A RESEARCH AND DEVELOPMENT STRATEGY FOR HIGH PERFORMANCE COMPUTING

Executive Office of the President
Office of Science and Technology Policy
November 20, 1987

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20506

This year the Federal Coordinating Council for Science, Engineering, and Technology (FCCSET) Committee on Computer Research and Applications began a systematic review of the status and directions of high performance computing and its relationship to federal research and development. The Committee held a series of workshops involving hundreds of computer scientists and technologists from academia, industry, and government. A result of this effort is the report that follows, containing findings and recommendations concerning this critical issue. It has been sent to the appropriate committees of Congress for their review.

A consistent theme in this report is the need for industry, academia, and government to collaborate and exchange information on future R&D efforts. Partners need to give one another signals as to their intent for future activities, and this report is a necessary first step in that process. The vision it represents must continue to grow. For that reason, I have asked the Committee to initiate the appropriate forums for discussing it further with the computing community.

Another theme has come out of this report: within four decades, the field of computer science has moved from a service discipline to a pervasive technology with a rigorous scientific basis. Computer science has become important to our national security and to our industrial productivity, and as such it provides the United States with many opportunities and challenges. Three of those opportunities are addressed in the report's findings and recommendations: High Performance Computers, Software Technology and Algorithms, and Networking. The fourth recommendation involves the Basic Research and Human Resources that will be required to conduct the other initiatives.

One thing is clear: the competition in an increasingly competitive global market cannot be ignored. The portion of our balance of trade supported by our high performance computing capability is becoming more important to the nation. In short, the United States must continue to have a strong, competitive supercomputing capability if it is to remain at the forefront of advanced technology. For that reason the Office of Science and Technology Policy is encouraging activities among the federal agencies together with the academic community and the private sector.

William R. Graham

William R. Graham
Science Adviser to the President and
Director, Office of Science and Technology Policy

CONTENTS

SUMMARY OF FINDINGS ON COMPUTER RESEARCH AND APPLICATIONS

1. **HIGH PERFORMANCE COMPUTERS: A strong domestic high performance computer industry is essential for maintaining U.S. leadership in critical national security areas and in broad sectors of the civilian economy.**

   - U.S. high performance computer industry leadership is challenged by government supported research and development in Japan and Europe.
   - U.S. leadership in developing new component technology and applying large scale parallel architectures are key ingredients for maintaining high performance computing leadership. The first generation of scalable parallel systems is now commercially available from U.S. vendors. Application-specific integrated circuits have become less expensive and more readily available and are beginning to be integrated into high performance computers.

2. **SOFTWARE TECHNOLOGY AND ALGORITHMS: Research progress and technology transfer in software and applications must keep pace with advances in computing architecture and microelectronics.**

   - Progress in software and algorithms is required to more fully exploit the opportunity offered by parallel systems.
   - Computational methods have emerged as indispensable and enabling tools for a diverse spectrum of science, engineering, design, and research applications.
   - Interdisciplinary research is required to develop and maintain a base of applications software that exploits advances in high performance computing and algorithm design in order to address the "grand challenges" of science and engineering.

3. **NETWORKING: The U.S. faces serious challenges in networking technology which could become a barrier to the advance and use of computing technology in science and engineering.**

   - Current network technology does not adequately support scientific collaboration or access to unique scientific resources. At this time, U.S. commercial and government sponsored networks are not coordinated, do not have sufficient capacity, do not interoperate effectively, and do not ensure privacy.
   - Europe and Japan are aggressively moving ahead of the U.S. in a variety of networking areas with the support of concentrated government and industry research and implementation programs.

4. **BASIC RESEARCH AND HUMAN RESOURCES: Federal research and development funding has established laboratories in universities, industry, and government which have become the major sources of innovation in the development and use of computing technology.**

Summary of Recommendations For a National High Performance Computing Strategy

1. **HIGH PERFORMANCE COMPUTERS:** **The U.S. Government should establish a long range strategy for Federal support for basic research on high performance computer technology and the appropriate transfer of research and technology to U.S. industry.**

2. **SOFTWARE TECHNOLOGY AND ALGORITHMS:** **The U.S. should take the lead in encouraging joint research with government, industry, and university participation to improve basic tools, languages, algorithms, and associated theory for the scientific "grand challenges" with widespread applicability.**

3. **NETWORKING:** **U.S. government, industry, and universities should coordinate research and development for a research network to provide a distributed computing capability that links the government, industry, and higher education communities.**

4. **BASIC RESEARCH AND HUMAN RESOURCES:** **Long term support for basic research in computer science should be increased within available resources. Industry, universities, and government should work together to improve the training and utilization of personnel to expand the base of research and development in computational science and technology.**

A RESEARCH AND DEVELOPMENT STRATEGY FOR HIGH PERFORMANCE COMPUTING

High performance computing refers to the full range of supercomputing activities including existing supercomputer systems, special purpose and experimental systems, and the new generation of large scale parallel architectures.

THE CHALLENGE

In the span of four decades, computing has become one of the most pervasive and powerful technologies for information management, communications, design, manufacturing, and scientific progress.

The U.S. currently leads the world in the development and use of high performance computing for national security, industrial productivity, and science and engineering, but that lead is being challenged. Through an increased foreign industrial capability, the U.S. technology lead in computing has diminished considerably in recent years, but the U.S. continues to maintain strength in basic science and technology. The technology is changing rapidly and the downstream rewards for leadership are great. Progress in computing can be accelerated through the continued pioneering of new hardware, software, algorithms, and network technology and the effective transition of that technology to the marketplace. A shared computing research and development vision is needed to provide to government, industry, and academia a basis for cooperative action. The successful implementation of a strategy to attain this vision and a balanced plan for transition from one generation of technology to the next can result in continued strength and leadership in the forthcoming decades.

