

**Report of the Task Force
on the
Future of the NSF Supercomputer Centers Program**

September 15, 1995

Arden L. Bement, Jr.
Purdue University

Edward F. Hayes -- Chairman
The Ohio State University

John Hennessy
Stanford University

John Ingram
Schlumberger, Austin

Nathaniel Pitts
National Science Foundation

Paul R. Young
National Science Foundation

Peter A. Kollman
UC San Francisco

Mary K. Vernon
University of Wisconsin

Andrew B. White, Jr.
Los Alamos National Laboratory

William A. Wulf
University of Virginia

Robert Voigt
National Science Foundation

Table of Contents

EXECUTIVE SUMMARY	VII
--------------------------	------------

1: INTRODUCTION	1
------------------------	----------

1.1: INTRODUCTION AND CHARGE TO THE TASK FORCE	1
1.2: BACKGROUND OF THE CENTERS PROGRAM	1
1.2.1: PHASE I – PURCHASING CYCLES	2
1.2.2: PHASE II – CENTERS ESTABLISHED	2
1.2.3: PHASE III – EVOLUTION IN MISSION	3
1.3: CONTRIBUTIONS OF CENTERS DURING PHASES II & III	4
1.4: BUDGET HISTORY OF THE CENTERS PROGRAM	5
1.5: REPORT OF THE BLUE RIBBON PANEL.....	6
1.6: REPORT OF THE NRC HPCC COMMITTEE	8
1.7: TAXONOMY OF COMPUTING RESOURCES	8

2: FACTORS	10
-------------------	-----------

2.1: DIRECT CENTERS PROGRAM STAFF INVOLVEMENT IN RESEARCH	10
2.2: GRANULARITY OF THE CENTERS PROGRAM USER-BASE	10
2.3: DISCIPLINE ORIENTATION OF THE CENTERS PROGRAM.....	11
2.4: LEVERAGE OF CORE CENTERS PROGRAM	11
2.5: CENTERS PROGRAM INVOLVEMENT IN INDUSTRIAL PARTNERSHIPS....	12
2.6: MULTIDISCIPLINARY ACTIVITIES	12
2.7: AVAILABILITY OF THE CENTERS PROGRAM RESOURCES.....	13
2.8: CENTERS PROGRAM INVOLVEMENT IN EDUCATION	13
2.9: CENTERS PROGRAM TIME-CONSTANT.....	14

3: ISSUES	15
------------------	-----------

3.1: SUNSETTING THE CENTERS PROGRAM.....	15
3.2: NEW COMPETITION OR RENEWAL	16
3.3: INDUSTRIAL PARTICIPATION AND USE OF THE CENTERS.....	17
3.4: EFFECT OF FREE CYCLES ON THE MARKET	19
3.5: THE TOTAL NEED FOR HIGH-END COMPUTING RESOURCES	21
3.6: MECHANISMS FOR JUDGING QUALITY OF THE SCIENCE	21
3.6.1: QUALITY OF RESEARCH PURSUED USING THE SUPERCOMPUTER CENTERS.....	22
3.6.2: METHOD OF ALLOCATING RESOURCES AT THE CENTERS	22
3.7: ALTERNATIVE FUNDING AND ALLOCATION MODELS.....	24
3.8: ROLE OF SMALLER CENTERS AND OTHER PARTNERSHIPS	25
3.9: TECHNOLOGY TRENDS	26
3.10: COMPUTER INDUSTRY TRENDS	26

3.11: APPROPRIATE ROLE OF THE CENTERS IN THE NII.....	28
3.12: CONTINUE TO SUPPORT TRADITIONAL VECTOR PROCESSING?	28
3.13: PARTNERING WITH OTHER FEDERAL AGENCIES	30
3.14: NON-COMPETITIVE FEDERAL FUNDING FOR SUPERCOMPUTER CENTERS	31
3.15: INTERNATIONAL COOPERATION	31
3.16: COMMERCIAL SUPPLIERS OF RESOURCES.....	32
4: FUTURE OPTIONS	33
OPTION A: LEADERSHIP CENTERS, SIMILAR TO THE CURRENT PROGRAM.	33
OPTION B: PARTNERSHIP CENTERS.....	34
OPTION C: SINGLE PARTNERSHIP CENTER.....	35
OPTION D: DISCIPLINARY CENTERS.....	36
OPTION E: TERMINATE THE PROGRAM.....	37
5: FUTURE DIRECTIONS AND PRIORITIES	39
5.1: RATIONALE FOR A CENTERS PROGRAM	39
5.2: PRIMARY ROLE OF THE CENTERS: FULL SERVICE ACCESS TO HIGH-END COMPUTING RESOURCES	39
5.2.1: ROLE OF RESEARCH PROGRAMS IN THE CENTERS	41
5.3: CONTEXT FOR THE NSF SUPERCOMPUTER CENTER PROGRAM	41
5.4: INTERACTIONS WITH INDUSTRY AND GOVERNMENT.....	43
5.4.1: VENDOR INTERACTION	43
5.4.2: OTHER INDUSTRIAL USERS AND TECHNOLOGY TRANSFER.....	43
5.4.3: INTERACTIONS WITH OTHER GOVERNMENT SUPPORTED CENTERS.....	44
5.5: THE ONGOING ROLE OF THE CENTERS	44
5.5.1: THE NEED FOR THE HIGH-END	44
5.5.2: TECHNOLOGY AND MARKET.....	45
5.5.3: MAINTAINING THE LEADING-EDGE CAPABILITY	45
6: RECOMMENDATIONS	47
6.1: CONTINUING NEED FOR THE CENTERS PROGRAM	47
6.2: SPECIFIC INFRASTRUCTURE CHARACTERISTICS FOR LEADING EDGE SITES.....	47
6.3: PARTNERING FOR A MORE EFFECTIVE NATIONAL INFRASTRUCTURE .	48
6.4: COMPETITION AND EVALUATION.....	49
6.5: SUPPORT OF RESEARCH AT THE CENTERS	50
6.6: ALLOCATION PROCESS FOR COMPUTER SERVICE UNITS	51
6.7: NSF LEADERSHIP IN INTERAGENCY PLANNING	51
<u>A: THE TASK FORCE ON THE FUTURE OF THE NSF SUPERCOMPUTER CENTERS PROGRAM</u>	<u>53</u>

A.1: CHARGE TO THE TASK FORCE	53
A.2: MEMBERSHIP OF THE TASK FORCE	55
<u>B: PAST AND FUTURE IMPACT OF COMPUTATIONAL SCIENCE AND ENGINEERING</u>	<u>57</u>
B.1: EXAMPLES OF PARADIGM SHIFTS	57
B.1.1: COSMOLOGY	57
B.1.2: ENVIRONMENTAL MODELLING	58
B.1.3: PROTEIN FOLDING	58
B.1.4: CONDENSED MATTER PHYSICS	59
B.1.5: QUANTUM CHROMODYNAMICS	59
B.1.6: DEVICE AND SEMICONDUCTOR PROCESS SIMULATION	60
B.1.7: SEISMOLOGY	61
B.1.8: TURBULENCE	62
B.2: DISTINCTIONS AND AWARDS ACCORDED USERS OF THE NSF SUPERCOMPUTING CENTERS.....	62
B.3: TESTIMONIALS FROM DISTINGUISHED SCIENTISTS	62
B.4: ADDITIONAL MATERIAL ON CENTERS' PROGRAM ACCOMPLISHMENTS	67
<u>C: BLUE RIBBON PANEL ON HIGH PERFORMANCE COMPUTING AND THE NSF RESPONSE</u>	<u>69</u>
<u>D: THE NATIONAL RESEARCH COUNCIL HPCC REPORT</u>	<u>73</u>
D.1: SUMMARY OF THE NRC COMMITTEE OBSERVATIONS AND RECOMMENDATIONS	73
D.2: RECOMMENDATION OF THE NRC COMMITTEE.....	73
<u>E: QUANTITATIVE DATA DESCRIBING THE NSF SUPERCOMPUTER CENTERS</u>	<u>75</u>
E.1: USAGE PATTERNS AT THE NSF SUPERCOMPUTER CENTERS	75
E.2: LONGITUDINAL ANALYSIS OF MAJOR PROJECTS	80
E.3: DURATION OF PROJECTS AT CENTERS	82
E.4: BUDGET INFORMATION ABOUT THE CENTERS.....	86
<u>F: EDUCATION AT THE NSF SUPERCOMPUTING CENTERS</u>	<u>89</u>
F.1: COMPUTATIONAL SCIENCE AND ENGINEERING - EDUCATION AND TRAINING.....	90
F.2: K-12 PROGRAMS	92
<u>G: WITNESSES AND OTHER SOURCES OF INFORMATION</u>	<u>95</u>

G.1: TASK FORCE WITNESSES AND INTERVIEWEES.....	95
H: SURVEY OF USERS	97
<hr/>	
H.1: BACKGROUND.....	97
H.2: MAJOR THEMES IN THE SURVEY RESPONSES TO THE OPEN-ENDED QUESTIONS:	97
H.3: SIGNIFICANT OPINIONS EXPRESSED BY FEWER RESPONDANTS.....	98
H.4: STATISTICAL RESULTS FROM THE SURVEY:	98
H.5: THE SURVEY INSTRUMENT.....	107

Tables and Figures

Figure 1.1: ASC program funding to Supercomputer Centers FY 1986-94.....	5
Figure 1.2: Other funding sources for the NSF SC Centers for FY 1986-94.....	6
Figure 3.1: Industrial support of the Centers program	18
Figure 3.2: Industrial support of NSF SC Centers separating base affiliate relationships and sales of computer time for FY 1986-94.....	19
Figure 3.3: Worldwide Distribution of Types of Top500 Computers.....	27
Figure 3.4: Increase in Normalized usage at Centers for both vector multiprocessors and parallel systems.	29
Figure 3.5: Distribution of Normalized usage at Centers between vector multiprocessors and parallel systems.....	30
Figure E.1 Usage Increase Over The History of the Program FY 1986-95.....	76
Figure E.2: Number of Users of the NSF SC Centers FY 1986-95	77
Figure E.3: Academic Status of Users and their Usage for FY 1994.....	78
Table E.1: Contents of the Annual Usage Reports of the National Science Foundations Supercomputer Centers	79
Figure E.4: FY 1994 Usage of NSF SC Centers by NSF Directorate and MPS Divisions	80
Figure E.5: NSF Average Funding History of 320 Selected PI's: FY 1984-93	81
Figure E.6: Reviewer's Scores of Selected PI's Proposals: FY 1984-94.....	82
Table E.2: The Top 50 Projects in FY 1992-95 in Order of their FY 1995 Usage Ranking	82
Table E.3: Ranking and NSF Funding of the Top 50 Project Leaders for FY 1992-95....	84
Table E.4: Composite Funding History FY 1985-94	87
Table F.1: Educational Impact at the NSF Supercomputing Centers.....	91

Executive Summary

The NSF supported Supercomputer Centers have played a major role in advancing science and engineering research. They have enabled collaboration among academic, industrial and government researchers on the solution of problems requiring demanding computational and visualization tools. In the 10 years of their existence, the Centers have fostered fundamental advances in our understanding of science and engineering, expanded the use of high-end computing in new disciplines, facilitated the major paradigm shift of the acceptance of computational science as a full partner in the scientific method, and facilitated the education of a new generation of computational scientists and engineers in support of that shift. These statements are documented in the body of the report as well as in its appendices.

Having accomplished so much in the last decade, it is natural to ask what the future role of the NSF should be in high-end computing. In October 1994, the National Science Board approved two-year continuation funding for the Supercomputer Centers. This provided time for the director of NSF to appoint this Task Force to analyze the alternatives. The possibilities considered include continuation, restructuring, or phasing-out of the current program, as well as creation of alternative models.

The Task Force believes that the future for computational science and engineering can be as bright or even brighter than in the past decade. If we seize the opportunity, over the next decade we can make major progress on multiple fronts.

There will be

- opportunities for exciting applications of our nation's exponentially increasing computational capacity, for example:
 - more complete models, and hence deeper understanding of physical systems by moving to three and higher dimensions;
 - progress in computational tools to aid drug and protein design;
 - computational predictions of scientifically and commercially significant materials;
 - multidisciplinary models of physical systems (e.g., combining fluid dynamics and electromagnetic models of the heart);
 - increased interconnectivity of supercomputers and high impact instrumentation; and
 - models of anatomical and physiological processes leading to new insights of benefit to human health.
- more quantitative computational results in unanticipated areas.
- more explosive growth of communications as a component of the computational science and engineering paradigm; and, importantly,
- continued progress in the tools and methods for developing code that is both portable and yet takes advantage of unique parallel architectures.

These advances will not automatically become available to American researchers. To position the U.S. academic community to participate in the exciting research possibilities

Executive Summary

enabled by these developments, the Task Force has the following recommendations leading to a restructured Centers program.

In order to maintain world leadership in computational science and engineering, NSF should continue to maintain a strong, viable Advanced Scientific Computing Centers program, whose mission is:

- ***providing access to high-end computing infrastructure for the academic scientific and engineering community;***
- ***partnering with universities, states, and industry to facilitate and enhance that access;***
- ***supporting the effective use of such infrastructure through training, consulting, and related support services;***
- ***being a vigorous early user of experimental and emerging high performance technologies that offer high potential for advancing computational science and engineering;***
- ***facilitating the development of the intellectual capital required to maintain world leadership.***

NSF should assure that the Centers program provides national “Leading-edge sites” that have a balanced set of high-end hardware capabilities, coupled with appropriate staff and software, needed for continued rapid advancement in computational science and engineering.

NSF, through its Centers program, should assure that each leading-edge site is partnered with experimental facilities at universities, NSF research centers, and/or national and regional high performance computing centers. Appropriate funding should be provided for the partnership sites.

NSF should announce a new competition of the High Performance Computing Centers program that would permit funding of selected sites for a period of five years. If regular reviews of the Program and the selected sites are favorable, it should be possible to extend initial awards for an additional five years without a full competition.

The Centers program should continue to support need-based research in support of the program’s mission, but should not provide direct support for independent research.

NSF should increase the involvement of NSF's directorates in the process of allocating service units at the Centers.

NSF should provide leadership in working toward the development of interagency plans for deploying balanced systems at the apex of the computational pyramid and ensuring access to these systems for academic researchers.

These recommendations are designed to set the Centers program on a new course that builds on its past successes, yet shifts the focus to the present realities of high-performance computing and communications, and provides flexibility to adapt to changing circumstances. It is our expectation, that at current NSF budget levels and absent new outside resources, there will be a reduction in the number of leading-edge sites to effect the benefits of the Task Force recommendations.

Executive Summary

In developing these recommendations, the Task Force obtained extensive input from academic, government, and industrial leaders; visited Centers and sought written input from the community. Some of this input is included as appendices, and the complete input is available on the Internet. The issues are complex and there are many strongly held opinions on the purpose, execution, and value of the program. The Task Force has tried to hear and understand all of the input, but in the end has, of necessity, formed its own judgment of what is best for the country. This report attempts to explain that judgment.

The report begins with a history of the Centers and how they fit into the nation's high performance computing infrastructure.

The second section attempts to identify factors the Task Force thinks are important in the evaluation process, including staff involvement in research, size of the user base, scientific discipline of the users, funding leverage, industrial partnerships, multidisciplinary activities, resource availability, and education.

The “hard issues” surrounding the Centers, particularly those not adequately discussed in previous reports, are discussed in the third section. This section examines such issues as: sunsetting the Centers; industrial use; effect of “free” computer cycles on the market; the total need for high-end computing; quality of the science; role of other centers; technology and computer industry trends; and the role of the Centers in the larger federal and international context.

Section four examines five options for a Centers program, ranging from the current system to termination of the program. Other options include partnership centers with stronger links between leadership centers and university or state facilities; a single partnership center; and disciplinary centers along the lines of the National Center for Atmospheric Research. The pros and cons of each option are discussed.

The fifth section of the report discusses future directions and priorities for the Centers program. The final section restates and explains each of the seven specific recommendations designed to support the Task Force vision for the future.

Executive Summary

Section 1: Introduction

1.1: Introduction and Charge to the Task Force

The director of the National Science Foundation established the “Task Force on the Future of the NSF Supercomputer Centers Program” in December 1994. Establishing the Task Force was one of the NSF administration’s responses to the resolution passed by the National Science Board at its October 1994 meeting that extended the cooperative agreements for the four NSF Supercomputer Centers by two years. During the period of the extension, the Foundation committed itself to explore thoroughly the needs for future infrastructure in high performance computing.

The Task Force was asked “. . .to analyze various alternatives for the continuation, restructuring, or phase-out of NSF’s current Supercomputer Centers program, or the development of similar future program(s), and to make recommendations among the alternatives.” Appendix A contains the complete charge to the Task Force and a listing of its members.

This report presents the Task Force’s analyses, findings, and recommendations on:

- The need for a federal government supported infrastructure
- The spectrum of options available for providing a computational infrastructure to support leading-edge academic research in computational science and engineering
- The factors, dimensions, and alternative models of possible infrastructures
- The Task Force’s preference among the alternatives
- A definition of the mission of the recommended program

The Task Force met from January 1995 through September 1995. Preliminary drafts of this report were circulated for comment. During that time, the Task Force members interviewed 30 academic, government, and industrial leaders, had visits to each NSF Center by one or more members, and sought written opinions from leaders of industry, senior government officials, representatives of state or regional centers, and knowledgeable members of the research community.

1.2: Background of the Centers Program

The NSF Supercomputer Centers program was established in 1984, following strong expressions of the need for such resources from the research community, and a study of the requirements set forth in a series of NSF Reports.¹ At that time, American researchers were at a serious disadvantage for gaining access to leading-edge high performance computers when compared to colleagues from other countries, or to domestic researchers whose research was supported by a United States mission agency (DoD, DoE, NASA). NSF leadership recognized that the lack of a suitable infrastructure was hampering important basic research and, with unprecedented support from Congress, moved promptly to establish the infrastructure.

¹ The need for Supercomputer Centers was discussed in the 1982 "Report on the Panel on Large Scale Computing in Science and Engineering," commonly referred to as the Lax report. The Curtis/Bardon report in 1983 outlined the plan of action to achieve the goals of the Lax report.

Section 1: Introduction

The situation in '82 (Lax, Bardon-Curtis, Fernbach reports)

- lack of academic access to high performance computers
- need for a *balanced* program
 - Supercomputer access, local computing,
 - training, networking, hardware, software,
 - and algorithms research

The NSF response

- Supercomputer program
- NSFnet
- computational science and engineering initiative
- expanded instrumentation and equipment programs

1.2.1: Phase I – Purchasing Cycles

A Program Office was set up in the Office of the Director of the National Science Foundation to purchase “cycles” from existing sources,² to distribute those cycles to NSF research directorates, and to oversee a program announcement soliciting proposals for the establishment of Centers. During this phase, over 5,000 hours of supercomputer time were made available to the research community, and more than 200 research projects at 80 institutions used the services.

While the announcement for establishing the initial program focused on providing supercomputer cycles, the original mission of the program was somewhat broader.

The major goals of this new Office [of Advanced Scientific Computing] include:

- Increasing access to advanced computing resources for NSF’s research community
- Promoting cooperation and sharing of computational resources among all members of the scientific and engineering community; and
- Developing an integrated scientific and engineering research community with respect to computing resources.³

In FY 1986, the Division of Advanced Scientific Computing, ASC, was formed within the newly created Computer and Information Sciences and Engineering Directorate, CISE, and the Supercomputer Centers program became a separate program activity within this division.

1.2.2: Phase II – Centers Established

Four Centers⁴ were established in 1985, and a fifth added in 1986, all providing “vector supercomputing services” for the research community and training for the many

²The original centers from which “cycles” were purchased were Purdue University, University of Minnesota, and Boeing Computing Services.

³ Memo from NSF Controller (Edward Hayes) to NSF Director (Edward Knapp), February 24, 1984, establishing the Office of Advanced Scientific Computing.

⁴ The original five centers were (1) The National Center for Supercomputing Applications (NCSA) at the University of Illinois, Urbana-Champaign, (2) The Cornell Theory Center (CTC), (3) The John von Neuman Center (JvNC), a consortium located at Princeton University, (4) The San Diego Supercomputer

Section 1: Introduction

researchers who lacked experience with these systems. These Centers were points of convergence where researchers learned to think in the new computational paradigm and to explore new vistas in resolution, accuracy, and parametric description of their problems.

Experiments in allocating resources, developing software support services, and starting standardized graphics and database descriptions to accelerate scientific visualization were all initiated during this phase.

Additionally, each Center established relationships with universities, both geographically close and far, to form consortia of members who had a stake in the resources of the Centers and in their future development. An important feature of these relationships was the formation of peer-review allocation boards, in which experts in computational science could direct attention to the performance of user's computer codes. Special attention was given to improvements of those codes with low performance. Direct interactions with experts at the Centers frequently facilitated significant performance improvements.

1.2.3: Phase III – Evolution in Mission

While the mission objectives of the Centers did not change significantly during their first five year period, there was directed evolution of the program's focus during the renewal period. Some of the changes were results of pressure from ASC management while others came from various advisory committees, in particular a Program Advisory Committee (PAC) that preceded the CISE Advisory Committee and the Program Plan Review Panel (PPRP).⁵

The Phase III renewal process resulted in the decision to close one of the original five Centers – The John von Neuman Center (JvNC). NSF had encouraged the original *high risk* plan of the JvNC to use the newly developed ETA computer, established by an offshoot of Control Data Corporation (CDC), the original modern supercomputer vendor. However, when ETA failed as an entity, there was a thorough review of JvNC and its future role in the program. When the review process identified major concerns, JvNC was not renewed.

There was a distinct effort at this time to expand outreach services, with initial efforts intended to forge closer ties to industries that could profit from exposure to high performance computing, and to include the community at large. Efforts to introduce tools to enable convenient access to Centers from the popular microcomputers and workstations led to such software development projects as NCSA Telnet and the Programmer's Workbench from the Cornell Theory Center, CTC. Each Center started to explore parallel computing, originally on their vector multiprocessors, later by adding new scalable parallel architectures to their stable of allocable computers.

Center (SDSC), located at the University of California, San Diego and operated by General Atomics, and (5) The Pittsburgh Supercomputing Center (PSC), directed by the University of Pittsburgh and Carnegie Mellon University and operated by Westinghouse, was added in 1986.

⁵ PPRP is the oversight panel appointed by the Foundation which meets annually to review the Centers plans and to make recommendations among competitive funding requests.

Section 1: Introduction

Adding parallel systems opened the doors to a new range of vendors that had not participated in the Phase II program. Each Center started to undertake a distinct set of activities and this difference in appearance threatened to fragment the program. However, during this period, NSF and its advisory committees stressed the need for changes in the coordination of the program. In response, the Centers formed themselves into a “MetaCenter”, with resources shareable on a national scale. NSF and the PPRP encouraged these cooperative activities to strengthen the program and remove the need to duplicate resources at four locations. The Center staffs started regular meetings and cooperative projects in networking, mass storage, outreach, etc. A joint brochure was prepared, and joint allocation procedures started. This MetaCenter model has generally been viewed as successful, and other agencies are now using it as a starting point for several new high performance computing initiatives.⁶

An important accomplishment of the Centers was an initiative, undertaken by all Centers to varying degrees, into scientific visualization, and on providing tools and standards for data exchange. An especially visible example of these activities is NCSA Mosaic, a “browser” for multi-media information using the protocols of the World-Wide-Web, initially developed at CERN. Mosaic, and its licensees and spin-offs, greatly expanded interest in the Internet, and networked information in general.

1.3: Contributions of Centers during Phases II & III

As the original program was defined, the Centers would be judged on the quality of the science and engineering research conducted (by other researchers) using the Centers’ resources. An assessment of the major research accomplishments during Phases II & III was prepared for the National Science Board in October 1994, and is available on the World Wide Web,⁷ a list of accomplishments that grows annually. Advances in cosmology and materials science brought about by researchers using the NSF Centers has been particularly noteworthy. Recent advances in computational physics have arguably pointed to this period as one of the most productive in modern physics. Computational biology, unknown a decade ago, is one of the most rapidly growing segments of the biological sciences. In addition, engineering has been increasing its share of Centers’ usage. Interestingly, in engineering not only is overall usage increasing, but the numbers of new users is also increasing, perhaps signaling increased future use of high performance computing in engineering research programs. Finally, use from the NSF Geosciences directorate has been high even with the excellent resources available to the atmospheric sciences community at the National Center for Atmospheric Research (NCAR), the fifth member of the NSF MetaCenter.

It is also the case that the Centers have benefited from strong and visionary leadership. Beyond the accomplishments achieved by the “users” of the Centers, many were initiated by the Centers. Appendix B has pointers to many of the overall scientific and

⁶ Aspects of the DoD HPC Modernization Program and the DoE ASCI program are planned around cooperative interaction between existing centers and laboratories. The program officers have informed NSF ASC division that they are basing these cooperative interactions on the NSF MetaCenter.

⁷ URL: <http://www.cise.nsf.gov/acir/hpc/>

Section 1: Introduction

technological accomplishments, as well as descriptions of some paradigm shifting applications and testimonials from senior scientists.

1.4: Budget History of the Centers Program

The program was started with a request for \$20M in FY 1985, which Congress increased to \$41.46M. This was aimed specifically at accelerating the inception of the NSF Centers, increasing their number, and appropriating the costs of the transfer of a computer from NASA to the NSF program.

The program planned to establish four National Centers with the initial FY 1985 allocation, but ultimately achieved five Centers before the end of FY 1986. Although funding growth following the initial FY 1985 appropriation came within the framework of overall NSF budget increases, the cooperative agreements for the Centers showed quite rapid growth until FY 1990-91. Growth has been more modest in recent years. In addition to the cooperative agreements, the ASC program provides support for special projects at the Centers reviewed by the PPRP, and a program called MetaCenter Regional Alliances, to assist researchers with complementary goals to establish closer links to the four major centers.

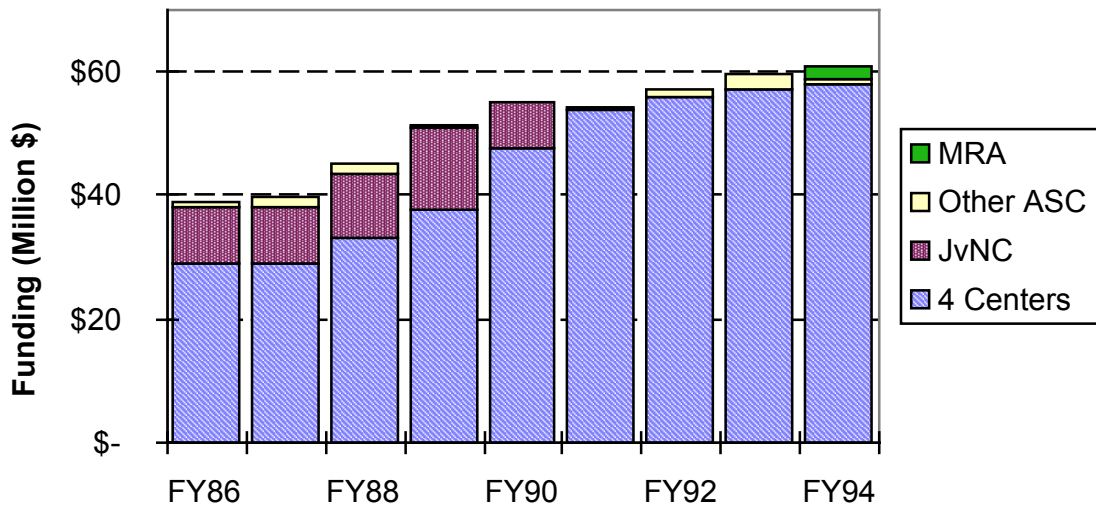


Figure 1.1: ASC program funding to Supercomputer Centers FY 1986-94⁸

While the NSF contribution through the base cooperative agreement has leveled off, the overall budgets have continued to grow. Other sources of funding (beyond the base cooperative agreements) have generally increased, as shown in Fig. 1.2. It is evident that these funding sources vary greatly from year to year, but the greatest contributor (after NSF) have been the vendors themselves, who have provided from 34% to 57% of the NSF contribution as discounts, equipment and software support, and vendor personnel

⁸ Funding at the John von Neuman Center is shown separately for FY 1986-90. Other ASC list all discretionary funds added to the base cooperative agreements. MRA indicates the MetaCenter Regional Alliances, a program started in FY 1994.

