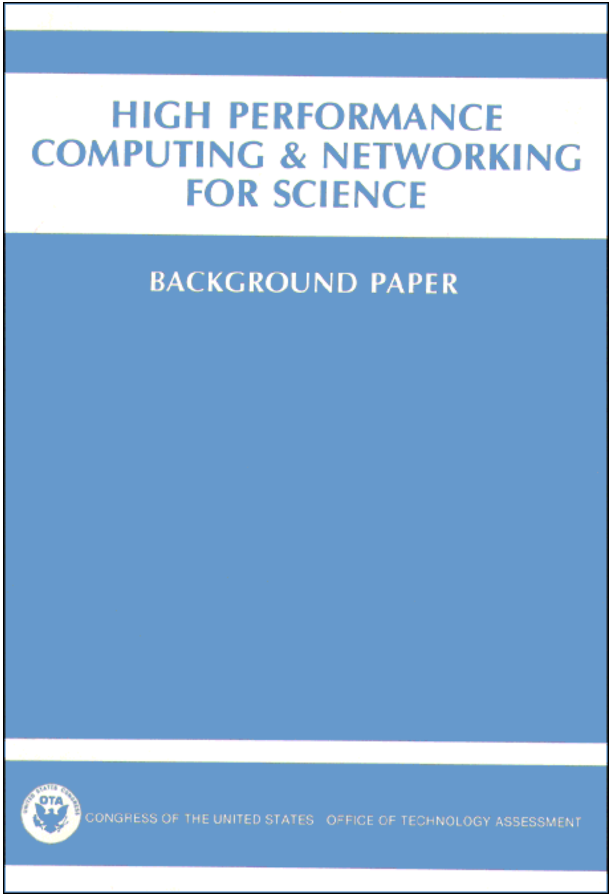
High Performance Computing and Networking for Science

September 1989



HIGH PERFORMANCE COMPUTING & NETWORKING FOR SCIENCE

BACKGROUND PAPER

CONGRESS OF THE UNITED STATES OFFICE OF TECHNOLOGY ASSESSMENT

Recommended Citation:

U.S. Congress, Office of ‘Technology Assessment, *High Performance Computing and Networkig for Science-Background Paper, OTA-BP-CIT-59* (Washington, DC: U.S. Government Printing Office, September 1989).

Library of Congress Catalog Card Number **89-600758**

For sale by the Superintendent of Documents
U.S. Government Printing Office, Washington, DC 20402-9325
(Order form can be found in the back of this report.)

Foreword

Information technology is fundamental to today's research and development: high performance computers for solving complex problems; high-speed data communication networks for exchanging scientific and engineering information; very large electronic archives for storing scientific and technical data; and new display technologies for visualizing the results of analyses.

This background paper explores key issues concerning the Federal role in supporting national high performance computing facilities and in developing a national research and education network. It is the first publication from our assessment, ***Information Technology and Research,*** which was requested by the House Committee on Science and Technology and the Senate Committee on Commerce, Science, and Transportation.

OTA gratefully acknowledges the contributions of the many experts, within and outside the government, who served as panelists, workshop participants, contractors, reviewers, detailees, and advisers for this document. As with all OTA reports, however, the content is solely the responsibility of OTA and does not necessarily constitute the consensus or endorsement of the advisory panel, workshop participants, or the Technology Assessment Board.

John H Gibbons

JOHN H. GIBBONS
Director

High Performance Computing and Networking for Science Advisory Panel

John P. (Pat) Crecine, *Chairman*
President, Georgia Institute of Technology

Charles Bender
Director
Ohio Supercomputer Center

Charles DeLisi
Chairman
Department of Biomathematical Science
Mount Sinai School of Medicine

Deborah L. Estrin
Assistant Professor
Computer Science Department
University of Southern California

Robert Ewald
Vice President, Software
Cray Research, Inc.

Kenneth flamm
Senior Fellow
The Brookings Institution

Malcolm Getz
Associate Provost
Information Services & Technology
Wnderbilt University

Ira Goldstein
Vice president Research
Open Software Foundation

Robert E. Kraut
Manager
Interpersonal Communications Group
Bell Communications Research

Lawrence Landweber
Chairman
Computer Science Department
University of Wisconsin-Madison

Carl Ledbetter
President/CEO
ETA Systems

Donald Marsh
Vice President, Technology
Contel Corp.

Michael J, McGill
Vice President
Technical Assessment & Development
OCLC, Computer Library Center, Inc.

Kenneth W. Neves
Manager
Research & Development Program
Boeing Computer Services

Bernard O'Lear
Manager of Systems
National Center for Atmospheric Research

William Poduska
Chairman of the Board
Stellar Computer, Inc.

Elaine Rich
Director
Artificial Intelligence Lab
Microelectronics and Computer Technology Corp.

Sharon J. Rogers
University Librarian
Gelman Library
The George Washington University

William Schrader
President
NYSERNET

Kenneth Toy
Post-Graduate Research Geophysicist
Scripps Institution of Oceanography

Keith Uncapher
Vice President
Corporation for the National Research Initiatives

Al Weis
Vice President
Engineering & Scientific Computer
Data Systems Division
IBM Corp.

NOTE: OTA is grateful for the valuable assistance and thoughtful critiques provided by the advisory panel. The views expressed in this OTA background paper, however, are the sole responsibility of the Office of Technology Assessment.

OTA Project Staff-High Performance Computing

John Andelin, *Assistant Director, OTA*
Science, Information, and Natural Resources Division

James W. Curhin, *Program Manager*
Communication and Information Technologies Program

Fred W. Weingarten, *Project Director*
Charles N. Brownstein, *Senior Analyst*[1]
Lisa Heinz, *Analyst*
Elizabeth I. Miller, *Research Assistant*

Administrative Staff
Elizabeth Emanuel, *Administrative Assistant*
Karolyn Swauger, *Secretary*
Jo Anne Price, *Secretary*

Other Contributors

Bill Bartelone
Legislative/Federal Program Manager
Cray Research, Inc.

Mervin Jones
Program Analyst
Defense Automation Resources Information Center

Timothy Lynagh
Supervisory Data and Program Analyst
Government Services Administration

---

[1] Detailee from NSF

List of Reviewers

Janice Abraham
Executive Director
Cornell Theory Center
Cornell University

Lee R. Alley
Assistant Vice President for
Information Resources Management
Arizona State University

James Almond
Director
Center for High Performance
Computing
Balcones Research Center

Julius Archibald
Department Chairman
Department of Computer Science
State University of New York
College of Plattsburgh

J. Gary Augustson
Executive Director
Computer and Information
Systems
Pennsylvania State University

Philip Austin
President
Colorado State University

Steven C. Beering
President
Purdue University

Jerry Berkman
Fortran Specialist
Central Computing Services
University of California at Berkeley

Kathleen Bernard
Director for Science Policy
and Technology Programs
Cray Research, Inc.

Justin L. Bloom
President
Technology International, Inc.

Charles N. Brownstein
Executive Officer
Computing & Information Science &
Engineering
National Science Foundation

Eloise E. Clark
Vice President, Academic
Affairs
Bowling Green University

Paul Coleman
Professor
Institute of Geophysics and
Space Physics
University of California

Michael R. Dingerson
Associate Vice Chancellor for
Research and Dean of the
Graduate School
University of Mississippi

Christopher Eoyang
Director
Institute for Supercomputing
Research

David Farber
Professor
Computer & Information
Science Department
University of Pennsylvania

Sidney Fernbach
Independent Consultant

Susan Fratkin
Director, Special Programs
NASULGC

Doug Gale
Director of Computer
Research Center
Office of the Chancellor
University of Nebraska-Lincoln

Robert Gillespie
President
Gillespie, Folkner
& Associates, Inc.

Eiichi Goto
Director, Computer Center
University of Tokyo

C.K. Gunsalus
Assistant Vice Chancellor
for Research
University of Illinois
at Urbana-Champaign

Judson M. Harper
Vice President of Research
Colorado State University

Gene Hemp
Senior Associate V.P. for
Academic Affairs
University of Florida

Nobuaki Ieda
Senior Vice President
NTT America, Inc.

Hiroshi Inose
Director General
National Center for Science
Information System

Heidi James
Executive Secretary
United States Activities
Board
IEEE

Russell C. Jones
University Research Professor
University of Delaware

Brian Kahin, Esq.
Research Affiliate on
Communications Policy
Massachusetts Institute of Technology

Robert Kahn
President
Corporation of National
Research Initiatives

Hisao Kanai
Executive Vice President
NEC Corporation

Hiroshi Kashiwagi
Deputy Director-General
Electrotechnical Laboratory

Lauren Kelly
Department of Commerce

Thomas Keyes
Professor of Chemistry
Boston University

Continued on next page

Doyle Knight
President
John von Neumann National
Supercomputer Center
Consortium for Scientific
Computing

Mike Levine
Co-director of the Pittsburgh
Supercomputing Center
Carnegie Mellon University

George E. Lindamood
Program Director
Industry Service
Gartner Group, Inc.

M. Stuart Lynn
Vice President for
Information Technologies
Cornell University

Ikuo Makino
Director
Electrical Machinery & Consumer
Electronics Division
Minisry of International
Trade and Industry

Richard Mandelbaum
Vice Provost for Computing
University of Rochester

Martin Massengale
Chancellor
University of N'ebraska-Lincoln

Gerald W. May
President
University of New Mexico

Yoshiro Miki
Director, Policy Research Division
Science and Technology Policy Bureau
Science and Technology Agency

Takeo Miura
Senior Executive Managing Director
Hitachi. Ltd.

J. Gerald Morgan
Dean of Engineering
New Mexico State University

V. Rama Murthy
Vice Provost for Academic
Affairs
University of Minnesota

Shoichi Ninomoiya
Executive Director
Fujitsu Limited

Bernard O'Lear
Manager of Systems
National Center for Atmospheric
Research

Ronald Orcutt
Executive Director
Project Athena
MIT

Tad Pinkerton
Director
Office of Information
Technology
University of Wisconsin-Madison

Harold J. Raveche
President
Stevens Institute of Technology

Ann Redelf
Manager of Information
Services
Cornell Theory Center
Cornell University

Glenn Ricart
Director
Computer Science Center
University of Maryland in
College Park

Ira Richer
Program Manager
DARPA/ISTO

John Riganati
Director of Systems Research
Supercomputer Research Center
Institute for Defense Analyses

Mike Roberts
Vice President
EDUCOM

David Roselle
President
University of Kentucky

Nora Sabelli
National Center for Supercomputing
Applications
University of Illinois at
Urbana-Champaign

Steven Sample
President
SUNY, Buffalo

John Sell
President
Minnesota Supercomputer Center

Hiroshi Shima
Deputy Director-General for
Technology Affairs
Agency of Industrial Science and
Technology , MITI

Yoshio Shimamoto
Senior Scientist (Retired)
Applied Mathematics Department
Brookhaven National Laboratory

Charles Sorber
Dean, School of Engineering
University of Pittsburgh

Harvey Stone
Special Assistant to the
President
University of Delaware

Dan Sulzbach
Manager, User Services
San Diego Supercomputer Center

Tatsuo Tanaka
Executive Director
Interoperability Technology
Association for Information Processing,
Japan

Ray Toland
president
Alabama Supercomputing Network
Authority

Kenneth Tolo
Vice Provost
University of Texas
at Austin

Kenneth Toy
Post-Graduate Research
Geophysicist
Scripps Institution of
Oceanography

August B. Tumbull. III
Provost & Vice President,
Academic Affairs
Florida State University

Continued on next page

List of Reviewers (continued)

Gerald Turner
Chancellor
University of Mississippi

Douglas Van Houweling
Vice Provost for Information
& Technology
University of Michigan

Anthony Villasenor
Program Manager
ScienCE Networks
Office of Space Science and
Applications
National Aeronautics and
Space Administration

Hugh Walsh
Data Systems Division
IBM

Richard West
Assistant Vice President,
IS&AS
University of California

Steve Wolff
Program Director for Networking
Computing & Information Science &
Engineering
National Science Foundation

James Woodward
Chancellor
University of North Carolina
at Charlotte

Akihiro Yoshikawa
Research Director
University of California, Berkeley
BRIE/IIS

NOTE: OTA is grateful for the valuable assistance and thoughtful critiques provided by the advisory panel, The views expressed in this OTA background paper, however, are the sole responsibility of the Office of Technology Assessment,

Contents

Figures

Tables

Chapter 1

Introduction and Overview Observations

The Office of Technology Assessment is conducting an assessment of the effects of new information technologies—including high performance computing, data networking, and mass data archiving-on research and development. This background paper offers a midcourse view of the issues and discusses their implications for current discussions about Federal supercomputer initiatives and legislative initiatives concerning a national data communication network.

Our observations to date emphasize the critical importance of advanced information technology to research and development in the United States, the interconnection of these technologies into a national system (and, as a result, the tighter coupling of policy choices regarding them), and the need for immediate and coordinated Federal action to bring into being an advanced information technology infrastructure to support U.S. research, engineering, and education.

RESEARCH AND INFORMATION TECHNOLOGY—A FUTURE SCENARIO

Within the next decade, the desks and laboratory benches of most scientists and engineers will be entry points to a complex electronic web of information technologies, resources and information services, connected together by high-speed data communication networks (see figure 1-1). These technologies will be critical to pursuing research in most fields. Through powerful workstation computers on their desks, researchers will access a wide variety of resources, such as:

- an interconnected assortment of local campus, State and regional, national, and even international data communication networks that link users worldwide;
- specialized and general-purpose computers including supercomputers, minisupercomputers, mainframes, and a wide variety of special architectures tailored to specific applications;
- collections of application programs and software tools to help users find, modify, or develop programs to support their research;
- archival storage systems that contain specialized research databases;
- experimental apparatus-such as telescopes, environmental monitoring devices, seismographs, and so on---designed to be set-up and operated remotely;
- services that support scientific communication, including electronic mail, computer conferencing systems, bulletin boards, and electronic journals;
- a “digital library” containing reference material, books, journals, pictures, sound recordings, films, software, and other types of information in electronic form; and
- specialized output facilities for displaying the results of experiments or calculations in more readily understandable and visualizable ways.