High performance computing technology has also become essential to progress in science and engineering. A **grand challenge** is a fundamental problem in science or engineering, with broad applications, whose solution would be enabled by the application of the high performance computing resources that could become available in the near future. Examples of grand challenges are: (1) Computational fluid dynamics for the design of hypersonic aircraft, efficient automobile bodies, and extremely quiet submarines, for weather forecasting for short and long term effects, efficient recovery of oil, and for many other applications; (2) Electronic structure calculations for the design of new materials such as chemical catalysts, immunological agents, and superconductors; (3) Plasma dynamics for fusion energy technology and for safe and efficient military technology; (4) Calculations to understand the fundamental nature of matter, including quantum chromodynamics and condensed matter theory; (5) Symbolic computations including speech recognition, computer vision, natural language understanding, automated reasoning, and tools for design, manufacturing, and simulation of complex systems. Many of these could be considerably advanced by the use of computer systems capable of trillions of operations per second.

THE STRATEGY

A **High Performance Computing Strategy**, involving close coordination of existing programs and augmented effort, is required to address this national challenge. This strategy involves the coordinated pursuit of computing technology goals through joint efforts of government, industry, and academia. The strategy will have impact in clarifying and focusing the direction of Federally-funded computing research, which continues to be the major source of innovation for computing technology and a primary catalyst for industrial development. Government support should be highly leveraged with resources provided by industry participants. To be effective, the strategy should also be defined and continually updated in cooperation with industry and academia by making them participants in developing and implementing a shared vision of the future to ensure continued U.S. leadership.

The high performance computing strategy is designed to sustain and focus basic Federally-funded research and promote the transfer of basic science from the laboratory to U.S. industrial development and finally to the marketplace. Technology development will be encouraged as appropriate to meet immediate needs as well as to create a foundation for long term leadership. Strong emphasis will be placed on continued transfer of the results of government funded R&D to industry and on cooperation with industry to insure the continued strength of American high technology trade in the international marketplace.

The basic elements of the strategy are research and development programs in high performance computer architecture, in custom hardware, in software and algorithms, and in networking technology, all supported by a basic research foundation. In each of these areas, major opportunities exist that require coordinated support and management, building on existing government programs. Access to high performance computing is essential for providing scientists and engineers at research institutions throughout the country with the ability to use the most advanced computers for their work. The strategy needs to concurrently address the appropriate Federal role in each of the basic elements of the R&D process—basic research, applied research, and industrial development—in order to meet long term, intermediate, and short term technology development goals. Explicit attention must be directed to the flow of technology from basic to applied areas and to the marketplace, as well as back into the research community to create the next generation of computing infrastructure, achieving a cumulative effect. Technology developments within individual element areas will contribute extensively to other activities. Simultaneous and coordinated pursuit of the areas is therefore an important element of the strategy.

Current Status and Trends

• **High performance computing systems.** Improvements in materials and component technology are rapidly advancing computer capability. Memory and logic circuits are continuing to improve in speed and density, but as fundamental physical limits are approached, advances are being sought through improved computer architectures, custom hardware, and software. Computer architecture has begun to evolve into large scale multiple processor systems, and in the past four years a first generation of scalable parallel systems has progressed from the research laboratory to the marketplace. Scalable architectures provide a uniform approach that enables a wide range of capacity, from workstations to very high performance computers. Application-specific integrated circuits, such as for real-time signal processing, are being incorporated into special purpose computers.

At current performance levels our ability to model many important science, engineering, and economic problems is still limited. Formulations of computational models presently exist that for realistic solutions would require speeds of teraflops (trillions of floating point operations per second) and equivalent improvement in memory size, mass storage, and input/output systems. In addition, symbolic processing is complementing and enhancing numeric approaches. Achievement of this performance level in the next 5 years appears to be a feasible goal, based on credible extrapolations of processor capability, number of processors, and software sophistication. In developing the new architectural approaches, however, careful collaboration will be required with the applications community to assess the various approaches and to achieve transition to the new approaches where appropriate. As transitions are made, the high performance computing industry should strive to maintain its continued leadership and competitveness.

• **Software technology and algorithms.** As high performance computing systems evolve and become more critical in science, engineering, and other applications domains, software technology becomes an increasingly central concern. As experienced in many U.S. space and defense programs, for example, software can become the dominant computational cost element in large systems because of the need to support evolution throughout the system life cycle from design and development to long term maintenance and transition to the next generation. Future software environments and tools should support the development of trustworthy systems capable of evolution, while increasing productivity of developers and users of the systems. Effective exploitation of the performance potential of the emerging parallel systems poses a special challenge both to software and to algorithm design.

High performance computing offers scientists and engineers the opportunity to use computer models to simulate conditions difficult or impossible to create and measure in the laboratory. This new paradigm of computational science and engineering offers an important complement to traditional theoretical and experimental approaches, and it is already having major impact in many areas. New approaches combining numeric and symbolic methods are emerging. The development of new instruments and data generation methods in fields as diverse as genetics, seismology, and materials accelerates demand for computational power. In addition, the opportunity is created to coordinate and focus effort on important grand challenges, such as computational fluid dynamics, weather forecasting, plasma dynamics, and other areas.

• **Computer network technology.** A modern high speed research network is one of the elements needed to provide high performance distributed computation and communication support for research and technology development in government, academia, and industry. A coordinated research network based on very high bandwidth links would enable the creation of large-scale geographically distributed heterogeneous systems that link multiple high performance workstations, databases, data generation sources, and extremely high performance servers as required, in order to provide rapid and responsive service to scientists and engineers distributed across the country. The existing national network is a collection of loosely coupled networks, called an internet, based on concepts pioneered in the U.S.

Technical issues being addressed include utilization of fiber optics to improve performance for the entire research and higher education enterprise of the nation. An additional issue of pressing concern, particularly within the governmental and industrial sectors, is that of computer and network security to ensure privacy and trustworthiness in a heterogeneous network environment. At present, responsibility for privacy and the assurance of trust are vested principally in the computers and switching nodes on the network. Further research, already actively underway, is urgently needed to develop models, methodology, algorithms and software appropriate to the scale of a coordinated research network.

- **Basic research and human resources in Computer and Computational Science.** Federal funding has historically been, and will likely remain, a major source of support for important new ideas in computing technology. Carefully managed and stable funding is required to maintain vigorous research in computer and computational science and sufficient growth in computer science manpower. It is important to maintain the strength of the existing major research centers and to develop new research activity to support the growth in computer and computational science. Interactions should be fostered among academia, industry, and national laboratories to address large problems and to promote transfer of technology. In the longer term, enhancement of the computing technology base will have significant impact in productivity, efficiency, and effectiveness of government, industry, and the research community.