Section 1: Introduction

assigned to the Centers.

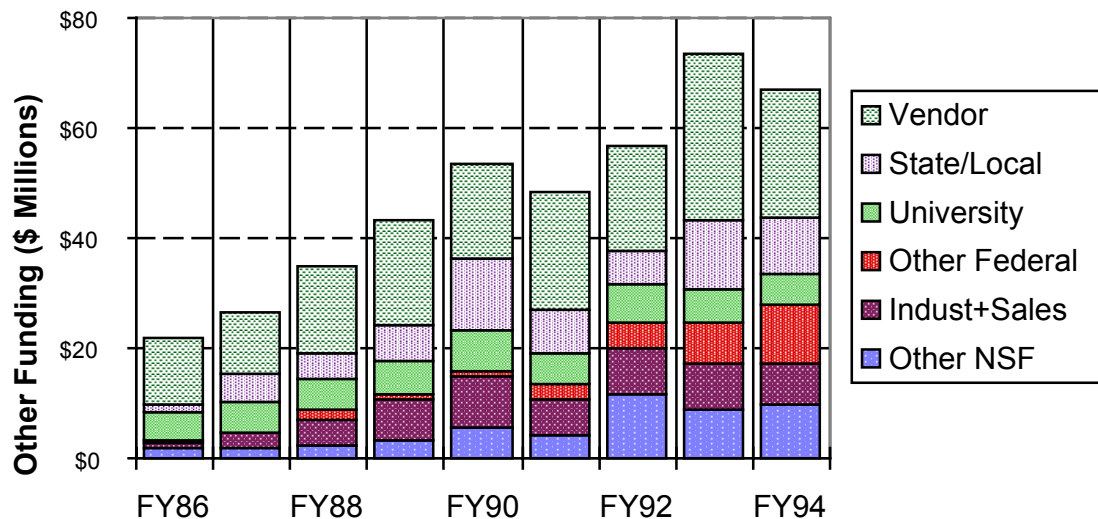


Figure 1.2: Other funding sources for the NSF SC Centers for FY 1986-94

In summary, the Centers have successfully attracted funds from a variety of sources to build a funding base for the Centers at about twice the NSF cooperative agreement levels. These extra funds maintain a core competency of personnel and hardware available for the research community as needed, and have been the underpinnings for the outreach programs of the Centers.

1.5: Report of the Blue Ribbon Panel

The current Task Force report is the latest of many studies of the Centers. Following the renewal of four of the Centers in 1990, the National Science Board (NSB) asked the director of NSF to appoint a blue ribbon panel

“... to investigate the future changes in the overall scientific environment due [to] the rapid advances occurring in the field of computers and scientific computing.” The resulting report, “From Desktop to Teraflop: Exploiting the U.S. Lead in High Performance Computing,” was presented to the NSB in October, 1993.

This report, which is discussed more extensively in Appendix C, points to the Foundation’s accomplishments in the seven years since the initial implementation of the recommendations of the Lax Report on high performance computing (HPC) and the establishment of the Supercomputer Centers. The report asserts that the NSF Centers have created an enthusiastic and demanding set of sophisticated users who make fundamental advancements in their scientific and engineering disciplines through the application of rapidly evolving high performance computing technology. Other measures of success cited include the thousands of researchers and engineers who have gained experience in HPC, and the extraordinary technical progress in realizing new computing environments. Some of these achievements are highlighted in Appendix E of the Blue Ribbon Panel Report.

Section 1: Introduction

The report notes that, through the NSF program and those of sister agencies, the U.S. enjoys a substantial lead in computational science and in the emerging, enabling technologies. It calls for NSF to capitalize on this lead, which not only offers scientific preeminence, but also aids the associated industrial lead in many growing world markets.

Addressing the opportunities brought about by the success of the program, the report puts forth a number of challenges and recommendations which are summarized in Appendix C. These recommendations were based on an environment with the following two characteristics, which have since changed:

- Parallel systems were just being introduced at the Centers. Because uncertainties surrounding systems software and architectural issues made it unclear how useful these systems would be for scientific computing, the report recommended investment in both the computational science and on the underlying computer science issues in massively parallel computing.⁹
- The report assumed that the administration and Congress would adhere to the stated plan of the HPCC and NSF budgets, which called for a doubling in five years.

Primary recommendations included the following:

- The NSF should retain the Centers and reaffirm their mission with an understanding that they now participate in a much richer computational infrastructure than existed at their formation.
- The NSF should assist the university community in acquiring mid-range systems to support scientific and engineering computing and to break down the software barriers associated with massively parallel systems.
- The NSF should initiate an interagency plan to provide a balanced teraflop system, with appropriate software and computational tools, at the apex of the computational pyramid.

These recommendations and the accompanying challenges could be summarized as calling for a broad based infrastructure and research program that would not only support the range of computational needs required by the existing user base, but would also broaden that base in terms of the range of capabilities, expertise, and disciplines supported.

As a follow up to the Blue Ribbon Panel Report, in 1993 the NSF director established an internal NSF High Performance Computing and Communications Planning Committee. In responding to the panel report, the committee was charged with establishing a road map and implementation plan for NSF participation in, and support of, the future HPC environment. The internal committee presented a draft of its report to the Director's Policy Group in March, 1994; a final version of the report was made available to the NSF in February, 1995.

The committee used the challenges of the panel report but, since at this point it had become clear that the budget would not be doubling, the committee used more realistic

⁹ At the time of this writing, as documented by the NRC-HPCC report, much progress has been made in understanding the use of parallel computers, but significant additional research, particularly on software (and algorithmic) issues, is still needed.

Section 1: Introduction

budget assumptions for its own report. The recommendations contained in the committee report were consistent with and supportive of the recommendations in the panel report; there were no major areas of disagreement. Both reports called for renewal of the current Centers without recompetition.

Additional recommendations in both reports called for a balanced approach to computing infrastructure ranging from workstations through access to the most powerful systems commercially available. (The so-called “Pyramid of Computational Capability”) The Supercomputer Centers were viewed as an essential ingredient in this infrastructure with continually evolving missions. Both reports also acknowledged the need for strong, continued support of research on enabling technologies such as algorithms, operating systems, and programming environments.

1.6: Report of the NRC-HPCC Committee

In 1994, Congress asked the National Research Council (NRC) to examine the status of the High Performance Computing and Communications Initiative. Deferring to the current Task Force, the NRC committee did not make explicit recommendations for funding levels or management structures for the Supercomputer Centers program. The committee did say:

"The committee recognizes that advanced computation is an important tool for scientists and engineers and that support for adequate computer access must be a part of the NSF research program in all disciplines. The committee also sees value in providing large-scale, centralized computing, storage, and visualization resources that can provide unique capabilities. How such resources should be funded and what the long term role of the Centers should be with respect to both new and maturing computing architectures are critical questions that NSF should reexamine in detail, perhaps via the newly announced Ad Hoc Task Force on the Future of the NSF Supercomputer Centers Program."

The other major point raised in the NRC report was on the use of HPCC funds for supporting computing on mature vector architectures. The NRC report recommends that HPCC funding be used exclusively for exploring new parallel architectures, rather than for supporting use of stable technologies. These issues are discussed further in Section 3 and in Appendix D.

1.7: Taxonomy of Computing Resources

The full range of the computational resource needs of the scientific and engineering research community vary widely. Not all of the needed resources can or should be provided by the Centers program. Some of the most important needs that the Task Force has identified are:

- Access to computing resources on different scale systems: workstations (desk top), mid-range, large centrally managed systems (state, regional, university), or highest-end systems (national).
- Access to computer resources on the highest-end systems: processing speed, memory size, mass storage, I/O bandwidth, internode communication bandwidth, network bandwidth.

Section 1: Introduction

- Access to more general resources: diversity of architectures, visualization, information processing, consulting, third party software, research teaming, code porting, and training.

The Task Force has found the following taxonomy helpful in characterizing the level of computing resource, where the dollar amounts are meant to represent the overall annual cost of the resource.

Level 1 -Workstations < \$100K

Level 2 -Mid Range ~ Departmental or interdisciplinary groups ~ \$100K to \$2M

Level 3 -State, regional, and university centers ~ \$2M to \$10M

Level 4 -National leadership centers > \$10M

Section 2: Factors

This section discusses nine factors that the Task Force thinks are useful in understanding and evaluating the mission, the issues, and the various options possible for defining the future program of the NSF Supercomputer Centers. Each factor is presented as a one-dimensional continuum described briefly in words, with the extremes marked below the description on a scale from left to right. The marker between the end points is an estimate of where the Task Force thinks the Centers program is at present. This description is followed by a discussion of where the Task Force believes that the Centers program should reside. These factors serve as yardsticks by which to measure and discuss the Centers program. While each of the ideas represented by these factors appears elsewhere in this report, the presentation we give here attempts to give a different focus on the basic elements of the Centers program.

The Task Force recognizes that taken alone or out-of-context these factors may be vague or misleading. Taken together, the factors provide a useful characterization of the Centers program and of its potential alternatives. Note that when we speak of the Centers program, we understand that this may include, in certain cases, only those activities supported significantly by the basic Centers cooperative agreements, while in other cases our discussion may include all aspects of the Centers program, and in particular the MetaCenter Regional Alliances. While this section focuses on the major factors, the next section devotes itself to the more controversial issues that have arisen during the program's life.

2.1: Direct Centers Program Staff Involvement in Research

Pure service <-----x-----> Pure research

This dimension measures the direct involvement of the Centers program personnel in research activities. By "direct" involvement, we mean that a staff member is a significant intellectual resource in a research effort, as opposed to primarily providing a service to one or many research efforts.

The early history of the Centers program and a superficial understanding of the mission might lead to the belief that "pure service" is the proper role for Center staff. However, the Task Force believes that the Center staff members must remain intellectually involved in research if they are to provide the best service in enabling world-class science and engineering. This is particularly true in the rapidly changing technological landscape in which the users find themselves. In fact, the Centers' staff help form this technology landscape, and thereby indirectly helps set the scientific and engineering research agenda. We return to this issue at several points in the report and address it specifically in the recommendations.

2.2: Granularity of the Centers Program User-base

Small <-----x-----> Large

If the Centers provide significant resources only for a few Grand Challenge investigators, then the granularity would be large. The Centers program will support both relatively large and relatively small consumers of resources and, by necessity, there are few

Section 2: Factors

relatively large users. Thus, this dimension measures the magnitude of the total computing resource that is allocated to the largest users. For example, we see in the usage data over the last five years of the Centers, that the large-user population and the resources allocated to them has remained relatively constant with about 80% of the cycles going to about 8% of the users. The total number of users has remained relatively constant,¹⁰ although the individuals vary significantly from year to year.

The Task Force believes that the mission of the Centers program should remain focused at the high end of computational science and engineering and, therefore, the granularity should be large. The support for entry-level applications should be met in other ways where possible, perhaps on smaller configurations distributed at regional, state, or local centers. The value of the Centers program is not in providing the most cost-effective cycles, but to enable the paradigm-shifting applications and technologies that occur at the high end of the spectrum. Further discussions of the allocation model and process appear in Sections 3 and 5 and in the recommendations.

2.3: Discipline Orientation of the Centers Program

Aligned <-----x--> **General**

The Task Force assumes that the Centers program should support all disciplines appropriately. However, this dimension characterizes the orientation or organization of individual Centers in the Program relative to a full complement of academic disciplines.

An example of a completely discipline-oriented program would be the National Center for Atmospheric Research (NCAR). While any measure of the NSF Centers program activities by discipline would vary from year to year, this variability should be the result of proposal pressure (e.g. some disciplines require more resources of a certain type than others) and explicit transient decisions made by the program management. Not a single one of the witnesses interviewed by the Task Force, including the NSF assistant directors, believed that the Centers should be organized as discipline-specific centers. On the basis of considerable, and unanimous, testimony, the Task Force believes that the benefits accrued from multidisciplinary activities, the resulting cross fertilization of ideas, and the leveraging of resources, far outweighs the advantage of having a single-community orientation for the centers within the program. We return to this topic in discussing the options for a future Centers program in Section 4.

2. 4: Leverage of Core Centers Program

NSF/CISE <-----x-----> **Other**

This dimension represents the percentage of funding that NSF/CISE expects to provide for the Centers program and, by the same token, the percentage of influence that NSF/CISE wishes to have on the Centers program.

The Task Force believes that supplemental funding must not significantly divert attention from the mission of the Centers program. Experience over the last ten years suggests that

¹⁰ It is interesting to note, however, that the number of users receiving more that 1000 normalized service units has increased by more than an order of magnitude over this same time period. This is primarily due to the rapidly increasing capabilities of the technology

Section 2: Factors

a minimum of 50% of the total funding, including extra vendor discounts, should be from NSF/CISE. A Task Force subcommittee met with representatives from NIH, ARPA, DoD, and DoE to specifically discuss the Centers program. During these discussions it became clear that it may, in fact, be very difficult for the Centers program to achieve the same leverage, especially in cash contributions, that it has in the past.

2.5: Centers Program Involvement in Industrial Partnerships

0% <-----x-----> 100%

This dimension characterizes the importance of partnership with industry. There are two types of industries that the Centers might collaborate with: the supply-side or technology (primarily vendors) and the demand-side or applications industries (primarily users of high performance computing). Industrial partnerships with industries that use high performance computing are transitioning from the provision of cycles by the Centers to a focus on training and access to new and more experimental software and hardware.

The Task Force believes that the Centers program should have as close a partnership with the technology industries as is necessary to fulfill its primary mission that focuses on supporting academic usage of supercomputing. In the current technological landscape, the Centers program should be tightly coupled with those vendors that most affect its ability to carry out its mission. This partnership should provide information on user requirements and feedback on performance to the vendors. However, vendor partnerships are collateral to the main mission of the Centers program and not of primary importance in and of themselves.

Similarly, the Task Force notes that outreach to emerging industrial users and interaction with industrial customers is a secondary, albeit important, component of the research activities of the Centers program. While such relationships may include provision of cycles to industrial users, the Task Force believes that industrial relationships that involve understanding the use of high performance computing in industry are probably more beneficial to both the Centers program and to industry. As with the vendor relationships, such industrial partnerships remain secondary to the primary mission.

We discuss the historical interactions of the Centers with industry and the appropriate role in the next section on issues.

2.6: Multidisciplinary Activities

0% <-----x>>>>-----> 100%

This dimension characterizes the magnitude of the Centers program's support of multidisciplinary programs.

The Centers program has been a major catalyst of multidisciplinary activities. The Task Force believes that the Centers program should continue to support both disciplinary and multidisciplinary activities as is appropriate to its mission. However, the Task Force believes that as the complexity of problems increases, the emphasis will gradually shift to more support of multidisciplinary activities. Moreover, this is the direction that some of the most outstanding young students and faculty appear to be moving. This is a special and important role of the Centers and should be encouraged. Many Task Force witnesses

Section 2: Factors

testified to the important role the current Centers play as a catalyst for interdisciplinary interactions.

2.7: Availability of the Centers Program Resources

Common <-----x-----> **Unique**

This dimension characterizes the extent to which the resources and activities of the Centers program are available to the general academic community from other sources.

The Task Force believes that the main focus of the resources and activities of the Centers program should be special, at least in so far as the Foundation's scientific and engineering community is concerned. When the Centers were founded, although there were significant computational resources available in the weapons and intelligence communities, the Centers program provided a unique resource for the academic community. Taken in light of its user-base, the Task Force believes that it is unlikely that the Centers program will find its mission invalidated anytime in the near future by available time on supercomputers from other sources. However a re-evaluation of this should be part of overall periodic evaluations of the entire program.

The motivations for focusing on the high-end and providing resources not available elsewhere is discussed as an issue in the next section and as a key factor in determining future directions for the Centers in Section 5.

2.8: Centers Program Involvement in Education

General <-----x----> **Targeted**

This dimension characterizes the Centers program's support of education, training, and knowledge transfer.

The Task Force believes that the Centers program has an important role in education and training, within its primary mission of enabling world-class science and engineering research, by providing high-end computing resources. Associated with this goal of supporting access to high-end resources is an education mission that naturally focuses on a more advanced student population for which supercomputing is an appropriate and valuable tool. Historically, the education and training component of the Centers program has been focused at the advanced undergraduate level and above. The Task Force believes that this is the proper focus for the future as well. At the same time, the Task Force recognizes the value of the efforts that have been targeted at teachers and students at grades 6-14. Such activities should be continued in the future at comparable levels.

The Task Force believes that some alternatives to the current program discussed in the Options section could strengthen the education component of the Centers program by enlarging the base of students that have access to computing facilities beyond what might be available in their own laboratory or university. While the task force believes that the education mission should stay primarily aimed at the high-end access that the Centers program enables, broadening the educational impact would be valuable. The educational impact of the Centers and the future educational role are discussed in more detail in Sections 3 and 5 and in the Recommendations.

Section 2: Factors

2.9: Centers Program Time-constant

Short <-----x-----> **Long**

This dimension characterizes the stability of major Centers program activities.

The Task Force recognizes the natural tension between stability and competition as regards major activities (e.g. Supercomputer Centers) of the Centers program. Stability is important to build up expertise and to provide users with a sense that the Centers program will continue to support efforts in accord with the main mission. Competition is important to maintain vitality and to provide the community with the very best resources available. The Task Force believes that each of these components should be made an explicit part of the Centers program. To deny a role to either would damage the Centers program as a whole. Some of the options discussed in Section 4 may improve the ability to recompute portions of the program without dramatically reducing the stability of the overall program.

Section 3: Issues

Over the life of the Supercomputer Centers program a number of issues have been raised repeatedly; and in some cases have not been adequately addressed in reports about the program. Several of these issues relate to the factors discussed in the previous section. Others represent areas of controversy. The purpose of this section is to address these issues, although not necessarily to reach a definitive conclusion with respect to each of them.

No one of these issues is so important that by itself it should determine the future of the program. Thus there is some danger in treating the issues separately. At the same time, given the complexity and diversity of the issues involved, the Task Force has found it useful to examine them separately. We will try to provide enough of the arguments for each side of these issues to help establish the basis for our final recommendations. While we began our discussion of these issues with the goal of presenting a balanced view of each issue, after we had obtained input from a broad range of perspectives and had completed our discussions of each issue, we decided that the discussion presented in the report would provide a better basis for understanding some of the Task Force's conclusions regarding these issues if we included our own conclusions in the presentation of the issue.

This section not only lays out a number of issues for the reader's consideration, but also often lays out the Task Force's best judgment of these issues based both on the full report and on the committee's overall deliberations.

3.1: Sunsetting the Centers Program

This issue is sometimes stated in an aggressive form as: "since other NSF centers like the Engineering Research Centers (ERC) and Science and Technology Centers (STC) are sunset, the supercomputer centers should be too." At least in part, this form of the question arises from a confusion between research centers and facilities. Although there are few "pure" examples of either type, for the present purposes it is useful to represent the two extremes of this dichotomy.

NSF has a clear policy with respect to pure research centers. They are reviewed at specified intervals and possibly renewed, but eventually they are sunset; the provisions for sunsetting are normally built into the original program plan. Of course, the fact that such a center is terminated does not preclude a new proposal from the same group. Likewise, NSF has a relatively clear policy for pure facilities, they are reviewed, and management of them may be "re-competed" (as in the recent case of the high magnetic field laboratory), but they are not sunset. The need for the *facility* does not go away, although its location or management can be changed.

Like many NSF facilities, the Supercomputer Centers fit neither of the pure models. In particular, although initially created in the facilities model to provide service to other researchers, the Centers have evolved to include a research component. To keep Center staff at the forefront of the technology, it is necessary for the Centers to have a research component, in effect, to participate in the development of the relevant technology. It is important to note that the vast majority of the funding for this research comes from

Section 3: Issues

competitively awarded grants and contracts, not from the base cooperative agreement for the Centers.

The original rationale for the Supercomputer Centers was that there were important scientific problems whose solution required access to the highest performance computers possible, that academic researchers did not have access to such resources, and that these resources were so expensive that the only alternative was to share facilities at a few national centers. At the time it was impossible to predict that any of these premises would become invalid at a specified time, and so no sunset provision was built into the program.

The Task Force believes that the first of these three premises is still true. There are important scientific, technological and societal problems whose solution requires the highest performance computation. Further, the rationale to pool the *highest* performance resources is also valid. Thus, at the present time, these premises still argue for continuation of a Centers program focused on the high-end program.

The second premise is more questionable. Clearly, academics now have access to high performance computing through a number of sources (including the NSF sponsored Centers). Moreover, the advent of scalable architectures and increasingly capable networking makes it feasible to deliver significant computing power for some problems to the individual researcher or research group locally. Hence, the argument for complete centralization is somewhat weakened.

But it should also be noted that over its 10 year life, the program has evolved to include more than just access to “fast” computation.

- Large memories and large archival storage are also crucial to an increasing number of research efforts. To support these efforts requires an appropriate aggregation of facilities.
- Reflecting the need to help make the emerging technology more usable by the computational science community, in the 1988 five-year renewal, NSF explicitly broadened the mandate of the Centers to include research activities aimed at this objective.
- Developing software for use by the computational science and engineering community, education and training in the new technologies, and leadership among the state and regional centers are all critical roles of the existing Centers that cannot be replaced merely by distributing smaller machines.

It is hard to see how these roles would be filled effectively without some number of national centers. Further discussion of the ongoing need for providing access to high-end computing resources appears in Section 5 and is a key focus of the Task Force’s recommendations.

3.2: New Competition or Renewal

While the existence of national centers for supercomputing can be justified by the need for their services, the question of the frequency of competition of such centers is often raised. In general, competition increases effectiveness and allows for flexibility.

Section 3: Issues

Furthermore, the facilities provided by the supercomputer Centers, quite unlike telescopes, become obsolete quickly and need to be replaced. In principle at least, the location of a Center could be moved easily.

The corresponding argument favoring renewal is that the “soft infrastructure” of the Center, its staff, cannot be easily replicated or moved. Since much of the value of the Centers is in their soft infrastructure, too frequent competitions could seriously disrupt the effectiveness of the program.

On balance, the Task Force sees value in both sides of this issue and is inclined to believe that infrequent competition with more periodic review and potential renewal is the best approach. Some of the options discussed in section 4 should improve the potential for recompeting portions of the program with less disruption. The recommendations also discuss schedules that might be used for renewal, new competition, and continuance of the program

3.3: Industrial Participation and Use of the Centers

There are three kinds of industrial funding that have been significant to the Centers program: funding from computer vendors, particularly suppliers of high performance hardware and software; funding from industrial users who wish to make use of the unique computational capabilities of the Centers; and funding from industries interested in technology transfer and training. In the early development of the supercomputer Centers there was significant industrial use; while the gross level of industrial usage has increased, rates for computer usage have fallen more rapidly, so that industrial revenue has declined. Some have interpreted this as a failure of the program.

Computer vendors have made contributions to the Centers program, occasionally in cash, but most often with in-kind contributions, including very substantial discounts on equipment and on-site personnel to interact with Centers staff and some users. In return, vendors get valuable feedback that assists them in their own strategic planning. Over the history of the Centers program these relationships have improved the research infrastructure available to academic researchers without any obvious negative consequences. One contributing factor has been that each of the four Centers has had a different set of participating vendors and the relationship has not been an exclusive one.

Section 3: Issues

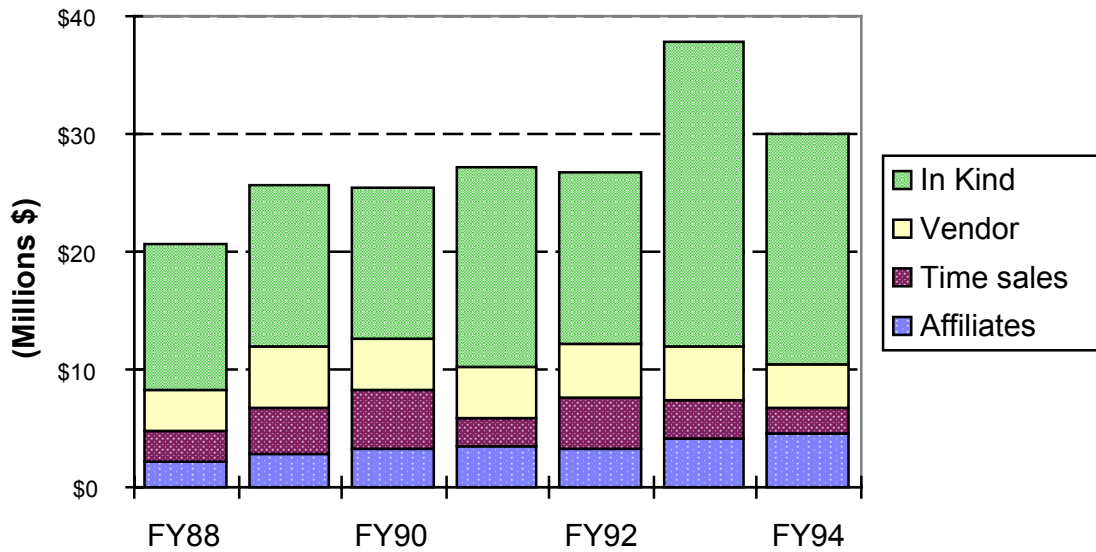


Figure 3.1: Industrial support of the Centers program.¹¹

Another major component of industrial support is from users of the technology. Such support includes: (1) fees for affiliates programs, mostly at NCSA and SDSC, and, (2) fees for use of Center resources. The affiliate support has continued to grow modestly, rising to a current level of about 10% of the cooperative agreement amounts, or 5% of the total budgets (see Fig. 3.2), while the usage fees have fallen from their peak in the late 80's. Thus, while total industrial revenues of the Centers has declined, one can infer that interest in the technology transfer portions of the program has remained strong.

¹¹ In kind support is reported by the Centers in dollars and is made up of three components, discounts beyond the normal educational discounts, vendor staff stationed at the Centers, and donated hardware and software.

Section 3: Issues

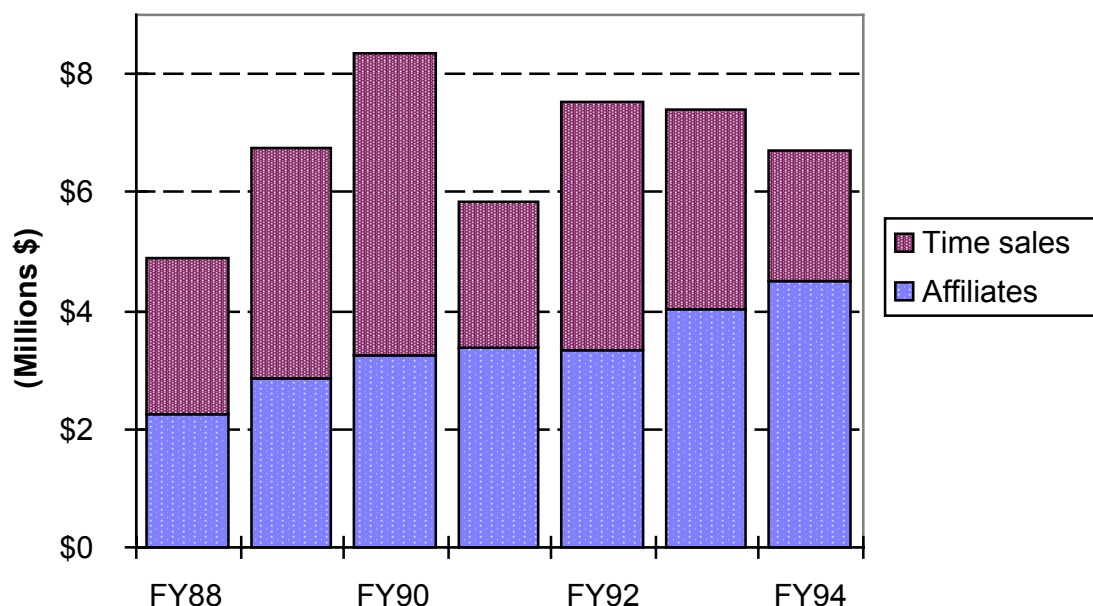


Figure 3.2: Industrial support of NSF SC Centers separating base affiliate relationships and sales of computer time for FY 1986-94

From an examination of the detailed usage data, the Task Force found that most of the time purchased at the Centers was purchased by a relatively small number of companies, and the decline of usage in FY1991 was from the cessation of use by three industrial firms. Each of these companies went on to purchase its own high-performance machines. From some points of view, this migration may be viewed as success of the program. Finally, while revenues from use are down, overall industrial use, in cycles, has significantly increased as rates have fallen.

