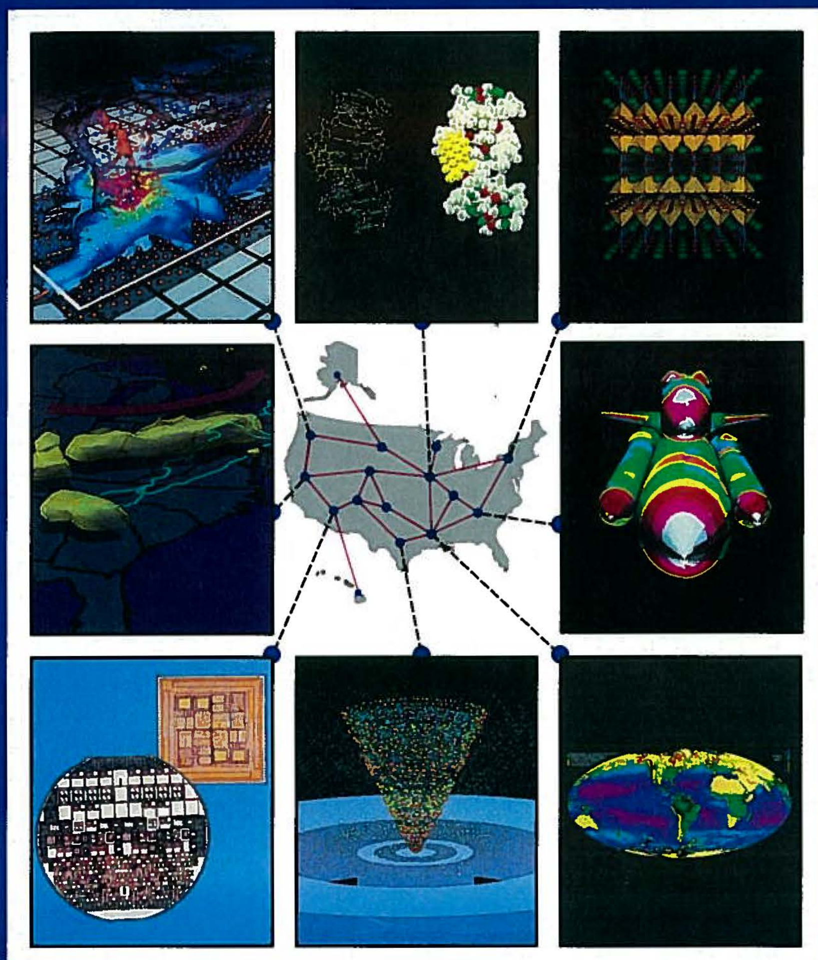
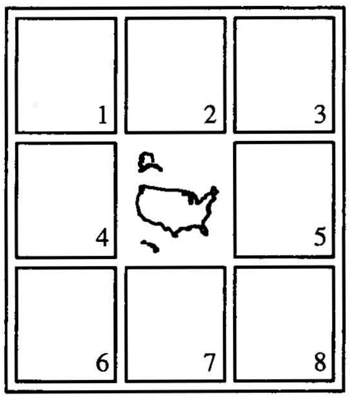
Grand Challenges:
High Performance Computing and Communications

The FY 1992 U.S. Research and Development Program



A Report by the Committee on
Physical, Mathematical, and Engineering Sciences

To Supplement the President's Fiscal Year 1992 Budget



On the Cover:

1. Numerically modelled thunderstorm.
2. Computational model of chemical carcinogen binding with DNA molecule.
3. Visualization of structure of superconducting material.
4. Simulation of acid rain pollutants over Ohio River basin.
5. Aerodynamic characteristics of space vehicle: computer simulated versus wind tunnel data.
6. Photo of wafer and multi–chip prototypes.
7. Numerical simulation of fuel jet.
8. Computer image of earth's biosphere components generated from satellite data.

The images used in this report were produced by ongoing scientific projects in areas of the planned HPCC program. They were selected to illustrate the breadth of subject matter of the HPCC program, and are elaborated upon in Chapter 4. The U.S. map suggests how the National Research and Education Network supports geographically distributed collaborative research activities.

Grand Challenges: High Performance Computing and Communications

The FY 1992 U.S. Research and Development Program

A Report by the Committee on
Physical, Mathematical, and Engineering Sciences

Federal Coordinating Council for Science,
Engineering, and Technology

Office of Science and Technology Policy

To Supplement the President's Fiscal Year 1992 Budget

Office of Science and Technology Policy
Federal Coordinating Council for Science, Engineering, and Technology
Committee on Physical, Mathematical, and Engineering Sciences

Acting Chairman

Charles Herzfeld, Department of Defense

Members

Barry Williamson, Department of the Interior
Charles E. Hess, Department of Agriculture
Robert White, Department of Commerce
William Raub, Department of Health and Human Services
James Decker, Department of Energy
John B. Childers, Department of Education
Norine Noonan, Office of Management and Budget
Eugene Wong, Office of Science and Technology Policy
Arnold Aldrich, National Aeronautics and Space Administration
Erich Bretthauer, Environmental Protection Agency
Fredrick M. Bernthal, National Science Foundation

