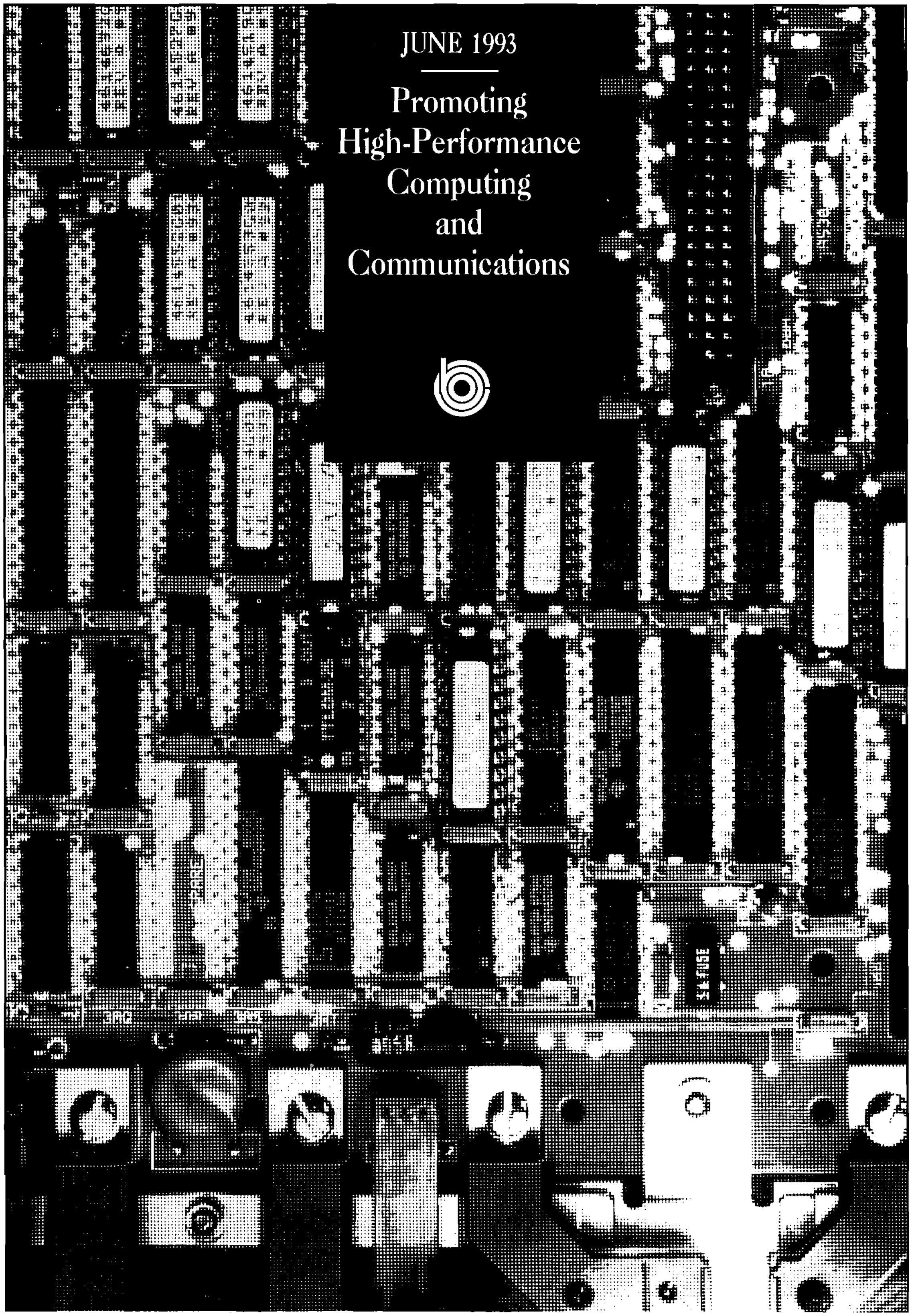
CONGRESS OF THE UNITED STATES
CONGRESSIONAL BUDGET OFFICE

A CBO STUDY

JUNE 1993

Promoting High-Performance Computing and Communications



PROMOTING HIGH-PERFORMANCE COMPUTING AND COMMUNICATIONS

The Congress of the United States
Congressional Budget Office

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328
ISBN 0-16-041828-3

NOTE

Cover photo shows a detail from a processing node of a parallel supercomputer. (Photo courtesy of Bolt Beranek and Newman, Inc., Cambridge, Massachusetts.)

Preface

The High Performance Computing and Communications (HPCC) program embodies many concerns regarding federal technology policy: the role of the federal government in promoting new industries, building infrastructure, supporting joint federal-private efforts, and encouraging the transfer of technology. This report focuses on the first of these concerns--the federal role in promoting new industries. As requested by the Senate Committee on Commerce, Science, and Transportation, the report concentrates on the obstacles--mainly on the demand side--that might prevent the high-performance computing and data communications industries from growing and using the technology being developed under the HPCC program. In keeping with the Congressional Budget Office's (CBO's) mandate to provide nonpartisan analysis, no recommendations are made.

Philip Webre of CBO's Natural Resources and Commerce Division wrote this report under the supervision of Jan Paul Acton and Elliot Schwartz. Serdar Dinc, then with CBO, provided research assistance in early stages of the project. The National Coordinating Office for High Performance Computing and Communications coordinated Executive Branch agency comments on earlier drafts. The author wishes to thank Fiona Branton, Alan Buzacott, Deborah Clay-Mendez, Kenneth Flamm, Stephen Gould, Mark Grabowitz, Peter Gregory, Anthony Hearn, Robert Kelly, Paul Muzio, W. Edward Steinmueller, Richard Van Atta, and Joan Winston for their helpful comments on earlier drafts.

Francis Pierce and Sherry Snyder edited the manuscript, with editorial assistance from Christian Spoor, and Donna Wood typed the many drafts. Martina Wojak-Piotrow prepared the report for publication.

Robert D. Reischauer
Director

June 1993

Contents

TABLES

FIGURES

Summary

In 1991, the Administration and the Congress initiated the multiagency High Performance Computing and Communications (HPCC) program to further the development of U.S. supercomputer technology and high-speed computer network technology. Although one reason for this legislation was to improve the performance of federal missions, an important consideration was the benefits such technology might yield for private industry. Such benefits can only be realized if the technology becomes widely used in industry--in short, if it is commercialized.

Commercial success, however, is only one of several goals for HPCC. A full analysis of the program, which is not attempted in this study, would take into account how HPCC technology will affect federal missions. Even if it has only a small impact on commercial technologies, the expenditures may be justified by its benefits for federal agencies.

The High Performance Computing and Communications Program

The technical goals of the High Performance Computing and Communications program are to create supercomputers capable of being scaled to a trillion mathematical operations per second and computer networks capable of transmitting a billion bits (a gigabit) of data per second. In each case, the goal is to achieve sustained performance of between one and three orders of magnitude better than that of current technology. With this additional power, federal agencies hope to be able to solve computational problems that have so far eluded solution in areas of federal interest such as changes in global climate, high-energy physics, and high-performance aircraft. These types of problems are referred to as "the grand challenges."

The program is divided into four areas:

- o Supercomputer hardware, which focuses primarily on so-called scalable massively parallel supercomputers that are often composed of hundreds and thousands of individual microprocessors;
- o Supercomputer software, which centers primarily on creating programming models, tools, and applications for massively parallel supercomputers, which cannot use conventional supercomputer software;
- o Networking, which focuses on upgrading the interagency National Research and Educational Network and other agency networks, as well as on advanced network research and development (R&D); and
- o Basic research and advanced education for programmers.

In many ways the HPCC program continues ongoing federal work in computer R&D, but with added emphasis. Thus, its budget has two components: existing funds and additional funds. The program is intended to double the annual federal resources being devoted to the development of high-performance computers and computer networks over a five-year period, going from $600 million in fiscal year 1992 to $1.2 billion in 1996. Funding over the five-year period is forecast to total $4.7 billion, of which $1.9 billion will be incremental to the previous levels of resources.

Although HPCC is often described as a computer network R&D program, 80 percent of the new funds are for R&D involving areas other than networks. Supercomputer hardware and software are to receive over two-thirds of the incremental funds, while network R&D will receive one-fifth of the new monies; the remainder goes to basic research and education. Summary Table 1 presents the division of additional spending by program area.

Summary Table 1.
Incremental Funding for the HPCC Program, Fiscal Years 1992-1996, by Program Component

	Funding	
Component	In Millions of Dollars	As a Percentage of Total
High-Performance Computing Systems	682	36
Advanced Software Technology and Algorithms	662	35
National Research and Education Network	390	20
Basic Research and Human Resources[a]	183	10
Total	1,917	100

SOURCE: Congressional Budget Office using data from Office of Science and Technology Policy, "The Federal High Performance Computing Program" (September 8, 1989), p. 46.

NOTES: HPCC = High Performance Computing and Communications.

Incremental funding is the additional money appropriated to the HPCC agencies above the level the agencies were receiving at the start of the HPCC program for their activities in this area. Numbers may not add to totals because of rounding.

a. In addition, 15 percent of each of the other three program components are to be devoted to this area.

Supercomputer Technology and Markets

The systems that now seem likely to be the first to develop the speeds sought by HPCC are massively parallel supercomputers. Conventional supercomputers are composed of a small number--usually less than 16--of very fast processing units. The massively parallel systems use a large number--as many as several hundred or thousand--of relatively slow processing units, all working in parallel on the same problem.

HPCC is supporting research into a type of such supercomputers called scalable, which means that the computers can be made more powerful by simply adding more computing nodes, unlike conventional supercomputers. The theoretical advantage of scalable architectures is that software applications can be developed on smaller computers and used to run smaller data sets on smaller computers. However, if the models or data become too massive, the problem can be shifted to larger computers without having to rewrite the software.

Most modern high-end supercomputers, including conventional supercomputers, are parallel in the sense of having more than one processor, and the trend clearly is to increase the number. Where the technology strategies differ between conventional and massively parallel machines is in how fast to increase the number of processors over time. In order to achieve its speed goals, HPCC has gone with computer designers who would increase the number rapidly.

Supercomputer Markets

Because most supercomputers are used for research purposes, the market for them is small relative to other parts of the computer market. Only a couple of hundred systems are sold each year. By comparison, mainframe computer systems--the most similar type of system physically, organizationally, and in cost--sell between 15,000 and 20,000 units a year. Overall, only 600 supercomputer systems are installed worldwide. According to the Department of Commerce, supercomputer sales totaled $1.8 billion worldwide in 1992.

Growth in the demand for supercomputers has slowed substantially in recent years. Although the market for them grew rapidly during most of the 1980s, the recent annual rate of growth has been only 6 percent. Since many supercomputer uses are military applications, overall demand may be depressed for a number of years.

Supercomputer Technology Spinoffs

Part of the reason for the focus on supercomputer technology is the commonly held belief that supercomputer technology leads all computer technology. New technologies, it is often argued, will appear first in supercomputers, where performance considerations outweigh cost and other considerations. After an innovation has proved its worth, conventional computers can adapt it to their use. The policy corollary to this belief is that small investments in supercomputer technology will produce large benefits in all areas of the computer market.

Although supercomputers and their high-performance predecessors originally led overall computer technical development, this role has been reduced in the last 10 years. In fact, supercomputers played a minor role in many of the most important developments in computer technology in the 1980s. Most significantly, in the two central areas of computer technology development during the 1980s--computer architecture and integrated circuit technology--conventional supercomputer technology was removed from the areas of rapid advance. The needs of personal computers and their high-powered cousins, the engineering workstations, drove most of the technology developments in these areas. Although some technology has come from supercomputers, the major thrusts of computer technology development have recently been elsewhere.

Nevertheless, in certain areas supercomputers have been at the forefront of computer technology, most notably in software. As engineering workstations--high-performance versions of the personal computer--have become rapid enough to provide timely mathematical solutions to engineering and scientific problems, software originally written for supercomputers has been translated for workstations.

Future Trends in Technology Spinoff

Unlike previous generations of supercomputer technology, the massively parallel supercomputers, such as those being developed in HPCC, use the same components as the engineering workstations. In this circumstance, the likelihood of innovation in workstation technology arising first in the supercomputer market may increase, although at present the technology is flowing the other way. Indeed, the very existence of massively parallel supercomputers is predicated upon the availability of high-performance technology from commercial workstations. At the computer component level, the technology flow is expected to continue from workstations to parallel supercomputers.

Two of the likeliest major spinoffs from the massively parallel supercomputer to other types of computers do not involve components. Rather, the novel challenges HPCC is confronting in advance of the rest of the industry involve creating parallel paths in solving software problems and facilitating communication between processors. When researchers have rewritten software applications to make them parallel for the massively parallel supercomputer, they have discovered that the applications also run better on conventional supercomputers.

This effect--making software run more rapidly by making the parallels implicit in most problems more explicit--may apply more generally. Similarly, discovering ways to enhance communications among the supercomputer's massive number of processors may be analogous to improving communications among conventional computers in computer networks. Transfer of these technical insights from supercomputer technology to conventional computer technology, however, may be limited by the differences in cost and performance objectives in the different markets.

Obstacles to the Development of Large Commercial Markets for HPCC Supercomputer Technology

The main obstacle to the rise of large commercial markets for HPCC-developed supercomputer technology is that cheaper workstations may preempt further substantial growth of the supercomputer market as a whole. At present, an increasing share of commercial design problems may be solved on workstations rather than on supercomputers. If the growth in the workstation market preempts further substantial growth in the supercomputer market, economic benefit from HPCC technology may not be sufficiently widespread to justify the large funds being devoted to it on purely commercial grounds. This situation may be analogous to the way the personal computer replaced the mainframe or minicomputer for many uses in the business world and reduced demand for larger business computers.

The other obstacle to the widespread use of HPCC technology in the near term can best be described as economic inertia: conventional supercomputer hardware is good and continues to improve, making conventional supercomputers hard to displace in the market. Furthermore, close to 20 years' worth of software has been refined to run on that hardware. The difficulty of writing software for massively parallel supercomputers will make them slow to supplant conventional supercomputers. Tests by the Department of Energy and the National Aeronautics and Space Administration suggest that currently available massively parallel supercomputers do not seem to be able to run most existing software as rapidly as the fastest conventional supercomputers, despite their higher theoretical peak speeds. Nevertheless, measures of cost per unit of performance are favorable to massively parallel supercomputers.

Massively parallel computers currently occupy a few different small markets, some with potential for significant growth. The market with the most potential involves data-base management. Roughly half of corporate mainframe use involves data-base manipulation, such as spotting sales trends. Such problems are also well suited to parallel architectures. One major producer of massively parallel computers is already selling to this specialized market. (Although its computer is parallel, this is not, properly speaking, a supercomputer as it lacks substantial number-crunching abilities.)

Some massively parallel supercomputer manufacturers are interested in entering the data-base market. Data-base manipulation, however, generally does not require the number crunching that is the strength of supercomputers. Whether supercomputers can reposition themselves into this market is not clear. Massively parallel supercomputers also occupy a significant fraction of the supercomputer market, currently supplementing conventional supercomputers in supercomputer centers.

Computer Network Technology and Markets

The National Research and Educational Network (NREN) is central to the HPCC program, both as part of the goal and as an enabling technology for other components of the program. By allowing supercomputers and other research assets of the federal government to be linked over a very rapid computer network, the program planners hope that the different agencies will be able to bring their best tools to bear on solving the grand challenges. In addition, NREN could demonstrate how other segments of society could use such power.

The NREN component of HPCC has two major subcomponents: upgrading the interagency interim NREN, and research and development focused on computer networks of 1 billion bits per second. The interagency interim NREN has several parts. The most commonly discussed is the effort to bring the National Science Foundation's NSFNET up to 45 million bits per second, which is largely completed. NSF has begun the bidding process for the next upgrade. Less well known efforts involve bringing specialized data bases developed by federal agencies on line to make them accessible from the NSFNET, and improving other federal agency networks.

The research and development component of NREN largely focuses on so-called "gigabit test beds." In these test beds, industry teams and government researchers unite to develop and test different components, protocols, and other aspects of networking technology. As technology is developed it presumably will be incorporated into the construction of the interim NREN and vendor equipment.

Computer Network Markets

The United States has a very large and thriving data communications market. According to the Department of Commerce, the worldwide market for computer network hardware and software exceeded sales of $8 billion in 1991, much if not most of it provided by U.S. companies to clients in the United States.

In addition, the Department of Commerce estimates that the U.S. market for value-added network services, including credit card validation and the like--perhaps the most common single type of commercial network service--reached $6 billion in 1991. The market for another popular type of service, called outsourcing, in which one firm contracts with another to manage its computer network, is estimated to be in the $3.5 billion to $7 billion range annually. By contrast, the market in switched network services, comparable to switched voice telephone service, is still in its infancy.

The Internet and Development of the Network Services Market

The Interim NREN or Internet, which uses the existing NSFNET as its backbone, is a network of computer networks serving the educational and research communities, all of which use the same computer protocol to communicate. The very rapid growth of traffic on Internet during its five years under the current contractor indicates that it is fulfilling a role in research and education (see Summary Figure 1). But its users are only a fraction of the potential numbers. A major question in upgrading Internet is whether delivering enhanced services to its most sophisticated users is more valuable than broadening its usage by less sophisticated users.

Because networks become more valuable as more people connect to them, Internet is becoming the core of a nationwide data network for many of the key centers of science and technology in the United States. Not only universities and research institutions, but also many technology companies have Internet connections directly or indirectly though commercial service providers.

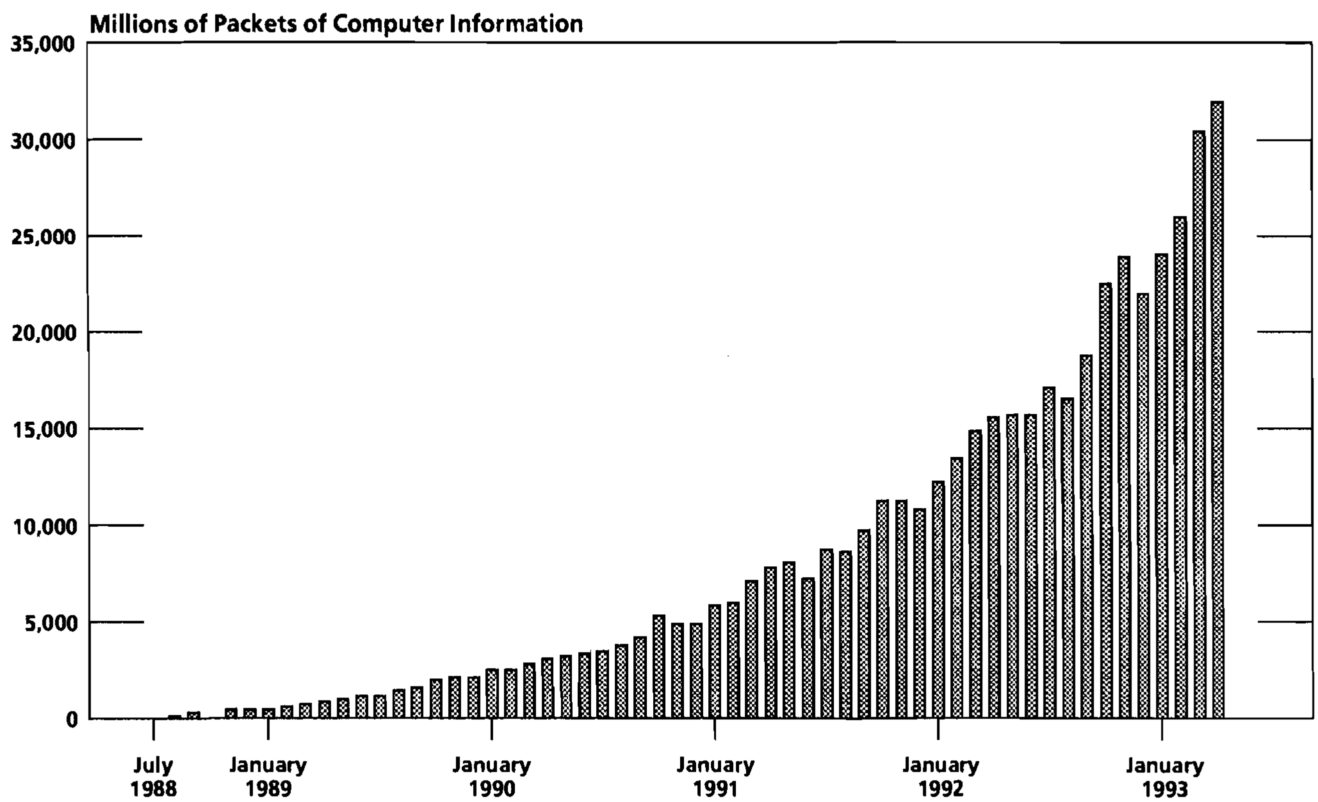
Obstacles to Developing a Large Commercial Market for HPCC Computer Network Technology

HPCC technology might be precluded from having substantial effect on the current market for data communications if demand for high-speed data communication services does not emerge as rapidly or in the directions envisaged by HPCC leaders. First, network users may find costs to be too high because the price of one of the most important network components, leased private telephone lines, is not dropping rapidly. Stagnant network costs could slow demand growth. Second, most users may find their needs satisfied or mostly satisfied by current or emerging technologies.

Costs of Private Network Lines. One cause of the proliferation of computers has been the rapid decline in computer prices. Prices of the metropolitan or regional area private telephone lines used by many computer networks are under the control of local telephone companies, which have a near-monopoly on them. The prices of these regulated lines, according to data filed before the Federal Communications Commission, are falling by only a small percentage a year, which is insignificant in comparison with the price declines in most electronics hardware. Although potential competitors have entered the market for local area lines, they represent only a fraction of total capacity. Despite commitments by local telephone companies to bring costs down by providing switched network services, their present commanding position in the leased private line business gives local telephone companies few incentives to spread switched network services aggressively, leaving most potential customers with expensive leased lines as their only option.

The long-distance carriers seem to be following the lead provided by the local telephone companies in setting short-distance private line prices. Although the prices posted with the Federal Communications Commission for private lines longer than 200 miles have been declining since 1990, the prices of leased private lines shorter than 100 miles have been rising. Thus, the average posted price of a private 1,500,000 bit-per-second leased line provided by the major long distance carriers roughly the distance between New York and Chicago (700 miles) has been falling roughly 13 percent a year since 1990,

Summary Figure 1.
History of Internet Traffic, 1988-1993



SOURCE: Congressional Budget Office from data provided by Merit/NSFNET, Ann Arbor, Michigan.

while the price of the same capacity line 75 miles long has been rising by 7 percent a year since 1990. (The long-distance carriers also provide time and quantity discounts, so no single user may face precisely these prices.) Such price increases may affect the decisions of small and medium-sized regional businesses, which are crucial in a growing market.

