Chapter 2
High Performance Computers

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

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

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

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

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

WHAT ISA HIGH PERFORMANCE COMPUTER?

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

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

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

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

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

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

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

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

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

A few policy-related conclusions flow from this discussion:

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

HOW FAST IS FAST?

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

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

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

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

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

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

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

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

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

THE NATIONAL SUPERCOMPUTER CENTERS

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

The Cornell Theory Center

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

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

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

The National Center for Supercomputing Applications

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

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

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

Pittsburgh Supercomputing Center

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

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

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

San Diego Supercomputer Center

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

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

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

John von Neumann National Supercomputer Center

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

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

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

OTHER HPC FACILITIES

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

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

Minnesota Supercomputer Center

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

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

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

Table 2-l—Federal Unctassified Supercomputer Installations

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

SOURCE: Offics of Technology Assessment estimate.

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

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

The Ohio Supercomputer Center

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

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

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

Center for High Performance Computing, Texas (CHPC)

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

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

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

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

Alabama Supercomputer Network

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

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

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

Commercial Labs

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

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

Federal Centers

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

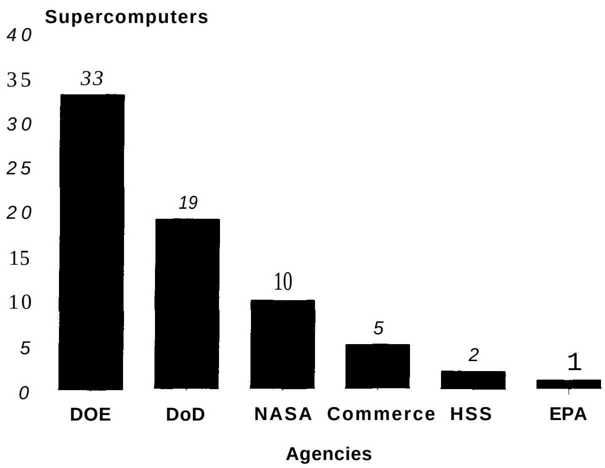
CHANGING ENVIRONMENT

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

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

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

Figure 2-1—Distribution of Federal Suparcomputers



SOURCE: office of Technology Assessment, 1989.

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

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

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

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

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

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

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

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

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

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

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

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

REVIEW AND RENEWAL OF THE NSF CENTERS

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

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

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

Some examples of possible choices are as follows:

L Provide Access to HPC

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

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

2. Educate and Train

- Provide facilities and programs to teach scientists and students how to use high performance computing in their research.

3. Advance the State of HPC Use in Research

- Develop applications and system software.
- Serve as centers for research in computational science.
- Work with vendors as test sites for advanced HPC systems.

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

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

THE INTERNATIONAL ENVIRONMENT

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

Japan

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Europe

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Other Nations

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

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

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

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

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

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