Chapter 2

Policy Considerations for High-Performance Computing

Currently the National Science Foundation (NSF) sponsors five leading edge computational centers, the four national supercomputer centers and the National Center for Atmospheric Researcher (NCAR) (see app. A). When the centers were established, one goal of the NSF initiative was to nationally provide researchers with access to leading edge technology. Prior to the NSF program, U.S. researchers and scientists had little opportunity---outside of Federal laboratories-to access supercomputers. Since their creation, the centers have been extremely successful in providing access to supercomputing resources to academic and industrial researchers.

The success of the NSF centers has made them the target of a debate over funding strategies for their support. It is noteworthy that they are not the only such facilities funded by the Federal Government or even by NSF. Computers, especially large-scale computers, always have required relatively large institutional structures to operate. The Department of Energy (DOE) and the Department of Defense (DoD) fund many more computational centers at a considerably higher cost than the NSF. Government establishment and support of scientific computing facilities date back to the earliest days of computing. Furthermore, high-performance computing is becoming increasingly important to all of science and engineering. The issue is not whether science, education, and engineering in the United States need high-performance computing centers, but rather how these centers should be supported, and how the costs of that support should be allocated over the long term.

It is imperative that the United States: 1) continue to steadily advance the capabilities of leading edge computer technology; 2) provide the R&D community with adequate computing resources; and 3) expand and improve the use of high-performance computing in science and engineering.

Advancing Computer Technology

Computers lie on the nearly seamless lines between basic research, applied research, and the development of new technologies. A program intended to advance the state-of-the-art of high-performance computing must include:

- physics research on fundamental devices, superconductors, quantum semiconductors, optical switches, and other advanced components;
- basic research in computer science and computer engineering, including theoretical and experimental work in computer architecture and a variety of other fields such as distributed systems, software engineering, computational complexity, data structures, programming languages, and intelligent systems;
- applied research and assembly of experimental laboratory testbed machines for exploring new concepts;
- experimentation, evaluation, and development of software for new prototype computers, e.g., the Connection Machine, Hypercube or neural nets;
- development of human resources and facilities for computing research needed to support a high-performance computing initiative, which requires additional trained researchers and research facilities;
- research and development of new technologies for data storage and retrieval (this may be the biggest technological bottleneck in the future); and
- creation of new algorithms tailored for advanced architectures to meet the needs of scientists and engineers for greater computational capabilities.

Difficulties and Barriers

Funding

The term “computational science” is used to define research devoted to applying computers to computationally intensive research problems in science and engineering. It is focused on developing techniques for using high-performance computing to solve scientific problems in fields such as chemistry, physics, biology, and engineering. Though growing, the base funding level for computer and computational science and engineering is currently low.

Defense Advanced Research Projects Agency (DARPA), NSF, DOE (particularly national laboratories such as Los Alamos and Lawrence Livermore Laboratory), the National Institute of Science and Technology (NIST), National Aeronautics and Space

Administration (NASA), and the National Institutes of Health have all contributed to improving the state-of-the-art of computer technology and its application to science and technology. There is, however, no clear lead agency to focus a national high-performance computing program.

A significant or substantial increase in the support of computational science as part of a high-performance computing initiative would require a relatively large additional investment. There is disagreement among researchers in the various disciplines about increasing funding for computational science. Some fear that investments in this area would reduce funds available for other research activities.

Procurement Regulations

In addition to the expense involved, obtaining prototype machines for experimental use has become more difficult because of some agency interpretations of Federal procurement law, In the past, research agencies have stimulated the development of advanced computer systems by purchasing early models for research use. Contracts for these machines were sometimes written before the machine was manufactured. The agency would then participate in the design and contribute expertise for software development. This cooperative approach was one key to advancing high-performance computing in the 1960s and 1970s. Unfortunately, the process has become more difficult as Federal procurement regulations for computing systems have become tighter and more complex.