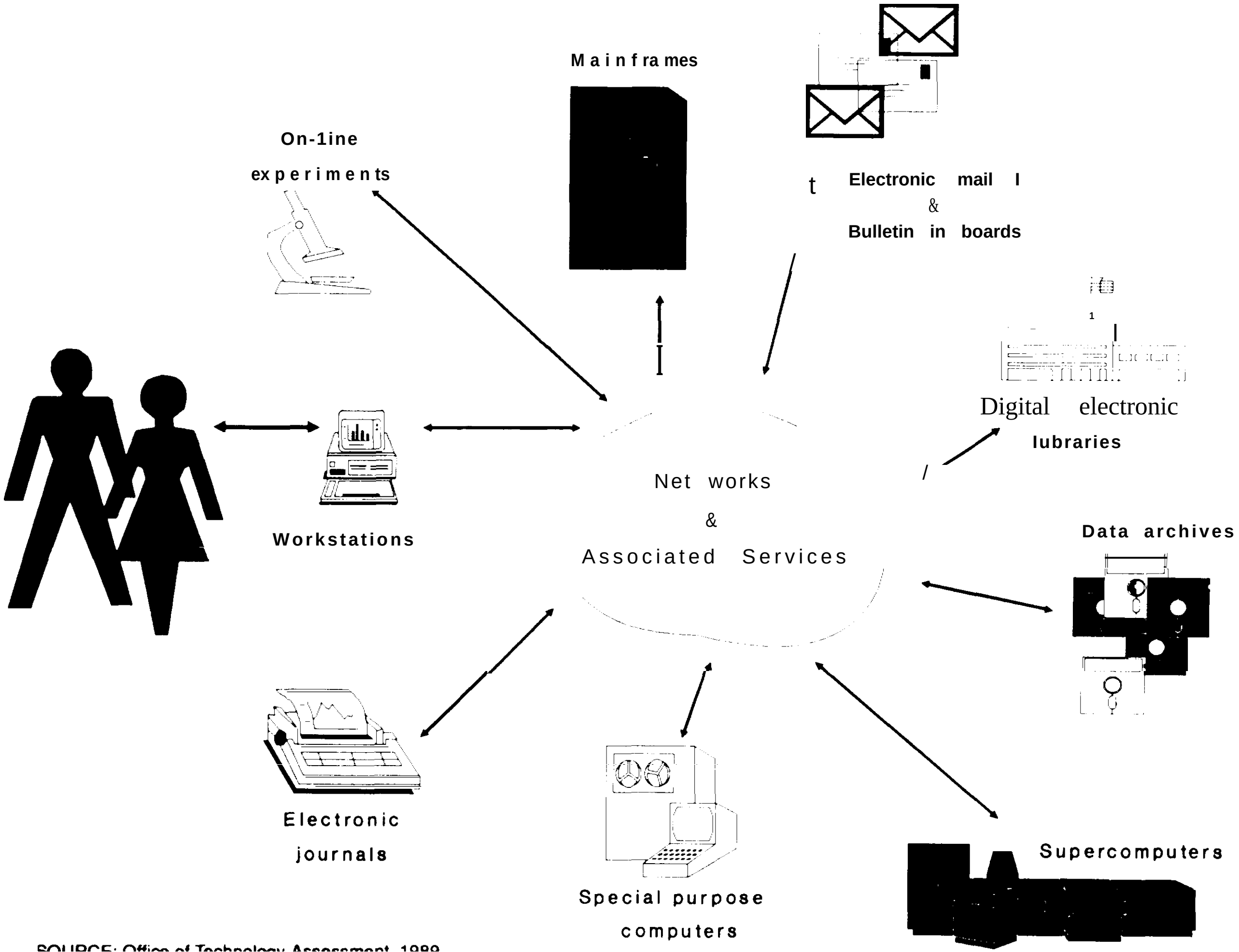
Many of these resources are already used in some form by some scientists. Thus, the scenario that is drawn is a straightforward extension of current usage. Its importance for the scientific community and for government policy stems from three trends: 1) the rapidly and continually increasing capability of the technologies; 2) the integration of these technologies into what we will refer to as an “information infrastructure”; and 3) the diffusion of information technology into the work of most scientific disciplines.

Few scientists would use all the resources and facilities listed, at least on a daily basis; and the particular choice of resources eventually made available on the network will depend on how the tastes and needs of research users evolve. However, the basic form, high-speed data networks connecting user workstations with a worldwide assortment of information technologies and services, is becoming a crucial foundation for scientific research in most disciplines.

MAJOR ISSUES AND PROBLEMS

Developing this system to its full potential will require considerable thought and effort on the part of government at all levels, industry, research institutions, and the scientific community, itself, It will present policy makers with some difficult questions and decisions.

Figure l-l—An Information Infrastructure for Research



SOURCE: Office of Technology Assessment, 1989.

Scientific applications are very demanding on technological capability. A substantial R&D component will need to accompany programs intended to advance R&D use of information technology. To realize the potential benefits of this new infrastructure, research users need advances in such areas as:

- more powerful computer designs;
- more powerful and efficient computational techniques and software; overly high-speed switched data communications;
- improved technologies for visualizing data results and interacting with computers; and
- new methods for storing and accessing information from very large data archives.

An important characteristic of this system is that different parts of it will be funded and operated by different entities and made available to users in different ways. For example, databases could be operated by government agencies, professional societies, non-profit journals, or commercial firms. Computer facilities could similarly be operated by government, industry, or universities. The network, itself, already is an assemblage of pieces funded or operated by various agencies in the Federal Government; by States and regional authorities; and by local agencies, firms and educational institutions. Keeping these components interconnected technologically and allowing users to move smoothly among the resources they need will present difficult management and policy problems.

Furthermore, the system will require significant capital investment to build and maintain, as well as specialized technical expertise to manage. How the various components are to be funded, how costs are to be allocated, and how the key components such as the network will be managed over the long term will be important questions.

Since this system as envisioned would be so widespread and fundamental to the process of research, access to it would be crucial to participation in science. Questions of access and participation are crucial to planning, management, and policymaking for the network and for many of the services attached to it.

Changes in information law brought about by the electronic revolution will create problems and conflicts for the scientific community and may influence how and by whom these technologies are used. The resolution of broader information issues such as security and privacy, intellectual property protection, access controls on sensitive information, and government dissemination practices could affect whether and how information technologies will be used by researchers and who may use them.

Finally, to the extent that, over the long run, modem information technology becomes so fundamental to the research process, it will transform the very nature of that process and the institutions—libraries, laboratories, universities, and so on—that serve it. These basic changes in science would affect government both in the operation of its own laboratories and in its broader relationship as a supporter and consumer of research. Conflicts may also arise to the extent that government becomes centrally involved, both through funding and through management with the traditionally independent and uncontrolled communication channels of science.

NATIONAL IMPORTANCE—THE NEED FOR ACTION

Over the last 5 years, Congress has become increasingly concerned about information technology and research. The National Science Foundation (NSF) has been authorized to establish supercomputer centers and a science network. Bills (S 1067 HR 3131) are being considered in the Congress to authorize a major effort to plan and develop a national research and education network and to stimulate information technology use in science and education. Interest in the role information technology could play in research and education has stemmed, first, from the government's major role as a funder, user, and participant in research and, secondly, from concern for ensuring the strength and competitiveness of the U.S. economy.

Observation 1: The Federal Government needs to establish its commitment to the advanced information technology infrastructure necessary for furthering U.S. science and education. This need sterns directly from the importance of science and technology to economic growth, the importance of information technology to research and development, and the critical timing for certain policy decisions.

Economic Importance

A strong national effort in science and technology is critical to the long-term economic competitiveness, national security, and social well-being of the United States. That, in the modem international economy, technological innovation is concomitant with social and economic growth is a basic assumption held in most political and economic systems in the world these days; and we will take it here as a basic premise. It has been a basic finding in many OTA studies.[1] (This observation is not to suggest that technology is a panacea for all social problems, nor that serious policy problems are not often raised by its use.) Benefits from of this infrastructure are expected to flow into the economy in three ways:

First, the information technology industry can benefit directly. Scientific use has always been a

---

[1] For example, U.S. Congress, Office of Technology Assessment, *Technology and the American Economic Transition*, OTA-TET-283 (Washington, DC: U.S. Government Printing Office, May 1988) and *Information Technology R&D Critical Trends and Issues*, OTA-CIT-268 (Washington, DC: U.S. Government Printing Office, February 1985),

major source of innovation in computers and communications technology. Packet-switched data communication, now a widely used commercial offering, was first developed by the Defense Advanced Research Projects Agency (DARPA) to support its research community. Department of Energy (DOE) national laboratories have, for many years, made contributions to supercomputer hardware and software. New initiatives to develop higher speed computers and a national science network could similarly feed new concepts back to the computer and communications industry as well as to providers of information services.

Secondly, by improving the tools and methodologies for R&D, the infrastructure will impact the research process in many critical high technology industries, such as pharmaceuticals, airframes, chemicals, consumer electronics, and many others. Innovation and, hence, international competitiveness in these key R&D-intensive sectors can be improved.

The economy as a whole stands to benefit from increased technological capabilities of information systems and improved understanding of how to use them. A National Research and Education Network could be the precursor to a much broader high capacity network serving the United States, and many research applications developed for high performance computers result in techniques much more broadly applicable to commercial firms.

Scientific Importance

Research and development is, inherently, an information activity. Researchers generate, organize, and interpret information, build models, communicate, and archive results, Not surprisingly, then, they are now dependent on information technology to assist them in these tasks. Many major studies by many scientific and policy organizations over the years-as far back as the President's Science Advisory Committee (PSAC) in the middle 1960s, and as recently as a report by COSEPUP of the National Research Council published in 1988[2]—have noted these trends and analyzed the implications for science support. The key points are as follows:

- Scientific and technical information is increasingly being generated, stored and distributed in electronic form;
- Computer-based communications and data handling are becoming essential for accessing, manipulating, analyzing, and communicating data and research results; and,
- In many computationally intensive R&D areas, from climate research to groundwater modeling to airframe design, major advances will depend upon pushing the state of the art in high performance computing, very large databases, visualization, and other related information technologies. Some of these applications have been labeled “Grand Challenges.” These projects hold promise of great social benefit, such as designing new vaccines and drugs, understanding global warming, or modeling the world economy. However, for that promise to be realized in those fields, researchers require major advances in available computational power.
- Many proposed and ongoing “big science” projects, from particle accelerators and large array radio telescopes to the NASA EOS satellite project, will create vast streams of new data that must be captured, analyzed, archived, and made available to the research community. These new demands could well overtax the capability of currently available resources.

Timing

Government decisions being made now and in the near future will shape the long-term utility and effectiveness of the information technology infrastructure for science. For example:

- NSF is renewing its multi-year commitments to all or most of the existing National Supercomputing Centers.
- Executive agencies, under the informal auspices of the Federal Research Internet Coordinating Committee (FRICC), are developing a national “backbone” network for science. Decisions made now will have long term influence on the nature of the network, its technical characteristics, its cost, its management, serv-

---

[2] Panel on Information Technology and the Conduct of Research, Committee on Science, Engineering, and Public Policy, *Information Technology and the Conduct of Research ” The User's* View (Washington, DC: National Academy Press, 1989).

ices available on it, access, and the information policies that will govern its use.

- The basic communications industry is in flux, as are the policies and rules by which government regulates it.
- Congress and the Executive Branch are currently considering, and in some cases have started, several new major scientific projects, including a space station, the Earth Orbiting System, the Hubble space telescope, the superconducting supercollider, human genome mapping, and so on. Technologies and policies are needed to deal with these "firehoses of data. " In addition, upgrading the information infrastructure could open these projects and data streams to broad access by the research community.

Observation 2: Federal policy in this area needs to be more broadly based than has been traditional with Federal science efforts. Plsnning, building, and managing the information technology infrastructure requires cutting across agency programs and the discipline and mission-oriented approach of science support. In addition, many parties outside the research establishment will have important roles to play and stakes in the outcome of the effort.

The key information technologies-high performance computing centers, data communication networks, large data archives, along with a wide range of supporting software-are used in all research disciplines and support several different agency missions. In many cases, economies of scale and scope dictate that some of these technologies (e.g., supercomputers) be treated as common resources. Some, such as communication networks, are most efficiently used if shared or interconnected in some way.

There are additional scientific reasons to treat information resources as a broadly used infrastructure: fostering communication among scientists between disciplines, sharing resources and techniques, and expanding access to databases and software, for instance. However, there are very few models from the history of Federal science support for creating and maintaining infrastructure-like resources for science and technology across agency and disciplinary boundaries. Furthermore, since the networks, computer systems, databases, and so on interconnect and users must move smoothly among them, the system requires a high degree of coordination rather than being treated as simply a conglomeration of independent facilities.

However, if information technology resources for science are treated as infrastructure, a major policy issue is one of boundaries. Who is it to serve; who are its beneficiaries? Who should participate in designing it, building and operating it, providing services over it, and using it? The answers to these questions will also indicate to Congress who should be part of the policymaking and planning process; they will govern the long term scale, scope, and the technological characteristics of the infrastructure itself; and they will affect the patterns of support for the facilities. Potentially interested parties include the following:

Users

Potential users might include academic and industrial researchers, teachers, graduate, undergraduate, and high school students, as well as others such as the press or public interest groups who need access to and make use of scientific information. Institutions, such as universities and colleges, libraries, and schools also have user interests. Furthermore, foreign scientists working as part of international research teams or in firms that operate internationally will wish access to the U.S. system, which, in turn, will need to be connected with other nation's research infrastructures.

Collaborators

Another group of interested parties include State and local governments and parts of the information industry. We have identified them with the term "collaborators" because they will be participating in funding, building, and operating the infrastructure. States are establishing State supercomputer centers and supporting local and regional networking, some computer companies participate in the NSF National Supercomputer Centers, and some telecommunication firms are involved in parts of the science network.

Service *Providers*

Finally, to the extent that the infrastructure serves as a basic tool for most of the research and development community, information service providers will require access to make their products available to scientific users. The service providers may include government agencies (which provide access to government scientific databases, for example), libraries and library utilities, journal and text-book publishers, professional societies, and private software and database providers.

Observation 3: Several information policy issues will be raised in managing and using the network. Depending on how they are resolved, they could sharply restrict the utility and scope of network **use** *in the scientific community.*

Security and privacy have already become of major concern and will pose a problem. In general, users will want the network and the services on it to be as open as possible; however, they will also want the networks and services to be as robust and dependable as possible-free free deliberate or accidental disruption. Furthermore, different resources will require different levels of security. Some bulletin boards and electronic mail services may want to be as open and public as possible; others may require a high level of privacy. Some databases may be unique and vital resources that will need a very high level of protection, others may not be so critical. Maintaining an open, easily accessible network while protecting privacy and valuable resources will require careful balancing of legal and technological controls.

Intellectual property protection in an electronic environment may pose difficult problems, Providers will be concerned that electronic databases, software, and even electronic formats of printed journals and other writings will not be adequately protected. In some cases, the product, itself, may not be well protected under existing law. In other cases electronic formats coupled with a communications network erode the ability to control restrictions on copying and disseminating.

Access controls may be called for on material that is deemed to be sensitive (although unclassified) for reasons of national security or economic competitiveness. Yet, the networks will be accessible worldwide and the ability to identify and control users may be limited.

The above observations have been broad, looking at the overall collection of information technology resources for science as an integrated system and at the questions raised by it. The remaining portion of this paper will deal specifically with high performance computers and networking.

Chapter 2

High Performance Computers

An important set of issues has been raised during the last 5 years around the topic of high performance computing (H-PC). These issues stem from a growing concern in both the executive branch and in Congress that U.S. science is impeded significantly by lack of access to HPC[1] and by concerns over the competitiveness implications of new foreign technology initiatives, such as the Japanese "Fifth Generation Project." In response to these concerns, policies have been developed and promoted with three goals in mind.