By contrast, prices of most other components of computer networks are falling. Consequently, overall computer network prices are still coming down. As prices of other network components fall, the prices of the leased lines will come to represent a larger fraction at the whole network, which may slow the overall decline in network costs and slow the growth in demand for network services. The rapid rise in the number of computers and local area networks for computers, however, may serve to increase the demand for network services, even if prices do not fall.

Trade-Offs Between Cost and Performance. Above a certain technological level, computer network users may find little advantage in higher performance speeds. The NSFNET switching facilities have been upgraded from 1.5 million bits per second to provide individual circuits capable of carrying 45 million bits per second. This new service will make a difference in most current applications. Above this level, however, networks reach a point of diminishing returns and expansion from 45 million bits per second to 1 billion bits per second may not create noticeable benefits for the vast majority of users who concentrate on electronic mail and simple file transfers. Without new applications to justify the cost, a mass market in such services may not develop.

Policy Directions

The Computer Systems Policy Project, a group representing U.S. computer producers, has argued in their evaluation of HPCC that the computing problems the program intends to address, the so-called grand challenges, are too distant from the problems of daily life. In their view, the program does not effectively demonstrate ways in which high-performance computer networks could actually help people.

In response to these arguments, the Congress has been considering legislation to expand the network component of HPCC to address these problems. One approach would expand HPCC to create applications in four distinct areas. To pay for these additional programs, the Congress would authorize an additional $1.0 billion to $1.6 billion over the next five years, targeted on four areas: medicine, education, manufacturing, and libraries.

- In medicine, the bills would direct the National Library of Medicine to develop network applications for communicating medical images, such as X-rays and CAT scans, by computer network; build test-bed networks to link hospitals and other medical centers for the sharing of data and records; and help provide long-distance medical care.

- In education, the National Science Foundation would be authorized to develop pilot projects to connect U.S. elementary and secondary schools to Internet, including generating software and training materials.

- In manufacturing, the National Institute of Standards and Technology would develop and transfer electronically networked manufacturing applications, including development of standards.

- Both the National Science Foundation and the National Aeronautics and Space Administration would be made responsible for developing prototype digital libraries and the associated technology, including, in the case of the National Aeronautics and Space Administration, making specialized government satellite data available over Internet.

Underlying the proposed expansion is an unresolved tension in the direction of HPCC. Is HPCC ultimately concerned with providing networks for the federal missions that need to link high-performance computers, or is its main objective to develop technology to make ubiquitous computer networks a reality? The former implies high-speed links between important national research centers, while the latter emphasizes an inexpensive and easy-to-use network available to a wider public.

Until now, the fact that parts of the HPCC budget were dispersed among the "discretionary spending" portions of several agency budgets mitigated this conflict among objectives. But given the cuts in discretionary spending mandated by the Budget Enforcement Act of 1990, policymakers may have to decide which aspects of HPCC are most critical. Since the agencies centrally involved in HPCC--the National Science Foundation and National Aeronautics and Space Administration--experienced real declines in their budget authority in 1993, they will have to make internal decisions as to which of the different components of HPCC will take precedence.

Conclusions

Most of the barriers to the commercialization of HPCC technology discussed in this report may be factors limiting the demand for these new technologies. New technologies often founder on the demand side, when the benefits they offer seem not to be worth the added expense for most potential users. Commercial demand is, of course, not the only reason for pursuing these technologies. In many cases, government agencies can use the capabilities of high-performance computing or communications--indeed, that is one of the primary goals of HPCC. In the short term, however, agency missions often conflict with commercial requirements.

The different elements of HPCC vary in their economic potential. Most obviously, the graduate students trained under the human resource programs are likely to find jobs in teaching and in industry, where they will be needed to educate the next generation of computer programmers and to write commercial computer programs. Other parts of the HPCC program may not have as smooth a road.

Chapter One

Introduction

The federal government has fostered the development of computer technology since World War II.[1] At the frontier of today's technology are high-performance computers and the communications networks that would allow more people to use them. The Administration and the Congress have taken existing programs, provided additional funds, and molded a multiagency program for research into and early development of high-performance computers, the software they use, and the networks that link them.

The High-Performance Computing Act of 1991 established the High Performance Computing and Communications (HPCC) program to promote U.S. technological leadership in high-performance computing and data communications. In programmatic terms, that goal means developing computer architectures--the design of a computer's functional operations and logic system--that can perform up to 1 trillion mathematical operations per second, and computer networks that enable users to transmit or receive up to 1 billion bits of computer data per second. Concomitant with this will be the development of software applications and the training of computer scientists and programmers who will use the new technologies.

The types of problems that require this level of computing power are referred to as the "grand challenges." These grand challenges include scientific problems such as studying ozone depletion and air pollution and modeling the ocean. But they also include potential industrial applications such as designing drugs (rather than discovering them) and modeling industrial chemical processes. Their common ground is that the solutions to these problems should be both demanding on computer resources and valuable to the nation.

The program's mission, however, is not simply to devise the fastest computers and computer networks, but also to ensure that the resulting technology is disseminated widely and applied to national needs, including those related to economic competitiveness, national defense, education, and the environment. A key goal of the program is to have the technology become an integral part of the problem-solving process in U.S. firms, thus improving U.S. productivity and competitiveness. Faster and cheaper computer power could help design teams in their search for new drugs, airplane designs, efficient and nonpolluting automobile engines, and so on. Program managers believe that the improvements derived from this new computer technology could allow U.S. firms to compete more effectively in world markets.

The HPCC program will clearly benefit the scientific and research communities. But will it have a major effect on the larger computing market beyond this initial user group? This question would not be as central if HPCC were solely a science program. The intent of the enabling legislation, however, is that HPCC will also help industry to create and sell better

1. For a history of federal government involvement in computer technology, see Kenneth Flamm, *Creating the Computer: Government, Industry and High Technology* (Washington, D.C.: Brookings Institution, 1988).

computers and data communications equipment. Meeting the first goal--building a scientific instrument to be used by federal agencies--will be relatively easier to accomplish and provides a familiar role for the government. Meeting the second goal will be more difficult; past federal efforts in this direction have had mixed success.

This report focuses on potential obstacles facing HPCC program managers in meeting the second goal, particularly on economic and institutional barriers that may hamper the creation of large commercial markets for the technology developed in the HPCC program. Most of the barriers discussed in this report relate to the demand for this new technology. It is on the demand side, in fact, that many new technologies founder: potential clients simply do not consider the advantages provided by the new product or services worth the cost and effort of switching from their current way of doing things.[2] Although there will also be challenges on the supply side, HPCC is addressing them directly through its research and development (R&D) component. This report assumes that the program can produce the desired technology. The question it seeks to answer is, What happens when the R&D is successful?

Public Purposes for Funding the HPCC Program

The federal government is funding the HPCC program for two general reasons. First, such technology may help federal agencies to accomplish some of their missions more effectively. Second, such funding is part of the overall federal role of funding R&D. These rationales, however, are shifting and are subject to some qualification.

2. Steven Schnaars, *Megamistakes: Forecasting and the Myth of Rapid Technological Change* (New York: Free Press, 1989).

The principal federal uses of supercomputers have been in the design of nuclear weapons and in intelligence. With the end of the Cold War, the need for new nuclear weapons has been substantially reduced. The intelligence function, especially the electronic surveillance that required supercomputers, has also been affected, though less drastically. In addition, supercomputers are used in energy, environmental, and space missions, but these have traditionally been secondary uses. Federal activity in the fields of national security and pure science is outside the bounds of this study, which focuses on the commercial aspects of HPCC technology.

Economists use two general arguments to justify financing R&D programs for technologies that have potential commercial uses. The first is the well-known tendency of private firms to underinvest in R&D. Economists generally argue that private firms, left to their own devices, tend to underfund R&D largely because the innovating firm is unable to capture all the benefits from its R&D investment.

The second argument invokes a traditional governmental role in the production of infrastructure. As computers become ubiquitous, their ability to communicate through networks will benefit more and more of the economy, just as the proliferation of telephones early in this century made telecommunications an essential part of the U.S. economy. The role of the federal government in helping to create that infrastructure--through R&D, demonstrations, regulation, and setting of standards--can be significant. (Questions of how the computer network infrastructure will be planned, financed, and provided are beyond the scope of this study.)

Governmental support for HPCC is also driven by the perception that the development of computer technology has been good for the country and will continue to be so. Because the federal government played a significant role in the past development of the computer industry, many people believe that it can do so in the future.

Commercialization Defined

In this study, the term "commercialization of HPCC technology" refers to the development of substantial nongovernmental markets for products incorporating technology developed under the HPCC program. A technology developed for the HPCC program would be commercialized if it were sold directly in large numbers or if it were incorporated into various products that were then sold in significant quantities.

The HPCC program leadership takes a less rigorous view of commercialization: it argues that if a contractor to an HPCC agency develops technology for the government, the technology is commercial because it is in the hands of a private party. Because most of the R&D performed on hardware for supercomputers or networks is being performed in the laboratories of private contractors, the technology is by this definition commercial, and the HPCC program has therefore succeeded in commercializing its technology.

The problem with this definition is that it ignores half of all markets--namely buyers. Numerous technologies have been sold by contractors to the government, but have had virtually no commercial buyers. The National Aeronautics and Space Administration's contractors have failed to sell a single space shuttle to nongovernmental buyers, for example. Similarly, the Department of Energy's contractors are not selling a lot of synchrotrons, although the expertise to build them is surely in private hands. Of course, weapons first developed for the Department of Defense have been sold worldwide, but usually to other governments.

To judge the effect of the HPCC program on the development of computer and data communications markets, which is the stated intent of the legislation, the appropriate analytic framework must involve private sales, not just private supplies. In the case of supercomputers and computer networks, however, this definition may be too narrow. In both these markets, federal buyers represent a large percentage of all buyers, independent of the HPCC program. For this reason, and also because many statistical sources may not break out the relevant data into governmental and nongovernmental sales, the definitions used in this study have been broadened to include all sales--governmental and commercial.

The HPCC Program as an Investment

Funding the HPCC program will be a substantial investment and should be undertaken only if the economic rewards are likely to provide an acceptable rate of return to society as a whole. Improving computer technology itself is not a sufficient rationale--computer technology is already improving rapidly. Rather, the HPCC program has to move computer technology ahead more rapidly than would otherwise occur or in directions in which it would not otherwise move.

On previous occasions, federal agencies committed themselves to speeding up electronic technology only to discover that commercial electronic technology was moving even more rapidly.[3] Improving technology is not enough; the HPCC program must have a very good chance of substantially advancing it more rapidly than would otherwise occur.

Similarly, the commercial benefits should be in proportion to the size of the investment. A federal program that spends several billion dollars to help create a market of a few hundred million dollars a year may not provide a substantial social rate of return to the taxpayers. (On an annual basis, a few hundred million dollars would normally be a reasonable return on a few billion dollars, but these few hundred million dollars of sales would require substantially more than just the federal

3. For a discussion of how civilian semiconductor technology usually outpaces that of the military despite special initiatives, see Anna Slomovic, "An Analysis of Military and Commercial Microelectronics: Has DoD's R&D Funding Had the Desired Effect?" (Ph.D. dissertation, RAND Graduate School, Santa Monica, California, 1991).

investment. They would also require additional product development by private investors, as well as investment in manufacturing the devices. The return on HPCC would be net of the return on these subsequent investments.)[4]

The HPCC and Previous Federal Computer R&D Programs

Many economic studies of federal R&D programs have found that, unlike private R&D spending, most federal R&D funding does not provide a substantial economic rate of return.[5] (Most federal R&D is mission-related and does not result in increased economic output as measured by typical indexes.) By contrast, federal efforts in computer R&D have enjoyed many technical and economic successes.

In the periods when the federal government had many of its largest successes in computer R&D, the computer industry was in its infancy. Now, however, that industry is very large and quite sophisticated. U.S. producers of computers, software, and related services sold roughly $260 billion worth of goods and services worldwide in 1989.[6] This amount compares with less than $5 billion in 1960 and $20 billion in 1970. Large, mature industries are likely to be more difficult to stimulate technologically than small, developing ones.

The federal role in computer R&D has been decreasing for several decades. In the 1960s, federal agencies spent between a quarter and a half of what private firms spent on R&D in the computer industry.[7] But in the late 1970s and 1980s, federal agencies did not keep up with the growth in private R&D in the computer industry: spending by federal agencies was between 10 percent and 15 percent of the private level.[8] The absolute levels of the private R&D efforts rose throughout this period. The National Science Foundation estimates that U.S. computer manufacturers alone spent roughly $10.5 billion of their own funds in 1989--the last year for which data are available--on R&D for computers, components, and software.[9] By contrast, in 1975, private firms spent $1.7 billion on such R&D. Given the reduced share of federal R&D, as well as the increased level of private R&D, one might assume that the federal role in promoting technological innovation was also reduced.

Most important, the computer and data communications industries have become much

4. See Congressional Budget Office, "A Review of Edwin Mansfield's Estimate of the Rate of Return from Academic Research and Its Relevance to the Federal Budget Process," CBO Staff Memorandum (April 1993), for a discussion of rates of return on R&D.

5. For a review of this literature, see Congressional Budget Office, *How Federal Spending for Infrastructure and Other Public Investments Affects the Economy* (June 1991), Chapter IV.

6. CBEMA Industry Marketing Statistics Committee, *The Information Technology Industry Data Handbook, 1960-2001* (Washington, D.C.: Computer and Business Equipment Manufacturers Association, 1991), p. 9. Other sources have a smaller, but still large, estimate of the U.S. computer industry. For example, McKinsey and Company, the management consultants, estimate 1992 worldwide sales at $208 billion. See McKinsey and Company, "The 1992 Report on the Computer Industry" (San Francisco, Calif., 1992), p. 2-13.

7. Before 1972, the National Science Foundation had no separate R&D category for the computer industry. Moreover, the NSF has not published federal spending for R&D in the computer industry for reasons of confidentiality since 1979. Thus, this analysis is based on the machinery industry, of which the computer industry is the largest component. The R&D and output statistics of the machinery industry are dominated by the computer industry, according to the National Science Board, *Science and Engineering Indicators--1991* (1991), p. 424. See *Science Indicators* for previous years. See also National Science Foundation, "Research and Development in Industry," various years. Using these sources, CBO was able to obtain a fairly consistent estimate of government and company R&D spending in the machinery industry and, to a less extent, the computer industry.

8. As noted, these figures are for the larger machinery industry; the computer industry R&D figures were probably closer throughout the entire period, while exhibiting the same downward trend.

9. National Science Board, *Science and Engineering Indicators--1991*, p. 424. Software producers (federal and private) spent an additional $2.7 billion in 1989. See Melissa Pollak, "Research and Development in the Service Sector," *The Service Economy* (Washington, D.C.: Coalition of Service Industries, July 1991), p. 4.

more competitive during the past decade. Many different companies compete for a relatively small part of this large market, and many federal R&D and regulatory programs have helped make this industry highly competitive. As a result, the pressures on firms in the industry are now quite different than when the computer industry was dominated by one company and data communications by another. In this new, more competitive market, the pressures on firms to introduce new technology are much greater than previously. Consequently, federal policies to stimulate innovation are less likely to play an important role overall.

Alternative Strategies for the HPCC Program

The strategy of the HPCC program for commercialization is consistent with a traditional federal approach. It aims to develop technology to fulfill the federal mission by using a mix of in-house expertise and outside contractors. Like previous federal programs, it emphasizes a single aspect of the technology--in HPCC's case, speed. Demonstration projects and other explicit efforts to transfer and commercialize the technology account for only a fraction of the resources devoted to the project.

Development of Everyday Applications

The Computer Systems Policy Project (CSPP), an industry organization representing the chief executive officers of major U.S. producers of computer systems, presented a different strategy for HPCC. Its recommendations for the network portion of the project included suggestions to "ensure the HPCC serves as a stepping stone to broader future information infrastructure. . . . This will require expanding activities under the NREN [National Research and Education Network] component to include research and development on the technologies needed to support broadly accessible and affordable networks."[10]

CSPP's specific concerns focused not so much on computer power or network speed, but on those aspects that might impede the spread of network technology. These aspects include security and privacy of data and individuals and issues of intellectual property rights, standards, and regulation. CSPP placed a great deal of emphasis on ensuring network compatibility and the broadest possible access to the system.

The technical choices the HPCC program is making to serve the interests of federal computational science researchers may not translate into features that the majority of commercial computer network users will find desirable. Although HPCC planners have expressed concern about the solution of a broad range of problems, including some that could contribute substantially to U.S. welfare, CSPP contends the initiative would directly contribute little to the lives of most citizens. The grand challenges are, in some sense, too grand.

CSPP would prefer to see more everyday applications of high-performance computing.[11] These uses would include giving broader access to electronic data bases, enhancing education through electronic field trips, and improving the life of housebound citizens. In the manufacturing arena, CSPP places as much emphasis on the integrative functions of data communications as on the "gee-whiz" aspects of designing products on computers. In this view, the HPCC program should not just aim to develop trillion-operation supercomputers

10. Computer Systems Policy Project, "Expanding the Vision of High Performance Computing and Communications: Linking America for the Future" (Washington, D.C., December 3, 1991), p. 11.

11. Computer Systems Policy Project, "Perspectives on U.S. Technology and Trade Policy: The CSPP Agenda for the 103rd Congress" (Washington, D.C., October 1, 1992), p. 8.

that talk to each other on billion-bit networks. Rather, it should include a wide array of computer resources, usable in a wide array of circumstances, and applicable to a wider array of national needs than originally envisioned.

In keeping with its alternative vision of the HPCC program, CSPP has recommended rearranging the program's budgetary priorities. Whereas most of the current funding focuses on a single design of supercomputer, CSPP would like to see a broader range of hardware and software configurations. Second, the HPCC program's software budget contains funds to purchase a great deal of hardware, while CSPP would emphasize more software funding. Third, it would expand the components for developing basic research and human resources. Finally, CSPP would shift the program's balance from advancing technology to applying it.

Although the HPCC program is commonly viewed as a network project, only 20 percent of the new funds for HPCC is for networks. CSPP believes that the network component of HPCC has the greatest potential and sees its low share of funding as an example of the imbalance in the HPCC budget.

In short, rather than being concerned about the program's meeting the grand challenges of the original HPCC documents, the CSPP is interested in ensuring that the technology has potential commercial use. Unless the new network technology addresses the concerns of their commercial customers, who represent the largest market share of demand for computer network goods and services, CSPP members feel it will not meet their needs.

Development of Self-Sustaining Markets

To a degree, CSPP's vision contains contradictory notions. The earliest users of advanced computer technology are likely to be involved in high-value activities, typically in business and medicine or in government, since only those activities can pay the costs high-performance systems usually involve. Electronic field trips for schools or the disabled, as envisioned by CSPP, will not be among the early uses of HPCC technology. Ordinary citizens are most likely to encounter such technology when they purchase a good or service that incorporates it--for example, medical records and images that are stored and transmitted electronically.

Indeed, HPCC technology is most likely to be commercially successful if it can first develop a broadly based market of high-value resources to store, distribute, and manipulate specialized information. By providing specialized information equipment and other resources to organizations willing to pay for them, it may enable CSPP members and other firms to bring down costs to the point that such equipment and resources become more widely used.

Competing Views of Commercialization

The differences in policy strategies discussed above largely result from divergent views of the process of technological change and the role of the government in it.[12] One view focuses on developing advanced technology, and the other on diffusing that technology.

In the first view, the federal government develops the technology for its own needs and leaves its further application to others. This view recognizes that there are many institutional obstacles to creating a market for new technology and that such market development is inherently risky. A basic tenet of this view is that, other than in a few areas such as setting of standards, the federal government can

12. For a more detailed discussion of different views of commercialization, see Congressional Budget Office, *Using R&D Consortia for Commercial Innovation: SEMATECH, X-ray Lithography, and High-Resolution Systems* (July 1990), pp. 5-10.

do little to overcome these obstacles and should leave their solution to the private sector, which has profit as the motivation for creating markets and overcoming obstacles.

In the case of HPCC, federal agencies will develop, or help industry to develop, very fast computers and networks and software tools for application to agency missions, including basic university research. Their diffusion to the rest of the economy, however, is left to the computer industry and telecommunications providers.

In the other vision of technological change, the federal government has a role to play in overcoming obstacles that could keep technology from spreading throughout the economy. The government might help to develop specific uses that could speed the spread of technology broadly into the economy--for example, developing less expensive, rather than faster, data communications technology, or making software easier to use. In this view, developing new markets and infrastructure is as filled with risk, uncertainty, and socially beneficial spillovers as is developing new technology. Given the costs, which are concentrated and large, and the social benefits, which are diffuse but in the aggregate can be greater than the costs, the federal government may in this view be uniquely positioned to sustain the investment necessary to spread new technology throughout the economy.

The federal track record in bringing new computer technology to market is mixed, however. The Advanced Research Projects Agency (ARPA) has certainly had much success in many areas of computing and electronics.[13] But it also has a 20-year history of promoting parallel high-performance computers, the market success of which is still quite limited (under $300 million in yearly sales). Even with some of its successful projects, such as ARPANET--the first computer network--ARPA has sometimes had to continue support for years beyond its original intent before a self-sustaining community of users developed.[14]

The Comparative Advantage of the Government

One trade-off that the Congress and the Administration face in the different facets of HPCC is the choice between pursuing near-term commercial implementation of HPCC technology and emphasizing the highest technical performance. For example, federal researchers stand at a comparative advantage in basic research on new ways of conceptualizing computer software. The reward system for federal R&D is based on technical excellence, often through peer review. By contrast, commercial providers of software achieve success--that is, profits--by being able to deliver software that addresses users' immediate concerns. Thus, although providing the next generation of hardware or software may ensure more commercial success for the technology developed under the HPCC program, federal researchers may not be in the best position to do that.[15]

Conversely, however, because HPCC program-sponsored supercomputers lack a large supply of commercial software, the sales of such computers are limited mainly to academic and government R&D centers. If the community of users is not sufficiently developed, the loss of federal backing could hurt the

13. The Defense Advanced Research Projects Agency (DARPA) recently changed its name to Advanced Research Projects Agency (ARPA), the name it was known by before 1972. For the role of ARPA in the development of computer technologies, see Richard Van Atta, Sidney Reed, and Seymour Deitchman, *DARPA Technical Accomplishments: A Historical Review of Selected DARPA Projects, Volumes I, II and III* (Alexandria, Va.: Institute for Defense Analysis, February 1990), especially Chapters 18 through 23 of Volume I and Chapters 13-19 of Volume II.