Policy Issues

Federal support of computing R&D is intertwined with the political debates over technology policy, industrial policy, and the appropriate balance of responsibility between the Federal Government and the private sector in developing computer technology. Computing researchers study basic, and often abstract, concepts including the nature of complex processes and algorithms. But, the results of their work can have important practical implications for the design of computer hardware and software.

Computing research is often based on the study of prototypes and artifacts rather than natural phenomena. Consequently, Federal support is sometimes viewed as technological-rather than scientific—in nature. Moreover, Federal defense procurement directly supports the U.S. computer and software industry. Because of this relationship with industry, the High Performance Computing Initiative invariably blends the role of traditional Federal science policy with Federal efforts to support precompetitive activities of a strategically important industry. This has led to confusion and debate over the goals and appropriateness of the proposed High Performance Computing Initiative.

Providing Access to Resources

Federal support for educational and research computer resources must broker their use among many different users with different needs at many different institutions. Policies that serve some users well may shortchange others. There are three general objectives that serve all: 1) provide funds for acquiring computer hardware and software; 2) assist in meeting operational expenses to maintain and manage facilities; and 3) ensure that scarce computational resources are distributed fairly to the widest range of users.

No single Federal program for supporting scientific computing is likely to serve the needs and policy objectives for all facilities and user groups. Support must come from a variety of coordinated programs. For example, since the inception of NSF's Advanced Scientific Computing programs, debates over support of the national supercomputer centers have reflected many different, and often contradictory, views of the roles the centers should play and the constituencies they serve.

Difficulties and Barriers

Diversity of Sources

Computers are expensive to buy and to operate. For larger machines, usage crosses many disciplines and users are associated with many different academic institutions and industrial organizations. Supplying computer time can be a significant burden on research budgets, and support is often found by pooling funds from several sources,

No Natural Limits

Researchers seem to have an insatiable appetite for computer time. This perplexes policymakers who are used to dealing with expenditures for freed cost items. One can estimate the number and kind of laboratory apparatus a chemist might need or microscopes a biology laboratory can use, based on the physical requirements of the researchers. However, the modeling of a complex organic chemical molecule for the design of a new pharmaceutical could saturate significant supercomputer resources. The potential use for supercomputer capacity appears to be limitless.

Administrators at research laboratories and government funding agencies have difficulties assessing computing needs and justifying new expenditures, either for purchase of additional computer time or for investments in upgrading equipment. It is even harder to predict future needs as researchers conceive new applications and become more sophisticated in developing innovative computer uses. These conflicting demands on the Federal science budget require careful balancing.

Disincentives to Investment

Support for computing resources may come from individual institutions themselves by underwriting the capital investment. The capital investment and operation costs are partially recaptured through fees charged back to the users. However, this model has not worked successfully, for a couple of reasons.

First, a multimillion dollar high-performance computer is a risky investment for an individual research institution. The risk is even greater for experimental machines whose potential use is difficult to anticipate. The institution must gamble that: 1) there is sufficient potential demand among research staff for the facilities; 2) federally supported researchers will have adequate funds to cover the costs; and 3) researchers with funds will choose to use the new computer rather than an outside facility,

Networks expand the possible user community of the facility, but they also provide access to competing systems at other institutions. In the past, researchers were, by and large, captive users of their own institutional facilities. Networks free them from this bondage. Now, researchers can use ' 'distributed' computer resources elsewhere on the network. Faced with a wider ''market' for computer time, research institutions may have less incentive to invest in more advanced systems, and instead upgrade local area networks to link with the NSFNET high-capacity backbone. On the other hand, networks can improve the efficiency and cost-effectiveness of computing by distributing computing capabilities.

pricing policies for computer time must be carefully scaled to recover the costs of capital investments in hardware. High-computing costs can result in loss of revenue as researchers seek better rates at other institutions. The government *requires* that federally supported researchers pay no more than nonsupported researchers for computer time. But to ensure that operations break even, computer centers are forced to charge a rate equal to the costs divided by usage. This policy seems reasonable and equitable on the surface, but it results in higher rates for computer time when machine usage is light and lower rates as it grows. This pattern produces an upside-down market similar to that of the electric utilities before capital costs forced them to shave peak loads by charging a premium for power during periods of high usage. This is the reverse of airline rates where fares are lower when seats are empty and higher when planes are full.

