provide online help, such as guides to services and mail addresses of individuals who can answer questions.
- *Information service providers* should create simplified common standards for accessing and querying information sources and eventually provide unified access to information.
- *Software vendors,* and *scientific and professional groups,* should create program libraries and make them accessible through the networks.

RECOMMENDATION II

The institutions supporting the nation's researchers, led by the federal government, should develop an interconnected national information technology network for use by all qualified researchers. Specifically

- The Office of Science and Technology Policy (OSTP) in the Executive Office of the President, and the federal agencies responsible for supporting and performing research and development, should plan and fund a nationwide infrastructure for computer-based research communication.
- Planning and development of this nationwide infrastructure should be guided by users of information technology in research, rather than by technical experts in information technology or hardware or software vendors. The Panel believes strongly that such a national network is too important to the future of research to be left only to the technical experts.
- The national research network should be founded on the fundamental premise of open access to all qualified researchers/scholars that has nurtured the world's scientific community for centuries.

- The national research network should be developed in an evolutionary manner, making full use of the existing successful networks for research.

RECOMMENDATION III

To facilitate implementation of Recommendations I and II, and to focus attention on the opportunities and impediments associated with research uses of information technology, the Panel recommends the establishment at a national level of a user's group to oversee and advise on the evolution and use of information technology in support of scientific, engineering, and clinical research.

Specifically, the National Research Council (NRC) should charge a standing committee or board (whether existing or newly created) with the mandate to oversee and advise on research use of information technology. The membership of this board should include a majority of users from a variety of research disciplines.

Introduction

This report is about how information technology has changed the conduct of scientific, engineering, and clinical research.

Information technology is that set of computer and telecommunications technologies that makes possible computation, communication, and the storage and retrieval of information. The term, therefore, includes

- Computer hardware of all kinds, from microprocessors dedicated to specific research tasks to the largest supercomputers;
- Communications networks that link researchers to each other and to resources of various kinds; and
- Computer software that researchers use to design and run scientific projects, and to manage the information that the projects yield.

The effect of information technology on the conduct of research has long been a concern of the Committee on Science, Engineering, and Public Policy (COSEPUP), a joint unit of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. A previous COSEPUP report, *Frontiers in Science and Technology: A Selected Outlook* (W. H. Freeman, 1983), discussed ways of improving scientific and technical communication and asked, “How can scientists and engineers be encouraged to use the new electronic modalities innovatively and effectively?”

The science policy community has occasionally discussed how computer and communication technologies affect research, although mostly as a corollary to other policy issues such as the need for support for advanced computing, the computing requirements of individual research disciplines, national security concerns about information dissemination, or the potential of developments in information technology research. For example, in September 1985, the House of Representatives' Committee on Science and Technology held hearings on “The

Impact of the Information Age on Science" as part of its study of U.S. science policy (U.S. Congress, 1986a,b).

COSEPUP members felt that the subject deserved a thorough examination. In July 1985, a COSEPUP planning group recommended a study, and in December, 1986, after approval and selection of a panel, the study formally began. The study has received financial support from the Department of Energy, the National Aeronautics and Space Administration, the National Bureau of Standards and National Oceanic and Atmospheric Administration of the Department of Commerce, the National Library of Medicine, and the National Science Foundation.

COSEPUP's charge to the Panel on Information Technology and the Conduct of Research was to

- Examine current and prospective applications of information technology—computers and communications—to improve productivity in selected fields of scientific and engineering research, including the biological, physical, social, and engineering sciences;
- Identify impediments to the effective use of these technologies in research—such impediments may be institutional, financial, behavioral, and technical;
- Examine the behavioral and cultural changes required to exploit the new opportunities offered by these information technologies; and
- Suggest appropriate actions by federal agencies, as well as by universities and other research institutions, manufacturers, vendors, scientific associations, and individual researchers.

In considering its charge, the Panel decided that the goal of the study would be to recommend how to stimulate research through the use of information technology, and the focus of the study would be on both current and prospective uses. Because little research on general scientific uses of information technology now exists, the Panel gathered additional information from several disciplines representing the range of scientific and engineering research. Experts were provided with a list of questions, and commissioned to prepare papers on the use of information technology in their fields. They were asked to represent their colleagues' views as well as their own. The Panel believes that these papers reflect the diversity in information technology uses within and across the scientific disciplines. The papers provided essential inputs to the Panel's discussions and eventual recommendations. They are listed in Appendix A, and are available from COSEPUP on request.

The report is written from the point of view, not of those who specialize in information technology, but of those who use it. For all researchers, information technology is beneficial; for some, it has become central. Some researchers want only better access to current technologies; others urgently want much more. The Panel does not presume to prescribe a single model for all researchers, but it does believe that the products of the information age are invaluable and should be available to researchers who need them. Therefore, in what follows, the Panel does not appraise developments in computing hardware or software but emphasizes instead how these developments affect the productivity of researchers.

The report is directed to two principal audiences. One audience includes the policymakers and leaders of those institutions responsible for the support and management of research. For this audience, the Panel describes issues and impediments, and recommends ways of helping the research community enhance its use of information technology. The second audience is the users themselves, that is, research scientists and engineers. The Panel hopes that researchers will find their concerns about information technology addressed clearly and in their own terms. In addition, the Panel hopes that practices in other disciplines may spark readers' ideas for their own research.

The Panel discovered early that there is almost no systematic information on the users and uses of information technology. For example, the Panel cannot estimate how many or what proportion of scientists use computers in different fields, how access to networks and computer facilities is distributed across disciplines, or to what extent useful applications are disseminated throughout the research community. Systematic collection of such information is essential to the development of intelligent policy. Researchers' experiences in using information technology can help guide decisions about policy and resource allocation. In turn, these decisions will shape the technological and institutional advances that break down impediments to the further use of information technology. This process will continue to change the nature of scientific, engineering, and clinical research itself.

Finally, the Panel has come to a view of new ways of managing scientific knowledge and conducting scientific research. In this view, scientists are more productive because they are using the power of computers both to augment their intellectual efforts and to improve communication. With artificial barriers to communication lowered, science itself is closer to the open, collaborative search that is its goal.

The next section of the report describes present trends, future potential, and impediments to the use of information technology in support of research, drawing on examples from a number of fields. The report's final section summarizes the Panel's findings and recommendations.

The Use of Information Technology in Research

In this chapter we examine the effect of information technology on the conduct of research. New technologies offer new opportunities, although pervasive use of computers in research has not come about without problems. Some of these problems are technological, some financial. Underlying many of them are complex institutional and behavioral constraints.

Nearly five decades ago, the first programmable, electronic, digital computer was switched on. That day science acquired a tool that at first simply facilitated research, then began to change the way research was done. Today these changes continue, and now amount to a revolution.

Electronic digital computers at first simply replaced earlier technologies. Researchers used computers to do arithmetic calculations previously done with paper and pencil, slide rules, abacuses, or roomfuls of people running mechanical calculators. Benefits offered by the earliest computers were more quantitative than qualitative; bigger computations could be done faster, with greater reliability, and perhaps more cheaply. But computers were large, expensive, required technically expert operators and programmers, and consequently were accessible only to a relatively small fraction of scientists and engineers.

One human generation and several computer generations later, with the advent of the integrated circuit (the semiconductor “chip”), computational speed increased by a factor of 1 trillion, computational cost decreased by a factor of 10 million, and the smallest useful calculator went from the size of a typewriter to the size of a wristwatch. At present, personal computers selling for a few thousand dollars can put significant computing power on the desk of every scientist. Meanwhile, advances in the software through which people interact with and instruct computers have made computers potentially accessible to people with no specific training in computation. More recently, computer technology has joined telecommunications technology to create a new entity, “information technology.” Information technology has done much to remove from the researcher the constraints of speed, cost, and distance.

On the whole, information technology has led to improvements in research. New avenues for scientific exploration have opened. The amount of data that can be analyzed has expanded, as has the complexity of analyses. And researchers can collaborate more widely and efficiently.

Different scientific disciplines use information technology differently. Uses vary according to the phenomena the discipline studies and the rate at which the discipline obtains information. In such disciplines as high energy physics, neurobiology, chemistry, or materials science, experiments generate millions of observations per second, and these must be screened and recorded as they happen. For these disciplines, computers that can handle large amounts of information quickly are essential and have made possible research that was previously impractical. Other disciplines, such as economics, psychology, or public health, gather data on events that accumulate slowly over relatively long periods of time. These disciplines also need computers with large capacities, but do not need the capability to react in “real time.” Most disciplines use information technology in ways that fall somewhere in the range between these two extremes.

Boxes supplement or expand points in the text: the first two below deal with specific disciplines.

HIGH ENERGY PHYSICS: SCIENCE DRIVES THE LEADING EDGE OF INFORMATION TECHNOLOGY.

An example helps to illustrate the direction in which many disciplines are moving: high energy physics could not be done without information technology, and offers an extreme example of the trends for computing and communication needs in many scientific disciplines.

Most high energy physicists work on the same set of questions: what is the behavior of the most elementary particles, and what is the nature of the fundamental forces between them? Their experiments are conducted in machines called accelerators, devices that produce beams of protons, electrons, or other particles that are accelerated to high-speeds and huge energies. There are two types of accelerators: those in which two beams of particles are made to collide with each other (colliders), and those in which a beam hits stationary targets. Physicists then reconstruct the collision to find new phenomena.

Remarkable results have emerged from high energy physics experiments conducted over the past two decades. For instance, a Nobel prize-winning experiment carried out at the proton-antiproton collider at the European Center for Nuclear Research (CERN) in Switzerland, discovered two new particles known as the W and the Z. Their existence had been predicted by a theory claiming that the weak and electromagnetic forces, seemingly unrelated at low energy levels, were in fact manifestations of a single force, called the electroweak interaction, which would appear at sufficiently high energies. This discovery is a significant step toward the description of all known interactions—gravity, electromagnetism, and the strong (nuclear) and weak (radioactive decay) forces—as manifestations of a single unifying force.

The process by which some tens of these

The Panel recognizes the diversity in research methods, and differences in needs for information technology. But the needs of researchers show sufficient commonalities across research fields to make a search for common solutions worthwhile.

THE CONDUCT OF RESEARCH

The everyday work of a researcher involves such activities as writing proposals, developing theoretical models, designing experiments and collecting data, analyzing data, communicating with colleagues, studying research literature, reviewing colleagues' work, and writing articles. Information technology has had important effects on all these activities, and more change is in the offing. To illustrate these effects, we examine three particular aspects of research: data collection and analysis, communications and collaboration, and information storage and retrieval. In each area, we discuss how researchers currently use information technology and what difficulties they encounter. In a final part of this section, we discuss new technological opportunities and their implications for the conduct of research.

new W and Z particles were isolated from millions of collision events in the CERN accelerator offers a striking illustration of the dependence of high energy physics on the most advanced aspects of information technology. Three steps are involved. First, data are acquired in real time as the experiment progresses; second, the data obtained are transformed into flight paths, from which the particles making the paths are identified; and third, the event itself is reconstructed, and those few events exhibiting the very special characteristics of the new phenomenon are identified. In each of these steps computers are vital: to trigger the identification of interesting events; to establish particle tracks from the data; and to carry out analysis and interpretation.

In the future, high energy physicists will demand more from information technology than it can now deliver. Proposed new particle accelerators, such as the Superconducting Super Collider (SSC), are expected to produce several million collisions every second, of which only one or two collisions a second can be recorded. Selecting this tiny fraction of the produced events in a manner that does not throw away other interesting data is a tremendous challenge. It is hoped that “farms” of dedicated microprocessors might be able to examine tens of thousands of collision events per second, so that sophisticated selection mechanisms can screen all collisions and select the very few that are to be recorded. The computer programs that need to be developed for these tasks are of unprecedented size and complexity, and will challenge the capabilities of both the physicists programming them and the information technology software support available to the programmers.

Even the small fraction of recorded events will result in some ten million collisions to be analyzed in a year. Processing one year's worth of saved data from the SSC would take a modern mid-sized computer 500 years;

Data Collection and Analysis

Current Use Collecting and analyzing data with computers are among the most widespread uses of information technology in research. Computer hardware for these purposes comes in all sizes, ranging from personal computers to microprocessors dedicated to specific instrumentational tasks, large mainframe computers serving a university campus or research facility, and supercomputers. Computer software ranges from general-purpose programs that computer numeric functions or conduct statistical analyses to specialized applications of all sorts.

The Panel has identified five trends in the use of information technology in data collection and analysis:

obviously, a faster processing rate is required. Although no computer currently on the market would handle this load in reasonable time, existing plans suggest that, by the time it is needed, some combination of dedicated microprocessors and large mainframe systems will be available.

High energy physicists are also highly dependent on networks. Accelerators are located in only seven main laboratories in the United States, Switzerland, West Germany, the Soviet Union, and Japan; the physicists who use them are located in many hundreds of universities and institutions scattered around the world. Almost every high energy experiment, large or small, is a result of international collaboration: for instance, one detector installed around one of the collision points of the accelerator at the Fermi National Laboratory is run by a collaboration of four foreign and thirteen U.S. institutions, involving some 200 physicists. Physicists at several institutions designed different parts of the detector; since the detector has to work as an integrated apparatus, the physicists had to coordinate their work closely. Different physicists are also interested in different aspects of the experiment, and subsequent analysis of the data depends crucially on adequate networking.

Future networking needs for high energy physics involve very high transmission speeds (as high as 10 megabits per second) between laboratories, with provision for exchange of collision event files, graphics, and video conferencing. Present long distance communication links are limited to lower transmission speeds (typically, 56 kilobits per second); each university physics group could use a 1.5 megabit per second line for its own research needs. The provision of these facilities would be of enormous benefit to university-based physicists and students who cannot travel frequently to accelerator sites.

- Increased use of computers for research. This trend coincides with large and continued increases in the speed and power of computers and corresponding declines in their costs.
- Dramatic increases in the amount of information researchers can store and analyze. For example, researchers can now process and manipulate observations in a database consisting of 18 years × 3,400 individuals × 1,000 variables per individual for each year, create sets of relationships among these observations,

and then subject the data to complex statistical analyses, all at a cost of less than $100. Two decades ago, that kind of analysis could not have been conducted, and a much simpler analysis would have cost at least ten times as much.

- The creation of new families of instruments in which computer control and data processing are at the core of observation. For example, in new telescopes, image-matching programs on specialized computers align small mirrors to produce the equivalent light-gathering power of much larger telescopes with a single mirror. For instruments such as radio-telescope interferometers, the computer integrates data from instruments that are miles apart. For computer-assisted tomographic scanners, the computer integrates and converts masses of data into three-dimensional images of the body.
- Increased communication among researchers, resulting from the proliferation of computer networks dedicated to research, from a handful in the early 1970s to over 100 nationwide at present. Different networks connect different communities. Biologists, high energy physicists, magnetic fusion physicists, and computer scientists each have their own network; oceanographers, space scientists, and meteorologists are also linked together. Networks also connect researchers with one another regionally; an example is NYSERNET, the New York State Education and Research Network. Researchers with defense agency contracts are linked with one network, as are scientists working under contract to the National Aeronautics and Space Administration (NASA). Such networks allow data collection and analysis to be done remotely, and data to be shared among colleagues.
- Increasing availability of software “packages” for standard research activities. Robust, standardized software packages allow researchers to do statistical analyses of their data, compute complex mathematical functions, simplify mathematical expressions, maintain large databases, and design everything from circuits to factories. Many of these packages are commercial products, with high-quality documentation, service, and periodic updates. Others are freely shared software of use to a specialized community without the costs or benefits of commercial software.

One example illustrating several of the above trends is a system that geophysicists have set up to predict earthquakes more accurately. Networks of seismographs cover the western United States. One such network in northern California is called CALNET. Information from the 264 seismographs in CALNET goes to a special-purpose computer called the real-time picker. The software on the real-time picker looks at data as they come in and identifies exceptional events: patterns that indicate a coming earthquake. Then it notifies scientists of the events by telephone and sends graphics displays of locations and magnitudes, all within minutes.

Difficulties Encountered The difficulties that researchers encounter using information technology to collect and analyze data vary in importance depending on the particular discipline.

One difficulty is uneven access to computing resources. Information technology is not equally accessible to all researchers who could benefit from its use, even though broadening access is a continuing focus of institutions and funding agencies. To take an example from the field of statistics: according to a 1986 report on the Workshop on the Use of Computers in Statistical Research, sponsored by The Institute for Mathematical Statistics, "…the quality and quantity of computational resources available to researchers today varies dramatically from department to department … Perceived needs appear to vary just as dramatically … [While] departments that already have significant computer hardware feel a strong need for operating support, … departments that do not have their own computational resources feel an equally strong need for hardware." (Eddy, 1986, p. iii.)

Exclusion from resources happens for a variety of reasons, all reducible in the end to financial constraints. Not all academic or research institutions have links to networks; in addition, access to networks can be expensive, so not everyone who wants it can afford it. In some cases, since access to networks often mediates access to resources such as supercomputers, exclusion from networks can mean exclusion from advanced computing.

One of the most frustrating difficulties for researchers is finding the right software. Software that is commercially available is often unsuited to the specialized needs of the researcher. In those fields in which industry has an interest, however, commercial software is being developed in response to a perceived market. Software could be custom designed for the researcher, but relatively few researchers pay directly for software development, partly because research grants often cannot be used to support it. Consequently, most research

See box on software, page 18.

RESEARCH MATHEMATICS AND COMPUTATION

Computation and theory in mathematics are symbiotic processes. Machine computing power has matured to the point where mathematical problems too complicated to be understood analytically can be computed and observed. Phenomena have been observed for the first time that have initiated entirely new theoretical investigations. The theory of the chaotic behavior of dynamic systems depends fundamentally on numerical simulations; the concept of a "strange attractor" was formulated to understand the results of a series of numerical computations. Recent advances in the theory of knots have relied on algebraic computations carried out on computers. These advances can be directly applied to such important topics as understanding the folding of DNA molecules. In the field of geometry, numerical simulation has been used recently to discover new surfaces whose analytic form was too difficult to analyze directly. The simulations were understood by the use of computer graphics, and led to the explicit construction of infinite families of new examples.

