Chapter 1

High-Performance Computing and Information Infrastructure for Science and Engineering

Introduction

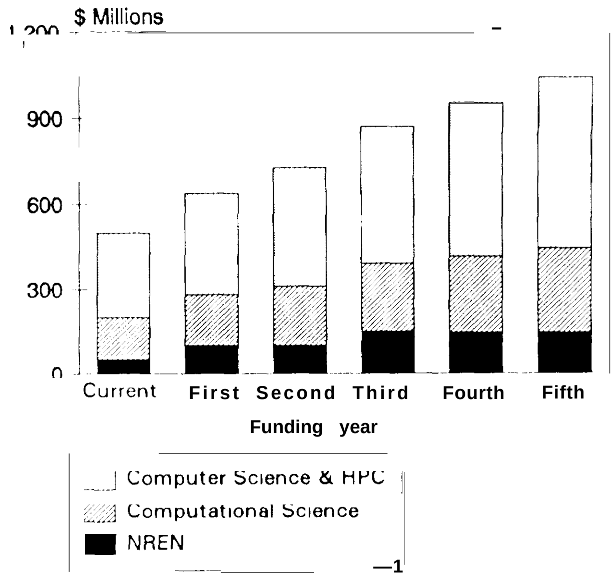
Information technology is a critical element for science and engineering. The United States is building a nationwide computer-communication infrastructure to provide high-speed data services to the R&D community, but the mere installation of hardware is not enough. Whether very fast data communication networks and high-performance computers deliver their promised benefits will depend on the institutions, processes, and policies that are established to design and manage the new system.

OTA was asked by the House Committee on Science, Space, and Technology and the Senate Committee on Commerce and Transportation to examine the role that high-performance computing, networking, and information technologies are playing in science and engineering, and to analyze the need for Federal action. An OTA background paper, released in September 1989, explored and described some key issues.[1] This background paper examines high-performance computing as part of the infrastructure proposed in the Office of Science and Technology Policy (OSTP) initiative. A detailed OTA report on the National Research and Education Network (NREN) is scheduled for release later in 1991.

Six years ago, Congress directed the National Science Foundation (NSF) to establish an Advanced Scientific Computing Program designed to increase access by researchers to high-performance computing. That program resulted in the establishment of five national centers for scientific supercomputing. Since then, one of the centers has been left unfunded; but the other four are still operating.

During the last 5 years, legislation has been introduced in Congress calling for the establishment of a high-capacity, broadband, advanced national data communications network for research. Over the years, congressional interest has grown as this concept has evolved into a plan for an integrated national research and education network (NREN) consisting of an advanced communication network linked to a variety of computational and information facilities and services.

Figure I—Estimated Proposed Funding Levels for Federal High-Performance Computing Program



SOURCE: Office of Science and Technology Policy, *The Federal High-Performance Computing Program* (Washington, DC: Office of Science and Technology Policy, 1989), app. C, p. 26.

In September 1989, at the request of Congress, OSTP submitted a draft plan developed by the Federal Coordinating Council for Science, Engineering, and Technology (FCCSET). The plan called for a ''National High-Performance Computing Initiative' that includes both a national network and initiatives to advance high-performance computing (see figure 1). In testimony to the 101st Congress, the director of OSTP stated that this Initiative was among the top priorities on the science agenda. On June 8, 1990, the National Science Foundation announced a $15.8 million, 3-year research effort aimed at funding 5 gigabit-speed testbed experimental networks. These test networks are the first step in developing a high-speed nation-

[1] U.S. Congress, Office of Technology Assessment, *High Performance Computing & Networking for Science*, OTA-BP-CIT-59 (Washington, DC: U.S. Government Printing Office, September 1989).

wide broadband advanced communication network in collaboration with the Defense Advanced Research Project Agency (DARPA).

OSTP set forth its plans for a Federal High-Performance Computing and Communications Program (HPCC) in a document released on February 4, 1991, supporting the President's Fiscal Year 1992 budget.[2] The Program proposes to invest $638 million in fiscal year 1992, an increase of about 30 percent over the 1991 level. These funds will support activities in four program areas: 1) high-performance computing systems; 2) advanced software technology and algorithms; 3) National Research and Education Network; and 4) basic research and human resources.

High-Performance Computing: A Federal Concern

Concern about information technology by high-level policymakers is a recent phenomenon. Researchers who see the importance of data exchange have managed to secure funding for computers and communications out of the limited Federal agency research budgets. Agencies such as DARPA and NSF have quietly developed computer and network-related programs without major administration initiatives or congressional actions. But the atmosphere is now different for several reasons.

First, researchers cannot consistently obtain needed information resources because of the cost. High-end scientific computers cost several million dollars to purchase and millions more per year to operate. Universities grew reluctant in the late 1970s and early 1980s to purchase these systems with their own funds, and government investment in computers slowed. In the meantime, researchers learned more about the use of high-performance computing. The machines became more powerful, doubling in speed about every 2 years. The scientific community slowly became aware of the lost opportunities caused by lack of access to high-performance computers and other powerful information technologies.

Second, information resources-computers, databases, and software-are being shared among disciplines, institutions, and facilities. These are being linked as common resources through networks to users at desktop workstations. A need has grown for better coordination in the design and operation of these systems; this will be particularly important for a national data communications network.