Executive Secretary

Jane Stutsman, National Science Foundation

FCCSET Directorate

Maryanne C. Bach, Executive Director
Charles H. Dickens, Senior Staff Associate
Jean Grace, Executive Assistant

High Performance Computing and Communications Working Group

David Nelson, Department of Energy, Chairman
Barry Boehm, Defense Advanced Research Projects Agency, Co–Chairman
Charles N. Brownstein, National Science Foundation, Co–Chairman
Lee Holcomb, National Aeronautics and Space Administration, Co–Chairman
Lawrence E. Brandt, National Science Foundation
John S. Cavallini, Department of Energy
Melvyn Ciment, National Science Foundation
Jack D. Fellows, Office of Management and Budget
Stephen M. Griffin, National Science Foundation
Paul E. Hunter, National Aeronautics and Space Administration
Gary M. Johnson, Department of Energy
Thomas A. Kitchens, Department of Energy
Norman H. Kreisman, Department of Energy
Albert T. Landberg, Jr., National Institute of Standards and Technology
Fred Scoresby Long, National Oceanic and Atmospheric Administration
Daniel R. Masys, National Library of Medicine
Joan H. Novak, Environmental Protection Agency
William L. Scherlis, Defense Advanced Research Projects Agency
Paul H. Smith, National Aeronautics and Space Administration
K. Speierman, National Security Agency
Stephen L. Squires, Defense Advanced Research Projects Agency
Stephen Wolff, National Science Foundation

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

MEMBERS OF CONGRESS:

I am pleased to forward with this letter "Grand Challenges: High Performance Computing and Communications, The FY 1992 U. S. Research and Development Program," a report by the Committee on Physical, Mathematical, and Engineering Sciences of the Federal Coordinating Council for Science, Engineering, and Technology, a supplement to the President's Fiscal Year 1992 Budget.

The report presents an ambitious and well-coordinated research and development program designed to sustain and extend U.S. leadership in all advanced areas of computing and networking. The program not only provides a far-sighted vision for the underlying technologies but also gives recognition to the importance of both human resources and those applications that serve major national needs. This is a program of national investment that will bring both economic and social dividends.

The program is strategically related to other key components of the President's overall approach to challenges in science, technology, and education. It provides for the use of improved computational and communications technologies to contribute to more effective solutions of grand challenge problems.

The goal of the Federal High Performance Computing and Communications (HPCC) Program is to accelerate significantly the commercial availability and utilization of the next generation of high performance computers and networks. Recent advances offer the potential for a thousand-fold improvement in useful computing capability and a hundred-fold improvement in available computer communications capability by 1996. These advances will come through improvements in hardware and software. This increased capability will greatly expand the availability of these resources for research and education. It is my personal view, moreover, that the successful implementation of this program will lay the foundation for changes in education at all levels.

Several years of effort on the part of senior government, industry, and academic scientists and managers are reflected in this program. Acting Chairman Charles Herzfeld and his interagency committee members, associates, and staff are to be commended on the excellent work that is manifest in both the program and the report.

D. Allan Bromley
Director

To obtain a copy of this document — send request to:

Committee on Physical, Mathematical, and Engineering Sciences
c/o National Science Foundation
Computer and Information Science and Engineering
1800 G Street, N.W.
Washington, D.C. 20550

Table of Contents

List of Figures

Grand Challenges: High Performance Computing and Communications

The FY 1992 U.S. Research and Development Program

Executive Summary

EXECUTIVE SUMMARY

- High performance computing and computer communications networks are becoming increasingly important to scientific advancement, economic competition, and national security. The technology is reaching the point of having a transforming effect on our society, industries, and educational institutions. The goal of the Federal High Performance Computing and Communications (HPCC) Program is to accelerate significantly the commercial availability and utilization of the next generation of high performance computers and networks in a manner consistent with the Strategic and Integrating Priorities shown in Figure 1.

- The HPCC Program is the result of several years of effort on the part of senior government, industry, and academic scientists and managers to design a research agenda to extend U.S. leadership in high performance computing and networking technologies.

- For FY 1992 the HPCC Program proposes to invest $638 million in the four complementary and coordinated components shown in Figure 1. This investment represents a $149 million, or 30%, increase over the FY 1991 enacted level.

- The HPCC Program is driven by the recognition that unprecedented computational power and capability is needed to investigate and understand a wide range of scientific and engineering "grand challenge" problems. These are fundamental problems whose solution is critical to national needs. Progress toward solution of these problems is essential to fulfilling many of the missions of the participating agencies. Examples of grand challenges addressed include: prediction of weather, climate, and global change; determination of molecular, atomic, and nuclear structure; understanding turbulence, pollution dispersion, and combustion systems; mapping the human genome and understanding the structure of biological macromolecules; improving research and education communications; understanding the nature of new materials; and problems applicable to national security needs.