The modern computer is the first laboratory instrument in the history of mathematics. Not only is it being used increasingly for research in pure mathematics, but, equally important, the prevalence of scientific computing in other fields has provided the me ers, although they are not often skilled software creators, develop their own software with the help of graduate students. The result meets researchers' minimum needs but typically lacks documentation and is designed for one purpose only. Such software is not fully understood by any one person, making it difficult to maintain or transport to other computing environments. This means that the software often cannot be used for related projects, and the scientific community wastes time, effort, and money duplicating one another's efforts. In sections to follow we examine how this problem is being addressed by professional associations, nonprofit groups, and corporations.

dium for communication between the mathematician and the physical scientist. Here modern graphics plays a critical role. This interaction is particularly strong in materials science, where the behavior of liquid crystals and the shapes of complex polymers are being understood through a combination of theoretical and computational advances.

In spite of all this, mathematics has been one of the last scientific disciplines to be computerized. More than other fields, it lacks instrumentation and training. This prevents the mathematician from using modern computing hardware and techniques in attacking research problems, and at the same time isolates him/her from productive communication with scientific colleagues.

Of course, mathematics is an important part of the foundation and intellectual basis of most of the methods that underlie all scientific use of computational machinery. To use today's high-speed computing machines, new techniques have been devised. The need for new techniques is providing a serious challenge to the applied mathematician, and has placed new and difficult problems on the desk of the theorist; algorithms themselves have become an object of serious investigation. Their refinement and improvement have become at least as important to the speed and utility of high-speed computing as the improvement of hardware.

Some disciplines are limited by available computer power because computers needed are not on the market. Some contemplated calculations in theoretical physics, quantum chemistry, or molecular dynamics, for example, could use computers with much greater capacity than any even on the drawing boards. In other cases, data gathering is limited by the hardware presently available. Most commercial computers are not designed to accommodate hardware and programs that select out interesting information from observational data, and scientists who want such computers must build them.

Another difficulty researchers encounter is in transmitting data over networks at high-speed. For researchers such as global geophysicists who use data collected by satellite, a large enough volume of information can be sent in a short enough time, but transmission is unreliable. Researchers often encounter delays and incur extra costs to compensate for “noise” on high-speed networks. Technological solutions such as optical fiber and error-correcting coding are currently expensive to install and implement and are often unavailable in certain geographic regions or for certain applications.

Communication and Collaboration among Researchers

Current Use Researchers cannot work without access to collaborators, to instruments, to information sources and, sometimes, to distant computers. Computers and communication networks are increasingly necessary for that access. Three technologies are concerned with communications and collaboration: word processing, electronic mail, and networks. Word processing and electronic mail are arguably the most pervasive of all the routine uses of computers in research communication. Electronic mail—sending text from one computer user to another over the networks—is replacing written and telephone communication among many communities of scientists, and is changing the ways in which these communities are defined. Large, collaborative projects, such as oceanographic voyages, use electronic mail to organize and schedule experiments, coordinate equipment arrivals, and handle other logistical

IF KITCHEN APPLIANCES WERE LIKE SOFTWARE

If kitchen appliances were like programs, they would all look alike sitting on the counter. They would all be gray, featureless boxes, into which one places the food to be processed. The door to the box, like the box itself, is completely opaque.

On the outside of each box is a general description of what the box does. For instance, one box might say: “Makes anything a meal”; another: “Cooks perfectly every time”; another: “Never more than 100 calories a serving.” You can never be exactly sure what happens to food when it is placed in these boxes. They don't work with the door open, and the 200-page user's manual doesn't give any details.

Working in a kitchen would be a matter of becoming familiar with the idiosyncracies of a small number of these boxes and then trying to get done what you really want done using them. For instance, if you want a fried-egg sandwich, you might try the “Makes anything a meal” box, since a sandwich is a sort of meal. But because you know from past experience that this box leaves everything coated with grease, you use the “Never more than 100 calories” box to postprocess the output. And so on. The result is never what you really want, but it is all you can do.

You aren't allowed to look inside the boxes to help you do what you really want to do. Each box is sealed in epoxy. No one can break the seal. If the box seems not to be working right, there is nothing you can do. Even calling the manufacturer is no help, because the box is not under warranty to be fit for any particular purpose. The manufacturer do have help lines, but not for help with broken boxes—rather to help you figure out how to use functioning boxes. But don't try to ask how your box works. The help-line people don't know, or if they do, they won't tell you. Several times a year you get a letter from the manufacturer telling you to ship them your old box and they will send you a new one. If you do so, you find yourself with a shinier box, which does whatever it did before a little faster, or perhaps it does a little more—but since you were never sure what it did before, you cannot be sure it's better now.

SOURCE: Mark Weiser, 1987. “Source Code,” *IEEE Computer,* 20(11): 66–73.

See box on document processing, page 19.

details. With the advent of electronic publishing tools that help lay out and integrate text, graphics, and pictures, mail systems that allow interchange of complex documents will become essential.

Networks range in size from small networks that connect users in a certain geographic area, to national and international networks. Scientists at different sites increasingly use networks for conversations by electronic mail and for repeated exchanges of text and data files.

The Panel has identified two major trends in the way information technology is changing collaboration and communication in scientific research:

- Information can be shared more and more quickly. For example, one of the first actions of the federal government after the discovery of the new high-temperature superconductors was to fund, through the Department of Energy's Ames Laboratory, the creation of a superconductivity information exchange. The laboratory publishes a biweekly newsletter on advances in high-temperature superconductivity research, available in both paper and electronic forms; the electronic version is sent out to some 250 researchers.
- Researchers are making new collaborative arrangements. The technology of networks provides increased convenience and faster turnaround times—often several completed message exchanges in one day. For shorter messages, special software allows real-time exchanges.

See box on collaboration, page 20.

DOCUMENT PROCESSING

[An] area of significant change is document processing. This began in the 1960s with a few simple programs that would format typed text. In the context of UNIX* in the 1970s, these ideas led to a new generation of document processing programs and languages, such as SCRIBE and the UNIX-based tools *troff, eqn, tbl,* and *pic.* The quintessence of these ideas are Knuth's TeX and METAFONT systems, which have begun to revolutionize the world's printing industry. In workstations, these ideas have produced WYSIWYG (wizzy-wig, or “what you see is what you get”) systems that display formatted text exactly as it will appear in print. International standards organizations are considering languages for describing documents, and some software manufacturers are constructing systems, such as the POSTSCRIPT protocols, embodying these ideas. The NSF-sponsored EXPRES project, at the University of Michigan and Carnegie Mellon University, illustrates a serious effort to develop a standard method of exchanging full scientific documents by network. Low-cost laser printers now make advanced document preparation and printing facilities available to many people with workstations and personal computers. It is now possible for everyone to submit high-quality, camera-ready copy directly to publishers, thus speeding the publication of new results; however, it is no longer true that a well-formatted document can be trusted to have undergone a careful review and editing before being printed.

SOURCE: Peter J. Denning, 1987, Position Paper: Information Technology in Computing.

As Lederberg noted a decade ago (Lederberg, 1978), digital communication allows scientists to define collegial relationships along the lines of specialized interests rather than spatial location. This is immensely beneficial to science as a whole, but causes some consternation among administrators who find more loyalty to disciplines than to institutions.

Technologies in the process of development show the networks' remarkable potential. Multimedia mail allows researchers to send a combination of still images, video, sound, and text. Teleconferencing provides simultaneous electronic links among several groups. Electronic chalkboards allow researchers to draw on their chalkboard and have the drawing appear on their computer and on the computers of collaborators across the country. Directory services, or “nameservers,” supply directories of the names and network addresses of users, processes, and resources on a given network or on a series of connected networks. Program distribution services include the supply of mathematical software to subscribers. A spectacular new technology is represented in the Metal Oxide Semiconductor Implementation System (MOSIS), a service that contracts for the manufacture of very large-scale integrated (VLSI) chips from circuit diagrams pictured on a subscriber's screen. Fabrication time is often less than 30 days. In one notable example, the researchers designing a radiotelescope in Australia designed custom chips for controlling the telescope. MOSIS returned the chips in a matter of days; the normal manufacturing process would have taken months and would have delayed the development of the instrument considerably.

NEW FORMS OF COLLABORATION THROUGH THE NETWORKS

The development of COMMON LISP (a programming language) would most probably not have been possible without the electronic message system provided by ARPANET, the Department of Defense's Advanced Research Projects Agency network. Design decisions were made on several hundred distinct points, for the most part by consensus, and by simple majority vote when necessary. Except for two one-day face-to-face meetings, all of the language design and discussion was done through the ARPANET message system, which permitted effortless dissemination of messages to dozens of people, and several interchanges per day.

The message system also provided automatic archiving of the entire discussion, which has proved invaluable in preparation of this reference manual. Over the course of thirty months, approximately 3000 messages were sent (an average of three per day), ranging in length from one line to twenty pages… It would have been substantially more difficult to have conducted this discussion by any other means, and would have required much more time.

SOURCE: Guy Steele, 1984. COMMON LISP: The Language. Bedford, MA: Digital Press, pp. xi-xii. Reprinted with permission. Copyright Digital Press/Digital Equipment Corporation.

To share complex information (such as satellite images) over the networks, researchers will need to be able to send entire pictures in a few seconds. One technique that is likely to receive more attention in the future is data compression, which removes redundant information and converts data and images to more compact forms that require less time to transmit.

Among the most important of potential applications of information technology is the emergence of a truly national research network—that is, a set of connections, or gateways, between networks—to which every researcher has access. The National Science Foundation has announced its intention to serve as a lead agency in the development of such a network, beginning with a backbone, called NSFNET, that links the NSF-supported supercomputing centers, and widening to include other existing networks.

Widespread access to networks will also offer much more than just communications links. They can become what the network serving the molecular biology community aims to be: a full-fledged information system.

Difficulties Encountered The principal difficulty with communicating across research communities via electronic mail and file transfer technologies is incompatibility. The networks were formed independently, evolved over many years, and are now numerous. Consequently, networks use different protocols, that is, different conventions for packaging data or text for transmission, for locating an appropriate route from sender to receiver over the physical network, and for signaling the start and stop of a message. For example, a physicist on the High Energy Physics network (HEPNET) trying to send data to a physicist on one of the regional networks would first have to ask “What network are you on?”; “How do I address you?”; and “What form do you want the information in?” In the gateway between two networks, the protocols of the first network must be removed from the message and the protocols for the second added. Under heavy traffic loads, the gateways can become bottlenecks. As a result, navigating from one network to a researcher on another is time-consuming, tiresome, and often unreliable; navigating over two networks to a researcher on a third is prohibitively complex.

Text can frequently be moved from one word processing system to another only with significant loss of formatting information—including the control of spacing, underlining, margins, or indentations. Graphics can only rarely be included with text. Such issues of compatibility may delay the expansion of electronic publishing as well as electronic proposal submission and review—the goals of the National Science Foundation's EXPRES project.

The issues are summarized succinctly by Denning: “Most word processors are inadequate for scientific needs: they cannot handle graphs, illustrations, mathematics and layout, and myriad file formats make exchange extremely difficult. With so many experts and so much competition in the market, it is hard to win agreement on standards. There is virtually no electronic support for the remainder of the process of scientific publication—submission, review, publication, and distribution. These issues can be expected to be resolved over the next few years, as document interchange formats are adopted by standards organizations and incorporated into software revisions and equipment upgrades. However, the transition process will not be painless" (Denning, 1987, pp. 26–27).

In addition, some networks limit use under certain circumstances; for instance, one network bars communication among researchers at industrial laboratories. The fear is that corporations would use a research network for commercial profit or even for sales or marketing. The Panel believes such fear is misplaced and that networks should be open for all research communication.

On the whole, the management of the networks is anarchic. Networks operate not as though they were a service vital to the health of the nation's research community but as small fiefdoms, each with strong disciplinary direction, with little incentive to collaborate. The National Science Foundation has taken an early leadership role, with such initiatives as NSFNET, which addresses many of the current networking problems, and the EXPRES project, which establishes standards for the electronic exchange of complex documents. Such efforts to provide integration and leadership are vital to increased research productivity.

Boxes on pages 22–27 examine network use alternatives.

FROM A NETWORK TO AN INFORMATION RESOURCE PROTOTYPE: BIONET

BIONET is a nonprofit resource for molecular biology computing that provides access to software, recent versions of databases relevant to molecular biology, and electronic communications facilities. Work is in progress to expand BIONET as a logical network reaching molecular biologists throughout the research community worldwide. Many existing physical networks are in use by molecular biologists, and it is BIONET's aim to utilize them all. BIONET is working on plans to provide molecular biologists with access to one or more supercomputers or parallel processing resources. Special programs will be developed to provide molecular biologists with an easy interface to submit supercomputer jobs.

Especially active are the METHODS-AND-REAGENTS bulletin board (for requesting information on lab protocols and/or experimental reagents) and the RESEARCH-NEWS bulletin board, which has become a forum for posting interesting scientific developments and also a place where scientists can introduce their labs and research interests to the rest of the electronic community. Bulletin boards have been instituted for the GenBank and EMBL nucleic acid sequence databases. Copies of messages on these bulletin boards are forwarded to the database staff members for their attention. These bulletin boards serve as a medium for discussing issues relating to the databases and as a place where users of the databases can obtain assistance. Along these same lines BIONET has developed the GENPUB program that facilitates submission of sequence data and author-entered annotations in computer-readable form directly to GenBank and EMBL via the electronic mail network.

The journals CELL and CABIOS have established accounts on BIONET and the Journal of Biological Chemistry and several others

Information Storage and Retrieval

Current Uses How information is stored determines how accessible it is. Scientific texts are generally stored in print (in the jargon, in hard copy) and are accessible through the indices and catalogs of a library. Some texts, along with programs and data, however, are stored electronically—on disks or magnetic tapes to be run in computers—and are generally more easily accessible. In addition, collections of data, known as databases, are sometimes stored in a central location. In general, electronic storage of information holds enormous advantages: it can be stored economically, found quickly without going to another location, and moved easily.

One kind of database holds factual scientific data. The Chemical Abstracts Service, for example, has a library of the molecular structures of all chemical substances reported in the literature since 1961. GenBank is a library of known genetic sequences. Both the National Aeronautics and Space Administration and the National Oceanic and Atmospheric Administration have thousands of tapes holding data on space and the earth and atmosphere.

will also soon be on board. Several journals have indicated an interest in publishing research abstracts on BIONET in advance of hardcopy articles.

Annotated examples of program usage have been included into the HELP ME system. The examples, formatted to be suitable for printing out as a manual, cover the major uses of the BIONET software for data entry, gel management, sequence, structure and restriction site analysis, cloning simulations, database searches, and sequence similarities and alignments. A manual of standard molecular biology lab protocols has also been added to HELP ME for users to reference.

One of BIONET's major goals is to serve as a focus for the development and sharing of new software tools. Towards achieving this goal, BIONET has made available to the community a wide variety of important computer programs donated by a number of software developers. A collaborative effort has occurred between the BIONET staff and the software authors to expand the usefulness of important software by making it compatible with a number of hardware and user community constraints.

BIONET provides an increasing number of databases online: lists of restriction enzymes; a bank of common cloning vector restriction maps and complete vector sequences; a database of regular expressions derived from published consensus sequences; the searchable full text of a recent revision of “Genetic Variations of Drosophila melanogaster” by Dan L. Lindsley and E.H. Grell (the Drosophila “Red Book”). Some of these can be used as input to search programs. BIONET invites curators of genetic and physical genome maps to use this resource for the collection, maintenance, and distribution of their databases.

SOURCE: Roode et al., 1988. “New Developments at BIONET,” *Nucleic Acids Research,* 16(5):1857–1859.

A second kind of database, a reference database, stores information on the literature of the sciences. For example, Chemical Abstracts Service has abstracted all articles published in journals of chemistry since 1970 and makes the abstracts available electronically. The National Library of Medicine operates services that index, abstract, and search the literature database (known as MEDLARS). In addition, it distributes copies of the database for use on local computers and has developed a communications package, called GRATEFUL MED, that simplifies searching the major MEDLARS files—over six million records through 1987. In addition to biomedicine and clinical medicine, the National Library of Medicine partially covers the literature of the disciplines of population control, bioethics, nursing, health administration, and chemistry. One of its most important databases, for instance, is TOXLINE, which references the chemical analysis of toxins. Information search services have grown up around these and other databases, including a number of commercial ones, and now constitute a substantial industry.

BIRTH OF A NETWORK: A HISTORY OF BITNET (EXCERPTED).

BITNET (Because It's Time NETwork) began as a single leased telephone line between the computer centers of The City University of New York (CUNY) and Yale University. It has developed into an international network of computer systems at over 800 institutions worldwide. Because membership is not restricted by disciplinary specialty or funding ability, BITNET plays a unique role in fostering the use of computer networking for scholarly and administrative communication both nationally and internationally.

In 1981, CUNY and Yale had been using internal telecommunications networks to link computers of their own. The New York/New Haven link allowed the same exchanges to take place between two universities. The founders of BITNET—Ira Fuchs, then a CUNY vice chancellor, and Greydon Freeman, the director of the Yale Computing Center—realized that the fledgling network could be used to share a wide range of data. Furthermore, the ease and power of electronic mail showed new potential for cooperative work among scholars; collective projects could now be undertaken that would have been difficult or impossible if conducted by postal mail or by phone.