Support Strategies

These disincentives and barriers have tended to limit investments in high-performance computers for research at a time when an increasing amount of important research requires access to more computational capacity. The Office of Science and Technology Policy's (OSTP) High Performance Computing Initiative, funding agencies' program plans, and pending legislation are aimed at balancing the Nation's R&D needs with high-performance computing capacity.

Four basic funding strategies to achieve this goal are described below:

Fully Support Federally owned and operated Centers

The most expedient strategy is to establish government-owned and operated facilities. The government could directly fund investments for hardware and software, and the centers' operational costs. There currently are several government funded and operated computational centers administered by the mission agencies. (See app. A, table A-l.) Government-fiianced computational centers provide a testbed for prototype machines and novel architectures that can help bolster the U.S. computer industry against foreign competition. Software development, critically needed for high-performance computing, is commonly a major activity at these centers.

A Federal high-performance computing initiative could select specific computational centers for full funding and operation by the Federal Government. A Federal agency might be needed to supervise the creation and management of the centers. Hardware would be owned or leased by the government. The center might be operated by a government contractor. The personnel, support staff and services, could either work directly for the government or a government contractor. These centers would be in addition to the existing mission agency computing centers.

Federally owned and operated computational centers currently exist under the management of several Federal mission agencies. The national laboratories---Los Alamos, Sandia, and Livermore —are operated by the DOE. Much of their work relates to national security programs, such as weapons research. NASA, DoD, and the Department of Commerce operate high-performance computing centers. NASA's centers primarily conduct aerospace and aerodynamic research. DoD operates over 15 supercomputers, whose research ranges from usage by the Army Corps of Engineers to Navy ship R&D to Air Force global weather prediction to intelligence activities of the National Security Agency (NSA). However, they do not fill the general needs of the science and education community. Access to these mission agency centers is limited, and only a small portion of the science community can use their facilities. The Federal Government could similarly own and operate computational centers for academic missions as well.

While federally owned and operated computing centers might risk experimentation with novel, untested computer concepts that academic or industrial organizations cannot afford, there is a possibility that this strategy could blossom into an additional layer of bureaucracy. The advantages of having direct government control over allocating computer time based on national priorities and acquiring leading edge technologies is offset by the risk of having government managers making decisions that should best be made by practicing scientists and engineers as is currently done at the NSF centers. Such shortcoming in systems management may be overcome by using nongovernment advisors or boards of governors, but centers could find it difficult to ensure stable year-to-year funding as national budgets tighten and competition for research dollars increases.

Fully or Partially Support Consortia or Institutionally Operated Centers

Federal science agencies can provide partial or full support to institutions for purchasing new computers. This is currently done by NSF and DOE. NSF provides major funding for four national supercomputer centers and the National Center for Atmospheric Research (NCAR) facility. DOE partially funds a supercomputer facility at Florida State University. The agencies provide funds for the purchase or leasing of computers and also contribute to the maintenance of the centers and their support staff. This has enabled the centers to maintain an experienced staff, develop applications software, acquire leading edge hardware, and attract computational scientists.

The government, through the NSF, provided seed funds and support to establish the centers and operate them. The NSF centers are complete computational laboratories providing researchers with leading edge technology, support services, software development, and computer R&D. The States and institutions in which the NSF centers are located have contributed about 35 percent of the expenses of the centers, and in addition the private sector has also contributed to the centers through direct funding and with in-kind contributions. Private firms are able to become partners with and use the centers' resources in return for their contribution. The national centers have attracted a user base exceeding that of the mission agency computational centers and including nearly every aspect of research, science, and education in U.S. universities.