1. To advance vital research applications currently hampered by lack of access to very high speed computers.
2. To accelerate the development of new HPC technology, providing enhanced tools for research and stimulating the competitiveness of the U.S. computer industry.
3. To improve software tools and techniques for using HPC, thereby enhancing their contribution to general U.S. economic competitiveness.

In 1984, the National Science Foundation (NSF) initiated a group of programs intended to improve the availability and use of high performance computers in scientific research. As the centerpiece of its initiative, after an initial phase of buying and distributing time at existing supercomputer centers, NSF established five National Supercomputer Centers.

Over the course of this and the next year, the initial multiyear contracts with the National Centers are coming to an end, which has provoked a debate about whether and, if so, in what form they should be renewed. NSF undertook an elaborate review and renewal process and announced that, depending on agency funding, it is prepared to proceed with renewing at least four of the centers[2]. In thinking about the next steps in the evolution of the advanced computing program, the science agencies and Congress have asked some basic questions. Have our perceptions of the needs of research for HPC changed since the centers were started? If so, how? Have we learned anything about the effectiveness of the National Centers approach? Should the goals of the Advanced Scientific Computing (ASC) and other related Federal programs be refined or redefined? Should alternative approaches be considered, either to replace or to supplement the contributions of the centers?

OTA is presently engaged in a broad assessment of the impacts of information technology on research, and as part of that inquiry, is examining the question of scientific computational resources. It has been asked by the requesting committees for an interim paper that might help shed some light on the above questions. The full assessment will not be completed for several months, however; so this paper must confine itself to some tentative observations.

WHAT ISA HIGH PERFORMANCE COMPUTER?

The term, "supercomputer," is commonly used in the press, but it is not necessarily useful for policy. In the first place, the definition of power in a computer is highly inexact and depends on many factors including processor speed, memory size, and so on. Secondly, there is not a clear lower boundary of supercomputer power. IBM 3090 computers come in a wide range of configurations, some of the largest of which are the basis of supercomputer centers at institutions such as Cornell, the Universities of Utah, and Kentucky. Finally, technology is changing rapidly and with it our conceptions of power and capability of various types of machines. We use the more general term, "high performance computers," a term that includes a variety of machine types.

One class of HPC consists of very large. powerful machines, principally designed for very large numerical applications such as those encountered in science. These computers are the ones often referred to as "supercomputers." They are expensive, costing up to several million dollars each.

---

[1] Peter D. Lax, *Report of the Panel on Large-Scale Computing in Science and Engineering* (Washington, DC: National Science Foundation, 1982).

[2] One of the five centers, the John von Neumann National Supercomputer Center, has been based on ETA-10 technology The Center has been asked to resubmit a proposal showing revised plans in reaction to the withdrawal of that machine from the market.

A large-scale computer's power comes from a combination of very high-speed electronic components and specialized architecture (a term used by computer designers to describe the overall logical arrangement of the computer). Most designs use a combination of "vector processing" and "parallelism" in their design. A vector processor is an arithmetic unit of the computer that produces a series of similar calculations in an overlapping, assembly line fashion, (Many scientific calculations can be set up in this way.)

Parallelism uses several processors, assuming that a problem can be broken into large independent pieces that can be computed on separate processors. Currently, large, mainframe HPC'S such as those offered by Cray, IBM, are only modestly parallel, having as few as two up to as many as eight processors.[3] The trend is toward more parallel processors on these large systems. Some experts anticipate as many as 512 processor machines appearing in the near future. The key problem to date has been to understand how problems can be set up to take advantage of the potential speed advantage of larger scale parallelism.

Several machines are now on the market that are based on the structure and logic of a large supercomputer, but use cheaper, slower electronic components. These systems make some sacrifice in speed, but cost much less to manufacture. Thus, an application that is demanding, but that does not necessarily require the resources of a full-size supercomputer, may be much more cost effective to run on such a "minisuper."

Other types of specialized systems have also appeared on the market and in the research laboratory. These machines represent attempts to obtain major gains in computation speed by means of fundamentally different architectures. They are known by colorful names such as "Hypercubes," "Connection Machines," "Dataflow Processors," "Butterfly Machines," "Neural Nets," or "Fuzzy Logic Computers." Although they differ in detail, many of these systems are based on large-scale parallelism. That is, their designers attempt to get increases in processing speed by hooking together in some way a large number-hundreds or even thousands-of simpler, slower and, hence, cheaper processors. The problem is that computational mathematicians have not yet developed a good theoretical or experiential framework for understanding in general how to arrange applications to take full advantage of these massively parallel systems. Hence, they are still, by and large, experimental, even though some are now on the market and users have already developed applications software for them. Experimental as these systems may seem now, many experts think that any significantly large increase in computational power eventually must grow out of experimental systems such as these or from some other form of massively parallel architecture.

Finally, "workstations," the descendants of personal desktop computers, are increasing in power; new chips now in development will offer the computing power nearly equivalent to a Cray 1 supercomputer of the late 1970s. Thus, although top-end HPCs will be correspondingly more powerful, scientists who wish to do serious computing will have a much wider selection of options in the near future,

A few policy-related conclusions flow from this discussion:

- The term "Supercomputer" is a fluid one, potentially covering a wide variety of machine types, and the "supercomputer industry" is similarly increasing y difficult to identify clearly.
- Scientists need access to a wide range of high performance computers, ranging from desktop workstations to full-scale supercomputers, and they need to move smoothly among these machines as their research needs dictate.
- Hence, government policy needs to be flexible and broadly based, not overly focused on narrowly defined classes of machines.

HOW FAST IS FAST?

Popular comparisons of supercomputer speeds are usually based on processing speed, the measure being "FLOPS," or "Floating Point Operation Per Second." The term "floating point" refers to a particular format for numbers within the computer that is used for scientific calculation; and a floating

---

[3] To distinguish between this modest level and the larger scale parallelism found on some more experimental machines, some experts refer to this limited parallelism as "multiprocessing."

point “operation” refers to a single arithmetic step, such as adding two numbers, using the floating point format, Thus, FLOPS measure the speed of the arithmetic processor. Currently, the largest supercomputers have processing speeds ranging up to several billion FLOPS.

However, pure processing speed is not by itself a useful measure of the relative power of computers. To see why, let’s look at an analogy.

In a supermarket checkout counter, the calculation speed of the register does not, by itself, determine how fast customers can purchase their groceries and get out of the store. Rather, the speed of checkout is also affected by the rate at which each purchase can be entered into the register and the overall time it takes to complete a transaction with a customer and start a new one. Of course, ultimately, the length of time the customer must wait in line to get to the clerk may be the biggest determinant of all.

Similarly, in a computer, how fast calculations can be set up and presented to the processor and how fast new jobs and their associated data can be moved in, and completed work moved out of the computer, determines how much of the processor’s speed can actually be harnessed. (Some users refer to this as “solution speed.”) In a computer, those speeds are determined by a wide variety of hardware and software characteristics. And, similar to the store checkout, as a fast machine becomes busy, users may have to wait a significant time to get their turn. From a user’s perspective, then, a theoretically fast computer can look very slow.

In order to fully test a machine’s speed, experts use what are called “benchmark programs,” sample programs that reproduce the actual work load. Since workloads vary, there are several different benchmark programs, and they are constantly being refined and revised. Measuring a supercomputer’s speed is, itself, a complex and important area of research. It lends insight not only into what type of computer currently on the market is best to use for particular applications; but carefully structured measurements can also show where bottlenecks occur and, hence, where hardware and software improvements need to be made.

One can draw a few policy implications from these observations on speed:

- Since overall speed improvement is closely linked with how their machines are actually programmed and used, computer designers are critically dependent on feedback from that part of the user community which is pushing their machines to the limit.
- There is no “fastest” machine. The speed of a high performance computer is too dependent on the skill with which it is used and programmed, and the particular type of job it is being asked to perform.
- Until machines are available in the market and have been tested for overall performance, policy makers should be skeptical of announcements based purely on processor speeds that some company or country is producing “faster machines. ”
- Federal R&D programs for improving high performance computing need to stress software and computational mathematics as well as research on machine architecture.

THE NATIONAL SUPERCOMPUTER CENTERS

In February of 1985, NSF selected four sites to establish national supercomputing centers: The University of California at San Diego, The University of Illinois at Urbana-Champaign, Cornell University and the John von Neumann Center in Princeton. A fifth site, Pittsburgh, was added in early 1986. The five NSF centers are described briefly below.

The Cornell Theory Center

The Cornell Theory Center is located on the campus of Cornell University. Over 1,900 users from 125 institutions access the center. Although Cornell does not have a center-oriented network, 55 academic institutions are able to utilize the resources at Cornell through special nodes. A 14-member Corporate Research Institute works within the center in a variety of university-industry cost sharing projects.

In November of 1985 Cornell received a 3084 computer from IBM, which was upgraded to a four-processor 3090/400VF a year later. The ***3090/***400VF was replaced by a six-processor ***3090/600E*** in May, 1987. In October, 1988 a second 3090/600E was added. The Cornell center also operates several other smaller parallel systems, including an Intel iPCS/2, a Transtech NT 1000, and a Topologix T1OOO. Some 50 percent of the resources of Northeast Parallel Architecture Center, which include two Connection machines, an Encore, and an Alliant FX/80, are accessed by the Cornell facility.

Until October of 1988, all IBM computers were "on loan" to Cornell for as long as Cornell retained its NSF funding. The second IBM 3090/600, procured in October, will be paid for by an NSF grant, Over the past 4 years, corporate support for the Cornell facility accounted for 48 percent of the operating costs. During those same years, NSF and New York State accounted for 37 percent and 5 percent respectively of the facility's budget. This funding has allowed the center to maintain a staff of about 100.

The National Center for Supercomputing Applications

The National Center for Supercomputing Applications (NCSA) is operated by the University of Illinois at Urbana-Champaign. The Center has over 2,500 academic users from about 82 academic afiliates. Each affiliate receives a block grant of time on the Cray X-MP/48, training for the Cray, and help using the network to access the Cray.

The NCSA received its Cray X-MP/24 in October 1985, That machine was upgraded to a Cray X-MP/48 in 1987. In October 1988 a Cray-2s/4-128 was installed, giving the center two Cray machines. This computer is the only Cray-2 now at an NSF national center. The center also houses a Connection Machine 2, an Alliant FX/80 and FX/8, and over 30 graphics workstations.

In addition to NSF funding, NCSA has solicited industrial support. Amoco, Eastman Kodak, Eli Lilly, FMC Corp., Dow Chemical, and Motorola have each contributed around $3 million over a 3-year period to the NCSA. In fiscal year 1989 corporate support has amounted to 11 percent of NCSA's funding. About 32 percent of NCSA's budget came from NSF while the State of Illinois and the University of Illinois accounted for the remaining 27 percent of the center's $21.5 million budget. The center has a full-time staff of 198.

Pittsburgh Supercomputing Center

The Pittsburgh Supercomputing Center (PSC) is run jointly by the University of Pittsburgh, Carnegie-Mellon University, and Westinghouse Electric Corp. More than 1,400 users from 44 States utilize the center. Twenty-seven universities are affiliated with PSC.

The center received a Cray X-MP/48 in March of 1986. In December of 1988 PSC became the first non-Federal laboratory to possess a Cray Y-MP. Both machines were being used simultaneously for a short time, however the center has phased out the Cray X-MP. The center's graphics hardware includes a Pixar image computer, an Ardent Titan, and a Silicon Graphics IRIS workstation.

The operating projection at PSC for fiscal year 1990, a "typical year," has NSF supporting 58 percent of the center's budget while industry and vendors account for 22 percent of the costs. The Commonwealth of Pennsylvania and the National Institutes of Health both support PSC, accounting for 8 percent and 4 percent of budget respectively. Excluding working students, the center has a staff of around 65.

San Diego Supercomputer Center

The San Diego Supercomputer Center (SDSC) is located on the campus of the University of California at San Diego and is operated by General Atomics. SDSC is linked to 25 consortium members but has a user base in 44 States. At the end of 1988, over 2,700 users were accessing the center. SDSC has 48 industrial partners who use the facility's hardware, software, and support staff.

A Cray X-MP/48 was installed in December, *1985.* SDSC's *first* upgrade, a Y-MP8/864, is planned for December, 1989. In addition to the Cray, SDSC has *5* Sun workstations, two IRIS workstations, an Evans and Sutherland terminal, 5 Apollo workstations, a Pixar, an Ardent Titan, an SCS-40 minisupercomputer, a Supertek S-1 minisupercomputer, and two Symbolics Machines.

The University of California at San Diego spends more than $250,000 a year on utilities and services for SDSC. For fiscal year 1990 the SDSC believes NSF will account for 47 percent of the center's operating budget. The State of California currently provides $1.25 million per year to the center and in 1988, approved funding of $6 million over 3 years to SDSC for research in scientific visualization. For fiscal year 1990 the State is projected to support 10 percent of the center's costs. Industrial support, which has given the center $12.6 million in donations and in-kind services, is projected to provide 15 percent of the total costs of SDSC in fiscal year 1990.

John von Neumann National Supercomputer Center

The John von Neumann National Supercomputer Center (JvNC), located in Princeton New Jersey, is managed by the Consortium for Scientific Computing Inc., an organization of 13 institutions from New Jersey, Pennsylvania, Massachusetts, New York, Rhode Island, Colorado, and Arizona. Currently there are over 1,400 researchers from 100 institutes accessing the center. Eight industrial corporations utilize the JvNC facilities.

At present there are two Cyber 205 and two ETA-1 OS, in use at the JvNC. The first ETA-10 was installed, after a 1-year delay, in March of 1988. In addition to these machines there is a Pixar H, two Silicon Graphics IRIS and video animation capabilities.

When the center was established in 1985 by NSF, the New Jersey Commission on Science and Technology committed $12.1 million to the center over a 5-year period. An addition $13.1 million has been set-aside for the center by the New Jersey Commission for fiscal year 1991-1995. Direct funding from the State of New Jersey and university sources constitutes 15 percent of the center's budget for fiscal year 1991-1995. NSF will account for 60 percent of the budget. Projected industry revenue and cost sharing account for 25 percent of costs. Since the announcement by CDC to close its ETA subsidiary, the future of JvNC is uncertain. Plans have been proposed to NSF by JvNC to purchase a Cray Research Y-MP, eventually upgrading to a C-90. NSF is reviewing the plan and a decision on renewal is expected in October of 1989,

OTHER HPC FACILITIES

Before 1984 only three universities operated supercomputers: Purdue University, the University of Minnesota, and Colorado State University. The NSF supercomputing initiative established five new supercomputer centers that were nationally accessible. States and universities began funding their own supercomputer centers, both in response to growing needs on campus and to increased feeling on the part of State leaders that supercomputer facilities could be important stimuli to local R&D and, therefore, to economic development. Now, many State and university centers offer access to high performance computers;[4] and the NSF centers are only part of a much larger HPC environment including nearly 70 Federal installations (see table 2-l).