Third, although the U.S. computer industry is relatively strong, there is concern about increasing competition from foreign firms, particularly Japanese. Over the last decade, the Japanese Government has supported programs, such as the Fifth Generation Project (it is now planning a Sixth Generation initiative) and National Superspeed Computer Project, designed to strengthen the Japanese position in high-performance computing. During the last 2 years there have been difficult trade negotiations between the United States and Japan over supercomputer markets in the respective countries. This has raised concern about the economic and strategic importance of a healthy U.S. high-perfomance computing industry.[3]

Fourth, concern for the Japanese challenge in high-performance computing goes beyond the competitiveness of the U.S. supercomputer industry. Computational simulation in engineering design and manufacturing is becoming a major factor in maintaining a competitive posture in high-technology industries such as automotive, aerospace, petroleum, electronics, and pharmaceuticals. It is in the availability and application of high-performance computing to increase productivity and improve product quality where the greatest future economic benefits may lie.

Finally, the infrastructure of this interlinked set of technologies is considered by some to be a strong basis for the development of a universal broadband information system. A very high-capacity digital communication network and information services, as visualized, would carry entertainment, educational, and social services to the home and support a broad range of business and education services. A

[2] OSTP, *Grand Challenges: High-Performance Computing and Communications,* (Washington DC: OSTP, 1991), p. 57.

[3] High-performance computers are considered to be strategically important to the United States because of the central role computers play in the economy, security, manufacturing, and research and development. See Office of Science and Technology Policy, *The Federal High Performance Computing Program* (Washington, *DC:* Office of Science and Technology Policy, 1989), pp. 12-13. See U.S. Congress, Office of Technology Assessment, *Making Things Better: Competing in Manufacturing,* OTA-ITE-443 (Washington, DC: U.S. Government Printing Office, February 1990), p. 241, for a comprehensive view of the U.S. competitive position in manufacturing, including computer-related industries. An assessment of the status of the U.S. supercomputing industry will be included in a forthcoming OTA report on Japan and International Trade.

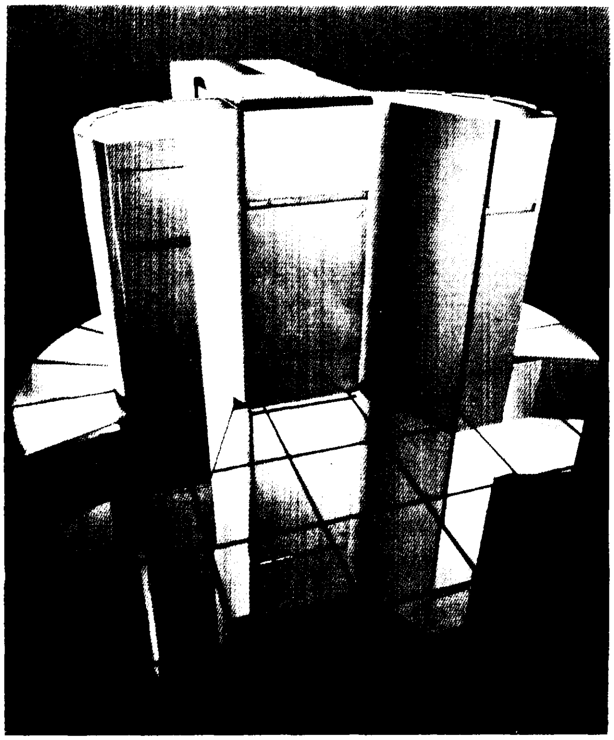


Photo credit: Cray Researeh, Inc.

The CRAY Y-MP/832 computer system is the top-of-the-line supercomputer of Cray Research, Inc. It contains 8 central processors and 32 million 64-bit words of memory.

nationwide network for research and education could be a starting point, and could gradually broaden to this vision.

Multiple Goals for an Initiative

The supporting arguments for a Federal High-Performance computing/networking initiative center on three objectives:

1. *To advance U.S. research and development* critical to U.S. industry, security, and education by providing researchers with the most powerful computers and communication systems available. This objective is based on a vision of computers and data communication technologies forming a basic infrastructure for supporting research. This goal has been proposed in several reports and policy papers.[4]
2. *To strengthen the U.S. computer industry* (particularly the high-performance computers and high-speed telecommunications) by testing new system concepts and developing new techniques for applications. Federal Coordinating Council for Science, Engineering, and Technology (FCCSET) and the Institute of Electrical and Electronics Engineers, among others, strongly endorse this views
3. *To enhance U.S. economic and social strength* by stimulating the development of a universal information infrastructure through development of new technologies that could serve as a system prototype.[6]

Strong sentiment exists among some Members of Congress for each of these three objectives. Furthermore, the goals are closely related and nearly inseparable-most discussion and proposals for computing and networking programs reflect elements of all three.

Not everyone in Congress or the executive branch agrees that all goals are equally important or even appropriate for the Federal Government. Some consider the current level of government spending to advance scientific knowledge to be adequate, and they believe that other needs have higher priority. Others point out that since information technology is now central to all R&D, it is important to create a modern information infrastructure in order to realize the benefits from government investment in science and engineering.

[4]Peter D. Lax, ' 'Report of the Panel on Large-Scale Computing in Science and Engineering" (Washington, DC: National Science Foundation, 1982), p, 10. EDUCOM, Networking and Telecommunications Task Force, *The National Research and Education Network: A Policy Paper* (Washington, DC: EDUCOM, 1990) p. 3.

"The goal of the National Research and Education Network is to enhance national competitiveness and productivity through a high speed . network infrastructure which supports a broad set of applications and network services for the research and instructional community.

[5]Executive Office of the President, Office of Science and Technology Policy, *The Federal High Performance Computing Program* (Washington, DC: September 1989), p. 1.

"[A goal of the High Performance Computing Program is to] maintain and extend U.S. leadership in high performance computing, and encourage U S. sources of production.''