As discussed in the factors section, the Centers program was created primarily to support fundamental, academic, research. Technology transfer to industry remains a secondary, but important role, while simple sale of cycles to industry was not, and should not be, a primary role for the program. Thus, the overall pattern shows continuing, valuable interactions with industry.

3.4: Effect of “Free” Cycles on the Market

A number of people have asserted that the existence of free Cycles at the Centers distorts not only *how* research gets done, but *what* research gets done.¹²

The following observations may be pertinent to this:

- Cycles at the Centers are no more or less “free” than those on a workstation bought on an NSF grant. Both are “free” in the sense that they are paid for by

¹² Some have also raised the concern that the program might distort the national market for high performance computing in detrimental ways. The magnitude of academic computing at the Centers, however, is not large enough to influence “the national market” (in the economic sense) in any material way.

Section 3: Issues

NSF. Both also have a cost to the researcher in terms of proposal preparation, ease of access, amount of time spent on system maintenance, etc. However, the cycles at the Centers may be considered “free” by an NSF program director in the sense that they come from another part of the budget.

- Every federal research program distorts the behavior of the PI’s to some extent, intentionally so. In this case, part of the rationale for the Supercomputer Center program was to encourage the development of computation as a modality of scientific investigation; to achieve this objective, high performance computational resources had to be both available and attractive.

The real concern is where the funding decisions are made. Some people worry that the existence of the Centers influences the decisions by both program directors and PI’s to use the center’s cycles rather than, for example, buying a more cost-effective local workstation, or joining with a group to invest in a network of workstations, or investing in other needed budget items such as hiring another graduate student. Others worry the opposite, that without central aggregation of resources PI’s would make apparently locally optimal decisions that were, in fact, neither globally optimal nor locally optimal in the long run. The aggregation necessary to acquire the highest performance machines and support would then not happen.

The fact that American academics had essentially no access to high performance computers prior to the creation of the Centers program is often cited in support of this latter view. However, as noted in the previous discussion, at the time the Centers program was created, there was no alternative to the aggregation of resources at a few national centers. With the advent of mid-range, scalable systems, there may now be.

The former view, that the Centers distort program directors’ and researchers’ decisions, is generally argued on its obvious rationality, but there is little evidence to support this assertion. For example, virtually every user of the Centers has powerful local workstations as well as access to the Centers.

But more importantly, the advocates of this first view generally sidestep the question of how to provide access for those users who require the highest possible performance at a given time, as well as how to most efficiently provide the high-end education and training functions and the development functions now facilitated by the Centers. We return to this question when we consider alternative mechanisms for funding the Centers and allocating their resources and when we consider alternatives to the current program.

A final perspective on this issue relates to the kind of cycles provided by the Centers. Some argue that the principal purpose of the program was to change the paradigm of research. If one takes this as the only objective of the program, then as technology matures it should no longer be necessary to provide that technology at the national centers. So, for example, vector supercomputing is now a mature technology and some argue that traditional vector machines should no longer be provided at the Centers; rather, resources should be focused exclusively on less mature technology such as scalable parallel machines. The validity of this view obviously depends on its premise about the objective of the program. Changing the paradigm was one of the initial objectives of the

Section 3: Issues

program, but not the only one, so the issue devolves to what the proper objective is now. This issue is discussed further in Section 3.12.

3.5: The Total Need for High-end Computing

Some critics of the current Centers program concede that there is a need for access to the highest possible computational resources, but not as much as the Centers program is currently providing.

The issue can be framed as to whether we need four Centers and how much equipment and service each Center should provide. Should the Centers provide access to the most powerful machines, since generally the cost-effectiveness of these machines is poorer than that of less powerful systems?

The crux of this issue is the total need for quality fundamental research and advanced education enabled by the Centers; is it sufficient to justify the program size and cost? Would we get more or less good science and engineering research if we reduced the Centers budget, particularly if we used the reduction to provide alternative computing facilities?

Another measure of the total need for cycles can be obtained by examining the allocation requests and the fraction of those allocations that the Centers are able to satisfy. The total demand exceeds supply by at least a factor of 2 and the fraction of the requests that are unsatisfied has been growing. Also, even with rapidly expanding resources on the parallel machines, and a lack of third party software, the use of these machines becomes very high shortly after they are installed.

The question of the “value” of science is a notoriously difficult one; it is no simpler in the case of the Centers than for other areas. In fact, the issue is more complicated for the Centers because they support a spectrum of scientific disciplines, and thus do not have the focused advocacy or consensus on the principles for evaluation of a single discipline. However, it should be noted that:

- There are examples where our current computational capability is far from what is required to solve scientifically and/or socially important problems. Turbulent flow and protein folding are often mentioned examples of this. Experience with these problems suggests that each increment in computational capability leads to new insights, and sometimes to fundamental changes in understanding of the underlying phenomenon.
- Each of the NSF Assistant Directors who talked to the Task Force stated that the quality of the science and engineering being done at the Centers is high, and voiced support of the program.

The next section addresses some of the issues related to judging the quality of the science enabled by the Centers.

3.6: Mechanisms for Judging Quality of the Science

Another criticism of the Centers is that the method of allocation of time is decoupled from the normal grant processing and merit review at NSF. Subscribers to this view argue that this could lead to support of science that is not up to NSF standards.

Section 3: Issues

There are really two issues here: (a) what is the quality of research using the Centers, and (b) does the method of allocating resources at the Centers ensure that excellent science is supported?

3.6.1: Quality of Research Pursued using the Supercomputer Centers

The NSF Division of Advanced Scientific Computing (ASC) has collected testimony and case histories on use of the NSF Centers. In the fall of 1994, a summary of research and technological accomplishments of the Centers was distributed to the National Science Board. This document, "High Performance Computing Infrastructure and Accomplishments," is available by request.¹³ As part of this study, the Task Force obtained additional highlights of major research advances carried out by computational scientists and engineers at the Centers. Information can be found in Appendix B.

The Task Force also directed an additional study from the Quantum Research Corporation data base on usage, and the NSF on-line grants data base. This study is presented in Appendix E. In this study, a four year sample of all substantial research projects was collected. There were 320 individual designated project leaders for the 1,428 identified projects. These 320 project leaders of research teams using the Centers were also named as principal investigator (PI) or Co-PI on 1,245 NSF grants.

The funding for the 320 project leaders or principal investigators averaged \$993,000 in total funding *from NSF* over the periods of this study: FY 1992-Feb 95 for their usage at the Centers, and FY 1988-94 for their NSF grants. The average of individual *NSF* grants identified for these researchers exceeded \$250,000 over this same period.¹⁴

Of particular interest is the quality of the research. The NSF data base contains detailed information on the reviewers' ratings of proposals (E-excellent, V-very good, G-good, F-fair, P-poor). Histograms were formed of these scores by NSF divisions and directorates. No significant difference could be detected in the distribution of the scores from the sample of grants from PI's that made use of the NSF Centers and a random sample of all NSF research grants.

The basic conclusions from this analysis are that:

- NSF grantees who make significant use of the supercomputer Centers have larger than average grants, and are better funded than grantees from the Foundation as a whole, and;
- Merit review of proposals shows similar rankings for Center users and for the general pool of awarded proposals.

3.6.2: Method of Allocating Resources at the Centers

At present, system units¹⁵ are allocated on the program's machines by a set of allocation committees, much the same as time on telescopes, accelerators, and other large facilities

¹³ The document is also available on the World Wide Web using URL: <http://www.cise.nsf.gov/acir/hpc/>

¹⁴ It should be noted that NSF centers do not restrict projects to NSF grantees, and in fact, 28% of all projects (and 21% of all usage) in FY 1994 were *not* in *direct* support of an NSF grant.

¹⁵ A system unit (SU) has been consistently defined as the equivalent of one CPU hour on a Cray X/MP.

Section 3: Issues

is allocated. Critics of this scheme point out that, unlike telescopes or accelerators, the allocation committees are not made up of peers from the same academic discipline, and hence are less likely to have the background to be qualified to judge the value of the science being proposed. These critics would prefer that the allocation be done by, say, NSF program directors as part of the normal merit review and resource allocation process. In the past a few grantees have also claimed that their research was subject to double jeopardy because it was reviewed by the normal NSF process and then again by the allocation committees.

Most of the proposals for time at the Centers are to support work that is funded by NSF or another merit-reviewed funding agency. Hence the allocation committees focus their review on the computational methods and the magnitude of the resource request. They rely on the normal merit review process to judge the science itself. This, it is claimed, avoids the double jeopardy issue. In the final analysis, of course, since the committees are allocating a finite resource, some judgment of the relative merit of the science may be made. This does present the possibility of second guessing the merit review of the science. Moreover, the double jeopardy issue is real in another sense. The PI is twice at risk of not obtaining the resources (money and computer time) necessary to complete the research.

The committee has examined these issues and believes that they are not currently a problem. Nevertheless, given tighter budgets and increased demand they could conceivably become a problem in the future, and so should be monitored by the Foundation.

The extent of the turnover in the largest allocations is evident from the longitudinal study of large users that is detailed in Appendix E. Of the top 24 projects in FY 1995 (through February), none were in the “large” usage category in FY 1992, and only 6 were there in FY 1994. If all projects under a specific faculty leader are aggregated to track the turnover in faculty leaders, one finds a similar picture: more than 2/3 of the faculty PI’s who were in the top 50 in FY 1995 did not rank in the top 50 in FY 1992. In fact, about half did not have active projects that crossed the 1,000 service unit threshold, defining the large users, during FY 1992-93.

NSF has a number of other facilities and infrastructure programs; in most cases they are specific to a discipline or to a few disciplines, e.g. telescopes, synchrotron light sources, and accelerators. The dominant mechanism for allocation is an allocation committee or board, very much like the Centers program, but the story is mixed. For those programs that use allocation committees, the primary difference is a more homogeneous scientific area, with broad overlap of the research interests of the users and the committee. For Polar Programs, which runs a variety of facilities for its researchers, such as research vessels and drilling equipment, the disciplinary expertise is maintained in the program and allocations of both facilities and funds are made by the normal merit review process.

At least in the past, there was reason to believe that few NSF program directors had the expertise to judge the quality of the proposed computational approaches – which is one reason why the current system came into use. In the long run, it would seem advantageous to both increase the capability of the program directors and to have at least

Section 3: Issues

some of the allocation of computational resources be part of the normal grant-making process. The Task Force notes, for example, that the disciplinary programs already have responsibility for allocation of workstations and most mid-range resources. These allocations would be more informed if the program directors had both the appropriate knowledge and the necessary insight into a proposal's computational methods and requirements.

Thus, although for the present, the allocation of computing resources may be best handled by an allocation committee, it is probably wise for NSF to facilitate greater participation of NSF program officers in the review process, and to move progressively toward putting more of the allocation process into a merit-based review system.

3.7: Alternative Funding and Allocation Models

The current structure of the Centers uses centralized cooperative funding agreements that provide the base funding for the Centers and an allocation process that is partially centralized (MetaCenter allocations process mentioned earlier) and partly distributed to the individual Centers. Several criticisms of these mechanisms have been raised including:

- criticism about distortion in the behavior of researchers as discussed in section 3.4,
- difficulties in evaluating the need for the Centers and potential inefficiencies among the Centers as discussed in section 3.5, and
- potential ineffectiveness or distortion in allocating resources at the Centers as discussed in section 3.6.

To address these criticisms, several proposals based on some form of high-performance-computing currency have been proposed. These strategies, variously called “green stamps” or just “stamps,” employ a method of budgeting for the Centers and allocating Center resources that ties the process closer to the disciplines and the funding decisions. These stamp proposals include versions with multiple types (colors) of stamps, as well as transition schemes that might eventually eliminate the stamps and treat all funding dollars (whether for computing or other) equally.

The key concept in a green stamp mechanism is the use of the stamps to represent both the total allocation of dollars to the Centers and the allocation of those resources to individual PI's. NSF could decide a funding level for the Centers, which based on the ability of the Centers to provide resources, would lead to a certain number of stamps, representing those resources, being available. Individual directorates could disperse the stamps to their PI's, which could then be used by the researchers to purchase cycles. Multiple stamp colors could be used to represent different sorts of resources that could be allocated.

The major advantages raised for this proposal are the ability of the directorates to have some control over the size of the program by expressing interest in a certain number of stamps, improvement in efficiency gained by having the Centers compete for stamps, and

Section 3: Issues

improvements in the allocation process, which could be made by program managers making normal awards that included a stamp allocation.

Other than the mechanics of overall management, most of the disadvantages of such a scheme have been raised in the previous sections. In particular, such a mechanism (especially when reduced to cash rather than stamps) makes it very difficult to have a centralized high-end computing infrastructure that aggregates resources and can make long-term investments in large-scale resources.

Stamps have often been proposed as a temporary mechanism with the intention of transitioning eventually to cash that may be used to purchase resources at the Centers or for any other use (students, travel, equipment, etc.). As discussed in section 3.4, the Task Force believes that such a change would too easily enable decisions made for short term gain to eventually destroy the aggregate, longer term values of the Centers program. Unless carefully implemented, such mechanisms might also lead to unstable, unpredictable funding levels for the program, and it would make re-aggregation of funding for the program difficult.

The Task Force believes that the goals of a green stamp mechanism, namely better participation from the disciplines in the overall program, improved efficiency at the Centers, and better coupling between disciplinary goals and allocation of center resources, are all useful goals. We believe, however, that if the Foundation were to move to a green stamp mechanism, great care would need to be taken to provide an environment that would encourage and nurture interdisciplinary research activities and would also facilitate reallocations from year to year based on changing research paradigms in the disciplines. NSF management needs to look at the green stamp approach and its staffing implications as well as alternative means of accomplishing these improvements in the program. Several suggested improvements were discussed in sections 3.4-3.6. Further suggested enhancements appear in sections 4 and 5. Recommendations for implementing these appear in section 6.

3.8: Role of Smaller Centers and Other Partnerships

Would NSF get greater leverage by using some (or all) of the present Centers budget to support existing state and regional centers, to support other agency centers, or to facilitate some other form of partnering? Equally, would the research community be better served by such a move?

This question is an instance of a familiar debate within NSF. The debate revolves around the Foundation's proper relative emphasis on getting "the best science" (which might involve concentration of its resources in a few institutions) vs building a broad national strength in science and engineering (which might involve spreading its resources more uniformly). We have found no consensus "right" answer to this debate.

However, in this instance, we may be able to satisfy both goals. The current Centers play a special leadership role and provide the top end of the pyramid of computing resources to the most demanding computational problems. One of their roles is to help make the newest commercial systems more usable for science and engineering, and to provide the

Section 3: Issues

sorts of education and training that would be difficult or impossible with only a larger number of smaller centers.

As the current MetaCenter collaboration is demonstrating, technology will increasingly enable integration of the operation of the various centers, permitting an application to execute wherever the most cost effective resources for that application can be obtained. This may enable the Foundation to separate some of the goals of the complete program from those of specific centers.

3.9: Technology Trends

The trends in technology have been copiously documented elsewhere, and are nothing short of amazing. At least some people question whether there is a continuing need for the Centers given that today's workstations are as powerful as yesterday's supercomputers. Alternatively, if today's workstations aren't powerful enough, why not wait awhile?

It's true. The ratio of cost to performance of microprocessors is much lower than that of current day supercomputers. Not only is it true that microprocessors' current absolute performance exceeds that of the supercomputers of only a few years ago, but this trend will continue for the foreseeable future. There is obvious validity to both of these arguments. But we also note the arguments in the NRC-HPCC report:

- The highest performance machines are a form of "time machine" in the sense that they let us solve today problems that we would have to wait years to solve otherwise. This time machine allows the U.S. research community to accelerate the progress in science, thus maintaining world leadership while at the same time providing long-term competitive advantages to the private sector. It is unfortunate that the magnitude of this value, like the value of all scientific investigation, can only be judged retrospectively, but experience suggests that it can be enormous.
- The highest performance machines are "time machines" in another sense too; they allow us to gain early experience with the form of machines and problems that may be "conventional" in the future. Again, there is value to the scientific and engineering community, and indeed to the society more broadly, to have someone gaining early experience at this "bleeding edge."

There is no alternative to eternal vigilance; the technology continues to move extraordinarily rapidly, but not uniformly so. The ever-changing balance between processor speed, memory size and speed, and communications bandwidth makes new architectures and variants of old ones sensible when they were not previously so, occasionally creating the need to rethink the optimal infrastructure. Thus, the Foundation will have to continually re-evaluate the best way to enable leading-edge computational science and engineering.

3.10: Computer Industry Trends

Some argue that the spate of recent failures of high-performance computer companies is indicative of a failure of the Centers program to create a healthy HPC industry, or (others argue to the contrary) a flaw in the very idea of high-performance computing.

Section 3: Issues

Note first that it was never the primary goal of NSF to create or sustain an HPC industry. Rather, the principal goal was to keep U.S. scientists and engineers at the forefront of research.

Second, it should be noted that, unlike every other segment of the computer industry, the market for the highest performance computers has been exceptionally inelastic – the current supercomputer market is about the same size as that in the late 1960's. All of the successful start-up companies in the computer industry succeeded in new markets (mini-computers, workstations, PC's, etc.), not by taking market share in an existing segment. The HPC companies, on the contrary, were fighting for a share of a fixed market – albeit one that the pundits predicted would expand. The presumption of an expanding market both lead to and exacerbated strategic errors on the part of failed companies.

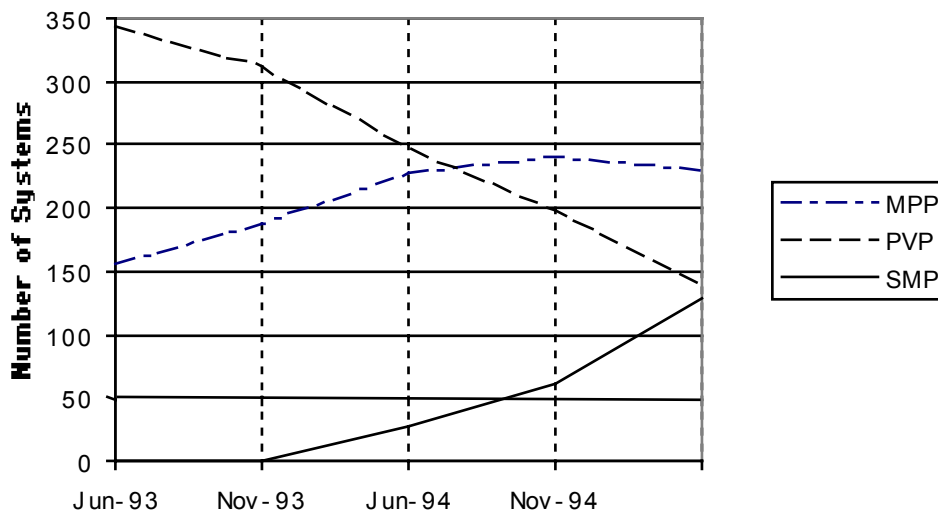


Figure 3.3: Worldwide Distribution of Types of “Top500” Computers¹⁶

Some might argue that the fact that the market has not expanded represents a failure of the Centers program to transfer technology and/or to effect the paradigm shift to computational science and engineering. There is no way to test this hypothesis, but the Task Force is skeptical of it; clearly the bleeding edge has paved the way for segments of the current workstation market – in systems software, in applications, and certainly in human resources. Indeed, Figure 3.3 documents the significant technology shift among the top 500 largest supercomputer sites to shared-memory multiprocessors, most of which are moderate scale machines, but built on the software and algorithms developed for larger scale machines. Another conclusion to be drawn from Figure 3.3 is the importance of paying attention to technology trends in planning for future balance in the overall Centers program, or the need for the *program* at all. This emphasizes the need for the program to be a savvy and insightful consumer of high-performance computing technology.

¹⁶ MPP - Massively Parallel Processors; PVP – Parallel Vector Processors; SMP - Shared-Memory Multiprocessors; data from URL <http://www.netlib.org/benchmark/top500.html>

Section 3: Issues

Moreover, as discussed in the NRC- HPCC report, the bleeding edge machines at the Centers have acted as a “time machine,” enabling research results sooner than they would otherwise have been obtained. While we may not be able to measure it precisely, there is significant value to this time advantage in terms of U.S. leadership in key areas of fundamental science and engineering.

3.11: Appropriate Role of the Centers in the NII

As noted in the discussion of sunseting the Centers, the focus of the Centers activities has evolved. They are not merely suppliers of computational resources; in order to properly support computational science, the Centers have assumed leadership in developing some aspects of the technology. Should this role further evolve to having the Center program play a leadership role in the National Information Infrastructure (NII)?

No one, at this point, seems to doubt the impact of the development of an NII – even if we don’t quite know precisely what that impact will be. The impact will be felt across the entire society, including the computational science and engineering community. It seems highly appropriate for the Centers to aggressively pursue this technology in support of the computational paradigm and as an enabler of computational science and engineering.

This observation, coupled with a more general one that there is a real need for NSF to assert leadership in the NII, has lead some people to suggest that the Centers should be assigned a broad leadership role. The contrarian view has several pieces:

- First, there is a general concern about *mission creep*; there are too many examples where successful organizations have ultimately failed to fulfill their primary role because their very success has encouraged them to be assigned additional responsibilities, but at significant cost in terms of diluted management and vision.
- Second, there is a concern that the Centers have an *unfair advantage* from their large base of support that would work to the disadvantage of individual PI’s (and ultimately to the disadvantage of the country).
- Finally, many think that there is nothing about the Centers role in the NII that is so *special* that NII research/development cannot be handled by normal program announcements, and competitive grants (as was done with digital libraries, for example).

The Task Force supports this latter view. Nevertheless the NII will be an indispensable part of the infrastructure of scientific and engineering research. Furthermore, we are sure that the Centers program needs to be deeply involved in this technology for the good of computational science and engineering. In addition, some NII experiments may require resources that can only be available at these national centers.

3.12: Should the Centers Continue to Support “Traditional” Vector Processing?

This issue was touched on briefly in the discussion of *free cycles*, in section 3.4 above. However, to elaborate – some people feel that the major goal of the NSF Centers has been to provide the infrastructure to enable a paradigm shift. According to this view, given finite resources, it would be better to invest those resources in leading-edge equipment

Section 3: Issues

(currently scalable parallel machines) to enable the next paradigm shift. Proponents of this view contend that, although good science may be being done on the Centers' vector machines, NSF is not getting a "double benefit," from investing in these machines by getting both good science and enabling a change in paradigm.

The Task Force notes that good science that requires the highest capability vector machines is being done at the Centers; a number of the grand challenge applications are in this category. High performance vector machines happen to be the most effective way to get some of that science done right now, and lesser capability machines would be inadequate to get this science done.

We also note that the Centers are in transition from complete dependence on vector machines to predominant use of scalable parallel ones. As scalable parallel machines become more mature and are better able to satisfy the needs of the full spectrum of applications, there is a natural path to make the transition complete.

Although an abrupt cessation of support for vector computing does not seem appropriate, we note that in times of tightening budgets, the Foundation will have to make some difficult decisions. Investments in future technologies that can support a wide range of scientific and engineering problems, such as parallel computation, should have priority over access to mature technology. The Task Force believes that the superior cost-effectiveness of parallel machines for a growing number of applications will tend to favor the deployment of more parallel machines. In a world of rapidly changing architectural forms, it is important that the NSF Centers program emphasize architectures that help move towards promising new forms of scientific computing as well as provide immediate scientific utility.

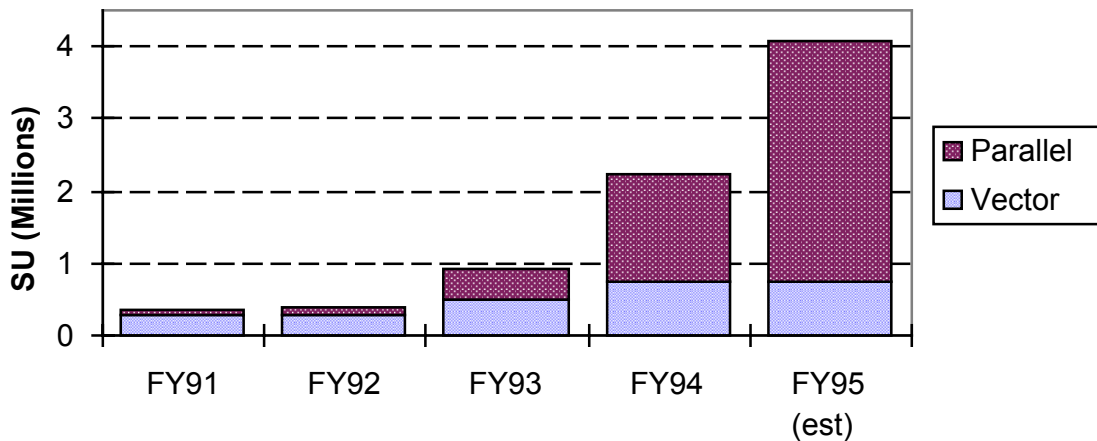


Figure 3.4: Increase in Normalized usage at Centers for both vector multiprocessors and parallel systems.

The capacity growth in parallel systems has been well documented, but even more startling is the dependence of the program on parallel computing – from 20% of the cycles in FY 1992 to 80% in FY 1995 – a complete reversal. Nevertheless, there is still

Section 3: Issues

substantial demand for the mature vector systems for those types of problems that currently do not perform well on scalable parallel systems.

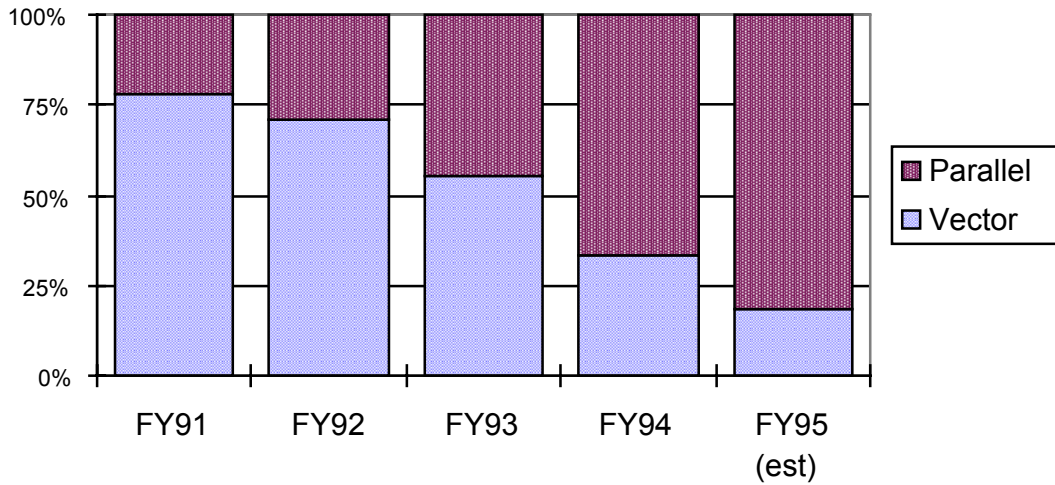


Figure 3.5: Distribution of Normalized usage at Centers between vector multiprocessors and parallel systems

3.13: Partnering with Other Federal Agencies

It has been suggested that NSF need not *go it alone* – that is, that more leverage and hence greater access to leading-edge machines could be achieved if the agencies that fund computation were to pool their resources.

To explore this possibility, a subcommittee of the Task Force met with senior officials from NASA, DoE, NIH, DoD, and ARPA.

While different agencies expressed different views about the long-term possibility of joint funding, NSF's sister agencies all indicated that, at the present time, given the uncertain budget climate, long term commitments to interagency centers projects are difficult. This does not mean, however, that the situation may not change. Some agencies expressed the view that high-end computation is so important to specific mission agency goals, and so clearly within overall agency budgets, that they will support very specific mission oriented, high-end centers within their agencies. Others, perhaps more concerned about the broad effect of budget cuts, expressed the view that when their budgets for the next five years are better known, and when their own planning is further developed, they will be interested in exploring – possibly for joint use, possibly for joint funding – either experimental mid-level high performance computing sites or a true interagency center “beyond the teraflop” range before the end of the century. These are possibilities that NSF management should continue to explore.