Supercomputer center operators perceive their roles in different ways. Some want to be a proactive force in the research community, leading the way by helping develop new applications, training users, and so on. Others are content to follow in the path that the NSF National Centers create. These differences in goals/missions lead to varied services and computer systems. Some centers are "cycle shops," offering computing time but minimal support staff. Other centers maintain a large support staff and offer consulting, training sessions, and even assistance with software development. Four representative centers are described below:

Minnesota Supercomputer Center

The Minnesota Supercomputer Center, originally part of the University of Minnesota, is a for-profit computer center owned by the University of Minnesota. Currently, several thousand researchers use the center, over 700 of which are from the University of Minnesota. The Minnesota Supercomputing Institute, an academic unit of the University, channels university usage by providing grants to the students through a peer review process.

The Minnesota Supercomputer Center received its first machine, a Cray 1A, in September, 1981. In mid 1985, it installed a Cyber 205; and in the latter part of that year, two Cray 2 computers were installed within 3 months of each other. Minnesota

---

[4] The number cannot be estimated exactly. First, it depends on the definition of supercomputer one uses. Secondly, the number keeps changing as States announce new plans for centers and as large research universities purchase their own HPCs.

Table 2-l—Federal Unctassified Supercomputer Installations

Laboratory	Number of machines
Department of Energy	
Los Alams National Lab	6
Livermore National Lab, NMFECC	4
Livermore National Lab	7
Sandia **National** Lab, Livermore.	3
Sandia National Lab, Albuquerque	2
Oak Ridge National Lab	1
idaho Falls National Engineering	1
Argonne National Lab	1
Knolls Atomic Power Lab	1
Bettis Atomic Power Lab	1
Savannah/DOE	1
Richland/DOE	1
Schenectady Naval Reactors/DOE	
Pittsburgh Naval Reactors/DOE	2
Department of Defense	
Naval Research Lab	1
Naval Ship R&D Center .,	1
Fleet Numerical Oceanography	1
Naval Underwater System Command	1
Naval Weapons Center	1
Martin Marietta/NTB	1
Air Force Wapons Lab	2
Air Force Global Weather	1
Arnold Engineering and Development	1
Wright Patterson AFB	1
Aerospace Corp.	1
Army Ballistic Research Lab	2
Army/Tacom ... , ,	1
Army/Huntsville.	1
Army/Kwajaiein	1
Army/WES (on order)	1
Army/Warren	1
Defense Nuclear Agency	1
NASA	
Ames	5
Goddard	2
Lewis	1
Langley	1
Marshal	1
Department Commerce	
National inst. of Standards and Technology	1
National Oceanic & Atmospheric Administration	4
Environmental **Protection Agency**	
Raleigh, North Carolina	1
Eeportment and Human Services	
National Institutes of Health	1
National Cancer institute	1

SOURCE: Offics of Technology Assessment estimate.

bought its third Cray 2, the only one in use now,at the end of 1988, just after it installed its ETA-10. The ETA-10 has recently been decommissioned due to the closure of ETA. A Cray X-MP has been added, giving them a total of two supercomputers. The Minnesota Supemomputer Centerhas acquired more supercomputers than anyone outside the Federal Government.

The Minnesota State Legislature provides funds to the University for the purchasing of supercomputer time. Although the University buys a substantial portion of supercomputing time, the center has many industrial clients whose identities are proprietary, but they include representatives of the auto, aerospace, petroleum, and electronic industries. They are charged a fee for the use of the facility.

The Ohio Supercomputer Center

The Ohio Supercomputer Center (OSC) originated from a coalition of scientists in the State. The center, located on Ohio State University's campus, is connected to 20 other Ohio universities via the Ohio Academic Research Network (OARNET). As of January 1989, three private firms were using the Center's resources.

In August, 1987, OSC installed a Cray X-MP/24, which was upgraded to an Cray X-MP/28 a year later. The center replaced the X-MP in August 1989 with a Cray Research Y-MP. In addition to Cray hardware, there are 40 Sun Graphic workstations, a Pixar II, a Stallar Graphics machine, a Silicon Graphic workstation and a Abekas Still Store machine. The Center maintains a staff of about 35 people.

The Ohio General Assembly began funding the center in the summer of 1987, appropriating $7.5 million. In March of 1988, the Assembly allocated ***$22*** million for the acquisition of a Cray Y-MP. Ohio State University has pledged $8.2 million to augment the center's budget. As of February 1989 the State has spent $37.7 million in funding.[5] OSC's annual budget is around $6 million (not including the purchase/leasing of their Cray).

Center for High Performance Computing, Texas (CHPC)

The Center for High Performance Computing is located at The University of Texas at Austin. CHPC serves all 14 institutions, 8 academic institutions, and 6 health-related organizations, in the University of Texas System.

---

[5] Jane Ware, "Ohioans: Blazing Computer," Ohio, February 1989, p. 12.

The University of Texas installed a Cray X-MP/24 in March 1986, and a Cray 14se in November of 1988. The X-MP is used primarily for research. For the time being, the Cray 14se is being used as a vehicle for the conversion of users to the Unix system. About 40 people staff the center.

Original funding for the center and the Cray X-NIP came from bonds and endowments from both The University of Texas system and The University of Texas at Austin. The annual budget of CHPC is about $3 million. About 95 percent of the center's operating budget comes from State funding and endowments. Five percent of the costs are recovered from selling CPU time.

Alabama Supercomputer Network

The George C. Wallace Supercomputer Center, located in Huntsville Alabama, serves the needs of researchers throughout Alabama. Through the Alabama Supercomputer Network, 13 Alabama institutions, university and government sites, are connected to the center. Under contract to the State, Boeing Computer Services provides the support staff and technical skills to operate the center. Support staff are located at each of the nodes to help facilitate the use of the supercomputer from remote sites.

A Cray X-MP/24 arrived in 1987 and became operational in early 1988. In 1987 the State of Alabama agreed to finance the center. The State allocated $2.2 million for the center and $38 million to Boeing Services for the initial 5 years. The average yearly budget is $7 million. The center has a support staff of about 25.

Alabama universities are guaranteed 60 percent of the available time at no cost while commercial researchers are charged a user fee. The impetus for the State to create a supercomputer center has been stated as the technical superiority a supercomputer would bring, which would draw high-tech industry to the State, enhance interaction between industry and the universities, and promote research and the associated educational programs within the university.

Commercial Labs

A few corporations, such as the Boeing Computer Corp., have been selling high performance computer time for a while. Boeing operates a Cray X-MP/24. Other commercial sellers of high performance computing time include the Houston Area Research Center (HARC). HARC operates the only Japanese Supercomputer in America, the NEC SX2. The center offers remote services.

Computer Sciences Corp. (CSC), located in Falls Church, Virginia, has a 16-processor FLEX/32 from Flexible Computer Corp., a Convex 120 from Convex Computer Corp, and a DAP21O from Active Memory Technology. Federal agencies comprise two-thirds of CSC's customers.[6] Power Computing Co., located in Dallas, Texas, offers time on a Cray X-MP/24. Situated in Houston, Texas, Supercomputing Technology sells time on its Cray X-MP/28. Opticom Corp., of San Jose, California, offers time on a Cray X-MP/24, Cray l-M, Convex C220, and cl XP.

Federal Centers

In an informal poll of Federal agencies, OTA identified 70 unclassified installations that operate supercomputers, confirming the commonly expressed view that the Federal Government still represents a major part of the market for HPC in the United States (see figure 2-l). Many of these centers serve the research needs of government scientists and engineers and are, thus, part of the total research computing environment. Some are available to non-Federal scientists, others are closed.

CHANGING ENVIRONMENT

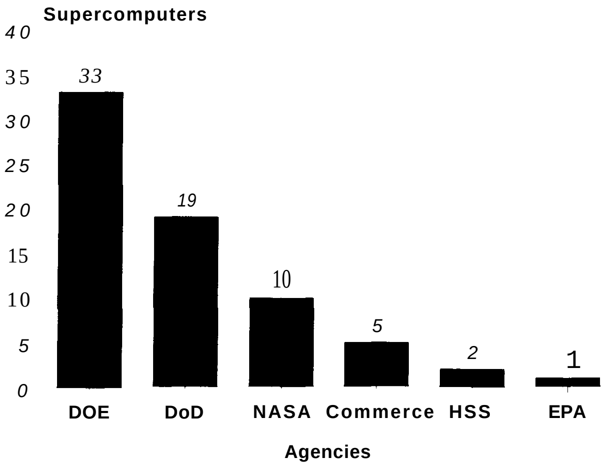
The scientific computing environment has changed in important ways during the few years that NSF's Advanced Scientific Computing Programs have existed. Some of these changes are as follows:

The ASC programs, themselves, have not evolved as originally planned. The original NSF planning document for the ASC program originally proposed to establish 10 supercomputer centers over a 3-year period; only 5 were funded. Center managers have also expressed the strong opinion that NSF has not met many of its original commitments for

---

[6] Norris Parker Smith, "More Than Just Buying Cycles," *Supercomputer Review,* April 1989.

Figure 2-1—Distribution of Federal Suparcomputers



SOURCE: office of Technology Assessment, 1989.

funding in successive years of the contracts, forcing the centers to change their operational priorities and search for support in other directions.

Technology has changed. There has been a burst of innovation in the HPC industry. At the top of the line, Cray Research developed two lines of machines, the Cray 2 and the Cray X-MP (and its successor, the Y-MP) that are much more powerful than the Cray 1, which was considered the leading edge of supercomputing for several years by the mid- 1980s. IBM has delivered several 3090s equipped with multiple vector processors and has also become a partner in a project to develop a new supercomputer in a joint venture with SSI, a firm started by Steve Chen, a noted supercomputer architect previously with Cray Research.

More recently, major changes have occurred in the industry. Control Data has closed down ETA, its supercomputer operation. Cray Research has been broken into two parts-Cray Computer Corp. and Cray Research. Each will develop and market a different line of supercomputers. Cray Research will, initially, at least, concentrate on the Y-MP models, the upcoming C-90 machines, and their longer term successors. Cray Computer Corp., under the leadership of Seymour Cray, will concentrate on development of the Cray 3, a machine based on gallium arsenide electronics,

At the middle and lower end, the HPC industry has introduced several new so-called "minisupercomputers"-many of them based on radically different system concepts, such as massive parallelism, and many designed for specific applications, such as high-speed graphics. New chips promise very high-speed desktop workstations in the near future.

Finally, three Japanese manufacturers, NEC, Fujitsu, and Hitachi have been successfully building and marketing supercomputers that are reportedly competitive in performance with U.S. machines.[7] While these machines have, as yet, not penetrated the U.S. computer market, they indicate the potential competitiveness of the Japanese computer industry in the international HPC markets, and raise questions for U.S. policy.

Many universities and State systems have established "supercomputer centers" to serve the needs of their researchers.[8] Many of these centers have only recently been formed, some have not yet installed their systems, so their operational experience is, at best, limited to date. Furthermore, some other centers operate systems that, while very powerful scientific machines, are not considered by all experts to be supercomputers. Nevertheless, these centers provide high performance scientific computing to the research community, and create new demands for Federal support for computer time.

Individual scientist and research teams are also getting Federal and private support from their sponsors to buy their own "minisupercomputers." In some cases, these systems are used to develop and check out software eventually destined to run on larger machines; in other cases, researchers seem to find these machines adequate for their needs. In either mode of use, these departmental or laboratory systems expand the range of possible sources researchers turn to for high performance computing. Soon, desktop workstations will have performance equivalent to that of supercomputers of a decade ago at a significantly lower cost.

---

[7] Since, as shown above, comparing the power and performance of supercomputers is a complex and arcane field, OTA will refrain from comparing or ranking systems in any absolute sense.

[8] See National Association of State Universities and Land-Grant Colleges, *Supercomputing for the 1990's: A Shared Responsibility* (Washington, DC: January 1989).

Finally, some important changes have occurred in national objectives or perceptions of issues. For example, the development of a very high capacity national science network (or "internet") has taken on a much greater significance. Originally conceived of in the narrow context of tying together supercomputer centers and providing regional access to them, the science network has now come to be thought of by its proponents as a basic infrastructure, potentially extending throughout (and, perhaps, even beyond) the entire scientific, technical, and educational community.

Science policy is also changing, as new important and costly projects have been started or are being seriously considered, Projects such as the supercollider, the space station, NASA's Earth Observing System (EOS) program, and the human genome mapping may seem at first glance to compete for funding with science networks and supercomputers. However, they will create formidable new demands for computation, data communications, and data storage facilities; and, hence, constitute additional arguments for investments in an information technology infrastructure.

Finally, some of the research areas in the so-called "Grand Challenges"[9] have attained even greater social importance-such as fluid flow modeling which will help the design of faster and more fuel efficient planes and ships, climate modeling to help understand long term weather patterns, and the structural analysis of proteins to help understand diseases and design vaccines and drugs to fight them.

REVIEW AND RENEWAL OF THE NSF CENTERS

Based on the recent review, NSF has concluded that the centers, by and large, have been successful and are operating smoothly. That is, their systems are being fully used, they have trained many new users, and they are producing good science. In light of that conclusion, NSF has tentatively agreed to renewal for the three Cray-based centers and the IBM-based Cornell Center. The John von Neumann Center in Princeton has been based on ETA-10 computers. Since ETA was closed down, NSF put the review of the JvNC on hold pending review of a revised plan that has now been submitted, A decision is expected soon.

Due to the environmental changes noted above, if the centers are to continue in their present status as special NSF-sponsored facilities, the National Supercomputer Centers will need to sharply define their roles in terms of: 1) the users they intend to serve, 2) the types of applications they serve, and 3) the appropriate balance between service, education, and research.

The NSF centers are only a few of a growing number of facilities that provide access to HPC resources. Assuming that NSF's basic objective is to assure researchers access to the most appropriate computing for their work, it will be under increasing pressure to justify dedicating funds to one limited group of facilities. Five years ago, few U.S. academic supercomputer centers existed. When scientific demand was less, managerial attention was focused on the immediate problem of getting equipment installed and of developing an experienced user community. Under those circumstances, some ambiguity of purpose may have been acceptable and understandable. However, in light of the proliferation of alternative technologies and centers, as well as growing demand by researchers, unless the purposes of the National Centers are more clearly delineated, the facilities are at risk of being asked to serve too many roles and, as a result, serving none well.

Some examples of possible choices are as follows:

L Provide Access to HPC

- Provide access to the most powerful, leading edge, supercomputers available,
- Serve the HPC requirements for research projects of critical importance to the Federal Government, for example, the "Grand Challenge" topics.
- Serve the needs of all NSF-funded researchers for HPC.
- Serve the needs of the (academic, educational, and/or industrial) scientific community for HPC.

---

[9] "Grand Challenge" research topics are questions of major social importance that require for progress substantially greater computing resources than are currently available. The term was first coined by Nobel Laureate physicist, Kenneth Wilson.

2. Educate and Train
   - Provide facilities and programs to teach scientists and students how to use high performance computing in their research.
3. Advance the State of HPC Use in Research
   - Develop applications and system software.
   - Serve as centers for research in computational science.
   - Work with vendors as test sites for advanced HPC systems.