- The HPCC Program nurtures the educational process at all levels by improving academic research and teaching capabilities. Advanced computing and computer communications technologies will accelerate the research process in all disciplines and enable educators to integrate new knowledge and methodologies directly into course curricula. Students at all levels will be drawn into learning and participating in a wide variety of research experiences in all components of this program.

- The FY 1992 Program and this document were developed by the HPCC Working Group under the direction of the Committee on Physical, Mathematical, and Engineering Sciences of the Federal Coordinating Council for Science, Engineering, and Technology.

Figure 1 **The High Performance Computing and Communications Program**

Goals: Strategic Priorities

Extend U.S. technological leadership in high performance computing and computer communications.

Provide wide dissemination and application of the technologies both to speed the pace of innovation and to serve the national economy, national security, education, and the global environment.

Spur gains in U.S. productivity and industrial competitiveness by making high performance computing and networking technologies an integral part of the design and production process.

Strategy: Integrating Priorities

Support solutions to important scientific and technical challenges through a vigorous R&D effort.

Reduce the uncertainties to industry for R&D and use of this technology through increased cooperation between government, industry, and universities and by the continued use of government and government-funded facilities as a prototype user for early commercial HPCC products.

Support the underlying research, network, and computational infrastructures on which U.S. high performance computing technology is based.

Support the U.S. human resource base to meet the needs of industry, universities, and government.

Program Components

High Performance Computing Systems
- Research for Future Generations of Computing Systems
- System Design Tools
- Advanced Prototype Systems
- Evaluation of Early Systems

Advanced Software Technology and Algorithms
- Software Support for Grand Challenges
- Software Components and Tools
- Computational Techniques
- High Performance Computing Research Centers

National Research and Education Network
- Interagency Interim NREN
- Gigabits Research and Development

Basic Research and Human Resources
- Basic Research
- Research Participation and Training
- Infrastructure
- Education, Training, and Curriculum

1. PROGRAM GOALS AND OVERVIEW

Introduction

High performance computing (HPC) is emerging as a powerful technology for industrial design and manufacturing, scientific research, communications, and information management. A robust U.S. high performance computing and computer communications capability contributes to leadership in critical technology and national security areas. Improved computational and communications technologies contribute to more effective approaches to problem solving, new products and services, and enhanced national competitiveness across broad sectors of the economy.

Recent advances offer the potential for a thousand–fold improvement in useful computing capability and a hundred–fold improvement in available computer communications capability by 1996. Based on several years of planning, under the auspices of the Federal Coordinating Council for Science, Engineering, and Technology (FCCSET), Federal agencies and the technical community have developed the Federal High Performance Computing and Communications (HPCC) Program to realize this potential and to meet the challenges of advancing computing and associated communications technology and practices. Agencies have realigned and enhanced their high performance computing research and development programs, coordinated their activities with other agencies, and shared common resources to develop the program presented in this document.

Needs and Benefits

High performance computing has become a vital enabling force in the conduct of science and engineering research over the past three decades. Computational science and engineering has joined, and in some areas displaced, the traditional methods of theory and experiment. For example, in the design of commercial aircraft, many engineering issues are resolved through computer simulation rather than through costly wind tunnel experiments. This trend has been powered by computing hardware and software, computational methodologies and algorithms, availability and access to high performance computing systems and infrastructure, and the growth of a trained pool of scientists and engineers. This process has been nurtured by Federal investment in advanced research, agency supercomputer centers, and national networks through DARPA, DOE, NASA, NSA, and NSF. These facilities have contributed to national mission areas such as energy, space, health, defense, environment, weather, and basic science and technology that could not be effectively addressed without the use of such advanced facilities.