As NSF develops its plans for future interactions with other federal agencies, there are three points to keep in mind:

- First, any joint funding of sites needs to be synergistic. Each agency has its own goals and its own uses for high performance computing, and pooling resources at

Section 3: Issues

any one site does not necessarily lead to greater resources available for each of the participating agencies. That said, there should be opportunities for increased diversity in the program with joint, synergistic funding. One possibility might be a greater range of midrange programs. Another might be the possibility for more leading-edge sites. However, perhaps the most important possibility is the potential for advanced apex computation, *beyond a teraflop*, which might be available to several agencies by funding one site, open to a full range of academic users, beyond the level that might be possible for any one agency to fund.

- Second, whatever the possibilities for joint funding of individual sites, *all* agencies are eager to exchange expertise in software development, algorithms, and other technical interchanges. This has worked well under the interagency HPC management (HPPCIT), and should be continued in the future.
- Finally, in discussion with the Task Force, and in earlier discussions with NSF management, ARPA, which has previously contributed to the NSF Centers by helping with early placement of scalable machines at NSF Centers, noted both that their budget was under increased pressure and that they are moving to a funding strategy that will place far greater priority on individual projects driven by direct agency mission requirements. Thus, ARPA is likely to eliminate or decrease funding that contributes directly to the Centers program base budgets. Since in the past few years ARPA has made substantial contributions toward the purchase of parallel machines at the NSF Centers, this will have a serious impact on NSF's overall ability to maintain four sites at the leading-edge of commercially available technology.

3.14: Non-competitive Federal Funding for Supercomputer Centers

In recent years there have been a few supercomputer centers that have been funded through Congressional mandates. While some argue that such centers are here to stay and that NSF should take them as a given as it plans for the future, this strategy has significant risks associated with it. First, in recent years both Republican and Democratic administrations have attempted to remove funding for such Congressional mandates each year when the President's Budget is submitted to the Congress. As a result, planning for such centers has been difficult, owing to federal funding uncertainties. The second complicating factor is that the environment for embracing activities that some would classify as *political pork* is highly charged with emotion on both sides of the issue.

As a general principle, the Task Force fully supports the use of peer review in the funding of Supercomputer Centers that receive funding from NSF. Moreover, we believe that it would be inconsistent with this principle to endorse the notion that non peer-reviewed centers would become a part of the NSF program – *unless* they were successful in an open competition for such a designation.

3.15: International Cooperation

Some major infrastructure programs at NSF have been accomplished with significant international involvement. In general these have been facilities that are basically unique in the world, and devoted primarily to pure research, with few if any, technology or

Section 3: Issues

economic spin-offs. Supercomputer Centers do not fit this model in many respects. First, many countries already have such facilities using hardware from a variety of vendors, including those in the U.S. Second, there is a perception that the economic spin-offs of computing technology are real and relatively immediate. Third, the investment to develop and operate a leading-edge Center is well within the budget of many U.S. agencies, not to mention countries. All of these factors argue against major international computer facilities to support computational science. We believe, however, that there is significant international cooperation in fundamental research in the Centers program, including both computational methods and in basic disciplinary research, and this should by all means continue.

3.16: Commercial Suppliers of Resources

When the Centers program began, resources were purchased from commercial suppliers of cycles. At the time there were a number of such vendors. The situation has changed significantly. There are now very few commercial suppliers, although a number of commercial entities would no doubt be glad to procure and operate such facilities. Some of the mission oriented agencies are considering such “out-sourcing” arrangements as cost saving measures for production computing cycles. The NSF Centers, as detailed elsewhere in this report, are far more than production centers, and it is hard to see how the other important missions of the Centers program could be accomplished in a commercial setting. Nevertheless, a new, open competition of the Centers program should test the appropriateness and efficiency of commercial suppliers for NSF users.

Section 4: Future Options

Following interviews with members of the HPCC and research user communities, and taking into consideration the factors that define the scope of the program and services demanded by the program, the Task Force considered a number of options including the following five representative options. These options range from continuing the program in much the same manner as the existing program to discontinuing NSF centralized support for advanced computing systems and services. Based on the Task Force's assessment of funding realities at NSF over the next decade, one option that was *not* considered was enlarging the scope of the program. With each option we have assembled a list of pros and cons based on the information the Task Force obtained from our interviews, our survey, and the background of Task Force members. The options presented are:

- A. Leadership Centers similar to the current program
- B. Partnership Centers
- C. Single Partnership Center
- D. Disciplinary Centers
- E. Terminate the program

Option A: Leadership Centers, similar to the current program

A number N of "Leadership Centers" are selected in response to a specific program announcement. The current program, before introduction of the Metacenter Regional Alliances, was approximately equivalent to this option with N=4. These centers would have access to sufficient computing hardware and software to enable them to provide the infrastructure for high performance computing in a broad cross-section of science and engineering applications, including computer science. These centers would provide educational leadership in high performance computing. There would be significant cost sharing with other Federal agencies, industry, the states, and the sponsoring universities.

Pros:

1. This approach has been successful over the past ten years and there is every reason to believe that it would continue to be an effective model for NSF leadership in providing the infrastructure necessary for continued research advances in computational science and engineering.
2. This approach has proven effective in providing the kind of education and training necessary to facilitate the significant shift towards massively parallel computing.
3. Continuation of significant cost sharing would leverage the NSF investment.
4. Having several very-high level centers provides considerable flexibility for NSF to encourage different high-end thrusts among the centers.

Section 4: Future Options

Cons:

1. As state, regional, private, other agency, and university centers move to acquire smaller versions of the hardware that is available at the leadership Centers, it will become increasingly important to stage usage among other centers and the Leadership Centers. This option divides up key decisions in a way that may lead to an overall program that is suboptimal both in terms of educational offerings and, importantly, resource sharing.
2. This option might yield an unbalanced pyramid of computational capability, with disproportionate NSF support at the top and bottom, and too little elsewhere.
3. This option may not provide as effective an infrastructure for experiments with distributed and mid-level parallel computing as that presented by Option B. As a side effect, there might be too few experimental alternatives to the high-end leadership Centers.

Option B: Partnership Centers

A number ($N > 1$) of leading-edge centers are organized as cooperatives among several sites connected with high speed communications networks. At least one site within each partnership would have highest-end commercially available computing capabilities in terms of computer power, memory, I/O and communications. Other sites (e.g., state or university centers) might have smaller versions of this, or related, hardware. The leading-edge site and its partners would present a coordinated computing resource. These partner sites would:

- promote effective regional education,
- facilitate the development and testing of software, algorithms, and applications, including networking and distributed technologies, and
- provide computing cycles for applications that do not need the resources of a high-end site.

NSF would issue a program announcement that would encourage partnerships among the leading-edge site(s) and multiple partners. The affiliated sites could span a wide range of functions and sizes, from narrowly focused sites on a single campus to large-scale state and/or federal centers, with tight coupling to a leading-edge site. The proposals, merit review, and NSF funding level would determine the strength of the connections between partner sites and leading-edge site(s).

Pros:

1. Such a distributed center should provide a more cost-effective means for providing the infrastructure necessary for continued research advances in computational science and engineering. This structure may also be one of the most effective ways to get cost sharing across a wide community. While this is not the approach that has been used over the past ten years, the explicit coordination of several sites – possibly including Science and Technology (S&T) Centers and Engineering Research (ERC) Centers, as well as Federal (DoD, DoE, NASA) and

Section 4: Future Options

state centers, and more narrowly focused university centers – could be a more effective model for NSF leadership in the future.

2. This approach should be more effective than the current structure in providing the increased level of education and training necessary to support wider use of emerging parallel computing technology. Such a distributed center provides more effective coupling to other sources of significant human capital and other resources.
3. The partnership sites could allow for different thrusts in particular applications areas. They might also develop hardware or software infrastructures uniquely adapted to particular application areas. This structure could encourage and facilitate experiments with new architectures and new software technologies at the partner sites. Moreover, this structure would offer the opportunity to experiment with more distributed computing models.
4. The coupled midrange machines may also provide capabilities not available on the highest-end machines in either software (e.g., the availability of some software) or hardware (e.g., better visualization capabilities, more memory per node, etc.).
5. The smaller, partner sites offer the opportunity to change the structure of the program on a shorter time scale.

Cons:

1. Support of more midrange machines at the partner sites and the required support of high-performance communication among this larger number of sites would reduce NSF's ability to support the highest level of computing at the very top of the pyramid of computational capability.
2. Effective coordination and management of both the vertical interaction and the horizontal interaction across the leading-edge sites would be more difficult.
3. The processes for managing resource allocations are likely to be more complicated and time consuming.
4. Trying to use resources in a distributed manner may not work or may not be cost-effective for some significant research areas.

Option C: Single Partnership Center

Same as Option B, with N=1. The Center would be organized as a cooperative among several sites, one of which would be the leading-edge site. The single leading-edge site would have the very highest end commercially available capabilities in terms of computing power, memory, I/O, and communications. Associated sites would have smaller versions of the systems for the purposes of software and algorithm development, production runs that don't require the largest facilities, education, and evaluating new hardware and software technologies.

Pros:

1. As compared with option B, assuming the same total funding for hardware at the leading-edge site(s), the hardware infrastructure at a single leading-edge site could

Section 4: Future Options

have N times the capability. In this case, the aggregation of memory capacity is probably the most significant potential gain, although there might be marginal savings in operational support costs. Alternatively, the systems at the leading-edge site could be purchased in a shorter time frame, allowing the center to keep up with rapid advances in the technologies. More likely, some combination of larger systems and accelerated payment schedule would prove most appropriate.

2. A single center might be more easily able to ensure appropriate geographic diversity in the total set of partnership sites.

Cons:

1. Lack of competition among multiple centers could lead to the single center being less responsive to what would be most useful to users.
2. This model may have less diversity of experimentation at the partnership sites because a high priority is likely to be placed on compatibility among the partnership sites and the leading-edge site.
3. Given the architectural convergence issue – the uncertainty in the optimal architecture for the highest levels of performance – and the continuing interest in achieving a programming model that is machine independent, it is risky, from a national perspective, to have a single center making the selections of capability for the leading-edge site(s).

Option D: Disciplinary Centers

The Centers program would consist of several “Disciplinary Centers.” The Scientific Computing Division of NCAR is an example of this model.

Pros:

Disciplinary centers have the advantage of being able to better focus on those research issues that are most important to the field.

1. It is easier to determine the appropriate funding level if one is making trade-offs within a single field. NSF has a long history of making such determinations between centers and small projects within a single discipline, such as astronomy or physics.
2. Disciplinary centers may be more effective in furthering international research links because these links already exist within the discipline.

Cons:

1. Some disciplines are not of sufficient size, may not have logical partners, or may only be starting to understand the value of high performance computing to their discipline.
2. This approach would not facilitate cross fertilization among fields and between scientists and engineers, and computer scientists and applied mathematicians.

Section 4: Future Options

3. The coupling with other high performance computing activities such as those at the existing university and state centers might not be very effective because these other centers usually do not have a disciplinary focus.
4. There would probably be a more limited set of computational options and a narrower base of support.

Option E: Terminate the Program

This option removes the direct subsidy for high performance computing. NSF ceases to provide centralized direct support to high performance computing centers. Funding for high-end computing would need to come from individual project grants.

Pros:

1. After 10 years, the paradigm shift that has enabled computational science and engineering to emerge with important modeling capability should enable computational scientists and engineers to pursue significant fundamental research activities without centrally funded facilities. Some argue that individual grantees can and should directly compete for the sorts of major computational resources that are now sheltered through the Centers program.
2. Funding at the project level would provide the principal investigator with more control. This should lead to greater responsiveness to the needs of the project and to the ability to better optimize resource allocations to produce the best research within a fixed budget.
3. The proposal review process would be enhanced because there would no longer be the potential for a project being in double jeopardy and, importantly, reviewers would see total project costs and would be better able to advise the NSF on the cost benefit of various competing projects.
4. NSF would have greater overall budget flexibility to achieve balance within the pyramid of computational capability, and it would no longer need to put great pressure on participating centers to put up significant cost sharing.

Cons:

1. This option would not ensure that there would be adequate high-end computing infrastructure available. A probable consequence would be significant stretching out of some break-through and paradigm-shifting research and major delays for research projects that require the highest level capability. It is not realistic to expect the highest level infrastructure to survive without significant centralized funding because neither the centers nor individual PI's have sufficient influence at NSF or other agencies to ensure the stable source of funding required for high performance computing equipment and personnel.
2. While some educational aspects of computational science and engineering can be met at centers with smaller shared-memory multi-processors (SMP) and massively parallel processors (MPP), this option would not provide the high-end research and software development infrastructure needed to bring new and emerging hardware and software advances to future users. There is still much to be learned

Section 4: Future Options

- about high performance computing – particularly massively parallel algorithms, software and productivity tools.
3. The resulting reduction of significant financial and intellectual leveraging from industry, other federal agencies, and state and regional centers would be a significant loss to the national effort in high performance computing.
 4. While decisions on budget trade-offs are important and healthy for research, they need to be made in the proper context. With the significant strains on federal budgets for fundamental research, over the next few years this option may lead to suboptimal trade-offs if individual investigators and program managers make trade-offs with funds that could appropriately be allocated for high performance computing infrastructure. While such individual decisions may not appear to be having much short term impact on the competitive position of U.S. computational scientists and engineers, the long term result is likely to be the loss of U.S. leadership in a research paradigm that is central to our economic health and well-being.

Section 5: Future Directions and Priorities

5.1: Rationale for a Centers Program

The rationale for a Centers program must always be the ability to support the computational needs of leading-edge science and engineering research. The unique position of the NSF Supercomputer Centers is in providing the highest-end commercially available systems in terms of computing power, memory, I/O, and communications of the form that can only be available at a few national Centers. The rationale for the program's existence must rest on the quality of the science and engineering research that the facilities and staff of the Centers enable. As computation continues to increase in importance as a research tool, the Centers will continue to be needed as long as the academic community cannot find adequate resources elsewhere.

The Centers also have a role in educating the advanced scientific and engineering communities about high performance computing. Providing expertise in the use of the leading-edge computational facilities to the scientific and research community is expected to be an important role for the foreseeable future.

The broad acceptance of high performance computing does not eliminate the justification for an educational or training role at the Centers, but it does reduce the need for a missionary role, which the Centers performed in earlier times. Although the advantages of high performance computing are well-understood, and the use of vector computers is now largely routine, the use of large-scale parallel machines is still in its infancy and will continue to be challenging for many more years. Thus, providing expertise and training in the use of high-end computational resources in support of science and engineering research will continue to be important for the NSF community.

5.2: Primary Role Of The Centers: Full Service Access To High-end Computing Resources

The primary role of the Centers program has been, and should remain, to provide the highest level of computational services to the scientific and engineering research community in as efficient a manner as possible. The cornerstone of these services is access to high-end machines and support that can reasonably be expected to be available at only a few national Centers. As demonstrated in the survey of users (Appendix G), access to highest-end systems remains one of the most important attributes for users. This viewpoint was reinforced by both the quantitative answers to the survey questions and by an analysis of the answers to the open-ended questions.

The Centers program must include the user assistance, access to tools, education, and training that allows effective use of the Centers' resources. For the rapidly changing parallel machine technologies, this training and access to machines for development is a critical part of the mission. For vector machines, the stability of the architectures and software reduces the need for such a training, education, and development component. The survey of users also attests to the importance of the training, education, and consulting services. The answers to open-ended questions strengthened the quantitative measure of the importance of these functions, with a number of users stating that the Centers' expertise and support is indispensable.

Section 5: Future Directions and Priorities

To support the easy and efficient use of emerging technologies, the Centers must have expertise in the choice of appropriate architectures and programming tools, as well as general knowledge of the application domains and computational techniques. In providing this type of support, a Center may engage in a wide range of activities, from acquiring and integrating software (both applications of widespread interest, and programming tools) to the creation of software environments in support of various computational communities.

While the four NSF Supercomputer Centers have not, to date, strongly focused on providing massive data storage, there is a general trend in scientific computation towards the generation and use of much larger scientific data sets. While some of these data sets may later be captured in national databases, there can be significant data storage needs that are an integral part of a particular computational study. This is another aspect of the computing environment that needs to be balanced with the growing computing power of the Centers. The Centers program will need to provide equipment and personnel to meet unique national storage needs in the future. Indeed, several Centers are already moving in this direction, including the use of experimental systems for new large-scale data storage systems.

Large-scale information repositories are a largely service function, as opposed to an NII research function. Providing support for such functions may sometimes be appropriate. But many such information repositories have important differences from the current Center services--the repositories tend to be discipline specific and long-lived. This does not preclude placing such repositories at the Centers, but it does mean that when such services are located at Supercomputer Centers, they should remain ancillary, and closely related to the primary mission of providing broad access to computational resources.

Because relatively large-scale data Centers are less costly than large-scale supercomputers, the need to centralize these facilities at national Centers is less obvious. However, if there are some overall efficiencies associated with supporting both the storage of supercomputer-generated data sets and the management of national data archives, this may generate disciplinary support for these services.

In the future, very high bandwidth data transmission will become increasingly important to the Centers to ensure that users can get adequate access to increasingly powerful resources. Indeed, the Centers have participated in gigabit networking research in the past, and are currently in the process of working towards increased connectivity among the Centers and to users through new "very high speed Backbone Network Service" (vBNS) connections. Improving network capabilities will have a major impact on the overall Supercomputer Centers program and will be important to the success of all Centers. We expect the Centers to be early adopters of such technology and to assist in its development where it is crucial to the service mission of the Center. At the same time, high bandwidth communications, by their very nature, must be dispersed, and, to be maximally effective, must serve an increasingly broad segment of the scientific and engineering communities, a community much larger than that served by the Centers. Thus, Center participation in networking research or broad network infrastructure development should be an auxiliary mission for the Centers.

Section 5: Future Directions and Priorities

5.2.1: Role of Research Programs in the Centers

Because the Centers have a role that involves providing assistance in the effective use of high performance computers and in selecting new computer systems, maintaining high quality expertise at the Centers is critical to carrying out the primary mission.

Maintaining this type of expertise requires that the Centers provide a stimulating intellectual environment. Such an environment is most easily maintained when the Centers have an active research program, either on their own or in collaboration with other academically-based research groups. Without such interactions, the Centers cannot maintain the intellectual vitality needed to keep the highest quality talent. In addition, a research program can help in informing the Centers about the needs of the community, improving the quality of the hardware and software acquisitions. Although this research component of the Centers' activity is secondary to the mission of providing high-end computational cycles and services, it is vital in the sense that the Centers cannot continue as high quality providers without the kind of expertise provided by people who maintain active research interests.

There is, however, a concern with a significant research role for the Centers, namely the danger of the Centers competing with other NSF grantees in a way that might be considered inequitable. When the Centers compete for research money, they compete against researchers whom they normally serve. This could create an awkward relationship between the Centers and the community they serve that the program must take into consideration. Such competition raises the concern that the Centers, with their large budget and staff, will have an advantage in competing for funding. This issue is not unique to the Supercomputer Centers program. For example, the potential exists at NCAR and the major astronomy Centers.

Such concerns must be mitigated by maintaining the research role as secondary to the primary role of serving the outside community and by taking steps to ensure that the Centers have no explicit advantage (particularly no advantage in cycle and staff allocation) in competing for research programs.

One particular role that has been advocated by some for the Centers is as laboratories or Centers for NII-related research. The Task Force believes that the Centers program has a unique role in providing high-end scientific computation and that this role should remain the primary focus of the program. The Task Force agrees with the views stated in the NRC-HPCC report: the needs of the majority of NII-related projects can be better and more cheaply supported in a widely distributed fashion than in a centralized fashion. Thus, while individual Centers may participate in some NII activities, such activities should not become a major focus for the Centers program.

5.3: Context for the NSF Supercomputer Center Program

The report of the Blue Ribbon Panel on High Performance Computing laid out a pyramid model of computing for scientific and engineering research. The pyramid of computational capability includes, starting at the top, high-end supercomputer facilities (the so-called "apex"), mid-range machines shared by a group, department, or school, and at the base, workstations on individual desks.

Section 5: Future Directions and Priorities

The high-end machines are the focus of the Centers program. The Task Force believes that for many leading-edge scientific and engineering applications access to the high-end remains critical if the United States is to maintain leadership in science and engineering. The Task Force finds the time machine argument for the high-end advanced in the NRC-HPCC report compelling. While there are areas and researchers who will be able to make significant progress without access to the highest commercially available systems, these constitute only a portion of the research frontier. There are several significant areas where academic researchers can obtain a five to ten year time advantage if they have well supported access to the highest-end computing systems.

At the same time, we observe that in the lower portions of the pyramid, computing cycles become significantly cheaper. Thus, investment efficiency demands that the investments in the apex be balanced by investments in the middle and lower portions of the pyramid of computational capability. This is the crux of the “balance” argument that is well-articulated in the report of the Blue Ribbon Panel.

The lowest portion of the pyramid, namely individual workstations, are typically purchased in the context of individual grants, possibly using university resources to supplement government funds. The Task Force believes that there are sufficient mechanisms for purchasing desktop machines.

The Task Force concurs with the Blue Ribbon Panel recommendation concerning the balancing of the pyramid. There may be a continuing imbalance in the middle of the pyramid because of bureaucratic barriers that limit the funds available to purchase mid-range machines. Such machines are more cost-effective than the apex machines because they are cheaper to purchase and to operate. Moreover, many of the smaller allocations at the current NSF Centers would be more efficiently, and probably more effectively, serviced on smaller machines either on-site in an academic institution or in a regional Center. Furthermore, such mid-range machines are often more effective for early development activities for major applications.

With these observations in mind, the Task Force calls again for NSF leadership to encourage the purchase of mid-range machines and to increase the funding available for such purchases. Without such leadership, users that could be served on mid-range machines are forced to use the Centers or to use less capable workstations. The use of high-end machines by those who would be appropriately served on smaller machines is an inefficient use of NSF dollars. The Task Force does not advocate using Centers program resources for this purpose, except in cases where such support directly impacts the Program’s primary mission.

Dramatic increases in the ability to connect users and resources will decrease the need to centralize certain functions. Thus, it would be beneficial to have distributed sites with mid-range systems that can serve as development vehicles for the larger-scale machines, as well as to handle the computational needs of smaller applications. The Task Force believes that the Centers program should include a component that supports and couples mid-range machines at universities and local Centers to the larger machines at the national Centers. Such an approach has the potential to improve both the efficiency of

Section 5: Future Directions and Priorities

use of the large machines and to increase the outreach and educational impact of the overall program

5.4: Interactions with Industry and Government

5.4.1: Vendor Interaction

The Centers play a role in interacting with vendors, both to provide information about the needs of the high-end scientific computing community and to provide feedback to the vendors on the suitability of their machines and needed enhancements. Nevertheless, one cannot expect the Centers to act as the major proving grounds for new technologies for two reasons. First, the Centers represent only a small portion of the market, focused at the high-end. Second, providing candid feedback requires the Centers to be harsh critics of the vendors. As quasi-public agencies, this role is very difficult, since it places a government funded group in the position of endorsing winners and identifying losers. Thus, the primary role of providing market input and criticism must come from industrial users.

The Centers do have a role in identifying strengths and weaknesses of machines and software in the particular computational environment that the Centers provide and in evaluating the efficacy of the basic architectural paradigms. This role, however, cannot be a major justification for the Centers existence and must necessarily be largely secondary to the role played by the broader industrial market for high performance computing. The current model, where the Centers provide informal feedback to vendors, to users, and to potential purchasers, is perhaps the best balance.

5.4.2: Other Industrial Users and Technology Transfer

The Centers play a continuing role in introducing and supporting initial experimentation with supercomputing by industrial users. But, the amount of actual industrial usage at the Centers remains small. This is a reasonable expectation, since the Centers are not intended to serve as industrial computing Centers. Fulfilling the need for providing initial exposure to high-end computing and experience in its use to industry is an important component of the Centers' mission. The Task Force does not expect this role to lead to major use of the Centers' facilities by industry.

The Task Force observed that *other* industrial funding of the Centers (primarily industrial affiliates programs), which is not focused on usage charges, has steadily increased; it accounted for \$5.3M in FY 94 or about 9.2% of the base NSF cooperative agreement -- the Center core program, or 4.3% of the total Centers' budgets. This funding indicates an interest in the broader mission of the Centers to educate users about high performance computing and to develop technologies to assist in its use. Related to this role is the transfer of technologies developed for research users to industry. While such transfers are certainly laudable, the mission of the Centers should remain primarily focused on NSF research users and in the education of a new generation of outstanding Ph.D's in new areas of computational science and engineering. In the long term, these new Ph.D's will be one of the most effective means of diffusing technical expertise to both industry and academia.

Section 5: Future Directions and Priorities

5.4.3: Interactions with Other Government Supported Centers

The Centers have interacted with a variety of other government-supported computational facilities, including NCAR, DoE facilities, and the regional and state supercomputing Centers. Such interactions facilitate cooperation in exploring new hardware and software systems as well as development of support software. In addition, close cooperation with state and regional Centers helps create facilities that complement one another and allows users to more easily scale up as their level of sophistication and applications software increases and/or their computational need increases. The expertise and experience of the NSF Centers can be of significant value to the smaller state and regional Centers. The Task Force believes that such synergistic interactions should continue and be encouraged.

5.5: The Ongoing Role of the Centers

The changes in computing and communications technologies, together with changes in the possible needs of the research community, affect possible future roles for the Centers. In this section, we first discuss the potential need for the Centers' high-end facilities and then examine the impact of technology changes and possible strategies for the Centers in the future.

5.5.1: The Need for the High-End

The Task Force believes that the need for high-end computational resources will continue as scientists and engineers increase their ability to use computation as a tool. While many of the users can be accommodated by high-end workstations and mid-range machines, the highest-end facilities will be needed by researchers with the potential for significant new breakthroughs.

Because it is difficult to predict which areas or researchers will need such facilities in the future, a pyramid of computing facilities spanning from workstations to high-end supercomputers, is an appropriate model for resource allocation.

This model includes not only small and mid-range machines at the researcher's home site or a regional Center, but also a pyramid of allocations at the Centers. This allocation pyramid leads to modest facility allocations to the majority of users and large allocations to those researchers judged by the allocation committees to have the best potential for the most significant new contributions. The Task Force believes that such an allocation policy most effectively uses the unique, high-end capability offered at the Centers.

This approach also facilitates and encourages cycling between periods of small allocations and the large allocations. The dynamism of the Centers program is enhanced by ongoing and significant turnover among the largest users. As shown in Appendix E, such turnover is indeed experienced among the current large users. (For example, between FY 94 and 95, there was a turnover of 2/3 of the top users.) The Task Force believes that such continuous change among the largest users of the Centers' resources is desirable and leads to higher productivity in the research enabled through the facilities. The challenge for the future is to enhance the actual and perceived fairness in the merit review of the largest requests by careful scrutiny of preliminary computational data that documents the technical readiness to use effectively the requested allocation.

Section 5: Future Directions and Priorities

Because this pattern of allocations leads to rather large allocations to a small number of individuals, it is appropriate that greater care and broader input be used in making such large allocations. The Task Force believes that encouraging greater involvement by NSF program directors in the large allocations would be beneficial. The program directors can bring additional expertise in evaluating the potential for a breakthrough in their disciplinary areas. Furthermore, greater program director involvement will increase the ties between the Centers program and the individual directorates served by that program.

5.5.2: Technology and Market

The reduction in the numbers of vendors and the design space of machines reduces the number of sites that are needed to have each of the major computational paradigms involved in the program. As the number of paradigms reduces to 2 or 3 over the next few years, it will be possible to have examples of each class of architecture with fewer sites. However, this probably will not reduce the cost of having the highest capability machines, which the Task Force believes should be the primary focus.

5.5.3: Maintaining the Leading-Edge Capability

At the present time, by maintaining high-end access, and by creating an effective education and training environment for the best of our future computational scientists and engineers, the Centers play a critical role in maintaining U.S. leadership in leading-edge science and engineering. Furthermore, the Task Force believes that this need for access to the high-end will continue at least for the near future. The primary role for the Centers should stay focused on providing full service access to the high-end of computation, including supporting education and training in the use of such capabilities. As discussed earlier, such a focus should naturally include a role for mid-range systems at distributed sites, including the coordinated deployment of smaller versions of the machines that reside at the leading-edge site(s).

While the Task Force recommends continued focus on the high-end, we also believe that as the Centers program continues, the attendant price/benefit ratio of high-end versus more distributed access to computational resources will require regular monitoring to be sure that the Foundation, and the nation, are receiving maximum benefits from their investment in high-end computational resources.