IMPACT

Computing technology is vital to national security. Advanced computer systems and software are now integral components in most major defense, intelligence, and aerospace systems. Computing technology has a central role in energy research, oil exploration, weapons research, aircraft design, and other national security technology areas.

Major advances in science and engineering have also accrued from recent improvements in supercomputing capability. The existence of machines with hundred megaflop (hundreds of millions of floating point operations per second) speed and multimillion word memories has allowed, for the first time, accurate treatment of important problems in weather prediction, hydrodynamics, plasma physics, stress analysis, atomic and molecular structure, and other areas. The emerging machines with 1 to 10 gigaflop (billions of flops) speed and 100 to 300 million word memories are expected to produce comparable advances in solving numeric and symbolic problems.

Many of these advances in science and engineering are the result of the application of high performance computing to execute computational simulations based on mathematical models. This approach to science and engineering is becoming an important addition to traditional experimental and theoretical approaches. In applications such as the National Aerospace Plane, supercomputing provides the best means to analyze and develop strategies to overcome technical obstacles that determine whether the hypersonic vehicle can fly beyond speeds of Mach seven, where wind tunnels reach their maximum capability. The list of applications for which supercomputing plays this kind of role is extensive, and includes nearly all high-technology industries. The extent of its usage makes supercomputing an important element in maintaining national competitiveness in many high technology industries.

The high performance computing strategy will have impact in many sectors of the economy. Nearly all sectors of advanced industry are dependent on computing infrastructure. Any improvement in computing capability will have substantial leveraged impact in broad sectors, particularly as applications software increases in power and sophistication.

The computer hardware industry alone amounted to $65 billion in 1986, and U.S. technical market dominance, long taken for granted, is now challenged in this and other critical areas, including networking, microsystems and custom high-performance integrated circuit technology. Foreign investment in computing research and technology has grown considerably in the last decade.

As stated in the report of the White House Science Council, *Research in Very High Performance Computing*, November 1985, "The bottom line is that any country which seeks to control its future must effectively exploit high performance computing. A country which aspires to military leadership must dominate, if not control, high performance computing. A country seeking economic strength in the information age must lead in the development and application of high performance computing in industry and research."

BACKGROUND

The Federal Coordinating Council on Science, Engineering and Technology (FCCSET) was established by Congress under the Office of Science and Technology Policy (OSTP) to catalyze interagency consideration of broad national issues and to coordinate various programs of the Federal government. The FCCSET in turn, established a series of committees, with interagency participation to assess and recommend action for national science and technology issues. The committees have become recognized as focal points for interagency coordination activity, addressing issues that have been identified by direct requests through the OSTP and indirect requests by member agencies (such as the NSF requirement to provide an update to the Lax Report on Large Scale Computing in Science and Engineering). These studies have enabled the FCCSET Committee on Computer Research and Applications to develop a national view of computing technology needs, opportunities, and trends.

From its inception, the FCCSET Committee on Supercomputing (the original name of this committee) was chartered to examine the status of high performance computing in the U.S. and to recommend what role the Federal Government should play regarding this technology. The committee issued two reports in 1983 that provided an integrated assessment of the status of the supercomputer industry and recommended government actions. The FCCSET Committee on Computer Research and Applications concluded that it would be proper to include an update of the earlier reports to address the changes that have occurred in the intervening period as a complement to the technical reports. The review was based upon periodic meetings with and site visits to supercomputer manufacturers and consultation with experts in high performance scientific computing. White papers were contributed to this report by industry leaders and supercomputer experts. The report was completed in September 1987 and its findings and recommendations are incorporated in the body of this report.

In developing the recommendations presented in this report, the FCCSET Committee reviewed findings and recommendations from a variety of sources, including those mentioned above. A related activity has been the preparation by the White House Science Council (WHSC) Committee on Research in Very High Performance Computing of the report *Research in Very High Performance Computing*, November 1985. The WHSC Committee, composed of respected individuals from academia, industry, and government, made recommendations related to the issues more recently addressed by the FCCSET Committee. In the areas addressed by both committees, there is a significant consistency of recommendations, and, indeed, progress in recent months further strengthens the case for the recommendations. The convergence of views expressed in the many reports, the strong interest in many sectors of government in developing a policy, the dramatic increase in foreign investment and competitiveness in computing and network technology, and the considerable progress in computing technology development worldwide are all indicators of the urgency of developing and implementing a strategy for nationwide coordination of high performance computing under the auspices of the government.

One of the of the direct requests that this report responds to is in Public Law 99-383, August 21, 1986, in which Congress charged the Office of Science and Technology Policy to conduct a study of critical problems and of current and future options regarding communications networks for research computers, including supercomputers, at universities and federal research facilities in the United States. The legislation asked that requirements for supercomputers be addressed within one year and requirements for all research computers be addressed within two years. Dr. William R. Graham, Director of the Office of Science and Technology Policy, subsequently charged the Federal Coordinating Council on Science Engineering and Technology (FCCSET) Committee on Computer Research and Applications to carry out the technical aspects of the study for OSTP.

It was recognized by the FCCSET Committee on Computer Research and Applications that networking technology needs to be addressed in the context of the applications of computing and the sources of computing power that are interconnected using the network technology. This report, therefore, presents an integrated set of findings and recommendations related to Federal support for computer and related research.

Three subcommittees carried out the work. Each of these committees contributed to the Findings and Recommendations contained in this report. The result is an integrated set of recommendations that addresses the technical areas.

- The **Subcommittee on Computer Networking, Infrastructure, and Digital Communications** invited experts in government, industry and academia to write white papers on networking trends, requirements, concepts applications, and plans. A workshop involving nearly 100 researchers, network users, network suppliers, and policy officials was held in San Diego, California in February 1987 to discuss the white papers and to develop the foundation for the report. Workshop leaders and other experts later met in Washington to summarize the workshop discussions and focused on six topics: access requirements and future alternatives, special requirements for supercomputer networks, internet concepts, future standards and services requirements, security issues, and the government role in networking. As a result of this work, the participants recommended that no distinction should be made between networks for supercomputers and other research computers and that the final report to the Congress should address networks generally. The requirements for both supercomputers and for other research computers are, therefore, addressed in this report.