[6]*Congressional Record* comments on introduction of bill. For example, Senator Gore stated the following, when introducing his bill, S1067:

' 'The nation which most completely assimilates high performance computing into its economy will very likely emerge as the dominant intellectual, economic, and technological force in the next century.

U.S. industry must produce advanced yet economical systems which will meet the needs of users found in each of the major sectors. . . . If this is not done by U.S. Government leadership, it will be done by foreign leadership to the detriment of U.S. national intercs~s. '

Some disagree with an initiative that resembles ''industrial policy' '—i.e., policy aimed at supporting specific private sector enterprises. They argue that the government should not intervene to support either the supercomputer or the telecommunications industry. Proponents of government intervention argue that the dominant position of the U.S. supercomputer industry has historically resulted from heavy Federal investments in computing for research and that the future health of the industry will require continued Federal attention.

Some ask why science should get early preferred access to what ultimately may become a universal communication service, and suggest that selectively providing such resources to science might delay broader adoption by the public that promises even greater payoffs. They are also wary of the government providing or subsidizing telecommunication services that should, in their view, be provided by the private sector. They argue that a universal network is best achieved through the expertise and resources of the commercial communication and information industries. Proponents of Federal action maintain that the science network will be an important prototype to develop and test new standards and technologies for extremely high-speed packet-switched data communication. Furthermore, in their view, a network oriented to research and education would be a valuable testbed for developing applications and better understanding how a universal network would be used.

These debates reflect in part different philosophies, values, and expectations about future events that must be resolved in a political process. The assumptions underlying each of these three goals—advance U.S. R&D, strengthen the U.S. computer and telecommunications industry, and enhance U.S. economic and social strength-are generally soundly based because:

1. Scientific users need access to advanced computers, communication systems, databases, and software services---Scientific and engineering research in the United States cannot retain its world-class position without the best available information and communication technologies. These include advanced computer systems, very large databases, and high-speed data communications, local workstations, electronic mail service, and bulletin boards. Such technology does not simply enhance or marginally improve the productivity of the research process; it enables research that could not be performed otherwise. Simulating the complex behavior of the Earth's climate, analyzing streams of data from an Earth satellite or visualizing the interactions of complex organic molecules are impossible without these new technologies. Furthermore, many more important applications await the as-yet-unrealized capabilities of future generations of information technology.
2. Major Federal research applications have stimulated the computer industry and will likely continue to do so---Scientific and engineering applications have stretched the capacities of information technologies and tested them in ways that other applications cannot. Eventually, the techniques and capabilities developed to serve these demands make their way into the broader community of computer users. Although the computer industry structure and markets are changing, this form of technology transfer will likely continue.
3. U.S. economic growth and societal strength can be assisted by the development of a national information infrastructure that couples a universal high-speed data communication network with a wide range of powerful computational and information resources--In a recent report, *Critical Connections: Communication for the Future*, OTA stated:

> Given the increased dependence of American businesses on information and its exchange, the competitive status among businesses and in the global economy will increasingly depend on the technical capabilities, quality, and cost of [their] communication facilities . . . Failure to exploit these opportunities is almost certain to leave many businesses and nations behind.[7]

In that report, OTA listed "modernization and technological development of the communication infrastructure' as one of the five key areas of future policy concern. The high-performance computing

7u,s. Congress, Office of Technology Assessment, *Critical Connections: Communication for the Future*, OTA-CIT-407 (Washington, DC: us. Government Printing Office, January 1990), p. 6.

and networking initiative reflects a mixture of three basic goals: 1) enhancing R&D, 2) accelerating innovation in U.S. information technology, and 3) stimulating the development of a universal broadband digital network in the United States. Achieving the last goal will ultimately bring information and educational opportunities to the doorstep of most American homes.

An Infrastructure for Science

Science and Information Technology Are Closely Linked

Whether computers are made of silicon chips, optical glass fibers, gallium arsenide compounds, or superconducting ceramics, and regardless of the architecture, the basic elements of computers are the same-data, logic, and language.

- *Data is* the substance that is processed or manipulated by the technology; often, but not always, numerical.
- *Logic* is the nature of the process, from basic arithmetic to extremely complex reasoning and analysis.
- *Language is* the means of communicating from the user to the machine what is to be done, and from the machine to the user the result of that action.

These three elements are also basic to science. They characterize the nature of research and the work of scientists.

- Researchers collect data from measurements of natural phenomena, experiments, pure mathematics, and, increasingly, from computer calculations and simulations. Data can take many forms, e.g., numbers, symbols, images, sounds, and words.
- Researchers build logical structures-theories, mathematical and computer models, and so on—to describe and understand the phenomena they are studying.
- Researchers communicate their work among themselves in common scientific languages. This communication is a continuing process—both formal and informal-that lies at the heart of the scientific method, It is based on exposing ideas to the critical review of peers, allowing the reproduction of experiments and analyses, and encouraging the evolution of understanding based on prior knowledge.

Scientists invented the computer to serve research needs during World War II. In the late 1960s, research needs led to the development of ARPANET —the first nationwide communication system designed specifically to carry data between computers. NSF operated a Computer Facilities Program in the late 1960s and early 1970s that assisted universities in upgrading their scientific computing capabilities for research and education. Today, computers are used throughout society, but researchers, joined by industry, are still driving the evolution of information technology and finding new applications for the most powerful computer and communication technologies.

The invention of the printing press in the 15th century created the conditions for the development and flourishing of modern science and scholarship. Not only did the press allow authors to communicate their ideas accurately, but the qualities of the medium stimulated entirely new methods and institutions of learning. Similarly, electronic information technology is again changing the nature of basic research. The character of the research-the way data are collected, analyzed, studied, and communicated—has changed because of technology.