Fuchs and Freeman approached the directors of other academic computer centers with major IBM installations to invite them to become members of the new network. The plan of shared resources that BITNET offered included two proposals: a) that each institution pay for its own communications link to the network; and b) that each provide facilities for at least one new member to connect. Software was used to create a store-and-forward chain of computers in which files, messages, and commands are passed on without charge from site to site to their final destination. BITNET became a transcontinental network in 1982 when the University of California at Berkeley leased its own line to CUNY. Berkeley agreed to allow other Califor

A database, taken together with the procedures for indexing, cataloging, and searching it, makes up an information management system. Some potentials of information management systems have been predicted for years, beginning with

Vannevar Bush's MEMEX (Bush, 1945). The box on pages 28–29 illustrates a current working information management system that links texts and databases in genetics and medicine.

nia institutions to link to the network through its line, in return for some expense sharing.

In 1984, IBM agreed to support CUNY and EDUCOM (a nonprofit consortium of colleges, universities, and other institutions founded in 1964 to facilitate the use and management of information technology) in organizing a centralized source of information and services to accommodate the growing number of BITNET users. EDUCOM set up a Network Information Center (BITNIC), whose ongoing functions include the handling of registration of new members; at the same time, CUNY established a Development and Operations Center (BITDOC), which develops tools for the network.

BITNET's success (it is now in all fifty states) led to the formation of a worldwide network of computers using the same networking software: in Europe and the Middle East (EARN, the European Academic Research Network), Canada (NetNorth), Japan, Mexico, Chile, and Singapore (all of which are members of BITNET). There is also active interest from other countries in the Far East, Australia and New Zealand, and South America. Although political and funding considerations have forced their administrative segregation, BITNET, EARN, and NetNorth form one topologically interconnected network.

Success has also meant some further structuring of what had once been essentially a buddy system. BITNET is now governed by a board of trustees elected by and from its membership. The members of the board each participate in various policy-making committees focusing on network usage, finance and administration, BITNIC services and activities, and technical issues. What began as a simple device for intercampus sharing is simple no longer.

SOURCE: Holland Cotter, 1988. Birth of a network: A history of BITNET. *CUNY/University Computer Center Communications,* 14:1–10.

Difficulties Encountered For all disciplines, both factual and reference databases promise to be significant sources of knowledge for basic research. But to keep this promise, a Pandora's box of problems will have to be solved.

Difficulties encountered with factual databases, stated succinctly, are: the researcher cannot get access to data; if he can, he cannot read them; if he can read them, he does not know how good they are; and if he finds them good, he cannot merge them with other data. Researchers have difficulty getting access to data stored by other researchers. Such access permits reanalysis and replication, both essential elements of the scientific process. At present, with a few exceptions, data storage is largely an individual researcher's concern, in line with the tradition that researchers have first rights to their data. The result has been a proliferation of idiosyncratic methods for storing, organizing, and indexing data, with one researcher's data essentially inaccessible to all other researchers.

Even if a researcher gets access to a colleague's data, he may not be able to read them. The formats with which data are written on magnetic tape—like the formats used in word processing systems—vary from researcher to researcher, even within disciplines. The same formatting problems prohibit the researcher from merging someone else's data into his own database. In order either to read or to merge another's data, considerable effort must be dedicated to converting tape formats.

Finally, when a researcher gets access to and reads another's database, he often has no notion of the quality of the data it contains. A number of proposals (see Branscomb, 1983, National Research Council, 1978) have been made for the creation of what are called evaluated databases, in which data have been verified by independent assessment.

In fields such as organizational science or public health, the costs of collecting and storing data are so large that researchers often have to depend on case studies of organizations or communities to test hypotheses. Researchers in these fields have proposed combining data from many surveys into databases of national scope. If differences in research protocols and database formats can be resolved, such national databases can increase the quality and effectiveness of research.

THE STUDY PANEL'S EXPERIENCE WITH ITS OWN ELECTRONIC MAIL IS INSTRUCTIVE.

Most of the members of the Panel use electronic mail in their professional work; some use it extensively, exchanging as many as seventy messages in one day. At their first meeting, Panel members and staff decided it would be useful to establish electronic communication links for the Panel. Using a network to which he had access, one of the Panel members devised a distribution-list scheme for the Panel. He designed a system that would allow Panel members to exchange messages or documents easily by naming a common group “address.” This group address would connect everyone by name from their own network. Panel members would not have to remember special codes or routes to other networks, but could use their own familiar network. Also, messages could be sent to one, several, or all of the Panel members at once.

Between December 1986 and March 1988, nearly 2,000 messages went out using the Panel's special electronic group address. In line with what has been found in systematic research on electronic mail by *ad hoc* task groups (Finholt, Sproull and Kiesler, 1987), most of the messages went from study staff managing the project to Panel members. Typically, staff used electronic mail to perform coordinating and attentional functions, e.g., to structure meetings, to ask Panel members for information or to perform writing tasks, and to provide members with progress reports. In addition, some Panel members sent mail through other network channels to each other; for instance, two Panel members exchanged electronic mail about computers in the oceanographic community through BITNET, ARPANET, and OMNET.

Although previous research and our own

The primary difficulty encountered with reference databases is in conducting searches. Most information searches at present are incomplete, cumbersome, inefficient, expensive, and executable only by specialists. Searches are incomplete because databases themselves are incomplete—updating a database is difficult and expensive—and because information is stored in more than one database. Searches are cumbersome and inefficient because different databases are organized according to different principles and cannot readily be searched except by commands specific to each database. Searches are expensive because access is expensive (as much as $300 per hour), because network linkages to the databases impose substantial surcharges, and because the inefficiency of the systems means that searches may have to be repeated.

A difficulty common to both scientific and reference databases is a pressing need for new and more compact forms of data storage. Disciplines such as oceanography, meteorology, space sciences, and high energy physics have already gathered so much data that more efficient means of storage are essential; and others are following close behind. One solution seems to lie in optical disk storage, for which various alternative technologies are under development. Currently, these new techniques lack commonly accepted standards.

informal observations agree in suggesting that the electronic group mail scheme helped the Panel to work more efficiently, the system was used much less extensively than had been originally envisioned. For example, when delivery of report drafts was crucial, the staff relied on overnight postal mail. Network service inadequacies and technical problems are partly to blame; for example, it took months before messages could be sent predictably and reliably to every Panel member. Because the networks do not facilitate access to service support (comparatable to telephone system operators, for example), Panel members had to rely on their own resources to remedy any system inefficiencies. For example, changes to electronic mail addresses in the system could not be made after a few months, so that new addresses had to be added to individual messages.

Such technical problems, though by no means insurmountable, were annoying. Analysis of a sample of messages received by Panel staff indicates that approximately 10 percent contained some complaint about delays, losses of material in transmission, or unavailability of the group mail system. Often, documents were difficult to read because document formatting codes embedded in the document files were removed prior to transmission. A message legible on one system might be filled with unintelligible characters when received on another. At considerable difficulty, some Panel members converted messages received electronically to formats they could read using their text editors. Then they would type in their own revisions, which once again would have to be converted to plain formats to be sent back through the networks. This experience suggests that much needs to be done to make internetwork communication by groups more efficient and easier to use.

Another difficulty is that stored data gradually become useless, either because the storage media decay or the storage technology itself becomes obsolete. Data stored on variant forms of punched cards, on paper tape, or on certain magnetic tape formats may be lost due to the lack of reading devices for such media. Even if the devices still exist, some data stored on magnetic tapes will be lost as the tapes age, unless tapes are copied periodically. Needless to say, such preservation activities often receive low priority.

An important archival activity that also receives a low priority is the conversion of primary and reference data from pre-computer days into machine readable form. In this regard, the efforts of the Chemical Abstract Service to extend their chemical substance and reference databases are praiseworthy.

See box on satellite-derived data, page 30.

HOW A LIBRARY USES COMPUTERS TO ADVANCE PRODUCTIVITY IN SCIENCE

In 1985 the William H. Welch Medical Library of the Johns Hopkins University began a unique collaboration with Dr. Victor A. McKusick, the Johns Hopkins University Press, and the National Library of Medicine to develop and maintain an online version of McKusick's book *Mendelian Inheritance in Man* (known as OMIM, for Online Mendelian Inheritance in Man). While the book contains 3,900 phenotypes (a specific disorder or substance linked to a genetic disease) and updates are issued approximately every five years, OMIM currently describes more than 4,300 phenotypes and is updated every week. A gene map is available, keyed to the phenotype descriptions.

Any registered user worldwide can dial up OMIM and search its contents through a simple three-step process: 1) state the search in simple English (e.g., relationship between Duchenne muscular dystrophy and growth deficiency hormone); 2) examine the list of documents, which are presented in ranked order of relevance; and 3) select one or more documents to read in detail. Having selected a document, the searcher can determine through a single keystroke whether the phenotype has been mapped to a specific chromosome. OMIM entries are also searchable in a related file, the Human Gene Mapping Library (HGML) at Yale University. By mid-1988, researchers will be able to use the same access code to enter and search three related databases: HGML in New Haven, the Jackson Laboratory Mouse Map in Bar Harbor, and OMIM in Baltimore.

OMIM is more than an electronic text. It is a dynamic databases with many applications. Searching the knowledge base is only one of its uses. It can be used as a working tool. For example, at the last biennial international Human Gene Mapping conference in Paris (September 1987) the results of the committees' deliberations were used to update and regenerate the database each evening. Every

Another difficulty in storing information is private ownership. By tradition, researchers hold their data privately. In general, they neither submit their data to central archives nor make their data available via computer. Increasingly, however, in disciplines like meteorology and the biomedical sciences, submission of primary data to data banks has become accepted as a duty. In the field of economics, the National Science Foundation now requires that data collected with the support of the Economics Program be archived in machine readable form, and that any professional article citing program support be accompanied by a fully documented disk describing the underlying data. In the social sciences, a 1985 report of the National Research Council's Committee on National Statistics recommended both that “sharing data should be a regular practice” and that a “comprehensive reference service for computer-readable social science data should be developed.” (Fienberg, Martin, and Straf, 1985.)

morning, the conferees had fresh files to consult. This information was available worldwide at the same time. In the future, these conferences can take place electronically as frequently as desired by the scientific community.

OMIM is a node in an emerging network of biotechnology databases, data banks, tissue repositories, and electronic journals. In a few years, it may be possible to enter any of these files from any one of the related files. Through this kind of linkage, OMIM may serve as a bridge between the molecular geneticists and the clinical geneticists. Currently, these databases are primarily text or numerical files. As technology improves and becomes ubiquitous, and as network band-width expands, databases will routinely include visual images and complex graphics. It may also be possible to jump from one point within a file to relevant and related points deep within other files.

OMIM and its future manifestations result from collaborative efforts and support from diverse groups. Dr. Victor A. McKusick is the scientific expert responsible for the knowledge base; his editorial staff adds new material and updates the database. The National Library of Medicine developed OMIM as part of its Online Reference Works program. The Welch Medical Library provides the computers, network gateways, database maintenance and management, and user support. Finally, the Howard Hughes Medical Institute provides partial support for access, maintenance, and future development of the system.

The Welch Library must work closely with both the author and the users to represent research knowledge in ways that best suit the users' purposes. It must be able to respond quickly to the changing needs of the author and the users. It is in a unique position to study and engineer a new kind of knowledge utility. The OMIM effort is part of a project to develop a range of online texts and databases in genetics and internal medicine, carried out in the Library's Laboratory for Applied Research in Academic Information.

In addition, peer review of articles and proposals has been constrained by the difficulty of gaining access to the data used for analysis. If writers were required to make their primary data available, reviewers could repeat at least part of the analyses reported. Such review would be more stringent, would demand more effort from reviewers, and raises a number of operational questions that need careful consideration; but it would arguably lead to more careful checking of published results.

Underlying the difficulties in information storage and retrieval are problems in the institutional management of resources. Who is to manage, maintain, and update information services? Who is to create and enforce standards? At present the research community has three alternative answers: the federal government, which manages such resources as MEDLINE and the GenBank; professional societies, such as the American Chemical Society, which manages the Chemical Abstracts Service, and the American Psychological Association, which manages Psychological Abstracts; and private for-profit enterprises such as the Institute for Scientific Information.

HANDLING SATELLITE-DERIVED OBSERVATIONAL DATA

At present both the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA) operate earth-orbiting satellites and collect data from them. Both NOAA and NASA store large volumes of primary data from the satellites on digital tape. Both have faced problems, although each organization's problems are different. NOAA, until 1985, had a system that, for purposes of satellite operations, stored environmental satellite data on a Terabit Memory System (TBM). The TBM technology was used from 1978 to 1985, at which time it became obsolete; the more than 1,000 tapes of data collected have been reduced by about 40 percent in transforming most of the useful materials to standard digital tape for storage. NASA has used standard digital tape and disk storage technologies and, since ceding the LANDSAT satellites, has recorded and saved data from its research earth-observing satellites as needed.

Both NASA and NOAA face real problems in making data accessible for scientific analysis. NASA has expended time, effort, and money building a number of satellite data distribution systems that provide digital data archives and a catalog of satellite data holdings, as well as images and graphical analyses produced from satellite data. For example, NASA's National Space Science Data Center received and filled some 2,500 requests for tapes, films, and prints in the first half of fiscal 1988, and also provided network access to specific databases. NOAA has been largely unable to get financial support for its proposed satellite data management systems. Selection of needed information from among the data available remains a problem. Some pilot systems under development at both agencies succeed in leading in the user through a catalog, but fail to contain much valuable new information and data. Both agencies continue to hold great amounts of environmental satellite data in their permanent achieves that are difficult to access, expensive to acquire, and as a result are ignored by many researchers who could benefit from their use. Much remains to be done to improve access to important satellite-derived data.

New Opportunities: Approaching the Revolution Asymptotically

The information technologies and institutions of the past that revolutionized scholarly communication—writing, the mails, the library, the printed book, the encyclopedia, the scientific societies, the telephone—made information more accessible, durable, or portable. The advent of digital information technology and management continues the revolution, suggesting a vision, still somewhat incoherent, of new ways of finding, understanding, storing, and communicating information.

Some technologies involved in the revolution are

- Simulations of natural (or hypothesized) phenomena;
- Visualization of phenomena through graphical displays of data; and
- Emerging use of knowledge-based systems as “intelligent assistants” in managing and interpreting data.

Simulations allow examination of hypotheses that may be untestable under normal conditions. Plasma physicists simulate ways of holding and heating a hot, turbulent plasma until it reaches the temperatures necessary for fusion. Cosmologists simulate the growth of galaxies and clusters of galaxies in an infant universe. Engineers simulate the growth of fractures in a metal airplane wing or nuclear reactor. Chemists' simulations may someday be sophisticated enough to screen out unproductive experiments in advance. Drug companies are considering the use of simulations to design drugs for a particular function, for example, a non-addictive drug that also kills pain. In general, simulations extend researchers' ability to model a system and test the model developed.

See box on simulation, below.

See box on visualization, pages 32-33.

USES OF SIMULATION IN ECONOMETRICS

Simulation techniques take estimated relationships or numerical models that appear to be consistent with observations of actual behavior and apply them to problems of predicting the changes induced by time, or of measuring the relationships among sets of economic variables. For example, simulation models have been utilized to study the effects of oil price changes on the rate of inflation, proposed policies regarding labor law, and future interest rates. In addition, exchanges among groups of agents in an economy have been used in dynamic input-output analysis to make inferences about the feasible or likely future course of economic growth in the entire economy or within specific industries or regions.

There is a growing interest in investigating the properties of models that represent the workings of firms, markets, and whole economies as nonlinear adaptive systems. Recently this has begun to expand the reliance placed by essentially theoretical researchers upon extensive applications of numerical simulation methods. Finally, in both extensions of the line of inquiry just noted and in other contexts, direct simulation of stochastic processes via Monte Carlo techniques can be used by economists to gain insights into the properties of stochastic systems that resist deductive techniques due to their (current) analytic intractability.

SOURCE: Paul A. David and W. Edward Steinmuller, 1987.
Position paper: “The Impact of Information Technology Upon Economic Science,” p.21.

Visualization techniques turn the results of numerical computations into images. The remarkable ability of the human brain to recognize patterns in pictures allows faster understanding of results in solutions to complex problems, as well as faster ways of interacting with computer systems and models. For instance, while small molecules have a few dozen atoms and are easy to visualize, large molecules, like proteins, have tens of thousands of atoms. A useful physical model of the structure of a protein might stand six feet high and cost several thousand dollars. Moreover a researcher could not slice a physical model to see how it looks inside; with visualization techniques, he could. Visualization is the single advanced technology most widely mentioned by Panel members and position paper writers. (For a critical analysis of opportunities in visual imaging, see McCormick, DeFanti, and Brown, 1987.)

VISUALIZATION IN SCIENTIFIC COMPUTING

Scientists need an alternative to numbers. A technical reality today and a cognitive imperative tomorrow are the use of images. The ability of scientists to visualize complex computations and simulations is absolutely essential to ensure the integrity of analyses, to provoke insights, and to communicate those insights with others.

Several visually oriented computer-based technologies already exist today. Some have been exploited by the private sector, and off-the-shelf hardware and software can be purchased; others require new developments; and still others open up new research areas. Visualization technology, well integrated into today's workstation, has found practical application in such areas as product design, electronic publishing, media production and manufacturing automation. Management has found that visualization tools make their companies more productive, more competitive, and more professional.

So far, however, scientists and academics have been largely untouched by this revolution in computing. Secretaries who prepare manuscripts for scientists have better interactive control and visual feedback with their word processors than scientists have over large computing resources that cost several thousand times as much.

Traditionally, scientific problems that required large-scale computing resources needed all the available computational power

Intelligent assistants can serve as interfaces between the researcher and the computer. Just as computers increase our power to collect, store, filter, and retrieve data, they can also help us reason about the data. Over the last three decades, computer scientists have been developing methods for symbolic information processing or artificial intelligence. While these programs are not fully intelligent in the sense that humans are, they allow computers to solve problems that are not reducible to equations.