- The **Subcommittee on Science and Engineering Computing** assessed computing needs related to computational science and engineering. The committee focused its deliberations on requirements for high performance computing, on networking and access issues, and on software technology and algorithms. Under the auspices of the Society for Industrial and Applied Mathematics (SIAM), and with the support of NSF and DOE, a workshop involving 38 recognized leaders from industry, academia, and national laboratories was held at Leesburg, Virginia in February 1987 on research issues in large-scale computational science and engineering. This workshop focused on advanced systems, parallel computing and applications. As a result of the workshop report, recommendations were made related to the role of computing technology in science and engineering applications.

- The **Subcommittee on Computer Research and Development** assessed the role of basic research, the development of high performance computing technology, and issues related to software technology. Contributing to this activity were two workshops. The National Science Foundation (NSF) Advisory Committee for Computer Research reviewed the field and produced an Initiatives Report in May 1987. This report recommended investment in three areas, including parallel systems and software technology. In September 1987, the Defense Advanced Research Projects Agency (DARPA) held a workshop on advanced computing technology in Gaithersburg, Maryland involving 200 researchers from academia, industry, and government. The workshop focused on large-scale parallel systems and software approaches to achieving high performance computing.

An important result of the activity of the FCCSET Committee on Computer Research and Applications and its subcommittees is that increased coordination among the Government elements is necessary to implement a strategy for high performance computing. The findings and recommendations presented here represent a consensus reached among the subcommittees and convey the powerful and compelling vision that emerged. As a result of this process, the next step would be for the members of the Committee on Computer Research and Applications to develop a plan to help ensure that the vision is shared between government, academia, and American industry. Subsequently, the Committee should develop an implementation plan for Federal government activities, including a detailed discussion of overall priorities.

1. HIGH PERFORMANCE COMPUTERS

• FINDING: A strong domestic high performance computer industry is essential for maintaining U.S. leadership in critical national security areas and in broad sectors of the civilian economy.

U.S. prominence in technology critical to national defense and industrial competitiveness has been based on leadership in developing and exploiting high performance computers. This preeminence could be challenged by dependency upon other countries for state of the art computers. Supercomputer capability has contributed for many years to military superiority. In addition, industrial applications now constitute more than half of the supercomputer market and are an important factor in U.S. industrial competitiveness. However, continued progress in computational science and engineering will depend in large part on the development of computers with 100 to 1000 times current capability for important defense, scientific, and industrial applications. These applications are represented by the grand challenges.

■ U.S. high performance computer industry leadership is challenged by government supported research and development in Japan and Europe.

The U.S. currently leads the world in research, development, and use of supercomputers. However, this leadership faces a formidable challenge from abroad, primarily from the Japanese. The 1983 FCCSET report stated that "The Japanese have begun a major effort to become the world leader in supercomputer technology, marketing, and applications." Most of the analyses and projections advanced in support of that statement have proven to be accurate.

Japanese supercomputers have entered the marketplace with better performance than expected. Japanese supercomputer manufacturers have attained a high level of excellence in high speed, high density logic and memory microcircuits required for advanced supercomputers. As a result, some U.S. computer manufacturers are dependent on their Japanese competitors for sole supply of critical microcircuits. Japanese manufacturers, universities, and government have demonstrated the ability to cooperate in developing and marketing supercomputers as well as in advancing high performance computing. Recent successes in dominating related high-technology markets underscore their financial, technical, and marketing capability.

- **U.S. leadership in developing new component technology and applying large scale parallel architectures are key ingredients for maintaining high performance computing leadership. The first generation of scalable parallel systems is now commercially available from U.S. vendors. Application-specific integrated circuits have become less expensive and more readily available and are beginning to be integrated into high performance computers.**

  The current generation of supercomputers achieve their performance through the use of the fastest possible individual components, but with relatively conservative computer architectures. While these computers currently employ up to eight parallel processors, their specific architectures cannot be scaled up significantly. Large scale parallel processing, in which the computational workload is shared among many processors, is considered to be the most promising approach to producing significantly faster supercomputers. The U.S. is currently the leader in developing new technology as well as components. However, exploiting these techniques effectively presents significant challenges. Major effort will be required to develop parallel processing hardware, algorithms, and software to the point where it can be applied successfully to a broad spectrum of scientific and engineering problems.

  Government funded R&D in universities and industry has focused on an approach to large-scale parallelism that is based on aggressive computer architecture designs and on high levels of circuit integration, albeit with somewhat slower individual components. Unlike current supercomputers, the resulting systems employ 100s to 10,000s of processors. Equally important, these architectures are scalable to higher levels of parallelism with corresponding increase in potential performance.

  The first generation of scalable parallel systems is now commercially available from U.S. vendors. These systems have demonstrated high performance for both numeric and non-numeric, including symbolic processing. Comparable systems do not yet exist outside the U.S. The second generation, with higher speed individual components and more parallelism, is already in development here. Experience with these systems has shown that, even with existing software, they are effective for certain classes of problems. New approaches to software for these large-scale parallel systems are in the process of emerging. These approaches suggest that parallel architecture may be effective for wide classes of scientific and engineering problems. An important benefit of the scalable architectures is that a single design, with its attendant components and software, may prove to be useful and efficient over a performance range of 10 to 100 or more. This allows one design to be used for a family of workstations, mini-supercomputers, and supercomputers.

• **<u>RECOMMENDATION:</u> The U.S. Government should establish a long range strategy for Federal support for basic research on high performance computer technology and the appropriate transfer of research and technology to U.S. industry.**

The program should build upon existing government supported efforts. However, government funding should not be viewed as a substitute for private capital in the high performance computer marketplace. A primary objective is to ensure continued availability of domestic sources for high performance computers that are required for Federal programs, both civilian and defense. These actions should include:

• Government should support, when appropriate for mission requirements, the acquisition of prototype or early production models of new high performance computers that offer potential for improving research productivity in mission areas. These computers could be placed in centers of expertise in order to allow sophisticated users to share initial experiences with manufacturers and other users, and to develop software to complement the vendor's initial offerings. These initial acquisitions should not require the vendor to supply mature operating systems and applications software typical of production computers. However, a criterion for acquisition should be that the hardware designs reflect a sensitivity to software issues, and that the computer has the potential for sustained performance in practical applications that approaches the peak hardware performance.