The allocation of resources at these centers differs from that of the mission agency centers. The process of obtaining computing time at these centers is more open and competitive than at government-operated centers. The competitive process is aimed at fair allocation of the computing resources through a peer review process. Government subsidization of the operation of the computing centers has increased the use of computational resources, and increased the user base. For example, before the NSF national centers, there were only three or four places in the United States where high-performance computers were available if the research was not funded by mission agencies. Now, a growing number of States and universities operate computational centers to support research.

Some individuals have proposed that certain high-performance computing centers be assigned specialized missions. For instance, one center might emphasize biomedical research, or fluid dynamics; another, the responsibility for one of the other ''grand challenges,' such as global warming. NCAR is often used as an example of a successful discipline-oriented computational center to be used as a model for further specialization.

NCAR'S computational center is partially funded by the NSF, but its research is specific to its mission in atmospheric science, In this way, it differs from the other four national NSF centers. NCAR'S research includes climate, atmospheric chemistry, solar and solar-terrestrial physics, and mesoscale and microscale meteorology. The center houses a ccre staff of researchers and support personnel, yet its computational tools and human resources are available to the international atmospheric research community. Computer networks enable researchers around the Nation to access NCAR's facilities. NCAR, through its staff, research, hardware, and networks, has become a focal point for atmospheric research,

The advantage of a subject or discipline-specific computational center is that it focuses expertise and concentrates efforts on selected, important national problems. The staff is familiar with the type of work done within the disciplines and often knows the best ways to solve specific problems using computational science. Computers can be matched to fill the specific needs of the center rather than attempt to use a general purpose machine to serve (sometimes inadequately) the needs of diverse users. Experts in the field would have a central focus for meeting, comparing and debating research findings, and planning future research strategies much as atmospheric scientists now do at NCAR.

There are also disadvantages to discipline-specific centers. The ''general' high-performance computing centers are a focal point for bringing together diverse users and disciplines. Researchers, scientists, computer scientists and engineers, and software engineers and designers work collaboratively at these centers. This interdisciplinary atmosphere makes the centers a natural incubator for the advancement of computational science, which is an essential component of research, by fostering communication among experts in various fields. It is noteworthy that NCAR, a mission-specific center, has a general purpose supercomputer identical to that at the general high-performance computing centers (i.e., a Cray Y-MP). Moreover, many atmospheric scientists also compute at the other NSF supercomputing centers.

The NSF centers were established to foster research and educational activities so that academic research could keep up with the needs and progress of the Federal research laboratories, the U.S. industrial research and engineering community, and foreign competitors, but subsidizing a select group of centers may create an impression of "elitism' within the science and technology community. The current funding of NSF centers authorizes only four federally funded centers. There has been no open competition for other computational centers in the NSF process since the selection in 1983-84, so equity within the community is often questioned. But the centers' plans are reviewed annually, and a comprehensive review was undertaken in 1989-90 that culminated in the closure of the Princeton University center. Some nonfederally funded State and university centers question why these installations are perpetually entitled to government funds while others are closed out of the competition. [1]

NSF's subsidization of its centers tends to establish a hierarchy within the computational community. However, objective competition among the centers would be hard to referee since the measures for determining eminence in computation are imprecise and subjective at best. The government must be leery of creating proclaimed "leaders' in computational science, because it risks setting limits instead of pushing the frontiers of computing.

Provide Supercomputing Funds to Individual Research Projects and Investigators

The Federal Government could choose to support computational resources from the grass-roots user level instead of institutional grants. Federal science agencies could provide funds to researchers as part of their research grants to buy and pay for computer services. In this way, the government would indirectly support the operational costs of the centers. Capital improvement would likely still need support

[1] Gillespie, Folkner & Associates, Inc., ''Access to High-Performance Computer Resources for Research,'' contractor report prepared for the Office of Technology Assessment, Apr. 12, 1990, p. 36,

from the Federal Government because of the unpredictability of funding through user control and the need for long-term planning for maintaining and upgrading computer technology.