Artificial intelligence programs have been written for many scientific tasks. These tasks are not expressible in terms of numerical operations alone, and, thus, require symbolic computation. The programs fall into a general class, called expert systems, because they are programmed to reach decisions in much the same way as experts do. Expert systems have been successfully applied to industrial areas such as manufacturing and banking. To date, only a few prototype systems have been written for scientific research. Prototypes include programs that assist in chemical synthesis planning, in planning experiments in molecular genetics, in interpreting mass aspects of organic molecules, in trou bleshooting particle beam lines for high energy physicists, and in automated theory formulation in chemistry, physics, and astronomy.

to perform the analyses or simulations. The ability to visualize results or guide the calculations themselves requires substantially more computing power.

Electronic media, such as videotapes, laser disks, optical disks, and floppy disks, are now necessary for the publication and dissemination of mathematical models, processing algorithms, computer programs, experimental data, and scientific simulations. The reviewer and the reader will need to test models, evaluate algorithms, and execute programs themselves, interactively, without an author's assistance. Scientific publication needs to be extended to make use of visualization-compartible media.

Reading and writing were only democratized in the past 100 years and are the accepted communication tools for scientists and engineers today. A new communication tool, visualization, in time will also be democratized and embraced by the great researchers of the future.

The introduction of visualization technology will profoundly transform the way science is communicated and will facilitate the commission of large-scale engineering projects. Visualization and Science go hand in hand as partners. No one ever expected Gutenberg to be Shakespeare as well. Perhaps we will not have to wait 150 years this time for the geniuses to catch up to the technology.

SOURCE: B. H. McCormick, T. A. DeFanti, and M. D. Brown, 1987. Visualization in Scientific Computing (NSF Report). *Computer Graphics* 21(6), ACM SIGGRAPH: New York, Association for Computing Machinery.

The methods to needed to assist with complex reasoning tasks are themselves the subject of considerable research in such fields as computer science, cognitive science, and linguistics. Research in these fields, in turn, is producing tools that facilitate research in other disciplines.

As these methods are used more widely in the future, some experts predict the conduct of research will change dramatically. Intelligent assistants, in the form of software, can carry out complex planning and interpretation tasks as instructed, leaving humans free to spend time on other tasks. When these reasoning programs are coupled to systems with data-gathering capabilities, much of the drudgery associated with research planning, data collection, and analysis can be reduced. Research laboratories and the conduct of research will become even more productive. When every researcher has intelligent assistants at his/her disposal and when the functions of these assistants are interlinked, science will expand the fronties of knowledge even more rapidly than it now does.

Future technologies will provide other forms of research support. Programs that recognize and follow natural-language commands, like “Give me the data from this file,” can simplify interaction between the researcher and computer systems. Spoken-language recognition offers the advantage of hands-free interaction. Speech production, in which computers generate connected sentences in responses to instructions, will, according to one author, lead to a revolutionary expansion in the use of computers in business and office environments (Koening, 1987). A variety of manipulative interfaces of different kinds are under active exploration (Foley, 1987). For example, the “data glove” is a glove on a computer screen that is an image of a specially-engineered glove on a researcher's hand. The data glove follows the motions of the researcher's hand, permitting a researcher, for instance, to manipulate a molecule directly on screen. When the data glove is coupled with feedback devices in the researcher's glove, a researcher can “feel” the fit between two molecular structure surfaces.

The Panel believes that the mature and emerging information technologies, taken together, suggest a vision of new approaches to scientific and engineering research. The vision focuses on an open infrastructure for research support and communication among researchers, along with the services for maintaining this infrastructure. Below are several examples of parts of the vision and of forms the vision could take. We discuss further steps in the report's final section on recommendations.

See boxes on pages 35–41.

INSTITUTIONAL AND BEHAVIORAL IMPEDIMENTS TO THE USE OF INFORMATION TECHNOLOGY IN RESEARCH

Underlying many of the difficulties we have discussed in the use of information technology in research are institutional and behavioral impediments. We have identified six such impediments that seem to affect research in most or all disciplines:

MOLECULAR GRAPHICS

The use of interactive computer graphics to gain insight into chemical complexity began in 1964. Interactive graphics is now an integral part of academic and industrial research on molecular structures and interactions, and the methodology is being successfully combined with supercomputers to model complex systems such as proteins and DNA. Techniques range from simple black-and-white bit-mapped representations of small molecules for substructure searches and synthetic analyses to the most sophisticated 3D color stereographic displays required for advanced work in genetic engineering and drug design.

The attitude of the research and development community toward molecular modeling has changed. What used to be viewed as a sophisticated and expensive way to make pretty pictures for publication is now seen as a valuable tool for the analysis and design of experiments. Molecular graphics complements crystallography, sequencing, chromatography, mass spectrometry, magnetic resonance, and the other tools of the experimentalist, and is an experimental tool in its own right. The pharmaceutical industry, especially in the new and flourishing fields of genetic and protein engineering, is increasingly using molecular modeling to design modifications to known drugs and to propose new therapeutic agents.

SOURCE: B. H. McCormick, T. A. DeFanti, and M. D. Brown, 1987. Visualization in Scientific Computing (NSF Report). *Computer Graphics* 21(6). ACM SIGGRAPH: New York, Association for Computing Machinery.

(1) Issues of costs and cost sharing;
(2) The problem of standards;
(3) Legal and ethical constraints;
(4) Gaps in training and education;
(5) Risks of organizational change; and
(6) Most fundamental, the absence of an infrastructure for the use of information technology.

Issues of Costs and Cost Sharing Many forces drive developments in information technology and its application to research. The result of these developments is constantly increasing requirements for higher performance computer and communications equipment, making current equipment obsolete. Universities and other research organizations are spending increasing fractions of their budgets on information technology to maintain competitive research facilities and to support computer-related instruction. At a number of private research universities, for example, tution has increased faster than inflation for a number of years, in part to cover some of these costs. It is unrealistic to rely on such funding sources to cover further cost increases that will be required to build local network infrastructures.

RESEARCH ON INTEGRATED INFORMATION SYSTEMS

Nearly a decade ago the Association of American Medical Colleges (AAMC) recognized the strategic importance of information technology to the conduct of biomedical research. In response to a study released by the AAMC in 1982, the National Library of Medicine has supported eleven institutions in efforts to develop strategic plans and prototypes of an Integrated Academic Information Management System (IAIMS). The objective of IAIMS is to develop the institutional information infrastructure that permits their individuals to access information they need for their clinical or research work from any computer terminal whenever and whenever it is needed, pull that information into a local environment, and read, modify, transform it, or otherwise use it for many different purposes.

Several pilot prototype models have emerged. The Baylor Medical College is developing a “virtual notebook,” a set of tools for researchers to collect, manipulate, and store data. Georgetown Medical Center has a model called BIOSYNTHESIS that automatically routes a user's query from one database to another. The knowledge sector development of a comprehensive patient management clinical decision support system called HELP is the IAIMS project focus at the University of Utab; and Johns Hopkins University is developing a knowledge workstation.

A related issue is who will pay for the costs of research computing support Historically, such costs have been partially recovered by bundling them into charges for use of time-shared mainframe computers. As usage has moved from campus mainframes to other options (ranging from supercomputer centers to workstations and personal computers), this source of revenue has been lost, while the needs for administrative staff and support personnel for consulting.

training, and documentation have continued. Efforts to move research support into indirect cost categories have not succeeded as many research institutions and universities fact caps on indirect cost rates and have no room to accommodate new costs.

Advances in communications and computing generate new service that require subsidy during the first years of their existence if they are to be successfully tested. This is particularly true of network-related services. Building services into a national network for research will require significant federal, state, and institutional subsidy, which cannot be recovered from user service charges until large-scale connectivity has been achieved and services are mature. Sources for these subsides must be determined.

A RESONABLE MODEL.

Although the Panel is unaware of anything precisely like the vision it holds for sharing information, proposals for the newly established National Center for Biotechnology Information (NCBI) at the National Library of Medicine may come close. The NCBI proposes to facilitate easy and effective access to a comprehensive array of information sources that support the molecular biology research community.

Many, but not all, of these sources are electronic. They encompass raw data, text, bibliographic information, and graphic representations. Ownership and responsibility for development and maintenance of these sources range from individual researchers to departmental groups, institutes, professional organizations, and federal agencies. Each was designed to serve specific needs and audiences, created in many different hardware configurations and software applications. Consequently, NCBI's mission requires experts in both information technologies and biotechnologies, NCBI staff must

- Provide directories to knowledge sources;
- Create useful network gateways between systems;
- Assist users in using databases effectively;
- Reduce incompatibilities in retrieval approaches, vocabulary, nomenclature and data structures;
- Promote standards for representing information that will reduce redundancy and detect inconsistencies or errors;
- Provide useful tools for manipulating and displaying data; and
- Identify new analytic and descriptive services and systems.

Some computing-intensive universities (e.g., Carnegie Mellon University and Brown University) and Medical centers (e.g., Johns Hopkins University, the University of Utah, Baylor University, and Duke University) are also attempting to develop instances of the vision.

Methods used for cost recovery can have significant impacts on usage. Two alternatives are to charge users for access to services or to charge users for the amount of service used. Networks such as BITNET have grown substantially in connectivity and use because they have fixed annual institutional charges for membership and connection, but charge no fees for use. Use-insensitive charge methods (often referred to as the library model) are attractive to institutions because costs can be treated as infrastructure costs and are predictable. Charges for amount of use, in contrast, can inhibit usage; a major inhibitor to use of commercial databases for information searches, for instance, is the unpredictability of user charges for time spent searching the databases. During the development of network services, it seems desirable to recover costs through fixed access charges wherever possible.

The Problem of Standards The development of standards for interconnection makes it possible for every telephone in the world to communicate with every other telephone. The absence of commonly held and implemented standards that would allow computers to communicate with every other computer and to access information in an intuitive and consistent way is a major impediment to scholarly communication, to the sharing of information resources, and to research productivity.

Standards for computer communication are being developed by many groups. The pace of these efforts is painfully slow, however, and the process is intensely political. The technologies are developing faster than our ability to define standards that can make effective use of them. Further, standards that are developed prematurely can inhibit technological progress; standards developed by one group (for example, and equipment vendor) in isolation create islands of users with whom effective communications is difficult or impossible.

Development of standards not only improves efficiency but also reduces costs. Open interconnection standards permit competition among vendors, which leads to lowered costs and improved capabilities. Proprietary standards restrict competition and lead to increased costs. Federal government procurement rules have been major source of pressure on vendors to support open standards.

Current mechanisms for reaching agreement on standards need examination and significant improvement. Such examination needs input from user groups, which will have to exert pressure on standards bodies and on the vendors who are major players in the standard-setting process.

Legal and Ethical Constraints The primary legal and ethical constraints to wider use of information technology are issues of the confidentiality of, and access to, data. The following discussion will only illustrate these issues; we believe they are too important and too specialized to be adequately addressed in a document as general as this one. In the report's final section, we recommend the establishment of a body that will study and advise on these issues.

Information technology has made possible large-scale research using data on human subjects. For the first time, researchers can merge data collected by national surveys with data collected in medical, insurance, or tax records. For instance, in public health research, long-term studies of workers exposed to specific hazards can be carried out by linking health insurance data on costs with Internal Revenue data on subsequent earnings, Social Security data on disability payments, and mortality data, including data and cause of death (Steinwachs, 1987, Position Paper: Information Technology and the Conduct of Public Health

Research). The scientific potential of such data mergers is enormous; the actual use of mergers is small, primarily because of concerns about privacy and confidentiality.

The right to confidentiality of personal information is held strongly in our society. Concerns about the conflict between researchers' needs and citizens' rights have been extensively explored by a number of scientific working groups, under the auspices of both governmental agencies such as the Census Bureau and private groups (for example, the National Academy of Sciences). As more information about individuals is collected and cross-linked, fears are raised that determined and technically sophisticated computer experts will be able to identify specific individuals, thus breaching promises of confidentiality and privacy of information. The Census Bureau, in particular, fears that publicity surrounding such breaches of confidentiality will undermine public confidence and inhibit cooperation with the decennial censuses.

Although there have been discussions and legislative proposals for outright restrictions on mergers of government survey or census data, a reasonable alternative seems to be to impose severe penalties on researchers who breach confidentiality by making use of information on specific individuals. The issue here, as elsewhere in public policy problems, is the balance of benefits against costs. Does better research balance the risk of compromising perceived fundamental rights to privacy? This is a topic that will need to be debated among both researchers and concerned constituencies in the general public.

THE FAR SIDE OF THE DREAM: THE LIBRARY OF THE FUTURE

“Can you imagine that they used to have libraries where the books didn't talk to each other?” [Marvin Minsky, MIT]

The libraries of today are warehouses for passive objects. The books and journals sit on shelves, waiting for us to use our intelligence to find them, read them, interpret them, and cause them finally to divulge their stored knowledge. “Electronic” libraries of today are no better. Their pages are pages of data files, but the electronic page images are equally passive.

Now imagine the library as an active, intelligent “knowledge server.” It stores the knowledge of the disciplines in complex knowledge structures (perhaps in a formalism yet to be invented). It can reason with this knowledge to satisfy the needs of its users. The needs are expressed naturally, with fluid discourse. The system can, of course, retrieve and exhibit (the electronic textbook). It can collect relevant information; it can summarize; it can pursue relationships.

It acts as a consultant on specific problems, offering advice on particular solutions, justifying those solutions with citations or with a fabric of general reasoning. If the user

A related issue is that of acceptable levels of informed consent for human subjects. At present, consent is usually obtained from each respondent to a survey; it is described as informed because the respondent understands what will be done with responses—usually, that they will be used only for some specific research project. Data-collecting organizations protect the confidenti ality of the information obtained from respondents, but guarantee only that information about specific individuals will not be released in such a way that they can be identified. The extent to which informed consent can be given to unknown future uses of survey data, in particular to their merger with other data sources, is of great concern to survey researchers. Controlling the eventual uses of merged, widely distributed data sets would be difficult.

can suggest a solution or a hypothesis it can check this, even suggest extensions. Or it can critique the user viewpoint, with a detailed rationale of its agreement or disagreement.

… The user of the Library of the Future need not be a person. It may be another knowledge system—that is, any intelligent agent with a need for knowledge. Such a Library will be a network of knowledge systems, in which people and machines collaborate.

Publishing is an activity transformed. Authors may bypass text, adding their increment to human knowledge directly to the knowledge structures. Since the thread of responsibility must be maintained, and since there may be disagreement as knowledge grows, the contributions are authored (incidentally allowing for the computation of royalties for access and use). Knowledge base maintenance (“updating”) itself becomes a vigorous part of the new publishing industry.

SOURCE: Edward A. Feigenbaum, 1986. Autoknowledge: From file servers to knowledge servers. In: *Medinfo 86.* R. Salamon, B. Blum, and M. Jorgensen, eds. New York: Elsevier Science Publishers B.V. (North-Holland).

Another concern that needs to be addressed is one of responsibility in computer-supported decision making.

Scientists, engineers, and clinicians more and more frequently will use complex software to help analyze and interpret their data. Who then is morally and legally responsible for the correctness of their interpretations, and of actions based on them? Experiments involving dangerous materials or human lives may soon be controlled by computers, just as many commercial aircraft landings are at present. Computers may be capable of faster or more precise determinations in some situations than humans. But software designers lack strong guidelines on assignment of responsibility in case of malfunction or unforesen disaster, and lack the expertise to guarantee against malfunctions or disasters. With complex software overlaid on complex hardware, it is impossible to prove beyond a doubt in all circumstances that both hardware, and software are performing precisely as they were specified to perform.

Gaps in Training and Education The training and education necessary for using information technology are lacking. Two decades ago many researchers dealt with computers only indirectly through computer programmers who worked in data processing centers. The development of information technology has brought computing into the researcher's laboratory and office. As a result, the level of computing competence expected of researchers, their support staff, and their students has increased manyfold.

Computers are changing what students need to learn. Undergraduate students of chemistry, for example, need more than the standard courses in organic, inorganic, analytic, and physical chemistry; in the view of many practicing chemists, they should also have courses in calculus, differential equations, linear algebra, and computer simulation techniques, and through formal courses or practical research experience, should be competent in mathematical reasoning, electronics, computer programming, numerical methods, statistical analysis, and the workings of information management systems (Counts, 1987, Position Paper: The Impact of Information Technologies on the Productivity of Chemistry).

Neither students nor researchers can obtain adequate training and education through one-time training courses. Because the numbers of new tools are multiplying, researchers need ways to continuously learn about, evaluate, and, if necessary, adopt these new tools. Using commercial programs and tutorial systems only partly alleviates the problem because the technologies often change faster than such supports can accommodate to the changes. Instructors in the uses of information technologies within the disciplines are rare. Senior researchers are especially hard hit. The Panel took no formal survey, but informal discussions suggest that most senior researchers have had exposure to no more than a one-semester programming course and have few of the skills needed to evaluate and use the available technology.

DOCUMENTS AS LINKED PIECES: HYPERTEXT

The vision of computing technology revolutionizing how we store and access knowledge is as old as the computing age. In 1945 Vannevar Bush proposed MEMEX, an electrooptical-mechanical information retrieval system that could create links between arbitrary chunks of information and allow the user to follow the links in any desired manner. In the early 1960s, Ted Nelson introduced “hypertext,” a form of nonsequential writing: a text branches and allows choices to the reader, best read at an interactive screen. In 1968, Doug Englebart demonstrated a simple hypertext system for hierarchically-structured documents—that is, a list of sections, each of which decomposes into a list of subsections, each of which decomposes into a list of paragraphs, and so on—to which annotations could be added during a multiple-workstation conference. Today hypertext refers to information storage in which documents are preserved as networks of linked pieces rather than as a single linear string of characters; readers can add links and follow links at will.