14. Advanced Research Projects Agency, "The Advanced Research Projects Agency, 1958-1974" (produced under contract by Richard J. Barber Associates, Washington, D.C., December 1975), pp. IX-57 through IX-59.

15. For a similar argument with regard to supercomputers, see Willie Schatz, "DARPA Rigs the Playing Field," *Upside* (May 1993), pp. 38-48.

efforts of private companies to create a market for their products.

This pattern of federal comparative advantage can be seen in other computer R&D efforts. In many of its more recent successes in high-performance computing, ARPA paid for university research to build a prototype and then left it to private initiative to bring this to market. University researchers operating on ARPA funds originally designed much of the technology underlying the specialized graphics, software, integrated circuits, and high-performance computers now commonly used in scientific and engineering computing circles.[16] Without ARPA funding, it is fair to argue, high-performance computing would probably be much less developed than it is today. But ARPA did its work without being directly concerned with the sales of the resulting product.

16. Van Atta, Reed, and Deitchman, *DARPA Technical Accomplishments*, Chapter 17 of Volume II.

Program Risks

By trying to be all things to all people, the HPCC program might lose its focus and dissipate its resources without effect. By focusing on one fairly narrow range of computing needs, the program may be more likely to advance the technology. The risk is that the technology will be irrelevant.

No strategy the HPCC planners choose will be without risk. On the one hand, federal R&D could be too near the marketplace, distorting choices and preempting private actions. On the other hand, the HPCC program could abandon such R&D prematurely to focus only on longer-term technology, leaving this technology an orphan, and perhaps stunting the growth of this market.

The HPCC Program

Policymakers designed the High Performance Computing and Communications program to improve the technology of supercomputers and data communications. This program focuses on hardware and software for supercomputers, data communications systems, and basic research and training. Several principal federal agencies are involved in developing different aspects of the technology.

The HPCC Budget

The HPCC program is essentially a continuation of ongoing federal work in the area of computer research and development with additional resources and more emphasis on coordinating federal efforts. Thus, the HPCC budget includes both existing and incremental funds.

The 1992 Plan

The program is intended to double the federal resources being devoted annually to developing high-performance computers and computer networks over the 1992-1996 period. The original funding request for 1992 was $640 million, rising to $1.2 billion in 1996. Funding over the five years would total $4.7 billion (see Table 1).

Since 1992, two factors have intervened to change the budget, though not substantially. First, other existing programs have been included in the HPCC tallies. For instance, the original 1992 plan included only $17 million for the National Library of Medicine, but the 1993 request included other parts of the National Institutes of Health, bringing the total to $45 million. The National Oceanic and Atmospheric Administration (NOAA) also increased its participation. The National Security Agency joined the HPCC program, adding $22 million to the total.

The second set of changes in the plan have resulted from policy decisions to change funding. For example, to allow for growth in funding for the space station within a constrained overall budget, the National Aeronautics and Space Administration (NASA) reduced its HPCC research and development for 1993 by $26 million. The Advanced Research Projects Agency has cut its funding by $8 million below the plan. The most significant change, however, has been in the National Science Foundation's appropriations for research and related activities for 1993, in which its HPCC activities received only a 12 percent increase rather than the planned increase of 23 percent.

Despite these changes, no new baseline budget exists for HPCC as a whole. The agencies regard their five-year budget projections for the program as confidential.

Incremental Funding

Establishing a firm estimate of the costs of the different components over the duration of the program is difficult. As noted above, this initiative includes many ongoing programs, some new ones, and additional resources that have been diverted to existing programs. Agencies may choose which of their programs to include in the HPCC. Specific agency participation is reflected in various pieces of authorizing legislation, if at all. Consequently, there is no single complete statement of the HPCC budget.

Details on the incremental funding that was included in the 1992 plan are presented by program area, not by agency. Not all program components received equally increased funding in the original plan. Table 2 presents the level of additional funding discussed in the HPCC planning documents.

Incremental funds totaled $1.9 billion over five years. The programs for High-Performance Computing Systems (HPCS) and Advanced Software Technology and Algorithms (ASTA) each received slightly more than one-third of the funds. The National Research and Education Network accounted for 20 percent, and Basic Research and Human Resources for the rest.

High-Performance Computing Systems

The goal of the HPCC program with regard to computer hardware is to develop computer architectures that can achieve a trillion mathematical operations per second--two or three orders of magnitude above current performance levels. The systems that seem most likely to develop this power first are called massively parallel supercomputers. Conventional supercomputers have a small number--usually fewer than 16--of very fast processing units working on a problem. In contrast, the massively parallel systems can have several hundred or even several thousand slower processing units.

What distinguishes the scalable architectures that the HPCC program is concentrating on from other massively parallel supercomputers is the idea that the same basic computer architecture should be able to be scaled

Table 1.
Original Five-Year Budget Plan for the HPCC Program
(Budget authority in millions of dollars, by fiscal year)

Agency	1992	1993	1994	1995	1996	1992-1996
Advanced Research Projects Agency	232	283	353	399	447	1,714
Department of Energy	93	110	138	157	168	666
Environmental Protection Agency	5	5	5	5	5	25
National Aeronautics and Space Administration	72	107	134	151	145	609
National Library of Medicine	17	17	17	17	17	85
National Institute of Standards and Technology	3	3	3	3	3	15
National Science Foundation	213	262	305	354	413	1,547
National Oceanic and Atmospheric Administration	2	3	3	3	3	14
Total	638	790	958	1,089	1,201	4,676

SOURCE: Congressional Budget Office using data from the Office of Management and Budget.

NOTE: Includes incremental funding.

Table 2.
Incremental Funding for the HPCC Program, Fiscal Years 1992-1996, by Program Component

Component	Funding: In Millions of Dollars	Funding: As a Percentage of Total
High-Performance Computing Systems	682	36
Advanced Software Technology and Algorithms	662	35
National Research and Education Network	390	20
Basic Research and Human Resources[a]	183	10
Total	1,917	100

SOURCE: Congressional Budget Office using data from Office of Science and Technology Policy, "The Federal High Performance Computing Program" (September 8, 1989), p. 46.

NOTES: HPCC = High Performance Computing and Communications.

Incremental funding is the additional money appropriated to the HPCC agencies above the level the agencies were receiving at the start of the HPCC program for their activities in this area. Numbers may not add to totals because of rounding.

a. In addition, 15 percent of each of the other three program components are to be devoted to this area.

linearly until it is the right size for the task at hand. Scaling refers to the addition of processors to the system so that the system's performance increases in proportion to the number of processors added.

Computer design visionaries believe that the proliferation of computer designs using the same microprocessors is part of the new wave in computer architecture.[1] In the future, they argue, computers will not be categorized as a microcomputer, minicomputer, mainframe, supercomputer, and so on, each with a different architecture. Rather, computers will be characterized by the number of standard microprocessors in the machine: one, a few, many, and very many. Most important software applications will be perfectly transportable between them, merely running faster as more processing nodes are added to the computers.

Most modern U.S. top-of-the-line supercomputers, including conventional supercomputers, are parallel in the sense of having more than one processor, and the trend is clearly to increase the number. Where the technology strategies differ between conventional and massively parallel machines is in how fast to increase the number of processors. To achieve the highest peak speeds, HPCC has opted to increase it rapidly.

The type of hardware-scaling that HPCC is exploring can, to some extent, be replicated in software today. Although many high-performance computer systems are not scalable, they have software that is compatible with that of more powerful computers. For example, the entry-level Cray Y-MP EL--a conventional supercomputer--has roughly 1 percent to 3 percent of the processor performance of the top-of-the-line Cray Y-MP C90 at between 1 percent and 3 percent of the cost. They both can run the same software, but at different speeds. Conventional supercomputers, however, cannot be upgraded easily. Buying a new machine is easier and less costly than making an existing machine work more rapidly. Computers with scalable architecture, therefore, allow users flexibility in tailoring the hardware, and consequently the speed, to the task at hand.

Components

The development of supercomputer hardware focuses on four areas:

- *Research for future generations of computing systems* addresses the concepts underlying advanced scalable architecture, components, packaging, and systems software. The goal of this research is to ensure that the requisite technology is available for new systems and to provide a solid techno-

1. David Patterson, "Expert Opinion: Traditional Mainframes and Supercomputers Are Losing the Battle," *IEEE Spectrum* (January 1992), p. 34.

logical base for the next generation of high-performance computers.

- *System design tools* focuses on the development of computer-aided design tools and frameworks for integrating multiple tools. The tools include computer-aided design, analysis, simulation, and testing. The frameworks are to allow computer designers to take data or output from one computer design tool and integrate it as seamlessly as possible into other design tools. The intent of this effort is to speed up construction of prototypes.

- *Advanced prototype systems* encompasses the development of experimental systems. The dual goals of these activities include both increasing the speed and reducing the size and cost of the scalable supercomputer hardware, thus making it accessible to a broader class of users. The federal government and the supercomputer vendors will share development costs.

- *Evaluation of early systems* places experimental machines with experts who use and evaluate them. Such evaluations include establishing universally agreed-upon performance criteria and widespread publication of results. Some of the early experimental systems may begin work on the grand challenges discussed in Chapter 1, such as some experiments in alleviating pollution or in high-energy physics.

Budget

The importance of the hardware component of the HPCC program is reflected in its share of overall funding. In the original budget plan, hardware accounted for roughly 36 percent of all additional funds, or $680 million over five years. It was scheduled to increase steadily in annual increments of between $30 million and $50 million through 1996. More than two-thirds of the hardware funding, roughly $460 million, was for the manufacture and testing of prototypes of advanced-design supercomputers.

Advanced Software Technology and Algorithms

Because these massively parallel supercomputers operate under a fundamentally different architecture than other computers, they do not have the software library, developed over 40 years, that other computers inherit. This element of the HPCC program will produce generic software and algorithms for research applications. It also contains funding to create complete supercomputer applications that can work on supercomputers attached to networks. The goals are to help solve the application problems posed by the grand challenges, make the software easier to use and more reliable, and apply the knowledge so gained to other computational problems.

Components

Development of the software for high-performance computers covers four areas:

- *Software support for grand challenges* aims to develop high-performance software to address issues related to the grand challenges, including global climate change, advanced materials research, and biotechnology. The problems are being selected on the basis of their national importance, their potential to advance computational technology, their potential to yield spill-over benefits, and the extent to which other parties--particularly private firms--might share the development costs.

- *Software components and tools* seeks to develop applications for high-speed computer networks and generic software for high-performance computers. Such software includes that which currently is available for conventional computers--for example, software that enables programmers to design fast software or helps them find errors in the software they have written. Because

HPCC involves many agencies and private vendors, this software is intended to operate on different brands of supercomputers.

- *Computational techniques* funds research on algorithms, parallel languages, and numerical analyses. The advent of massively parallel supercomputers requires the development of techniques optimized in ways different from those of single-processor machines.

- *High-performance computing research centers* supports the deployment of new HPCC computer hardware to research centers and provides fresh opportunities for research as well as transfer of technology to large numbers of researchers. The intent is to provide these centers with the new hardware and connect them through the National Research and Education Network. Doing so will permit a wide range of software experiments.

Budget

The software element accounted for about 35 percent of the incremental HPCC funding, or roughly $660 million. About half of that will be spent for R&D on tools and techniques, and about 30 percent will fund the research centers. The grand challenges themselves account for less than a quarter of incremental spending for software.

National Research and Education Network

The National Research and Education Network (NREN) is intended to make data communications services available to more researchers and to government and educational institutions. It would also increase the speed and quality of such services. A by-product of this effort would be the further development of the commercial high-speed data communications industry by using the research and education communities as the first customers--the communities in which the applications and other ancillary technologies are most likely to be developed.

Components

This portion of the HPCC program has two components:

- *The Interagency Interim National Research and Education Network* is based on the National Science Foundation Network (NSFNET), the Department of Energy (DOE) Energy Sciences Network, the National Aeronautics and Space Administration's Science Internet, and other networks supporting research and development. This component includes a phased process of upgrading and expanding the U.S. portion of the research-oriented Internet so that it can serve more of its intended community with, insofar as feasible, data transmission rates of 1 billion bits (one gigabit) per second.[2] The interagency network has been upgraded from transmission speeds of 1.5 million bits per second, commonly called T-1, to 45 million bits per second, commonly called T-3. The DOE is also upgrading its research network.

- *Gigabit research and development* funds R&D on network switches, protocols, and other hardware and software needed to deploy the initial gigabit network. Much of this R&D is to be accomplished through six joint private/public research test beds. A key feature of this program is its close collaboration with industry to encourage the transfer of the resulting technology to the U.S. telecommunications industry.

2. Internet is an international network of computer networks all using a common communications protocol and roughly centered in and around research and educational institutions. For a more substantial discussion of Internet, see Chapter 4.

Budget

In the plan, the network portion of the HPCC program would receive only 20 percent--or $390 million--of the incremental funding. This funding would be equally divided between setting up the interim NREN and doing R&D on the next generation of networks.

Basic Research and Human Resources

The central goal of this portion of the HPCC program is to encourage the long-term growth of high-performance computing by ensuring that the infrastructure for basic research is developed and that students now being trained as the next generation of programmers have the opportunity to pursue such research.

Components

This element of the HPCC program has four components:

- *Basic research* aims to increase awards for R&D on high-performance computing initiated by individual investigators. Such basic research may relate to one or more of the three main HPCC areas--massively parallel supercomputer hardware, software, and networking.

- *Research participation and training* provides more advanced training through workshops, fellowships, and other activities at the postsecondary level.

- *Infrastructure* provides upgrades of both equipment and software for computer science research facilities at universities.

- *Education, training, and curriculum* involves the preparation of materials and courses for improving computer education, including more teacher training.

Budget

In the original plan, only 10 percent of all new funds for HPCC were directly assigned to basic research and human resources. However, HPCC planning documents stipulated that 15 percent of HPCS, ASTA, and NREN funding would be devoted to this general area.

Agency Responsibilities

The HPCC program divides its tasks not only into the four elements, but also along agency lines. The program includes eight agencies with some level of responsibility for portions of HPCC R&D (see Table 3). Although the agency count has been rising, there is no single lead agency. The Office of Science and Technology Policy is responsible for coordinating the individual agency efforts.

The Advanced Research Projects Agency has major responsibility for R&D on hardware and generic software tools for both scalable supercomputers and networks. ARPA is funding the development of scalable massively parallel supercomputers and much of the basic software needed to operate these new systems successfully, such as tools for debugging and compiling software. ARPA is also paying half the federal portion of the gigabit test beds, where much of the network hardware R&D will occur.

The Department of Energy's mission and history has endowed it with much of the federal government's expertise with supercomputers. DOE therefore has the task of evaluating the scalable supercomputer systems developed under the HPCC program. The department is also the potential user of these supercomputer systems for research into nuclear weapons, global warming, and other energy matters and is responsible for developing software applications in these areas. As a potential user of high-speed computer networks, DOE will begin a major upgrade of its Energy

Sciences Network to link its facilities as part of the HPCC program and will probably be the first federal agency to provide near-gigabit speeds over a network.

The other agency with responsibility for overall hardware is the National Science Foundation (NSF). Its principal responsibility is to coordinate deployment of the National

Table 3.
Agency Responsibilities Under the HPCC Program

Agency	High-Performance Computing Systems	Advanced Software Technology and Algorithms	National Research and Education Network	Basic Research and Human Resources
Advanced Research Projects Agency	Technology development and coordination for supercomputer systems	Technology development for parallel algorithms and systems software	Technology development and coordination for gigabits networks	University programs
Department of Energy	Technology development Systems evaluation	Energy applications research centers Energy grand challenge and computation research Software tools	Access to energy research facilities and data bases Gigabits applications research	Basic research and education programs
National Aeronautics and Space Administration	Aeronautics and space application test beds	Software coordination Computation Research in: Aerosciences Earth and space sciences	Access to aeronautic and spaceflight research centers	Research institutes and university block grants
National Science Foundation	Basic architecture research Prototyping experimental systems	Research in: Software tools, data bases Grand challenges Computer access	Facilities coordination and deployment Gigabits research	Programs in: Basic research Education Training/curricula
National Institute of Standards and Technology	Research in systems instrumentation and performance measurement Research in interfaces and standards	Research in: Software indexing and exchange Scalable parallel algorithms	Coordinate performance assessment and standards Programs in protocols and security	
National Oceanic and Atmospheric Administration		Ocean and atmospheric computation research Software tools Techniques	Oceanographic and atmospheric mission facilities Data-base access	
Environmental Protection Agency		Research in environmental computations, data bases, and application test beds	State environmental mission assimilation	Technology transfer to states University programs
National Institutes of Health	Systems evaluation and performance analysis	Medical application test beds for medical computation research	Development of intelligent gateways Access for academic medical centers	Basic Research Internships for parallel algorithm development Training and career development

SOURCE: Congressional Budget Office based on data from the Federal Coordinating Council for Science, Engineering, and Technology.

Research and Education Network in several stages, depending on the pace of the R&D accomplishments. NSF also houses the basic research efforts in both hardware and software.

The remaining agencies' responsibilities for developing hardware and related software applications are largely associated with their individual missions: NASA, space; NOAA, oceanographic and atmospheric research; the Environmental Protection Agency, the environment. Only the National Institute of Standards and Technology, which sets some network standards, has crosscutting responsibility.

Role of the Federal Coordinating Council on Science, Engineering, and Technology

The federal entity responsible for coordinating HPCC research and development is the Federal Coordinating Council for Science, Engineering, and Technology (FCCSET) within the Office of Science and Technology Policy. Agencies participating in HPCC meet regularly under the auspices of FCCSET. However, FCCSET has no budgetary power or other authority to realign priorities or make binding decisions. Agencies attend meetings and present proposals for new R&D on data communications or scalable supercomputers, but FCCSET neither sets priorities nor decides which programs should go forward. Thus the National Aeronautics and Space Administration can decide whether to reduce HPCC funding to accommodate the space station or other high-priority programs, but no one has the authority to decide HPCC funding trade-offs among agencies.

Much of the high-speed data communications work being done within each agency is also not necessarily represented at FCCSET. For example, the National Library of Medicine represents the National Institutes of Health, even though much of the work regarding medical imaging technology at NIH is done elsewhere and, in the past, had been done independent of the HPCC program. Similarly, efforts by the armed services to transmit medical images at relatively high speeds have been undertaken outside HPCC. (Part of the problem in describing the HPCC budget comes from the fact that member agencies continue to increase the portion of their activities described as belonging in the HPCC program.)

The HPCC program has addressed this problem by creating a National Coordination Office for High-Performance Computing and Communications, which provides the Congress and the public with a point of entry into the initiative. The National Coordinating Office is responsible to a FCCSET committee and its director is a special assistant to the Director of the Office of Science and Technology Policy. This new office is intended to coordinate interagency budget and reporting requirements, but it has neither a budget nor a full-time staff. Furthermore, the director is also the director of the National Library of Medicine, which presumably will continue to occupy a substantial portion of his time.

Chapter Three

HPCC and the Development of Supercomputer Markets

One intent of the High Performance Computing and Communications program is to ensure continued leadership of U.S. firms in the high-performance computing arena, including supercomputers and other fast computers, in the belief that such leadership would enable these firms to increase their performance in the other segments of the computer market. Policymakers also expect to make the benefits of additional computer technology more widely available by stimulating development of more applications for supercomputers. They hope that increasing access to high-performance computers and increasing the quality of the software for those computers will accelerate the widespread diffusion of this technology throughout the economy.

HPCC and Supercomputers

The stated goal of HPCC program officers is to apply leverage to federal monies by funding only those areas where they feel private dollars are not likely to go and that appear to be on a path of rapid technological advance. Consequently, they are more likely to fund revolutionary than evolutionary advances.

To date, the HPCC supercomputer program has developed some early prototype systems that operate at 100 billion operations per second. Given the rapid rate at which semiconductor technology advances and the intense competition among producers of supercomputer hardware, HPCC will most likely deliver a prototype of a supercomputer with a peak performance of 1 trillion operations per second. Whether this machine will be able to perform useful work for a wide variety of users is more open to question.

The Promises of New Supercomputer Technology

HPCC planners expect to produce supercomputers that cost the same as today's supercomputers but deliver between a hundred and a thousand times the power. This added power would permit supercomputers to solve problems that are beyond the capabilities of current supercomputers. Such solutions might give U.S. industry a competitive edge by enhancing the design and analytic capabilities of U.S. firms. From a policy perspective, then, HPCC will yield substantial benefits if the power of the scalable massively parallel supercomputer expands the uses and, hence, the market for high-performance computers. In this sense, HPCC is trying to shift both the supply curve for supercomputers (that is, expand the supply at a given price) and the demand curve by creating new uses for the improved supercomputers.

The new supercomputer technology could also directly strengthen the competitiveness of the U.S. computer industry. Planners believe that supercomputer technology leads all computer technology--that is, improvements appear on these machines first, are tested and improved, and then move onto other computers. In this way, the program would further enhance the lead of U.S. technology in computers overall.

Unsettling Questions

This vision of success is not in complete correspondence with the market for high-performance computers. Far from driving computer technology, supercomputers have largely been involved in areas of technology that are not growing rapidly. Even now, massively parallel supercomputer technology draws largely on mainstream computer component technology rather than leading it. But it would appear that HPCC has the potential to improve software and processor communications technology in ways that might have wider application.

HPCC's vision of the development of computer technology also ignores the most important development in high-performance computing of the 1980s: the engineering workstation. Sales of these technical computers--high-performance versions of the personal computer--have grown substantially, just as supercomputer sales have slowed, suggesting that workstations are competing with supercomputers for many uses. If this analysis is correct, HPCC is attempting to increase computer speed at a time when the market, including the scientific market, may be valuing other factors such as cost and convenience more highly than raw speed.

Finally, the massively parallel supercomputers, after a decade of federal R&D support, have yet to demonstrate their utility as general-purpose computers. Translating existing software for them has proved difficult. For a variety of reasons that will be explored below, they match the performance of conventional supercomputers only in the most specialized problems. They are generally cheaper, however, than the high-end conventional supercomputers and may gain a following on that basis.

Massively parallel computer systems already have a substantial market in commercial data-base management as specialized (single-purpose) machines. Half of all mainframe computer use involves data-base manipulation; even if their only sales were to customers with such uses, parallel architecture computers would have a substantial market. Whether they can move into general-purpose use is at present unclear.