Some believe that funding the researcher directly for purchasing computer services would create competition among computational centers that could lead to improvements in the efficiency of the operation of computer centers and make them more responsive to the needs of the users. If scientists could choose where to ''purchase' supercomputing services, they would likely choose the center that provides the best value and customer service. Scientists could match the services they seek with the specialties of each center to meet their individual needs. Proponents of funding computer services through individual research grants believe that creating efficient, market-oriented computational centers should be a goal of the high-performance computing progam.

Centers vying for users might be captured by the largest users since they would have the most computing funds to spend. Well-funded users could force centers to cater to their needs at the expense of smaller users by the sheer purchasing power they represent. The needs of small users and new users could be slighted as centers compete for the support from big users. Competition among centers for users could have a downside if it should lead to isolation and lack of cooperation, and interfere with communication among the centers.

Upgrades and new machines involve large financial investments that user-derived finds may not be able to provide, The uncertainty of future funding in a competitive environment would make long-range planning difficult. High-performance computers generally must be upgraded about every 5 years because the technology becomes outdated and maintenance too costly. National centers aimed at maintaining leading edge technology must upgrade whenever state-of-the-art technology emerges. Therefore, supplemental funding would be required for capital outlays even if user funds were used to offset operational expenses.

Critics of direct funding of researchers for supercomputer time claim that the money set aside for supercomputing should be dedicated solely for that use. They believe that if researchers were given nonearmarked funds for computer services, they might use them instead to buy minisupercomputers or graphic workstations for themselves, or to fund graduate students. They believe that much of the money would never reach the supercomputing centers, leading to unstable and unpredictable budgets. Direct funding of researchers for computing time was tried in the 1970s, and led to many of the problems identified in the Lax report.

Proponents of user-controlled funding believe that researchers can best decide whether supercomputing is necessary or not for their projects, and if minisupercomputers would suffice, then perhaps that is the best option.

Provide Incentives for State/Private Institutions TO Supply Computational Services

Universities are heavily investing in information technologies and computational resources for the sciences. These non-Federal efforts should be encouraged. The government could provide matching funds to State and private institutions to contribute to the capital costs for computers and startup. Even a small amount of government seed money can help institutions leverage funds needed to establish a computing center. Supplemental assistance may be needed periodically for upgrading and maintaining up-to-date technology.

Some believe that temporary financial seeding of new centers is the best way for the Federal Government to subsidize supercomputing. Providing matching funds for several years to allow time for a center to become self-sufficient maybe the best strategy for the Federal Government to assist in achieving supercomputing excellence.

After the seed period expires, centers must eventually upgrade their machines. Without additional finds to purchase upgrades they might fall behind new centers that more recently purchased state-of-the-art technology. Should this happen, a number of computational centers might be created, but none of them may end up world-clam centers.

Expanding and Improving Usage

High-performance computers are general analytical tools that must be programmed to solve specific computational problems. Learning how to use the potential power of high-performance computers to solve specific problems is a major research effort itself. Research on how to apply high-performance computers to problems goes hand-in-hand with research on how to design the computers themselves. A Federal program to advance high-performance computing must strike a careful balance by supporting programs that advance the design of high-performance computers while at the same time advancing the science and engineering of computing for the R&D community.

It is important to distinguish *computational science* from *computer science and engineering.* Computer science is the science in which the object of intellectual curiosity is the computer itself, Computational science is the science in which the computer is used to explore other objects of intellectual curiosity. The latter discipline includes fields of basic research aimed at problems raised in the study of the computer and computing. They are not driven by specific applications. Although distinct, the two fields are closely related; researchers in each area depend on results and questions raised in the other.