• Government agencies should seek opportunities to cooperate with industry in jointly funded R&D projects, concentrating especially on those technologies that appear scalable to performance levels of trillions of operations per second (teraops) for complex science, engineering, and other problems of national importance. Systems are needed for both numeric and symbolic computations.

However, since government mission requirements typically exceed those of industrial applications, cooperating with industry in R&D for computers to meet these missions will help to assure that the necessary computers are available. It will also drive supercomputer development at a faster pace than would be sustained by commercial forces alone, an important factor retaining and increasing U.S. leadership in this area.

• Government agencies should fund basic research to lay the foundation for future generations of high performance computers. Steps should be taken to ensure that development of state-of-the-art computers continues to be monitored for appropriate export controls.

2. SOFTWARE TECHNOLOGY AND ALGORITHMS

- **FINDING: Research progress and technology transfer in software and applications must keep pace with advances in computing architecture and microelectronics.**
  - **Progress in software and algorithms is required to more fully exploit the opportunity offered by parallel systems.**
  - **Computational methods have emerged as indispensable and enabling tools for a diverse spectrum of science, engineering, and design research and applications.**
  - **Interdisciplinary research is required to develop and maintain a base of applications software that exploits advances in high performance computing and algorithm design in order to address the "grand challenges" of science and engineering.**

A **grand challenge** is a fundamental problem in science and engineering, with broad application, whose solution will be enabled by the application of the high performance computing resources that could become available in the near future.

As high performance computing systems evolve and are applied to more challenging problems, it is becoming increasingly clear that advances in software technology and applications are essential to realize the full performance potential of these systems. Software development, analysis, and adaptation remain difficult and costly for traditional sequential systems. Large scale complex systems including parallel systems pose even greater challenges. Market pressures for the early release of new computing system products have created a tradition of weak systems software and inadequate programming tools for new computers.

Current approaches to software development provide only limited capabilities for flexible, adaptable, and reusable systems that are capable of sustained and graceful growth. Most existing software is developed to satisfy nearer term needs for performance at the expense of these longer term needs. This is particularly the case for applications in which specific architectural features of computers have been used to obtain maximum performance through low level programming techniques. The lack of portability of these programs significantly raises the cost of transition to newer architectural approaches in many applications areas. Approaches are beginning to emerge in the research community that have a potential to address the reuse and portability problems.

Experiments with parallel computers have demonstrated that computation speeds can increase almost in direct proportion to the number of processors in certain applications. Although it is not yet possible to determine in general the most efficient distribution of tasks among processors, important progress has nonetheless been made in the development of computational models and parallel algorithms for many key problem areas.

Access to advanced computing systems is an important element in addressing this problem. Experience has shown that the quality of systems and applications software increases rapidly as computing systems are made more available. Initial generic operating systems and extensions to existing programming languages can provide access through coupling high performance computers with existing workstations using either direct or network connections. However, in order to achieve the full potential impact of large scale parallel computing on applications, major new conceptual developments in algorithms and software are required.

The U.S. leads in many areas of software development. The Japanese, however, also recognize the need for high quality software capability and support in order to develop and market advanced machines. They have demonstrated the ability to effectively compete, for example in the area of sophisticated vectorizing compilers.

The U.S. will need to encourage the collaboration of computer scientists, mathematicians, and the scientists in critical areas of computing applications in order to bring to bear the proper mix of expertise on the software systems problem. Such collaboration will be enhanced by network technology, which will enable geographically dispersed groups of researchers to effectively collaborate on "grand challenges." Critical computer applications include problems in fluid dynamics, plasma physics, elucidation of atomic and molecular structure, weather prediction, engineering design and manufacturing, computer vision, speech understanding, automated reasoning, and a variety of national security problems.

- **<u>RECOMMENDATION:</u> The U.S. should take the lead in encouraging joint research with government, industry, and university participation to improve basic tools, languages, algorithms, and associated theory for the scientific "grand challenges" with widespread applicability.**

Software research should be initiated with specific focus on key scientific areas and on technology issues with widespread applicability. This research is intended to accelerate software and algorithm development for advanced architectures by increased early user access to prototype machines. It would also provide settings for developing advanced applications for production machines. Software technology needs to be developed in real problem contexts to facilitate the development of large complex and distributed systems and to enable transition of emerging parallel systems technology into the computing research community and into the scientific and engineering applications communities.

As part of a mixed strategy, longer term and more basic software problems of reliability and trust, adaptability, and programmer productivity must continue to be addressed. Languages and standards must be promoted that permit development of systems that are portable without sacrificing performance.

In applications areas including computational science and engineering, technology should be developed to support a smooth transition from the current software practice to new approaches based on more powerful languages, optimizing compilers, and tools supported by algorithm libraries. The potential of combining symbolic and numeric approaches should be explored. Progress in these areas will have significant impact on addressing the "grand challenges" in computational science and engineering. Although there are many pressing near term needs in software technology, direct investment in approaches with longer term impact must be sustained if there is to be significant progress on the major challenges for software technology while achieving adequate system performance.

3. NETWORKING

- **FINDING: The U.S. faces serious challenges in networking technology which could become a barrier to the advance and use of computing technology in science and engineering.**

  - **Current network technology does not adequately support scientific collaboration or access to unique scientific resources. U.S. commercial and government sponsored networks presently are not coordinated, do not have sufficient capacity, do not interoperate effectively, and do not ensure privacy.**
  - **Europe and Japan are aggressively moving ahead of the U.S. in a variety of networking areas with the support of concentrated government and industry research and implementation programs.**

Computer network technology provides the means to develop large scale distributed approaches to the collaborative solution of computational problems in science, engineering, and other applications areas. Today, researchers sharing a local area network are able to exploit nearly instantaneous communication and sharing of data, creating an effect of linking their workstations and high performance servers into a single large scale heterogeneous computing facility. This kind of capability is now appearing in larger scale campus-wide computer networks, enabling new forms of collaboration. National networks, on the other hand, have low capacity, are overloaded, and fail to interoperate successfully. These have been expanded to increase the number of users and connections but the performance of the underlying network technology has not kept pace with the increased demands. Therefore, the networks which in the 1970s had significant impact in enabling collaboration, are now barriers. Only the simplest capabilities, such as electronic mail and small file transfers, are now usable. Capacity, for example, is orders of magnitude less than the rates required, even if the network is used only for graphics.