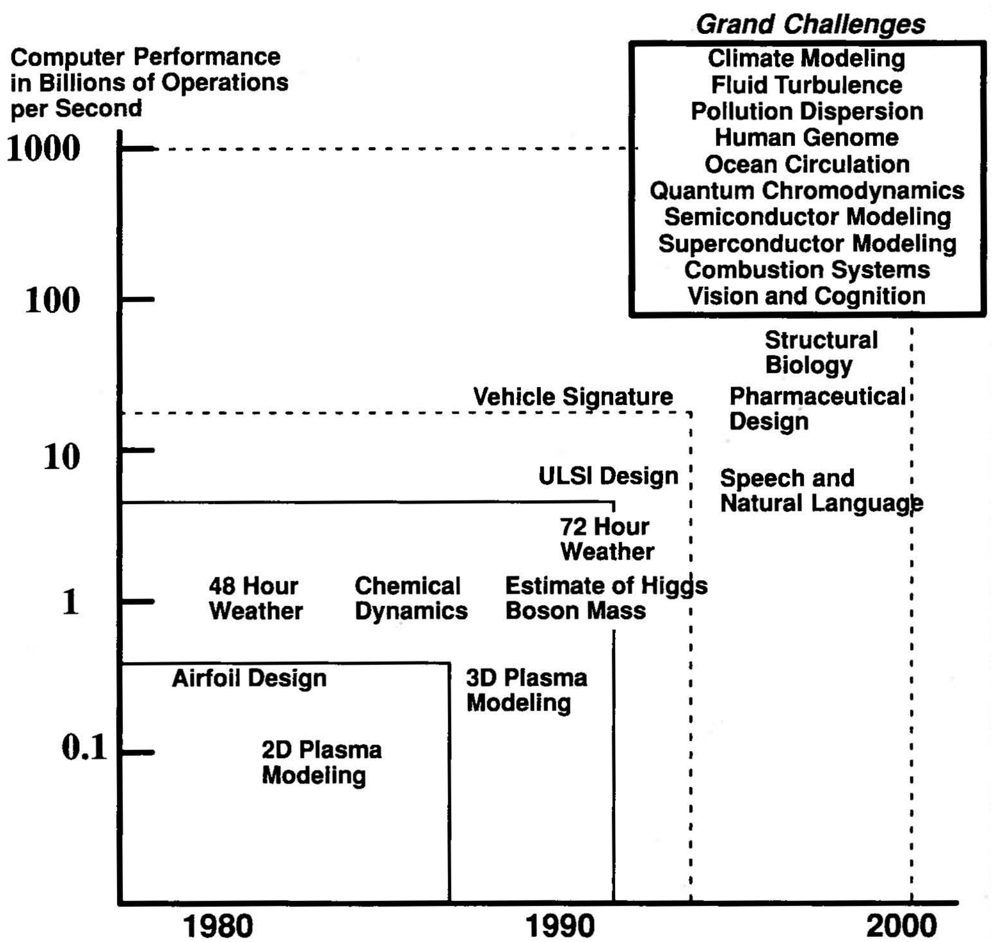
High performance computing technology is knowledge and innovation intensive. Its development and use engages the entire scientific and engineering community. Building upon fundamental research of the early 1980's, a new computing technology of scalable parallel processing computers emerged. By the mid–1990's, this innovative approach to high performance computing systems promises to achieve sustained performance improvements of a thousand–fold compared to current systems.

In a growing number of science and technology fields, progress and productivity in modern research are increasingly dependent on the close interaction of people located in distant places, sharing and accessing computational resources across networks. Although the U.S. is the world leader in most of the critical aspects of computing technology, this lead is being challenged.

The Federal HPCC Program is a strategic Federal investment in the frontiers of computing and computer communications technologies and is formulated to satisfy national needs from a variety of perspectives including: technology, science applications, human resources, and technology transition. Needs are derived from the agency missions and based on the underlying science, engineering, and technology base required to carry out these missions. Many of these mission needs are related to solving very intensive large scale computing problems. These fundamental problems often cut across various agencies and missions and are called grand challenge problems (Figure 2).

The industrial and academic sectors provide major sources of innovation, cost effective development, and support of information technologies and their application to grand challenge problems. As these technologies are developed, the results support the Federal agency missions and become available nationally. The program provides for development of these revolutionary technologies within a framework of a partnership among government, industry, and academe and allows for rapid transition of laboratory results into new products that will then be applied within the program.

Figure 2 **Performance Requirements for Grand Challenge Problems**



Program Description

The Program consists of four integrated components representing the key areas of high performance computing and communications:

High Performance Computing Systems (HPCS)— the development of the underlying technology required for scalable parallel computing systems capable of sustaining trillions of operations per second on large problems.

Advanced Software Technology and Algorithms (ASTA)— the development of generic software technology and algorithms for grand challenge research applications to realize the performance potential of high performance computing systems in a networked environment.

National Research and Education Network (NREN)— the development of a national high speed network to provide distributed computing capability to research and educational institutions and to further advanced research on very high speed networks and applications.

Basic Research and Human Resources (BRHR)— support for individual investigator and multidisciplinary long term research drawn from diverse disciplines, including computer science, computer engineering, and computational science and engineering; initiation of activities to significantly increase the pool of trained personnel; and support for efforts leading to accelerated technology transition.

Advances in high performance computing enable advances in almost every other science and engineering discipline. There is a complex web of research interdependencies among the four components, and each area contributes to progress in other areas. Because of these dependencies, achieving and maintaining balance between the research components is a primary goal and the most important priority in the current context and environment. The HPCC Program is designed to provide balanced support both for technology areas including components, systems, software, and algorithms, and for applications, infrastructure, and human resources to achieve rapid overall research progress and productivity.

The component activities are planned to produce a succession of intermediate benefits on the way to meeting the long range programmatic goals. The HPCC Program builds on Federal programs already in place, providing additional resources in selected areas. Computational science and engineering grand challenges as illustrated in Figure 2 are the focal points for these efforts.

Goals

The goals of the High Performance Computing and Communications Program are to:

Extend U.S. technological leadership in high performance computing and computer communications.

Provide wide dissemination and application of the technologies both to speed the pace of innovation and to serve the national economy, national security, education, and the global environment.

Spur gains in U.S. productivity and industrial competitiveness by making high performance computing and networking technologies an integral part of the design and production process.