As the use of HPC expands into more fields and among more researchers, what are the policies for providing access to the necessary computing resources? The Federal Government needs to develop a comprehensive analysis of the requirements of the scientific researchers for high performance computing, Federal policies of support for scientific computing, and the variety of Federal and State/private computing facilities available for research.

We expect that OTA's final report will contribute to this analysis from a congressional perspective. However, the executive branch, including both lead agencies and OSTP also need to participate actively in this policy and planning process.

THE INTERNATIONAL ENVIRONMENT

Since some of the policy debate over HPCs has involved comparison with foreign programs, this section will conclude a brief description of the status of HPC in some other nations.

Japan

The Ministry of International Trade and Industry (MITI), in October of 1981, announced the undertaking of two computing projects, one on artificial intelligence, the Fifth Generation Computer Project, and one on supercomputing, the National Superspeed Computer Project. The publicity surrounding MITI's announcement focused on fifth generation computers, but brought the more general subject of supercomputing to the public attention. (The term "Fifth Generation" refers to computers specially designed for artificial intelligence applications, especially those that involve logical inference or "reasoning.")

Although in the eyes of many scientists the Fifth Generation project has fallen short of its original goals, eight years later it has produced some accomplishments in hardware architecture and artificial intelligence software. MITI's second project, dealing with supercomputers, has been more successful. Since 1981, when no supercomputers were manufactured by the Japanese, three companies have designed and produced supercomputers.

The Japanese manufacturers followed the Americans into the supercomputer market, yet in the short time since their entrance, late 1983 for Hitachi and Fujitsu, they have rapidly gained ground in HPC hardware. One company, NEC, has recently announced a supercomputer with processor speeds up to eight times faster than the present fastest American machine.[10] Outside of the United States, Japan is the single biggest market for and supplier of supercomputers, although American supercomputer companies account for less than one-fifth of all supercomputers sold in Japan.[11]

In the present generation of supercomputers, U.S. supercomputers have some advantages. One of American manufacturer's major advantages is the availability of scientific applications software. The Japanese lag behind the Americans in software development, although resources are being devoted to research in software by the Japanese manufacturers and government and there is no reason to think they will not be successful.

Another area in which American firms differ from *the* Japanese has been in their use of multiprocessor architecture (although this picture is now changing). For several years, American supercomputer companies have been designing machines with multiprocessors to obtain speed. The only Japanese supercomputer that utilizes multiprocessors is the NEC system, which will not be available until the fall of 1990.

---

[10] The NEC machine is not scheduled for delivery until 1990, at which time faster Cray computers may well be on the market also. See also the comments above about computer speed.

[1] Marjorie Sun, "A Global Supercomputer Race for High Stakes," *Science*, February 1989, vol. 243, pp. 1004-1006.

American firms have been active in the Japanese market, with mixed success.

Since 1979 Cray has sold 16 machines in Japan. Of the 16 machines, 6 went to automobile manufacturers, 2 to NTT, 2 to Recruit, 1 to MITI, 1 to Toshiba, 1 to Aichi Institute of Technology, and 1 to Mitsubishi Electric. None have gone to public universities or to government agencies.

IBM offers their 3090 with attached vector facilities, IBM does not make public its customers, but report that they have sold around 70 vector processor computers to Japanese clients. Some owners, or soon to be owners, include Nissan, NTT, Mazda, Waseda University, Nippon Steel and Mistubishi Electric.

ETA sold two supercomputers in Japan. The first was to the Tokyo Institute of Technology (TIT). The sale was important because it was the first sale of a CDC/ETA supercomputer to the Japanese as well as the first purchase of an American supercomputer by a Japanese national university. This machine was delivered late (it arrived in May of 1988) and had many operating problems, partially due to its being the first installment of an eight-processor ETA 1O-E. The second machine was purchased (not delivered) on February 9, 1989 by the University of Meiji. How CDC will deal with the ETA 10 at TIT in light of the closure of ETA is unknown at this time.

Hitachi, Fujitsu, and NEC, the three Japanese manufacturers of supercomputers, are among the largest computer/electronic companies in Japan; and they produce their own semiconductors. Their size allows them to absorb the high initial costs of designing a new supercomputer, as well as provide large discounts to customers. Japan's technological lead is in its very fast single-vector processors. Little is known, as of yet, what is happening with parallel processing in Japan, although NEC's recent product announcement for the SX-X states that the machine will have multiprocessors.

Hitachi's supercomputer architecture is loosely based on its IBM compatible mainframe. Hitachi entered the market in November of 1983. Unlike their domestic rivals, Hitachi has not entered the international market. All 29 of its ordered/installed supercomputers are located in Japan.

NEC's current supercomputer architecture is not based on its mainframe computer and it is not IBM compatible. They entered the supercomputer market later than Hitachi and Fujitsu. Three NEC supercomputers have been sold/installed in foreign markets: one in the United States, an SX-2 machine at the Houston Area Research Consortium, one at the Laboratory of Aerospace Research in Netherlands, and an SX-1 has recently been sold in Singapore. Their domestic users include five universities.

On April 10, 1989, in a joint venture with Honeywell Inc., NEC announced a new line of supercomputers, the SX-X. The most powerful machine is reported to be up to eight times faster than the Cray X-MP machine. The SX-X reportedly will run Unix-based software and will have multiprocessors. This machine is due to be shipped in the fall of 1990.

Fujitsu's supercomputer, like Hitachi's, is based on their IBM compatible mainframes. Their first machine was delivered in late 1983. Fujitsu had sold 80 supercomputers in Japan by rnid-1989. An estimated 17 machines have been sold to foreign customers. An Amdahl VP-200 is used at the Western Geophysical Institute in London. In the United States, the Norwegian company GECO, located in Houston, has a VP-200 and two VP-1OOs. The most recent sale was to the Australian National University, a VP-1OO.

Europe

European countries that have (or have ordered) supercomputers include: West Germany, France, England, Denmark, Spain, Norway, the Netherlands, Italy, Finland, Switzerland, and Belgium. Europe is catching up quickly with America and Japan in understanding the importance of high performance computing for science and industry. The computer industry is helping to stimulate European interest. For example, IBM has pledged $40 million towards a supercomputer initiative in Europe over the 2-year period between 1987-89. It is creating a large base of followers in the European academic community by participating in such programs as the European Academic Supercomputing Initiative (EASI), and the Numerically Intensive Computing Enterprise (NICE). Cray Research also has a solid base in academic Europe, supplying over 14 supermomputers to European universities.

The United Kingdom began implementing a high performance computing plan in 1985. The Joint Working Party on Advanced Research Computing's report in June of 1985, "Future Facilities for Advanced Research Computing," recommended a national facility for advanced research computing. This center would have the most powerful supercomputer available; upgrade the United Kingdom's networking systems, JANET, to ensure communications to remote users; and house a national organization of advanced research computing to promote collaboration with foreign countries and within industry, ensuring the effective use of these resources."[12] Following this report, a Cray XMP/48 was installed at the Atlas Computer Center in Rutherford. A Cray 1s was installed at the University of London. Between 1986 and 1989, some $11.5 million was spent on upgrading and enhancing JANET[13]

Alvey was the United Kingdom's key information technology R&D program. The program promoted projects in information technology undertaken jointly by industry and academics. The United Kingdom began funding the Alvey program in 1983. During the first 5 years, 350 million pounds were allocated to the Alvey program. The program was eliminated at the end of 1988. Some research was picked up by other agencies, and many of the projects that were sponsored by Alvey are now submitting proposals to Esprit (see below).

The European Community began funding the European Strategic Programme for Research in Information Technology (Esprit) program in 1984 partly as a reaction to the poor performance of the European Economic Community in the market of information technology and partly as a response to MITI's 1981 computer programs. The program, funded by the European Community (EC), intends to "provide the European IT industry with the key components of technology it needs to be competitive on the world markets within a decade. "[14] The EC has designed a program that forces collaboration between nations, develops recognizable standards in the information technology industry, and promotes pre-competitive R&D. The R&D focuses on five main areas: microelectronics, software development, office systems, computer integrated manufacturing, and advanced information.

Phase I of Esprit, the first 5 years, received $3.88 billion in funding.[15] The finding was split 50-50 by the EC and its participants, This was considered the catch-up phase. Emphasis was placed on basic research, realizing that marketable goods will follow. Many of the companies that participated in Phase I were small experimental companies.

Phase II, which begins in late 1989, is called commercialization. Marketable goods will be the major emphasis of Phase II. This implies that the larger firms will be the main industrial participants since they have the capital needed to put a product on the market. The amount of funds for Phase II will be determined by the world environment in information technology and the results of Phase I, but has been estimated at around $4.14 billion.[16]

Almost all of the high performance computer technologies emerging from Europe have been based on massively parallel architectures. Some of Europe's parallel machines incorporate the transputer. Transputer technology (basically a computer on a chip) is based on high density VLSI (very large-scale integration) chips. The T800, Inmos's transputer, has the same power as Intel's 80386/80387 chip, the difference being in size and price. The transputer is about one-third the size and price of Intel's chip.[17] The transputer, created by the Inmos company, had its initial R&D funded by the British government. Eventually Thorn EMI bought Inmos and the rights to the transputer. Thorn EMI recently sold Inmos to a French-Italian joint venture company, SGS-Thomson, just as it was beginning to be profitable.

---

[12] "Future Facilities for Advanced Research Computing," the report of a Joint Working Party on Advanced Research Computing, United Kingdom, Juty 1985.

[13] Discussion paper on "Supercomputers in Australia," Department of Industry, Technoloy and Commerce, April 1988, pp. 14-15.

[14] "Esprit," commission of the European Communities, p. 5.

[15] "Esprit," Commission of the European Communities, p. 21.

[16] Simon Perry, "European Team Effort Breaks Ground in Software Standards," *Electronic Business*, Aug. 15, 1988, pp. 90-91.

[17] Graham K. EMS, "Transputers Advance Parallel Processing," *Research and Development*, March 1989, p. 50.

Some of the more notable high performance computer products and R&D in Europe include:

- T.Node, formerly called Supernode P1085, is one of the more successful endeavors of the Esprit program. T.Node is a massively parallel machine that exploits the Inmos T800 transputer. A single node is composed of 16 transputers connected by two NEC VLSI chips and two additional transputers. The participants in the project are The University of Southampton, Royal Signals, Radar Establishment, Thom-EMI (all British) and the French firm Telemat. The prototype of the French T. Node, Marie, a massively parallel MIMD (multiple instruction, multiple data) computer, was delivered in April of 1988. The product is now being marketed in America.
- Project 415 is also funded by Esprit. Its project leader is Philips, the Dutch electronics group. This project, which consists of six groups, focuses on symbolic computation, artificial intelligence (AI), rather than “number crunching” (mathematical operations by conventional supercomputers). Using parallel architecture, the project is developing operating systems and languages that they hope will be available in 5 years for the office environment.[18]
- The Flagship project, originally sponsored by the Alvey program, has created a prototype parallel machine using 15 processors. Its original participants were ICL, Imperial College, and the University of Manchester. Other Alvey projects worked with the Flagship project in designing operating systems and languages for the computer. By 1992 the project hopes to have a marketable product. Since cancellation of the Alvey program, Flagship has gained sponsorship from the Esprit Program.
- The Supernum Project of West Germany, with the help of the French Isis program, currently is creating machinery with massively parallel architecture. The parallelism, based on Intel’s 80386 microprocessors, is one of Esprit’s more controversial and ambitious projects. Originally the project was sponsored by the West German government in their supercomputing program. A computer prototype was recently shown at the industry fair in Hanover. It will be marketed in Germany by the end of the year for around $14 million.
- The Supercluster, produced and manufactured by Parsytec GmbH, a small private company, exemplifies Silicon Valley initiative occurring in West Germany. Parsytec has received some financial backing from the West German government for their venture. This start-up firm sells a massively parallel machine that rivals superminicomputers or low-end supercomputers. The Supercluster architecture exploits the 32-bit transputer from Inmos, the T800. Sixteen transputer-based processors in clusters of four are linked together. This architecture is less costly than conventional machines, costing between $230,000 and $320,000.[19] Parsytec has just begun to market its product in America.

Other Nations

The Australia National University recently purchased a Fujitsu VP-1OO. A private service bureau in Australia, Leading Edge, possesses a Cray Research computer. At least two sites in India have supercomputers, one at the Indian Meteorological Centre and one at ISC University. Two Middle Eastern petroleum companies house supercomputers, and Korea and Singapore both have research institutes with supercomputers.

Over half a dozen Canadian universities have high performance computers from CDC, Cray Research, or IBM. Canada’s private sector has also invested in supercomputers. Around 10 firms possess high performance computers. The Alberta government, aside from purchasing a supercomputer and supporting associated services, has helped finance Myrias Computer Corp. A wholly owned U.S. subsidiary, Myrias Research Corp. manufactures the SP-2, a minisupercomputer.

One newly industrialized country is reported to be developing a minisupercomputer of its own. The

---

[18] Julia Vowler, *Supercomputing* Review, “European Transputer-based Projects Issue Challenge to U.S. Supercomputing Supermacy,” November/December 1988, pp. 8-9.

[19] JohnGosh, “A New Transputer Design From West German Startup,” *Electronics*, Mar. 3,1988, pp. 71-72.

first Brazilian rninisupercomputer, claimed to be capable of 150 mips, is planned to be available by the end of 1989. The prototype is a parallel machine with 64 processors, each with 32-bit capacity. The machine will sell for $2.5 million. The Funding Authority of Studies and Projects (FINEP) financed the project, with annual investment around $1 million.

Chapter 3

Networks

Information is the lifeblood of science; communication of that information is crucial to the advance of research and its applications. Data communication networks enable scientists to talk with each other, access unique experimental data, share results and publications, and run models on remote supercomputers, all with a speed, capacity, and ease that makes possible the posing of new questions and the prospect for new answers. Networks ease research collaboration by removing geographic barriers. They have become an invaluable research tool, opening up new channels of communication and increasing access to research equipment and facilities. Most important networking is becoming the indispensable foundation for all other use of information technology in research.

Research networking is also pushing the frontiers of data communications and network technologies. Like electric power, highways, and the telephone, data communications is an infrastructure that will be crucial to all sectors of the economy. Businesses demand on-line transaction processing, and financial markets run on globally networked electronic trading. The evolution of telephony to digital technology allows merging of voice, data, and information services networking, although voice circuits still dominate the deployment of the technology. Promoting scientific research networking—dealing with data-intense outputs like satellite imaging and supercomputer modeling—should push networking technology that will find application far outside of science.

Policy action is needed, if Congress wishes to see the evolution of a full-scale national research and education network. The existing "internet" of scientific networks is a fledgling. As this conglomeration of networks evolves from an R&D enterprise to an operational network, users will demand round-the-clock, high-quality service. Academics, policy makers, and researchers around the world agree on the pressing need to transform it into a permanent infrastructure. This will entail grappling with difficult issues of public and private roles in funding, management, pricing/cost recovery, access, security, and international coordination as well as assuring adequate funding to carry out initiatives that are set by Congress.