Other countries have recognized the value of national computing networks, and, following the early U.S. lead, have developed and installed national networks using current technology. As a result, these countries are now much better prepared to exploit the new opportunities provided by distributed collaborative computing than the U.S. is at the present time. The basic technologies for later generations are also being developed in the U.S., but there have been no major efforts to apply them to address the needs.

Applications include (1) distributed access to very large databases of scientific, engineering, and other data, (2) high bandwidth access to and linking among shared computational resources, (3) high bandwidth access to shared data generation resources, (4) high bandwidth access to shared data analysis resources, such as workstations supporting advanced visualization techniques.

A longer term goal is the creation of large scale geographically distributed heterogeneous systems that link multiple superworkstations and high performance supercomputers to provide service to scientists and engineers distributed across the country. A well-coordinated national network could link these resources together when required on an *ad hoc* basis to provide rapid response to computational needs as they arise. This could reduce the number of sites needed for the physical presence of supercomputers. Present access to computer networks by researchers is dependent upon individual funding or location. There is unnecessary duplication in the links from various agencies to each campus. The development of improved networking facilities could greatly stimulate U.S. research and provide equitable access to resources.

Many scientific research facilities in the U.S. consist of a single, large, and costly installation such as a synchrotron light source, a supercomputer, a wind tunnel, a particle accelerator, or a unique database. These facilities provide the experimental apparatus for groups of scientific collaborators located throughout the country. Wide area networks are the logical mechanism for making data from such facilities more easily accessible nationwide. An important issue is that of computer and network security to ensure privacy and trustworthiness in a heterogeneous network environment. At present, responsibility for privacy and the assurance of trust are vested principally in the computers and switching nodes on the network.

Existing government-supported wide-area networks include ARPANET, HEPNET, MFENET, NSFNET, NASNET, MILNET, and SPAN, as well as private and commercial facilities such as TYMNET, TELENET, BITNET, and lines leased from the communication carriers. Longer-range estimates vary, but it is expected that by the year 1995 the nation's research community will be able to make effective use of a high capacity national network with capacity measured in billions of bits per second. Without improved networks, speed of data transmission will be a limiting factor in the ability of researchers to carry out complex analyses. The digital circuits most widely available today with transmission speeds of 56 kilobits per second (kb/s) are impediments to leading edge research and to optimal remote high performance computer use.

Point-to-point connections require interconnects through multiple vendors with cumulative costs. Greater network speed can reduce the time required to perform a given experiment and increase both the volume of data and the amount of detail that can be seen by researchers. Scientists accessing supercomputers would benefit because access speed is often critical in their work. Improved functionality frees scientists to concentrate directly on their experimental results rather than on operational details of the network. Increased network size extends these opportunities to thousands of individuals at smaller academic institutions throughout the nation. These modernization measures would significantly enhance the nation's position in scientific research. A national network would help maintain the U.S. leadership position in computer architectures, microprocessors, data management, software engineering, and innovative networking facilities, and promote the development of international networking standards based on U.S. technology.

Integrated Systems Digital Networks (ISDN–voice and data) have been installed abroad on a national or regional scale. Research abroad is being conducted on service up to 1 Gb/s. Within the next five years, Integrated Services Digital Network (ISDN) circuits ranging from 64 kb/s to 1.5 Mb/s will be available in the larger metropolitan areas of the U.S. However, these services will fall short of the requirements for computer networks. By 1988 more than fifty Campus Area Networks will be operational at speeds approaching 100 Mb/s. Wide area networks operating at 1.5 Mb/s or less will not be able to handle the data volume expected.

Japan and Europe have extensive efforts with experimental nets in intermediate (40Mb) and high (gigabit) range. Japan is studying operational aspects of fiber nets using their national research network as a testbed, which includes exploring the feasibility of fiber optic services to residences.

To estimate the network bandwidth needed to support research at a major installation, the kinds and volume of traffic that would be used have been estimated at a representative campus, extrapolated ten years into the future. Three models were used to compute three independent estimates of the requirements for bandwidth needed by type of work, information needs by type of user, and information flowing at the installation boundary. In each model, the peak bandwidth was estimated for each type of service. For example, in the Task model, the need is dominated by that of at least one researcher to receive full color and full–motion high resolution images. A high–resolution color image contains about 30 megabits of information, so that a display rate of 30 frames per second requires a bandwidth of nearly one gigabit per second (Gb/s). In the User model, a research university with 35,000 students and 3,000 faculty and research staff using a mix of bandwidths again requires an aggregate bandwidth of approximately one Gb/s. In the Edge of the Installation model, bandwidth is estimated by the types of remote facilities being accessed and the expected number of simultaneous users; typical facilities include particle accelerators, supercomputers, and centers for imaging and/or animation. The aggregate bandwidth needed is one Gb/s. Thus three independent means of estimating bandwidth arrive at nearly the same requirement for a large research installation, and one Gb/s can confidently be used as a lower bound on the bandwidth of a national research network.

- **<u>RECOMMENDATION:</u> U.S. government, industry, and universities should coordinate research and development for a research network to provide a distributed computing capability that links the government, industry, and higher education communities.**

  A research network should be established in a staged approach that supports the upgrade of current facilities and development of needed new capabilities. Achievement of this goal would foster and enhance the U.S. position of world leadership in computer networking as well as provide infrastructure for collaborative research. The FCCSET Committee on Computer Research and Applications should provide a forum for interagency cooperations. Elements of the plan should include:

  - *Stage 1.* Upgrade existing facilities in support of a transition plan to the new network through a cooperative effort among major government users. The current interagency collaboration in expanding the Internet system originated by DARPA should be accelerated so that the networks supported by the agencies are interconnected over the next two years.

  - *Stage 2.* The nation's existing networks that support scientific research should be upgraded and expanded to achieve data communications at 1.5 Mb/sec for 200 to 300 U.S. research institutions.