Analysis of the Supercomputer Market

The supercomputer market represents only a minute fraction of annual computer sales and an even smaller fraction of installed computers. Roughly 10 million to 15 million personal computers are sold each year. Engineering workstations account for 300,000 to 400,000 units. Computer producers also sell 10,000 to 20,000 mainframe computers each year.[1]

By contrast, only a few hundred supercomputers are sold each year.[2] These machines have a high value, however, totaling about $1.8 billion in worldwide sales in 1992.[3] As of 1990, slightly more than 600 supercomputers were operational worldwide.[4]

Growth in the demand for supercomputers has slowed substantially in recent years. According to one recent industry estimate, the world market for supercomputers grew from sales of less than $90 million to more than $1 billion between 1980 and 1988. Since then,

1. Egil Juliussen and Karen Juliussen, *The Computer Industry Almanac, 1989* (New York: Simon & Schuster, 1988), pp. 3.4-3.5.

2. John B. Jones and Theresa Liu, "Cray Computer Corporation" (Montgomery Securities, San Francisco, Calif., July 23, 1991), p. 4.

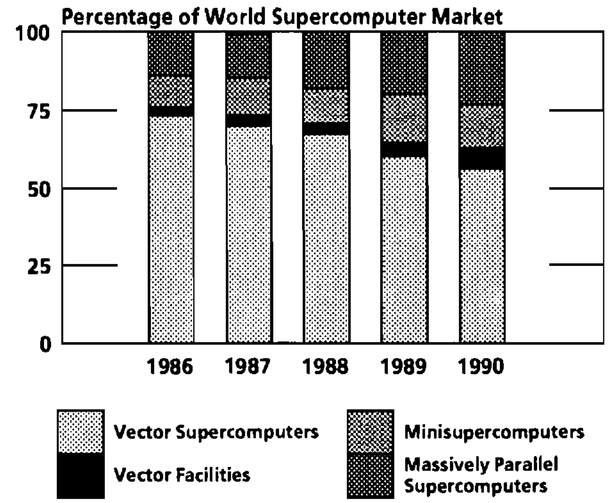
3. Department of Commerce, *1991 U.S. Industrial Outlook* (1991), p. 26-6.

4. Los Alamos National Laboratory, Department of Energy, "High Performance Computing and Communications: Investment in American Competitiveness" (prepared under contract by Gartner Group, Stamford, Conn., March 15, 1991), p. 75.

however, demand for supercomputers has grown only 6 percent annually.[5] Demand for entry-level supercomputers, which are typically much less expensive and almost as powerful, has remained high.

The recent slowdown is attributable in part to the decline in federal demand brought about by the end of the Cold War and the recession, but also to the increased power of engineering workstations, which will be discussed below.[6] Assuming that defense spending continues to decrease throughout the 1990s, demand for supercomputers may not resume its upward climb for 5 to 10 years unless emerging environmental or other applications grow sufficiently to replace military demand.

Figure 1.
Supercomputer Market Segments



SOURCE: Congressional Budget Office from Department of Commerce.

Supercomputer Market Segments

The supercomputer market has four general submarkets: vector supercomputers, minisupercomputers, massively parallel supercomputers, and vector facilities (see Figure 1).

Vector supercomputers usually consist of between 1 and 16 very fast processors with special features that allow them to work on strings of numbers (vectors) quite rapidly. The market for large vector supercomputers is traditionally associated with the name Cray Research, but is also occupied by a handful of Japanese and, until recently, U.S. firms.[7] In 1991, roughly 60 of these vector supercomputers were shipped worldwide.[8]

The minisupercomputer (or departmental supercomputer) is sold to cost-sensitive users or those with limited needs.

The largely experimental massively parallel machines using arrays of inexpensive microprocessors have started to sell more widely. Currently, most such sales are for smaller versions of the massively parallel machines with fewer than 50 or 100 microprocessors. Such configurations allow purchasers to experiment with the technology at relatively low cost.

At the rapidly growing end of the market is the vector facility, which takes a mainframe computer and supercharges it with special capabilities. Vector facilities can rapidly do

5. Los Alamos National Laboratory, "High Performance Computing and Communications," p. 73. Other analyses differ in both absolute size and rate of growth. However, there seems to be more general agreement that growth in the demand for the largest supercomputers has slowed. See Department of Commerce, *U.S. Industrial Outlook 1991*, p. 28-5. At least part of the difference in size of total market comes from the imprecise nature of the definition of the term supercomputer.

6. Between 1984 and 1989, public agencies bought an average of 17 of the top-of-the-line supercomputers per year, compared with only 11 in 1990. Private purchases fell from an average of 13 to 10. See Department of Commerce, Technology Administration, "Global Markets for Supercomputers: The Impact of the U.S.-Japan Supercomputer Procurement Agreement" (October 1992), p. 14.

7. Sometimes vector supercomputers are also called vector parallel supercomputers because many of them have between 4 and 16 processors. Most of the time each processor is assigned to a different user--that is, they are not used in parallel.

8. Department of Commerce, *1992 U.S. Industrial Outlook* (1992), p. 27-6.

computations on strings (vectors) of numbers. The rest of the computer program remains unaffected. Since programs that need supercomputers typically have large strings of computations, vector facilities can give a mainframe some supercomputer capabilities. The Department of Commerce data do not break out the cost of the vector facility from that of the mainframe computer.

Sources of Supercomputer Demand

The government is the largest buyer of the large vector supercomputers and the massively parallel processor-based supercomputers. This demand has been driven largely by the need to design nuclear weapons at the Department of Energy's national laboratories. Since the national laboratories were the first to adopt many of the new computing technologies, their use carried with it a stamp of approval. In addition, the national laboratories served the socially useful function of being on the so-called "bleeding edge" of technology: they made the mistakes and absorbed the costs of making the new models of supercomputers usable.

At the opposite end of the spectrum are very price-sensitive users of supercomputers. These users, typically private firms, have a more mundane use for the supercomputer: designing better products. The cost-sensitive user currently has three options: the minisupercomputer, a vector facility attached to a mainframe computer, or a high-end workstation. Despite the presence of high-end workstations, the low end of the supercomputer market is growing more rapidly than the high end.

The market for massively parallel supercomputers has been growing rapidly, up roughly a third in 1992 to almost $300 million, according to Department of Commerce estimates.[9] Most of this growth, however, reflected price increases or increases in average computer size; unit sales rose only 8 percent. Most demand is still in federal laboratories or supercomputer centers, where the machines are available as part of the general computational services the centers provide. Commercial users are still relatively rare. Although there has been a great deal of discussion about "massively" parallel machines, most units currently sold can best be described as "moderately" parallel--that is, they have fewer than 64 processing nodes. Their moderate parallelism reduces the size of the initial investment and the complexity in testing this new technology.

The Supercomputer Market and Technology Spinoffs

Supercomputers are widely perceived to be the motivating force behind advances in computer technology.[10] According to this view, new computer technologies are introduced through this small but crucial market at the apex of the technology pyramid; successful technologies then trickle down to the broader market at the base. Although the market for supercomputers is small, it is so performance-driven that it allows computer companies to test new ideas, thus providing a platform for new technologies that mainstream computer users will see only later.

A practical corollary of this view is that investments in supercomputer technology may provide society with a higher return than similar investments in other parts of computer technology because these investments will trickle down to other users. In this way, adherents believe, federal investments in supercomputer technology may benefit U.S. computer manufacturing firms by providing them

9. Department of Commerce, *1993 U.S. Industrial Outlook* (1993), p. 26-7.

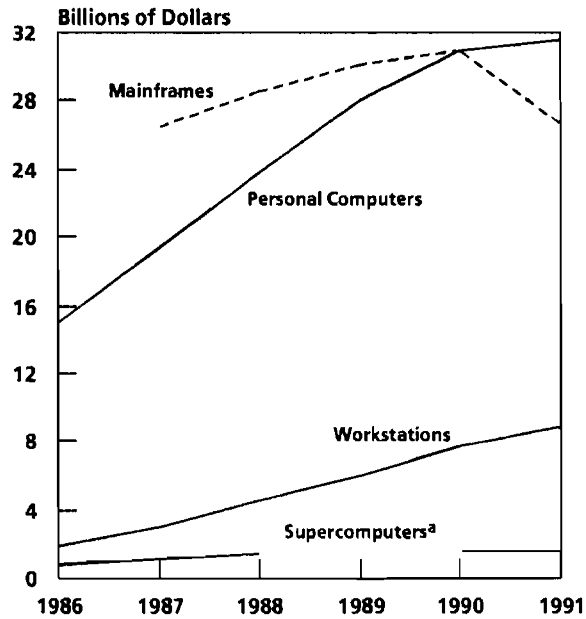
10. For example, see Office of Technology Assessment, *Competing Economies: America, Europe and the Pacific Rim* (October 1991), p. 253.

with access to technology before their foreign competitors.

Contradicting this view, however, is the secondary role supercomputer technology has played in two major advances of the 1980s: improved integrated circuits, and reduced-instruction-set computer (RISC) architecture.[11] These advances have been driven instead by the other end of the spectrum of computer technology--personal computers and engineering workstations. Of course, some new technology has come from supercomputers, but the major thrusts of computer technology development lie elsewhere.

The workstation and personal computer markets have been at the heart of technology advances because the demand for such computers has grown rapidly enough to justify devoting resources to solving the problems these markets confront (see Figure 2). Around 1980, all three market segments--personal computers, workstations, and supercomputers--were small. Although the supercomputer market has grown to $1.8 billion, it has been dramatically outpaced by growth in both the personal computer and workstation markets, which in 1991 exceeded $30 billion and $8.8 billion, respectively.[12] Given the levels and growth in sales, companies anxious to become market leaders have pushed every aspect of the technology, with the result that these markets see whole turnovers in their technology every few years.

Figure 2.
Relative Computer Market Sizes



SOURCE: Congressional Budget Office from Department of Commerce data.

a. The Department of Commerce provided no estimate for 1989 supercomputer shipments consistent with those of other years.

The manufacture of supercomputers, on the other hand, has not led to significant advances in general computer technology. The trade-offs in cost and performance that supercomputer producers face are different from those of most other computer producers. For example, vector supercomputers must often be equipped with specialized liquid cooling facilities. Even when the manufacturers use more conventional technology, its application is likely to differ substantially from the practices of most electronics equipment producers. In limited areas, however, supercomputers have made a contribution to technology, most notably in software. As engineering workstations have become rapid enough to provide timely mathematical solutions to engineering and scientific problems, software originally written for supercomputers has been translated for workstations.

The trickle-down theory of technology may have corresponded to reality in the past. In

11. In RISC architecture, the computer can perform fewer instructions, but it performs them rapidly. Complex instructions are then put together from simple individual instructions. Conventional computer architecture, by contrast, includes many more and complex instructions, which are performed more slowly.

12. Supercomputer sales are based on Department of Commerce, *U.S. Industrial Outlook 1992*, p. 27-6. Personal computer sales are from International Data Corp, a market research firm, and appeared in Jim Seymour, "Platforms: How the PC Stacks Up," *PC Magazine* (May 12, 1992), p. 115. Engineering workstation sales are based on Dataquest estimates in Rick Whiting, "A Clash of the Giants Shakes the Low-End Workstation Market," *Electronic Business* (April 27, 1992), p. 87. Because the workstation and personal computer markets overlap, the two numbers should probably not be added together.

the 1950s and 1960s, high-performance computers led the general advance in computer technology. For example, the IBM Stretch computer, a high-performance predecessor to the famous IBM 360, pioneered a technique called pipelining to speed up computers.[13] Now all modern computers have elements of pipelining. But more recent years may provide a better indication of conditions in the near future. Recent spinoffs from supercomputer technology have been concentrated in the mainframe computer market, where sales have leveled off and their relative market value has declined to roughly one-quarter of computer sales.

Recent Trends in Computer Technology Development

The more significant advances in technology have occurred in integrated circuits, computer architecture, printed circuit boards, and computer disks. Appendix A describes major areas of computer technology in greater detail.

Integrated Circuits. The most notable semiconductor advances of the 1980s--such as those in computer memories and microprocessors--were made with the complex but relatively slow integrated circuits used in personal computers and workstations. By contrast, most vector supercomputers have used simple, fast integrated circuits. Semiconductor manufacturing technology has also been driven by the needs of the complex, slow integrated circuits. The economies of scale of these large markets, combined with improved manufacturing technology, allowed these integrated circuits to incorporate more functions while becoming faster, making up most of the original difference in speed between the integrated circuits used for small computers and those used for supercomputers.

Computer Architecture. The most important advance in computer architecture has been the proliferation of reduced-instruction-set computing designs; conventional supercomputer design, on the other hand, has concentrated on vector arrays. Although the Advance Research Projects Agency funded much of the early research on RISC as part of its efforts in high-performance computing, including massively parallel processors, the rapid development of this technology during the last decade is largely the result of private initiative in the workstation market. The most commercially successful massively parallel supercomputers now incorporate the RISC microprocessors developed for commercial workstations.

When RISC designs first emerged, designers copied many of the features of existing mainframe and supercomputer architectures. However, RISC design moved aggressively, surpassing existing computer designs in many areas.

Printed Circuit Boards and Other Packaging. Producers of conventional electronics equipment have had to accommodate fewer numbers of relatively more complex integrated circuits on their printed circuit boards. Conversely, supercomputer designers accommodate many relatively simple integrated circuits on their printed circuit boards. Because of the heat generated by many small but fast integrated circuits, supercomputer designers have had to turn to water and even freon cooling, while most computers remain air-cooled.

Computer Disks. Although supercomputer designers demanded speed from their disks, the bulk of demand for computer disks focused on capacity, size, and price. The result has been a computer disk industry largely geared to serving the growing market for small computers. Now the newest disk technology--known as redundant arrays of inexpensive disks (RAID)--is coming to supercomputers,

---

13. Pipelining can best be thought of as an assembly line in which many different instructions are in various stages of completion at all times. See David Patterson and John Hennessy, *Computer Architecture: A Quantitative Approach* (San Mateo, Calif.: Morgan Kaufmann Publishers, 1990), p. 338. For a more complete history of the Stretch computer and its contributions to computer technology, see Kenneth Flamm, *Creating the Computer: Government, Industry and High Technology* (Washington, D.C.: Brookings Institution, 1988), pp. 90-95.

after first being used in other computer markets.

Future Directions of Technology Spinoffs

No one can predict the origins of commercially successful ideas and technology. The history of technology development is replete with innovative tools and ideas originating from unlikely circumstances. In most cases, however, technology stays where it was invented; spinoffs, though desirable, are the exception, not the rule. If supercomputer technology remains true both to its own history and to the history of technology development in general, it will probably not result in many component spinoffs into other parts of the computer industry, especially the rapidly growing segments of that industry.

Component Technology. Unlike previous generations of supercomputer technology, the massively parallel supercomputers, such as those being developed by HPCC, use many of the same components as engineering workstations. Indeed, the very existence of massively parallel supercomputers is predicated on the availability of high-performance commercial semiconductor technology. Much of the improvement in the performance of massively parallel supercomputers during the last decade can be attributed to the use of increasingly sophisticated workstation microprocessors.[14]

In fact, within the market for massively parallel supercomputers, the use of microprocessor technology designed specifically for supercomputers is often regarded as a weakness. Such technology, because it is not continually pushed by the brutally competitive forces of the workstation market, is felt to be in danger of falling behind its commercial counterparts. Technologists in the field therefore see component spinoffs flowing mainly from workstation technology to supercomputer technology.

Some HPCC advocates argue that the price of computer hardware will decrease substantially below recent trends because of massively parallel supercomputers. But massively parallel supercomputers are likely to represent only a small fraction of the market for most computer components, and hence are not likely to be the source of economies of scale or of learning curves. Consequently, HPCC will probably not result in a drastic drop in the price of computer hardware, as the expansion of the personal computer and computer workstation markets did in the last decade.

Areas of Likely Spinoff. Two of the likeliest areas of spinoff from the massively parallel supercomputer to other types of computers do not involve components but communication between processors and the creation of parallel solution paths in software problems.

Most massively parallel supercomputers are composed of many RISC microprocessors. In getting these individual microprocessors to communicate and coordinate rapidly, HPCC will have to confront the exigencies of processor communication to a much greater extent than does the computer market as a whole. Both the speed and frequency of communication between microprocessor nodes that designers of these systems must deal with present greater challenges than computer designers have had to face until now. As more computers are joined in computer networks, the problems of coordination will become more severe, and the experience gained in trying to tie microprocessors together within a massively parallel supercomputer may become relevant.

The transfer of technology may be limited, however. Communication and coordination among microprocessors in a massively parallel supercomputer must be rapid and tight. By contrast, most computer network applications involve slower communication and looser coordination. Whether the principles needed to solve problems of tight coordination and rapid

14. For this history of improvement, see John Van Zandt, *Parallel Processing in Information Systems* (New York: John Wiley and Sons, 1992).

communication can be economically applied to more general computer communications is not clear. As noted above, supercomputers push performance to such extremes that many of the solutions they devise are uneconomic or inappropriate for other types of computers.

The other area of substantial innovation in massively parallel supercomputer technology involves making problems parallel. This process involves rethinking problems and rearranging them so that their component elements can be worked on simultaneously. The methods of finding parallels in problems may have wider applicability. For example, researchers who have rewritten software applications to make them parallel for the massively parallel supercomputer have discovered that these applications also run better on vector supercomputers, sometimes faster than on the parallel system.[15] This effort to make internal parallelisms more explicit may be one of the areas in which massively parallel supercomputer researchers advance general computer technology.

Obstacles to Widespread Commercial Use of HPCC Technology

The main obstacle to developing widespread commercial markets for HPCC-developed supercomputer technology is not that a competing supercomputer technology will emerge, but rather that cheaper workstation technology will preempt further substantial growth of the supercomputer market as a whole. At present, more commercial design problems are being solved on workstations than on supercomputers.

Another obstacle to the widespread use of HPCC-sponsored supercomputer technology in the near term can best be described as economic inertia: current supercomputer hardware and software are very good and continue to improve. Furthermore, close to 20 years' worth of software has been written to run rapidly on that hardware. Given the difficult problems of writing software for massively parallel supercomputers, one cannot expect the massively parallel supercomputers to supplant vector supercomputers soon; they do not yet seem able to run most software as rapidly as the vector supercomputers. Nevertheless, supercomputer centers are buying massively parallel supercomputers as supplements to their vector supercomputers so they can serve clients who need the additional power. This market is still developing.

The most apparent area of near-term commercial demand for massively parallel computers is in data-base management. The market for parallel special-purpose computers for managing data bases is almost as large as the market for massively parallel supercomputers. It remains to be seen whether the producers of massively parallel supercomputers can reposition their products to serve this market, or whether the firms that currently serve this market can break into the general-purpose computer market.

Most analysts believe that parallel supercomputers and, more generally, parallel computers, are indeed the wave of the future. Experts disagree on how near and how parallel that future is and on whether the economy will get there through incremental steps or major leaps.[16] By trying to leapfrog the development of intermediate technology, producers of massively parallel supercomputer hardware risk leaving software producers and users with no simple path for the evolution of their products.

15. George Cybenko and David Kuck, "Revolution or Evolution?" *IEEE Spectrum* (September 1992), p. 41.

16. Cybenko and Kuck, "Revolution or Evolution?" pp. 39-41.

Growth in the Engineering Workstation Market May Slow Further Substantial Growth in the Supercomputer Market

As growth in the supercomputer market has slowed, the engineering workstation market has continued to grow at a very high rate, as shown in Figure 2. Since both types of computers are used for numerical analysis and are--to some extent--interchangeable, workstations may be performing work that previously would have been performed on a supercomputer. Both markets were small in the early 1980s, but the workstation market has grown to several times the size of the supercomputer market.

The increasing power of engineering workstations has spurred their growth relative to that of supercomputers. The speed and low cost of many high-end engineering workstations make them a logical substitute for a supercomputer. They are quite popular among engineers and designers; the workstation is already nearby and, at the margin, it is virtually free to use. Consequently, advances in workstation technology and speed may further limit the supercomputer market.

Even the positive developments in software for supercomputers can become a liability for supercomputer hardware sales. As supercomputer software becomes transferable to workstations, it establishes a self-reinforcing trend. Software programs that previously could run only on supercomputers have already been transferred to workstations, including some programs that perform specialized numerical analysis, which had been the forte of the supercomputer.[17] Of course, not all problems that could once be solved only on supercomputers are going to workstations. Many problems are simply beyond the power of a workstation to solve economically. But the workstation permits a much finer segmentation of the market for numerical analysis: the little problems go to workstations, the big problems go to supercomputers, and some analyses can be performed on either.

In fact, one might argue that, in the United States, computer tasks go to supercomputers only when the data needed for their solution exceed the ability of most workstations to retrieve and use the data rapidly.[18] These very often are tasks such as designing nuclear weapons, sifting through electronic intelligence, or other traditional supercomputer problems where cost is not of paramount concern. The market is more likely to favor tasks that could have been solved on a supercomputer had supercomputers been ubiquitous and cheap. Thus, in essence, the expanding need for numerical processing is more likely to be satisfied on a workstation than on a supercomputer. To some extent, the workstation may do to the supercomputer what the personal computer did to the mainframe--limit its demand.

Resolution of some of the software problems of the massively parallel supercomputers that HPCC is working on may inadvertently help workstations take over part of the supercomputer market. As noted above, many of these massively parallel supercomputers are composed of the same microprocessors that power the workstations; each microprocessor may have its own memory and some peripheral chips, just like workstations. In short, these massively parallel supercomputers are really many "workstations" within one box. Software developments that permit many of these workstations to work together within one box may also permit many freestanding workstations to do the same. Indeed, HPCC is funding research to speed such developments.

Software already exists that allows clusters of freestanding workstations on a network to use each other's computational resources to produce a computing machine that mimics

17. Jeffrey Young, "Crashware," *Forbes*, April 13, 1992, pp. 110-112.

18. Mark Furtney and George Taylor, "Of Workstations and Supercomputers," *IEEE Spectrum* (May 1993), pp. 64-68.

much of the power of a supercomputer.[19] With this software, problems are divided among different workstations, just as they are within a massively parallel supercomputer, while the different parts of the software communicate back and forth between different workstations. This type of software is currently in a relatively primitive state and works best with problems having parts that are fairly independent. Clustered workstations are also multipurpose: they can be used as stand-alone computers during the day and as parts of supercomputers when not occupied.

But using workstations in this manner has serious limitations. Communication between freestanding workstations is a couple of orders of magnitude slower than between the workstations inside a massively parallel supercomputer. Consequently, until network communications technology becomes much faster, there will be whole classes of problems for which such clusters are inappropriate.