Research networking faces two particular policy complications. First, since the network in its broadest form serves most disciplines, agencies, and many different groups of users, it has no obvious lead champion. As a common resource, its potential sponsors may each be pleased to use it but unlikely to give it the priority and funding required to bring it to its full potential. There is a need for clear central leadership, as well as coordination of governments, the private sector, and universities. A second complication is a mismatch between the concept of a national research network and the traditionally decentralized, subsidized, mixed public-private nature of higher education and science. The processes and priorities of mission agency-based Federal support may need some redesigning, as they are oriented towards supporting ongoing mission-oriented and basic research, and may work less well at fostering large-scale scientific facilities and infrastructure that cut across disciplines and agency missions.

In the near term, the most important step is getting a widely connected, operational network in place. But the "bare bones" networks are a small part of the picture. Information that flows over the network, and the scientific resources and data available through the network, are the important payoffs. Key long-term issues for the research community will be those that affect the sort of information available over the network, who has access to it, and how much it costs. The main issue areas for scientific data networking are outlined below:

- research-to develop the technology required to transmit and switch data at very high rates;
- private sector participation-role of the common carriers and telecommunication companies in developing and managing the network and of private information firms in offering services;
- scope—who the network is designed to serve will drive its structure and management;
- access—balancing open use against security and information control and determining who

will be able to gain access to the network for what purpose;
- standards-the role of government, industry, users, and international organizations in setting and maintaining technical standards;
- management-public and private roles; degree of decentralization;
- funding-an operational network will require significant, stable, continuing investment; the financial responsibilities demarcated must reflect the interests of various players, from individual colleges through States and the Federal Government, in their stake in network operations and policies;
- economics-pricing and cost recovery for network use, central to the evolution and management of any infrastructure. Economics will drive the use of the network;
- information services-who will decide what types of services are to be allowed over the network, who is allowed to offer them; and who will resolve information issues such as privacy, intellectual property, fair competition, and security;
- long-term science policy issues—the networks' impacts on the process of science, and on access to and dissemination of valuable scientific and technical information.

THE NATIONAL RESEARCH AND EDUCATION NETWORK (NREN)

> "A universal communications network connected to national and international networks enables electronic communication among scholars anywhere in the world, as well as access to worldwide information sources, special experimental instruments, and computing resources. The network has sufficient bandwidth for scholarly resources to appear to be attached to a world local area network."
>
> EDUCOM, 1988.

> ". . . a national research network to provide a distributed computing capability that links the government, industry, and higher education communities."
>
> **OSTP,** 1987.

> "The goal of the National Research and Education Network is to enhance national competitiveness and productivity through a high-speed, high-quality network infrasoucture which supports a broad set of applications and network services for the research and instructional community."
>
> EDUCOM/NTTF March 1989.

> "The NREN will provide high-speed communication access to over 1300 institutions across the United States within five years. It will offer sufficient capacity, performance, and functionality so that the physical distance between institutions is no longer a barrier to effective collaboration. It will support access to high-performance computing facilities and services . , . and advanced information sharing and exchange, including national file systems and online libraries . . . the NREN will evolve toward fully supported commercial facilities that support a broad range of applications and services."
>
> FRICC, Program Plan
> for the NREN, May 23, 1989.

This chapter of the background paper reviews the status of and issues surrounding data networking for science, in particular the proposed NREN. It describes current Federal activities and plans, and identifies issues to be examined in the full report, to be completed in summer 1990.

The existing array of scientific networks consists of a hierarchy of local, regional and national networks, linked into a whole. In this paper, "NREN" will be used to describe the next generation of the national "backbone" that ties them together. The term "Internet" is used to describe a more specific set of interconnected major networks, all of which use the same data transmission protocols. The most important are NSFNET and its major regional subnetworks, ARPANET, and several other federally initiated networks such as ESNET and NASNET. The term internet is used fairly loosely. At its broadest, the more generic term internet can be used to describe the international conglomeration of networks, with a variety of protocols and capabilities, which have a gateway into Internet; which could include such things as BITNET and MCI Mail.

The Origins of Research Networking

Research users were among the first to link computers into networks, to share information and broaden remote access to computing resources. DARPA created ARPANET in the 1960s for two purposes: to advance networking and data communications R&D, and to develop a robust communications network that would support the data-rich conversations of computer scientists. Building on the resulting packet-switched network technology, other agencies developed specialized networks for their research communities (e.g., ESNET, CSNET NSFNET), Telecommunications and electronic industries provided technology and capacity for these networks, but they were not policy leaders or innovators of new systems. Meanwhile, other research-oriented networks, such as BITNET and Usenet, were developed in parallel by academic and industry users who, not being grantees or contractors of Federal agencies, were not served by the agency-sponsored networks. These university and lab-based networks serve a relatively small number of specialized scientific users, a market that has been ignored by the traditional telecommunications industry. The networks sprang from the efforts of users—academic and other research scientists-and the Federal managers who were supporting them.[1]

The Growing Demand for Capability and Connectivity

Today there are thousands of computer networks in the United States. These networks range from tempoary linkages between modem-equipped[2] desktop computers linked via common carriers, to institution-wide area networks, to regional and national networks, Network traffic moves through different media, including copper wire and optical cables, signal processors and switches, satellites, and the vast common carrier system developed for voice communication. Much of this hodgepodge of networks has been linked (at least in terms of ability to interconnect) into the internet. The ability of any two systems to interconnect depends on their ability to recognize and deal with the form information flows take in each. These "protocols" are sets of technical standards that, in a sense, are the "languages" of communication systems. Networks with different protocols can often be linked together by computer-based "gateways" that translate the protocols between the networks.

National networks have partially coalesced, where technology allows cost savings without losing connectivity. Over the past years, several agencies have pooled funds and plans to support a shared national backbone. The primary driver for this interconnecting and coalescing of networks has been the need for connectivity *among* users. The power of the whole is vastly greater than the sum of the pieces. Substantial costs are saved by extending connectivity while reducing duplication of network coverage. The real payoff is in connecting people, information, and resources. Linking brings users in reach of each other. Just as telephones would be of little use if only a few people had them, a research and education network's connectivity is central to its usefulness, and this connectivity comes both from ability of each network to reach the desks, labs, and homes of its users and the extent to which various networks are, themselves, interconnected.

The Present NREN

The national research and education network can be viewed as four levels of increasingly complex and flexible capability:

- physical wire/fiberoptic common carrier ''highways";
- user-defined, packet-switched networks;
- basic network operations and services; and
- research, education, database, and information services accessible to network users

In a fully developed NREN, all of these levels of service must be integrated. Each level involves different technologies, services, policy issues, research opportunities, engineering requirements, clientele, providers, regulators, and policy issues. A more detailed look at the policy problems can be drawn by separating the NREN into its major components.

Level 1: Physical wire/fiber optic common carrier highways

The foundation of the network is the physical conduits that carry digital signals. These telephone wires, optical fibers, microwave links, and satellites are the physical highways and byways of data transit. They are invisible to the network user. To provide the physical skeleton for the intemet, government, industry, and university network man-

---

[1] John S. Quarterman and Josiah C. Hoskins, "Notable Computer Networks," *Communications of the ACM,* vol *29,* No. 10, October 1986, pp. 932-971; John S. Quarterman, *The Matrix: Networks Around the World,* Digital Press, August 1989

[2] A "Modem" converts information in a computer to a form that a communication system can carry, and vice versa. It also automates SOme Simple functions, such as dialing and answering the phone, detecting and correcting transmission errors.

agers lease circuits from public switched common carriers, such as AT&T, MCI, GTE, and NTN. In doing so they take advantage of the large system of circuits already laid in place by the telecommunications common carriers for other telephony and data markets. A key issue at this level is to what extent broader Federal agency and national telecommunications policies will promote, discourage, or divert the evolution of a research-oriented data network.

Level 2: User-defined subnetworks

The internet is a conglomeration of smaller foreign, regional, State, local, topical, private, government, and agency networks. Generally, these separately managed networks, such as SURANET, BARRNET, BITNET, and EARN, evolved along naturally occurring geographic, topical, or user lines, or mission agency needs. Most of these logical networks emerged from Federal research agency (including the Department of Defense) initiatives. In addition, there are more and more commercial, State and private, regional, and university networks (such as Accunet, Telenet, and Usenet) at the same time specialized and interlined. Many have since linked through the Internet, while keeping to some extent their own technical and socioeconomic identity. This division into small, focused networks offers the advantage of keeping network management close to its users; but demands standardization and some central coordination to realize the benefits of interconnection.

Networks at this level of operations are distinguished by independent management and technical boundaries. Networks often have different standards and protocols, hardware, and software. They carry information of different sensitivity and value. The diversity of these logical subnetworks matters to institutional subscribers (who must choose among network offerings), to regional and national network managers (who must manage and coordinate these networks into an internet), and to users (who can find the variety of alternatives confusing and difficult to deal with). A key issue is the management relationship among these diverse networks; to what extent is standardization and centralization desirable?

Level 3: Basic network operations and services

A small number of basic maintenance tools keeps the network running and accessible by diverse, distributed users. These basic services are software-based, provided for the users by network operators and computer manufacture in operating systems. They include software for password recognition, electronic-mail, and file transfer. These are core services necessary to the operation of any network. These basic services are not consistent across the current range of computers used by research. A key issue is to what extent these services should be standardized, and as important, who should make those decisions.

Level 4: Value-added superstructure: **links to research, education, and information services**

The utility of the network lies in the information, services, and people that the user can access through the network. These value-added services provide specialized tools, information, and data for research and education. Today they include specialized computers and software, library catalogs and publication databases, archives of research data, conferencing systems, and electronic bulletin boards and publishing services that provide access to colleagues in the United States and abroad. These information resources are provided by volunteer scientists and by non-profit, for-profit, international, and government organizations. Some are amateur, poorly maintained bulletin boards; others are mature information organizations with well-developed services. Some are "free"; others recover costs through user charges.

Core policy issues are the appropriate roles for various information providers on the network. If the network is viewed as public infrastructure, what is "fair" use of this infrastructure? If the network eases access to sensitive scientific data (whether raw research data or government regulatory databases), how will this stress the policies that govern the relationships of industry, regulators, lobbyists, and experts? Should profit-seeking companies be allowed to market their services? How can we ensure that technologies needed for network maintenance, cost accounting, and monitoring will not be used inappropriately or intrusively? Who should set prices for various users and services? How will intellectual property rights be structured for electronically available information? Who is responsible for the quality and integrity of the data provided and used by researchers on the network?

Research Networking as a Strategic High Technology Infrastructure

Research networking has dual roles. First, networking is a strategic, high technology infrastructure for science. More broadly applied, data networking **enables** research, education, business, and manufacturing, and improves the Nation's knowledge competitiveness. Second, networking technologies and applications are themselves a substantial growth area, meriting focused R&D.

Knowledge is the commerce of education and research. Today networks are the highways for information and ideas. The y expand access to computing, data, instruments, the research community, and the knowledge they create. Data are expensive (relative to computing hardware) and are increasingly created in many widely distributed locations, by specialized instruments and enterprises, and then shared among many separate users. The more effectively that research information is disseminated to other researchers and to industry, the more effective is scientific progress and social application of technological knowledge. An internet of networks has become a strategic infrastructure for research.

The research networks are also a testbed for data communications technology. Technologies developed through the research networks are likely to enhance productivity of all economic sectors, not just university research. The federally supported Internet has not only sponsored frontier-breaking network research, but has pulled data-networking technology with it. ARPANET catalyzed the development of packet-switching technology, which has expanded rapidly from R&D networking to multibillion-dollar data handling for business and financial transactions. The generic technologies developed for the Internet-hardware (such as high-speed switches) and software for network management, routing, and user interface-will transfer readily into general data-networking applications. Government support for applied research can catalyze and integrate R&D, decrease risk, create markets for network technologies and services, transcend economic and regulatory barriers, and accelerate early technology development and deployment. This would not only bolster U.S. science and education, but would fuel industry R&D and help support the market and competitiveness of the U.S. network and information services industry,

Governments and private industries the world over are developing research networks, to enhance R&D productivity and to create testbeds for highly advanced communications services and technologies. Federal involvement in infrastructure is motivated by the need for coordination and nationally oriented investment, to spread financial burdens, and promote social policy goals (such as furthering basic research).[3] Nations that develop markets in network-based technologies and services will create information industry-based productivity growth.

Federal Coordination of the Evolving Internet

NREN plans have evolved rapidly. Congressional interest has grown; in 1986, Congress requested the Office of Science and Technology Policy (OSTP) to report on options for networking for research and supercomputing.[4] The resulting report, completed in 1987 by the interagency Federal Coordinating Council for Science, Engineering, and Technology (FCCSET), called for a new Federal program to create an advanced national research network by the year 2000.[5] This vision incorporated two objectives: 1) providing vital computer-communications network services for the Nation academic research community, and 2) stimulating networking and communications R&D which would fuel U.S. industrial technology and commerce in the growing global data communications market.

The 1987 FCCSET report, building on ongoing Federal activities, addressed near-term questions over the national network scope, purposes, agency authority, performance targets, and budget. It did not resolve issues surrounding the long-term operation of a network, the role of commercial services in

---

[3] Congressional Budget Office, *New Directions for the Nation's Public Works*, September 1988, p, xiii; CBO, *Federal Policies for Infrastructure Management,* June 1986,

[4] P.L. 99-383, Aug. 21, 1986.

[5] OSTP, *A Research and Development Strategy for High Performance Computing*, Nov 20, 1987

providing network operations and services, or interface with broader telecommunications policies.

A 1988 National Research Council report praised ongoing activities, emphasized the need for coordination, stable funding, broadened goals and design criteria, integrated management, and increased private sector involvement.[6]

FCCSET'S Subcommittee on Networking has since issued a plan to upgrade and expand the network.[7] In developing this plan, agencies have worked together to improve and interconnect several existing networks. Most regional networks were joint creations of NSF and regional consortia, and have been part of the NSFNET world since their inception. Other quasi-private, State, and regional networks (such as CICNET, Inc., and CERFNET) have been started.

Recently, legislation has been reintroduced to authorize and coordinate a national research network.[8] As now proposed, a National Research and Education Network would link universities, national laboratories, non-profit institutions and government research organizations, private companies doing government-supported research and education, and facilities such as supercomputers, experimental instruments, databases, and research libraries. Network research, as a joint endeavor with industry, would create and transfer technology for eventual commercial exploitation, and serve the data-networking needs of research and higher education into the next century.

Players in the NREN

The current Internet has been created by Federal leadership and funding, pulling together a wide base of university commitment, national lab and academic expertise, and industry interest and technology. The NREN involves many public and private actors. Their roles must be better delineated for effective policy. Each of these actors has vested interests and spheres of capabilities. Key players are:

- universities, which house most end users;
- networking industry, the telecommunications, data communications, computer, and information service companies that provide networking technologies and services;
- State enterprises devoted to economic development, research, and education;
- industrial R&D labs (network users); and
- the Federal Government, primarily the national labs and research-funding agencies

Federal funding and policy have stimulated the development of the Internet. Federal initiatives have been well complemented by States (through finding State networking and State universities' institutional and regional networking), universities (by funding campus networking), and industry (by contributing networking technology and physical circuits at sharply reduced rates). End users have experienced a highly subsidized service during this "experimental" stage. As the network moves to a bigger, more expensive, more established operation, how might these relative roles change?