  - *Stage 3.* Develop a system architecture for a national research network to support distributed collaborative computation through a strong program of research and development. A long-term program is needed to advance the technology of computer networking in order to achieve data communication and switching capabilities to support transmission of three billion bits per second (3 Gb/s) with deployment within fifteen years.

  - Develop policy for long term support and upgrading of current high performance facilities, including timetables for backbone and connection development, industry participation, access, agency funding, tariff schedules, network management and administration. Support should be given to the development of standards and their harmonization in the international arena.

  Until the national research network can replace the current system, existing networks should be maintained and modified as they join the national network. Remedial action should be initiated as soon as possible. Upgrading the backbone to at least 1.5 Mb/s should be accomplished by 1990. This will ensure that the new generation of high performance computing can be effectively interconnected.

  Industry should be encouraged to participate in research, development, and deployment of the national research network. Telecommunication tariff schedules which have been set for voice transmission should be reviewed in light of the requirements for transmission of data through computer networking.

Prompt effective coordination is needed to increase user participation in the standards development process, to get requirements for standards expressed early in the development process, and to speed the implementation of standards in commercial off-the-shelf products. It is essential that standards development be carried out within the framework of overall systems requirements to achieve interoperability, common user interfaces to systems, and enhanced security.

4. Basic Research and Human Resources

- **<u>FINDING:</u> Federal research and development funding has established laboratories in universities, industry, and government which have become the major sources of innovation in the development and use of computing technology.**

Many of the advances in computer science and technology in the U.S. were made possible by Federal programs of research support to universities and industry. For example, the advances that have occurred since 1983 in the area of parallel computing are the direct result of Federal research investment through agencies including DARPA and NSF. In the area of application of supercomputers to science and engineering, the majority of this investment came from the NSF Advanced Scientific Computing centers. In the area of parallel architectures, the major investment came from the DARPA Strategic Computing Program. Programs sponsored by DOE, NASA, and Defense to support critical mission needs have been a major source of investment in computational applications research. In industry, support for basic research is only a small fraction of industry research most of which is focused on nearer term product development. This can be attributed in part to the long term and high risk nature of basic research, but a more significant inhibitor of investment is the difficulty in the computer industry of maintaining proprietary protection for certain kinds of key fundamental advances.

- **<u>RECOMMENDATION:</u> Long term support for basic research in computer science should be increased within available resources. Industry, universities, and government should work together to improve the training and utilization of personnel to expand the base of research and development in computational science and technology.**

Maintain vigorous research in Computer Science and sufficient growth of computer science manpower to support the scientific/technological basis of the computer field. Foster interactions among academia, industry, and national laboratories by creating interdisciplinary teams to address large scale problems. Extend the technology base to attain significant impact on competitiveness and industrial productivity.

Innovative very high performance computing systems should be made available to universities and basic research laboratories in order to assist in the evaluation and exploitation of new technology and new industrial innovations.

Continue the following successful approaches to basic research and development: (1) The practice of loosely coordinated and flexible basic research supported through various federal sectors and applied to a diversity of institutions, (2) The mixed strategy of peer review to support a broad range of exploratory basic research throughout the academic community and the complementary technical program management approach of larger scale experimental systems programs which exploit new opportunities as they emerge, (3) Support for individuals and small groups in theoretical areas, (4) The practice of supporting the relevant basic research as part of larger experimental systems projects.

IMPLEMENTATION

Success of the National High Performance Computing Strategy will require an attitude of cooperation in which academia, industry and government work effectively together in developing and assessing new technology and in achieving the transition of promising new ideas into the marketplace. The rapid pace of developments in computing technology creates a number of implementation challenges that must be addressed explicitly if the Strategy is to have maximum impact.

The FCCSET Committee on Computer Research and Applications provides an appropriate forum for coordination of Federal agency programs. Specifically:

- The subcommittee on Computer Networking, Infrastructure, and Digital Communications will develop a coordinated implementation plan for the national research network.

- The subcommittee on Science and Engineering Computing will review the *grand challenges* through the use of high performance computing systems, including the research that will be involved.

- The subcommittee on Computer Research and Development will review the need for advanced software, algorithms, and hardware for future high performance computing systems.

All of the subcommittees will consider appropriate action to secure a foundation of basic research and human resources. In all three subcommittees we expect some overlap of responsibility and interchange of ideas to be compatible with success.

As has been firmly stated, the full cooperation through a shared vision between government, industry and the research community will be a necessary ingredient for the successful implementation of this strategy. The FCCSET Committee on Computer Research and Applications therefore calls for timely consideration of the vision and strategy by representative bodies of the research community and industry.

It is essential, however, that implementation of the strategy be undertaken in a timely manner. There is a need to follow through on the breakthroughs that occurred partially as a result of federal investment in the early 1980s. The fast pace of development dictates that appropriate Federal efforts are needed to help ensure continued excellence in high speed networking technology and leadership in high performance computing. Foreign investment in technology development in these key areas has increased dramatically. The prudent strategy is to maintain a consistent strong lead in research and to transfer the results as quickly as possible to American industry.

Cost Estimates

Many of the basic elements of the high performance computing strategy are already being implemented as part of ongoing agency programs at DOE, DARPA, NSF, NASA, and other Federal agencies, and important progress is being made. The FCCSET Committee activity has contributed to achieving a shared vision, and early coordination is already occurring in anticipation of implementation of the strategy. Implementation of the strategy involves three principal funding components, including the national research network, joint research to address the "grand challenges," and basic research in high performance computing architecture, custom hardware design, software, algorithms, and supporting technologies. Multiple agencies are involved in the implementation and funding of each of these components.

The funds that would be associated with each of these components are described below. Obviously, any incremental funding must be evaluated and approved within the context of current activities and research needs in other high priority fields. Currently, the Federal government is spending about $500M per year on all aspects of high

Summary of Additional Funds
(Millions of Dollars)

Current Funds			Yr 1	Yr 2	Yr 3	Yr 4	Yr 5
50 [a]	National Research Network	Stage 1	5	5	5	0	0
		Stage 2	5	5	55	55	55
		Stage 3	40	40	40	40	40
150 [b]	Joint Research in Computational Science and Engineering		30	60	90	120	150
300	Basic Research in Computer Science and High Performance Computing		60	120	180	240	300
500	TOTAL (above current funds)		140	230	370	455	545
	Funding Increase by Year (noncumulative)		140	90	140	85	90

a Estimated network research and support in grants and contracts.
b Estimated operating costs for existing computational science facilities.

performance computing. Funding for the activities recommended in this report would increase this base by $140M in additional resources for the first year, growing to an additional $545M per year in 5 years.