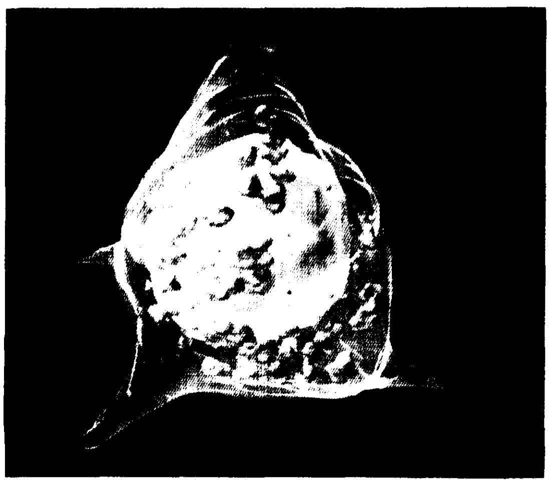
Computational research has joined experimentation and theory as a major mode of investigation. Scientists now use computer models to analyze very complex processes such as the flow of gases around a black hole or the wind patterns around the eye of a hurricane or typhoon (see boxes A and B). These and other areas of research, such as global climate change, can be accomplished only with high-performance computing. They cannot use conventional mathematical and experimental approaches because of the complexity of the phenomena.

Research is generating data at unprecedented rates. The human genome database is projected to eventually contain over 3 billion units of information. Earth observation experiments in space will collect and send to Earth trillions of units of data daily. A single image of the United States, with resolution to a square yard, contains nearly a trillion data points. Current data storage technologies are unable to store, organize, and transmit the amounts of data that will be generated from these projects. Long-term storage capabilities must be researched and developed. ‘‘Big science’ projects, such as those mentioned above, should devote a portion of their budgets for R&D in high-capacity data storage

Box A—Black Holes: The Mysteries of the Universe

A black hole is an object in space, whose mass is so dense that nothing is able to escape its gravitational pull, not even light. Astronomers think the universe is populated with black holes that are the remains of collapsed stars. Much of the research conducted on the universe has implications for other areas of study, such as physics. Below is a visualization of the three dimensional flow of gases past a black hole. The computer codes used to create this image can also be used to determine the accuracy of computer generated models of three dimension fluid flows.

Collaborative efforts between two NSF-funded supercomputer centers, the National Center for Supercomputing Application (NCSA) in Illinois and The Cornell Theory Center, resulted in a video of the phenomenon pictured, The computer code used to derive the data was written by researchers at Cornell and was run on Cornell's IBM 3090 computer. NCSA remotely accessed the data via NSFNET from Cornell. At the Illinois center researchers worked with a scientific animator who, using Waverfront Technologies Graphic Software tools and a graphics packaged designed at NCSA, processed the data on a Alliant computer. The research team created a rough contoured image of the cell with the Alliant. The researchers returned to Cornell and modified the contouring graphics programs, creating a videotape on Silicon Graphics workstation at Cornell. The project was the first joint effort between NSF supercomputer centers. Utilizing the expertise of two centers was instrumental in graphically depicting three-dimensional fluid flow.



systems if such large projects are to be successful. New forms of institutions and procedures to manage massive data banks are also needed.

Journal articles have been the major form of communication among scientists. Publishers are beginning to develop electronic journals, accessed directly over communication networks or distributed in computer readable form. The different nature of electronic storage means that these new 'publications" will likely look, behave, and be used differently than printed journals. They may contain information in a variety of forms: high-definition video, moving images, sound, large experimental data sets, or software. Using so-called "hypermedia' and multimedia techniques, these electronic journals can be linked to other related articles, films, and so on. They can evolve and change over time, containing later annotations by the original author or others, or references to later articles that advanced or stemmed from the original work.

Scientists communicate with each other continually by letter, telephone, conferences, and seminars, and by meeting personally around the departmental coffee pot. These modes of communication are often as important to research as formal publications. All of these modes will likely continue in some form, but digital communication systems provide many new powerful ways to communicate-electronic mail, computer conferences, and bulletin boards. Proximity, time, and travel are less important in using these new communication paths. With bulletin boards and electronic mail, information can be exchanged much faster than by mail and with accuracy and detail that is impossible to achieve with telephone. Since the participants need not travel, computer conferences can accommodate large numbers of participants, Sessions can take place over weeks or months, and people can participate in the electronic meeting wherever they are at whatever time is convenient.