Advances in network technology also provide the most likely route through which computing resources will be "scaled." As was noted in Chapter 2, HPCC is developing high-performance computer hardware that can be scaled linearly by simply adding more computing nodes. The basic architecture will not change--it will just grow faster. Advanced network technology, however, also allows users to create "virtual" scaling by simply adding more computer resources as they are needed.[20] Although virtual scaling is not likely to allow as fine increments in capabilities as purely scalable hardware, it may be more economic--allowing more investigators to make part-time use of a single resource--than are simple additions to hardware.

HPCC is devoting substantial resources to resolving the software problems involved in assigning tasks to groups of processors. It has joined workstation producers in funding much of the research on workstation clusters. Should generic solutions be found, the workstation market may benefit substantially. Thus, efforts by HPCC to develop network software and some classes of hardware may affect the commercialization of HPCC-supported hardware in other markets.

Vector Supercomputing Technology May Limit the Near-Term Growth of the Massively Parallel Supercomputer Market

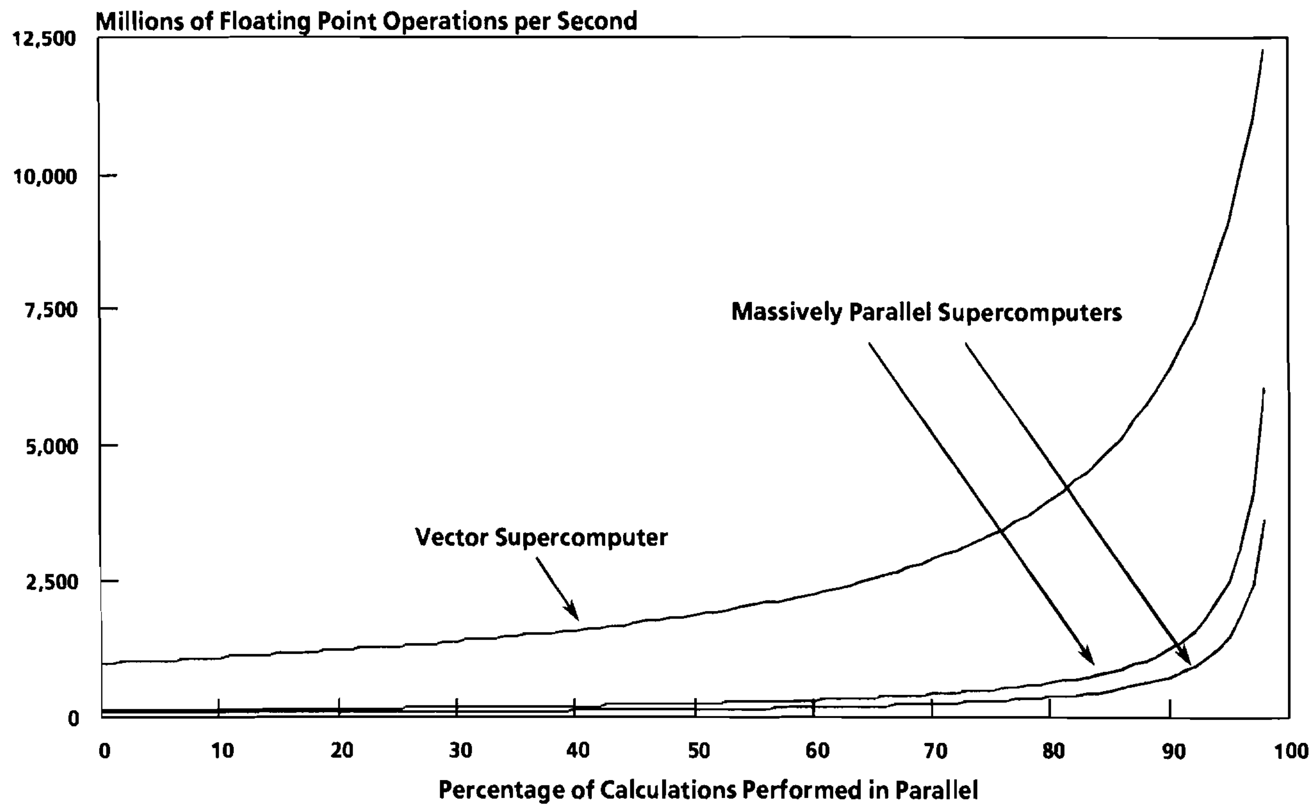
The quest to make massively parallel supercomputers the dominant supercomputer architecture faces a well-entrenched incumbent: the vector supercomputer. Displacing vector supercomputers from their current position is not likely to be easily or quickly accomplished. It is not yet certain that the higher theoretical peak speeds of the massively parallel supercomputers will translate into features that the marketplace will value. The difficulty in programming the massively parallel supercomputers as well as their lower performance on many kinds of problems may limit their share of the supercomputer market. Finally, many massively parallel supercomputers are incompatible not only among manufacturers, but in some cases also between generations.

Hardware for Vector Supercomputers. Because of differences in design, massively parallel supercomputers can provide very high speeds for certain types of problems, and vector supercomputers can provide higher speeds for most problems. The central point to remember is that a vector supercomputer is composed of a few--16 at most--extremely fast central processing units, while a massively parallel supercomputer is composed of hundreds or thousands of slower central processing

19. Alexander Wolfe, "Parallel Tools Kick Applications into High Gear," *Electronic Engineering Times* (March 30, 1992), pp. 30, 60, and 82.

20. For examples of adding computer resources to create a larger computing resource through networks, see the presentations of Larry Smarr of the National Center for Supercomputing Applications, University of Illinois, and Gene Feldman of the NASA/Goddard Space Flight Center at the Revolution in Supercomputing Symposium, January 27, 1993, Washington, D.C., sponsored by Silicon Graphics, Inc.

Figure 3.
Simulated Performance of Three Supercomputers



SOURCE: Congressional Budget Office based on Peter Gregory, "Will MPP Always Be Specialized?" *Supercomputing Review* (March 1992), pp. 28-31.

NOTE: The vector supercomputer simulation is based on Cray Research's C-90. The massively parallel supercomputer simulations are based on Thinking Machines Corporation's CM-5 and Intel's Paragon.

units. The high theoretical speed occurs only when a problem can be divided into very many components.[21] Then the massively parallel supercomputer can take advantage of its many processing units. But when a problem cannot be divided into many parallel components, the speed of the vector supercomputer's central processing unit becomes a winning advantage.

Most actual problems are not entirely massively parallel or entirely composed of a few components, however. Typically, some aspects will be parallel and some will be nonparallel. A supercomputer consultant recently analyzed the performance of different types of supercomputers with different combinations of parallel and nonparallel computation.[22] The analysis in Figure 3 compares the simulated performance of different supercomputers as the problem to be solved becomes more parallel.

The analysis shows that, for several of the most common models of supercomputer, the

21. For a discussion of the limits to parallelism, see Jack Worlton, "Existing Conditions," in *Supercomputers: Directions in Technology and Applications* (Washington, D.C.: National Academy Press, 1988) pp. 47-49.

22. This section is based on Peter Gregory, "Will MPP [Massively Parallel Processing] Always Be Specialized?" *Supercomputing Review* (March 1992), pp. 28-31. Appendix B of this report presents the formula used in the calculations. A similar point is made in Patterson and Hennessey, *Computer Architecture*, p. 576.

vector supercomputers are faster than massively parallel supercomputers unless as much as 99.2 percent--depending on the computer--of the computations are performed in parallel. That is, even if only 8 of every 1,000 computations performed in the course of solving a problem are not performed in parallel, then the vector supercomputer will be faster.

Another major issue in the speed of parallelism is the length of time required to communicate between nodes versus the length of time it takes to perform the work. Problems that require either relatively little communication between nodes or communication only to a few nodes are especially appropriate for parallel computation. But problems whose solution has many interactive parts are more difficult for parallel computers because the computers spend their time communicating between the nodes rather than computing within the nodes.[23]

The difficulty of internode communication further complicates the issue of parallelism: not only must the problems be massively parallel, but they must also have a certain type of parallelism, one in which the different components are fairly independent. One obstacle for parallel supercomputers is that whole categories of problems most commonly solved on supercomputers have low degrees of parallelism and require a high degree of communication between nodes.[24]

Thus, theoretical maximum speed may not be a helpful guide to actual operations. In benchmark tests conducted at NASA's Ames Research Laboratory, the actual performance of several brands of massively parallel supercomputers across a wide range of problems proved to be between 1 percent and 5 percent of theoretical peak efficiency.[25] In most problems, the massively parallel supercomputers were able to obtain only a fraction of the speed of the older vector supercomputer, in many cases not even matching the efforts of an older Cray's processor. Interestingly, the latest generation of massively parallel supercomputers has closed in on the performance of an older, single-processor CRAY Y-MP for several benchmarks, a feat that largely eluded the earlier generations.

The NASA analysis found, however, that because of the lower purchase price of massively parallel supercomputers, the best of them exceeded the best vector supercomputers on a cost-performance basis. In addition, some Department of Energy researchers concluded that at least part of the lower performance of the massively parallel supercomputers was attributable to their immature software tools, a weakness HPCC intends to correct.[26]

The common wisdom in the supercomputer industry is that a new computer system has to be 5 to 10 times as fast (or cost-effective) as the old one to justify the costs of adapting to the new technology. In the absence of such a cost differential, the massively parallel supercomputers will replace vector supercomputers only slowly, if at all. Massively parallel supercomputers exhibit this differential in only a small number of applications.

Software for Vector Supercomputers. The vast majority of problems solved today on supercomputers are solved on vector supercomputers, not massively parallel supercomputers. For the past 20 years, scientists and engineers have written and optimized their programs for these vector computers, so that there is now a large stable of mature software

23. Robert Pool, "Massively Parallel Machines Usher in Next Level of Computing Power," *Science*, April 3, 1992, p. 50.

24. Furtney and Taylor, "Of Workstations and Supercomputers," p. 68.

25. See David Bailey, "Experience with Parallel Computers at NASA Ames," Technical Report RNR-91-007 (NASA Ames Research Center, November 14, 1991); David Bailey and others, "NAS Parallel Benchmark Results," Technical Report RNR-92-002 (December 7, 1992). For similar results, see Olaf Lubeck, Margaret Simmons, and Harvey Wasserman, "The Performance Realities of Massively Parallel Processors: A Case Study" (Los Alamos National Laboratory, Los Alamos, New Mexico, June 19, 1992).

26. Lubeck, Simmons, and Wasserman, "The Performance Realities of Massively Parallel Processors."

able to solve a wide variety of problems.[27] Most of this software is continuously being upgraded and refined to improve its performance and make it easier to use.

A recent survey at the Pittsburgh Supercomputing Center indicated that their library had more than 300 third-party software programs for their vector supercomputer but only two such software packages for their massively parallel supercomputer.[28] As a result, their vector supercomputer was oversubscribed, and their massively parallel supercomputer was underused.

Existing software would have to be rewritten in order to run on massively parallel supercomputers. Some of it can be easily rewritten, but much cannot. In fact, some of the most popular software for supercomputers is likely to be difficult to translate. The technical software written for supercomputers can involve thousands of lines of code. Moreover, those programs must not only be rewritten but also optimized for the new computer architecture. (This rewriting assumes that HPCC's software effort is successful in producing software tools and techniques for massively parallel supercomputers. If robust software tools are not developed, such rewriting may not take place.[29])

Even if the software code is rewritten, there is no guarantee that it will improve performance. For some kinds of problems the massively parallel supercomputers are fast, but for others they are not as fast as vector supercomputers. Even if they are faster than conventional supercomputers, they may be only modestly faster--by a factor of two or some small number--which would appear to limit the incentives to rewrite software for them. If rewriting several hundred thousand lines of code only halves computation time, such an investment would probably not be worthwhile given the relative costs of computer time and programmer time. The current code, although not perfect, may be "good enough."

This obstacle should not be underestimated. In most other computer markets, a backlog of existing software usually gives incumbent computer designs a substantial, if not insuperable, advantage. For example, most users would not switch from an IBM-compatible personal computer even though the leading alternative is commonly described as "friendlier." In technical markets, an investment in current software may weigh less heavily because so many users write their own software, although it is unlikely that this completely offsets the advantage of having a library of existing applications.

Economic Inertia and Technology Development. Technological conservatism is seen everywhere in the computer world. A familiar example is the QWERTY keyboard, the conventional alphanumeric keyboard on virtually every typewriter and computer terminal.[30] More productive keyboard layouts have been available since the 1930s, but have never caught on, even though they could easily pay for themselves and the retraining costs in enhanced typing productivity. Typists have little incentive to learn a new layout until it becomes prevalent, while any business that introduced a new system would have to retrain its entire staff.

Tendencies toward technological conservatism may be gradually overcome, but only if there is a clear reward for doing so. The original CRAY 1 overcame such obstacles by pro-

---

27. For a list of more than 600 software applications for vector supercomputers, see "Directory of Applications Software for Cray Research Supercomputers" (Cray Research, Mendata Heights, Minn., 1990).

28. Frank Wimberly, "Connection Machine Software Packages Becoming Available," *Pittsburgh Supercomputing Center News* (May/June 1992), p. 7. Third-party software is software that has been written by neither the manufacturer nor the user of the supercomputer. Third-party software is typically intended for sale.

29. The General Accounting Office recently argued that ARPA was neglecting the development of software tools. See General Accounting Office, *Advanced Research Projects Agency Should Do More to Foster Program Goals* (May 17, 1993).

30. Paul David, "Clio and the Economics of QWERTY," *American Economic Review* (May 1986), pp. 332-337.

viding fast performance when it was introduced in 1976. The massively parallel processors have not yet been able to match current supercomputers for most applications, so that there is no clear reward for investing in them. For most applications, existing supercomputers are adequate.

In some commercial applications, where there may be clear rewards for rewriting software code, massively parallel supercomputer technology may penetrate first. The simulation of oil reservoirs, for example, is more accurate on parallel machines, and the problems involved in the simulations lend themselves to such technology. Moreover, from an economic standpoint, coming up with an accurate estimate of the size of an oil field carries a substantial payoff. The contracts for oil production can be in billions and tens of billions of dollars. Even a small improvement in the estimate can easily pay for the rewriting effort.[31] Another example involves financial calculations for certain investment instruments for which the rewards can be large and immediate. Whether these isolated instances of commercial use can be turned into larger markets is as yet unclear.

Other applications, such as rational drug design and computational chemistry, may provide openings for massively parallel supercomputers because they are so new that vector supercomputers have not yet been applied to them to a great extent. Consequently, investigators can start from scratch without worrying about leaving behind a lifetime's work.

Potential Entry into Business Computer Markets

One way to move massively parallel technology into general use despite the preemption of much of its potential market by the workstation and the vector supercomputer may be to shift from the technical market to the business market. Recently, a leading manufacturer of massively parallel supercomputers--Thinking Machines Corporation--announced its interest in providing "back end" power for mainframe computers.[32] In such a system, a massively parallel supercomputer would supplement the power of a conventional mainframe when accessing large data bases. Other, smaller producers of these supercomputers have positioned their product similarly. As an adjunct to a mainframe, massively parallel supercomputers would not conflict directly with conventional supercomputers. But like the vector facilities discussed above, they might be able to create a niche by exploiting weaknesses in certain computer markets.

So far, the commercial sales of massively parallel supercomputers have been largely for a special or single purpose. In a widely publicized sale, American Express bought two CM-5 massively parallel supercomputers from Thinking Machines Corporation to reduce the time it took to query its massive data base. Even in this case, however, the CM-5s are not acting independently but in conjunction with a mainframe computer.[33]

One special-purpose massively parallel computer made by the Teradata unit within NCR has made substantial inroads into business markets over the last decade.[34] (Because Teradata computers lack specialized number-crunching capacity, they are not properly considered supercomputers.) This line of machines combines between several dozen and several hundred of the microprocessors that power IBM-compatible personal computers (the so-called 80286, 80386, and 80486) into a facility that runs a specialized data-base application, again as an addition to a general-purpose mainframe computer. Teradata's

31. Ironically, simulation of oil reservoirs is often mentioned as a problem that is especially well suited for networks of engineering workstations, as discussed above. Thus, two types of technology will be competing for the same set of problems.

32. Johanna Ambrosio, "Super CPU Maker Widens Aim," *Computerworld* (October 19, 1992), p. 6.

33. Dwight Davis, "Supercomputers Knock at IS Doors," *Datamation* (December 1, 1992), pp. 79-82.

34. Dwight Davis, "Oracle's Parallel Punch for OLTP [On Line Transaction Processing]," *Datamation* (August 1, 1992), p. 67.

sales in this specialized market almost equal the combined sales of all the massively parallel supercomputer producers. IBM recently announced its intention of entering this market of specialized parallel data-base computers.[35] A major software company, Oracle, has also devoted substantial efforts to rewriting its software for several brands of massively parallel supercomputer to serve this market. For its part, Teradata has been trying to expand its market beyond its specialized one.

It is not yet clear whether the data-base market has a path that will lead to general-purpose commercial computing. Such a path might emerge as more and more functions are incorporated into the specialized computers until eventually they become general-purpose computers. This path might be more likely to penetrate the business markets than one directly through the supercomputer market because most U.S. firms do not use supercomputers in their business. A survey of Fortune 500 firms found that only 15 percent of them use supercomputers, and an additional 6 percent consider them an option.[36] By contrast, data-base management accounts for 50 percent of corporate mainframe use, and almost all Fortune 500 firms have mainframes.

Thus, to the extent that parallel business computers use HPCC technology, it might enter mainstream commercial uses through alternative avenues. Like the massively parallel supercomputers, however, the software problems associated with turning these special-purpose computers into general-purpose computers are substantial. Also, because the levels of technical competence are lower in the business market than in the scientific and technical markets, the effect of these software problems on demand is likely to be more severe. And many of the same questions regarding parallelism discussed above in connection with massively parallel supercomputers would also apply to these parallel mainframes, although data-base management is inherently very parallel.

35. Laurence Hooper, "IBM to Push New Generation of Mainframes," *Wall Street Journal*, December 22, 1992, p. A3.

36. Mohammad Amini, "The Status of Supercomputing in Corporate America," *Supercomputing Review* (September 1991), pp. 40-41.

Obstacles in Perspective

HPCC is not unique in facing obstacles in the marketplace. Most new products and many new technologies fail to gain widespread acceptance when first introduced. Even when they finally succeed, it may be for reasons the inventors could not have foreseen. For example, the creation of small personal computers stimulated the demand for large flat-panel displays. These had been invented in the 1960s to service a home TV market that did not materialize. Their inventors could not have foreseen laptop computers.

The designers of massively parallel supercomputers hope to increase the demand for supercomputers by making their hardware less expensive. But the software for these machines, when it exists, is difficult to write and difficult to use. Viewed from the perspective of the costs of a supercomputer over its lifetime in actual use, the massively parallel supercomputer reduces the cost of the least expensive element (initial purchase price) while increasing the cost of the most expensive elements (software and support).

Firms introducing products based on HPCC supercomputer technology may be able to overcome these obstacles, although the resulting market is not likely to experience the automatic growth of a new market. It is not merely a case of waiting until sufficient applications are written for massively parallel supercomputers and all the kinks are ironed out. The target it seeks to hit is moving: both workstation technology and vector supercomputer technology are improving.

Most important, as noted above, most analysts agree that computer technology is likely be parallel in the future. They disagree about how best to get there and how parallel is parallel. HPCC has positioned itself toward one end of the spectrum.

The presence of these obstacles does not mean that there will be no market in the near term for such massively parallel supercomputers: the analysis found two potential markets. Rather, the obstacles will serve to limit demand within those potential markets. Overcoming some of these obstacles is a goal of the HPCC program. Other obstacles may be overcome by the firms that actually sell products incorporating HPCC technology. How these firms position themselves in the market will be crucially important, but is beyond the scope of this report.

Chapter Four

HPCC and Computer Network Markets

In addition to its effect on the research and education communities, the High-Performance Computing and Communications program may also have a significant impact on the emerging data communications markets in the United States. The commercial markets currently consume most of the billions of dollars worth of network equipment and data communications services the U.S. industry produces every year. Consequently, they are likely to be among the early consumers of the network technology developed under HPCC.

HPCC and Computer Networks

The National Research and Educational Network, discussed in Chapter 2, is a central part of the HPCC, both as a goal in and of itself and as an enabling technology for other components of the program. The aim is to make the network available to more researchers and also to enable it to transmit data at the rate of 1 billion bits (1 gigabit) per second (it currently operates at about 45 million bits per second). By linking supercomputers and other research assets of the government over very rapid computer networks, federal agencies should be able to bring the best tools to bear on the grand challenges. In addition, the NREN could demonstrate how other segments of society might use and benefit from such a network.

As noted in Chapter 2, the NREN component of HPCC has two major subcomponents: upgrading the interagency interim NREN, and promoting the development of computer networks that transmit data at the rate of 1 billion bits per second. Upgrading interim NREN has several parts. The main focus of the Internet program has been on bringing the existing NSFNET up to 45 million bits per second speed. NSFNET is a computer network supported by the National Science Foundation that serves as the central link of Internet. Recently, the NSF published a solicitation to bid on upgrading the NSFNET to 155 million bits per second speed, with an eye toward eliminating the federal subsidy on part of it. Less well known are efforts to bring some specialized agency data bases on line to make them accessible from the NSFNET.

In another part of the interim NREN, the Department of Energy is accelerating development of the Energy Sciences Network--the computer network for the department's R&D facilities--beyond the capabilities of the rest of the interim NREN. DOE had selected a contractor for the development of a 622 million bit-per-second network for service among a dozen or so nodes on the Energy Sciences Network, using protocols that have only recently entered commercial trials. However, the General Accounting Office set this selection aside and the contract will have to be rebid.

The research and development component of the NREN largely focuses on so-called gigabit test beds.[1] These test beds are a public/private venture in which industry teams and government researchers develop and test components, protocols, and other aspects of networking technology. As technology is proved, it will be incorporated into the construction of the interim NREN and into commercial offerings. In the test beds, the federal government has leveraged its investment several times over with private participants: while the federal cost is projected to be roughly $15 million, the private investment is reported to be many times that.

Meeting Goals

The network component of HPCC is more diffuse than the supercomputer R&D component because part of the effort is involved in R&D and part in developing an existing network. Each needs to be judged by different criteria. The R&D component obviously needs to be judged primarily by technical standards.

Technical criteria, however, are not sufficient for evaluating a network; marketing and demonstration criteria also need to be taken into account--for example, growth in the use of Internet. The combination of R&D programs with an operating network is intended to make computer networks faster and to demonstrate to more people that computer networks can contribute to their work, education, and other aspects of their lives.

Network R&D and especially Internet are intended to encourage the rapid spread of computer networks throughout the economy. Like telephones and FAX machines, computer networks become more valuable as more people hook up to them. HPCC planners hope that as the capacity becomes more widely available, more sectors of the economy will become linked by computer and will invent and discover new ways of using the network.