Broader applications of computers often flow from advances made in research computing. Research in visualization, driven by the need to better understand the output of scientific calculations, has led to computer graphics technology that has revolutionized the movie and television industry and has provided new tools for doctors, engineers, architects, and others that work with images.

To advance the science of using high-performance computing, Federal programs must support five basic objectives:

1. **Expand** thc capabilities of human resources-Individuals educated, trained, or skilled in applying the power of high-performance computers to new problems in science and technology are in high demand. They are sought by businesses, industries, and an assortment of institutions for the skills they bring to solving complex problems. There is a shortage of scientists, engineers and technicians with such skills. A Federal high-performance computing initiative must ensure that the pipeline for delivering trained personnel remains full.
2. **Develop software and hardware resources and technologies-The** research and development of technologies that can be applied to major research problems—' grand challenges' —must continue. Special efforts are needed to ensure progress in the development of software in order to harness the power of high-performance computing for the solution of R&D problems.
3. Strengthen the scientific underpinnings of computation-This can be accomplished through the support of computer science and engineering as well as computational science.
4. **Construct a broadly accessible, high-speed advanced broadband** network-Such a network will provide the scientific and educational community with access to the facilities, the data, and the software needed to explore new applications.
5. Develop new algorithms for computational science----Algorithms are mathematical formulas used to instruct computers (part of computer programs and hardware). They are the basis for solving computational problems. New and better algorithms are needed to improve the performance of hardware and software in the computing environment.

Difficulties and Barriers

Computer and computational sciences compete with many other disciplines, for science funding. They are relatively young fields and are growing from a small funding base. Funding levels for computing research is relatively small compared with the more mature disciplines. Stimulating growth in computer and computational science encounters a ' 'chicken and egg' problem.

The size and level of activity of a research field is partially related to funds available. A Federal initiative designed to increase the research activity in computer and computational sciences must anticipate additional demands for Federal research funds. Furthermore, to maintain a healthy level of research activity, adequate funds to ensure future growth must be provided or talent will abandon the field to seek research money elsewhere. The small number of researchers working in computer and computational science may be cited as justification for *not* increasing levels of support, yet low levels of support limit the number of researchers and research positions.

Computational science is, in all but a few disciplines, a relatively new field. New researchers looking to establish their careers need assurance that their work will be recognized and accepted by their peers. Peer acceptance affects both their ability to obtain research funds and to publish articles in scientific journals. If computational methods are new to the field, the researcher may face a battle to gain acceptance within the traditional, conservative disciplines.

In many cases, researchers are in the early stages of understanding how to program radically new types of computers, such as massively parallel computers and neural nets. Researchers wishing to use such a computer need the assistance of those who can program and operate these computers for the duration of a project. There is currently a scarcity of such talent.

A NSF program dedicated to computational science and engineering may be needed. The program could find computational scientists from a cross section of traditional disciplines such as biology, chemistry, and physics. Funds for programs aimed at developing human resources, such as fellowships, young investigator grants, and so on, may also need to be earmarked for computational science. Direct funding for computational sciences would overcome the tendency of the disciplines to favor the funding of conventional research and their reluctance to try new methodologies,

Computational Centers

The most difficult issues, which programs in NSF's Advanced Scientific Computing Division are addressing, stem from the problems in putting leading edge technology in the hands of knowledgeable users who can explore and develop its potential.

In the mid-1980s, NSF formed five national supercomputer centers. Three of them-the University of California at San Diego, Pittsburgh, and University of Illinois at Champaign-Urbana-were based on Cray supercomputers. One, at Cornell University, installed modified IBM computers, and the Princeton Center was based on a machine to be built by ETA, a subsidiary of Control Data that has since gone out of business. Subsequently, NSF did not renew the Princeton Center for a second 5-year period.