These goals will be realized by achieving: computational performance of one trillion operations per second (10^{12} ops, or teraops) on a wide range of important applications; development of associated system software, tools, and improved algorithms for a wide range of problems; a national research network capable of one billion bits per second (10^9 bits, or gigabits); sufficient production of Ph.D.'s and other trained professionals per year in computational science and engineering to enable effective use and application of these new technologies.

Strategy

The goals will be met through coordinated government, industry, and university collaboration to:

Support solutions to important scientific and technical challenges through a vigorous R&D effort.

Reduce the uncertainties to industry for R&D and use of this technology through increased cooperation between government, industry, and universities and by the continued use of government and government–funded facilities as a prototype user for early commercial HPCC products.

Support the underlying research, network, and computational infrastructures on which U.S. high performance computing technology is based.

Support the U.S. human resource base to meet the needs of industry, universities, and government.

At the program component level the strategy will: exploit and extend scalable parallelism and engage in intensive software development in HPCS; use common requirements of the grand challenges to foster HPC software progress in ASTA to strengthen HPC software development and coordination; evolve from the current Internet network to the NREN using a series of testbed systems; and strengthen academic activities in computer science and computational science and engineering as part of BRHR.

Program Execution Strategy

The strategy is based on the strengths of partnerships among the Federal agencies and other organizations. Major portions of the program will be cost–shared and leveraged by the participation of industry and universities. This general approach operates well today and provides strong evidence that the HPCC Program can be successful in the future. The specific elements of the approach are to:

> *Create a balanced, critical–mass program.* The program must achieve sufficient scope and balance among the components. A technology program that created extremely fast processors without comparable memory, input–output, and mass storage systems would not succeed. Neither would a program that created powerful computers without adequate software, network access, and capable people. Similarly, a program that created only high performance networks would not satisfy the increased performance requirements needed for grand challenges. The HPCC Program must operate at a sufficient scale and coverage of technology areas that the new technologies can be effectively applied to grand challenge problems with acceptable levels of risk. The HPCC Program achieves balance by the extensive participation of experienced users, applications developers, and researchers in the HPCC disciplines throughout the design, development, and implementation process.

> *Build on agency strengths.* The strategy builds on agency strengths by giving appropriate agencies the responsibility to coordinate activities for areas of demonstrated capability. It also ensures that the strengths of the other agencies are included by integrating their participation in various task areas. DOE, NASA, NSF, DOC (NOAA, NIST), and DOD (DARPA, NSA) have decades of experience in applying the world's most powerful computers, and thus provide a valuable perspective on high performance computing requirements and applications. NSF has the demonstrated technical and operational expertise needed for deploying high performance national networks in the research community, and is uniquely positioned to support basic research in computer science, computational science, and other scientific areas that can benefit from high performance computing in

interdisciplinary programs. DARPA, having pioneered technologies for high performance computing systems and microelectronic components, has a strong existing research program in teraops computing and gigabit networking, and offers rapid technology transition into commercial products in support of DOD science, technology, and applications base. EPA, NIH, and NOAA provide complementary HPCC requirements perspectives and application bases. NIST has extensive experience in HPC systems and networks instrumentation and evaluation, and provides a means for standards development.

Accelerate Technology Transfer. The transfer of technology from research to development and to application can be a very slow process due to a number of barriers to the use of new technology: high initial cost, inadequate and user–unfriendly software and systems, and lack of standards. The HPCC Program relies upon substantial industry participation to overcome these barriers and yield the benefits derived from moving new technologies to industry. The strategy accelerates technology transition by using a participative development process for each of the task areas, and by stimulating the growth of shared knowledge and capabilities. An example of this is the creation of network–accessible repositories of scalable HPC software and associated user groups to provide usage feedback and improvement of the HPC software base.

Overcome Barriers. To overcome high costs of creating successive generations of high performance computers, the program will emphasize scalable computer designs. Scalable computing and networking technologies enable exploratory use with small, lower cost prototype systems needed to eventually support the acquisition of larger systems. The development of user–friendly software and systems as part of the investment in the HPC software base is a major, integral part of the program. The central role of mission agencies, and the broad academic scientific research community involvement, ensures that the hardware, software, and networking technology developed will be responsive to user needs. A strategy of cooperative government, industry, and university activities is used to manage and coordinate the coupling of these sectors to achieve maximum synergy.

2. HIGH PERFORMANCE COMPUTING AND COMMUNICATIONS PROGRAM COMPONENTS

The HPCC Program is composed of four integrated and coordinated components that are designed to enhance scientific productivity and support long term agency needs. The emerging scientific computing environment is that of advanced workstations with high resolution color displays connected to a high speed computer communications network and high performance computing resources. The regional and national networks provide a means to gain access to additional high performance computing systems and research resources. Realizing the full potential of these teraops computing and gigabits networking systems will require advanced software technology and algorithms and people educated and trained to use these tools and resources in this dynamic field.