1. Gary Stix, "Gigabit Connection," *Scientific American* (October 1990), pp. 118 and 120.

Most computer power in the United States exists in unconnected islands. Even where computers are connected, the links are either isolated, as with organizational networks, or temporary, as with modem-based dial-up services. Computer networks have already demonstrated that they can provide a competitive advantage to companies having them. Many of the most successful retailing and service companies use computer networks as a source of competitive advantage.[2]

Furthering U.S. Competitiveness

Advocates of HPCC believe that a proliferation of network services could enhance economic performance and welfare. The competitiveness of U.S. industry would be furthered in several ways. First, U.S. service and retail companies, which already use computer networks far more than other sectors, are likely to be the first to take advantage of any new network technology. Of course, if networks became more widely used internationally, U.S. firms might lose some of the advantages they currently enjoy relative to their foreign competition.

Second, U.S. firms manufacture most of the large computer networks.[3] Consequently, if federal R&D and demonstration projects increase the demand for these networks, U.S. firms are likely to capture a large portion of that demand. Even in the market for small networks, U.S. firms have a disproportionate influence, largely because they dominate the markets for personal computer hardware and software. An offsetting factor is that computer networks compete with mainframe computers, a market in which U.S. firms are also quite dominant. If sales of U.S.-produced networks

2. Thomas Malone and John Rockart, "Computers, Networks and the Corporation," *Scientific American* (September 1991), pp. 128-136.

3. National Research Council, Computer Science and Telecommunications Board, *Keeping the U.S. Computer Industry Competitive: Systems Integration* (Washington, D.C.: National Academy Press, 1992), p. 7.

increased, some of this increase would come at the cost of reduced sales by other U.S. corporations, such as mainframe producers.

HPCC proponents argue that computer networks can help people in many different walks of life share scarce resources, reducing their cost and improving their quality. Examples of such specialized resources include services, such as medical expertise; or information, such as library catalogs; or people, such as teachers.

Box 1.
Segments of the Market for Networking Products

The market for computer network products and services can be divided into four overlapping segments. While hardware and software are generally complementary, network services are generally in distant competition with hardware and software.

Hardware

File Servers
Cabling
Network Interface Cards
Bridges and Routers
Leased Lines

Software

Operating Systems
Applications

Services

Value Added Networks
Outsourcing
Switched Networks

Internetworking

Internet
Commercial Providers

SOURCE: Congressional Budget Office.

Unlike the supercomputer component, the vision for the network component of HPCC is in line with economic trends. There is a tradeoff, however, between performance and sales--very fast networks are not likely to be cheap, and expensive networks are not likely to be widely used. All users would like faster, more responsive networks, although most of them do not even begin to use the maximum technical capabilities currently available. Advocates argue that the very presence of such capabilities will lead to new applications, but critics maintain that new applications may not be economically viable or may not find wide use outside a few specific sectors.

Regulation--particularly if the regulation is dispersed among numerous state public utility authorities with many diverging objectives--may retard the adoption of advanced network technology. Although the general trend of regulatory intervention seems favorable in many ways, the regulated carriers will need time to adjust their behavior and to put the infrastructure in place.

Analysis of the Markets for Computer Network Hardware, Software, and Services

The United States has a very large and thriving data communications market. Some analysts believe that data transmission amounts to a substantial part of all U.S. telecommunications. Many large companies are in the business of providing the hardware, connecting lines, software, and services for computer networks.

Market Segments

The market for computer networks comprises four partially overlapping segments (see Box 1). The first two segments consist of providers of the hardware and software for computer networks, including telephone companies that lease lines over which private firms run their computer networks as well as firms that manufacture network hardware or write network software. According to the Department of Commerce, worldwide sales of computer network hardware and software exceeded $8 billion in 1991.[4] Dataquest, a market research firm, estimates that spending for digital private lines increased from less than $1 billion in 1986 to more than $3 billion by 1990--an annual increase of roughly 25 percent.[5]

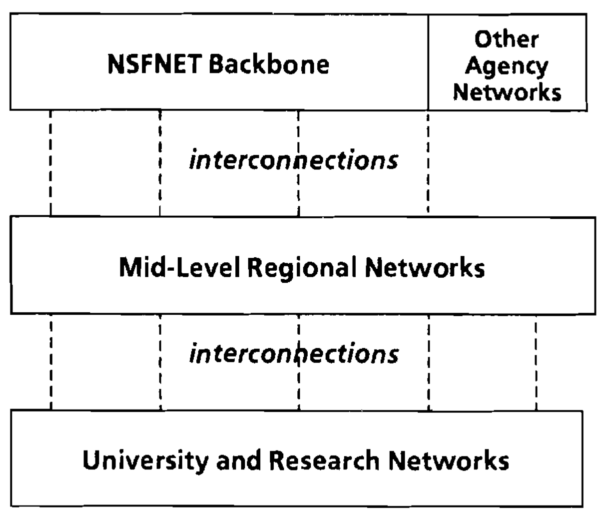
Firms in the third segment provide network services to companies, organizations, and individuals who do not wish to set up their own networks. This set of firms is in distant competition with the first two, just as restaurants are in distant competition with grocery stores. The Department of Commerce estimates that the U.S. market for the most common single type of commercial network service--value added networks--reached $6 billion in 1991.[6] The market for another popular type of service, called outsourcing (in which one firm contracts with another to manage its computer network), is estimated to be from $3.5 billion to $7 billion annually. By contrast, the market for simple switched network services, comparable with switched voice telephone service, is still in its infancy.

The fourth segment has one principal member--Internet, a federally funded group of networks associated mainly with universities and research organizations. Logically it is a subset of the third (network) segment, but it is now too large to be treated on a par with commercial providers of network services. Unlike those providers, Internet is not a single network but a network of networks: it serves network providers directly and final users indirectly through the networks to which they belong.

Internet and the Development of the Network Services Market

Internet is a network of computer networks serving the education and research communities, all of which use the same protocol (TCP/IP, for Transmission Control Protocol/Internet Protocol) to communicate. This network is an outgrowth of the ARPANET. Internet has three tiers: national NSFNET links regional midlevel networks, which in turn link local research and university networks. Other federal agencies' networks are also connected to many of the research and educational communities, but such connections are specific to different agency missions. The contractor that runs the facilities for the National Science Foundation is a joint venture of IBM, MCI, and Merit, the Michigan-based midlevel network. Although the NSFNET provides the upper-level connections, the entire agglomeration of connections and networks is commonly called Internet (see Figure 4).

Figure 4.
Composition of Internet



SOURCE: Congressional Budget Office.

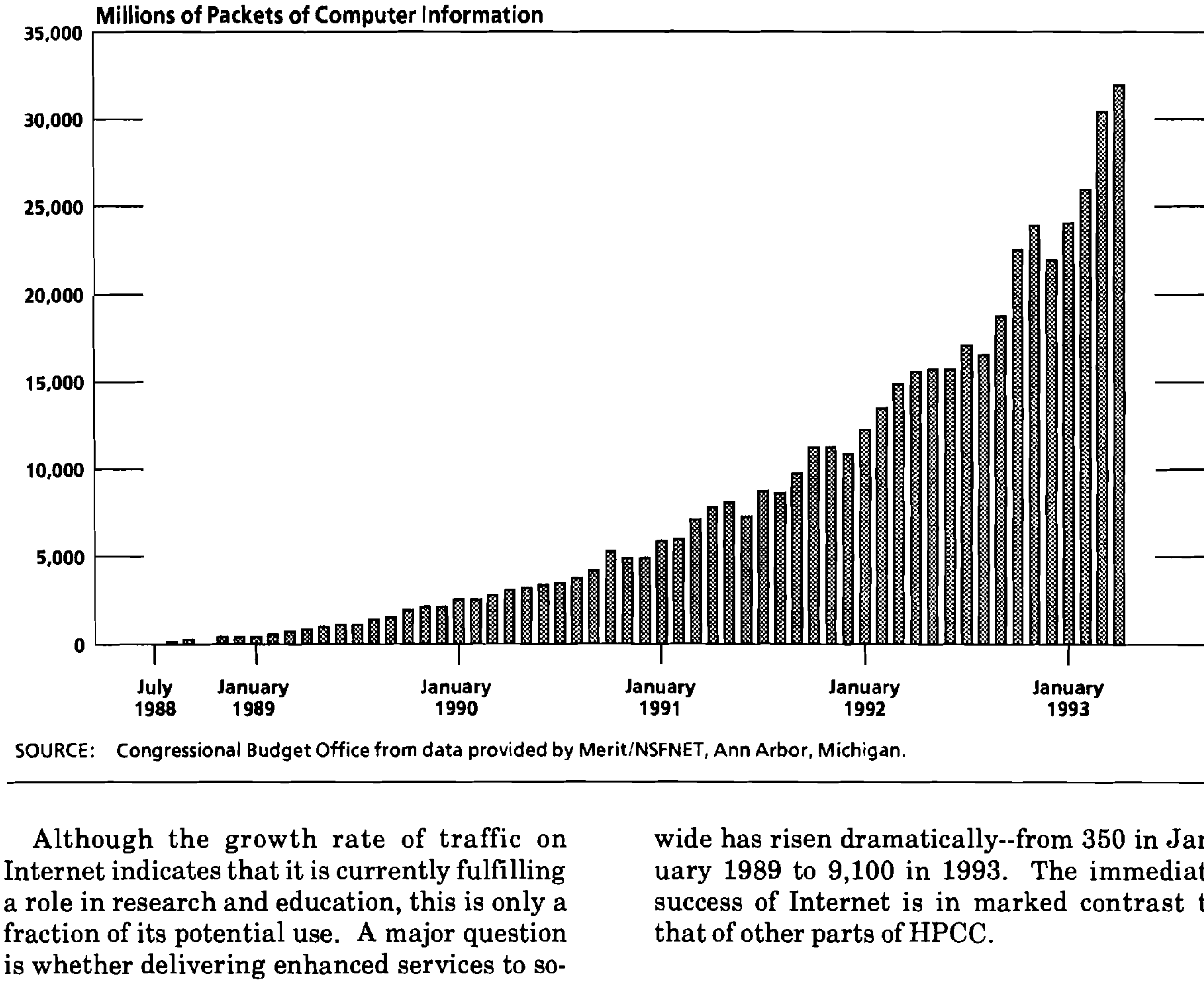
NOTE: NSFNET = National Science Foundation computer network.

4. Department of Commerce, *1992 U.S. Industrial Outlook* (1992), p. 27-18. Spending on modems used by computers to connect to networks is not included in the network market data discussed here. The market for modems exceeded $900 million in 1991.

5. "Dataquest Market Statistics: Long Distance Telephone Services" (Dataquest, Inc., San Jose, Calif., May 1991), p. 2. Much of the private-line usage is for private voice networks, but this is decreasing relative to the whole.

6. Department of Commerce, *1992 U.S. Industrial Outlook*, p. 28-8. The Congressional Budget Office found no estimate of the total market for network services.

Figure 5.
History of Internet Traffic, 1988-1993



SOURCE: Congressional Budget Office from data provided by Merit/NSFNET, Ann Arbor, Michigan.

Although the growth rate of traffic on Internet indicates that it is currently fulfilling a role in research and education, this is only a fraction of its potential use. A major question is whether delivering enhanced services to sophisticated users is more important than broadening usage to the less sophisticated.

Background. Internet has achieved a very high rate of growth during its five years under the current contractor.[7] When it began operating in July 1988, the NSFNET backbone traffic was well under 100 million packets of computer information delivered electronically. In August 1992, the NSFNET backbone delivered 16.5 billion packets of computer information (see Figure 5). Similarly, the number of networks connected to NSFNET worldwide has risen dramatically--from 350 in January 1989 to 9,100 in 1993. The immediate success of Internet is in marked contrast to that of other parts of HPCC.

The uses of the NSFNET backbone are similar to those of other computer networks: electronic mail and file exchanges account for almost three-quarters of the data transmitted. These have been a decreasing share of the total, however, although no single other use is displacing them.

Despite its popularity, Internet presents obstacles to some users. Individuals not affiliated with a research firm or university may have difficulty obtaining access to the network. Security is also a concern to many potential members that are research firms. Potential users are put off by the unavailability of addresses of potential correspondents: pro-

7. Information on growth and usage of NSFNET come from MERIT Network, Ann Arbor, Michigan.

grams for electronic white pages are still being developed, as is a directory of services. Within universities, many faculty and research staff members are unaware of having access to Internet or, if aware, may be intimidated by it (for example, users must learn yet another layer of software). Training on how to use Internet is limited.

Nevertheless, a remarkable number of users log on regularly. Internet offers access to electronic message boards and journals. Professional academic societies have mailing lists that permit members to exchange information rapidly, often serving as a grapevine in "hot" research areas.

Impact on Commercial Markets. Internet has already had a substantial effect on the commercial networking market. It specializes in internetworking, that is, connecting different networks. When the first contract for the NSFNET was signed in 1987, the internetworking market barely existed. Now, however, the number of networks has multiplied, and companies and groups investing in them want interconnections, rapidly creating a substantial market for appropriate hardware and services. In fact, the market for internetworking hardware exceeded $700 million in 1991. As one of the first providers of internetworking services to a large community of users, Internet was in a position to help define standards and protocols and develop the generic technology that would see wider use.

Internet already provides network connections to a substantial fraction of the users and developers of technology in the United States. In addition to universities and research institutions, many technology companies also have Internet connections directly or indirectly through a commercial service provider. Because networks become more valuable as more people connect to them, this particular network is becoming the core of a nationwide data network. The major public computer networks--such as Compuserve and MCI Mail--already have Internet gateways, and the pressure to provide connections to Internet clients only grows as more individuals, corporations, and organizations join the network. Although many questions remain about how best to transfer the responsibility for Internet's non-R&D operations to private hands, it has had a major influence on the development of nationwide computer network markets.

Commercial Internetworking. Internet was one of the first major providers of switched internetworking services--that is, providing network capability on demand parallel to telephone communication. But the commercial development of switched internetworking services has been slow for several reasons. Because the market for switched voice communications is much larger than that for switched data communications, most of the potential providers--the regional Bell operating companies and the long-distance companies--have been slow to enter it. Standards have also been slow to develop. Finally, the number of computer networks needing to be interconnected has not been very large until recently.[8] In these circumstances, Internet provided services for a stable pool of customers.

The most rapidly growing type of switched networking service, called frame relay, still has quite a limited market, and most others are still in trials or even awaiting them.[9] The NSFNET contractor is in fact involved in some of these trials, separate from its contract with NSFNET but using many of the same facilities. A very important contribution of Internet, particularly the NSFNET, may have been in allowing the contractor--especially its joint venture members, IBM and MCI--to develop expertise in providing switched internetworking services ahead of other firms that did not have a built-in clientele.

8. Elisabeth Horwitt, "Answering Users' Interconnectivity Needs," *Computerworld* (March 30, 1992), pp. 69 and 72.

9. One recent survey found that as of September 1992, there were only 189 frame relay customers in the United States. Nathan Muller, "What Does Frame Relay Really Cost?" *Datamation* (December 1, 1992), p. 58.

Obstacles to Wider Use of HPCC Technology

What is the potential demand for switched services at speeds above 45 million bits per second? Current commercial technology can serve at speeds below that. No one expects the market to jump from 45 million to 1 billion bits per second. Some designs are already running at 155 million and 622 million bits per second. This additional capacity may be shared by a number of users, none of whom might require all of it.

Unlike the massively parallel supercomputer technology discussed in the previous chapter, the technology being developed and tested in HPCC gigabit test beds has clear potential for influencing mainstream data communications technology. The participating companies are reportedly testing switches that may then become the basis for commercial offerings, although probably operating at sub-gigabit speeds.

Similarly, Internet has helped the demand for internetworking services to jell. As this market develops further, Internet technology and modes of operation, not to mention Internet itself, may become the heart of further developments.

Most analysts believe that a large market for rapid data communications will ultimately develop. The disagreements are about the path and the pace. Is the large market more likely to be achieved by creating many network applications that each need only a small network capacity and by allowing the evolution of their needs to create a large demand for rapid services? Or is it more likely to result from creating requirements, and equipment to serve the most demanding requirements, which are relatively few in number?

Telephone Companies and the Price of Leased Telephone Lines

The data communications market as a whole is growing quite rapidly, probably as a result of the rapid proliferation of computers and the substantial decline in the cost of network equipment. Because of these changed circumstances, new network services may achieve the commercial success that eluded network services in the 1970s and 1980s. It is not only a question of new technical capabilities, but also of price.

An analogy may be seen in the way falling computer prices stimulated the demand for computers. In 1970, when the IBM 370 was being sold for several million dollars, the demand for it was large for the time, but minuscule in comparison with today's sales figures. Currently, IBM-compatible personal computers, which have roughly the same computational capabilities as the IBM 370, sell by the millions for under $2,500. There have been some increases in computational power to be sure, but they are far outweighed, especially at the lower end of the price spectrum, by decreases in price.

If the HPCC program leads to similar decreases in the cost of data communications, more and more commercial uses will be found for them. Until quite recently, communications costs had not matched computation costs in their decline.[10] One recent analysis suggests that, through 1980, communications costs had declined between 3 percent and 5 percent a year, but that computing costs were falling more than 20 percent per year during this same period, even though major pieces of computing equipment and communications switching equipment share many components.[11]

10. This analysis is from Kenneth Flamm, "Innovation, Regulation and Competition," in Robert Crandall and Kenneth Flamm, eds., *Changing the Rules: Technological Change, International Competition, and Regulation in Communications* (Washington, D.C.: Brookings Institution, 1989), pp. 13-61.

11. Ibid., p. 29.

The trajectory of costs in the future, however, will crucially depend on the market structure. The above-cited analysis argued that it was not lack of investment or R&D that kept communications prices from falling as rapidly as those of computers, but rather regulation and lack of competition. Providers of network hardware, however, are in a highly competitive market. With the breakup of the Bell system, there is now competition in the long-distance market, and so some prices of leased digital long-distance lines, as discussed below, have fallen.

Local Leased Telephone Lines. The local telephone companies still have near-monopoly control over local access, which the nationwide providers of leased lines and network services need in order to reach their customers. The prices local telephone companies charge these providers may not drop sufficiently to encourage a large market.

Nonetheless, some providers of alternative services have bypassed local service providers and may be in a position to force competitive reactions from the local telephone companies in some areas. How much discretion the local companies will be allowed will depend crucially on the actions of regulatory agencies. The regulatory climate seems favorable for encouraging competition overall, but a fully competitive market may be slow to emerge and may take years to develop nationwide. The slow development of alternative providers may allow local providers a substantial degree of protection from the hazards of competition and enable them to hold onto a sizable chunk of the entire commerce in this area.

As noted above, private leased lines are a major element in the data communications market, especially at faster speeds. Local telephone companies have not lowered their prices for these lines significantly. According to filings with the Federal Communications Commission (FCC), the Bell operating companies' prices for data services averaged a 2 percent annual decline between July 1990 and July 1992 (the most recent data available).[12] Although some local exchange carriers had larger declines, none had annual declines larger than 5 percent.

Local carriers still account for the vast majority of private lines. According to one estimate, they provided 80 percent of all private lines in 1989.[13] Two-thirds of all data lines were local, at least 90 percent of which were provided by local exchange carriers. The local metropolitan fiber systems, satellite systems, and other systems still accounted for only a small fraction of leased data lines. According to FCC data, these alternative local carriers together served fewer than 6,000 customer locations nationwide at the end of 1991.[14]

Long-Distance Leased Lines. Interexchange (or long-distance) carriers as a group accounted for roughly 20 percent of leased data circuits in 1989, according to the previously cited survey of private lines. Their percentage has probably risen since then. Though more competitive than the local exchange carriers, American Telephone and Telegraph still accounts for the bulk of this business.

The interexchange carriers, in contrast with local exchange carriers, seem to have brought some of their prices on leased lines down substantially. Most important, in posted prices filed before the FCC, there is no dominant single price trajectory. Posted prices for short-range private lines, where the interexchange carriers compete with the local exchange carriers, have not fallen and in some instances may have risen since 1990. In the highly competitive longer-distance (above 200

12. The data are taken from the individual company Tariff Review Plans using the Special Access Basket lines detailing High Capacity and DDS (Dedicated Digital Service) indices.

13. Allan Turnbill, "The Future of Private Lines" (Probe Research, Cedar Knolls, N.J., 1989). See Chapter 13, especially pp. 396 and 402. Private lines can carry either data or voice traffic. Surveys and other techniques can be used to provide estimates of the use.

14. Jonathan Kraushaar, "Fiber Deployment Update, End of Year 1991" (Industry Analysis Division, Common Carrier Bureau, Federal Communications Commission, March 13, 1992), Table 15.

miles) markets, prices have been falling, although not by as much as computer prices.

Private-line prices have a fixed monthly component and a mileage-based monthly component. While the mileage price has been falling dramatically--by more than half in the last two years--the fixed component has undergone a dramatic rise--20 to 50 percent since 1990. For short distances, the rise in the fixed component dominates the decrease in the mileage-based component. Thus, posted prices on private lines shorter than 200 miles have risen, while posted prices on longer private lines have fallen, with coast-to-coast posted prices falling most. By way of illustration, the posted price for a 1,500,000 bit-per-second leased line roughly the distance from Chicago to New York (700 miles) fell by an average of 13 percent per year between 1990 and 1992, while prices for leased lines 75 miles long have risen by an average of 7 percent per year.

The posted prices are subject to volume discounts, commitment discounts, and other discounts depending on private negotiation. Such discounts serve the larger customers the best. The consumer who is least likely to benefit from such arrangements is a regional business or organization that wishes to tie together only a few sites: neither the local-exchange carriers nor the interexchange carriers have reduced posted prices in that market substantially.

Data communications prices may come down if consumers buy switched network services rather than lease private data lines. Under this arrangement, customers would pay only for the actual use, just as they do currently for long-distance voice conversations.[15] The local-exchange carriers have jointly committed themselves to providing these services. This service is completely new for the local-exchange carriers, however, and it is not clear how rapidly these services will displace their leased-line services or force those prices down in competition.

The near-monopoly of local-exchange carriers on local circuits, both for purely local transmissions and for access to interexchange carrier lines, is forecast to decline. Whether the decline will be fast enough or large enough to cause prices to decline is not clear. If competition is not sufficient to lower prices substantially, then the market for data communications may not develop as quickly as the computer market.

Though central to the successful commercialization of HPCC technology, the issue of regulation of the telecommunications markets is beyond the scope of this analysis, other than to note the central role competition could play in bringing the technology to market.