There have been many changes in the high-performance computing environment since the establishment of those centers. These changes include: 1) the evolution of the mini-supercomputer, 2) the establishment of other State and institutional supercomputing centers, 3) the increase in use and interest in applications of high-performance computing to research, 4) the emergence of the Japanese as a force in the design, manufacturing, and use of high-performance computers, and 5) the emergence of a national network. Because of these changes—particularly in light of budget pressures and the high cost of the program-questions are being asked about the future directions of NSF support for these centers.

The basic conflict arises from several concerns:

1. the need for the NSF programs that support computational centers to determine what their ultimate goals should be in an environment where technological changes and user needs are constantly changing;
2. the need of computer centers and their researchers for stable, predictable, and long-term support in contrast to the reluctance of the government to establish permanent institutions that may make indefinite claims on Federal funding; and
3. the view that any distribution of NSF high-performance computing funds should be openly competitive and based on periodic peer review.

Purposes for Federal High-Performance Computing Programs

Leading Edge Facilities

Leading edge facilities provide supercomputing to academe and industry and provide facilities for testing and experimenting with new computers. Academics are provided an opportunity to train with leading edge technology; researchers and engineers learn about new computer technology.

A leading edge facility's responsibilities go beyond merely providing researchers access to CPUs (central processing units). Manufacturers of high-performance computers rely on these centers to test the limits of their equipment and contribute to the improvement of their machines. Leading edge technology, by its nature, is imperfect. Prototype machines and experimental architectures are provided a testbed at these centers. Scientists' experiences with the technology assist the manufacturers in perfecting new computing equipment. Bottlenecks, defects, and deficiencies are discovered through use at the centers. Moreover, user needs have led to the creation of new applications software, computer codes, and software tools for the computers. These needs have forced the centers to take the lead in software development.

Several computational centers have industrial programs with large corporate sponsors. These corporations benefit from leading edge computational centers in two ways. First, industry gains access to the basic research conducted at universities on supercomputers. Second, industry learns how to use leading edge computer technology. The support services of these facilities are available to corporate sponsors and are a major attraction for these corporations. Corporate researchers are trained and tutored by the centers' support staff, and work with experienced academic users. They gain a knowledge of supercomputing, and this experience is taken back to their corporations. Participating corporations often leave the programs when they gain sufficient knowledge to operate their own supercomputer centers.

A high-performance computing plan that establishes and maintains leading edge facilities benefits a broad range of national interests. Academics learn how to use the technology, manufacturers use their experiences to improve the technology, and industry gains an understanding of the value of supercomputing in the work place.

Increasing the Supply of Human Resources

An important aspect of any high-performance computing program is the development of human resources. National supercomputer centers can cultivate human resources in two ways. First, researchers and scientists are taught how to use high-performance computers, and new users and young scientists learn how to use modern scientific tools. Second, national centers provide an atmosphere for educating and cultivating future computer support personnel. Users, teachers, and technicians are critical to the future viability of supercomputing.

Producing proficient supercomputer users is an important goal of a high-performance computing program. Researchers with little or no experience must be trained in the use of the technologies. Education must begin at the graduate level, and work its way into undergraduate training. Bringing supercomputer usage into curricula will help familiarize students with these tools. The next generation of scientists, engineers, and researchers must become proficient with these machines to advance their careers. The need for competent users will increase as supercomputers proliferate into the industrial sector. Already there are reports of a shortage of supercomputer trained scientists and engineers.[2]

Support staff is an essential element of computational centers. The support services, which include seminars and consultation and support, educate the next generation of users. Support personnel are the trouble-shooters, locating and correcting problems, and optimizing computer codes. The NSF national centers have excellent staff, some of whom have moved to responsible positions at State and university-operated centers. The experience they gained at the NSF national centers contributes to the viability of new high-performance computing operations in industry and elsewhere in academe. The importance of the services that support personnel provide is often overlooked by policymakers, yet their contributions to supercomputing are invaluable. The greatest asset of a proficient high-performance computing center is the staff, not the computer. A high-performance computing program must emphasize the importance of developing human resources by producing educated users and users who will educate.