High Performance Computing Systems (HPCS)

The HPCS component produces scalable parallel computing systems through the development of prototype systems. The program is designed to attack computational science problems by developing innovative systems that will provide a one–hundred to one–thousand–fold increase of sustained computing capability over machines that follow the more conventional design evolution path (See Figure 3). DARPA will coordinate the research and development effort that will produce teraops systems.

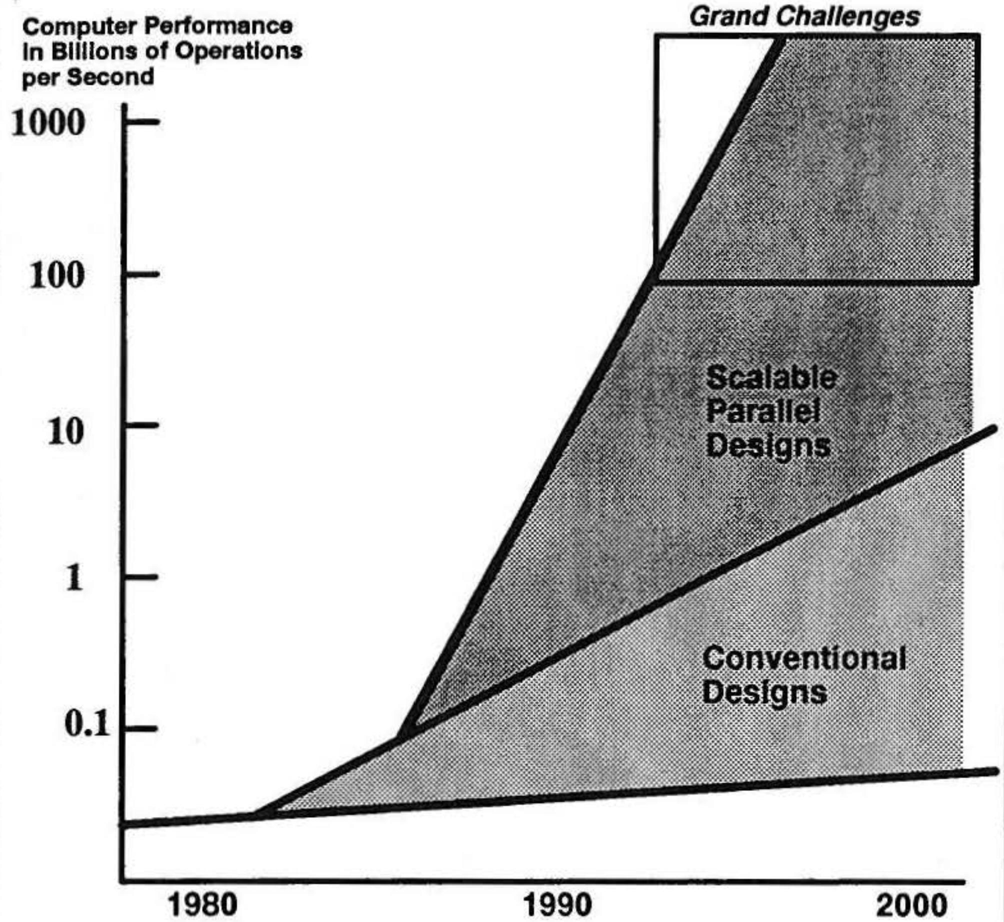
The program is structured to focus on technological challenges in the earliest stages of the product development cycle. Critical underlying technologies are developed in prototype form along with associated design tools. This allows empirical evaluation of alternative solutions as the prototype systems mature. Evaluation is performed throughout the development cycle, with experimental results being fed back into the design process to refine successive generations of systems. There is risk inherent in creating new technologies, and each project will be managed according to its proximity to commercial introduction. Larger projects which are close to yielding commercial products are performed on a cost shared basis with industry.

HPCS is composed of the four subcomponents shown below.

Research for Future Generations of Computing Systems. This activity produces the underlying component, packaging, and scaling concepts. These projects ensure that the required advanced technologies will be available for the new systems while providing a foundation for the more powerful systems that need to follow.

Figure 3 **Computer System Performance Trends for Grand Challenge Problems**

– Power and Time to System Development –



1 Teraops = 1000 Billion Operations per Second

System Design Tools. This activity develops computer aided design tools and frameworks for enabling multiple tools to work together to enable design, analysis, simulation, and testing of systems components. The tools will enable rapid prototyping of new system concepts.