Likely Effects of Price Changes. The creation of a substantial market usually involves shifts in both demand and supply. HPCC is attempting to influence both. By subsidizing early users through Internet, HPCC helps increase the potential uses of network technology and thus shifts the demand curve so that more people seek to use networks. The independent proliferation of computers similarly shifts the demand curve by increasing the number of potential users.

For many components of networks, primarily hardware, the supply curve also has been shifting as the general progress in electronics has lowered prices. Prices of network-related electronic equipment have declined by a weighted average of 16 percent a year since 1990, according to industry data.[16]

Although the total cost of the system and the network services has fallen, the price of one major component--the leased local data

15. Muller, "What Does Frame Relay Really Cost?" pp. 55-62.

16. Figure calculated using U.S. prices and world market shares of the different types of network equipment. Price data from Dataquest, "Local Area Networks Market Share and Forecast" (May 1992); and "Networking Market Trends" (August 1992). Market share data from Department of Commerce, *1993 U.S. Industrial Outlook* (1993), pp. 24-26 and 25-26.

line--has not budged, although the number of such lines has increased. Such lines are crucial if the widespread internetworking conceived by HPCC is to become a commercial reality. But as the other components decrease in cost, local data lines will come to represent an ever larger share of the total cost.

Under some circumstances, HPCC technology could become widely used without major shifts in relative prices of leased lines. Computer networks may be subject to what economic historian Nathan Rosenberg calls "learning by using": as people learn to use networks, they rely more and more on them, and the networks increase in value over time. Under these circumstances, demand may expand without price decreases, provided the alternative local telephone systems also expand to match any capacity constraint of the Bell operating companies.

Demand for Very Fast Services

When the capacity of a network to carry data files is initially increased, the user sees immediate benefits.[17] Large files can be transferred much more rapidly, for instance. But as the capacity of the network rises further, the speed of file transfers is limited not by the capacity of the network but by other factors, such as the speed of light or distance.[18] Consequently, there is a point of diminishing returns to increasing network capacity. For modest-size files, the point of diminishing returns comes at a network capacity well below the gigabit-per-second level.

To analyze the relationship between capacity and speed, the Congressional Budget Office simulated the behavior of three networks of different capacity transmitting files roughly the size of the average FAX halfway across the country.[19] FAXes were chosen because they are currently the most common form of data communications.

At roughly 20 percent capacity use, a network of 1.5 million bits per second--roughly the capacity of the older Internet trunks--would transmit a FAX-sized file in a little less than a second. This rate is much faster than the current analog transmission speed of 9,600 baud per second.[20] If the capacity of the network was improved to 45 million bits per second--the capacity of the new trunks on Internet--transmission would take one-twentieth of a second. If a 1.2 billion bit-per-second network was transmitting the files, transmission would take one-eightieth of a second. In this last case, a 26-fold increase in capacity brings only a fourfold decrease in the system's response time. (Appendix C contains the formula used in calculating these results.)

Well before a network reaches the point at which there is no physical advantage to its increase in capacity, it is likely to reach the point at which there is no economic advantage either. For many users, the difference between receiving files in one-eightieth of a second rather than one-twentieth of a second may not be worth the extra cost.

If communications costs and transmission times decline, however, file size will probably grow in response as people are able to include more information in their messages. If the average file size grows by a factor of 10, the benefits of the higher speed become more apparent: at 20 percent capacity, the 1.2 billion bit network is almost 15 times faster than the 45 million bit network. If the average file size

17. This analysis is taken from Leonard Kleinrock, "The Latency/Bandwidth Tradeoff in Gigabit Networks," *IEEE Communications Magazine* (April 1992), pp. 36-40.

18. Light does not travel through fiber-optic networks at the same velocity as in a vacuum. Most fibers have an index of refraction of roughly 1.5, which reduces the speed of light by a third.

19. This analysis assumes the average file on the network is the size of a FAX, that is, 1 million bits. Most users would not use FAX technology to transmit data through a computer network because it would take too long to print. Actual data files are likely to be much smaller.

20. A baud is a measure of modern transmission speed and does not correspond to bits of data per second.

grows by a factor of 25, the faster network is 20 times as fast.

Thus, the commercial success of the HPCC's fast network technology will depend on the spread of applications that can make economic use of the capacity. Except in some obvious areas, such as medical imaging and video conferencing, such uses may be slow to become widely used.

Advances in Other Areas of Technology

Rapid advances in computer technology in other areas can temporarily substitute for acceleration of transmission speeds in several ways. Software and hardware designers are designing systems that reduce the volume of data that must be transmitted, using a technology called data compression or simply compression. The effective result is the same as an increase in communications speed. If, for example, the size of a file to be transmitted is compressed to half its size, then the effective speed of communication--from the user's point of view--is doubled. Advances in compression technology have also reduced the cost substantially. In 1992, AT&T announced a set of integrated circuits costing $400 that would perform many of the same compression functions previously performed by printed circuit boards costing $25,000.[21]

Electronic systems designers have substantial incentive to improve compression technology and already use it to allow consumers to defer expensive investments or increase capabilities. Several of the more popular pieces of personal computer software claim to expand the capacity of existing computer hard disks by compression. Similarly, the growing field of multimedia applications depends on compression technology. Already some experimental systems have used compression to produce full-motion video over the copper telephone (twisted-pair) wires connecting most homes. Home trials are expected to begin soon.

Compression technology has two closely related advantages over accelerating transmission speeds. First, it can use existing infrastructure. Second, it does not have to advance the whole system to achieve its end. Designers of compression technology can cater to selected final users without concerning themselves much with the rest of the system. They can target high-value and specialized niches much more rapidly than can providers of network infrastructure, who, by definition, must compromise among users. To some degree, they can beat the network providers to the early, most valuable customers.

The potential of compression technology to substitute for network technology has several limitations, however. The techniques using the highest compression ratios--those that reduce the file to be transmitted by the largest percentage--currently discard part of the data. Some economic uses, such as video transmissions to the home, can accommodate these losses, while others, such as medical images, cannot. But even the compression techniques that lose data are limited in their ability to compress it. Although they may slow the migration to higher network capacity, they cannot prevent it.

Parts of HPCC are concentrating on compression R&D and presumably will play a role in increasing that technology's power and availability. This particular trade-off between different aspects of HPCC is not unique and can work in any direction. Depending on the level of technology development, software may be interchangeable with hardware, and communications capacity may be interchangeable with computing power. Such substitutability is common in complex technology development programs.

A similar advance in electronic technology that reduces the need for high-speed networks is the splitting of electronic signals. This tech-

21. John Keller, "AT&T Unveils a Chip Set That Lets PCs Be Used for Video Communication," *Wall Street Journal*, April 3, 1992, p. B2.

nique involves combining and splitting the signals that come down the transmission line, substituting many small channels for a single large one.

The technology is quite mature and has been in use for a long time. When combined with some new technologies that shift signals in time (called delay equalization) and with changes in pricing and service structures of long-distance carriers, it substitutes for higher transmission speeds. For instance, by combining seven 56,000 bit-per-second lines, a video conference system can be transmitted without the need for special lines. The long-distance carrier Sprint provides a commercial service (Healthcare Application Network Delivery System) that uses this type of system to transmit video conferences and medical images.

End users also employ these techniques to split a large leased line into many small lines. In fact, the leased lines described above are often used not for single large applications but to agglomerate many small applications, even mixing voice and data transmissions. Some of this equipment also has limited switching capacity.

Compression techniques and splitting provide instances in which an advance in computational technology is substituted for an advance in communications technology. Since the growth in computer technology does not seem to be slowing, there is substantial reason to believe that these and other advances may continue to delay the need for faster communications technology. In both compression technology and splitting technology, however, market growth is hampered by lack of compatibility and common standards. Much will depend on the extent to which producers can reach agreement.

As noted above, computer power can be substituted for communications power only to a limited extent. As computers become ubiquitous, the need to link them rises. This linking is likely to be guided by economic considerations, which does not mean that it will inevitably favor the consumer uses often described by data communications forecasters.

Perspective on the Obstacles to the Development of a Commercial Market for HPCC Technology

Obstacles to development are not insurmountable. If high-speed data transmission is costly, firms and individuals will be encouraged to find ways around the expensive component or to make do with less. Although declining prices increase demand, so also does expansion in the pool of potential users. A case in point is the growth of local area networks that connect all the computers within a building or campus. Connecting two aggregations of computers and computer users is likely to be more valuable than connecting individual computers that happen to be at great distances from each other. Thus, increases in the number of local area networks are likely to expand the demand for larger networks.

These shaping forces are creating opportunities for entrepreneurs to make use of the type of technology that is being developed under HPCC. In this perspective, the scenario for network technology looks better than that for supercomputer R&D. This is not to say that all the firms that use HPCC network technology will be successful, or that none of the firms attempting to market HPCC supercomputer technology will be successful. The outcome will depend on events that are beyond the scope of this study.

Chapter Five

Policy Directions and Conclusions

Commercial demand is not the only reason high-performance computing and communications technologies are being pursued. Federal agencies can use some of the capabilities of high-performance computing or communications--indeed those are among the primary goals of HPCC. In the short term, however, agency missions are often in conflict with commercial success. Federal agencies differ from most potential users of the new technologies in having more technical expertise and being somewhat less sensitive to costs. In setting policy for HPCC, policymakers must decide how much weight to assign to each set of goals: the short-term goals of federal agencies and the long-term goals of commercialization.

The Computer Systems Policy Project (CSPP), in its evaluation of HPCC, argued that the computing problems the program intended to address, the so-called grand challenges, were too distant from those most people encounter to provide much of a demonstration of the ways in which computer networks could actually help people. They proposed that the program focus on more mundane applications.

Other commentators have suggested that the federal government should help create this new information infrastructure, just as it has played a role in creating other components of infrastructure such as the Interstate Highway System. In this case, however, the roles of the federal government and private parties might be reversed. When the Interstate Highway System was built, its everyday benefits were apparent to anyone who drove a car or truck. The federal government's unique contributions were the financial resources to build the highways and the legal power to get the land and rights of way. In the HPCC case, many private parties can afford to build major portions of the information infrastructure, but most people do not yet have a use for it. The appropriate federal role may be to help develop early applications of the new technology.

In response to these concerns, the Congress is considering whether to expand the network component of HPCC. The most commonly discussed forms of this expansion would create applications in four distinct areas under the rubric of an information infrastructure development program.[1] The four areas are medicine, education, manufacturing, and libraries.

- *Medicine.* The National Library of Medicine would develop network applications for communicating medical images, such as X-rays and CAT (computer-aided tomography) scans, by computer network; build test-bed networks linking hospitals and other medical centers to enable them to share data and records; and develop network applications to provide long-distance medical care.

- *Education.* The National Science Foundation would be authorized to develop pilot projects to connect U.S. elementary and

1. Several versions of this option have been introduced in the 103rd Congress. S. 4, introduced in January, incorporates such an option as Title VI; most of this legislation is unrelated to HPCC. H.R. 1757 and parts of H.R. 820 also incorporate much of this expansion.

secondary schools to Internet and to generate software and training materials.

- *Manufacturing.* The National Institute of Standards and Technology would develop and transfer electronically networked manufacturing applications, including standards development.

- *Libraries.* Both the National Science Foundation and the National Aeronautics and Space Administration are responsible for developing prototype digital libraries and the associated technology. NASA would make specialized government sensing data available over Internet.

To pay for these programs, the proposals would typically authorize an additional $1.0 billion to $1.6 billion over the next five years. If funded at the proposed level, total spending for high-performance computing and communications technology would rise to between $5.0 billion and $5.5 billion over the 1992-1996 time frame, depending on when the proposal was passed, for an average of $1.0 billion to $1.2 billion per year.

Aside from these Congressional initiatives, HPCC has shifted the focus of some research and development to accommodate the CSPP critique. For example, in choosing "grand challenge" supercomputing projects, the Department of Energy has begun to select projects that will use the high speed of the massively parallel supercomputer to provide answers to some applied research questions of near-term economic significance. DOE has included projects using computational chemistry to study various aspects of industrial pollution and alternative ways of alleviating it. Another project employs computational fluid dynamics to study industrial processes with the aim of increasing energy efficiency and lowering the output of pollutants. DOE is also funding R&D toward the modeling of petroleum reservoirs and groundwater. These projects are in addition to the long-range ones to which DOE is committed as part of its grand challenge problems, such as those involving high-energy physics.

Networks and Medicine

The medical program is designed to improve the use of information technology in manipulating and transferring medical information, from centralized recordkeeping to medical imaging. The major questions regarding the sharing of patients' records are legal and organizational as well as technical and economic: How does one organize a cost-effective system that allows medical personnel timely access to useful information, yet preserves the privacy of patients and satisfies the needs of medical organizations?

Medical imaging is seen as an early potential user of high-speed data transmission. Hospitals and medical centers have to get X-rays into the hands of radiologists and other doctors for interpretation. Other medical images that doctors need to see on a regular basis include angiograms, sonograms, CAT scans, and magnetic resonance imaging (MRI) scans. (Cardiologists are already faxing electrocardiogram results to colleagues for diagnosis.)

Hospitals have been transmitting medical images over telephone lines to radiologists, who use personal computers in their offices or homes. In this way a team of radiologists can serve more than one hospital and make more efficient use of their time. Because ordinary X-rays contain so much data, they are difficult to transmit over phone lines.[2] Many other radiological images, for instance CAT scans and MRI scans, can be easily transmitted this way. Transmission time is slow, however, requiring roughly half an hour to send a complete scan.

Because conventional telephone lines take so long to transmit sophisticated images, medical centers cannot use them for most of their imaging needs. A large medical center

2. An ordinary X-ray can be converted by a laser scanner into an array of data 2,000 numbers wide by 2,000 numbers high by 12 numbers deep (for shades of gray), or roughly 50 million bits of data. Even if the resolution is reduced to 1,024 by 1,024, an X-ray still has over 12 million bits of data.

at the height of its work day might produce a medical image every few seconds. For this reason, research in medical imaging has turned to computer networks.

Types of Medical Imaging Communications

Medical imaging transmission needs can be divided into two types: for local areas within major medical centers and, for wider areas, between medical centers of differing sizes or from medical centers and hospitals to individual doctors and small practices. The two types have different communications needs, and each is likely to find a different cost-effective solution.

Local Communications. Several private demonstration projects have begun. For example, Washington University, which has a major radiology institute, has teamed up with private companies to develop a high-speed network to give radiologists rapid access to medical images.[3] The objective is to create a network initially capable of delivering 155 million bits per second to 128 ports, and ultimately of delivering 620 million bits per second to 256 ports

Wide-Area Communications. Sprint has demonstrated a system for transmitting medical images between medical centers that joins together Sprint's current switched services (56,000 bits per second). It allows customers to join together temporarily several low-capacity telephone lines to provide a single large-capacity telephone line without having to build new capacity to accommodate the peak load.

This service permits communications between centers that have a moderate amount of radiological data to transmit. For instance, the armed services have contracted to have a switched service so that much of their radiological work can be concentrated in regional centers. A small fort or post may not have sufficient demand to justify a radiologist, and for a center that sends relatively few images (several dozen a day) the cost of using dedicated private lines would be exorbitant. Local exchange carriers are also beginning to offer some switched data services that allow users to pay only for the time the lines are in actual use.

Many radiological practices and individual radiologists still use ordinary telephone lines, modems, and personal computers to transmit medical images. Currently available technology has a lot of unused potential, often because the load does not warrant special investments. The ultimate transmission vehicle for small practices is still unclear.

Obstacles to Market Development

By concentrating on medical imaging, HPCC may be able to avoid the fate of earlier telemedicine demonstration projects. Earlier projects were unable to show that adopting sophisticated broadband telecommunications equipment led to significant increases in delivered health care despite higher delivery costs.[4] Furthermore, many of the benefits that did occur were easily obtained using a telephone. Most of these early experiments involved replacing face-to-face contact between doctor and patient with some electronically intermediated contact, whereas much HPCC research will be concentrating on efficient use of information within the health care delivery system.

Parallel with the need for transmission of medical images is that of storing, cataloging,

3. Jerome Cox and Jonathan Turner, "Project Zeus: Design of a Broadband Network and its Application on a University Campus" (Department of Computer Science, Washington University, St. Louis, Mo., 1991).

4. For a review of earlier efforts to use sophisticated telecommunications technology in the provision of medical services, see David Conrath, Earl Dunn, and Christopher Higgins, *Evaluating Telecommunications Technology in Medicine* (Dedham, Mass.: Artech House, 1983), pp. 197-217.

and retrieving these medical images. (These functions might fall under the purview of the library systems described below.)

None of the current systems has proved adequate for all the uses to which it might be put. For example, no software exists to help researchers search through X-rays for patterns to help them test theories about the development of certain medical conditions. Such limitations have prevented the electronic storage of medical images from becoming as widely used as it might be, given the stage of hardware development. Moreover, without software that can electronically store, search, and manipulate the images, the transmission of medical images becomes much less valuable. To develop fully the market for electronic exchange of medical images, these software problems will have to be addressed.

Networks and Education

Most versions of the education proposal would have the National Science Foundation develop pilot programs to connect primary and secondary schools to NSFNET and Internet, and develop software, systems, and training for teachers and students. Although NSF could develop materials and perhaps subsidize Internet connections, the ultimate responsibility for setting up computer networks at the local school level would most likely lie with state and local authorities, who in fact have shown themselves quite willing to move rapidly.

Existing State Networking Programs

Twenty to thirty states already have state-level computer networks dedicated to education. Although they vary in sophistication, the Texas Education Network (TENet) is often described as the most comprehensive. Started by a $1.2 million grant from the state legislature, TENet had 13,000 accounts in the first year of service and was growing rapidly. Teachers, administrators, parents, and other interested people connect to TENet an average of 3,000 times per day. TENet provides data bases to help teachers develop work plans for their classes. Some curriculum plans have been developed cooperatively over the network. Matchmaker services help teachers and classes find correspondent teachers and classes to undertake joint projects--for example, comparing the cost of family groceries across the state. In addition, TENet provides the usual computer network services--electronic mail, access to Internet, and file transfer.

The Texas state legislature funded TENet only after a comprehensive plan had been drawn up to develop computer use for education in kindergarten through grade 12. To ensure success in reaching its broad target group, TENet piggybacked on the Texas Higher Education Network that serves the state universities. This move reduced TENet's costs and ensured the expertise needed to run such a system. TENet had a comprehensive training strategy to ensure that there would be a trained representative in every school district in Texas. The system administrators also made every effort to ensure that the network included information that schoolteachers would find of immediate interest--curriculum development materials, for example. They rewrote software to make it easier for teachers to use, and made it affordable to encourage participation. Along with TENet, the State of Texas also provided $100 million to fund technology grants to local schools to encourage them to buy technology training, hardware (including calculators), and software or educational systems to help spread technology throughout the state educational system.

Obstacles to Market Development

Whatever the efforts of HPCC in developing applications, the trend toward extensive use of computer networks at the school classroom level will probably be limited, at least in the near future, for a number of reasons. Only one-third of elementary and secondary school teachers have had as much as 10 hours of computer training.[5] Existing training focuses more on computer literacy than on how to use the computer in education. Most classrooms at the elementary and secondary level do not even have telephone lines. Rewiring all the elementary and secondary schools would be expensive, and most education budgets are already under pressure. Most of the existing computers in public schools are not appropriate for network use and would require upgrading or replacement.

Despite these limitations on near-term market development, federal agencies may be useful in providing leadership to the states' efforts in educational technology. Although many school districts are conducting sophisticated experiments, progress is uneven nationwide, and every state and district has had to relearn expensive lessons. As the main sponsor of Internet, the federal government is also in a unique position to ensure that this resource is made available to education. As use of computer networks becomes more widespread, schools may have to introduce students to computer networks in order to prepare them for their lives as adults in a networked society.

The proposals under consideration suffer from several shortcomings. First, they are defined from the perspective of technology rather than education. The National Science Foundation, not the Department of Education, is the organizational locus. Policy seems to be focused on getting educators to use Internet rather than using Internet to solve the problems educators face.

And, unlike the Texas educational technology endeavor, these proposals largely fail to address the program's main constraint--the lack of teacher training in advanced computer technology. They include funding for the preparation of training materials, but not for training.

As noted above, entering Internet is a daunting experience. Without training or considerable experience, most teachers will have difficulty using any resource available through the network, regardless of its quality. Policymakers must decide whether the use of technology in education would be advanced more by spending several hundred million dollars to develop specialized network resources or by training teachers to use existing network resources.

5. Lynn Olson, "Contrary to Predictions, Use of Technology in Schools Elusive," *Education Week*, Special Report (January 8, 1992), p. 2.

Networks and Manufacturing

Federal agencies have been trying to promote computer networking in manufacturing for years without much success. The National Institute of Science and Technology (NIST) has been very active in fostering several manufacturing-oriented protocols for computer communications. These protocols include GOSIP, the government version of the OSI (Open Systems Interconnection) protocol, and MAP (Manufacturing Automation Protocol), but the market for capital goods using these protocols has been very limited thus far. Commonly cited reasons for the market's failure to take off include the inability of the vendors to agree on standards, expensive and relatively primitive technology, and cyclical problems in the major industries that would use the network.

Networks and Libraries

By contrast, libraries are readily taking to computer networks. The fit is quite natural: networks increase both the resources available at any library and the potential user base of any library. For example, one of the services provided on the TENet allows Texas teachers on-line access to the U.S. Department of Education Research and Information Center (ERIC) data base, giving teachers in distant communities access to much more information than would be available to them locally. Since many teachers use TENet in the evening, having libraries on-line means that they need to have fewer resources at home. For example, TENet makes encyclopedias available. Many libraries also have their card catalog on-line.

Similarly, the National Library of Medicine allows doctors to search the literature from remote locations. As in medical imaging, networks and medicine have begun to work well together in this area.

In addition to these public data bases, many private firms provide electronic data-base services, including many of the commercial network services discussed in Chapter 4. For example, Compuserve has information on airline schedules, and various dial-up services sell data for a fee. Most of these services, as well as the public data bases, are text retrieval services and lack experience with the much larger data sets or protocol issues that come with electronic imaging.

As noted in the medical section above, substantial questions arise in making digital libraries useful to researchers. In the medical area, for instance, the inability of currently available software to sift through medical images according to predetermined criteria has limited the usefulness of networks to radiological research.

HPCC is already substantially involved with the creation of a national storage system as part of NSF efforts to integrate all the resources at its supercomputer centers. This effort includes promoting standards for filing systems that would be accessible by network and creating systems for distributed mass storage--that is, permitting the integration of data that are stored in different computers at different sites. Using such data requires universally accepted conventions and protocols.