Nelson's XANADU system is perhaps the most ambitious hypertext system proposed. XANADU would make all the world's knowledge accessible in a global distributed database to which anyone can add information,

For all researchers, learning advanced computing means taking a risk. They must interrupt their work and pay attention to something new and temporarily unproductive. They must become novices, often where sources of appropriate instruction and help are unclear or inaccessible. The investment of time and level of frustration are likely to be high. Understandably, many researchers cannot find the time and the confidence to learn technical computing; some justify their choices with negative attitudes, for example: “I get enough communications as it is; I don't need a computer network,” or “If I put my data on the computer, others will steal it,” or “We are doing fine as things are; why change at this point?”

and in which anyone can browse or search for information. A document is a set of one or more linked nodes of text, plus links to nodes already in the global database; a document may be mostly links, constructed out of pieces already in the database. Users pay a fee proportional to the number of characters they have stored. Anyone accessing an item in the global database pays an access charge, a portion of which is returned to the owner as a royalty. Individuals can store private documents that cannot have public links pointing to them and can attach annotations to public documents that become available to everyone reading those documents. Documents can be composed of different parts including text, graphics, voice, and video. INTERMEDIA, a hypertext system with some of these properties, has been implemented at Brown University and has been used to organize information in a humanities course for presentation to students. Small-scale hypertext systems, such as Apple's Hypercards for the Macintosh, are available on personal computers; their promoters claim these systems will change information retrieval as radically as spreadsheets changed accounting a few years ago.

SOURCE: Peter and Dorothy Denning, personal communication, 1987.

Given these natural but negative attitudes, organizations are sometimes slow in responding to demands for new information technologies. Some research organizations view these attitudes as unchangeable and wait to introduce advanced computing until existing researchers move or retire. Others are actively replacing personnel or creating new departments for computational researchers. Still others are attempting to change attitudes by giving researchers the necessary time and support systems. While we have no data on changes in productivity, there is some evidence that in organizations following the latter course, existing researchers at all ranks can achieve as high computing competence as new personnel (Kiesler and Sproull, 1987).

Because people are now being introduced to computing skills at earlier stages of schooling, the lag in computer expertise is disappearing. Over time, alternatives to personal expertise in the form of user-friendly software or individual assistance from specialists will also develop.

Risks of Organizational Change Changing an organization to make way for advanced information technology and its attendant benefits entails real risks. Administrators and research managers are often reluctant to incur the costs—financial, organizational, behavioral—of new technology. In some cases, administrators and research managers relegate computer resources—hardware, software, and people-based support services—to a lower priority than the procurement and maintenance of experimental equipment. The result can be a long-term suppression of the development and use of the tools of information technology.

See box on electronic laboratory notebook, page 42.

In other cases, administrators are misled into underestimating the time and resources required to deploy new information technology. Efforts to develop effective networks have been insufficiently supported by government planners and research institution administrators, who have been led to assume that technology and services to provide network access are easily put in place. Some administrators have promoted change, but without adequate planning for the resources or infrastructure needed to support users. Problems such as these are exacerbated by overly optimistic advice given the administrators by technological enthusiasts. This particular impediment probably cannot be overcome. It can, however, be alleviated by establishing collaborative arrangements to develop plans for and share the costs of change. EDUCOM, for example, is a consortium of research universities with large computing resources that promotes long-range planning and sharing of resources and experiences.

LEGAL CONSTRAINTS TO AN ELECTRONIC VERSION OF A LABORATORY NOTEBOOK

Today, the paper laboratory notebook is the only legally supportable document for patent applications and other regulatory procedures connected with research. Some organizations, however, routinely distribute electronic versions of laboratory notebook information to managers and other professionals who would otherwise have to visit the research site physically or request photocopies. The benefits of legal electronic notebooks are speculative but attested to by those using them informally (Liscouski, 1987). First, they would help give researchers access to information or expertise that is otherwise lost because people have moved or reside in different departments. Second, they would allow research managers and researchers to observe and compare changes in results over time. Third, they would eliminate or make easier the assembly of paper versions of documents needed for government agencies. The barrier to an electronic notebook is social—its lack of acceptance as a legal document. Such acceptance could take place if legal conditions for an electronic system—storage, format, security—were delineated. However, researchers, scientific associations, and government agencies have failed to develop such guidelines. This failure is probably connected to the traditions of privacy in laboratory notebooks, to the inability to forecast how an electronic system would stand up in court, (and related to that, the risk and unacceptable cost to any single institution of developing a system), and to the uncertainty of the ultimate benefits on some widely accepted index of research effectiveness. Whatever the reasons, the end result is that a complete and accepted electronic notebook remains undeveloped.

Absence of Infrastructure Most fundamental of all the institutional and behavioral impediments to the use of information technology is the absence of an infrastructure that supports that use. Just as use of a large collection of books is made possible by a building and shelves in which to put them, a cataloguing system, borrowing policies, and reference librarians to assist users, so the use of a collection of computers and computer networks is supported by the existence of institutions, services, policies, and experts—in short, by an infrastructure. On the whole, information technology is inadequately supported by current infrastructures.

An infrastructure that supports information technology applications to research should provide

- Access to experts who can help;
- Ways of supporting and rewarding these experts;
- Tools for developing software, and a market in which the tools are evaluated against one another and disseminated;
- Communication links among researchers, experts, and the market; and
- Analogs to the library, places where researchers can store and retrieve information.

Several different kinds of experts in information technology help researchers. Some are specialists in research computing. Some are programmers who develop and maintain software specific to research. Others are specialists who carry out searches. Still others are "gatekeepers," who help with choices of software and hardware. Gatekeepers are members of an informal network of helpers centered around advocates and specialists, experts in both a discipline and in information technology who become known by reputation. Overdependence on gatekeepers creates other problems: as with any informal service, some advice received may be narrowly focused or simply wrong and the number of persons wanting free information often becomes larger than the number of persons able to provide it. As a result, the gatekeepers may become overloaded and eventually retreat from their gatekeeping roles.

To hold on to expert help of all types, research and funding institutions must find ways of supporting and rewarding it. While institutions and disciplines have evolved ways of rewarding researchers—publication in refereed journals, promotion, tenure—no such systems yet reward expert help.

Another aspect of the needed infrastructure is some formal provision for developing and disseminating software for specific research applications. Tools for constructing reliable, efficient, customized, and well-documented software are not used in support of scientific research. Computer science, as a supporting discipline, needs to facilitate rapid delivery of finished software, and easy extension and revision of existing software. The Department of Defense has recently pioneered the creation of a Software Engineering Institute at Carnegie Mellon University. Efforts to create tool building and research resources for nondefense software are worth encouraging.

Development and dissemination of scientific software could be speeded in many cases by adoption of emerging commercial standards. These standards are supported by many vendors for a variety of computing environments. The temptation to narrowly match software to specific applications should be resisted in favor of standard approaches.

Software, once developed, needs to be evaluated and disseminated. The research establishment now evaluates research information principally through peer review of funding proposals and manuscripts submitted for publication. Software needs to be dealt with in a similar manner. EDUCOM has recently announced its support of a peer-review process for certain kinds of academic software. Other prototypes of systems for evaluating and disseminating software already exist (see boxes on BIONET and on IBM's software market). These prototypes couple an electronic "market," through which software can be disseminated, with a conferencing capability that allows anyone with access to contribute to the evaluation of the market wares. The system provides an extremely important feature: those contributors who are most successful in the open market can automatically be identified and given credit in much the same way as authors of books and research papers now are.

AN EXAMPLE OF A SOFTWARE MARKET INFRASTRUCTURE: IBM RESEARCH

IBM's internal computer network connects over 2,000 individual computers worldwide, providing IBM's researchers, developers, and other employees with communications facilities such as electronic mail, file transfers, and access to remote computers. In recent years, software repositories and online conferencing facilities have grown and flourished, and become one of the primary uses of the network. With a single command, any IBMer has access to some 3,000 software packages, developed by other IBMers around the world and made available through the network. Many of these packages are computer utilities and programming tools, but others are tools for research. They include statistical and graphics applications, simulation systems, and AI and expert system shells, as well as many everyday utilities to make general use of the computer simpler. The high level of interconnection offered by the network and the centralization of information offered by the repositories allows scientists with a particular need to see if software to satisfy that need is available, to obtain it if it is, and to develop it if it is not, with confidence that they are not duplicating the efforts of some colleague.

The online conferences (public special-purpose electronic bulletin boards), which are as widespread and accessible as the software repositories, allow users of the software (and of commercial and other software) to exchange experiences, questions, and problems. These conferences provide a form of peer review for the software developer. For internally developed software, they provide a fast and convenient channel between the software author and the users; authors with an interest in improving their programs have instant access to user suggestions and to

The infrastructure for information technology also depends on communication links. The Panel believes that one of the most important services that computer networks can provide is the link between users and expert help. Existing links often take the form of electronic bulletin boards on various networks; other mechanisms also exist. Until more formal mechanisms come about, open communication with pioneers, advocates, and enthusiasts is one of

See software market, box below.

the best ways to allow new technologies to be disseminated and evaluated by research communities.

eager testers. Users with a special need or a hard question have equally fast access to the author for enhancements or answers.

The conferences also allow users with common interests to exchange other sorts of information in the traditional bulletin board style. AI researchers debate the usefulness of the concept of intentionality or discuss how software engineering methodologies apply to expert systems development; computer graphics and vision workers talk about the number of bits required to present a satisfactory image to the human eye.

Over 100 individual conferences support thousands of separate discussions about computer hardware and software and virtually all other aspects of IBM's undertakings. The software repositories provide a “reviewed” set of tools and applications for a broad population on a wide spectrum of problems.

The organization that originally sets up a repository or a conference generally provides user support for it (answering “how to do it” questions), and installation and maintenance of local services is usually handled either by an onsite group that has an interest in the specialty served by the facility, or on a more formal basis by the local Information Systems department.

The benefits of these repositories and conferences are at least as widely distributed and probably even harder to quantify, but the success of these software libraries and online conferences within IBM should serve as an encouraging sign for others with the same sorts of needs. A market can be made to succeed, provided that high levels of standardization and compatibility in both hardware and software can be achieved. Such levels of interoperability have, so far, been easier to achieve at commercial institutions such as IBM Research than at research universities. Such as IBM Research than at research universities.

A final piece of infrastructure largely missing is housing and support for the storing and sharing of information. Such a function could be performed by disciplinary groups or, more generally, at the university level. Many university libraries have a professional core staff whose members hold faculty rank and function not only as librarians but also as researchers and teachers. Some university computer centers operate similarly. National laboratories, like astronomical observatories and accelerator facilities, have a core staff of astronomers or physicists whose main task is to serve outside users while also maintaining their own research programs.

The existence of such a professional staff involved in the storage and retrieval of information for a discipline would provide a means of recognizing, rewarding, and providing status to these people. In some cases, a university might wish to consider integrating its information science department with its computer center and its library.

Panel Findings and Recommendations

The vision of a more productive research enterprise, and the recognition of difficulties in reaching that vision, prompted this report. In preceding sections, we reviewed many aspects of the use of information technology in research. In this section we present our findings and recommendations.

FINDINGS

The Panel finds that

1. Information technology has already had a significant and widespread impact on the conduct of research. For the future, that impact amounts to a revolution.

Computer and communication technologies are valuable to every scientific discipline and essential to a growing number of them. The Panel examined the uses of information technology in ten science and engineering disciplines. Although these fields use information technology in different ways, the Panel also found many similarities. These can be summarized as follows:

- As the power and speed of computers have increased, numerical computations of increasing complexity have become practical. The result has been more realistic simulation of systems either physically impossible or too costly to study directly.
- Another consequence of increased computational power has been the capability to collect, store, retrieve, manipulate, and analyze enormous quantities of information. Electronic storage of information means that huge information archives—for example, the Human Genome Project—are not only feasible but also accessible to more researchers from different disciplines.
- An additional consequence of increased computational power has been the capability to present the results of numerical computations as visual images. The

remarkable ability of the human brain to recognize patterns in pictures allows faster understanding of computational results and speedier, more efficient interaction with models when numbers are turned into visual images.

- Information technology allows computers to take over monitoring and control of scientific instruments. This in turn makes scientific observation more convenient, more reliable, and often lower in cost; in some cases, it has led to new computer-based instruments that extend the bounds of observation.
- Information technology has greatly expanded the capabilities for communication among researchers. Communications systems mediated by computers have led to the rapid and relatively inexpensive exchange of everything from memoranda to massive data files.

The Panel further finds that:

2. Significant impediments to the widespread use of information technology in research require careful attention. Some impediments are technical and financial. Other impediments, which up to now have received the least attention, are behavioral and institutional.

- Technical impediments are serious in a few fields. Fields doing large-scale experiments, using satellites to gather data, and using graphics to analyze large amounts of data will for the foreseeable future need computers, software, and networks that are bigger, faster, more capable, and more efficient. The needs of these fields deserve, and will continue to receive, attention from computer scientists and engineers, from information technology manufacturers, and from institutions such as the federal government and universities.
- Financial impediments are chronic. Despite the decreasing cost of hardware, no sources will ever supply enough money to provide every researcher the best information technology environment. The institutions that fund researchers will continue to do their best, and information technology will continue to need more funding.

While the Panel cannot suggest detailed means for increasing the total resources devoted to information technology, it does feel strongly that the provision of such resources is critical to progress in American clinical, engineering, and scientific research. In the resource-constrained environment likely to confront research in the future, difficult decisions will be made on reallocating necessary resources. With a declining population pool from which to draw new scientists, clinicians, and engineers, and with the increasing complexity of research, the Panel believes that increased support of information technology in research deserves high priority. If the nation's best researchers are not given the appropriate tools, our role in the international scientific community will continue to diminish in importance. The limited resources of federal agencies must continue to be allocated on the basis of scientific merit and significance of the proposed research activities.

- The Panel believes that serious impediments to increasing the use of information technology in research are behavioral and institutional. These can be sorted into four major categories: problems of access, problems with learning and use, attitudes of individual researchers, and problems of management.
- Problems of access. Computer networks, hardware, and software are not necessarily accessible to the researchers who want them, nor can everyone who wants them afford them. Network access is still limited and inconvenient. The limitation is especially important since network access permits both collaboration with distant colleagues and access to computing resources. For many researchers in a large number of fields, the hardware to which they have access is adequate. Some researchers, however, require large amounts of time on specialized, expensive hardware; supercomputers leads researchers to reconsider scientific problems, one can anticipate increased demands for access and network use; researchers in the future may find time on supercomputers limited and networks congested.

Finding the right software is a more serious problem. Software that is commercially available is often unsuited to the specialized needs of the researcher, and professionals who create software for research groups are rare. Consequently, many researchers, who are usually not skilled software creators, develop their own. Too often, members of the research community waste time, effort, and money duplicating one another's efforts.

What appears to be lacking are institutional means for providing the services of skilled professionals to create and maintain appropriate software. Funding sources do not consider support for such professionals essential; institutions give them neither recognition nor career status.

Also needed is an institutional system to collect, review, document, and disseminate scientific and engineering application software. Prototypes exist but need broader evaluation. Such a system could evaluate software's effectiveness and eventually lead to standardization of software for wider use.

- Problems with learning and use. Learning to use information technology requires a large investment in time and effort before the investment pays off, and help is hard to find. Researchers who need information technology face difficult choices: they must either learn to use whatever hardware and software the market offers, create their own, or do without. When researchers learn to use existing information technology, they receive haphazard help in learning; instruction or specialists in information technology are often unavailable. Neither the researchers' disciplines nor their institutions provide incentives for learning. Furthermore, researchers must invest time at the expense of their research productivity during the learning period.

Once researchers learn the necessary information technology, they face problems in using it. Networks pose a significant problem. Most of the approximately 100 research-oriented computer networks in the United States were established to serve the needs of small and widely scattered communities of researchers. As a result, a researcher on one network wanting to communicate with a researcher on another faces problems of compatibility. The technical aspects of these problems are tractable; the institutional and behavioral aspects are less so. No one agrees on how procedures for using the networks might be standardized. The networks are not well coordinated with one another, and users have limited opportunity to suggest improvements.

- Problems of the attitudes of individual researchers are twofold. For reasons enumerated above, researchers often approach new information technology applications cautiously. When senior researchers, who are involved in peer review and decisions about publications and research proposals, are resistant, their attitudes can lead to the rejection of innovative applications of information technology to research. Another problem is that of proprietary attitudes toward research data in many disciplines. So long as primary data are viewed as the exclusive property of the researcher who collects them, they will not be available to other researchers. Even if data are made available, they will be left in idiosyncratic formats with insufficient explanatory documentation; and the effort required to make these data usable to the research community will not receive high priority. Increased access to data does, however, raise issues of how large volumes of primary data should be stored, and of the need for validating stored information.
- Problems of managing information technology. Some basic questions need to be addressed: who is to manage, maintain, and update information services? Who will create standards? How will costs be charged, and who will pay for them? No current institutional framework provides the ideal answer to these questions. Federal agencies, professional societies and scientific associations, and private profit-making groups need to consider how to address the needs of research users of information technology.

RECOMMENDATIONS

Recommendation I

The institutions supporting the nation's researchers must recognize and meet their responsibilities to develop and support policies, services, and standards that help researchers use information technology more widely and productively. Specifically, we recommend that

- *Universities* provide accessible, expert help in learning and using information technology.
- *University departments,* and *scientific and professional groups,* establish career ladders for scientific programming positions.
- *Funding agencies* provide support for scientific programming and for help services in learning and using information technology systems for research.

- *Scientific associations* establish disciplinary standards for the storage and indexing of scientific data.
- *University departments,* and *scientific and professional groups,* implement mechanisms for the evaluation, merit (peer) review, and dissemination of software useful in the conduct of research.
- *Vendors,* in collaboration with *scientific groups,* establish standards for simplified and consistent user-machine interfaces.
- *Network administrators* provide simple user interfaces and addressing schemes, add gateways to other networks, improve system reliability and capacity, and provide online help, such as guides to services and mail addresses of individuals who can answer questions.
- *Information service providers* create simplified common standards for accessing and querying information sources, and eventually provide unified access to information.
- *Software vendors,* and *scientific and professional groups,* create program libraries and make them accessible through the networks.