Perhaps one of the most significant challenges facing NSF and the Centers program is how to provide access to the next generation of high-end machines in an environment where resources are limited and the largest machines may cost \$20-30M. Several of the alternatives discussed in the previous section (Options) provide a method for addressing this important issue. For example, Alternative B proposes creating a *small* number of *partnership centers* that have the highest-end machines and support at leading-edge sites, then focusing other sites on smaller machines and on related aspects of the mission in a coordinated partnership fashion. The Task Force believes that this alternative provides a good approach to allow the Centers program to continue to focus on their primary role: providing access to the highest-end resources for NSF's most technologically demanding investigators, while simultaneously making the emerging technologies available to a wider class of users through partnerships with university, regional, and national Centers.

Section 6: Recommendations

6.1: Continuing Need for the Centers Program

Recommendation: In order to maintain world leadership in computational science and engineering, NSF should continue to maintain a strong, viable Advanced Scientific Computing Centers program, whose mission is:

- ***providing access to high-end computing infrastructure for the academic scientific and engineering community;***
- ***partnering with universities, states, and industry to facilitate and enhance that access;***
- ***supporting the effective use of such infrastructure through training, consulting, and related software support and development services;***
- ***being a vigorous earlier user of experimental and emerging high performance technologies that offer high potential for advancing computational science and engineering.***
- ***facilitating the development of the intellectual capital required to maintain world leadership.***

The Task Force's chief finding is that there are significant areas of computational science and engineering where the current Centers have made possible not only major research results, but also paradigm shifts in the way that computational science and engineering contribute to advances in fundamental research and associated advanced education and training across many areas supported by the Foundation. This, together with the evolution of the underlying enabling technology, is still continuing, and still requires support in order to enable world leadership in computational science and engineering across many disciplines. In some areas, such as cosmology, ocean modeling, fluid dynamics, and materials research, advances that are possible only through the use of the most advanced computer modeling are already an essential component in maintaining U.S. leadership in the field. For these fields there is a continuing need to have access to leading-edge computational capabilities – including computing speeds beyond the teraflop level, significant memory and storage, plus advanced graphics and visualization capabilities coupled with high speed networking. The Task Force also believes that there will be significant growth in the number of disciplinary and interdisciplinary areas (for example, ecological modeling, and multi-disciplinary design optimization) that will be significantly advanced as computing capabilities advance and as the relevant scientific and engineering communities develop a cadre of knowledgeable users.

The Task Force is convinced that, at the present time, a Centers program is the best mechanism through which NSF can efficiently meet its responsibility to maintain world leadership for those areas of research in computational science and engineering and computer science that require leading-edge computational infrastructure.

6.2: Specific Infrastructure Characteristics for Leading-edge Sites

Recommendation: NSF should assure that the Centers program provides national “Leading-edge Sites” that have a balanced set of high-end hardware capabilities,

Section 6: Recommendations

coupled with appropriate staff and software, needed for continued rapid advancement in computational science and engineering.

In order to maintain world leadership in computational science and engineering, and in order to have a balanced program in which leading research Centers can continue their educational and research mission, the infrastructure at leading-edge sites should have several key components. High-end hardware systems should be one to two orders of magnitude beyond what is available at leading research universities. These systems need to be balanced in terms of processor speed, memory, and storage systems. These should be accompanied by appropriate staff, software (including mission-specific software development), and, increasingly, high speed data communications that will enable leading-edge sites to work effectively with other computational Centers, other NSF Centers, and the research and education community as a whole.

Access to this leading-edge capability is necessary for the most advanced computational science and engineering, and in the immediate future such access is likely to become increasingly important for experiments in computer science and engineering as well. However, in the current budget climate, the costs of this infrastructure are such that it can be available only at a very limited number of national sites. Thus, to be effective, it is essential that the NSF Centers program provide access to a few well-balanced leading-edge sites that contain the key hardware, software, and intellectual components.

Balanced high-end sites not only enable leading-edge computational science and engineering that can now be performed no where else, but balance is also required to provide the most effective educational and training environment for future applications and technology. Finally, balanced leading-edge Centers provide a critical environment for the testing of new software, algorithms, and hardware. As parallel computation has come to play a dominant role in the Centers program, and as enabling research has focused on the problem of dealing with scalability of both applications software and the underlying systems software, the Centers have increasingly benefited from interaction with computer scientists and engineers. We expect this trend to continue as long as issues of scalability remain critical. Thus we expect this interaction to be part of the balance needed for the overall program in the immediate future.

6.3: Partnering for a More Effective National Infrastructure

Recommendation: NSF, through its Centers program, should assure that each Leading-edge Site is partnered with experimental facilities at universities, NSF research Centers, and/or national and regional high performance computing centers. Appropriate funding should be provided for the partnership sites.

Such partnerships will increase the impact and efficiency of the leading-edge sites by promoting regional education, facilitating the development and experimentation with new hardware, software, and applications technology, and providing cycles for applications development runs that do not need the high-end capabilities of the leading-edge sites.

The national program in high performance computing is enriched by the presence of NCAR, several computationally oriented NSF Science and Technology and Engineering Research Centers, Centers funded by other federal agencies, as well as a number of state

Section 6: Recommendations

and university high performance computing Centers. Over the past several years the NSF Supercomputer Centers have developed an array of relationships with these Centers. The future of high performance computing is likely to benefit from greater partnering of these Centers with NSF leading-edge sites. There are two specific ideas that merit particularly careful evaluation through a competitive awards process:

- The nature of the formal partnering with the *leadership sites* to provide a more robust and cost effective infrastructure for high performance computing.
- Support for high speed network connections among these Centers in order to facilitate more effective interaction and to provide an infrastructure for experiments which require a coupling of high bandwidth communications with very high performance computing.

While the Metacenter Regional Alliances program currently facilitates some coupling, that program does not provide for coordinated planning and allocation of resources nor for enhanced networking. A coordinated plan for experimenting with new high performance technologies and for resource allocation should be more cost effective and should offer more responsive services to users.

6.4: Competition and Evaluation

Recommendation: NSF should announce a new competition of the High Performance Computing Centers program that would permit funding of selected sites for a period of five years. If regular reviews of the Program and the selected sites are favorable, it should be possible to extend initial awards for an additional five years without a full competition.

The Task Force is concerned about the effects of a new competition on the environment that has been built up at the existing Supercomputer Centers. Dedicated, outstanding staff as well as relationships with other Centers, with the nationwide academic community, and with industrial partners are difficult to build and hard to maintain in the face of uncertainty. At the same time, competition and continuing evaluation are consistent with long-standing NSF policy, and are an important mechanism for restructuring programs and for insuring that NSF has efficient and innovative programs.

As noted earlier, there are several major reasons for recommending continuation of the overall program at the present time. These include:

- i. the need to provide leading-edge resources to enable world leadership in computational science and engineering,
- ii. the need for advanced educational and training opportunities to provide a cadre of outstanding computational scientists and engineers in high-performance computing,
- iii. the need to advance ways of dealing with issues of scalability, both of applications software and of the underlying systems software.

The Task Force notes that (i) and (ii) provided the original impetus for founding the program in the mid 1980's, and it might be thought that these two factors will always be seen as valid reasons for continuing the program. On the other hand, we note that the

Section 6: Recommendations

cost-effectiveness of mid-range machines compared with the cost-effectiveness of the high-end computing available at major Centers has improved since the mid 1980s. Many experts believe that this trend will continue. This argues that (i) may not always serve as a sufficient driver for the program.

Similarly, given a sufficient population of well-trained computational scientists and engineers at universities, and given a stable underlying technology, (ii) can not serve as a long-term driver for the program. As noted in the recent NRC-HPCC report, (iii) continues to serve as a driver of the program *because* of the recent advent of parallel computation. The Task Force accepts this argument, and notes that it is dependent on the current lack of stability in the underlying technology, and so needs periodic re-evaluation.

Finally, issues of scalability are critical at the present time. The Task Force agrees with the NRC-HPCC report that the Centers program has an important role to play, both in making the relatively new parallel technology available and usable to computational scientists and engineers, and in serving as a “time machine” which enables American scientists and engineers to foresee, and hence quickly use, future developments in the technology. We also agree with the NRC-HPCC report that *this* currently important role for the Centers in pioneering massively parallel computation is not likely to continue.

On a regular basis, there should be a review of the overall program, articulating the purpose and need for the program. It is particularly important that such reviews take into account how rapid changes in the technology may affect the overall need for the program, as well as the balance within the program.

A full competition at the present time will permit realignment of the program by:

- providing an incentive for creative new ideas and commitments to the goals of the program,
- allowing broadening of the high-end base for computational science and engineering by encouraging and enhancing partnerships among a variety of high performance computing Centers with leading-edge sites, and
- encouraging the coordinated development of high-end resources on university campuses.

Only with the past help of ARPA has the program been able to acquire high-end parallel systems at all four Centers. Without increased budgets or newly emerging partnerships, it is unlikely that NSF can maintain four sites in a world leadership role. Thus, a new competition offers an appropriate way to migrate to a smaller number of leading-edge sites, capable of maintaining the nation’s ability to do world leading computational science and engineering. It is our expectation, that at current NSF budget levels, and absent new outside resources, there will be a reduction in the number of leading-edge sites to effect the benefits of the Task Force recommendations.

6.5: Support of Research at the Centers

Recommendation: The Centers program should continue to support need-based research projects in support of the program’s mission, but should not provide direct support for independent research.

Section 6: Recommendations

Having staff at the Centers who are experienced and knowledgeable both in the development and in the application of the most advanced hardware and software is a clear advantage, not only to the users of the Centers, but to users of high-performance computing nationwide. Currently, the Centers provide little or no direct support for independent research efforts of the staff from base NSF funding. The Task Force believes that this practice should continue.

There are two mechanisms that have worked well in the past and that should be encouraged:

- Staff should become involved in specific research projects that are necessary to improve services to external users (i.e., need-based research projects in support of the Centers mission)
- Independent of the Centers base funding, staff should be encouraged to submit competitive proposals individually or on a collaborative basis to other programs and, if funded, participate on a non-interference basis with other assigned duties.

In addition, the institutions at which the NSF Centers are located should be free to compete for center and group research funding from NSF and other sources, but such competitions should be decoupled from the basic Center's support obligations and duties. The Task Force believes that these mechanisms should be sufficient to keep a talented staff interested and up-to-date.

6.6: Allocation Process for Computer Service Units

Recommendation: NSF should increase the involvement of the directorates in the process of allocating service units at the Centers.

The Task Force has examined the current allocation process and is convinced that it has worked quite well. Nevertheless, particularly in times of tight budgets, it is important to the overall program and to the long-term health of computational science and engineering that NSF take steps to improve the merit review process, particularly for large allocations. It is important to establish better mechanisms to involve program staff in the allocation process. Program staff needs to understand computing technology across the full range of computationally capable machines, from workstations to the highest end, and thus be able to better evaluate the computational needs of their grantees. The merit review process can also be further enhanced by the information and insights available from program staff dealing with currently funded NSF projects that will be impacted by the allocation (or non-allocation) of computational resources at the Centers.

6.7: NSF Leadership in Interagency Planning

Recommendation: NSF should provide leadership in working toward the development of interagency plans for deploying balanced systems at the apex of the computational pyramid and ensuring access to these systems for academic researchers.

With the recommended configuration of leading-edge sites and affiliated partners, the NSF Centers program will continue to be a major player in terms of its technical expertise and, possibly, in terms of its computing capability within the overall matrix of federally supported supercomputing activities. As such, it is important that NSF management take

Section 6: Recommendations

a strong and continuing leadership position in shaping shared investments accessible to a wide range of academic researchers, at the highest end of the pyramid. The Task Force believes that continued interagency planning discussions will benefit significantly from the NSF expertise and perspective and, importantly, that NSF should work for an appropriate level of access by top academic researchers to very high-end computing capabilities, even when the highest end capabilities are justified primarily on the basis of mission specific requirements of other agencies.

The Blue Ribbon Panel on High Performance Computing called for a similar leadership role for NSF, but the recommended planning effort was focused on achieving the near term goal of a balanced teraflop system. The present recommendation anticipates a longer term need for NSF leadership at the apex of the pyramid, and goes beyond specifying such a detailed level of capability. (As the recent NRC-HPCC Report pointed out, “the teraflop machine was intended as a direction, not a goal.”)

Appendix A: Task Force on the Future of the NSF Supercomputer Centers Program

A.1: Charge to the Task Force

The Task Force is being called together during NSF's 1995 Fiscal Year to advise NSF on several important issues related to the review and management of the NSF Supercomputer Centers program.

NSF is asking the Task Force to analyze various alternatives for the continuation, restructuring, or phase-out of NSF's current Supercomputer Centers program, or the development of similar future program(s), and to make recommendations among the alternatives.

In making the recommendations the Task Force should consider:

- a. How to best meet future needs of the science and engineering research communities for high-end computational resources in support of computational science and engineering.
- b. The appropriate role for NSF and any recommended program in:
 1. facilitating access to leading-edge technologies, including parallel and distributed computation, in scientific and engineering applications;
 2. interacting with vendors in developing hardware and software for high performance systems for scientific and engineering applications; and
 3. working with industrial users in understanding leading-edge high performance computing and communications technologies.
- c. The potential needs of more information intensive users (National Information Infrastructure, NII) as well as high end computational users (High Performance Computing, HPC).
- d. The appropriate role for the Centers in fostering interdisciplinary and intradisciplinary collaborations.
- e. The appropriate educational role of any recommended program for:
 1. pre-college and undergraduate education
 2. graduate education
 3. postdoctoral education
 4. more mature researchers needing training/orientation in leading-edge high performance computation and communications technologies
 5. industrial users
- f. The appropriate range of potential grantees and suppliers in any recommended program. This includes the appropriate role for leverage of NSF program funds by interacting with partners such as:
 1. other federal agencies
 2. state agencies or centers
 3. technology vendors
 4. universities
 5. industrial users
 6. other sources as appropriate (including possible non-U.S. partners)
- g. Expected budget realities for the first five years of any recommended program.

Appendix A: Task Force on the Future of the NSF Supercomputer Centers Program

Expected Milestones:

Informal oral progress reports to Committee on Programs and Plans of the National Science Board on a regular basis.

Formal report giving advice to internal NSF program committee in June or July, to give staff time to prepare a detailed program description to present to the board no later than November, 1995.

Appendix A: Task Force on the Future of the NSF Supercomputer Centers Program

A.2: Membership of the Task Force

Arden L Bement, Jr.
Basil S. Turner Distinguished Professor of Engineering
Purdue University

Edward F. Hayes -- **Chairman**
Vice President for Research
The Ohio State University

John Hennessy
Professor and Chair, Computer Science Department
Stanford University

John Ingram
Schlumberger Research Fellow
Schlumberger, Austin

Peter A. Kollman
Professor of Chemistry and Pharmaceutical Chemistry
UC San Francisco

Mary K. Vernon
Professor of Computer Science
University of Wisconsin

Andrew B. White, Jr.
Director of Advanced Computing Laboratory
Los Alamos National Laboratory

William A. Wulf
AT&T Professor of Engineering and Applied Science
University of Virginia

NSF Staff (non-voting)

Paul Young
Assistant Director-Computer and Information Science & Engineering (CISE)

Bob Voigt
Acting High Performance Computing and Communications (HPCC) Coordinator-CISE

Nathaniel Pitts
Director - Office of Science and Technology Infrastructure

Support Staff

Robert Borchers
Director, Division of Advanced Scientific Computing

Richard Kaplan
Director, Supercomputer Centers Program

B: Past and Future Impact of Computational Science and Engineering

The ability to precisely determine the value of ongoing research is notoriously difficult. Nonetheless, some judgment must be made in the overall allocation of resources. This appendix contains the findings of the Task Force in four areas that are relevant in making this judgment. These areas are:

- paradigm shifts enabled by computation,
- quality of the researchers involved in the program,
- testimonials by distinguished scientists, and
- material on other accomplishments of the Centers program.

B.1: Examples of Paradigm Shifts

In the 10 years of their existence, the Centers have fostered fundamental advances in our understanding of science and engineering, enabled new research which could have been done in no other way, expanded the use of high-end computing in new disciplines, enabled the major paradigm shift to the acceptance of computational science as a full partner in the scientific method, and facilitated the education of a new generation of computational scientists and engineers in support of that shift.

The sections below give examples of some of the scientific and engineering areas where computational models make significant impact on a field of research. Appendix E of the NSF Blue Ribbon Panel Report¹⁷ also contains a discussion of this topic.

Several common themes emerge from these examples of the impact of supercomputing. First, the rapid growth of supercomputing together with its availability to the research community during the past ten years have enabled computational science and engineering to contribute to significant new advances in a wide set of scientific and engineering fields. Second, high performance computing is making it possible to perform complex simulations in three dimensions, rather than just two. This important shift has dramatically enhanced the usefulness of computational approaches. Third, supercomputer-based simulations have combined multiple disciplines and different physical phenomena to yield new scientific discoveries and understanding. Last, increases in supercomputing capability and advances in computational techniques are beginning to enable computer-based simulations to predict new science and make new discoveries.

B.1.1: Cosmology

A number of exciting theoretical and observational advances have placed theoretical cosmology on a firm scientific footing, moving the field to one with well defined physical theories which make testable predictions. Numerical algorithms which can accurately simulate the formation of cosmological structures such as galaxies and clusters of galaxies starting from primordial initial conditions were developed and refined and can now be combined into predictive numerical codes. More recently, supercomputers have finally become powerful enough and with sufficient memory to begin modeling the universe in full 3D plus time rather than in 2D as done a decade ago (See Ostriker

¹⁷ From Desktop to Teraflop: Exploiting the U.S. Lead in High Performance Computing, edited by Lewis Branscomb, October, 1993

B: Past and Future Impact of Computational Science and Engineering

testimonial in section B.3). Codes can also now be tested on departmental machines, whether scalar, vector or truly parallel (using for example PVM) and then run on center machines at a scientifically interesting level.

B.1.2: Environmental Modeling

The U.S. spends in excess of \$30 billion/year on air pollution controls and, despite this expenditure, progress towards meeting air quality goals has been very slow. Work carried out using resources at the Pittsburgh Supercomputing Center and at the National Center for Supercomputing Applications led to a clearer understanding of the processes responsible for the formation of photochemical air pollution or smog. The necessity of including elementary physical and chemical processes in the air quality models has been recognized for many years, but the complexity of these processes and the significant differences in time scales, temperatures and physical dimensions present major computational challenges.

This work would not have been feasible without the computing capacity to carry out very detailed simulations of the atmospheric dynamics and chemistry over multi-day periods. The results, which were enthusiastically endorsed in the National Research Council Study "Rethinking the Ozone Problem", led to a change in the Clean Air Act and are now a routine part of the design of air pollution control strategies throughout the world.

B.1.3: Protein Folding

One of the most challenging problems in simulations of complex biological systems is to predict correctly the global minimum conformation. The ability to be able to do this would have enormous implications in biotechnology and medicine, with great benefits to human health. However, the problem is extraordinarily difficult because such macromolecules have a tremendous number of conformations. Even a small protein of 100 amino acids has of the order of 10^{20} possible conformations which need to be considered. There are many promising approaches to use simplified models to give some insight into the protein folding problem, but it is likely that to fully solve it, models of the protein, including all the atoms, will have to be employed at some stage in this process. To give one some sense of where a brute force approach to this problem stands, one can now simulate protein dynamics with all atom representations for a few nanoseconds. On the other hand, it takes real proteins milliseconds to seconds to fold in the laboratory. Thus, currently our computer power and tools are about 6 to 9 orders of magnitude too limited to make accurate apriori predictions.

Nevertheless, one should not underestimate the progress that has been made in the last few years, with very important contributions through simulations carried out at the supercomputer Centers. Each order of magnitude in computer power has let one improve the force field/energy representation that is crucial for ultimately solving this problem. It also gives more flexibility for the development of short cuts that might circumvent the brute force approach to the problem.

Thus, there is continued important progress in accurately simulating the structure, dynamics, and folding of macromolecules and thus elucidating important aspects of their function during the next decade. A wide variety of approaches will be needed to make

B: Past and Future Impact of Computational Science and Engineering

progress on this, with an absolutely crucial element being access to the highest end of the computational spectrum (see also the testimonials in Section B.3).

B.1.4: Condensed Matter Physics

The impact of high performance computing on theoretical condensed matter physics in the last ten or fifteen years has been remarkable. In the late 70's computational approaches were of relatively minor importance, used by a few pioneers to obtain impressive but isolated results. Now computational approaches are arguably the driving force for most of the field.

Consider, for example, the theoretical efforts to understand high-temperature (high-Tc) superconductivity. The entire field of high-Tc superconductivity extends back less than a decade, yet the shift in theoretical efforts between the inception of the field and the present are tremendous. Shortly after the discovery of these remarkable compounds in 1986, a rather large number of theories were proposed to explain the effect. At the same time, numerical simulations began to be used on several models related to high-Tc. In the beginning, theories were tested mostly against available experimental data. However, many of the experimental techniques proved less accurate than hoped because they were highly susceptible to impurities in the crystals and to surface effects. Meanwhile, thanks to improvements both in algorithms and in computational facilities, the numerical approaches have improved enormously. Today, the numerical simulations are of equal importance with experiments in testing theories. For example, most of the properties of the Hubbard model, the leading model in high-Tc studies, cannot be calculated analytically in any reliable fashion. During the late 1980's, working mostly at the San Diego Supercomputer Center, the basic magnetic properties of the model were mapped out. Today work continues on this model to determine whether it exhibits superconductivity, or whether additional terms must be included.

Numerical simulations, because of the tremendous increase in the quality of the data they provide, are now also stimulating new theories. As an example, the recent high-Tc theory of Dagotto and coworkers is based almost entirely on numerical results. These calculations were primarily exact diagonalizations of Hubbard and other models. Without central supercomputer facilities, calculations such as these would be impossible. The very difficult problem of high-temperature superconductivity is not yet solved, but when it is, it will almost certainly be largely due to numerical simulations.

B.1.5: Quantum Chromodynamics

Quantum Chromodynamics (QCD) has been accepted as the fundamental theory of the strong interactions of particle physics for some time. In principal this theory should allow one to calculate some of the most important quantities in nature, and to test ideas about the fundamental force laws. Work in this area is directly tied to major experimental programs in high energy and nuclear physics. However, it has proven extremely difficult to extract many of the predictions of QCD. At present the only promising way of doing so is through large scale numerical simulations that tax the power of the largest available supercomputers.

B: Past and Future Impact of Computational Science and Engineering

The NSF Supercomputer Centers have enabled a great deal of progress in the numerical study of QCD. Work has included the development of new computational techniques, and the detailed study of a variety of problems including the behavior of nuclear matter at high temperatures, the calculation of the masses of strongly interacting particles, and the weak decays of these particles.

The study of high temperature QCD provides one example of the progress that has been made at the Centers, and the work that remains to be done. Under ordinary laboratory conditions one does not directly observe quarks and gluons, the fundamental entities of QCD. Instead one observes their bound states, protons, neutrons, and hosts of short lived particles produced in high energy accelerator collisions. However, at very high temperatures, one expects to find a transition to a new and as yet unobserved state of matter consisting of a plasma of quarks and gluons. Work at the Centers has provided an accurate determination of the temperature at which this transition occurs, and has provided insight into the nature of the transition and the properties of the quark-gluon plasma. This information will be important for the interpretation of heavy-ion collisions being planned to detect the plasma, but much more extensive calculations are needed for a detailed determination of the equation of state of the plasma.

With present calculational techniques definitive calculations of many of the important properties of QCD will require computers capable of sustaining several teraflops. New calculational approaches presently being tested may reduce this estimate, but it seems clear that QCD calculations will strain the capabilities of the more powerful supercomputers for years to come. Thus, the NSF Centers will continue to play a vital role in the advancement of this field.

B.1.6: Device and Semiconductor Process Simulation

As the feature sizes of semiconductor devices in integrated circuits are reduced to the submicron region, it is not possible to understand the detailed physics with simplified models because of a variety of non-linear effects. Such effects are not simply interesting from the pure research point of view, but have extremely relevant implications for the reliability of devices and the design of high density structures. Since the cost of production lines for a new semiconductor process exceeds a billion dollars, accurate simulation of semiconductor devices is key to maintaining leadership in the semiconductor industry.

In the late 1980's and early 1990's, vector supercomputers reached a level of capability (both in cycles and memory) that made particle simulations based on a Monte Carlo approach viable. Such approaches, which work from first-principles physics, are needed to deal with the complex nonlinear behavior that occurs in submicron devices. The availability of supercomputers with sufficient computational power and memory made such simulations tractable. Both the electronic device research community and the semiconductor industry now rely on such advanced simulation tools to understand the behavior of new devices.

The complexities of modern submicron fabrication increasingly require that simulations be done in 3-D rather than just 2-D. Indeed, 3-D simulation has been able to show effects in device structure that were not observable based on purely 2-D simulation. To

B: Past and Future Impact of Computational Science and Engineering

accommodate 3-D simulation, however, requires a significant increase in both memory and computational power. Modern multiprocessor machines with shared memory are the ideal platforms for such simulations and can enable the development of full-scale 3-D simulation for submicron devices. This capability is required today in research laboratories engaged in developing 0.1 micron devices and will be vital to the semiconductor industry in the next few years. Accurate simulation of 3-D submicron structures is likely to be one of the most demanding applications in the future, requiring machines with up to teraflops of computational capability and terabytes of memory.

Perhaps the most exciting potential development in device and technology computer aided design is the drive towards full computational prototyping of a new semiconductor process before the process is physically realized in a integrated circuit fabrication facility. Such a capability could dramatically decrease the time it will take to bring up and tune a new semiconductor process. Computational prototyping of a semiconductor process involves not only device simulation, but also the simulation of a series of complex manufacturing processes. It requires the integration of multiple simulation techniques and science from different disciplines. Such a capability could shorten the time and reduce the cost of developing new semiconductor processes. Fully achieving this vision will require computational capabilities beyond the teraflop range.

B1.7: Seismology

In reality the paradigm shift in both exploration and earthquake seismology, began some time ago. As the oil industry discovered the enormous advantage to be gained from three dimensional subsurface description, it found the courage to make the substantial commitment necessary to undertake the acquisition of the rather enormous quantities of data required. This can easily amount to a terrabyte of raw data for a fair sized section. The processing of this data has often required six to nine months, a virtually unacceptable commercial delay. However, the gain in accuracy of the description has more than paid for the expense and delay and there is hope that new computational methods coupled with the evolving computing and storage resources can make the technique more feasible.

Similar considerations hold for earthquake seismology and are, if anything, more stringent. The scale of the volumes considered are larger and although the frequencies are lower, the need for three dimensional descriptions is even greater. Furthermore, the key to the future in both of these areas is in more accurate description of the media and the equations which govern the phenomena. In addition both applications can benefit from time lapse techniques which again multiplies the amount of data to be acquired. These are inverse problems on a grand scale and ultimately must be reformulated to incorporate the advances in mathematical simulation currently underway. This is beyond the scope of current systems. Such problems cannot be tackled without significant advances in the handling of massive quantities of data, high speed communications, time dependent visualization of three dimensional data sets, and the solution of very large systems of partial differential equations. There is a very significant role for the NSF centers program to play in this evolution, if the facilities have sufficient capacity.

B: Past and Future Impact of Computational Science and Engineering

B.1.8: Turbulence

This is a grand challenge problem because of the difficulty of the fluid dynamics and the ubiquity of the phenomena. There is now a broad consensus that major discoveries in key applications of turbulent flows will be within grasp of teraflop class computers.

High performance computer systems have already enabled a radical step forward in the modeling of turbulent flows for real applications, ranging from the typical interior channel and pipe flows of mechanical engineering to the estuary and coastal flows which are the province of environmental engineers. The improved accuracy and resolution of the models have allowed e.g., simulations of the San Francisco estuary which have improved the understanding of phenomena like salinity variations and tidal flows.

The use of computational fluid dynamics (CFD) to study turbulence not only has led to new understandings of physics, but also, perhaps more than any other field, has inspired major advances in numerical techniques. Refinements in techniques such as adaptive methods, unstructured grids, preconditioned iterative methods, mesh generation, and spectral methods have been motivated by CFD and have enabled computational advances in many other fields.