Program Balance

The proposals to expand the network component of HPCC could provide additional funding for current programs, or they could replace current activities. Replacing current activities with those suggested in these proposals would reduce HPCC's current focus on supercomputer R&D and result in a more even division between supercomputer R&D and network R&D.

The two options reflect the unresolved tension in the direction of HPCC. Is it ultimately concerned with aiding federal missions by providing networks that link high-performance computers, or with developing technology to make computing inexpensive, communications technology easy to use, and computers therefore ubiquitous?

Until now, the fact that the HPCC budget was spread among different parts of the discretionary spending category mitigated this tension and prevented it from surfacing completely. Starting in fiscal year 1994, however, the Budget Enforcement Act puts all discretionary spending in a common pot. Policymakers, for example, during the current budget debate can remove supercomputer R&D funds from the Advanced Research Projects Agency to pay for network R&D in the NSF, NIST, and NASA. Given the cuts in discretionary spending mandated by the act, policymakers may have to decide which part of HPCC is most critical. Furthermore, the agencies centrally involved in HPCC--NSF and NASA--experienced real declines in budget authority in 1993 and will have to make internal decisions regarding the different components of HPCC.

Conclusions

The different elements of HPCC have different potentials. Clearly, graduate students trained under the human-capital part of HPCC are likely to find jobs in academia where they will educate the next generation of computer programmers, and in industry where they will write computer programs. It is less easy to foresee the outcome of other parts of HPCC.

Supercomputer Technology

The scalable massively parallel supercomputer systems that HPCC is supporting must overcome very large obstacles before they can succeed commercially. As noted in Chapter 3, growth in the workstation market may preempt substantial growth in the supercomputer market. If that occurs, firms may not be willing to make the large investments in product development necessary to bring massively parallel supercomputer systems to market. If they do bring them to market, they may not sell enough of them to pay for those investments.

The supercomputer market is already occupied by a successful incumbent, the vector supercomputer, sales of which have risen from $100 million to $1.5 billion over the last decade. This incumbent is armed with a backlog of well-written, fine-tuned, ready-to-run software. Nor is it standing still; rather, it is advancing in technology, even becoming more parallel, albeit at a less rapid pace than the massively parallel supercomputers.

Dislodging incumbents in computer markets is not easy. New entrants usually go around their competitors by introducing new capabilities or sizes or price levels. The minicomputer avoided direct competition with the mainframe computer, as the personal computer did with the minicomputer. Whether the additional speed provided by massively parallel processors on the limited number of problems they are uniquely qualified to answer will provide a sufficient distinction in the marketplace is not yet clear.

Network Technology

Demand for computer communications, by contrast, is growing quite rapidly, and there is no dominant incumbent technology. Further, a large component of the federal effort in this area is addressed not so much to technology development as to institution building, an area in which the federal government has a comparative advantage because it is perceived as a disinterested party. Although Internet has done much to push back the frontiers of technology, its major achievement has been to create a network of individuals and organizations who have learned how to work together in maintaining lines of communication. But even though the field of data communications is growing rapidly, this growth will not necessarily translate into a demand for services providing the types of speed HPCC is designed to provide.

The next arena for policy discussion is whether to begin a second phase of HPCC that would develop more applications for its network capability. Members of Congress and industry have suggested developing and demonstrating such applications in areas as diverse as education, medicine, and manufacturing. Legislation has been introduced that would authorize $1.0 to $1.6 billion over the next five years for this purpose. In some of the areas the federal program could build on, there is already substantial activity. In addition, federal, state, and local governments already play a substantial role in both education and medicine.

Appendixes

Appendix A

Supercomputer Component Technology Spinoffs

This appendix describes major areas of computer technology and shows in each case how supercomputer technology requirements differ from the requirements of the rapidly growing parts of the computer market. In general, it is these other parts of the market that are leading the development of technology. Table A-1 outlines the areas of computer technology that are touched on and notes how the requirements of the supercomputer market differ from those of other computer markets.

The word "technology" as used in this report means the ability to perform certain work at a given cost level: technology is defined in both technical and economic terms. That two computers or electronic components have the same capabilities or features does not mean they have the same technology. For example, in 1973 the XEROX Palo Alto Research Center invented the ALTO computer, which had many of the features now commonly seen in a state-of-the-art personal computer, including graphical-user interfaces, networks, a pointing device like the "mouse," and so forth. Although the technical features are the same, the technology of the computers is different: the XEROX machine would have cost $25,000, while the personal computer with these features now costs less than $1,000.[1] The components have changed, the software has changed, the way they work together has changed, and the way they are manufactured has changed.

Thus, it is not surprising when features that appeared first on supercomputers eventually appear elsewhere. When other computer makers copy a popular feature, this is a response to perceived demand and not necessarily a spinoff of the same technology. In fact, the ability to provide the same feature in a more convenient package at lower cost is characteristic of technological advance.

Integrated Circuits

The largest advances in integrated circuits have been found in complementary metal-oxide-semiconductor (CMOS) technology.[2] This type of integrated circuit permits low power usage with a high transistor count (or density) at relatively low cost, but it is relatively slow. Throughout the 1980s, most of the microprocessors and memory chips at the heart of the personal computer increasingly employed CMOS or its metal-oxide-semiconductor (MOS) relatives. The slowness of the chips did not matter to most personal computer or workstation buyers given the relatively low cost and high number of the components. (A higher transistor count allows an integrated circuit to perform more tasks or retain more information in its memory.)

By contrast, conventional supercomputer designers have had to work with emitter-

1. Robert Cringely, *Accidental Empires* (New York: Addison-Wesley Publishing Co., 1992), pp. 83-84.

2. For a simple if dated description of CMOS and other semiconductor devices, see James D. Meindl, "Microelectronic Circuit Elements," *Scientific American* (September 1977), pp. 70-81.

Table A-1.
Major Computer Technologies

	Integrated Circuits				Printed Circuit Boards				
	Type	Charac-teristics	Transister Count	Archi-tecture	Packed	Layers	Cooling	Disks	Operating Software
Supercomputers	ECL	Fast	Relatively Few	Vector Arrays	Densely	Many	Liquid	Fast	Unix
Other Major Computer Markets	CMOS	Cheap	Many	RISC	Less Densely	Few	Air	Cheap	Unix and Other

SOURCE: Congressional Budget Office.

NOTES: ECL = emitter-coupled logic; CMOS = complementary metal-oxide-semiconductor; RISC = reduced-instruction-set computing.

coupled logic (ECL), a fundamentally different integrated circuit technology that, although it has fewer components and uses more power, is substantially faster. Since supercomputers are optimized for speed, they usually have not had the option of using the CMOS devices. Thus, the ECL markets and MOS markets have been relatively separate; even the producers of these different integrated circuits are not necessarily the same.[3]

As the speed of workstations has increased, CMOS devices have proved too slow. The BiCMOS technology for integrated circuits was developed to meet the need for dense but fast components. This technology is much faster than the conventional CMOS, but still has quite a high transistor count. The engineering workstation market has been the primary driver of the development of this technology.[4] Now both mainframe and supercomputer producers are looking at BiCMOS technology, especially in memory, for their next generation of computers.[5] In this sense, supercomputer technology is beginning to benefit from the technology developed for the ubiquitous and inexpensive personal computer and its more expensive cousin, the engineering workstation.

Within their market segment, however, supercomputer designers have improved or been among the first to adopt new ECL technology as it became available, and thus have influenced mainframe computer technology, often by pushing their vendors. For instance, Cray Research bought its ECL gate arrays from vendors and pushed them to increase the complexity of the metal layers and to increase the number of layers from three to four.[6] The vendors subsequently made these improvements available to other ECL clients, primarily in mainframe computers. Before its current generation of supercomputers, however, Cray Research mainly bought its integrated circuits as off-the-shelf components, preferring to concentrate its research and development efforts on novel packaging and interconnections.

3. See Harold Dozier, "Superchips for Supercomputing," *IEEE Spectrum* (September 1992), pp. 66-68.

4. Occasionally, a CMOS workstation microprocessor design will be implemented in ECL technology, but such implementations are not driving the technology of workstation microprocessors.

5. David Patterson, "Expert Opinion: Traditional Mainframes and Supercomputers Are Losing the Battle," *IEEE Spectrum* (January 1992), p. 34.

6. A gate array is a kind of semicustom integrated circuit composed of a matrix of logic gates (or switches). The layers of metal connecting these gates give each integrated circuit its unique electronic character.

Computer Architecture

The other major innovation in computer technology of the last decade was reduced-instruction-set computer (RISC) architecture, which is at the heart of contemporary workstations and allows them great speed despite the use of CMOS components. Most mainframe and personal computers, by contrast, still use complex-instruction-set computer (CISC) architecture.

The central difference between the two types of architecture is in the number and types of instructions a computer understands. A RISC machine understands relatively fewer instructions, but executes them rapidly. A CISC computer has a larger vocabulary with more complex instructions, but executes them more slowly.

The insight that led to RISC is that most computer tasks actually involve a relatively small number of operations, such as adding two numbers, so it is advantageous to be able to perform these common operations rapidly, rather than to be able to perform a wider array of operations more slowly. By way of comparison, when RISC was first introduced, the major CISC families were capable of performing between 200 and 300 instructions, while the early RISC computers had only between 40 and 120 instructions.[7]

IBM first started work on RISC architecture in the 1970s, and other research was undertaken at Berkeley and Stanford, although the latter two projects were unaware of the IBM efforts.[8] RISC architecture became widespread as the price of computer memory fell dramatically throughout the 1980s. (RISC software programs, because they are composed of many simple instructions, typically take up more memory than equivalent programs designed for CISC.) While many supercomputers have had elements of RISC--mainly a simplified instruction set--in them since the 1960s, supercomputers did not lead in RISC design. The engineering workstation, which is the major market for RISC, was the primary force behind software and hardware development in this market.

As RISC architecture first developed in microprocessors for engineering workstations, designers copied features from existing mainframe and supercomputer designs. The high-performance features most commonly copied included pipelining, which, as discussed in Chapter 3, involves creating an assembly line of instructions, and cache memory, which functions as a clipboard for frequently used items in the computer memory and gives fast access to these items.

However, as CMOS integrated circuit technology progressed, RISC designers were able to advance the use of these features very aggressively. According to John Hennessy, one of the original RISC designers, "a number of [RISC] microprocessor-based machines have used pipelining and cache techniques that are more advanced than those in use in the biggest machines."[9] The ease of redesigning large-scale integrated circuits, relative to redesigning conventional central processing units, encouraged rapid development of RISC architecture. Thus, although RISC architecture inherited much of its technology from existing high-performance computers, it moved at a much faster pace and soon began advancing technology on its own.

The Defense Department's Advanced Research Projects Agency (ARPA) paid for much of the early R&D on RISC at Berkeley and Stanford as part of its work on very-large-

7. David A. Patterson, "Reduced Instructions Set Computers," *Communications of the ACM* [Association of Computing Machinery] (January 1985), pp. 8-21.

8. David Patterson and John Hennessy, *Computer Architecture: A Quantitative Approach* (San Mateo, Calif.: Morgan Kaufmann Publishers, 1990), pp. 189-191. For the IBM contribution, see Ronald Jurgen, "John Cooke: Vision with Enthusiasm," *IEEE Spectrum* (December 1991), pp. 33-34.

9. John Hennessy and Norman Jouppi, "Computer Technology and Architecture: An Evolving Interaction," *Computer* (September 1991), p. 22.

scale integrated circuits in the late 1970s.[10] ARPA also paid for much of the early R&D involved in developing the first engineering workstation and funded much of the early R&D in massively parallel supercomputers. This juncture is no coincidence: ARPA was funding several different avenues of research to achieve the same goal--faster computers. In the words of one technology analysis,

> Where RISC architects sought to achieve efficiencies by incorporating within microprocessor structure the best possible juxtapositions of cache memory, registers, message routers, and logic units so that they would interact in the most efficient way, parallel architects sought to use separate processors in communication (interconnect) structures that would achieve those kinds of efficiencies.[11]

Having grown primarily in workstation design, RISC architecture is now moving to other areas of the computer market. Mainframe producers have announced their interest in such architecture. In addition, the next generation of IBM-compatible personal computers is likely be powered by an Intel microprocessor (the Pentium) that contains elements of RISC architecture. Lastly, many, if not most, of the massively parallel supercomputers are based on RISC microprocessors originally developed for workstations. As noted in the main text, much, if not most, of the improvement in the performance of massively parallel supercomputers during the last decade can be attributed to the use of ever more sophisticated workstation RISC microprocessors.[12]

A growing number of workstations and even high-end personal computers use more than one microprocessor, in a trend reminiscent of but independent from the development of massively parallel supercomputers.[13] Sun Microsystems, Silicon Graphics, and others already make or have announced products containing more than one processor. Thus, while supercomputer technology is advancing by linking massive numbers of microprocessors, workstation technology is similarly advancing by linking small numbers of the same RISC microprocessors. Because manufacturers of vector supercomputers have had substantial experience with multiprocessors, computer designers are likely to have drawn on that experience in producing multiprocessor workstations.[14]

Printed Circuit Boards and Other Packaging

Supercomputer producers have confronted problems in the area of printed circuit boards well ahead of other producers of electronics equipment. In order to make the supercomputers run faster, producers have placed the electronic components close together and have had to design sophisticated multilayer boards to accommodate them.[15] The fast components used by supercomputers put out more elec-

10. Richard Van Atta, Sidney Reed, and Seymour Deitchman, "DARPA Technical Accomplishments: A Historical Review of Selected DARPA Projects," Volume II (Alexandria, Va.: Institute for Defense Analysis, February 1990), pp. 17-A-1 through 17-A-4. Chapter 17 and its appendixes concentrates on DARPA's efforts in integrated circuit and computer design technology in the later 1970s and early 1980s.

11. Ibid., p. 17-23.

12. For this history of improvement, see John Van Zandt, *Parallel Processing in Information Systems* (New York: John Wiley and Sons, 1992).

13. Norris Parker Smith, "Workstations and Supercomputers," *Supercomputing Review* (January 1992), pp. 50-51.

14. For a history of multiprocessing in workstations, see Patterson and Hennessy, *Computer Architecture*, p. 588ff.

15. Supercomputer Systems, an IBM-backed Cray spinoff, announced that the printed circuit board for its forthcoming supercomputer would have 78 layers. Terry Costlow, "SSI Supercomputer to Use Superboards," *Electronic Engineering Times* (July 20, 1992), pp. 1 and 64. (Since this announcement, the company has had to close for reasons unrelated to its circuit boards.) By contrast, one printed circuit board for the next generation microprocessor for IBM-compatible personal computers will have only six layers. Rick Boyd-Merritt and Alan Patterson, "Asians [personal computer producers] Get Chance to Plug in Pentium," *Electronic Engineering Times* (January 4, 1993), p. 60.

tronic noise than slower, low-power devices; consequently, the supercomputer designers have become expert at isolating both the integrated circuits and the circuits that connect them, using special materials and other sophisticated techniques.

As conventional electronic equipment grows in sophistication, its producers will confront many of the same problems. For instance, as workstation microprocessors become faster, they outrun the electronic signal speeds that can be transmitted easily through conventional printed circuit boards. Consequently, some have begun to turn to multichip modules, where the central processing unit (or microprocessor) and several peripheral chips are contained within specially designed modules.[16] Vendors who have had to supply supercomputer producers may find some of their expertise applicable in these new areas.

Because supercomputers were designed to use large numbers of relatively simple and hot integrated circuits, their producers have confronted problems of packaging and heat dissipation long before most other computer producers have. Although most computers can be cooled by air and fans, supercomputers often need water and refrigerant systems.[17] As other computers increase their component count and speed they will confront the same package density and cooling problems that supercomputer producers have for years.

However, conventional computer producers will be constrained in ways that supercomputer producers are not. For example, workstation makers will want to retain the small size and relative portability of their systems. Given these constraints, it is not clear how much of the unique supercomputer packaging and cooling technology will migrate to the workstation or personal computer markets. Again, the mainframe computer market seems positioned as the most likely beneficiary.

16. Philip Koopman, Jr. and Daniel Siewiorek, "ICs: The Brains of a Workstation," *IEEE Spectrum* (April 1992), pp. 52-53.

17. In fact, so much waste heat is generated that it can be used for other purposes. Supercomputer waste heat is sometimes used to heat buildings.

Computer Disks

In the past, supercomputers have been at the forefront in many aspects of magnetic disk storage technology. Specifically, they have demanded ever-increasing speed from suppliers of hard disk drives. This has meant faster read/write heads, multiple platters, and multiple spindles, all of which have the potential to be carried over to the general computer market.

However, as with microprocessors, the needs of the more numerous desktop systems have dominated the industry. The small systems have emphasized the achievement of size and price goals rather than absolute performance. In fact, parallel with the use of large numbers of inexpensive microprocessors to create a supercomputer, computer designers have now begun to provide redundant arrays of inexpensive disk (RAID) subsystems as an alternative to the larger disk drives used by most mainframe and supercomputer producers.[18] Arrays of disks have been available on smaller systems and are beginning to migrate to mainframes. ARPA funded some of the early R&D in this area.

Supercomputer Software

At present, the vector supercomputers and minisupercomputers made in the United States largely use the same UNIX operating software as the engineering and technical computers. This is the same software that runs on the engineering workstations that

18. Terry Costlow, "RAIDs are Underway," *Electronic Engineering Times* (January 6, 1992), pp. 1 and 50.

dominate the design and technical uses. UNIX is as ubiquitous in that world as the IBM-originated operating systems (DOS, MVS) are in the world of business and finance. For example, since the mid-1980s, Cray Research has used UNICOS, a Cray version of UNIX. Because it is a version of UNIX, UNICOS can run many UNIX programs, which makes using these supercomputers simpler for those who are already familiar with UNIX software.[19]

19. Cray Research, "Directory of Applications Software for Cray Research Supercomputers" (Mendota Heights, Minn., January 1990).

Much of the software originally developed for supercomputers has found its way into the workstation market. The scientific visualization of it originated in supercomputer applications and only in recent years, as workstations have increased in power, has the software begun to migrate. Many of the numerical analysis applications have similarly migrated from supercomputers to workstations.

Appendix B

Supercomputer Speed Calculations

This appendix discusses the assumptions and calculations underlying the comparison of vector supercomputer speed with massively parallel supercomputer speed. The Congressional Budget Office (CBO) replicated the calculations of Peter Gregory in "Will MPP [Massively Parallel Processing] Always be Specialized?" which appeared in the March 1992 edition of *Supercomputing Review.*[1]

The simulations are based on Amdahl's Law and published estimates of the speed of components of different supercomputers, not on actual performance. Amdahl's law states that "the performance improvement to be gained from using some faster mode of execution is limited by the fraction of the time the faster mode can be used."[2] In other words, the speedup attributable to the use of a massively parallel processor will be limited by the percentage of a problem that can use parallel calculations.

The formula used to construct Figure 3 on page 27 was the following:

$$P = \frac{S}{(1\text{-}FP) + (FP/N)}$$

where

P is computer performance (in millions of floating point operations per second),

S is the speed of the individual processor (in millions of floating point operations per second),

FP is the fraction of operations that can be performed in parallel, and

N is the number of processors.

Numerical problems presented to supercomputers for solution typically have two types of processes: scalar operations, which involve single numbers, and vector operations, which involve strings of numbers. Each processor has a different speed for each type of operation. Each problem has a unique combination of scalar and vector operations. Thus, the formula was used to create polar cases, where problems were 100 percent vector and 100 percent scalar. Actual performance will lie between. Either way, except for instances where problems permitted parallel calculations that were almost entirely parallel, the vector supercomputer outperformed the massively parallel supercomputer. For purposes of expositional simplicity, CBO only presented the vector calculations. The addition of scalar performance would not substantially change the conclusions. In only one of the existing computers did inclusion of scalar performance shift the crossover point by more than 1 percentage point, and there it only served to reinforce the advantage of vector supercomputers.

The computers Peter Gregory chose for comparison were the Cray Research Y-MP C90, Thinking Machines CM-5, and the Intel Paragon. These machines are the premier ma-

---

1. See pp. 28-31.

2. David Patterson and John Hennessy, *Computer Architecture: A Quantitative Approach* (San Mateo, Calif.: Morgan Kaufmann Publishers, 1990), p. 8.

chines of three leading supercomputer producers. The assumptions about their speeds (in millions of floating point operations per second) and processor counts are shown below:

Computer	Scalar Speed	Vector Speed	Processor Count
Y-MP C90	29	1,000	16
CM-5	4	128	1,000
Paragon	7	75	2,048

When CBO performed sensitivity tests, even doubling the scalar speeds of the massively parallel supercomputers did not substantially alter the central conclusion. The vector performance is based on published peak speed.

The variable FP is defined as the percentage of operations that can be performed in parallel over the entire application. It assumes that parallel means that the entire number of processors of each machine will be used. This assumption is more favorable to the massively parallel supercomputer as it is easier to find problems that are divisible 16 ways than 1,000 or 2,000 ways. It also assumes instantaneous communication between nodes, an assumption that again favors the massively parallel supercomputer, since it has more nodes.

Appendix C

Network Speed Calculations

This appendix discuses the assumptions and calculations underlying the discussion of computer network response time.[1] The calculation of computer network response time is based on the following formula:

$$T = \frac{(F/C)}{(1\text{-}p)} + D$$

where

T is response time (in milliseconds),

F is the average file size on a network (in 1,000 bits),

C is the network capacity (in million bits per second),

p is capacity utilization (in percentage of total capacity), and

D is propagation delay because of the amount of time it takes light signals to travel between locations.

Unlike Kleinrock, the Congressional Budget Office does not assume that light travels at the same speed in a fiber optic cable as in a vacuum. Light travels through fiber at roughly two-thirds the speed it travels through interstellar space. Consequently, CBO assumes that it takes roughly 10.5 milliseconds for a light signal to propagate through a fiber halfway across the continental United States. Kleinrock assumed it would travel across the continental United States in 15 milliseconds.

CBO assumed that the average size of files on the computer network was 1,000,000 bits, roughly the size of a FAX.[2]

1. The discussion is based largely on Leonard Kleinrock, "The Latency/Bandwidth Tradeoff in Gigabit Networks," *IEEE Communications Magazine* (April 1992), pp. 36-40.

2. Lawrence Roberts, "Gigabit Networks: Emerging Commercial Applications & Opportunities" (presentation at the 1992 conference of the Technology Transfer Institute, Santa Monica, Calif., July 8-29, 1992).

CONGRESSIONAL BUDGET OFFICE
Second and D Streets, S.W.
Washington, D.C. 20515

583

ISBN 0-16-041828-3
90000
9 780160 418280