Universities

Academic institutions house teachers, researchers, and students in all fields. Over the past few decades universities have invested heavily in libraries, local computing, campus networks, and regional network consortia. The money invested in campus networking far outweighs the investment in the NSFNET backbone. In general, academics view the NREN as fulfillment of a longstanding ambition to build a national system for the transport of information for research and education. EDUCOM has long labored from the "bottom" up, bringing together researchers and educators who used networks (or believed they could use them) for both research and teaching.

Networking Industry

There is no simple unified view of the NREN in the fragmented telecommunications "industry." The long-distance telecommunications common carriers generally see the academic market as too specialized and risky to offer much of a profit opportunity.

---

[6] National Research Council, *Toward a National Research Network* (Washington, DC, National Academy Press, 1988), especially pp. 25-37.

[7] FCCSET or Federal Coordinating Council for Science, Engineering, and Technology, The *Federal High Performance Computing Program*, Washington, DC, OSTP, Sept. 8, 1989.

[8] S. 1067, "The National High-Performance Computer Technology Act of 1989," May 1989, introduced by Mr. Gore, Hearings were held on June 21, 1989, H.R. 3131, "The National High-Performance Computer Technology Act of 1989," introduced by Mr. Walgren.

However, companies have gained early experience with new technologies and applications by participating in university R&D; it is for this reason that industry has jointly funded the creation and development of NSFNET.

Various specialized value-added common carriers offer packet-switched services. They could in principle provide some of the same services that the NREN would provide, such as electronic mail. They are not, however, designed to meet the capacity requirements of researchers, such as transferring vast files of supercomputer-generated visualizations of weather systems, simulated airplane test flights, or econometric models. Nor can common carriers provide the "reach" to all carriers.

States

The interests of States in research, education, and economic development parallel Federal concerns. Some States have also invested in information infrastructure development. Many States have invested heavily in education and research networking, usually based in the State university system and encompassing, to varying degrees, private universities, State government, and industry. The State is a "natural" political boundary for network financing. In some States, such as Alabama, New York, North Carolina, and Texas, special initiatives have helped create statewide networks.

Industry Users

There are relatively few industry users of the internet; most are very large R&D-intensive companies such as IBM and DEC, or small high-technology companies. Many large companies have internal business and research networks which link their offices and laboratories within the United States and overseas; many also subscribe to commercial services such as MCI Mail. However, these proprietary and commercial networks do not provide the internet's connectivity to scientists or the high bandwidth and services so useful for research communications. Like universities and national labs, companies are a part of the Nation's R&D endeavor; and being part of the research community today includes being "on" the internet. Appropriate industry use of the NREN should encourage interaction of industry, university, and government researchers, and foster technology transfer. Industry internet users bring with them their own set of concerns such as cost accounting, proper network use, and information security. Other non-R&D companies, such as business analysts, also are likely to seek direct network connectivity to universities, government laboratories, and R&D-intensive companies.

Federal

Three strong rationales-support of mission and basic science, coordinating a strategic national infrastructure, and promotion of data-networking technology and industrial productivity-drive a substantial, albeit changing, Federal involvement. Another more modest goal is to rationalize duplication of effort by integrating, extending, and modernizing existing research networks. That is in itself quite important in the present Federal budgetary environment. The international nature of the network also demands a coherent national voice in international telecommunications standardization. The Internet's integration with foreign networks also justifies Federal concern over the international flow of militarily or economically sensitive technical information. The same university-government-industry linkages on a domestic scale drive Federal interests in the flow of information.

Federal R&D agencies' interest in research networking is to enhance their external research support missions. (Research networking is a small, specialized part of agency telecommunications. It is designed to meet the needs of the research community, rather than agency operations and administrative telecommunications that are addressed in FTS 2000.) The hardware and software communications technologies involved should be of broad commercial importance. The NREN plans reflect national interest in bolstering a serious R&D base and a competitive industry in advanced computer communications.

The dominance of the Federal Government in network development means that Federal agency interests ha-e strongly influenced its form and shape. Policies can reflect Federal biases; for instance, the limitation of access to the early ARPANET to ARPA contractors left out many academics, who consequent y created their own grass-roots, lower-capability BITNET.

International actors are also important. As with the telephone system, the internet is inherently international. These links require coordination, for example for connectivity standards, higher level network management, and security. This requirement implies the need for Federal level management and policy.

The NREN in the International Telecommunications Environment

The nature and economics of an NREN will depend on the international telecommunications context in which it develops. Research networks are a leading edge of digital network technologies, but are only a tiny part of the communications and information services markets.

The 1990s will be a predominantly digital world; historically different computing, telephony, and business communications technologies are evolving into new information-intensive systems. Digital technologies are promoting systems and market integration. Telecommunications in the 1990s will revolve around flexible, powerful, "intelligent" networks. However, regulatory change and uncertainty, market turbulence, international competition, the explosion in information services, and significant changes in foreign telecommunications policies, all are making telecommunications services more turbulent. This will cloud the research network's long-term planning.

High-bandwidth, packet-switched networking is at persent a young market in comparison to commercial telecommunications. Voice overwhelmingly dominates other services (e.g. fax, e-mail, on-line data retrieval). While flexible, hybrid voice-data services are being introduced in response to business demand for data services, the technology base is optimized for voice telephony.

Voice communications brings to the world of computer telecommunications complex regulatory and economic baggage. Divestiture of the AT&T regulated monopoly opened the telecommunications market to new entrants, who have slowly gained long-haul market share and offered new technologies and information services. In general, however, **the** post-divestiture telecommunications industry remains dominated by the descendants of old AT&T, and most of the impetus for service innovations comes from the voice market. One reason is uncertainty about the legal limits, for providing information services, imposed on the newly divested companies. (In comparison, the computer industry has been unregulated. With the infancy of the technology, and open markets, computer R&D has been exceptionally productive,) **A crucial concern for long-range NREN planning is that scientific and educational needs might be ignored among the regulations, technology priorities, and economics of a telecommunications market geared toward the vast telephone customer base.**

POLICY ISSUES

The goal is clear; but the environment is complex, and the details will be debated as the network evolves

There is substantial agreement in the scientific and higher education community about the pressing national need for a broad-reaching, broad-bandwidth, state-of-the-art research network, The existing Internet provides vital communication, research, and information services, in addition to its concomitant role in pushing networking and data handling technology, Increasing demand on network capacity has quickly saturated each network upgrade. In addition, the fast-growing demand is overburdening the current informal administrative arrangements for running the Internet. Expanded capability and connectivity will require substantial budget increases. The current network is adequate for broad e-mail service and for more restricted file transfer, remote logon, and other sophisticated uses. Moving to gigabit bandwidth, with appropriate network services, will demand substantial technological innovation as well as investment.

There are areas of disagreement and even broader areas of uncertainty in planning the future national research network. There are several reasons for this: the immaturity of data network technology, services, and markets; the Internet's nature as strategic infrastructure for diverse users and institutions; and the uncertainties and complexities of overriding telecommunications policy and economics.

First, the current Internet is, to an extent, an experiment in progress, similar to the early days of the telephone system. Technologies, uses, and potential markets for network services are still nascent.

Patterns of use are still evolving; and a reliable network has reached barely half of the research community. Future uses of the network are difficult to identify; each upgrade over the past 15 years has brought increased value and use as improved network capacity and access have made new applications feasible.

The Internet is a conglomeration of networks that grew up ad hoc. Some, such as ARPANET, CSNET, and MFENET, were high-quality national networks supported by substantial Federal funding. Other smaller networks were built and maintained by the late-night labors of graduate students and computer centers operators. One of these, BITNET, has become a far-reaching and widely used university network, through the coordination of EDUCOM and support of IBM. The Internet has since become a more coherent whole, under Federal coordination led by NSF and DARPA and advised by the Internet Activities Board. Improvements in service and connectivity have been astounding. Yet the patchwork nature of the Internet still dominates; some campus and regional networks are high quality and well maintained; others are lower speed, less reliable, and reach only a few institutions in their region. Some small networks are gatewayed into the Internet; others are not. This patchwork nature limits the effectiveness of the Internet, and argues for better planning and stronger coordination.

Second, the network is a strategic infrastructure, with all the difficulties in capitalizing, planning, financing, and maintaining that seem to attend any infrastructure.[9] Infrastructures tend to suffer from a "commons" problem, leading to continuing underinvestment and conflict over centralized policy. By its nature the internet has many diverse users, with diverse interests in and demands on the network. The network's value is in linking and balancing the needs of these many users, whether they want advanced supercomputer services or merely e-mail. Some users are network-sophisticated, while many users want simple, user-friendly communications. This diversity of users complicates network planning and management. The scope and offerings of the network must be at least sketched out before a management structure appropriate to the desired mission is established.

Third, the network is part of the telecommunications world, rampant with policy and economic confusion. The research community is small, with specialized data needs that are subsidiary to larger markets. It is not clear that science's particular networking needs will be met.

Planning Amidst Uncertainty

Given these three large uncertainties, there is no straightforward or well-accepted model for the "best" way to design, manage, and upgrade the future national research network. Future network use will depend on cost recovery and charging practices, about which very little is understood. These uncertainties should be accommodated in the design of network management as well as the network itself.

One way to clarify NREN options might be to look at experiences with other infrastructures (e.g., waterways, telephones, highways) for lessons about how different financing and charging policies affect who develops and deploys technology, how fast technology develops, and who has access to the infrastructure. Additionally, some universities are beginning trials in charging for network services; these should provide experience in how various charging practices affect usage, technology deployment and upgrading, and the impacts of network use policies on research and education at the level of the institution.

Table 3-1 lists the major areas of agreement and disagreement in various "models" of the proper form of network evolution.

Network Scope and Access

Scope

Where should an NREN reach: beyond research-intensive government laboratories and universities to all institutions of higher education? high schools? nonprofit and corporate labs? Many believe that eventually— perhaps in 20 years-de facto data networking will provide universal linkage, akin to a sophisticated phone system.

---

[9] Congressional Budget Office, *New Directions for the Nation's Public Works,* September 1988; National Council on Public Works Improvement, *Fragile Foundation A Report on America's Public Works,* Washington, IX, February 1988.

Table 3-1-Principal Policy Issues in Network Development

Major areas of agreement	Major areas of disagreement and uncertainty
Scope and access	
1. The national need for a broad state-of-the-art research network that links basic research, government, and higher education.	la. The exact scope of the NREN; whether and how to control domestic and foreign access. lb. Hierarchy of network capability. Cost and effort limit the reach of state-of-the-art networking; an "appropriate networking" scenario would have the most intensive users on a leading edge network and less demanding users on a lower-cost network that suffices for their needs. Where should those lines be drawn, and who should draw them? How can the Federal Government ensure that the gap between leading edge and casual is not too large, and that access is appropriate and equitable?
Policy and management structure	
2. The need for a more formal mechanism for planning and operating the NREN, to supersede and better coordinate informal interagency cooperation and ad hoc university and State participation, and for international coordination.	2a. The form and function of an NREN policy and management authority; the extent of centralization, particularly the role of Federal Government; the extent of participation of industry users, networking industry, common carriers, and universities in policy and operations; mechanisms for standard setting,
Financing and coat recovery	
3. The desirability of moving from the current "market-establishing" environment of Federal and State grants and subsidies, with services "free" to users, to more formal cost recovery, shifting more of the cost burden and financial incentives to end users.	3a. How the transition to commmercial operations and charging can and should be made; more generally, Federal-private sector roles in network policy and pricing; how pricing practices will shape access, use, and demand.
Network Use	
4. The desirability of realizing the potential of a network; the need for standards and policies to link to information services, databases, and nonresearch networks.	4a. Who should be able to use the network for what purposes, and at what entry rest; the process of guiding economic structure of services, subsidies, price of for multi-product services; intellectual property policies.

S0URCE: Office of Technology Assessment, 1989.

The appropriate breadth of the network is unlikely to be fully resolved until more user communities gain more experience with networking, and a better understanding is gained of the risks and benefits of various degrees of network coverage. A balance must be struck in network scope, which provides a small network optimized for special users (such as scientists doing full-time, computationally intensive research) and also a broader network serving more diverse users. The scope of the internet, and capabilities of the networks encompassed in the internet, will need to balance the needs of specialized users without diluting the value for top-end and low-end users. NREN plans, standards, and technology should take into account the possibility of later expansion and integration with other networks and other communities currently not linked up. After-the-fact technical patches are usually inefficient and expensive. This may require more government participation in standard-setting to make it feasible for currently separated communities, such as high schools and universities, to interconnect later on.

Industry-academic boundaries are of particular concern. Interconnection generally promotes research and innovation. Companies are dealing with risk of proprietary information release by maintaining independent corporate networks and by restricting access to open networks. How can funding and pricing be structured to ensure that for-profit companies bear an appropriate burden of network costs?

Access

Is it desirable to restrict access to the internet? Who should control access? Open access is desired by many, but there are privacy, security, and commercial arguments for restricting access. Restricting access is difficult, and is determined more by access controls (e.g., passwords and monitoring) on the computers that attach users to the network, than by the network itself. Study is needed on whether and how access can be controlled by technical fixes within the network, by computer centers attached to the network, informal codes of behavior, or laws,

Another approach is not to limit access, but minimize the vulnerability of the network—and its information resources and users-to accidents or malice. In comparison, essentially anyone who has a modest amount of money can install a phone, or use a public phone, or use a friend's phone, and access the national phone system. However, criminal, fraudulent, and harassing uses of the phone system are illegal. Access is unrestricted, but use is governed.

Controlling International Linkages

Science, business, and industry are international; their networks are inherently international. It is difficult to block private telecommunications links with foreign entities, and public telecommunications is already international. However, there is a fundamental conflict between the desire to capture information for national or corporate economic gain, and the inherent openness of a network. Scientists generally argue that open network access fosters scientifically valuable knowledge exchange, which in turn leads to commercially valuable innovation.

Hierarchy of Network Capability

Investment in expanded network access must be balanced continually with the upgrading of network performance, As the network is a significant competitive advantage in research and higher education, access to the "best" network possible is important. There are also technological considerations in linking networks of various performance levels and various architectures. There is already a consensus that there should be a separate testbed or research network for developing and testing new network technologies and services, which will truly be at the cutting edge (and therefore also have the weaknesses of cutting edge technology, particularly unreliability and difficulty of use).

Policy and Management Structure

Possible management models include: federally chartered nonprofit corporations, single lead agencies, interagency consortium, government-owned contractor operations, commercial operations; and Tennessee Valley Authority, Atomic Energy Commission, the NSF Antarctic Program, and Fannie Mae. What are the implications of various scenarios for the nature of traffic and users?

Degree of Centralization

What is the value of centralized, federally accountable management for network access control, traffic management and monitoring, and security, compared to the value of decentralized operations, open access and traffic? There are two key technical questions here: to what extent does **network** technology limit the amount of control that can be exerted over access and traffic content? To what extent does technology affect the strengths and weaknesses of centralized and decentralized management?