Like simulations of many other physical phenomena, large eddy simulation of turbulence is based on a well-founded underlying model of the physics, and accurate solution of partial differential equations over four-dimensional fields (three space, and time). Such calculations, requiring the most powerful computing systems, are being carried out at NSF and NASA supercomputer centers.

B.2: Distinctions and awards accorded some users of the NSF Supercomputing Centers

In addition to the evidence on the quality of the research enabled by the Centers discussed in the Report and in Appendix E, the Task Force asked the centers to compare their “faculty” user list with other sources of information in several categories: Members of the National Academies of Science and Engineering, Nobel Laureates, and other awards or recognition. The summary of this investigation is shown below:

Recognition	Number of accounts
Nobel Laureates	9
Members of the National Academy of Science	89
Members of the National Academy of Engineering	66

B.3: Testimonials from Distinguished Scientists who have used the Centers

In order to get opinions from some scientists who have used the Centers to further their own research, the Task Force asked a number of them to write about their experiences, opinions of the program and personal observations about computational science. Excerpts from these letters are reproduced below.

B: Past and Future Impact of Computational Science and Engineering

Jeremiah P. Ostriker
Cosmology
Princeton University

Fifteen years ago, my view was that not only had supercomputing not made a positive scientific impact on astrophysics and cosmology, but that its net impact was negative because so many talented young people became absorbed by the "black hole" of computing technology and thus were lost to science. Today, my view is the exact opposite. What made me change my mind? Several things. First and foremost, a number of exciting theoretical and observational advances have placed my field of theoretical cosmology on a firm scientific footing. Cosmology has moved from something akin to theology to a hard science with well defined physical theories which make testable predictions. Second, numerical algorithms which can accurately simulate the formation of cosmological structures such as galaxies and clusters of galaxies starting from primordial initial conditions were developed and refined and can now be combined into predictive numerical codes. Third, and more recently, supercomputers have finally become powerful enough and with sufficient memory to begin modeling the universe in full 3D rather than in 2D, as we did decade ago. Fourth, the NSF supercomputing centers, in addition to providing the raw cycles, also provide the human infrastructure of technically trained people who are responsive to the needs of academic scientists such as myself. Finally, codes can now be tested on departmental machines, whether scalar, vector or truly parallel (using for example PVM) and then run on NSF center machines at a scientifically interesting level. There are universal languages, whereas in the past one needed very special tools to approach "supercomputers", the acquisition of which almost disqualified one for normal scientific life.

The "computational culture" at the NSF centers, specifically at the Pittsburgh and Illinois centers, underpins the HPCC grand challenge project in numerical cosmology which I lead, and allows our six institution consortium to cross-compare, integrate, and scale up our cosmological models to a level of physical complexity previously unheard of. With the aid of supercomputers, our models are just now crossing the threshold of realism to where we can begin to test various specific theories of structure formation within the Big Bang framework. ... As observations and computers continue to improve, we expect to be able to rule out a number of competing models with a high degree of confidence.

Gregory J. McRae
Joseph R. Mares Professor of Chemical Engineering
MIT

... The Centers program, by almost any standard, is one of the most outstanding success stories of the National Science Foundation. During their relatively short existence the Centers have played an absolutely vital role in facilitating new science and engineering, training thousands of researchers in computational science and, through outreach programs, have contributed to strengthening the U.S. industrial base. An example is appropriate.

One of the points that often confuses the present debate is to say that it would be even much more cost effective to use the same resources to provide individual researchers with

B: Past and Future Impact of Computational Science and Engineering

powerful workstations. After all the argument goes that, with the rapid increases in chip speeds and the equally dramatic drop in costs, if we just wait we will have on our desktop all the power needed. There is much more to problem solving than just faster chips. There are such technical issues as memory costs, I/O bandwidth, file storage, software support, system maintenance etc. These costs typically dwarf the purchase price of a basic workstation. What is often ignored, and an important role that is currently being played by the Centers, is supplying specialist knowledge, training and personnel support. The Center staffs are a veritable gold mine of knowledge and, more importantly, are accessible to the community as a whole. Individual researchers cannot afford to support experts in data base management, graphics and communications that form part of the team needed to attack many of the "grand challenge" problems.

My greatest concern however is that we must not lose the very valuable infrastructure and staffing support system that has been built. These resources are critical to the scientific community and must be maintained if we are to tackle large scale problems in science and engineering. If the Centers are to attract, and retain the very best people, they must be seen as a stable place to build a career. The most important question facing the panel should not be how to cut, but how to enhance and expand the effectiveness of the present system.

To summarize, in my view the NSF Supercomputer Centers program has been a brilliant success -- the Centers are an indispensable resource to the Nation and should be preserved at all costs. There are an investment in our future, the next generation of scientists and the economic competitiveness of the nation.

C. Roberto Mechoso **Professor Department of Atmospheric Sciences** **University of California, Los Angeles**

I have learned that you are collecting the experiences of major investigators who have used NSF center resources, and I am delighted to respond.

As you are no doubt aware, meteorological and climatological modeling of the type pioneered here at UCLA laid the foundation for the now very considerable predictive skills of our National Weather Service. The extension of these methods to the modeling of seasonal and interannual climatic variability, with the objective of attaining similar skill, is imperative for guiding plans that affect the future of U.S. agricultural and energy production, trade, and commerce.

Our research group uses the computational facilities of the National Center for Atmospheric Research (NCAR). Since we are always in need of computer resources to conduct the complex climatological experiments that our most sophisticated models would permit, we became one of the first research groups to take advantage of SDSC resources.

We quickly found out, however, that to be granted computer resources at SDSC amounted to a great deal more than hardware alone. From the beginning, SDSC maintained a professional, full time consulting team. The backup supplied by these consultants made it possible to optimize codes that had originally been written for much

B: Past and Future Impact of Computational Science and Engineering

less powerful machines. This was important, not only to the efficiency and clarity of our codes, but also to our continued ability to obtain computational resources. The SDSC Allocation Committee has also from time to time suggested ways in which our group might both improve code performance and obtain better scientific output, and their independent review of our computational program has often been of value.

The research collaboration with SDSC was a particularly good experience for me and my colleagues, and for our postdoctoral researchers and students. It advanced the state of the science of climate dynamics, and it did more: it enabled us to see forward to the coupling not only of ocean and atmosphere models, but also to the coupling of such models with models of atmospheric chemical processes.

Herbert A. Hauptman

Nobel Laureate

President Hauptman-Woodward Medical Research Institute

As a practicing crystallographer trying to develop improved methods for molecular structure determination, essential for the rational design of drugs, I have found the Pittsburgh Supercomputing Center to be of the greatest importance. Attendance at schools sponsored by PSC have proved to be essential in the development of improved techniques of structure determination and initiating collaborations with potential users. In addition, the availability of the parallel supercomputers has facilitated the development of more powerful methods and has been crucial to my research over the past several years.

Andrew McCammon

Joseph E. Mayer Professor of Theoretical Chemistry and Pharmacology University of California at San Diego

Senior Fellow San Diego Supercomputer Center

Referring to comments he made at the 1994 Smithsonian awards dinner.

... I focused my brief remarks at the program on NSF's key role in shaping my own career: an NSF-sponsored summer program for high school students at the Scripps Institution of Oceanography in 1964 (31 years ago!!!), predoctoral and postdoctoral fellowships, my first major Federal grant as a struggling new assistant professor, up through the present.

Most of the attendees at the program seemed to be from industry, and I think they were interested in this example of how what started out as "pure research" - an inquiry into the nature of motions in proteins - led to the development of tools that have put promising candidates for the treatment of a number of diseases into clinical trials I also argued briefly for the continuing importance of high-performance computing. Indeed, I look forward to working even more closely with NSF's ASC arm as a Senior Fellow of SDSC.

Although we're still getting our group's feet on the ground after our recent move to La Jolla, I have gotten involved with hardware planning at SDSC, and I'm excited about their forward-looking approach to data- and numerically-intensive computing.

B: Past and Future Impact of Computational Science and Engineering

Mary Ostendorf
Department of Physics
University of Illinois

The use of the computational facilities at NCSA was of inestimable value to me in my theoretical studies with Philippe Monthoux of the mechanism for high temperature superconductivity and the pairing state to which it gives rise. NCSA support enabled us to explore in depth the consequences of a momentum-dependent magnetic interaction between planar quasi-particles and the role played by strong coupling corrections. We found that taking the momentum dependence of the interaction into account was crucial for obtaining superconductivity at high temperatures and that when this was done the strong coupling corrections which otherwise would have proved fatal for the theory were of manageable size. On the basis of these calculations we were able to predict unambiguously that a magnetic mechanism would give rise to $d_{x^2-y^2}$ pairing, a state which has subsequently been confirmed in many different recent experiments.

Arthur J. Freeman
Morrison Professor of Physics
Northwestern University

... Without any doubt they(The Centers) have been a major force in giving the U.S. leadership in vast areas of Computational Science and Engineering. As a member of several committees that led to the establishment of the Centers, I was very much aware of the enormous need for computational facilities in the U.S. and the sorry state of affairs that existed prior to the establishment of the NSF Centers in our scientists having to go to Europe to do their work. It is very clear that the Centers have impacted strongly on both basic science and applications. They have provided supercomputing facilities in a very cost effective manor, and have more than paid for themselves in terms of the development of both basic science and the resultant industrial applications.

The availability of supercomputers at the NSF Centers radically changed the science that I and my research group were able to perform... . In both we have been highly successful thanks to the computational facilities provided by the Centers.

The one problem with the existing Centers is that their very success has put increasing demands on their resources without concomitant increases in their funding. While this is a difficult thing to propose in this era of funding cuts, I believe that funding in fact should be increased for them as a cost effective way of increasing their impact for technology and industrial applications.

Steven R. White
Department of Physics and Astronomy
University of California, Irvine

The impact of high performance computing on theoretical condensed matter physics in the last ten or fifteen years has been remarkable. In the late 70's computational approaches were of relatively minor importance, used by a few pioneers to do some impressive things, but still a very small-time operation. Now computational approaches are arguably the driving force for most of the field.

B: Past and Future Impact of Computational Science and Engineering

I have mentioned only examples from my sub-field, but access to supercomputer time at the NSF centers has been equally important in other areas of theoretical condensed matter physics. For example, some of the biggest users of supercomputer time at the NSF centers use density functional theory to predict the properties of a wide variety of materials. The group of Cohen and Louie at Berkeley (who predicted a new material that may be harder than diamond, for example) and Joannaopolous' group at MIT are leading examples. The work of these groups would virtually cease without the NSF centers. Access to supercomputer time has now become an indispensable tool for much of the most important work in theoretical condensed matter physics.

B.4: Additional Material on Centers' Program Accomplishments

The division of Advanced Scientific Computing (ASC) of Computer and Information Science and Engineering (CISE) directorate, together with the four NSF supercomputing centers, prepared a document as supporting material for the presentation of the Centers Program renewal. This document, entitled "High Performance Computing Infrastructure and Accomplishments" was an extensive listing of the major accomplishments in and by the centers in five areas: technology, education, outreach, science and engineering and MetaCenter concept recognition

This document highlighted significant accomplishments of the Centers and their users over the life of the program, with a paragraph of explanatory text. The document itself is available on the world wide web with a URL

<http://www.cise.nsf.gov/acir/hpc/>

but more extensive presentations are at each of the centers web sites, and can be accessed most easily from the following URL's:

<http://www.ncsa.uiuc.edu/> <http://www.tc.cornell.edu/>
<http://pscinfo.psc.edu/> <http://www.sdsc.edu/>
<http://www.tc.cornell.edu/Research/MetaScience/>

Additionally, the centers web pages are attracting pointers from many other sites, increasing the outreach of the program.

While these web pages were started to organize documentation and account information about each center, they have grown to present highlights of research results, science outreach information, and in the case of the Cornell Theory Center, even its quarterly project report to ASC.

Appendix C. Blue Ribbon Panel on High Performance Computing and the NSF Response.

Following the renewal of four of the five NSF Supercomputer Centers in 1990, the National Science Board (NSB) maintained an interest in the Centers' operations and activities. In 1992 at the request of the NSB, the Director of NSF appointed a blue ribbon panel "... to investigate the future changes in the overall scientific environment due [to] the rapid advances occurring in the field of computers and scientific computing." The resulting report, "From Desktop to Teraflop: Exploiting the U.S. Lead in High Performance Computing," was presented to the NSB in October, 1993.

The Report points to the Foundation's accomplishments in the decade since it implemented the recommendations of the Peter Lax Report on High Performance Computing (HPC)¹⁸ and established the Supercomputer Centers. These Centers have created an enthusiastic and demanding set of sophisticated users who are making fundamental advancements in their scientific and engineering disciplines through the application of the rapidly evolving HPC technology. Other measures of success cited include the thousands of researchers and engineers who have gained experience in HPC, and the extraordinary technical progress in realizing new computing environments.

The Report notes that, through the NSF program and those of sister agencies, the U.S. enjoys a substantial lead in computational science and in the emerging, enabling technologies. It calls for the NSF to capitalize on this lead, which not only offers scientific preeminence but also the associated industrial lead in many growing world markets.

The Report puts forth four Challenges, summarized below, that address the opportunities brought about by the success of the program. These Challenges and the accompanying recommendations were based on an environment with the following two characteristics that have since changed:

- Parallel systems were just being introduced at the Centers and elsewhere. Because of uncertainties surrounding systems software and architecture issues made it unclear how useful these systems would be for scientific computing, the report recommended investment in both the computational science and the computer science issues in massively parallel computing.
- The report assumed that the administration and the Congress would adhere to the stated plan of the HPCC budget, which called for a doubling in five years.

Challenge 1: How can NSF, as the nation's premier agency funding basic research, remove existing barriers to the rapid evolution HPC, making it truly usable by all the nation's scientists and engineers?

Challenge 2: How can NSF provide scalable access to a pyramid of computing resources, from the high performance workstations needed by most scientists to the critically needed teraflop-and-beyond capability required for solving Grand Challenge problems?

¹⁸ "Large Scale Computing in Science and Engineering", edited by Peter Lax, 1982

Appendix C. Blue Ribbon Panel on High Performance Computing and the NSF Response.

Challenge 3: The third challenge is to encourage the continued broadening of the base of participation in HPC, both in terms of institutions and in terms of skill levels and disciplines.

Challenge 4: How can NSF best create the intellectual and management leadership for the future of HPC in the U.S.?

These Challenges and the accompanying 14 recommendations could be summarized as calling for a broad based infrastructure and research program that would not only support the range of computational needs required by the existing user base, but would also broaden that base in terms of the range of capabilities, expertise, and disciplines supported. Some of the key recommendations include:

- The NSF should take the lead in expanding access to all levels of the pyramid of computing resources.
- The NSF should initiate an interagency plan to provide a balanced teraflop system, with appropriate software and computational tools, at the apex of the computational pyramid.
- The NSF should assist the university community in acquiring mid-range systems to support scientific and engineering computing and to break down the software barriers associated with massively parallel systems.
- The NSF should retain the Centers and reaffirm their mission with an understanding that they now participate in a much richer computational infrastructure than existed at their formation. This included use of ever more powerful workstations and networks of workstations.

As a follow up to the Blue Ribbon Panel Report, the NSF Director established an NSF High Performance Computing and Communications Planning Committee of NSF staff in 1993. In responding to the Panel Report, the Committee was charged with establishing a road map and implementation plan for NSF participation in and support of the future HPC environment. The Committee presented a draft of its report to the Director's Policy Group in March, 1994; a final version of the report was made available to the NSB in February, 1995.

The Committee used the four Challenges put forth in the Panel Report as a basis for its report, and the recommendations contained in the Committee Report were consistent with and supportive of the recommendations in the Panel Report; there were no major areas of disagreement. For example, the cornerstone of the vision put forth in the Committee Report was that by the year 2000 NSF would provide a completely transparent, scalable, interoperable National Computing Infrastructure supporting its research, education and training, and technology transfer activities.

Recommendations in both reports called for a balanced approach to computing infrastructure ranging from workstations up through access to the most powerful systems commercially available. The Supercomputer Centers were viewed as a fundamental ingredient in this infrastructure with continually evolving missions, and both reports called for their renewal without recompetition.

Appendix C. Blue Ribbon Panel on High Performance Computing and the NSF Response.

Both reports also acknowledged the need for strong, continued support of research on computational science technologies such as algorithms and on enabling technologies such as operating systems and programming environments.

Appendix D: The National Research Council HPCC Report

In early 1994, acting through the Defense Authorization Act for FY 1994 (Public Law 103-160), Congress asked the National Research Council (NRC) to examine the status of the High Performance Computing and Communications Initiative (HPCCI). The final report, *“Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure”*, known as “Brooks-Sutherland”, contains a number of recommendations about the HPCC program as a whole, as well as specific recommendations about the NSF Supercomputer Centers. Following a brief summary of the observations and recommendations of the NRC committee, we discuss specific recommendations vis-à-vis the conclusions of this task force.

D.1: Summary of the NRC Committee Observations and Recommendations

The NRC-HPCC report observes that the centers have played a major role in establishing parallel computing as a full partner with the prior paradigms of scalar and vector computing. The report also states that the centers have played an important role in promoting early use of new architectures by providing access to such architectures and by educating and training users. Brooks-Sutherland stated that advanced computation will remain an important tool for scientists and engineers and that support for adequate computer access must be a part of the NSF research program in all disciplines. The Brooks-Sutherland committee avoided recommending the appropriate overall funding level for the centers. Nonetheless, the NRC committee questioned the exclusive use by the NSF of HPCCI- specific funds for support of general computing access, when the computing does not simultaneously help drive the development of high- performance computing and communications technology.

D.2: Recommendation of the NRC Committee

The NRC Committee made one recommendation about the centers.

Recommendation 9. The mission of the National Science Foundation supercomputer centers remains important, but the NSF should continue to evaluate new directions, alternative funding mechanisms, new administrative structures, and the overall program level of the centers. NSF could continue funding of the centers at the current level or alter that level, but it should continue using HPCCI funds to support applications that contribute to the evolution of the underlying computing and communications technologies, while support for general access by application scientists to maturing architectures should come increasingly from non- HPCCI funds.

It is the view of this Task Force that the justification for the centers in the context of the overall NSF program and in the context of the HPCCI program are legitimately different. In the context of the overall NSF program, the Centers are the providers of high-end computing services to the science and engineering research community. Thus, a legitimate case can be made for supplying access to maturing architectures, if such architectures are the current best choice for some applications. From the view of the HPCCI program, the use of mature architectures does not contribute significantly to the development and advancement of HPCCI technologies. The goals of the Centers program within NSF and the goals of the HPCCI program, while sharing many elements can appropriately differ in the choice of resource deployment. Nonetheless, this task force believes, as stated earlier in this report, that under tight budget constraints, priority should

Appendix D: The National Research Council HPCC Report

be given to deployment and use of architectures that both serve the computational needs of the research community and simultaneously help advance HPCCI technology, and so our recommendations are consistent with the spirit of this recommendation.

The NRC committee also recommended an examination of the supercomputer centers program to include identification of:

- Emerging new roles for the centers in supporting changing national needs, and
- Future funding mechanisms, including charging mechanisms and funding coupled to disciplinary directorates.

These issues are addressed in this report. In considering new roles for the centers, the task force concluded that such roles should remain closely affiliated with the primary role as provider of high-end services. In addition, specific recommendations and suggestions have been given for increasing the involvement of NSF's directorates and divisions in the allocation process.

Finally the NRC committee recommended that "the centers, and the researchers who use their facilities, should compete for research funds by the normal means established by the funding agencies. "

As indicated by our fifth explicit recommendation (6.5), the Task Force agrees with this view.

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

During the period when the “Task Force on the Future of the NSF Supercomputer Centers Program” was meeting, substantial amounts of quantitative data were collected to answer questions about the existing centers program. This Appendix contains the data that was deemed germane to the Task Force’s charge. The data collected involved 4 areas:

1. usage history at the NSF Supercomputer Centers,
2. usage patterns of a cohort of “large projects”, including research funding and quality estimates of project leaders,
3. duration of these large projects at the centers, and
4. budget history of the current NSF centers.

A more extensive collection of data in these areas is available at URL:

<http://www.cise.nsf.gov/acir/hpc/>

E.1: Usage Patterns at the NSF Supercomputer Centers

Each of the existing NSF Supercomputer Centers¹⁹ has maintained its database of users since Phase II of the program was started in 1986. In 1989, the importance of usage data in a uniform format was recognized, and the Advanced Scientific Computing Division engaged Quantum Research Corporation (QRC) to analyze the data and ultimately to organize a database of usage for the centers program as a whole. Quantum’s efforts in this area are well documented by regular monthly and annual reports, and most recently, as a World Wide Web document accessible at URL:

<http://usage.npaci.edu/>.

Without repeating information in these web pages, the contents of the database may be summarized as follows:

- Every Project performed at the Supercomputer Centers, including:
 - project title,
 - center used,
 - nominal sponsor of the research, and
 - NSF division code of research area (even when the research was not sponsored by NSF).
- User information:
 - Every User ID (linked to projects),
 - Principal investigator(s) (PI) of every project, and
 - Home institution and department of PI(s).
- Computer used for each project, including its classification as:
 - Vector Multiprocessor, or
 - Parallel Computer

¹⁹The original five centers were (1) The National Center for Supercomputing Applications (NCSA) at the University of Illinois, Champaign-Urbana, (2) The Cornell Theory Center (CTC), (3) The John von Neuman Center (JVNC), a consortium located at Princeton University, (4) The San Diego Supercomputer Center (SDSC), located at the University of California, San Diego and operated by General Atomics, and (5) The Pittsburgh Supercomputing Center (PSC), operated by University of Pittsburgh, Carnegie Mellon University, and Westinghouse Corp.

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

- Normalized usage – converted to an equivalent System Unit (SU) of a 9.5 nsec Cray XMP processor
- Identification of training usage
- Identification of industrial usage (although individual industrial users and projects are not tallied separately).

The centers have provided substantially more in services than computing cycles, but the only data that has been collected has been CPU usage figures – no memory use, I/O requirements, special software access, etc. are recorded. Furthermore, as the machines have developed and machine types have proliferated, the conversion of usage (measure in *time*) to SU has relied on some standardized benchmarks, which may not be relevant for particular applications. The methodology is fully documented at the URL mentioned at the start of this section.

Comparative usage measurements can *help* explain the magnitude of the relative usage, and provide a graphic indication of the growth of the *capacity* of the program.

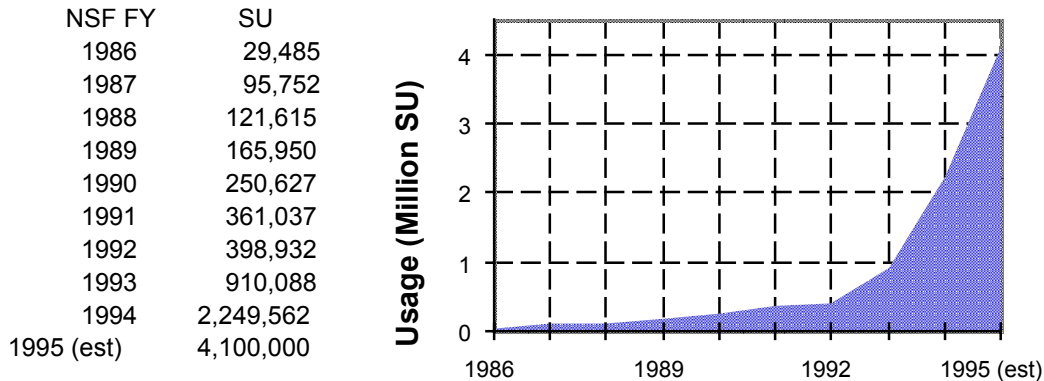


Figure E.1 Usage Increase Over The History of the Program FY 1986-95

Clearly usage grew at a lesser rate (about 30%/yr) from 1986-92 than from 1993 to present (an average of 10%/month, and about 115%/yr). 1993 saw the introduction of the original massively parallel systems, and their potential to deliver *raw* computing cycles that far outstrips vector multiprocessors. However, such systems often require extensive new coding and algorithms to reduce any inefficiencies that can degrade the performance of MPP systems on specific research codes. Simple measures of “raw cycles” does not present the complete capacity story, but since it is the only data available, it is presented with that caveat.

A different pattern emerges when the *number* of users of the centers is examined.

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

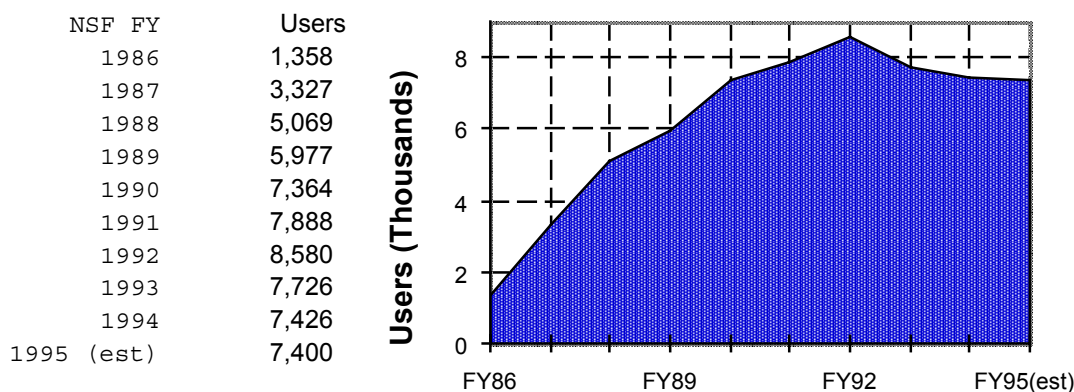


Figure E.2: Number of Users of the NSF SC Centers FY 1986-95

The annual number of users peaked in 1992, and has been decreasing gradually for the past two years. Part of this drop can be explained by an affirmative policy to assist *small* users who do not take advantage of all the features of the Centers' resources to migrate their usage to “workstation” class machines. For many applications, workstations perform the calculations nearly as fast at a much lower cost. Indeed, smaller systems with impressive amounts of computing power have been purchased in increasing numbers by universities and departments over the past several years. In many instances, the wall-clock time for a calculation or a series of test cases to be completed is substantially less for an investigator, compared to a large center where the system is shared with thousands of other users.

The data shown so far demonstrate that there are about 7-8,000 users of the Centers per year. The complete data base has entries for more than 21,000 user names – entries of all users who have had accounts at one of the centers since the data base was established. The number of scientists and engineers “touched” by the centers during their existence, then is closer to 20,000, with the number split between researchers and students.

The largest segment of the usage came from academic researchers. For the most recent complete year (FY 1994), the fraction of usage and users could be broken down as shown in figure E.3.

A small number of users account for most of the machine usage measured. For example, in 1994 4% (332) users accounting for 64% of the computer usage, used more than 1,000 SU of computer time at all the centers. Small users consume resources other than computer time: consultant services, training, etc., or in some cases, appear to have logged in but not to have used any measurable computer time on the supercomputer systems.

In this example the *counts* of users are led by graduate students rather than faculty, who are the leaders in the *usage* category. The sample shown in E.3 counts only those users who used more than 10 SU in 1994, and are probably most representative of *research* users.

In the Annual Report: Annual Summary of Usage Statistics for the National Science Foundation Supercomputer Centers Fiscal Year 1994, substantial additional detail (and

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

figures) are presented on the following selected areas shown in table E.1. In this table, [C] stands for results presented in a Chart, and [T] for results presented in Tabular form.

Research usage has been tracked to the discipline that is home of the major portion of the research, and presented in figure E.4. The methodology is slightly different than that used in the reports cited above, but has the advantage that every project is linked to an NSF directorate or division. The presentation includes only *research* usage, and omits training, system functions, etc. The MPS directorate is generally on the left semi-circle and accounts for 57% of the centers' usage.