Advancing Computational Science

High-performance computer centers are a focal point for bringing together diverse users and disciplines. Researchers, scientists, computer scientists and engineers, and software engineers and designers work collaboratively at these centers. This interdisciplinary atmosphere makes the centers a natural incubator for the advancement of the computational sciences, which is an essential component of supercomputing. A national high-performance computing program could promote the computational sciences by fostering communication among experts in various fields.

Researchers and scientists know what questions to ask, but not necessarily how to instruct computers to answer them. Computational scientists know how to instruct computers. They create the computer instructions sets, computer codes, and algorithms for computers so that researchers can most efficiently utilize the technology. The development of computer codes and software is often a collaborative effort, supported by previous codes, software tools, and support staff, many of whom are computational scientists. Providing the methodology for utilizing

[2]Michael Schroeder, "How Supercomputers Can Be Super Savers, " *Business Week*, Oct. 8, 1990, p. 140

these tools is as important as providing the tools themselves.

Developing New Software Applications

New algorithms and codes must be developed to allow optimum use of supercomputer time. One of the more frequent criticisms of many high-performance computing operations has been the use of suboptimal codes. Supercomputer time is wasted when outdated or less than optimal codes are used. Creating codes is a specialty in itself. The development of codes is so labor intensive and time consuming that using an outdated code, as opposed to creating a new one, is sometimes more time efficient, although it may waste costly supercomputer time. A high-performance computing program could advance the usage of new and efficient codes by promoting computational science.

Providing Access to More Supercomputing CPUs

Supercomputing CPUs offer researchers computing power and speed unattainable from conventional mainframes. High-performance computer centers provide, at a minimum, access to supercomputing cycles. Supercomputing CPUs currently are a scarce resource in high demand. Any Federal high-performance computing program will increase the amount of supercomputing cycles available to researchers. It is uncertain, however, how much increase in CPUs the government should provide. Supercomputers are used in the advancement of all scientific disciplines, for both ‘ ‘big’ and ‘ ‘little’ science projects. All areas of research benefit from high-performance computing. Notwithstanding any reasonable level of effort, the government will be unable to provide enough supercomputing resources to meet all researchers’ needs. They will always seek more and faster supercomputing power.

Computer facilities whose main goal is to provide supercomputing CPUs are often called ‘‘cycle shops. ’ The NSF centers are *not* cycle shops. At cycle shops, support services are minimal: A skeletal support staff, enough personnel to keep the machines up and ruining, is all that is required. This limits cycle shops’ usefulness to primarily experienced users. Only proven technology can be used. Training, education, and software development are not major activities at such facilities. User applications have to be ‘canned’ and ready for use. These centers are the antithesis of leading edge facilities. Cycle shops are more economical for experienced users in need of large amounts of CPU time. This is not the majority of users, however.

Improving Data Storage Capabilities

Increasing importance is being placed on data storage capabilities. Researchers now realize the limits of current data storage technologies, A high-performance computing program can stimulate research in high-capacity storage and retrieval technologies.

Data storage technologies do not have the public appeal and visibility that supercomputers do. For this reason, they have been overlooked in supercomputing R&D, yet data storage is an integral part of high-performance computing, Supercomputers often use and produce large data sets. Computational centers are increasingly running into data memory and storage problems. New technologies for gathering data, e.g. satellites and automated sensors, are placing even greater demands on storage facilities. These data are often used in computing, and are converted into new data sets that require additional storage.

The Federal Government could take the initiative in R&D on new storage technologies, emphasizing its importance to high-performance computing. The amount of data handled at supercomputing centers will increase as the user base multiplies, and as data sharing increases through the use of high-capacity communications networks through the National Research and Education Network (NREN). Storage technologies are currently pushing their limits, and breakthroughs are needed if they are not to become the limiting factor in high-performance computing.