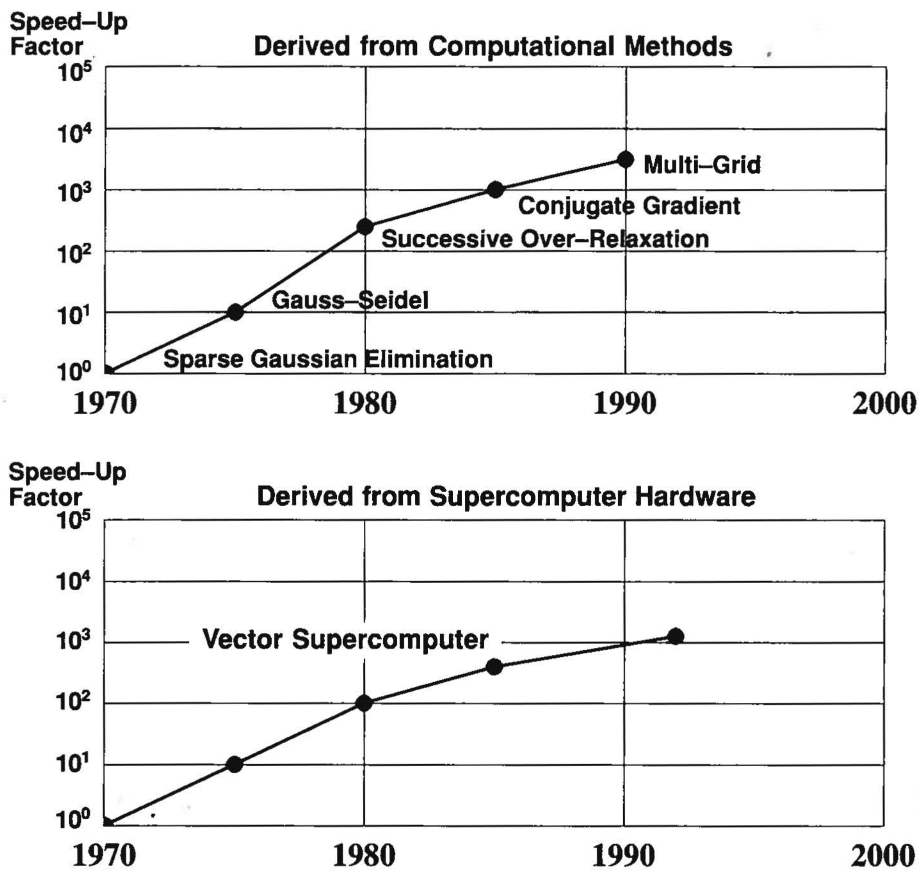
Advanced Prototype Systems. This activity consists of focused development of experimental systems that are designed and developed on a cost shared basis with industry. The 100 gigaops systems provide a basis for the teraops systems. New models of computation will be introduced and successive generations of computer systems, along with systems software, will encompass broader grand challenge domains. Modular technologies will enable a wide variety of system configurations using common components. Systems with 100 gigaops sustained performance will be developed by the early 1990's. The teraops level will be reached by the mid 1990's.

Evaluation of Early Systems. Experimental systems will be placed at sites with high levels of expertise to provide feedback to systems architects and software designers. Performance evaluation criteria for systems and results of evaluation will be made widely available. Because of scalability, early systems can be acquired at smaller scales to evaluate their potential performance. As noted below, the ASTA component will support, on a cost shared basis, the acquisition of large scale systems for experimental use in grand challenge applications.

Advanced Software Technology and Algorithms (ASTA)

Dramatic improvements in algorithm design and software technology are essential to achieving sustained teraflops computing system performance. Improvements in hardware, especially for scalable parallel systems, must be matched with new and innovative algorithms and software to enable researchers to expand the boundaries of computational capabilities. In Figure 4, this point is illustrated for broad classes of scientific computing problems by showing that comparable improvements in throughput have resulted from advances in computational methods as from improved hardware technology. In this case, hardware improvements yielded a speed–up of 1000 over 20 years, while software and algorithms improvements yielded a speed–up of 3000 over the same period, for an aggregate speed–up of more than a million.

Figure 4 **Performance Improvement for Scientific Computing Problems**



The ASTA component of the Federal High Performance Computing and Communications Program has three goals:

Enable solution of grand challenge application problems in science and engineering.

Improve system user–friendliness, reliability, and software productivity.

Use experience gained on leading edge applications to help guide future software efforts.

The emphasis is on the development of advanced algorithms and software technology required to address applications problems on the scale of grand challenges. The ASTA component is comprised of four subcomponents, as shown below.

Software Support for Grand Challenges. The HPCC Program is designed to demonstrate new computing technology capabilities by confronting an expanding number of grand challenge application problems. The goal is to reduce the risks to researchers inherent in adopting innovative high performance computing technologies. Grand challenge application problems will be selected based upon their scientific importance, the potential for cost sharing with sources directly concerned with the specific scientific and engineering applications, and the potential for leveraging across sectors.

Software Components and Tools. The multidisciplinary research teams will have common needs in software technology and programming environments, advanced compiler technology, optimization and parallelization tools, interoperability and data management, visualization, debugging and analysis, and performance measurement. Advances in these generic software areas will minimize the need for specialized researchers to master advanced, complex computing science skills. Coordination of the development of advanced software technology and algorithms among the participating agencies will be important to ensure effective and efficient use of resources. A particular focus of this element will be to develop advanced software applications that exploit the NREN using distributed file systems and national software libraries.