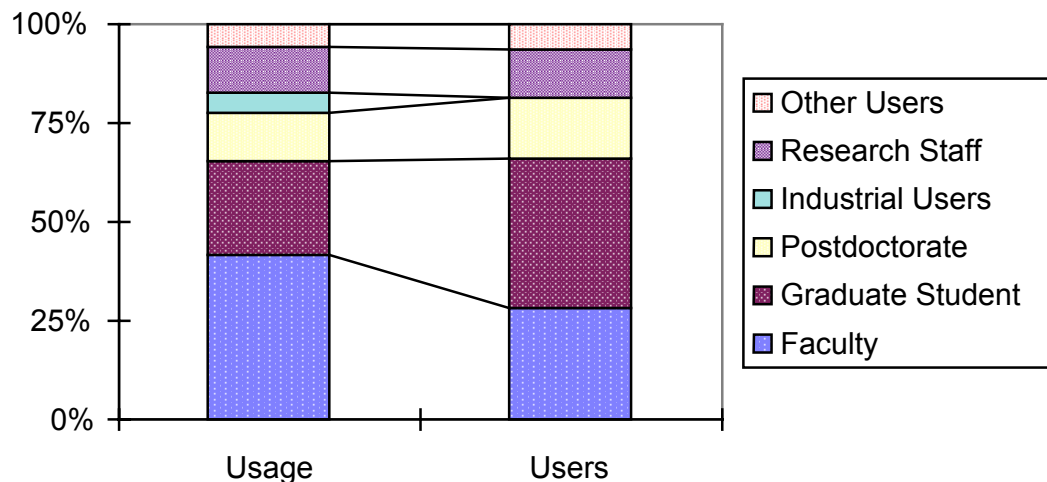


Figure E.3: Academic Status of Users and their Usage for FY 1994²⁰

²⁰ Industrial usage is all reported to the data base linked to one name: "Industrial User". Thus, there is *no* count in this system for the actual number of users.

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

General Subject	Detailed Information
1. Summaries for all Centers and detail for each center	Normalized CPU usage by FY [C] Active Users by FY [C] Active Grants by FY [C] Summary of NSF SC Usage [T]
2. Normalized CPU usage and active users by center and annual CPU usage level FY 1990-94 (All centers and detail for each center)	Distribution of active users and usage by annual CPU usage level: FY94 [C] Distribution of normalized CPU usage: FY94 [C] Distribution of normalized CPU usage and active users by normalized CPU usage level: FY 1990-94 [T]
3. Normalized CPU usage and active users by center and academic status: FY 1990-94 (All centers and detail for each center)	Normalized CPU usage by academic Status: FY 1994 [C] Active users by academic status: FY 1994 [C] Distribution of normalized CPU usage and active users by academic status: FY 1990-94 [T]
4. Normalized CPU usage by center and state: FY 1990-94 (All centers and detail for each center)	Normalized CPU usage by state: FY 1994 [C] Normalized CPU usage by state: FY 1990-94 [T]
5. Active users by center and state: FY 1990-94 (All centers and detail for each center)	Active users by state: FY 1994 [C] Active users by state: FY 1990-94 [T]
6. Normalized CPU usage by center and NSF Directorate and Division: FY 1990-94 (All centers and detail for each center)	Normalized CPU usage by NSF directorate: FY 1994 [C] Normalized CPU usage by NSF directorate and division: FY 1990-94 [T]
7. Active grants by center and NSF directorate and division	Active grants by NSF directorate: FY 1994 [C] Active grants by NSF directorate and division: FY 1990-94 [T]

Table E.1: Contents of the Annual Usage Reports of the National Science Foundations Supercomputer Centers

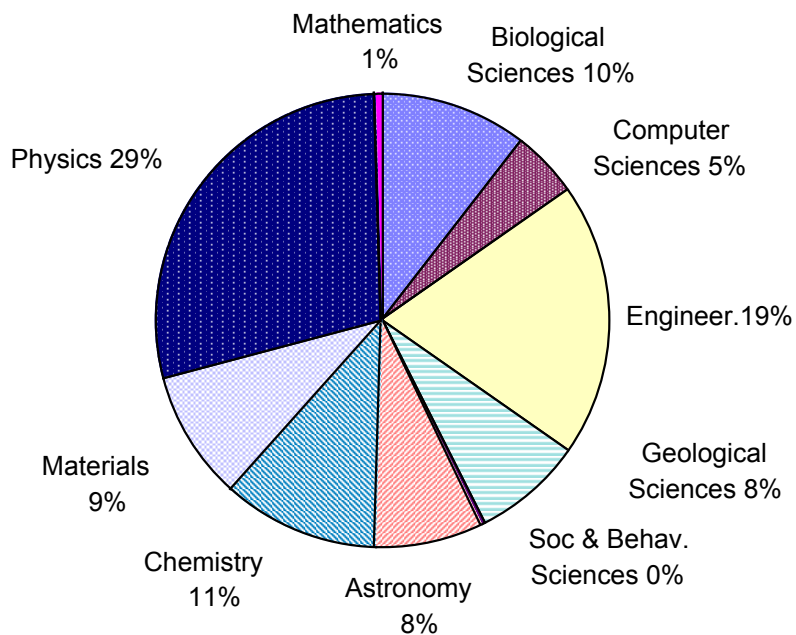


Figure E.4: FY 1994 Usage of NSF SC Centers by NSF Directorate and MPS Divisions

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

E.2: Longitudinal Analysis of Major Projects

There is much useful information in the Supercomputer Centers usage data as published. However, most major projects are performed by *groups* of faculty, post-docs, and graduate students. The data base supports the relations between accounts and the Project Leader, although the published reports do not make this linkage. Additionally, the users are sufficiently identified to determine whether they are NSF grant recipients.

The Task Force performed the following study:

For the Fiscal years 1992-95 (through February 95):

1. look at *projects* that accumulated more than 1,000 SU,
2. present these results from all uses (usage, title, NSF division/directorate) under the Project Leader,
3. identify the Project Leader in the NSF grants data base, and return all grants identified by the system (whenever the name of the Project Leader is listed in the system), and
4. enumerating for these grants, the grant amount, title, and the rating scores of the reviewers.

Some modifications to the grants data were made.

- Two center directors appeared on the list of NSF Grants (removed)
- Small grants (less than \$20,000) were removed from the data set
- If a grant had multiple PI's, the value of the grant was evenly divided.

However, this collection of projects (and PI's) provided substantial illumination about the usage patterns of grantees, and the quality of their research.

Some summary information from this study follows:

Number of Center Projects:	1428
Number of Distinct Project Leaders:	320
Value of 1245 grants held by project leaders:	\$317,786,170
Reviewer's average rating of proposals:	between E & VG

Figure E.5 shows the average funding history of the computational scientists and engineers who were project leaders using the NSF centers. The *total* base level funding for this cohort was \$15-20M per year, until just before the period studied, when it increased to a maximum of \$40M in 1991, and stayed above \$30M/year thereafter. (Data for FY 1994 was not complete in the system at the time of this study). Figure E.5 reflects these totals divided by the number of PI's in this sample (320, approximating the average new award per investigator).

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

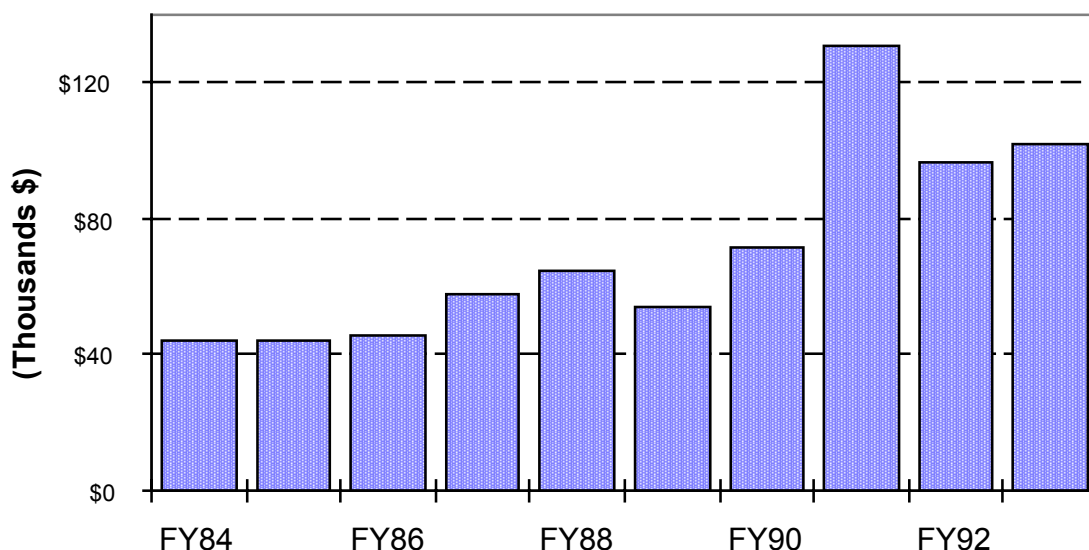


Figure E.5: NSF Average Funding History of 320 Selected PI's: FY 1984-93

A list of the grants in this sample and their titles (with PI's names removed) is appended to this document to illustrate the diversity of usage of the Centers. The large value in FY 1991 corresponds to several of the PI's in this group receiving large *Center Grants* from the newly established Science and Technology Centers program.

Individual PI's who were also project leaders had substantial amounts of total funding from NSF. One cannot make a quantitative evaluation of the *quality* of research from such data. The quality of research at the Centers is addressed in Appendix B.

Another approach is to examine how the original proposals were assessed by reviewers. All NSF proposals are reviewed, and given scores of Excellent (E), Very Good (V), Good (G), fair (F), or Poor (P). Figure E.6 shows the overall distribution of these scores for all the funded proposals of the selected PI's²¹. Recognizing that this metric is problematic in three regards:

1. it attempts to quantify what are clearly qualitative reactions
2. it measures opinions before the research is performed, rather than following the research.
3. it measures all NSF grants for the selected PI's, rather than the grants most closely related to the research at the Centers.

Without dwelling too long on the deficiencies of this method, one merely observes that the overall ratings are quite high, and comparable to random samples of ratings extracted from the same database.

²¹ The addition of reviewer's scores to the NSF grants database is performed by NSF program offices, and was instituted during the period under review. In many cases, the data were incomplete, but the results are presented for comparison.

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

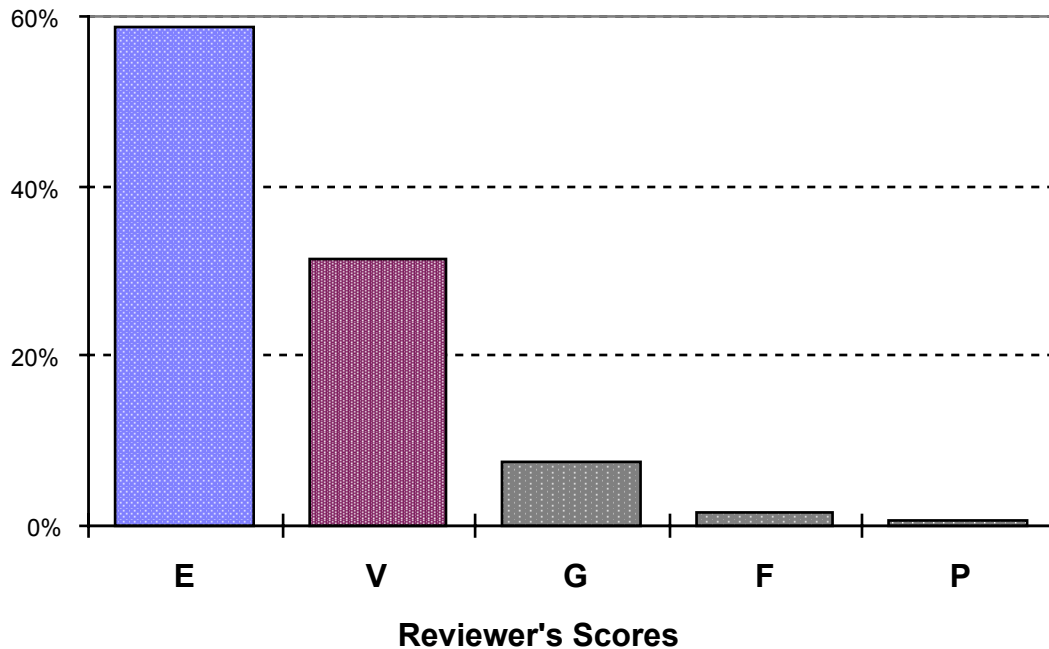


Figure E.6: Reviewer's Scores of Selected PI's Proposals: FY 1984-94

E.3: Duration of Projects at Centers

There is a general perception among some observers that there are users and projects that just keep *going on forever* at big computer centers. One way to test this assertion is to look at the cohort of projects, and rank their usage every year. The complete table of rankings is available under the TF_Report URL as reported at the beginning of this section, but the first 50 projects in 1995 ranking are shown in the table below:

Table E.2: The Top 50 Projects in FY 92-95 in Order of their FY 95 Usage Ranking

PI ID	Center	92 Rank	93 Rank	94 Rank	95 Rank	Title
33	P			37	1	The Formation Of Galaxies And Large-Scale Structure
291	P		26	10	2	Coupling Of Turbulent Compressible Convection With Rotation
195	P			80	3	Computer Simulation Of Biomolecular Structure, Function And Dynamics
255	N		2	8	4	Computational Relativity
215	N			28	5	Mca94-The Formation Of Galaxies And Large Scale Structure
281	N			1	6	Lattice Gauge Theory On MIMD Parallel Computers
281	S			9	7	Lattice Gauge Theory On MIMD Parallel Computers
132	N			43	8	Ab-Initio Simulations Of Materials Properties
31	P		29	16	9	Quantum Molecular Dynamics Simulations Of Growth Of Semiconductors And Formation Of Fullerenes
147	N			32	10	Penguin Operator Matrix Elements In Lattice QCD With Staggered Fermions
186	S			18	11	Theory Of Biomolecular Structure And Dynamics
44	C			136	12	Implementation On Parallel Computer Architectures Of A Particle Method Used In Aerospace Engineering

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

198	P				13	Synchronization And Segmentation In A Unified Neural Network Model Of The Primary Visual Cortex
81	C			156	14	The Deconfinement Transition In Lattice Quantum Chromodynamics In The Wilson Fermion Scheme
124	N				15	Electrostatic Properties Of Membrane Proteins
253	N		69	11	16	Modeling Of Biological Membranes And Membrane Proteins
300	N			376	17	Massively Parallel Simulations Of Colloid Suspension Rheology
153	P		21	34	18	Density Functional Ab Initio Quantum Mechanics And Classical Molecular Dynamics To Simulate Chemical Biomolecular Sys...
3	N			150	19	Cellular Automaton Analysis Of Suspended Flows
318	N			7	20	Massively Parallel Simulation Of Large Scale, High Resolution Ecosystems Models
191	P		73	12	21	Coherent Structures And Statistical Dynamics Of Rotating, Stratified Geophysical Turbulence At Large Reynolds Number
238	N				22	Direct Numerical Simulation Of Turbulent Reacting Flows Near Extinction
310	N			51	23	The Numerical Simulation Of Convective Phenomena
153	S			45	24	Simulations On Complex Molecular Systems
125	N			79	25	Application Of Effective Potential Monte Carlo Theory To Quantum Solids
148	C	57	32	358	26	Theoretical Study Of Lepton Anomalous Magnetic Moments
222	S		68	26	27	Salt Effects In Solutions Of Peptides And Nucleic Acids
34	P				28	Molecular Simulation Studies Of Biological Molecules
134	S			52	29	Crystal Growth Of Si And Al
157	C			105	30	Computer Simulations Of Critical Behavior
217	C			48	31	Monte Carlo Simulations Of Phase Coexistence For Polymeric And Ionic Fluids
18	N		318	275	32	Simulation Of High Rayleigh Number Thermal Convection With Imposed Mean Shear An D System Rotation
205	N			33	33	Computational Micromechanics
33	C			335	34	The Formation Of Galaxies And Large-Scale Structure
120	N	152	185	180	35	A Numerical Approach To Black Hole Physics
243	P	125	65	116	36	Dynamical Simulations Of Kinked DNA And Crystallographic Refinement By Simulated Annealing Of DNA Eco Ri Endonuclease...
206	N	216	320	171	37	Tempest
166	N		39	4	38	Supercomputer Simulations Of Liquids And Proteins
53	N		3	20	39	Simulations Of Quantum Systems
109	C		331	58	40	Molecular Simulation Of Fluid Behavior In Narrow Pores And Pore Networks
150	C		118	74	41	Quantum And Classical Simulations Of Molecular Aggregates
95	P		11	31	42	Electronic Structure Simulations Of Magnetic And Superconducting Materials
53	P			87	43	Simulations Of Quantum Systems
150	P		30	13	44	Quantum And Classical Simulations Of Molecular Aggregates
240	P			83	45	Modeling The Generation And Dynamics Of The Earth's Magnetic Field
30	N				46	Large Scale Simulations Of Polarizable Aqueous Systems
100	P			70	47	Numerical Simulation Of Reacting Shear Flow
132	S			69	48	Ab-Initio Simulations Of Materials Properties
140	S			41	49	Hybrid Spectral Element Algorithms: Parallel Simulation Of Turbulence In Complex Geometries
316	C		246	262	50	Differential Diffusion And Relative Dispersion In Isotropic Turbulence

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

Table E.2 is organized in three sections:

- The first two columns show an identifier number for the project leader, and the center where the project is rooted (C=CTC, N=NCSA, P=PSC, S=SDSC).
- The next four columns show the project ranking (1 = largest) for the four years of the study.
- The final column is the (abbreviated) project title.

It is interesting to compare the rankings of the projects in 1995 with earlier years. Most of these projects were not in the top rankings in FY92 or 93, and in many cases, rated much lower during FY94. Complete inspection of the full data reveals projects that represent the top usage in FY 1992, 93 are generally absent in FY 1994, and 95, and the converse. This pattern probably occurs because this group of PI's organizes their research projects into well defined tasks, most likely related to dissertation topics for graduate students.

In table E.3, the same data are aggregated for project leaders. This table now occupies ten columns, and is extracted for the first 50 PI's based on their 1995 ranking.

Column 1 is the ID number of the PI
 Columns 2-5 show the usage (in normalized SU) for the FY 1992-95
 Columns 6-9 show the ranking of the PI in FY 1992-95
 Column 10 shows the NSF funding for the faculty PI in the NSF data base (adjusted as described earlier). Those entries where NSF funding is absent belong to PI's with funding from other agencies, but the reference base for these grants is different and our ability to get all funding sources is limited.

Only 14 of the top 50 in 1995 were in the top 50 in 1992, indicating a turnover of 72% during the period of this study. Additionally, these top project leaders also have substantial amounts of NSF funding; in fact, only 2 of the top 50 lack NSF funding (one is funded by NIH, and the second by ONR).

If one looks at the usage (rather than the ranking) and recalls that the cut-off for inclusion in this cohort is the total use of 1,000 SU over the period, one concludes that these PI's are truly large users, but they have patterns that show starts and stops in their computer usage, driven by ideas and opportunities as in other kinds of research.

Table E.3: Ranking and NSF Funding of the Top 50
 Project Leaders for FY 1992-95

PI ID	92 usage	93 usage	94 usage	95 usage	92 rank	93 rank	94 rank	95 rank	NSF Funding
33	0	827	24,235	108,263	201	121	16	1	1,037,930
281	33,347	34,533	148,958	60,993	2	2	1	2	8,651,897
291		9,997	24,565	56,174		13	15	3	1,162,300
195			4,830	46,739			53	4	
255		28,907	29,728	44,678		4	8	5	74,000
215		308	14,890	42,529		200	28	6	3,590,533
132	8,425	9,356	16,176	32,698	3	15	25	7	171,200
31	193	7,755	20,841	21,706	151	18	19	8	423,000
186	2,982	3,143	25,891	20,271	9	39	13	9	1,872,122

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

147		1,735	33,760	19,863		63	5	10	270,600
153	6,557	12,130	20,152	18,202	5	7	20	11	2,400,350
44			1,938	16,857			113	12	298,404
198		4	1,040	15,369		256	185	13	141,189
253	7,079	8,774	25,898	14,779	4	16	12	14	3,283,161
53	1,044	24,605	25,145	14,407	53	5	14	15	847,700
81			1,653	13,873			126	16	1,194,693
124				13,664				17	4,026,985
300			195	11,832			281	18	356,708
3			1,758	10,765			123	19	291,226
318			27,046	10,743			11	20	1,455,666
150	1,712	10,909	28,675	10,675	28	10	10	21	8,921,630
157		899	6,695	8,514		114	43	22	1,554,367
310	2,637	5,456	23,786	8,231	15	25	18	23	2,727,003
191		3,391	24,188	8,142		37	17	24	1,843,641
34	2,680	2,974	1,280	7,698	14	40	156	25	911,002
238				7,658				26	1,198,788
222	1,541	4,972	15,574	7,267	32	26	27	27	2,087,256
109	2,950	30,910	32,651	6,815	10	3	6	28	2,831,747
125			3,677	6,699			67	29	368,624
148	1,065	4,449	344	6,632	52	29	266	30	
217	678	3,750	6,874	6,536	74	34	42	31	282,000
166	546	4,677	50,697	6,252	93	28	3	32	635,000
120	265	639	2,728	6,179	130	141	89	33	341,183
18		52	1,781	6,127		247	122	34	61,699
134			5,702	5,950			47	35	928,000
96	559	1,480	12,266	5,559	89	76	33	36	994,304
95	1,396	11,528	13,126	5,269	35	8	32	37	1,790,102
95	1,396	11,528	13,126	5,269	35	8	32	37	1,790,102
140		3,428	9,666	5,183		36	37	38	1,828,880
205	117	444	10,439	5,130	172	170	35	39	1,739,197
317	2,379	5,483	18,272	4,952	17	24	22	40	409,800
243	416	2,046	2,408	4,838	110	55	98	41	39,840
206	53	43	1,468	4,803	189	251	139	42	1,404,901
100	559	1,908	5,441	4,424	90	58	48	43	189,433
240			3,301	4,394			74	44	993,361
30				4,181				45	3,316,728
64	3,535	7,736	8,534	4,158	8	19	38	46	4,045,000
65	1,905	6,808	4,880	3,907	23	21	52	47	508,097
184		60	4,267	3,890		245	59	48	6,223,125
316		353	895	3,754		192	200	49	99,758
106	745	713	10,914	3,353	69	131	34	50	1,810,461

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

E-4: Budget Information about the Centers

The Task Force had detailed investigations of the budgets, and budget history, of the four NSF supercomputer centers as part of its study of the program. The information presented is the Income History of all centers from the following sources:

- The Base
- Cooperative Agreement
- Other ASC funds available only to the centers (by competition)
- Other NSF funds (peer reviewed) distributed through the cooperative agreement
- Other Federal Funds (other agencies)
- State and Local Support
- University Support
- Foundation SupportIndustrial Support
- Affiliates
- Purchase of Computer Time
- Vendor Support (both real money and *in-kind* support at center facilities)

Appendix E: Quantitative Data Describing the NSF Supercomputer Centers

Funding History of the NSF Supercomputer Centers Composite History of Four NSF Centers (CTC, NCSA, PSC, SDSC) Through the NSF CISE / ASC Centers Program And other Private Sources

	FY85	FY86	FY87	FY88	FY89	FY90	FY91	FY92	FY93	FY94	10 Year
Base Coop Agreement	14,600,000	23,505,000	29,102,601	33,016,560	37,772,079	47,483,369	53,939,007	56,104,498	57,247,073	57,951,215	410,721,402
Other NSF Funding	1,636,108	2,047,121	2,007,256	2,135,924	3,179,075	5,610,095	4,083,616	11,689,450	8,937,314	9,881,107	51,207,066
All other Federal	0	532,923	305,528	1,677,046	972,340	920,995	2,973,918	4,540,261	7,323,222	10,397,366	29,643,599
University	4,533,824	4,910,642	5,679,313	5,573,472	6,340,258	7,212,766	5,447,752	7,038,405	6,302,352	5,618,532	58,657,316
State/Local	250,000	1,460,944	4,942,991	4,937,195	6,550,000	13,161,072	7,914,226	5,904,976	12,273,744	10,383,722	67,778,870
Grants/Found. & Corp	252,115	377,100	1,159,315	2,454,826	3,150,545	3,514,381	3,432,281	3,383,501	4,187,820	4,735,450	26,647,334
Vendor (incl in-kind)	8,100,000	12,194,900	10,924,577	15,833,200	18,825,508	17,084,936	21,249,782	19,137,761	30,437,702	23,326,851	177,115,217
Misc	0	303,900	3,537,366	4,004,899	5,175,544	6,194,521	3,987,354	5,341,686	5,816,839	3,338,514	37,700,623
Totals	29,372,047	45,332,530	57,658,947	69,633,122	81,965,349	101,182,135	103,027,936	113,140,538	132,526,066	125,632,757	859,471,427
Base Coop Agreement	49.7%	51.9%	50.5%	47.4%	46.1%	46.9%	52.4%	49.6%	43.2%	46.1%	47.8%
Other NSF Funding	5.6%	4.5%	3.5%	3.1%	3.9%	5.5%	4.0%	10.3%	6.7%	7.9%	6.0%
All other Federal	0.0%	1.2%	0.5%	2.4%	1.2%	0.9%	2.9%	4.0%	5.5%	8.3%	3.4%
University	15.4%	10.8%	9.8%	8.0%	7.7%	7.1%	5.3%	6.2%	4.8%	4.5%	6.8%
State/Local	0.9%	3.2%	8.6%	7.1%	8.0%	13.0%	7.7%	5.2%	9.3%	8.3%	7.9%
Grants/Found. & Corp	0.9%	0.8%	2.0%	3.5%	3.8%	3.5%	3.3%	3.0%	3.2%	3.8%	3.1%
Vendor (incl in-kind)	27.6%	26.9%	18.9%	22.7%	23.0%	16.9%	20.6%	16.9%	23.0%	18.6%	20.6%
Misc	0.0%	0.7%	6.1%	5.8%	6.3%	6.1%	3.9%	4.7%	4.4%	2.7%	4.4%

Table E.4: Composite Funding History FY 1985-94

Appendix F: Education at the NSF Supercomputing Centers

Advanced education and training have been an important part of the Supercomputing Centers program. A significant role was foreseen from the very beginning:

One of the purposes of an Advanced Scientific Computing Center will be to educate students and researchers in the use of supercomputers in order to supply the talent necessary for industry, government, and academia to take full advantage of these machines.²²

When the current centers were established, vector supercomputers were the epitome of high end computing and they were more complex to program than the computers generally available to the research community in the early 1980's. It was well understood that unless codes used in research took advantage of the vector features of these systems, the promise of high execution speeds characteristic of high performance computing could not be achieved. Hence, it would be necessary to train researchers to use these systems efficiently and to build up a cadre of researchers well-educated in the uses of high-end computation.

As shown in Fig 3.5 of the Task Force report, the corresponding change from vector to parallel computing in computational science and engineering is a recent phenomenon, and the inherent difficulties now faced in efficiently exploiting parallel computation are substantially greater than the problems faced in migrating to vector computing in the last half of the 1980's. As computational power increases, and as the ability to use computation in advancing scientific and engineering disciplines is more broadly appreciated, new users of supercomputing, and whole new disciplines, need educating in the techniques and ideas of supercomputing. Furthermore, until parallelism is better understood and until there is a broader national cadre of researchers who can lead the education of both disciplinary and interdisciplinary computational scientists and engineers, there will remain an important role for advanced education and training as part of a national supercomputer centers program. One of the overall program goals should be to develop and distribute this expertise nationally, so that there will no longer be the same need for a nationally centered program to train people in the use of parallelism and to increase the numbers of expert researchers in advanced scientific computation.

As discussed in section 3.1, the Supercomputer centers do not fit the "pure model" of either a NSF facility or a research center. Since they are primarily facilities, they do not have the same broad educational mission as do research centers. Nevertheless, the Centers educational and support functions have played a critical role in the emergence of computation as a full partner in the paradigm of how science and engineering are done, and they should continue to play such a role as this computational paradigm gets established in more disciplines and as parallel computation plays an increasingly broad role in this paradigm shift. Thus the Center program's primary educational role should be as an enabler, shifting both computational science and engineering and the use of parallelism to a broader national base. It follows that the program's primary educational role should be focused on advanced education and training to enable these two broad enabling shifts. At the same time, the program has had, and can continue to have,

²² Project Solicitation--Advanced Scientific Computing Centers. NSF 1985

Appendix F: Education at the NSF Supercomputing Centers

significant impact at the K-12 level for heightening interest in, and information about, science and engineering at an informal level.

One can get some idea of the overall educational role of all of the Supercomputing centers by browsing URL

<http://www.tc.cornell.edu/Edu/>.