Mechanisms for Interagency Coordination

Interagency coordination has worked well so far, but with the scaling up of the network, more formal mechanisms are needed to deal with larger budgets and to more tightly coordinate further development.

Coordination With Other Networks

National-level resources allocation and planning must coordinate with interdependent institutional and mid-level networking (the other two legs of networking).

Mechanisms for Standard Setting

Who should set standards, when should they be set, and how overarching should they be? Standards at some common denominator level are absolutely necessary to make networks work. But excessive standardization may deter innovation in network technology, applications and services, and other standards.

Any one set of standards usually is optimal for some applications or users, but not for others. There are well-established international mechanisms for formal standards-setting, as well as strong international involvement in more informal standards development. These mechanisms have worked well, albeit slowly. Early standard-setting by agencies and their advisers accelerated the development of U.S. networks, In many cases the early established standards have become, with some modification, de facto national and even international standards. This is proving the case with ARPANET's protocol suite, TCP/IP. However, many have complained that agencies' relatively precipitous and closed standards determination has resulted in less-than-satisfactory standards. NREN policy should embrace standards-setting. Should it, however, encourage wider participation, especially by industry, than has been the case? U.S. policy must balance the need for international compatibility with the furthering of national interests.

Financing and Cost Recovery

How can the capital and operating costs of the NREN be met? Issues include subsidies, user or access charges, cost recovery policies, and cost accounting. As an infrastructure that spans disciplines and sectors, the NREN is outside the traditional grant mechanisms of science policy. How might NREN economics be structured to meet costs and achieve various policy goals, such as encouraging widespread yet efficient use, ensuring equity of access, pushing technological development while maintaining needed standards, protecting intellectual property and sensitive information while encouraging open communication, and attracting U.S. commercial involvement and third-party information services?

Creating a Market

One of the key issues centers around the extent to which deliberate creation of a market should be built into network policy, and into the surrounding science policy system. There are those who believe that it is important that the delivery of network access and services to academics eventually become a commercial operation, and that the current Federal subsidy and apparently "free" services will get academics so used to free services that there will never be a market. How do you gradually create an information market, for networks, or for network-accessible value-added services?

Funding and Charge Structures

Financing issues are akin to ones in more traditional infrastructures, such as highways and waterways. These issues, which continue to dominate infrastructure debates, are Federal private sector roles and the structure of Federal subsidies and incentives (usually to restructure payments and access to infrastructure services). Is there a continuing role for Federal subsidies? How can university accounting, OMB circular A-21, and cost recovery practices be accommodated?

User fees for network access are currently charged as membership/access fees to institutions. End users generally are not charged. In the future, user fees may combine access/connectivity fees, and use-related fees. They may be secured via a trust fund (as is the case with national highways, inland waterways, and airports), or be returned directly to operating authorities. A few regional networks (e.g., CICNET, Inc.) have set membership/connectivity fees to recover full costs. Many fear that user fees are not adequate for full funding/cost recovery.

Industry Participation

Industry has had a substantial financial role in network development. Industry participation has been motivated by a desire to stay abreast of data-networking technology as well as a desire to develop a niche in potential markets for research networking. It is thus desirable to have significant industry participation in the development of the NREN. Industry participation does several things: industry cost sharing makes the projects financially feasible; industry has the installed long-haul telecommunications base to build on; and industry involvement in R&D should foster technology transfer and, generally, the competitiveness of U.S. telecommunications industry. Industry in-kind contributions to NSFNET, primarily from MCI and IBM, are estimated at $40 million to $50 million compared to NSF's 5 year, $14 million budget.[10] It is anticipated that the value of industry cost sharing (e.g., donated switches, lines, or software) for NREN would be on the order of hundreds of millions of dollars.

---

[10] Eliot Marshal, "NSF Opens High-Speed Computer Network," *Science*, p. 22.

Network Use

Network service offerings (e.g., databases and database searching services, news, publication, and software) will need some policy treatment. There need to be incentives to encourage development of and access to network services, yet not unduly subsidize such services, or compete with private business, while maintaining quality control. Many network services used by scientists have been "free" to the end user.

Economic and legal policies will need to be clarified for reference services, commercial information industry, Federal data banks, university data resources, libraries, publishers, and generally all potential services offered over the network.[11] These policies should be designed to encourage use of services, while allowing developers to capture the potential benefits of network services and ensure legal and economic incentives to develop and market network services.

Longer Term Science Policy Issues

The near-term technical implementation of the NREN is well laid out. However, longer-term policy issues will arise as the national network affects more deeply the conduct of science, such as:

- patterns of collaboration, communication and information transfer, education, and apprenticeship;
- intellectual property, the value and ownership of information;
- export control of scientific information
- publishing of research results
- the "productivity" of research and attempts to measure it
- communication among scientists, particularly across disciplines and between university, government, and industry scientists.
- potential economic and national security risks of international scientific networking, collaboration, and scientific communication;
- equity of access to scientific resources, such as facilities, equipment, databases, research grants, conferences, and other scientists. (Will a fully implemented NREN change the concentration of academic science and Federal funding in a limited number of departments and research universities, and of corporate science in a few large, rich corporations; what might be the impacts of networks on traditional routes to scientific priority and prestige?)
- controlling scientific information flow. What technologies and authority to control network-resident scientific information? How might these controls affect misconduct, quality control, economic and corporate proprietary protection, national security, and preliminary release of tentative or confidential research information that is scientifically or medically sensitive?
- cost and capitalization of doing research; to what extent might networking reduce the need for facilities or equipment?
- oversight and regulation of science, such as quality control, investigations of misconduct, research monitoring, awarding and auditing of government grants and contracts, data collection, accountability, and regulation of research procedures.[12] Might national networking enable or encourage new oversight roles for governments?
- the access of various publics to scientists and research information;
- the dissemination of scientific information, from raw data, research results, drafts of papers through finished research reports and reviews; might some scientific journals be replaced by electronic reports?
- legal issues, data privacy, ownership of data, copyright. How might national networking interact with trends already underway in the scientific enterprise, such as changes in the nature of collaboration, sharing of data, and impacts of commercial potential on scientific research? Academic science traditionally has emphasized open and early communication, but some argue that pressures from competition for research grants and increasing potential for commercial value from basic research have

---

[11] OMB, Circular A-130, 50 Federal Register 52730 (Dec. 24, 1985); A. 130. H.R. 2381, The Information Policy Act of 1989, which restates the role of OMB and policies on government reformation dissemination.

[12] U.S. Congress, Office of Technology Assessment, *The Regulatory Environment for Science*, OTA-TM-SET-34 (Washington, DC: U.S. Government Printing Office, February 1986).

dampened free communication. Might networks counter, or strengthen, this trend?

Technical Questions

Several unresolved technical challenges are important to policy because they will help determine who has access to the network for what purposes. Such technical challenges include:

- standards for networks and network-accessible information services;
- requirements for interface to common carriers (local through international);
- requirements for interoperability across many different computes;
- improving user interfaces;
- reliability and bandwidth requirements;
- methods for measuring access and usage, to charge users that will determine who is most likely to pay for network operating costs; and
- methods to promote security, which will affect the balance-between network and information vulnerability, privacy, and open access.

Federal Agency Plans: FCCSET/FRICC

A recently released plan by the Federal Research Internet Coordinating Committee (FRICC) outlines a technical and management plan for NREN.[13] This plan has been incorporated into the broader FCCSET implementation plan. The technical plan is well thought through and represents further refinement of the NREN concept. The key stages are:

Stage 1: upgrade and interconnect existing agency networks into a jointly funded and managed T1 (1.5 Mb/s) National Networking Testbed.[14]
Stage ***2:*** integrate national networks into a T3 (45 Mb/s) backbone by 1993.
Stage 3: push a technological leap to a multigigabit NREN starting in the mid-1990s.

The proposal identifies two parts of an NREN, an operational network and networking R&D. A service network would connect about 1,500 labs and universities by 1995, providing reliable service and rapid transfer of very large data streams, such as are found in interactive computer graphics, in apparent real time. The currently operating agency networks would be integrated under this proposal, to create a shared 45Mb/s service net by 1992. The second part of the NREN would be R&D on a gigabit network, to be deployed in the latter 1990s. The first part is primarily an organizational and financial initiative, requiring little new technology. The second involves major new research activity in government and industry.

The “service” initiative extends present activities of Federal agencies, adding a governance structure which includes the non-Federal participants (regional and local networking institutions and industry), in a national networking council, It formalizes what are now ad-hoc arrangements of the FRICC, and expands its scale and scope. Under this effort, virtually all of the Nation’s research and higher education communities will be interconnected. Traffic and traffic congestion will be managed via priority routing, with service for participating agencies guaranteed via “policy” routing techniques. The benefits will be in improving productivity for researchers and educators, and in creating and demonstrating the demand for networks and network services to the computing, telecommunications, and information industries.

The research initiative (called stage 3 in the FCCSET reports) is more ambitious, seeking support for new research on communications technologies capable of supporting a network that is at least a thousand times faster than the 45Mb/s net. Such a net could use the currently unused capabilities of optical fibers to vastly increase effective capability and capacity, which are congested by today’s technology for switching and routing, and support the next generation of computers and communications applications. This effort would require a substantial Federal investment, but could invigorate the national communication technology base, and boost the long-term economic competitiveness of

---

[13] FRICC, *Program Plan for the National Research and Education Network,* May *23, 1989.* FRICC has members from DHHS, DOE, DARPA, USGS, NASA, NSF, NOAA, and observers from the Internet Activities Board. FRICC is an informal committee that grew out of agencies’ shared interest in coordinating related network activities and avoiding duplication of resources. As the de facto interagency coordination forum, FRICC was asked by NSF to prepare the NREN program plan.

[14] See also *NYSERNET NOTE,* vol. 1, No. 1, Feb. 6, 1989. NYSERNET has been awarded a multimillion-dollar contract from DARPA to develop the National Networking Testbed.

the telecommunications and computing industries. The gigabit network demonstration can be considered similar to the Apollo project for communications technologies, albeit on a smaller and less spectacular scale. Technical research needed would involve media, switches, network design and control software, operating systems in connected computers, and applications.

There are several areas where the FRICC management plan-and other plans-is unclear. It calls for, but does not detail any transition to commercial operations. It does not outline potential structures for long-term financing or cost recovery. And the national network council's formal area of responsibility is limited to Federal agency operations. While this scope is appropriate for a Federal entity, and the private sector has participated influentially in past Federal FRICC plans, the proposed council does not encompass all the policy actors that need to participate in a coordinated national network. The growth of non-Federal networks demonstrates that some interests—such as smaller universities on the fringes of Federal-supported R&D-have not been served. The FRICC/FCCSET implementation plan for networking research focuses on the more near-term management problems of coordinated planning and management of the NREN. It does not deal with two extremely important and complex interfaces. At the most fundamental level, the common carriers, the network is part of the larger telecommunications labyrinth with all its attendant regulations, vested interests, and powerful policy combatants. At the top level, the network is a gateway into a global information supermarket. This marketplace of information services is immensely complex as well as potentially immensely profitable, and policy and regulation has not kept up with the many new opportunities created by technology.

The importance of institutional and mid-level networking to the performance of a national network, and the continuing fragmentation and regulatory and economic uncertainty of lower-level networking, signals a need for significant policy attention to coordinating and advancing lower-level networking. While there is a formal advisory role for universities, industry, and other users in the FRICC plan, it is difficult to say how and how well their interests would be represented in practice. It is not clear what form this may take, or whether it will necessitate some formal policy authority, but there is need to accommodate the interests of universities (or some set of universities), industry research labs, and States in parallel to a Federal effort. The concerns of universities and the private sector about their role in the national network are reflected in EDUCOM'S proposal for an overarching Federal-private nonprofit corporation, and to a lesser extent in NRI's vision. The FRICC plan does not exclude such a broader policy-setting body, but the current plan stops with Federal agency coordination.

Funding for the FRICC NREN, based on the analysis that went into the FCCSET report, is proposed at $400 million over 5 years, as shown below. This includes all national backbone Federal spending on hardware, software, and research, which would be funneled through DARPA and NSF and overseen by an interagency council. It includes some continued support for mid-level or institutional networking, but not the value of any cost sharing by industry, or specialized network R&D by various agencies. This budget is generally regarded as reasonable and, if anything, modest considering the potential benefits (see table 3-2).[15]

AREN Management Desiderata

All proposed initiatives share the policy goal of increasing the nation's research productivity and creating new opportunities for scientific collaboration. As a technological catalyst, an explicit national NREN initiative would reduce unacceptably high levels of risk for industry and help create new markets for advanced computer-communications services and technologies. What is needed now is a sustained Federal commitment to consolidate and fortify agency plans, and to catalyze broader national involvement. The relationship between science-oriented data networking and the broader telecommunications world will need to be better sorted out before the NREN can be made into a partly or fully commercial operation. As the engineering challenge of building a fully national data network is surmounted, management and user issues of economics, access, and control of scientific information will rise in importance.

---

[15] For example, National Research Council, *Toward a National Research Network* (Washington, DC: National Academy Press, 1988), pp. 2~31

Table 3-2-Proposed NREN Budget ($ millions)

	FY90	FY91	FY92	FY93	FY94
FCCSET Stage 1 & 2 (upgrade; NSF)	14	23	55	50	50
FCCSET Stage 3 (gigabit+; DARPA)	16	27	40	55	**60**
Total	30	50	95	95	110
S. 1067 authorization	50	50	100	100	100
HR. 3131 authorization	50	50	100	100	100

SOURCE: Office of Technology Assessment, 1989.

The NREN is a strategic, complex infrastructure which requires long-term planning. Consequently, network management should be stable (insulated from too much politics and budget vagaries), yet allow for accountability, feedback, and course correction. It should be able to leverage funding, maximize cost effciency, and create incentives for commercial networks. Currently, there is no single entity that is big enough, risk-protected enough, and regulatory-free enough to make a proper national network happen. While there is a need to formalize current policy and management, there is concern that setting a strong federally focused structure in place might prevent a move to a more desirable, effective, appropriate management system in the long run.

There is need for greater stability in NREN policy. The primary vehicle has been a voluntary coordinating group, the FRICC, consisting of program officers from research-oriented agencies, working within agency missions with loose policy guidance from the FCCSET. The remarkable cooperation and progress made so far depends on a complex set of agency priorities and budget fortunes, and continued progress must be considered uncertain.

The pace of the resolution of these issues will be controlled initially by the Federal budget of each participating agency. While the bulk of the overall investment rests with midlevel and campus networks, it cannot be integrated without strong central coordination, given present national telecommunications policies and market conditions for the required network technology. The relatively modest investment proposed by the initiative can have major impact by providing a forum for public-private cooperation for the creation of new knowledge, and a robust and willing experimental market to test new ideas and technologies.

For the short term there is a clear need to maintain the Federal initiative, to sustain the present momentum, to improve the technology, and coordinate the expanding networks. The initiative should accelerate the aggregation of a sustainable domestic market for new information technologies and services. These goals are consistent with a primary purpose of improving the data communications infrastructure for U.S. science and engineering,