Rationale. Information technology is now becoming an essential component of the research environment. The services needed by research users include

- Access to computers;
- Access to networks, both local and wide-area;
- Long-term storage of and access to data;
- Hardware maintenance and augmentation;
- Help in learning to use existing software and services;
- Production of new software and customization of existing software; and
- Collection, review, documentation, and dissemination of software.

In some instances these services may be efficiently provided by a central organization; in others, by decentralized groups. The services may be provided either by augmenting the responsibilities of existing groups or by creating new groups. Currently such services are being provided in a variety of ways, with highly variable degrees of success and efficiency, across many laboratories, professional societies, universities, and federal agencies. In the most successful models, the researchers feel their needs are paramount. These models need to be publicized, evaluated, and disseminated so that policy and allocation decisions are well informed by the views of research users of information technology.

With regard to policies, the Panel recognized a number of issues, cited earlier, that need further discussion; the appropriate groups may wish to consider the specific recommendations that follow.

- The federal government, through policies of its research-supporting agencies, should ensure proper support for software development for scientific research. Software developed should meet minimal standards of compatibility, reliability, and documentation, and should be made available to other researchers.

- As a general principle, data collected with government support rightfully belongs in the public domain, although the right of researchers to reasonable time for first publication must be respected. Federal agencies, scientific societies and professional associations, university consortia, and other private groups may wish to make specific recommendations based on their reexamination of the implications of the adoption of such a policy. Such recommendations might include the creation and maintenance of data banks and indices to data by research communities, the potential of evaluated databases, and the possibility of including reanalyses of data as part of peer review of publications.
- There is a pressing need for new and more compact forms of data storage. One particular area in which these techniques would be useful is that of image compression. One solution seems to lie in optical disk storage. Unfortunately, these new techniques are still immature and lack commonly accepted standards. Federal agencies should encourage engineering research on optical storage techniques for scientific purposes.
- Efforts to create tool building and research resources for nondefense software should be encouraged.
- The federal government should fund pilot efforts of two kinds. One would implement information storage and software dissemination concepts for selected disciplines. The other would implement the concept of software markets, with emphasis on the development of generic tools useful in several or many disciplines.
- In addition, these pilot projects should be coupled with exploration of such policy issues as protection of confidentiality of information about human subjects, protection of intellectual property, and information security concerns in a global electronic information environment.

Recommendation II

The institutions supporting the nation's researchers, led by the federal government, should develop an interconnected national information technology network for use by all qualified researchers.

Specifically, we recommend that

- The Office of Science and Technology Policy (OSTP) in the Executive Office of the President and the federal agencies responsible for supporting and performing research and development plan and fund a nationwide infrastructure for computer-based research communication.
- Planning and development of this nationwide infrastructure be guided by users of information technology in research, rather than by technical experts in information technology or hardware or software vendors. The Panel believes strongly that such a national network is too important to the future of research to be left only to the technical experts.

- The national research network be founded on the fundamental premise of open access to all qualified researchers/scholars that has nurtured the world's scientific community for centuries.
- The national research network be developed in an evolutionary manner, making full use of the existing successful networks for research.

Rationale. The Panel views the federal government's role in developing a national network for research as analogous to its role in developing the nation's network of roads, streets, and highways. Here, the federal government has planned and funded the interstate highway system and the national highways, and has imposed or encouraged certain national standards. State and local governments have planned and funded a network of highways, roads, and streets that is fully interconnected and compatible with the federal framework. In the national research network, analogs would be research institutions, scientific and professional associations, and corporate groups operating individually or banded together in consortia.

Funding for this undertaking, as outlined in the report of the Federal Coordinating Council on Science, Engineering, and Technology (FCCSET) Committee issued recently (Office of Science and Technology Policy, 1987) should be made available to ensure an advanced national network infrastructure and services for the nation's research communities. Appropriate division of responsibilities among federal and state agencies and research institutions warrants careful attention.

The Panel notes that the Director of the National Science Foundation has announced the Foundation's intention to serve as lead agency in developing a national network. The Panel strongly supports the concept of a lead agency, believing that leadership in coordination of support for a national research network is an essential element in the nation's science policy. The Panel believes the National Science Foundation would be an appropriate lead agency, given its legislatively mandated responsibility for supporting research and education across the full range of science and engineering disciplines.

The creation of a national network will take a considerable amount of time. In the meantime some valuable and much-used networks—such as BIONET, OMNET, BITNET, or HEPNET—exist within particular scientific or academic communities. Many of these networks have been improved with advice from their users, and can be an invaluable source of advice to the designers of any national science network. It is particularly important, therefore, that the existing academic and disciplinary networks continue to receive support until such time as they can either be integrated into or supplanted by the national network.

The Panel's recommendation for a national research network is similar in substance and spirit to the more detailed recommendations contained in a recent report of the National Research Network Review Committee of the NRC's Computer Science and Technology Board (National Research Council, 1988).

Crucial to setting up and running a national network infrastructure is participation by users. The Panel urges that agencies work with an advisory board largely composed of users. A research network of national scope must be oriented toward the research user of information technology. Its philosophy and structure should be such that learning, entering, using, and leaving the network are simple. This should involve, among other things, an easily interpreted program for helping the user, instructions for different levels of use, and simple connections between the network and many varieties of terminals, personal computers, and workstations. The network should be capable of transmitting graphics, symbols, and large amounts of data quickly. It should have gateways to networks in other countries. It should be supported by a professional staff whose main task is to help users.

The Panel recognizes that constraints on access to information are sometimes warranted in cases where the privacy of personal information, the protection of human subjects or of intellectual property, or national security concerns are of overriding importance. Nevertheless, the Panel believes that the interests of the global research community are best served by establishing open and unfettered access as a fundamental presumption in the operation of a national research network.

Recommendation III

To facilitate implementation of Recommendations I and II, and to focus continuing attention on the opportunities and impediments associated with research uses of information technology, the Panel recommends the establishment at a national level of a user's group to oversee and advise on the evolution and use of information technology in support of scientific, engineering, and clinical research.

Specifically, the National Research Council (NRC) should charge a standing committee or board (whether existing or newly created) with the mandate to oversee and advise on research use of information technology. The membership of this board should include a majority of users from a variety of research disciplines.

Rationale. The problems and needed changes addressed by Recommendations I and II are diverse and do not have short-term solutions, and, therefore, require some institutional setting for ongoing examination and discussion. Leadership and coordination in the application of information technology in research would be provided best by a single organization.

Many organizations currently promote developments in information technology at a national level. Among these are: the National Science Foundation and its Computer and Information Science and Engineering (CISE) Directorate and Division of Advanced Scientific Computing, especially through the NSFNET initiative, EXPRES program, and the Panel on Graphics, Image Processing, and Workstations; the National Research Council, especially its Computer Science and Technology Board, Numerical Data Advisory Board, Committee on National

Statistics, and Board on Telecommunications and Computer Applications; a variety of initiatives within the National Library of Medicine; and the Federal Coordinating Council on Science, Engineering, and Technology (FCCSET) of the Office of Science and Technology Policy (especially its committees on Networking and on High Performance Computing). These activities, however, are fragmented, specialized, and tend to give little attention to the behavioral aspects of research uses of information technology.

The Panel recommends location of the user's group within the National Research Council (NRC) rather than a federal agency because it believes that the group ought to be free to focus on the interests and concerns of users, unconstrained by immediate concerns for the distribution of funds among disciplines or agencies. Several boards or committees now existing at the NRC have titles related to the charge to be given this new board. However, it is the Panel's view that none of the existing groups represents adequately the needs of users. Whether one of the existing groups might be reconstituted to create a body of the required nature or whether a new body might be founded is, in the view of the Panel, far less important than the nature of the resulting body. What is needed is a group of researchers who use information technology, supplemented by a few expert providers of information technology.

The National Research Council unit (whether existing or new) would have several specific functions. One function would be to advise policymakers on a broad range of issues related to research uses of information technology. Some of these issues are: international implications of networking; national security concerns associated with scientific databases; and issues of cost for network communications. Ad hoc study panels would be established as appropriate.

Another important function of the unit would be to collect, analyze, and disseminate data on how researchers use information technology. In the course of its deliberations the Panel became painfully aware that these data presently do not exist. Such data would inform not only the actions of those responsible for the support of research but would also apprise researchers themselves of new opportunities offered by information technology. A model for such activities among present National Research Council operating groups might be the Office of Scientific and Engineering Personnel.

This mechanism would also provide a forum to facilitate the transfer of technology: by sponsoring workshops for scientists on newly developed information technology and on coordination of approaches to simplified standards; by exchanging information with technology developers; and by coordinating interaction among scientific organizations and professional associations. The board would disseminate information on current uses of information technology and new developments.

The unit would also convene meetings between researchers and those responsible for supporting research. By providing a forum for discussion, it would ensure that the needs of the research community are brought to the attention of appropriate officials and administrators.

Appendix A

List of Position Papers

Drafts of these papers can be obtained by contacting the study director, Dr. John R. B. Clement, or the Committee on Science, Engineering, and Public Policy, at the following address:

National Academy of Sciences, Room NAS 246
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

RICHARD J. BLAKELY, U.S. Geological Survey, Menlo Park, California: “Computers and Solid-Earth Geophysics”

RICHARD W. COUNTS, Quantum Chemistry Exchange Program, Indiana University:

“The Impact of Information Technologies on the Productivity of Chemistry”

PAUL A. DAVID and W. EDWARD STEINMUELLER, Center for Economic Policy Research, Stanford University: “The Impact of Information Technology Upon Economic Science”

PETER J. DENNING, Research Institute for Advanced Computer Science, NASA Ames Research Center, Moffet Field, California: “Information Technology in Computing”

JOHN HUBBARD, Department of Mathematics, Cornell University: “Computers in Mathematics”

HARVEY NEWMAN, Physics Department, California Institute of Technology: “Computing and Data Communications for High Energy Physics”

CYNTHIA H. NULL, Phychology Department, The College of William and Mary, and BERT F. GREEN, Psychology Department, The Johns Hopkins University: “Computers in Behavioral Science”

DONALD M. STEINWACHS, School of Hygiene and Public Health, The Johns Hopkins University: “Information Technologies and the Conduct of Public Health Research”

Appendix B

Biographies of Panel Members

DONALD N. LANGENBERG, Chancellor, University of Illinois at Chicago, Chicago, Illinois (*Chairman*).

For much of his career, Dr. Langenberg was Professor of Physics and, for a time, also Professor of Electrical Engineering and Science at the University of Pennsylvania. There he directed a materials research facility, the Laboratory for Research on the Structure of Matter, and served as Vice Provost for Graduate Studies and Research. From 1980 to 1982, he served as Deputy Director of the National Science Foundation and, for several months, as Acting Director. He became the first Chancellor of the University of Illinois' Chicago campus (UIC) in 1983. His research interests were in condensed matter physics and included the Fermi surfaces of metals and semiconductors, tunneling, Josephson effects and nonequilibrium phenomena in superconductors, and precision measurement and the fundamental physical constants. He has served on many boards and advisory committees, and is currently a member of the Board of Directors of the American Association for the Advancement of Science (AAAS). Dr. Langenberg holds a B.S. degree from Iowa State University, an M.S. degree from the University of California at Los Angeles, and a Ph.D. degree from the University of California at Berkeley, all in physics. He also holds an honorary M.A. and an honorary D.Sc. from the University of Pennsylvania. Unfortunately, he is, as noted in the Preface, computer illiterate.

W. RICHARDS ADRION, Chair, Computer and Information Science Department, University of Massachusetts, Amherst, Massachusetts.

Dr. Adrion came to the University of Massachusetts in 1986 from the National Science Foundation, where he was most recently Chief Scientist for the Directorate for Computer and Information Science and Engineering, and before that Deputy Director for Computer Research. Previously, he served as manager of the Software Engineering Group at the National Bureau of Standards, and has been on the faculty at both Oregon State University and The University of Texas at

Austin, and as adjunct professor at The American University, George Washington University, and Georgetown University. His research interests are in the areas of programming systems and software engineering, especially programming environments, program verification, and object bases for software development. Dr. Adrion earned bachelor's and master's degrees in electrical engineering at Cornell University, and a Ph.D. degree in the same subject at The University of Texas at Austin.

JOSEPH BALLAM, Professor of Physics, Stanford Linear Accelerator, Stanford University, Stanford, California.

Dr. Ballam is emeritus associate director of the Stanford Linear Accelerator Center and head of its Research Division. He was on the physics faculty of Princeton University and a professor of physics at Michigan State University. His research interests include elementary particles, cosmic rays, and experimental high energy physics. Dr. Ballam received a B.S. degree from the University of Michigan and a Ph.D. degree from the University of California at Berkeley.

BRUCE G. BUCHANAN, Professor and Co-Director, Center for Parallel, Distributed, and Intelligent Systems, University of Pittsburgh, Pittsburgh, Pennsylvania.

For many years Dr. Buchanan was professor of Computer Science Research and co-director of the knowledge Systems Laboratory at Stanford University. In 1988 he moved to the University of Pittsburgh. Professor Buchanan's main research interest is in artificial intelligence, in particular the design of intelligent computer programs that assist scientists and physicians. These include programs and methods for knowledge acquisition and machine learning, scientific hypothesis formation, and construction of expert systems. He was one of the principals in the design and development of DENDRAL, Meta-DENDRAL, MYCI, E-MYCIN, and PROTEAN systems. Dr. Buchanan holds a B.A. degree in mathematics from Ohio Wesleyan University, and M.S. and Ph.D. degrees in philosophy from Michigan State University.

WILLIAM J. EMERY, Professor, Aerospace Engineering Science, Colorado Center for Astrodynamics Research, University of Colorado, Boulder, Colorado.

Trained as a physical oceanographer, Dr. Emery leads a group concentrating on satellite remote sensing of atmosphere and ocean. In cooperation with NOAA's Program for Regional Observing and Forecasting Services (PROFS), his group operates a satellite receiving system to collect data from operational weather satellites. His research interests include large-scale ocean and atmosphere problems with an emphasis on the analysis of large volumes of data. As a consequence of the need to analyze satellite images, he has developed new image-processing tools for SUN, DEC/SPX, and Macintosh II workstations in order to provide students with easy access to a variety of display systems. Dr. Emery has a B.S. from Brigham Young University in mechanical engineering and a Ph.D. from the University of Hawaii in physical oceanography.

DAVID A. HODGES, Professor of Electrical Engineering and Computer Science, University of California, Berkeley, California.

Dr. Hodges teaches and researches microelectronics technology and design, communications and computer systems, and computer-integrated manufacturing systems at the University of California, where he has been a member of the faculty since 1970. Before then, he held positions at Bell Telephone Laboratories, in the components area at Murray Hill and as head of the systems elements department at Holmdel. He is a member of the National Academy of Engineering and has served on several National Academy of Engineering and National Research Council committees. Dr. Hodges' degrees are in electrical engineering; he earned his B.S. degree at Cornell University and his M.S. and Ph.D. degrees at the University of California, Berkeley.

DAVID A. HOFFMAN, Professor, Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts.

David Hoffman is Professor of Mathematics at the University of Massachusetts at Amherst, and a member of the Geometry Analysis Numerics and Graphics Center. His recent research interests include the use of computation and computer graphics in the study of extremal surfaces in mathematics and polymer physics. He has held positions at the University of Michigan, Stanford University, and IMPA, Rio de Janeiro. Dr. Hoffman has a B.A. degree from the University of Rochester and M.S. and Ph.D. degrees from Stanford University in mathematics.

F. THOMAS JUSTER, Professor of Economics, University of Michigan, Ann Arbor, Michigan.

Dr. Juster is also Research Scientist in the Survey Research Center at the University of Michigan's Institute for Social Research. He has served on a number of National Research Council committees, including the Committee on National Statistics, and chairs both the NRC Committee on the Supply and Demand for Mathematics and Science Teachers as well as the American Economic Association Committee on the Quality of Economic Data. Dr. Juster is a fellow of the American Statistical Association. His research includes the design of economic and social accounting systems, the development of measures of economic welfare, consumer behavior and forecasting, and analysis of household saving and asset accumulation behavior. Dr. Juster received a B.S. degree from Rutgers University in education and a Ph.D. degree from Columbia University in economics.

SARA B. KIESLER, Professor, Social Sciences and Social Psychology, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Dr. Kiesler has been on the faculty of the College of Humanities and Social Sciences and of the Robotics Institute at Carnegie Mellon since 1979; previously, she was a staff member of the National Research Council and a professor at the University of Kansas. She was the senior staff director of three National Research

Council committees that produced reports on aging, basic research in education, and behavioral and social sciences. She has served on several committees of the National Academies of Sciences (NAS) and Engineering (NAE). Her research interests include the study of the introduction and impact of computer and computer-communication technologies in groups and organizations. Dr. Kiesler received a B.S. degree from Simmons College in social science, an M.A. degree from Stanford University, and a Ph.D. degree from Ohio State University, both in psychology.

KENNETH M. KING, President, EDUCOM, Princeton, New Jersey.

Kenneth King currently serves as President of EDUCOM, a consortium of 550 universities founded to develop and work toward common goals in information technology and communications. Previously, he was Vice Provost for Computing at Cornell University, where he was responsible for all academic and administrative computing. This included active programs in the development of instructional software for microcomputers, in networking, in library system development, and a national supercomputer center. Before that, he was professor, vice chancellor for university systems, and university dean for computer systems at the City University of New York. He also was director of Columbia University's computer center, and manager of Columbia's IBM Watson Scientific Computing laboratory. Dr. King received a B.A. degree in physics from Reed College and a Ph.D. degree in theoretical physics from Columbia University.

ROBERT LANGRIDGE, Professor of Pharmaceutical Chemistry in the School of Pharmacy and Professor of Biochemistry and Biophysics in the School of Medicine, University of California, San Francisco, California.