F.1: Computational Science and Engineering - Education and Training

Many Centers' programs are specifically targeted to faculty, post-docs, graduate students, and advanced undergraduates, as well as to industrial scientists and engineers. In addition to the training that comes by working directly with researchers and graduate students who use the Center's resources, educational mechanisms include stand alone documentation, special short courses, training sessions, institutes, and conferences, often augmented by networks of affiliated universities and industries. Some idea of the extent of this activity can be seen from Table F.1, which gives data indicating overall educational and outreach activities.

The Cornell Theory Center's (CTC's) "Smart Nodes" program, is one noteworthy example of how the Centers' academic outreach programs work. This program gives access to training and education to 96 associated colleges, universities, and industrial laboratories, and CTC's use of web technology to make virtual workshops globally available for classes on parallel computation, including access to the Center's SP2 parallel environment.

The National Center for Supercomputer Applications at Illinois (NCSA) provides a similar example. The NCSA Academic Affiliates Program involves over 120 universities around the country. These affiliates received direct training from NCSA through week-long workshops with training and consulting support delivered to local users on their own campuses. The program also supports university course accounts through the use of block grants.

The Cornell Theory Center's (CTC's) "Smart Nodes" program, is one noteworthy example of how the Centers' academic outreach programs work. This program gives access to training and education to 96 associated colleges, universities, and industrial laboratories, and CTC's use of web technology to make virtual workshops globally available for classes on parallel computation, including access to the Center's SP2 parallel environment.

Appendix F: Education at the NSF Supercomputing Centers

	General Visits	Course Support		Independ. Use	Special Consult	Center Sponsored		Academic Affiliates
		Events	People			Workshop	Seminar	
K-12 Teachers	2,984	4	148	818	28	519	62	
K-12 Students	1,273	4	8,042	89		271	2	
Undergrads	797	31	941	227		72	275	600
Graduate Students	263	53	1,134	20	124	683	1,437	129
Industry Tech Transfer	1,119	1	44	397	233	275	279	215
Academic Researchers	856	45	626	18	214	821	715	42
Computat. Support Professionals	941	1	21	151	95	706	149	44
Totals	8,233	139	10,956	1720	666	3347	2,919	1030

Table F.1: Educational Impact at the NSF Supercomputing Centers --Number of Participants for 1994²³

Historically black colleges and universities that are members of the SCCs' affiliate programs include South Carolina State College, Tuskege University, Howard University, Prairie View A&M University, Southern University, and Tennessee State University.

Table F.1 shows that, in addition to providing direct access for graduate students involved in research projects, the Centers have had a significant impact on classroom teaching of undergraduates and graduate students. For example, the Centers provide access to supercomputing for classes at universities around the country. Table F.1 shows a typical year, with 84 university classes supported, serving over 2,000 undergraduate and graduate students. In addition, Centers personnel provided staff support for over 300 classes at affiliated universities and colleges, and nearly 2,500 undergraduate and graduate students attended special workshops and seminars sponsored by the Centers.

All of the Centers have been active in the NSF's Research Experience for Undergraduates Program. As one example, in FY94, NCSA worked with 63 REU students, who were directly involved in research and development projects in scalable metacomputing, software tools development, scientific applications development, education research and virtual environments development. A dozen of these students were from primarily

²³Definitions of column headings:

General Visits: A Center visit that had an educational value for Center staff or the visitor.

Course Support: Education accounts, other people who use the facility for training events, etc. This is broken down into number of events and number of participants.

Independent Use: Users of Center resources for educational purposes for a largely independent basis (e.g., student employees, REU programs, SuperQuest).

Center Workshops: Includes any Center Sponsored workshop. Often including a hands-on component.

Center Seminars: Center sponsored events including seminars, symposiums, conferences.

Special Consulting: One-on-one, long term consulting relationships including industrial and strategic users.

Academic Affiliates: Course taught or supported by those Computational Support Professionals at other institutions that the Centers train (e.g., Smart Node Consultants).

Appendix F: Education at the NSF Supercomputing Centers

undergraduate institutions and spent an entire semester working as part of research teams with applications scientists at NCSA.

Finally, while no exact figures are available, the Task Force estimates that, since the start of the program, on average between one hundred and two hundred Ph.D.'s have been awarded each year to students whose research has been dependent on computation carried out at the Centers.²⁴

Beyond student support, some idea of the technology transfer role the Centers serve is indicated in a typical year by over 5,500 visits and attendance at Center sponsored workshops and seminars just by people from industry, academic researchers, and computational support professionals.

As an example of how this works, the NCSA Industrial Program has established particularly close long-term partnerships with thirteen major corporations. Efforts have focused on educating partners to the rapidly evolving high performance computing and communication tools and technologies and on working jointly to solve major industry projects involving product enhancements and process improvements. Industrial Partners have benefited through technology transfer involving workshops, seminars, specially designed training sessions, and numerous consultations which are conducted both at NCSA and at the industrial headquarters locations.

The Industrial Partners have also included university undergraduate and graduate students in their research and development projects, giving students practical industrial experience.

Overall, it is clear that the Centers program has had, and can continue to have, significant impact on technology transfer, on research training, and on the numbers of students who are educated in the uses of high end computation for computational science and engineering. On average, since the late 1980's, over 900 new users have used the Center's programs each year.

F.2: K-12 ProgramsAs can be seen from the Figure F.1, the attraction of a "center" as a magnet to develop interests in secondary school students in science and engineering should not be underestimated. Visualization can help show young people the results of scientific work, while the approach of breaking down complex problems into calculations or numerical simulations can give new insights into ideas often approached by more abstract methods. While the Centers' K-12 activities are generally not aimed at computational problems *per se*, they are a tool in heightening general scientific and technological awareness.

In addition to nearly 3,000 general visits to the Centers by K-12 teachers, in 1994, over 850 K-12 teachers attended Center sponsored workshops and seminars, a surprisingly large number (818) made independent use of the Center's facilities, and 148 directly used the Centers in courses impacting over 8,000 K-12 students.

²⁴ This estimate is based on the average number of graduate students users of the Centers, assuming a time to Ph.D. of six years and about 50% success ratio. It compares well with the Centers' own estimate and records.

Appendix F: Education at the NSF Supercomputing Centers

One highlight of the K-12 effort has been SuperQuest, the Cornell Theory Center's program for students *and their teachers* to compete for places in special summer programs focused on interdisciplinary programs in computational science. Other centers have also participated. For example, as part of the 1994 program, NCSA worked with 27 teachers from 14 different schools .

CTC also has a special science exploratory home page,

<http://www.tc.cornell.edu/Edu/MathSciGateway/>

for locating science and mathematics resources for students and teachers in grades 9 through 12. In a recent three month period, there were over 25,000 visits to the Gateway. In addition, CTC's web pages,

<http://www.tc.cornell.edu/Kids.on.Campus/WWWDemo/>

was recently ranked as one of the "50 best places to go on-line: by NetGuide Magazine.

The Pittsburgh center, with support from NSF's EHR directorate is linked with the Pittsburgh public schools in a project called "Common Knowledge: Pittsburgh" (CK:P). The Center is providing the technical expertise for cost effective configuring of user devices, servers, and networking of the schools and developing techniques to make these systems more easily managed. The project will assist the Pittsburgh public schools in developing technology plans and in institutionalizing the requisite support within the school system. SDSC has established a special program called Science Scholars for secondary school girls in the greater San Diego Area who have an interest in a scientific career. Each year, the program provides about 25 girls with computer instruction, mentoring, science enrichment, and tutoring. There has been a special emphasis to see that under-represented minorities students have the necessary transportation and computer equipment. In each of the two years of the programs existence, about 3/4 of the participants have been Native American, Hispanic, or African American. For the 1995-1996 year the program has expanded to involve teachers (working with local groups at their school), with a special emphasis on Native American girls living on Indian reservations in San Diego county.

In addition to local impact, the Centers affiliates programs have allowed the Centers to have broader national impact. For example NCSA has worked with Montana State University, TENET, Pan Educational Institute, North Central Regional Educational Laboratory (NCREL), Champaign and Urbana School Districts, Minnesota Supercomputer Center, and others to transfer tools, technology and methodologies to the broader K-12 community supported by these institutions. In addition to working through their regular affiliates program, NCSA has additional programs for K-12 outreach, working directly with a number of school districts around the country, and working with OSTP to extend the US Tech Corp. from Massachusetts, to five additional states, and eventually to the entire country.

In addition to K-12 outreach, the Centers has established regional community outreach programs. The Cornell Theory Center works directly with the state government in industrial outreach, and NCSA has a number of state-wide programs. For example, at the request of Illinois's lieutenant Governor, NCSA's highly effective "Silicon Prairie"

Appendix F: Education at the NSF Supercomputing Centers

program of outreach to Champaign County Illinois, has now been extended to a state-wide project, "Illinois Learning Mosaic".

Appendix G: Witnesses and other Sources of Information

In the course of its deliberations, the Task Force queried a variety of sources of information and data. The Task Force heard from a number of witnesses both in open session and in direct discussions with subgroups (Section G.1). All of the Centers were visited by Task Force members, as well. Over 100 companies and individuals, prominent in high performance computing, were invited to provide written input to the Task Force and many did. This list included all of the industrial affiliates of the Centers and all of the major vendors of computing hardware. Quantitative data are in Appendix E. The results of a survey gathered via the world wide web are in Appendix H.

G.1: Task Force witnesses and interviewees

NSF ASSISTANT DIRECTORS

Joseph Bordogna
Mary Clutter
Robert Correll
William Harris

ACADEMIC EXPERTS AND RESEARCHERS

Geoffrey Fox, Syracuse University
Ken Kennedy, Director CRPC - Rice University
Ed Lazowska, Univ of Washington
Mike Norman , NCSA and UIUC
Jerry Ostriker, Princeton Univ
Juri Toomre, Univ of Colorado
Paul Woodward, Univ of Minnesota

OTHER GOVERNMENT AGENCIES

Erich Bloch, Former Director NSF
Richard Dubois, NIH Research Resources
Howard Frank, CSTO/ARPA
Lee Holcomb, NASA
Caroline Holloway, NIH Research Resources
Anita Jones, Director, DDR&E/DOD
David Nelson, DOE/ER
John Toole, Director, National HPCC Coordination Office
Alvin Trivelpiece, Director Oak Ridge National Lab, Former Energy Research
Director, DOE
Gil Weigand, DOE Defense Programs

CENTERS (NSF and other) AND ALLOCATION COMMITTEES

Jay Blaire, CTC and Metacenter allocations
John Connolly, University of Kentucky, (Former head ASC/NSF)
Dennis Duke, Florida State University
Jeff Huskamp, Director, North Carolina Supercomputer Center
Mal Kalos, Director CTC
Sid Karin, Director, SDSC

Appendix G: Witnesses and other Sources of Information

George Karniadakis, NCSA/PSC allocations committee
Tom Jones, Univ. of Minnesota
Robert Leary, SDSC allocations
Paul Messina, Caltech
Ralph Roskies/Michael Levine, Co-Directors, PSC
Larry Smarr, Director, NCSA

Appendix H: Survey of Users

H.1: Background

To aid the Task Force in assessing the opinions of prospective supercomputer users, including researchers who currently do not use the NSF Supercomputer Centers, a survey form was made available on the internet by Ohio State University. Pointers to the survey form were distributed through Computer Research Association (CRA), the Supercomputer Centers, internet bulletin boards and news groups on high performance computing, and publications of interest to those in high performance computing. The form and its questions are shown in H.5. A total of 513 responses were received (139 by e-mail). Responses were validated by the following criteria:

- The respondent had a valid e-mail address
- Only one response per e-mail address
- E-mail was sent to the address provided to verify that the identifier was truly the respondent - only one reply was received indicating that the addressee was not the one who completed the form.

Below is a summary analysis of the responses. The Survey data is available at URL

<http://www.cise.nsf.gov/acir/hpc/>

The responses to the multiple choice questions are summarized in H.4. The statistics show that respondents represent the range of disciplines supported by NSF. 75% of the respondents have used the existing centers within the past year, and 78% believe that the centers have had a significant or critical influence on their research. Furthermore, approximately 50% of the respondents estimate that 40-100% of their computing needs in the next few years will be met by resources at national centers. 60-75% of the respondents felt that processor speed, memory capacity, and system availability and reliability are either very important or critically important features of center facilities. 40-50% of the respondents feel that large mass storage, network bandwidth for remote access, and internode communications bandwidth are either very or critically important. Smaller but significant percentages of respondents (20-30%) indicated that I/O bandwidth, consulting services, training, and software support were either very or critically important. These results were remarkably consistent, qualitatively, across various cross-cuts of respondents (by discipline, status of researcher, users vs. non-users). The statistics are also consistent with the major themes and other significant opinions expressed in the answers to the open-ended questions about the center's program at the end of the survey. The major themes are summarized in H.2 below; further significant opinions that were expressed are summarized in H.3.

H.2: Major themes in the survey responses to the open-ended questions

The opinions below are listed in order of decreasing frequency of the number of survey responses that expressed the given opinion. A collection of excerpts of the survey responses that express the opinions summarized below are available at the URL listed above.

1. The NSF Centers have enabled important research that otherwise would not have been possible, across the range of science and engineering disciplines represented by the respondents to the survey.

Appendix H: Survey of Users

2. The focus of the Centers Program should be access to the highest performance systems possible (suitable for Grand Challenge efforts).
3. Shared hardware resources and staff support are most appropriate and cost effective; the Centers are providing important leadership in the form of information, access, and prototyping of the latest high performance hardware and software; improvements may be needed, but drastic overhaul of the Program is not.
4. Staff support and software expertise at the Centers is very helpful, if not indispensable for many researchers.
5. There is a need for better hardware, software, and technical support for computer science research.
6. The systems at the Centers are oversubscribed; many researchers find it difficult to get large enough allocations of time and/or dedicated time on the largest possible configurations. (Note: this is a reflection of the success of the Centers.)
7. The allocation process is too time consuming. Suggestions for improving this included: expanding the MetaCenter allocation process and having the MetaCenter allocation committee meet more frequently; allocating generic SU's that can be used on any platform; provide allocations without expiration date; on-line application submittals.

H.3: Significant opinions expressed by fewer respondents

1. Centers have been instrumental in facilitating important interdisciplinary interactions.
2. The Program should include a greater funding for midrange systems -- an effective way to support smaller users and increase scientific throughput; may be possible to get enormous leverage of University matching funds
3. Better software libraries, more convenient access to mass storage, and more reliable file systems are needed.
4. More consulting support is needed, particularly for remote users. Suggestions: remote training workshops, more extensive on-line documentation, expanded hours for phone consultation, and "MetaCenter consulting" services.
5. Internet bandwidth is insufficient, particularly for remote visualization.
6. Allocation committees need to consider proposals more carefully and/or include a better balance of disciplinary representation;
7. Access to Center resources for college courses has been important to the teaching mission in science and engineering

H.4: Statistical Results from the Survey:

The following section summarizes the statistical responses to the survey. Each entry is a brief restatement of a survey question, with two groups presented:

1. all respondents and
2. those respondents who have used the NSF centers during the previous year.

Futher summary statistics are available on the World Wide Web at URL:

Appendix H: Survey of Users

<http://www.cise.nsf.gov/acir/hpc/>

Appendix H: Survey of Users

Status as a researcher	All Responses	Used NSF Center in past year
Research Staff.	56	46
Faculty Principal Investigator	187	138
Other Faculty Member	16	12
Graduate Student	115	89
Post-Doctoral	53	45
Undergraduate Student	16	9
Industrial Research Staff	17	12
Center Research Staff	21	17
Other	32	17
Total	513	385

Used any NSF Center in past year?	All Responses	Used NSF Center in past year
Yes	385	385
No	128	100%

Experience at:	All Responses	Used NSF Center in past year
CTC	185	157
NCSA	171	149
PSC	151	136
SDSC	136	122
NCAR	33	29
Federal	157	127
State	53	37
University	134	93
Lab or Dept.	112	78

Field of Research	All Responses	Used NSF Center in past year
Physics	68	55
Chemistry	57	45
Astronomy	38	35
Materials Research	27	22
Mathematics	15	12
CISE - Computer Science	102	64
ENG - Engineering	100	75
BIO - Biological Sciences	33	25
GEO - Geological Sciences	21	20
SBE - Social Sciences	7	6
Other	45	26
Total	513	385

1. Expect my area of research will require access to Super-computers in future?

	All Responses	Used NSF Center in past year
Yes	477	374
No	15	3
No Opinion	21	8

2. Fraction of computing needs met by:

2a. Desktop Workstations	All Responses	Used NSF Center in past year
0-20%	171	148
21-40%	123	105
41-60%	89	67
61-80%	74	45
81-100%	38	15
Do Not Know	18	5

Use of SC Centers Influence on Research	All Responses	Used NSF Center in past year
Critical Influence	244	216
Significant Influence	148	117
Moderate Influence	78	44
No Influence	43	8
Mean response	Significant Influence	Significant Influence

Appendix H: Survey of Users

2b. Dept. or University Systems	All Responses	Used NSF Center in past year
0-20%	203	154
21-40%	144	112
41-60%	62	49
61-80%	19	14
81-100%	12	8
Do Not Know	73	48

2c. National Centers	All Responses	Used NSF Center in past year
0-20%	118	66
21-40%	98	80
41-60%	80	70
61-80%	70	62
81-100%	99	88
Do Not Know	48	19

3. For the "Highest Performance Systems", rank projected needs in following areas:

3a. Processor Speed	All Responses	Used NSF Center in past year
No Opinion	23	5
Unimportant	8	2
Important	89	63
Very Important	178	136
Critical	215	179
Mean response	Very Important	Very Important

3b. Large memory size	All Responses	Used NSF Center in past year
No Opinion	24	8
Unimportant	23	13
Important	140	106
Very Important	165	129
Critical	161	129
Mean response	Very Important	Very Important

3c. Large mass storage	All Responses	Used NSF Center in past year
No Opinion	37	16
Unimportant	55	39
Important	191	134
Very Important	128	110
Critical	102	86
Mean response	Very Important	Very Important

3d. Network Bandwidth for remote access	All Responses	Used NSF Center in past year
No Opinion	56	30
Unimportant	55	38
Important	183	140
Very Important	127	103
Critical	92	74
Mean response	Very Important	Very Important

3e. I/O Bandwidth (to local peripherals)	All Responses	Used NSF Center in past year
No Opinion	80	49
Unimportant	83	61
Important	187	149
Very Important	107	81
Critical	56	45
Mean response	Important	Important

3f. Intermode Communication Bandwidth	All Responses	Used NSF Center in past year
No Opinion	86	52
Unimportant	47	32
Important	125	92
Very Important	120	97
Critical	135	112
Mean response	Very Important	Very Important

Appendix H: Survey of Users

	All Responses	Used NSF Center in past year	Mean response	Important
3g. System availability & reliability				
No Opinion	31	12		
Unimportant	7	3		
Important	129	100		
Very Important	177	135		
Critical	169	135		
Mean response	Very Important	Very Important		
4c. Information Processing				
No Opinion	132	84		
Unimportant	120	91		
Important	160	126		
Very Important	67	57		
Critical	34	27		
Mean response	Important	Important		
4d. Consulting Services				
No Opinion	132	84		
Unimportant	120	91		
Important	160	126		
Very Important	67	57		
Critical	34	27		
Mean response	Important	Important		
4e. 3rd Party or Commercial software				
No Opinion	103	66		
Unimportant	136	98		
Important	179	144		
Very Important	69	54		
Critical	26	23		
Mean response	Important	Important		
4. How important do you believe that National Centers will be in providing each of the following services (located at the centers) for your research or educational projects?				
4a. Diversity of computer architectures				
No Opinion	82	44		
Unimportant	84	59		
Important	174	137		
Very Important	110	91		
Critical	63	54		
Mean response	Important	Important		
4b. Visualization facilities & software				
No Opinion	70	40		
Unimportant	128	94		
Important	167	132		
Very Important	103	84		
Critical	45	35		

Appendix H: Survey of Users

4f. Expertise in Team Building	All Responses	Used NSF Center in past year
	No Opinion	69 18%
	Unimportant	148 38%
	Important	91 24%
	Very Important	61 16%
Mean response	24 5% Important	4% Important
4g. Assistance in code porting	All Responses	Used NSF Center in past year
	No Opinion	47 12%
	Unimportant	103 27%
	Important	143 37%
	Very Important	67 17%
Mean response	30 6% Important	6% Important
4h. Training	All Responses	Used NSF Center in past year
	No Opinion	44 11%
	Unimportant	93 24%
	Important	159 41%
	Very Important	71 18%
Mean response	29 6% Important	5% Important
4i. Communications software	All Responses	Used NSF Center in past year
	No Opinion	65 17%
	Unimportant	66 17%
	Important	167 43%
	Very Important	62 16%
Mean response	38 7% Important	6% Important

4j. Data Services (massive databases)	All Responses	Used NSF Center in past year
	No Opinion	78 20%
	Unimportant	127 33%
	Important	104 27%
	Very Important	51 13%
Mean response	31 6% Important	6% Important

4k. Data Dictionaries & Repositories	All Responses	Used NSF Center in past year
	No Opinion	93 24%
	Unimportant	84 22%
	Important	113 29%
	Very Important	69 18%
Mean response	28 5% Important	7% Important

4l. Internet software development	All Responses	Used NSF Center in past year
	No Opinion	81 21%
	Unimportant	69 18%
	Important	134 35%
	Very Important	79 21%
Mean response	31 6% Important	6% Important

Appendix H: Survey of Users

5. I expect to obtain good results for my codes on the following computers or their successors in the next few years (check all that apply)

	All Responses	Used NSF Center in past year
Vector Multiprocessors	228 44%	194 50%
Massively Parallel Systems	332 65%	285 74%
Shared Memory Microprocessors	230 45%	183 48%
Tightly Coupled Net of Workstations	218 42%	161 42%
Loosely Coupled Workstations	9 2%	7 2%
Do Not Know	28 5%	11 3%

7. Should NSF use a disciplinary Model at centers?

	All Responses	Used NSF Center in past year
Yes	91 18%	51 13%
No	291 57%	247 64%
No Opinion	131 26%	87 23%

8. Have NSF Centers facilitated Interdisciplinary Interactions?

	All Responses	Used NSF Center in past year
Yes	215 42%	182 47%
No	182 35%	131 34%
No Opinion	116 23%	72 19%

Appendix H: Survey of Users

H.5: The Survey Instrument

NSF SUPERCOMPUTER CENTERS - USER SURVEY

General Instructions

Save this form as "text" from your web browser, and then fill in the responses with your favorite editor. Remember to edit it and save it as "text" or ASCII. When finished, mail it to survey@osc.edu.

Thank you for your interest in the NSF Supercomputer Centers Program. Members of the "Task Force on the Future of the NSF Supercomputer Centers Program" have prepared some questions about your experiences with either the NSF or similar Supercomputing Centers, your opinions about the future of high performance computing, and your research needs in these areas over the next few years.

Background Questions:

USED FOR CONFIRMATION

Upon filling out this questionnaire, an electronic message will be sent to you at the e-mail address you specify below. This is a simple attempt to prevent others from using your name while responding to the

survey. All respondents must have valid internet e-mail addresses. Those who have had their e-mail addresses misappropriated will be instructed to send a message to have the entry voided. Additionally, "bounced" e-mail resulting from improper addresses or hosts will also cause those entries to be voided.

Your name:

- Dr.
- Prof.
- Mr.
- Ms.
- Miss

Mrs. First: _____ Last: _____

E-Mail address: _____

Status as a researcher:

- Research Staff.
- Faculty Principal Investigator
- Other Faculty member
- Graduate Student
- Post-Doctoral
- Undergraduate Student
- Industrial Research Staff
- Center Research Staff
- Other

EXPERIENCE

Field of Research matched to NSF Division or Directorate:

- Physics
- Chemistry
- Astronomy

Appendix H: Survey of Users

- Materials Research
- Mathematics
- CISE - Computer Science
- ENG - Engineering
- BIO - Biological Sciences
- GEO - Geological Sciences
- SBE - Social Sciences
- Other

I have High Performance Computing experience at: (check all that apply)

- CTC - Cornell Theory Center
- NCSA - National Center for Supercomputing Applications
- PSC - Pittsburgh Supercomputing Center
- SDSC - San Diego Supercomputer Center
- NCAR - National Center for Atmospheric Research
- Other Federal (NASA, DoE...) Supercomputer Center(s)
- State or regional Supercomputing Center
- My University's Supercomputing System or Facility
- My own, or my department's, supercomputing facility

To what degree has the use of supercomputing centers influenced your research?

- No Influence
- Moderate Influence
- Significant Influence
- Critical Influence

Have you (or your collaborators or students) used any NSF Center during the past year?

- Yes
- No

I estimate that my total supercomputer computer usage last year was equivalent to _____ Hours on a _____ computer using _____ processors and _____ MB memory.

Appendix H: Survey of Users

Opinions

The intent of the following section is to determine your opinions of the future of high performance computing -- We encourage you to speculate about your needs five years from now. To set the correct perspective, think about where supercomputing was five years in the past, and how your expectations have changed over the past five years.

1) Do you expect that your area of research will require access to supercomputing resources in the future?

- No Opinion
- Yes
- No

2) During the next several years, what fraction of your computing needs do you think could met by:

- * (A) Desk top workstations
 - 0-20%
 - 21-40%
 - 41-60%
 - 61-80%
 - 81-100%
 - Do not know
- * (B) Mid Range Systems (Department/University owned with a market value from \$100K to \$2M)
 - 0-20%
 - 21-40%
 - 41-60%
 - 61-80%
 - 81-100%
 - Do not know
- * (C) Highest Performance Systems (National)
 - 0-20%
 - 21-40%
 - 41-60%
 - 61-80%
 - 81-100%
 - Do not know

3) For the "Highest Performance Systems", how do you rank your projected needs in the following areas:

- * (A) Processing speed
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (B) Large Memory size
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (C) Large Mass Storage availability
 - No Opinion

Appendix H: Survey of Users

- Unimportant
- Important
- Very Important
- Critical
- * (D) Networking Bandwidth for remote access
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (E) Large I/O Bandwidth (to local peripherals)
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (F) Network Bandwidth for internode communication
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (G) System availability and reliability
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (H) System security
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical

Appendix H: Survey of Users

4) How important do you believe that National Centers will be in providing each of the following services (located at the centers) for your research or educational projects.

- * (A) Diversity of computer architectures
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (B) Visualization facilities and software (including Virtual Environments)
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (C) Information processing
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (D) Consulting Services
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (E) Third party or commercial software
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (F) Expertise in your research topic/Research Team Formation
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (G) Assistance in Code Porting
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (H) Training
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (I) Communication software
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (J) Data Services (massive databases)
 - No Opinion
 - Unimportant

Appendix H: Survey of Users

- Important
- Very Important
- Critical
- * (K) Repository for Data and Data Dictionaries
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical
- * (L) Internet software development
 - No Opinion
 - Unimportant
 - Important
 - Very Important
 - Critical

5) I expect to obtain good results for my codes on the following computers or their successors in the next few years (check all that apply)

- Vector systems (such as Cray C/T90, Convex C3/4, or IBM Mainframe:ES9000)
- MPPs (such as CM-5, T3D, SP2, or Intel Paragon)
- Shared memory microprocessor systems (such as SGI Power Challenge, Convex Exemplar or Cray CS6400)
- Networks of microprocessors or workstations
- No experience

6) I expect that in the future, my largest runs at a National Supercomputer Center will last _____ Hours on a

- Vector Multiprocessor
 - Loosely coupled Network of Workstations
 - Tightly coupled Network of Workstations
 - Massively Parallel Processors
 - Shared Memory Microprocessors
 - Don't know
- using _____ (number of) processors and _____ MB memory.

7) Would an NSF Supercomputer Centers Program organized on disciplinary models (i.e.; a center for Physics, another for Chemistry, etc.) be of greater value to you as a user?

- yes
- no
- No Opinion

8) Have the National Supercomputing Centers been important in facilitating your interactions with researchers from other disciplines?

- yes
- no
- No Opinion

9) Please comment on the existing NSF Supercomputer Centers program. You may enter observations of the advantages, disadvantages, areas which should be improved, or services that are so important, they shouldn't be touched.

Appendix H: Survey of Users

10) Do you have any other comments, or suggestions for a different kind of program or national infrastructure that would be better at meeting the high performance computing needs of the computational science and engineering, information processing, and computer science researchers?

THE FORM IS NOW COMPLETED. Please send your responses by e-mail to survey@osc.edu.

Thank you for your patience and cooperation. When the survey is complete, we will try to send a synopsis of the responses to the e-mail address you provided.