A National Infrastructure for Research and Education

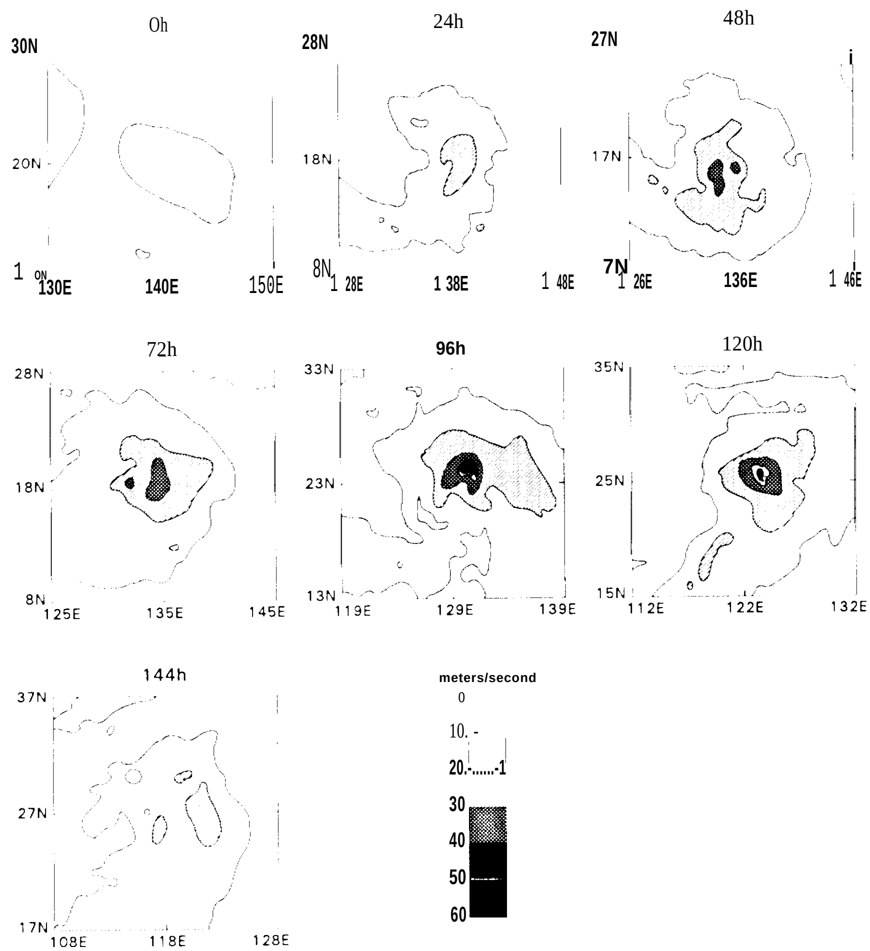
These information technologies and applications are merging into an interconnected network of resources referred to in telecommunication's parlance as an "Information Infrastructure. " This infrastructure is a conceptual collection of linked resources made up of:

Box B—Supertyphoon Hope

A typhoon is a tropical storm confined to the western Pacific Ocean, referred to as hurricanes in the Western Hemisphere. The cyclonic storms are usually accompanied by extremely low atmospheric pressure, high winds of over 100 knots, and vast amounts of rainfall. These storms can wreak havoc when they reach land, at which time they dissipate. A series of computer-generated images that trace the 6-day evolution of supertyphoon Hope are shown here. The supertyphoon's course was simulated using a computer model. Researchers were able to measure the precision of their model by comparing its results to the data gathered during the storm in 1979. Developing accurate weather models continues to be difficult despite advances in technology. Weather models must incorporate many variables such a winds, temperatures, the oceans, and atmospheric pressures.

Researchers processed their weather model at the National Center for Atmospheric Research (NCAR) remotely from their home institute, Florida State University (FSU). Using a NASA computer network, the group accessed an IBM 4381 computer, which served as the front end machine for a Cray X-MP. After the data were processed at NCAR, it was transferred to magnetic tapes and mailed to FSU for further analysis. (An increase in bandwidth now allows the researchers to separate large data sets into sections and send them over high bandwidth computer networks.) At FSU the data were translated into images using a Silicon Graphics Workstation. Data collected from the storm in 1979 are stored on computers at NCAR. The data were used to measure the accuracy of FSU'S weather model. The computer-generated storm was accurate within hours of the actual events of Supertyphoon Hope.

The facilities at NCAR were especially suited for the needs of the FSU researchers since both the staff and resources were geared towards atmospheric and ocean modeling. Researchers frequently visit NCAR for user conferences and have become familiar with many of the technical support staff.



A Nationwide High-Speed Broadband Advanced Information Communication Network

This computer-to-computer network is composed of many parts—local networks on campuses and in research facilities, State and regional networks, and one or more national ‘‘backbone’ networks interconnecting them all. This domestic backbone would link to networks in other countries. Some of the domestic networks will be private commercial networks; others may be operated by private non-profit organizations, and still others may be government-funded and/or managed.

Specialized and General Purpose Computers

Users will be able to access the newest, most powerful supercomputers. There will be a variety of specialized machines tailored to specific uses and applications because of the developing nature of current computer architectures. They will be used for database searches, graphical output, and artificial intelligence applications such as pattern recognition and expert systems. Researchers will have access to a heterogeneous computing environment where several specialized machines, each with their own strengths, are linked through a software network that will allow users to simultaneously exploit the power of each computer to solve portions of a single application.

Collections of Specialized Applications Programs

Some application programs are extremely large and represent years of development effort. They may be maintained and updated centrally and made available over the network. Groups can also make available libraries of commercial or public domain software that could be distributed to local computers.

Remote Access to Research Instruments

Some research facilities house one-of-a-kind instruments, unique because of their cost or their site location, such as telescopes, environmental monitoring devices, oceanographic probes, seismographs, space satellite-based instruments, and so on. Remote use and control of these instruments is possible through the network infrastructure.

Services To Support and Enhance Scientific Communication

These services include electronic mail and conferencing systems, bulletin boards, and electronic journals through which researchers can communicate with each other. They also will include scholarly bibliographic reference and abstracting services, and online card catalogs linked to key research libraries.

“Digital Libraries” and Archives

These resources would contain collections of reference materials-books, journals, sound recordings, photographs and films, software, and other types of information-archived in digital electronic form. These also include major scientific and technical databases such as the human genome database, time-series environmental data from satellites, and astronomical images. Some visionaries see the network eventually providing access to a connection with a “Global Digital Library, ’ a distributed collection of facilities that would store electronically most of the world’s most important information.