Dr. Langridge also directs the Computer Graphics Laboratory at his university's School of Pharmacy. He has been a professor in the biochemistry department at Princeton University and a professor in the department of biophysics at the University of Chicago. His research includes computer graphics, biomolecular structure and function, protein engineering, and drug design. In particular, he uses computer graphics to visualize the motions of molecules in three dimensions and in time. Dr. Langridge holds a B.Sc. degree from the University of London in physics and a Ph.D. degree from King's College, the University of London, in crystallography.

NINA W. MATHESON, Associate Professor of Medical Information and Director, William H. Welch Medical Library, The Johns Hopkins University, Baltimore, Maryland.

Before going to Johns Hopkins, Ms. Matheson was special consultant to the National Library of Medicine in Bethesda, Maryland, and assistant director of health information management studies at the Association of American Medical Colleges in Washington, D.C. She had previously been assistant research professor in the department of health care sciences and director of the Himmelfarb

Health Science Library at The George Washington University, where she introduced automated library operating systems. She has served as President and member of the Board of both the Medical Library Association and the Association of Academic Health Sciences Library Directors, and is a member of the National Library of Medicine's Board of Regents. Ms. Matheson's B.A. and M.L. degrees are in English and Library Science, from the University of Washington.

DAVID A. PENSAK, Corporate Advisor, Computing Technology, E.I. du Pont de Nemours & Co., Inc., Wilmington, Delaware.

Dr. Pensak has been at du Pont since 1974, where he manages the corporation's computer science research and development. His organization is chartered with identifying those areas of science and technology in which a two order of magnitude (or greater) increase in complexity or capacity of modelling would permit revolutionary advances in understanding (and the construction of the appropriate hardware and software to achieve these goals). He was a member of the American Chemical Society's Task Force on Large Scale Computing. His research includes structure-activity correlations, theory of catalysis, interactive graphics, programming language design and human engineering, systems, artificial intelligence, and parallel computer architectures. Dr. Pensak has a B.A. degree from Princeton University in chemistry and M.A. and Ph.D. degrees from Harvard University, also in chemistry.

ALLAN H. WEIS, Vice President, Data Systems Division, IBM Enterprise Systems, White Plains, New York.

Mr. Weis has worldwide responsibility for the strategy, development, and technical support of IBM's large system for numerically intensive computing.

His career at IBM, which began in 1961, has included assignments in research, development, new business areas, and information systems. Most recently, Mr. Weis was responsible for the computing and communication system at IBM's research laboratories, and for the direction of a number of advanced technology programs.

Mr. Weis majored in electrical engineering at the University of Kansas, and received his M.S. from the Massachusetts Institute of Technology on an Alfred P. Sloan Fellowship. He is a member of the Board of Directors of NYSERNET, and is on the Executive Committee of the NSFNET.

Bibliography and Selected Readings

Barbacci, Mario R., A. Nico Habermann, and Mary Shaw. The Software Engineering Institute: Bridging practice and potential. *IEEE Software,* volume 2, number 6, pp. 4–21, November 1985.

Bardon, Marcel. *A National Computing Environment for Academic Research* (Kent K. Curtis, Chair). NSF Working Group on Computers for Research. Washington, D.C.: National Science Foundation, July 1983.

Bikson, Tora K., Barbara A. Gutek, and Don A. Mankin. Implementing Computerized Procedures in Office Settings. Santa Monica, CA: The Rand Corporation. R-3077-NSF/IRIS. October, 1987.

Bird, Peter. Formation of the Rocky Mountains, western United States: A continuum computer model. *Science,* volume 239, pp. 1501–1507, 25 March 1988.

Bowler, Ken C., Alastair D. Bruce, Richard D. Kenway, G. Stuart Pawley, and David J. Wallace. Exploiting highly concurrent computers for physics. *Physics Today,* special issue: Computational physics, volume 40, number 10, pp. 40–48, October 1987.

Bowman, Carlos M., John A. Nosal, and Anne E. Rogers. Effect of new technology on information transfer in the 1990s. *Journal of Chemical Information and Computing Sciences,* volume 27, pp. 147–151, 1987.

Branscomb, Lewis M. Improving R&D productivity: The federal role. *Science,* volume 222, pp. 133–135, 14 October 1983.

Brooks, Frederick P., Jr. No silver bullet: Essence and accidents of software engineering. *IEEE Computer,* volume 20, number 4, pp. 10–19, April 1987.

Brown, Frank R., and Norman H. Christ. Parallel supercomputers for lattice gauge theory. *Science,* volume 239, pp. 1393–1400, 18 March 1988.

Bush, Vannevar. As we may think. *The Atlantic Monthly,* volume 176, number 1, pp. 101–108, July 1945.

Caming, H.W. William. Protection of personal data in the United States. *The Information Society,* volume 3, number 2, pp. 113–130, 1984.

Cotter, Holland. Birth of a network: A history of BITNET. *CUNY/University Computer Center Communications,* volume 14, number 1–2, pp. 1–10, January-February 1988.

Davis, Ruth M. Where will technology put the library of the 21st century? *Bulletin of the Medical Library Association,* volume 75, number 1, pp. 1–6, January 1987.

DeLisi, Charles. Computers in molecular biology: Current applications and emerging trends. *Science,* volume 240, pp. 47–52, April 1, 1988.

Denning, Peter J. Paradigms crossed. Editorial. *Communications of the ACM,* volume 30, number 10, p. 808, October 1987.

__. The science of computing: Electronic publishing. *American Scientist,* volume 74, pp. 582–585, 1986.

__. The science of computing: A new paradigm for science. *American Scientist,* volume 75, pp. 572–573, November-December 1987.

Drake, Miriam. From print to nonprint materials: Library information delivery systems. *EDUCOM Bulletin,* pp. 28–30, Spring 1988.

Eddy, William F. (chairman). *Computers in statistical research: Report of a workshop on the use of computers in statistical research.* Washington, D.C.: The Institute for Mathematical Statistics, 1986.

EDUCOM Networking and Telecommunications Task Force. *A national higher education network: Issues* *and opportunities.* NTTF Paper Number One. Princeton, NJ: EDUCOM, May 1987.

Ezzell, Carol. AIDS workers going for computer link. *Nature,* volume 332, p. 194, 17 March 1988.

Feigenbaum, Edward A. Autoknowledge: From file servers to knowledge servers. In: Salamon, R., B. Blum, and M. Jorgensen (eds.) *Medinfo 86,* pp. xliii–xlvi. Elsevier Science Publishers B.V. (North-Holland), 1986.

Fienberg, Stephen E., Margaret E. Martin, and Miron L. Straf (eds.) *Sharing research data.* Washington, D.C.: National Academy Press, 1985.

Foley, 1987. Manipulative interfaces.

Freeman, John. Data quality and the development of organizational social science: An editorial essay. *Administrative Science Quarterly,* volume 31, pp. 298–303, 1986.

Gillespie, Robert G. (with Deborah A. Dicaro). *Computing and higher education: An accidental revolution.* Washington, D.C.: National Science Foundation, 1981.

Green, Bert F. Adaptive testing by computer. In: R.B. Ekstrom (ed.), *Measurement, technology, and individuality in education.* New Directions for Testing and Measurement, number 17. San Francisco, Jossey-Bass, March 1983.

Haggin, Joseph. Process control no longer separate from simulation, design. *Chemical and Engineering News,* volume 62, number 14, pp. 7–16, April 2, 1984.

__. Workshop underscores wide use of computers in chemical engineering. *Chemical and Engineering News,* volume 65, number 45, pp. 27–29, November 9, 1987.

Hey, Anthony J. G., John H. Merlin, Martin W. Ricketts, Michael T. Vaughn, David C. Williams. Topological solutions in gauge theory and their computer graphic representation. *Science,* volume 240, pp. 1163–1168, 27 May 1988.

Hoffman, David. The computer-aided discovery of new embedded minimal surfaces. *The Mathematical Intelligencer,* volume 9, number 3, pp. 8–21, 1987.

Holland, William R., and James C. McWilliams. Computer modeling in physical oceanography from the global circulation to turbulence. *Physics Today,* special issue: Computational physics, volume 40, number 10, pp. 51–57, October 1987.

Holmstrom, Engin Inel. *Access to supercomputers.* Higher Education Panel Report Number 69. Washington, D.C.: American Council on Education, January 1986.

Hut, Piet, and Gerald Jay Sussman. Advanced computing for science. *Scientific American,* special issue: The next computer revolution, volume 257, pp. 144–153, October 1987.

IEEE Computer Society. *Proceedings: First international conference on supercomputing systems.* St. Petersburg, Florida, December 16–20, 1985. IEEE-CS order number 654. Washington, D.C.: IEEE Computer Society Press, 1985.

Kahn, Robert E. Networks for advanced computing. *Scientific American,* special issue: The next computer revolution, volume 257, pp. 136–143, October 1987.

Karplus, Martin. Molecular dynamics simulation of proteins. *Physics Today,* special issue: Computational physics, volume 40, number 10, pp. 68–72, October 1987.

Kerr, Richard A. Pluto's orbital motion looks chaotic. *Science,* volume 240, pp. 986–1987, 20 May 1988.

Kiesler, Sara B., and Lee S. Sproull (eds.) *Computing and change on campus.* Cambridge, England: Cambridge University Press, 1987.

Koening, 1987. Speech production and its impact on use of computers.

Lax, Peter D. *Report of the Panel on Large-scale Computing in Science and Engineering* (P.D. Lax, Chair). Washington, D.C.: National Science Foundation, December 26, 1982.

Leaf, Jesse J. Databases turn computers into science libraries. *Computers in Physics,* January/February 1988, volume 2, number 1, pp. 24–31.

Lederberg, Joshua. Digital communications and the conduct of science: The new literacy. *Proceedings of the IEEE,* volume 66, number 11, pp. 1314–1319, November 1978.

Lievrouw, Leah A. The communication network as interpretive environment: “Sense-making” among biomedical research scientists. Unpublished doctoral dissertation, University of Southern California, April 1986.

Liscouski, Joseph G. Note on barriers to information technology in research. Personal communication, January 23, 1987.

Londer, Randi. Access to supercomputers. *Mosaic,* volume 16, number 3, pp. 26–32, 1985.

Lunin, Lois F., and Peter B. Schipma (eds.) Perspectives on CD-ROM for information storage and retrieval. *Journal of the American Society for Information Science,* special issue, volume 39, number 1, January 1988.

Mace, Scott. University net thrives as 8,000-node system. User profile. *InfoWorld,* November 2, 1987, p. 24.

Malone, Thomas W., Joanne Yates, and Robert I. Benjamin. Electronic markets and electronic hierarchies. *Communications of the ACM,* volume 30, number 6, pp. 484–497, June 1987.

Manuel, Tom. What's holding back expert systems? Inside technology, special report. *Electronics,* pp. 59–63, August 7, 1986.

Marshall, Eliot. The scourge of computer viruses. *Science,* volume 240, pp. 133–134, April 8, 1988.

Mathis, Robert F. The last ten percent. *IEEE Transactions of Software Engineering,* volume SE-12, number 6, pp. 705–712, June 1986.

McCormick, Bruce H., Thomas A. DeFanti, and Maxine D. Brown. Visualization in scientific computing. (NSF Report) *Computer Graphics,* volume 21, number 6, pp. 1–14. New York, NY: Association for Computing Machinery, ACM SIGGRAPH, November 1987.

Metzger, Norman. A third kind of science. *Mosaic,* volume 16, number 3, pp. 2–7, 1985.

Morris, James H., Mahadev Satyanarayanan, Michael H. Connor, John H. Howard, David S. H. Rosenthal, and F. Donelson Smith. Andrew: A distributed personal computing environment. *Communications of the ACM,* volume 29, number 3, pp. 184–201, March 1986.

National Academy of Sciences, National Academy of Engineering, Institute of Medicine. Committee on Science, Engineering, and Public Policy. *Frontiers in science and technology: A selected outlook.* New York, NY: W.H. Freeman and Company, 1983.

__. Research briefing on computer architecture. 1984. Reprinted in: *New Pathways in Science and Technology.* New York, NY: Vintage Books, 1985.

National Academy of Sciences/National Research Council, Ad Hoc Committee on Resources for the Mathematical Sciences . *Renewing U.S. mathematics: critical resource for the future* (E. E. David, Chair). Washington, D.C., National Academy Press, 1984.

National Commission for Employment Policy. *Computers in the workplace: Selected issues.* Report No. 19. Washington, D.C.: National Commission for Employment Policy, March 1986.

National Library of Medicine. *Long range plan executive summary.* Report of the Board of Regents. Washington, D.C.: National Institutes of Health, January 1987.

__. Long range plan. Report of Panel 2: *Locating and gaining access to medical and scientific literature.* Washington, D.C.: National Institutes of Health, December 1986.

__. Long range plan. Report of Panel 4: *Medical informatics.* Washington, D.C.: National Institutes of Health, December 1986.

National Research Council, Committee on the Applications of Mathematics, Office of Mathematical Sciences, Commission on Physical Sciences, Mathematics, and Resources. *Computational modeling and mathematics applied to the physical sciences.* Washington, D.C.: National Academy Press, 1984.

National Research Council, Committee on Computer-Assisted Modeling, Board on Basic Biology, Commission on Life Sciences. *Computer-assisted modeling: Contributions of computational approaches to elucidating macromolecular structure and function.* Washington, D.C.: National Academy Press, 1987.

National Research Council, Committee on Data Needs, Numerical Data Advisory Board, Assembly of Mathematical and Physical Sciences. *National needs for critically evaluated physical and chemical data.* Washington, D.C.: National Academy of Sciences, 1978.

National Research Council, Computer Science and Technology Board, Commission on Physical Sciences, Mathematics, and Resources. *The national challenge in computer science and technology.* Washington, D.C.: National Academy Press, 1988.

National Research Council, National Research Network Review Committee, Computer Science and Technology Board, Commission on Physical Sciences, Mathematics, and Resources. *Toward a National Research Network.* Washington, D.C.: National Academy Press, 1988.

National Research Council, Panel on Education, Numerical Data Advisory Board, Commission on Physical Sciences, Mathematics, and Resources. *Improving the treatment of scientific and engineering data through education.* Washington, D.C.: National Academy Press, 1986.

National Research Council, Panel on Mathematical Sciences, Board on Mathematical Sciences, Commission on Physical sciences, Mathematics, and Resources. *Mathematical sciences: A unifying and dynamic resource* (Phillip A. Griffiths, Chair). Washington, D.C.: National Academy Press, 1986.

National Research Council, Physics Survey Committee, Board on Physics and Astronomy, Commission on Physical Sciences, Mathematics and Resources. *Physics through the 1990s: An overview.* Washington, D.C.: National Academy Press, 1986.

Norris, William C. Computer generations and technological cooperation. *The Bridge,* volume 16, number 4, pp. 19–24, Winter 1986.

NYSERNET. *Market and economic impact study: Lessons from four networks.* Second report. New York, NY: New York State Education and Research Network, Inc., April 5, 1988.

Office of Science and Technology Policy, Executive Office of the President . *A research and development strategy for high performance computing.* Committee on Computer Research and Applications, Federal Coordinating Council on Science, Engineering, and Technology (FCCSET). Washington, D.C.: Executive Office of the President, November 20, 1987.

__. *Report of the Federal Coordinating Council on Science, Engineering, and Technology Panel on Advanced Computer Research in the Federal Government.* Washington, D.C.: Executive Office of the President, June 1985.

__. *Research in very high performance computing: Policy recommendation and research require-*

ments statement. Washington, D.C.: Executive Office of the President, November 1985.

Peterson, Ivars. Packing it in: Fractals play an important role in image compression. *Science News,* volume 131, number 18, pp. 283–285, May 2, 1987.

__. Picture this: The sounds of speech lead to novel ways of representing complex data. *Science News,* volume 131, pp. 392–395, June 20, 1987.

__. Twists of space: An artist, a computer programmer and a mathematician work together to visualize exotic geometric forms. *Science News,* volume 132, pp. 264–266, October 24, 1987.

__. Reaching for the supercomputing moon. *Science News,* volume 133, number 11, pp. 172–173, March 12, 1988.

__. A digital matter of life and death. *Science News,* volume 133, number 11, pp. 170–171, March 12, 1988.

__. A computer at your fingertips. *Science News,* volume 133, number 22, p. 351, May 28, 1988.

__. Highways for information. *Science News,* volume 133, number 25, pp. 394–395, June 18, 1988.

Porter, Alan L., and Frederick A. Rossini. Current and future uses of the computer: Industrial R&D in the United States. *R&D Management,* volume 16, number 4, pp. 279–289, 1986.

Quarterman, John S., and J. C. Hoskins. Notable computer networks. *Communications of the ACM,* volume 29, number 10, pp. 932–971, October 1986.

Raveche, Harold J., Duncan H. Lawrie, and Alvin M. Despain. *A national computing initiative: The agenda for leadership.* Report of the Panel on Research Issues in Large-Scale Computational Science and Engineering. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1987.

Rheinboldt, Werner C. (chairman) *Future directions in computational mathematics, algorithms, and scientific software.* Report of the Panel on Future Directions in Computational Mathematics, Algorithms, and Scientific Software . Philadelphia, PA: Society for Industrial and Applied Mathematics, 1985.

Roberts, Leslie. Academy backs genome project. *Science,* volume 239, pp. 725–726, 12 February 1988.

Roode, David, Rob Liebschutz, Sunil Maulik, Terry Friedemann, David Benton, and David Kristofferson. New developments at BIONET. *Nucleic Acids Research,* volume 16, number 5, part a, pp. 1857–1859, 1988.

Rumble, John R. Why can't we access more numeric data via computers? In: Williams, M. E., and T. H. Hogan (eds.). *Proceedings of the National Online Meeting,* pp. 325–330. Medford, N.J.: Learned Information, Inc., 1984.

Simon, Herbert A. Whether software engineering needs to be artificially intelligent. *IEEE Transactions on Software Engineering,* volume SE-12, number 7, pp. 726–732, July 1986.

Steele, Guy. *COMMON LISP: The language.* Bedford, MA: Digital Press, 1984.