National Research Network. Current operating costs for the present collection of research-support networks operated by DARPA, NSF, DOE, and NASA is approximately $50M per year; the figure is uncertain because many subnetworks are funded by increments on research grants and contracts, rather than being centrally supported. Currently the interconnection of existing agencies' networks is planned within existing budgets. A significant increase in investment is needed to achieve the required capability. This investment could occur in three concurrent stages.

The *first stage* activity would involve an immediate upgrade to 1.5 Mbit/sec of the existing research-support networks. This would cost $15M over three years.

The *second stage* would expand upgraded network services (45Mbit/sec) to 200 to 300 research installations, using primarily fiber-optic trunk facilities. Development costs for this stage would be $5M per year of additional funding. Operation of the upgraded network would commence in three to five years, with operating costs of approximately $50M per year. Since the transition from the first stage to the second stage network could not be instantaneous, initially the full operating cost of the second stage network would necessitate additional funding; that requirement will diminish to the extent that the first stage network is phased out.

The goal of the *third stage* would be to deliver one to three Gbit/sec to selected research facilities, and 45 Mbit/sec to approximately 1000 research sites. Research and development costs for this project are estimated at $400M of new funds, spent over ten years; after ten years, operating costs would be about $200M per year unless some tariff relief is achieved.

Joint Research in Computational Science and Engineering. Current operating costs for existing computational science laboratory facilities is approximately $150M per year. Additional investment would be required to upgrade the existing facilities and/or to establish additional joint research activities, with government, industry, and university participation, to address approximately specific problem areas, including selected *grand challenges*. Many of these joint research efforts will involve multiple physical sites connected by the research network. The investment in these research activities supports pursuit of the grand challenges. This includes personnel to develop computational approaches in terms of theory, algorithms, and software, and the acquisition of modern computing equipment. Estimated Federal costs average $15M per year to establish and sustain each grand challenge. The joint research activities would be introduced at the rate of two per year. Overall investment will be approximately $30M per year initially, increasing to $150M per year in five years as new grand challenges are added.

Basic Research in Computer Science and High Performance Computing. Current Federal investment in advanced computer research is estimated at $300M in FY88. Over the past four years, investment in these areas has grown at 15% per year. The rate of increase appears to be declining, however, at a time when increased investment appears to be needed. Sufficient resources should continue to be allocated to take full advantage of the high performance computing opportunities that now exist including design and prototype development of systems capable of trillions of operations per second. A second important element is stable funding, which is required to preserve the long-term strength of the research community.

Other countries are also devoting considerable resources in this area. For example, the Japanese government supports two projects which directly address supercomputer development: The Fifth Generation Project and the Superspeed Project. Support for each of these is estimated to be in excess of $100M per year. In addition to this government support, Japanese industry is investing considerably more to develop high performance computers. Japanese government and industry are also investing amounts comparable to those recommended here to develop high bandwidth research networks.

ACKNOWLEDGMENTS

Office of Science and Technology Policy guidance was provided by Michael Marks. Stephen L. Squires, Defense Advanced Research Projects Agency, acted as Executive Secretary for this report. Technical assistance was provided by William L. Scherlis, Defense Advanced Research Projects Agency; along with Kathleen Bernard, Office of Science and Technology Policy; Charles N. Brownstein, National Science Foundation; Leslie Chow, National Aeronautics and Space Administration; and Michael Crisp, Department of Energy.

FCCSET Committee on Computer Research and Applications

Paul G. Huray (Chair)
Office of Science and Technology Policy

Subcommittees

Science and Engineering Computing

James F. Decker (Chair)
Department of Energy

James Burrows
National Bureau of Standards

John S. Cavallini
Health and Human Services

Melvyn Ciment
National Science Foundation

John Connolly
National Science Foundation

Craig Fields
Defense Advanced Research Projects Agency

Harlow Freitag
Supercomputer Research Center

Randolph Graves
National Aeronautics and Space Administration

Norman H. Kreisman
Department of Energy

Lewis Lipkin
National Institutes of Health

Allan T. Mense
Strategic Defense Initiative Office

David B. Nelson
Department of Energy

C. E. Oliver
Air Force Weapons Lab

John P. Riganati
Supercomputer Research Center

Paul B. Schneck
Supercomputer Research Center

K. Speierman
National Security Agency

Computer Research and Development

Saul Amarel (Chair)
Defense Advanced Research Projects Agency

Donald Austin
Department of Energy

C. Gordon Bell
National Science Foundation

James Burrows
National Bureau of Standards

Bernard Chern
National Science Foundation

Peter Freeman
National Science Foundation

Lee Holcomb
National Aeronautics and Space Administration

Charles Holland
Office of Naval Research

Robert E. Kahn
Computer Science Technology Board

Daniel R. Masys
National Institutes of Health

Robert Polvado
Central Intelligence Agency

David Sadoff
Department of State

William L. Scherlis
Defense Advanced Research Projects Agency

K. Speierman
National Security Agency

Stephen L. Squires
Defense Advanced Research Projects Agency

Charles F. Stebbins
Air Force Systems Command

Daniel F. Weiner, II
Joint Tactical Fusion Program

Computer Networking, Infrastructure and Digital Communications

C. Gordon Bell (Chair)
National Science Foundation

Ronald Bailey
National Aeronautics and Space Administration

Sandra Bates
National Aeronautics and Space Administration

James Burrows
National Bureau of Standards

John S. Cavallini
Health and Human Services

Thomas Kitchens
Department of Energy

James Oberthaler
National Institutes of Health

Dennis G. Perry
Defense Advanced Research Projectes Agency

Arnold Pratt
National Institutes of Health

Shirley Radack
National Bureau of Standards

Rudi F. Saenger
Naval Research Laboratory

Daniel VanBelleghem
National Science Foundation

Stephen Wolff
National Science Foundation