Facilities for Analyzing and Displaying the Results of Computations

Researchers who simulate large, complex systems must develop ways to interpret these simulations to replace examining enormous quantities of computer-generated numbers. These researchers need to interact with the model; directly controlling the computer is the next step. Researchers are developing new ways to ‘‘see’ their models by visualizing them directly or by use of methods such as holograms that provide three-dimensional views. Other researchers are developing tactile systems that, through special gloves and visors, allow a person to “feel’ simulated objects or act as if they were moving about in a simulated environment (’‘virtual reality’ ‘). Researchers on the network will draw on specialized centers of technology and expertise to help them develop interfaces with computations and databases.

The Current Federal Picture

Recent Studies

Over the last 8 years, the key Federal science agencies and private groups have assessed the role of information technology in science and engineering research. These studies have concluded that computers and communication technology are critically important to R&D and the competitive position of the United States in the global economy. The studies pointed out shortcomings in the current system and recommended Federal actions.

The Lax Report

In 1982, the Panel on Large Scale Computing in Science and Engineering issued a report that became known as the ''Lax Report' after its chairman, Peter Lax. It was jointly funded by the NSF and the Department of Defense. The panel noted that the U.S. research establishment seriously lacks access to high-performance computing. It found that this deficiency harms U.S. preeminence in R&D and threatens the current strong position of the U.S. computer industry. To remedy this, the panel recommended a national supercomputer program consisting of four basic components:

1. establish national supercomputing centers and develop ''a nation-wide interdisciplinary network through which users will have access to facilities';
2. support research in software and algorithms—particularly work on parallelism and other new computer architectures for high-performance computing in the future;
3. support education and training programs for new users in order to assist the research community in using supercomputer applications; and
4. support research aimed at developing new, faster, supercomputers.

Variation of these four elements are repeated in the subsequent proposals and initiatives.

The Bardon/Curtis Report

At the request of Congress, NSF undertook an internal review of the Lax Report designed to form a program plan. The 1983 report, which became known as the Bardon/Curtis Report, offered an ambitious program plan with several recommendations for NSF:

1. ' 'greatly increase its [NSF'S] support for local computing facilities, including individual workstations, systems for research groups, specialized computer facilities, and local area networks';
2. establish 10 supercomputer centers;
3. support the establishment of networks to link users with the supercomputer centers; and
4. ' 'support a program of academic research in the areas of advanced computer systems design.

National Academy of Sciences/COSEPUP

In 1989, the Committee on Science, Engineering, and Public Policy of the National Academy of Sciences (NAS) published a report, *Information Technology and the Conduct of Research,*[9] which examined the needs of science for new technological initiatives. This report, prepared by a panel chaired by Donald Langenberg, emphasized the changing form of research and its increased dependence on new information technologies. The report advised against leaving the design and operation of these programs only to the technical experts. Systems designers must learn what the users need, then design the system. The panel made two recommendations to do this:

- "The institutions that support U.S. research, including universities, industry, and Government should develop and support policies, services, and standards that help researchers use information technology more widely and productively, and
- "The institutions supporting the nation's researchers, led by the Federal government, should develop an interconnected national information technology network for use by all qualified researchers.

The panel recognized that industry, the universities, libraries, professional societies, and the Federal Government share responsibilities for this. The Federal role, according to the committee, should include leadership and coordination in the development of technologies and services to support the research and education needs in addition to funding.

National Association of State Universities and Land Grant Colleges (NASULGC)

In 1989, NASULGC issued a report, *Supercomputing for the 1990's: A Shared Responsibility,* on the need to make high-performance computing available for academic research, that contained recommendations for the Federal Government and universities, It points out that a computing infrastructure would have to include facilities operated by a variety of institutions beyond the Federal Government. Federal policy, it suggests, should be tailored to encouraging and leveraging private, regional, and local efforts. Recommendations for Federal action include:

'NAS repro

- supporting the national supercomputer centers and maintaining them at the technological leading edge;
- "fostering and encouraging university, state, and regional supercomputing facilities' (of which the report identified 27);
- supporting the development of a national network; and
- assuring the ''constancy' of support.

These recommendations reflect concern that the support of NSF national centers would draw funds away from computing at non-NSF centers. The report notes that the non-NSF centers will also play an important role in the future of scientific computing. NASULGC further observes that changing policies and unpredictable funding disrupts operations and discourages the development of facilities.

EDUCOM

EDUCOM is an association of higher education institutions. It functions as a clearing house for information and expertise about computers and communication technologies. EDUCOM'S university consortium created and manages BITNET, a private, shared network that serves the networking needs of academics at a low cost. Its Networking and Telecommunications Task Force examined the Federal networking and computing initiatives and has produced several policy statements and reports.

A statement, released in March 1990, focuses on the network. It makes a series of specific recommendations on implementing the NREN, but its statement of the basic goal for a network is broad:

> [NREN]. . . is to enhance national competitiveness and productivity through a high-speed, high-quality network infrastructure which supports a broad set of applications and network services for the research and education community.[9]

Common Themes

The series of reports strikes some common themes. Three points, in particular, are important to the current policy debate.

First, the network has become the key element. Seen frost as simply a means to access expensive or highly specialized computing resources (similar to the initial intentions for ARPANET), the network has become the basic foundation for the information infrastructure, connecting researchers and students not only to computers, but providing access to a wide range of services.

The network is actually an internet, a family of networks (networks within networks), the design and operation of which needs coordination and leadership from the Federal science agencies. OTA's forthcoming report on the NREN will explore these issues in depth.