Straub, Galen K. Role for PCs in theoretical physics? In: Kahaner, David K. (ed.). *Science on small computers. SIAM News,* p. 12, March 1988.

Thomsen, Dietrick E. Seeking supernovas systematically. *Science News,* volume 132, number 10, pp. 156–157, September 5, 1987.

__. Big telescopes on a roll. *Science News,* volume 132, pp. 170–171, September 12, 1987.

Turner, Judith A. Science agency picks Michigan group to run computer network: $50 million “NSFNET” will link supercomputers, researchers. *The Chronicle of Higher Education,* pp. A1, A14, November 25, 1987.

__. Project linking different computers may alter the way researchers collaborate. *The Chronicle of Higher Education,* volume 34, number 27, pp. A13, A20, March 16, 1988.

__. Uncertainties strain computer networks for U.S. professors. *The Chronicle of Higher Education,* pp. A12–A13, April 6, 1988.

__. E-mail technology has boomed, but manners of its users fall short of perfection. *The Chronicle of Higher Education,* pp. A12–A16, April 13, 1988.

__. Plan for $5-million prototype of electronic research library announced. *The Chronicle of Higher Education,* p. A27, June 1, 1988.

U.S. Congress. General Accounting Office. *Satellite Data Archiving: U.S. and Foreign Activities and Plans for Environmental Information.* Report RCED-88–201. Washington, D.C.: United States General Accounting Office, September 1988.

U.S. Congress. General Accounting Office. *Space operations: NASA's use of information technology.* Report IMTEC-87–20. Washington, D.C.: United States General Accounting Office, April 1987.

U.S. Congress. House Committee on Science and Technology. Task Force on Science Policy. *The impact of the information age on science.* Science Policy Study—Hearings volume 10 (September 10–12, 1985). No. 103. Washington, D.C.: U.S. Government Printing Office, 1986a.

U.S. Congress. House Committee on Science and Technology. Task Force on Science Policy. *The impact of the information age on science.* Science Policy Study—Background Report No. 5. Washington, D.C.: U.S. Government Printing Office, September 1986b.

U.S. Congress. Office of Technology Assessment. *Information technology R&D: Critical trends and issues.* New York, N.Y.: Pergamon Press, 1985.

U.S. Congress. Office of Technology Assessment. *Science technology and the Constitution—Background paper,* OTA-BP-CIT-43. Washington, D.C.: U.S. Government Printing Office, September 1987.

University of Michigan. *EXPRES Newsletter.* Volume 1, Number 1. Available from: Violet Elder, CSMIL, C2420 Business Administration Building, 904 Monroe St., Ann Arbor, MI 48109 . June 1987.

Warlick, Charles H. Academic computing facilities and services in higher education—a survey. *EDUCOM Bulletin,* volume 21, number 3, pp. 2–7, Fall/Winter 1986.

Weiser, Mark. Source code. *IEEE Computer,* volume 20, number 11, pp. 66–73, November 1987.

Weiss, Rick. Organic origami: Scientists study the art of protein folding. *Science News,* volume 132, pp. 344–346, November 28, 1987.

Westbrook, J. H., H. Behrens, G. Dathe, and S. Iwata (eds.) *Materials data systems for engineering.* Proceedings of a CODATA workshop, September 22–27, 1985. Karlsruhe, FRG: Fach-informations-zentrum, 1986.

Westbrook, J. H., and L. R. McCreight (eds.) *Computerized aerospace materials data (Computerized materials property and design data for aerospace technology).* New York, N.Y.: American Institute of Aeronautics and Astronautics, 1987.

Westbrook, J.H., and J. R. Rumble, Jr. (eds.) *Computerized materials data systems.* Office of Standard Reference Data. Gaithersburg, MD: National Bureau of Standards, 1983.

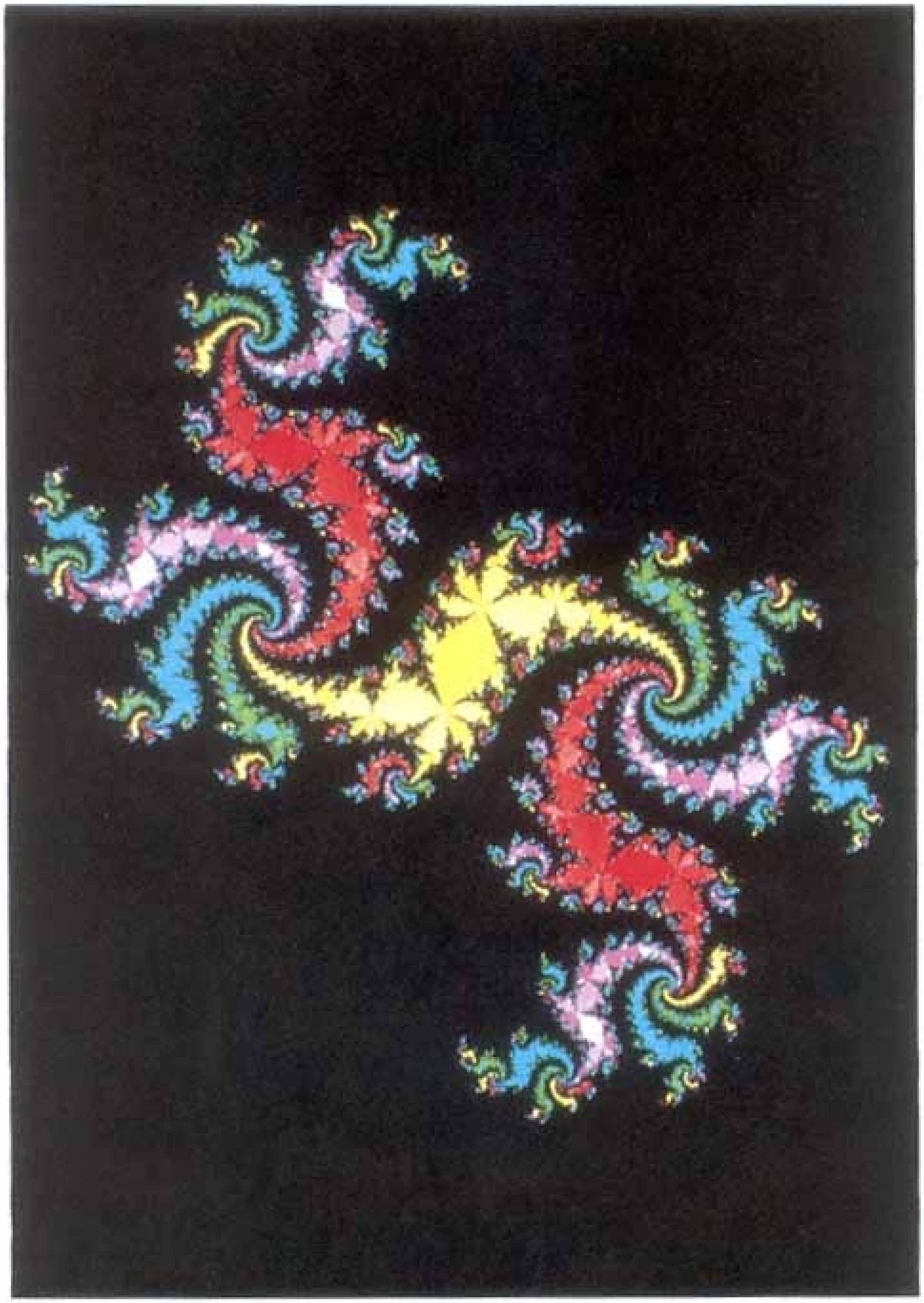
Winkler, Karl-Heinz A., Jay W. Chalmers, Stephen W. Hodson, Paul R. Woodward, and Norman J. Zabusky. A numerical laboratory. *Physics Today* (Special Issue: Computational Physics), volume 40, number 10, pp. 28–37, October 1987.

Wooster, Harold. Historical note: Shining palaces, shifting sands: National information systems. *Journal of the American Society for Information Science,* volume 38, number 5, pp. 321–335, September 1987.

Wright, Robert. Virtual reality. The information age. *The Sciences,* volume 27, number 6, pp. 8–10, November/December 1987.

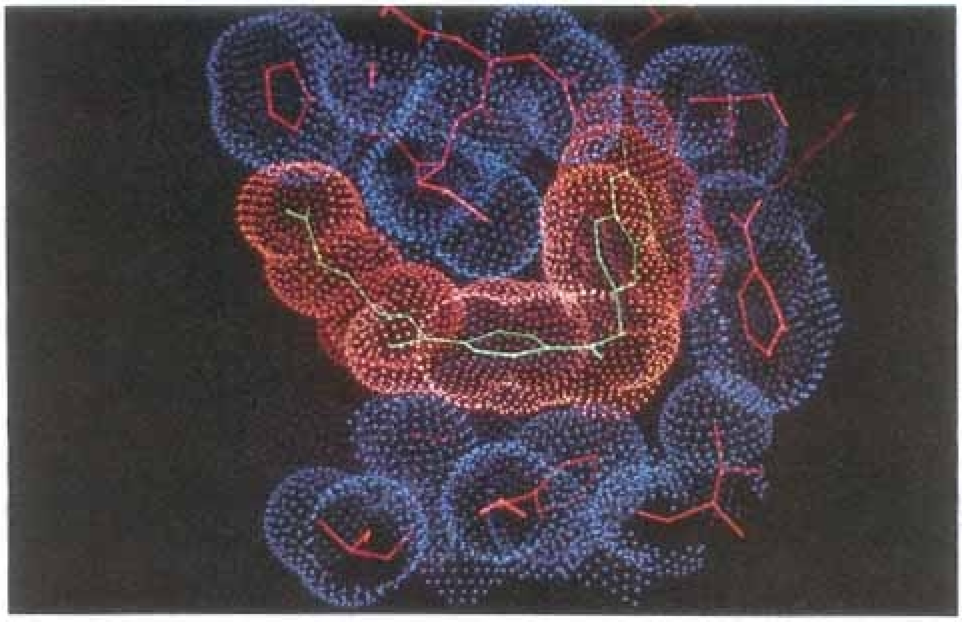
Zurer, Pamela S. Computers gaining firm hold in chemical labs. Special Report. *Chemical and Engineering News,* pp. 21–46, August 19, 1985.

Index



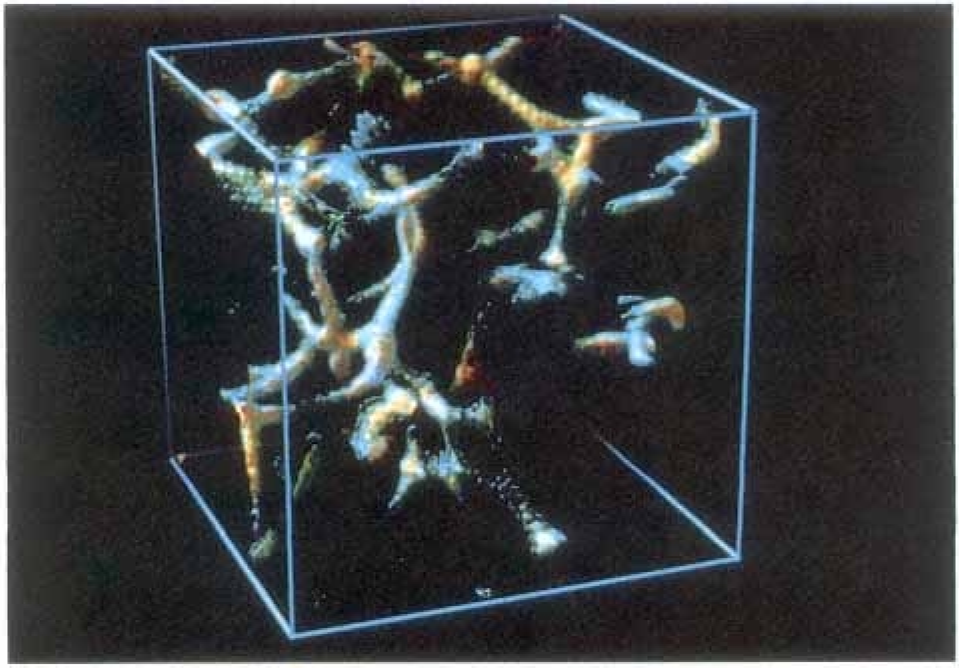
A "fractal dragon" generated by the IBM Fellow Benoit B. Mandelbrot, the originator of fractal geometry. This is an example of the "Julia set," which in turn is an example of a "speller set" of a dynamic system. This may seem to be an extremely complicated shape, yet it has a very simple equation based on the formula From the front cover of the book The Fractal Geometry of Nature by Benoit B. Mandelbrot, 1982, W. H. Freeman and Company.

The "fractal planetrise" by IBM scientist Richard F. Voss. In spite of its startling realism, every element in this picture is artificially generated. The striking resemblance between some fractal images and familiar landscapes illustrates the fact that fractals describe aspects of nature that have formerly eluded mathematical description. From the back cover of the book The Fractal Geometry of Nature by Benoit B. Mandelbrot, 1982, W. H. Freeman and Company. Courtesy of the IBM Corporation.

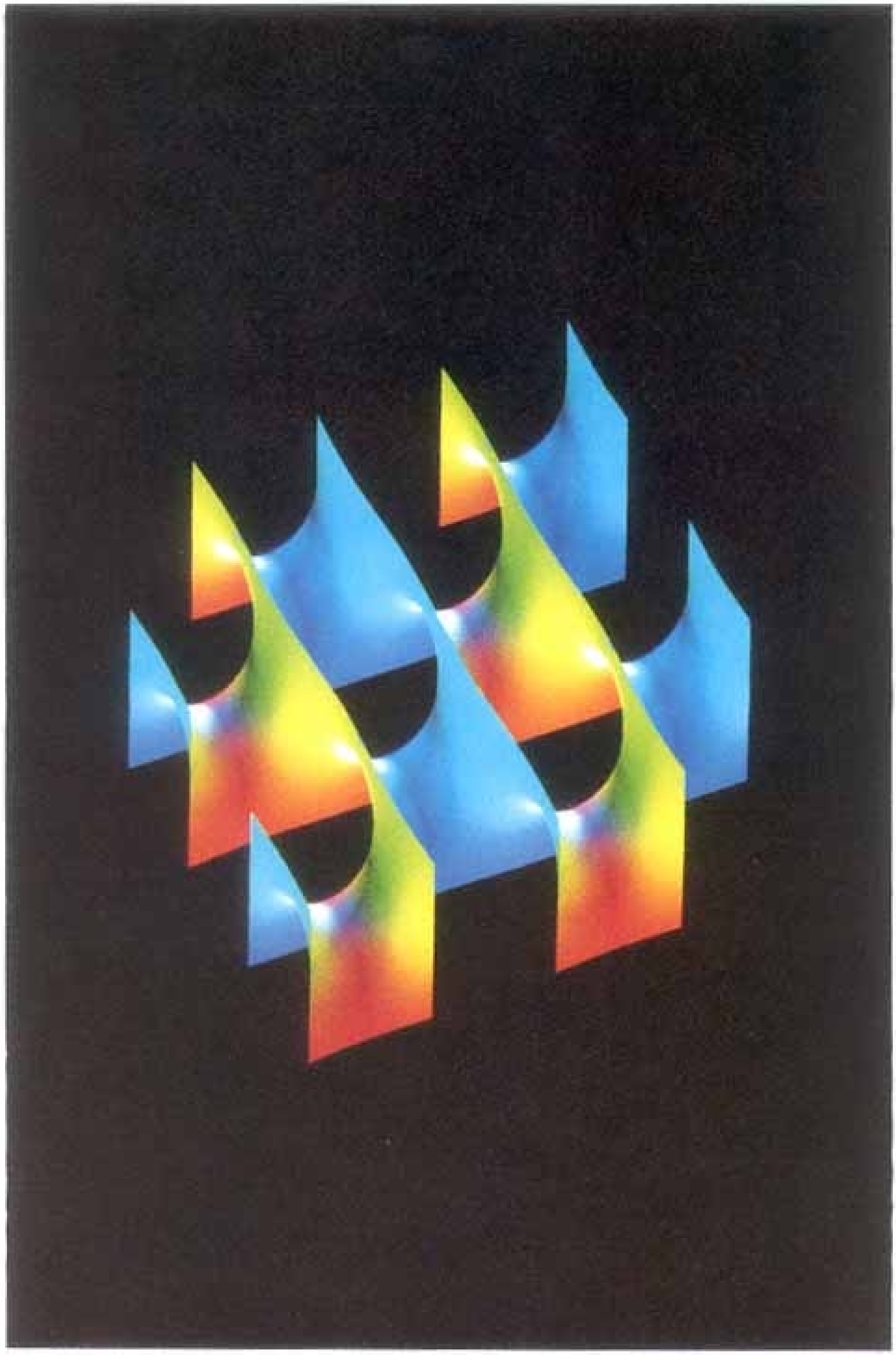


Van der Waals surface of dihydrofolate reductase and methotextrate. Red is oxygen, blue is nitrogen, green is carbon, and yellow is phosphorus. Produced by the Computer Graphics Laboratory, University of California, San Francisco. © Regents, University of California.

This realistic-looking landscape is actually only a few hundred atoms square and approximately 10 atoms high. Generated by IBM scientist Richard F. Voss, who used a computer to add color, lighting, and shading to a scanning tunneling microscope image of thermally roughened silicon. In addition to the esthetic benefits of such pictures, scientists and engineers can use such images to understand the properties of critical materials such as silicon. Color coding, for example, can emphasize and delineate specific atomic areas of interest, such as atomic trace impurities or surface defects. Courtesy of the IBM Corporation.



Frame from a computer-animated film depicting clustering of matter in the early evolution of the universe. The film itself was produced by a collaborative effort between an astrophysicist and a Hollywood special-effects graphics firm. © Joan M. Centrella, Drexel University.



Scherk's doubly periodic minimal surface, which has recently been proposed as a model microstructure for grain boundaries in copolymers. Nature, August 8, 1988.
© J. T. Hoffman.