Second, educational needs are now part of the NREN plan although it is undecided how wide the range of users and institutions will be. In any event, this will affect network architecture and operating policies. Once referred to as a National Research Network (NRN), it is now known as a National Research and Education Network (NREN). This evolution was natural. It is impossible to separate education from research at the graduate level. There are also strong arguments for including undergraduates, secondary schools, and even primary schools in the system. To better coordinate the educational community's views on how the NREN may assist education, the Department of Education and State and local educators must be actively involved in the policy process.

The question of scope of the network extends to research in non-science scholarly disciplines, some of which are not well-funded by Federal programs. The wide range of services offered, including access to libraries, bibliographic services, electronic mail, bulletin boards, computer conferencing, and so on, extends the network's potential scholarly beneficiaries beyond just scientists and engineers.

A third commonly raised issue in the reports is the need to look beyond mere hardware. Computers need software that make them accessible and usable by researchers. A network needs software tools and data-bases that allow scientists to communicate effectively with one another. Databases need inquiry systems (search engines), indices, tables of contents, directories, hypertext, and other tools to enable users to search, identify, and retrieve the information they need for their work. To properly develop an infrastructure that is useful to all science, attention must be paid to software as well as hardware.

Other groups also rely on access to scientific information. Public interest groups with concerns of

[9] EDUCOM, The National Research and Education Network, op. cit., footnote 4, p, 3.

public safety and health, the environment, energy, defense policy, and so on, rely on access to scientific publications and databases and attendance at conferences and seminars. The press, particularly the specialized scientific and technical press, must access conferences, journals, and other forms of electronic communication.

In most cases, these applications, which enable effective access and use of information, are neither simple nor obvious. Developing them will require significant research and software development as well as better understanding of how information technology can best assist scholars in their work. The answers to these questions will also depend on the nature and breadth of the constituency for the network. Different users will have different skills, analytical strategies, and research styles depending, in part, on the traditions of their particular disciplines and their level of training.

The Government's Role Is Changing

The Federal Government has major responsibilities for the health of basic and academic research in the United States, both as a user of the products and from its role in supporting science and engineering to advance the economy and improve the quality of life. The government is already participating heavily in the development and management of the existing R&D information infrastructure-in using it, funding it, and in setting policies. The government must deal with additional responsibilities resulting from the new infrastructure. The challenge will be to organize and assemble a government entity to: 1) identify and determine promising technological directions for the high-performance computing infrastructure; 2) evaluate progress over the course of the High-Performance Computing (HPC) initiative; and 3) make course corrections at the appropriate times.

First, facilities need to be highly interconnected. They must connect physically and logically with the network. Digital signals must conform to standards and protocols in the same way that electrical appliances must plug into standard 110 volt, 60 Hz AC power outlets. Users must be able to transmit and receive communications, programs, and data seamlessly and transparently to and from each nook and cranny in the system. Government policies and programs must be coordinated if interconnectability is to be achieved.

Second, many of the shared resources cut across agencies, institutions, disciplines, and programs. This feature of sharing is most obvious in the physical network; but it is also true for many of the computing facilities, data archives, and network services such as directories, bulletin boards, and electronic mail. Thus, many policy decisions regarding the use and access to these resources and services must be made at an interagency level. Furthermore, new private networks and service corporations now provide networking services to public and private customers. The policies of these private entities will become more important as the network moves towards full commercialization.

Third, the facilities will be expensive and require large capital investments. The Federal Government will be asked to share these costs with States, local governments, other countries, research and educational institutions, industrial users, service providers, and individual users. Private entities may be expected to contribute substantially as well. But while technological risks may be acceptable to private companies. The commercial risks may be unacceptable without government support.

Fourth, many of the resources on the network will be unique and of great national-even international— importance, e.g., a supercomputer dedicated to global climate modeling or the human genome database. Access to these scarce resources must reflect a cooperative set of goals, to determine access, decide who can use it, and to set national priorities. Federal policies are needed to balance and ensure equitable access and security. The Executive Office of the President (OSTP) and congressional committees may be called on to referee conflicts among competing interests from time to time. A well-organized Federal management system responsible for policy oversight and operation of the HPC and network infrastructure can anticipate or avoid many problems, and thus reduce the need for political resolutions.

Finally, the government must assist in advancing the state of computer and communication technologies to hasten the development of more powerful high-performance computers, faster data communications, and more effective software. Several studies by NAS, OSTP, and others have identified "Grand Challenges in Research" of critical national importance, but which are currently unachievable because

33-982 - 91 - 2

of inadequate computing power.[10] What is needed are computers that are hundreds-even thousands—of times faster than the best now available. Similarly, the “data explosion” demands better and faster storage technology to archive large data sets. The rapidly growing communication needs of science require switched wide-area digital communication networks capable of moving billions of units of information per second to and from researchers.

The Structure of Federal Policy

Researchers foresee computers that will soon perform a trillion arithmetic steps (’ ‘teraflops’ per second, data communication systems that can transmit billions of units of data (’ ‘gigabits per second, and electronic storage systems that can store correspondingly large amounts of information and absorb and disgorge it at rates matched to the speed of the computers and communication lines. New hardware will require new streamlined software to operate the high-speed computers and communications networks efficiently,

Developing,]Managing, and Funding Major Resources

A striking trend in information technology is the development of inexpensive input and output devices-computers, telephones, facsimile machines, and so on—that are affordable and easy to use. But the opposite is happening in the development of advanced computers and communications technology. High-performance computers, data networks, and archiving facilities are extremely expensive to build and operate. They are complicated technologies that require experts to develop and operate. Such expensive and complicated systems must be located in a few central facilities and made available to network users through such installations as the National Supercomputer Centers, NSFNET, the planned Human Genome Data Archive, and so on.

Allocating Resources and Assuring Equitable Access

The network raises a number of allocation and access issues that must be resolved. The number of high-performance computers and elaborate research instruments, such as telescopes or particle accelerators, are limited because of their high capital and operating costs. Universal access is not feasible, yet these facilities are critical to certain types of research. An equitable, fair process for allocating time on these facilities is crucial.

Network utility features, such as electronic mail, bulletin boards, journals, and so on, are basic to research in any field. Without them, one is locked out of the profession of science. Every researcher must have access to these services.

Updating Information Policies

Policies that currently govern the existing networks were developed to resolve conflicts over access and control of the information, e.g., protecting the privacy and confidentiality of communications and data on the network, or enforcing intellectual property rights. There are many more information policy questions concerning the rights and responsibilities to various electronic forms of communication that must still be addressed. Should ‘‘electronic mail’ be protected like first class mail? Should bulletin board operators be legally responsible for messages placed on their boards? Does the First Amendment of the U.S. Constitution protect the sender? Should intellectual property protections be granted to electronic databases? If so, what form? Answers to these and other information policy questions will determine how the network is used and what services will be offered.

Adapting Science Policy

Just as an information infrastructure may change the way science is *done,* it may also lead to the need to change Federal science policy to accommodate these changes. High-performance computers may

[10] OSTP defines Grand Challenges as ‘‘. . .a fundamental problem in science or engineering, with potentially broad economic, political, and/or scientific impact, that could be advanced by applying high performance computing resources. ’ Examples include: 1) Computational fluid dynamics for the design of hypersonic aircraft or efficient automobile bodies and recovery of oil; 2) Computer based weather and climate forecasts, and understanding of global environmental changes; 3) Electronic structure calculations for the design of new materials such as chemical catalysts, immunological agents and superconductors; 4) Plasma dynamics for fusion energy technology and for safe and efficient military technology; 5) Calculations to improve the understanding of the fundamental nature of matter, including quantum chromodynamics and condensed matter theory; and 6) Machine vision to enable real-time analysis of complex images for control of mechanical systems. See Office of Science and Technology Policy, *The Federal High Performance Computing Program* (Washington, DC: 1989), p. 8.

change research priorities and create new ways for research groups to organize and work together.

Determining the Type of Technology

The purpose of the technology will drive future policies. Management questions arising from the design and operation of a nationwide. ultra-high speed communication network may differ in nature from the problems of supporting and operating National Supercomputer Centers. But decisions related to both will collectively determine how effectively a national in format ion infrastructure will be used in research and education.

On the other hand, information policy issues—that relate to the information that flows through the network - connected faci1ities—are seamless1y linked. Although the technical means for protecting and controlling in formation moving over the system may differ from computer to computer, Or application to application, in formation policies are less dependent on the nature of the technology than on the generic issues. Privacy protection, access control. data security, and intellectual property protection are problems that need to be addressed across the board. Similarly, changes in the framework of Federal Government support and oversight of science policy wi11 affect all technologies, disciplines, and agencics.

Determining Access

Depending on how one views the NREN, it is seen serving widely different user groups, ranging from a few federally funded high-end researchers engaged in ' 'Big Science, to the scholarly community, to education from kindergarten through secondary schools. Both technical design decisions and policy will affect these various users in different ways. Who the intended user will be is a criticaly important consideration in making NREN policy. It is a subject dealt with in detail in a forthcoming OTA Report that focuses on the network as a broadband advanced communication infrastructure to simultaneously deliver data, video, and voice service.

Major Strategic Concerns

The mutual dependence and interconnectedness of a national information infrastructure will force the Federal Government to develop long-term strategies to guide the overall development of the NREN: this must be done in concert with a coordinated program to provide high-performance computer-based tools for science, research and education.

Breadth of Scope

Long-Range Planning Needs

Creating the infrastructure, the network, and its related resources, is not a one-time job. There are misconceptions that information infrastructure is a static concept only needing to be plugged in. This is not so: new applications will appear and the capabilities of technology will continue to grow and change. The system will, therefore, be a continually evolving assembly of technologies and services. Therefore planning and operating the NREN must be considered a dynamic process. An institutional framework must be developed to ensure its success.

Studies on information technology and science (including OTA'S) rely on anecdotal examples, ' 'gee-whiz' speculation about future applications, and the subjective views of the research and education community. These arguments are persuasive and sufficient to justify the support for the NREN, but they do not contribute sufficiently to long-term management and planning for the *operation* of the infrastructure. The Federal investment in computer, communication, and data resources for science and engineering should be based on a periodic assessment of needs and changing technologies. This assessment should include:

- surveys of existing resources-public, private nonprofit, and commercial, such as:
  - —specialized and general purpose high-performance computing facilities;
  - —Federal, State, and local data communication networks;
  - —scientific and technical databases; and
  - —software packages for research uses.
- utilization levels of existing facilities by categories such as:
  - —field of research;
  - —government, academic, or industrial use; and
  - —research, graduate education, undergraduate education, or pre-college education.
- Barriers to efficient use of facilities, such as:
  - —policy or legal barriers;
  - —lack of standards for interconnection of systems; and
  - —user difficulties such as lack of training, inadequate user interfaces, or lack of software and services.

- Projections of future computing needs (particularly, assessments of the need for new, large-scale research initiatives.)

Although NSF attempts to keep tabs on the computational and information needs of the science community, the pace of technological development and massive new science projects make such information more important during periods of tight budgets. These data are difficult to compile, and special efforts are needed to provide such planning data to those decisionmakers responsible for anticipating future national computing needs.