Computational Techniques. The focus of the HPCC Program on scalable parallel computing systems dictates that significant advances in computational techniques will be needed. The design and theory of algorithms are as important as hardware or networking improvements in reaching teraflops computational performance. Research in computational techniques will include parallel algorithms, numerical and mathematical analyses, parallel languages, computational models, and program refinement techniques. Higher level languages will be developed to enable computational scientists to work directly at the level of the abstract computational problem being addressed and to more easily explore specific implementation approaches.

High Performance Computing Research Centers. High performance computing research centers and testbed facilities will provide a large number of researchers access via the NREN to both conventional and innovative supercomputing architectures. Many of these centers will introduce innovative architectures and will be the focus of technology transfer activities and form the base for training new generations of computer and computational scientists.

National Research and Education Network (NREN)

The NREN component of the HPCC Program dramatically expands and enhances the U.S. portion of an existing worldwide infrastructure of interconnected computer networks called the Internet. A substantial fraction of the domestic Internet is supported and loosely coordinated by Federal agencies, principally DARPA, DOE, NASA, and NSF.

Collaboration among scientists is an important and integral facet of the U.S. research environment. It can be greatly improved by increasing the level of network connectivity and by introducing new capabilities into the existing infrastructure. The NREN design will not only address broad network connectivity, but will also provide the basis for necessary higher level capabilities and services.

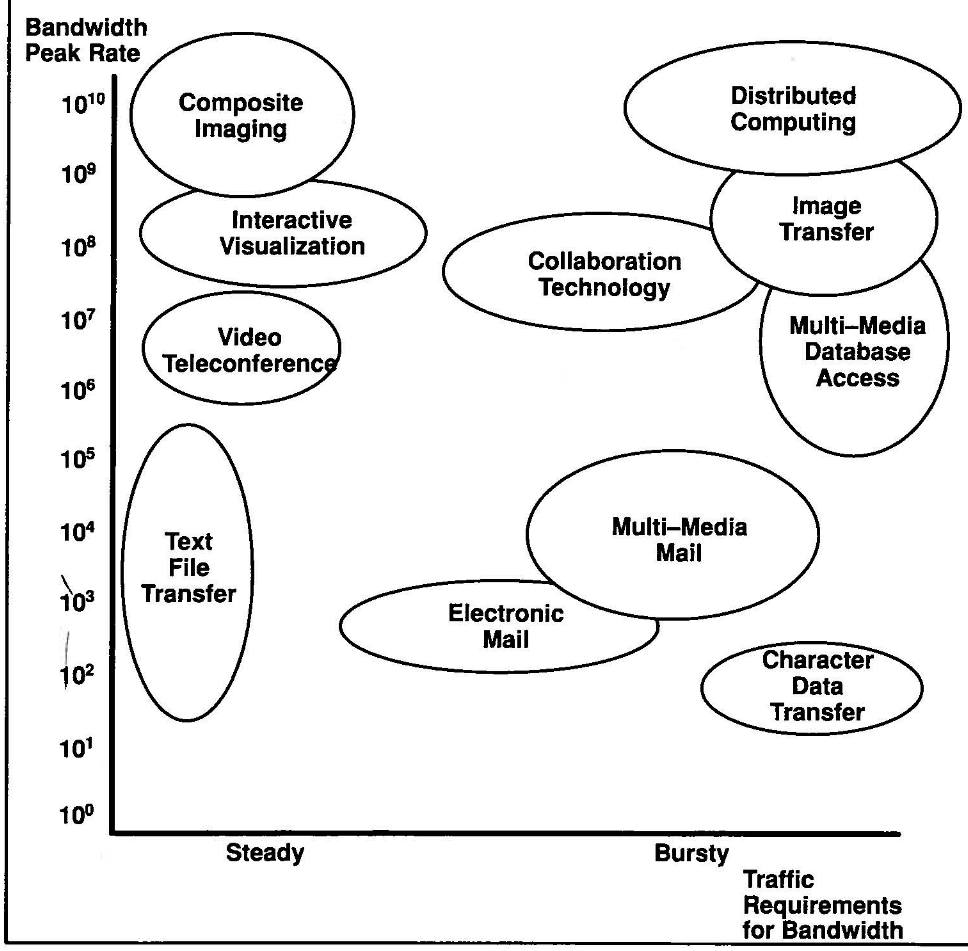
Many educational institutions, government laboratories, and industrial research facilities are currently connected to the Internet. Yet, it still falls short of a widespread, uniform, and high performance national infrastructure. In order to satisfy the HPCC Program goals, the NREN must not merely provide network access to research and educational institutions at all levels and locations, it must also deliver new capabilities. Some of these, such as distance learning, may initially be extensions of current technology. All capabilities will benefit from, and many will be enabled by, a program of research into very high speed technology. This technology is needed to support access to digital libraries, large scale distributed computing resources, as well as to perform computationally intensive applications that require real time visualization of modeling and simulation results, rapid interrogation and retrieval of scientific data from specialized data bases, remote control of experiments and simulations, and teleconferencing.

In addition to serving the needs of the scientific and research communities, the NREN will provide valuable experience necessary for the successful development of a broader, privately–operated national information infrastructure. Such an infrastructure would allow consumers, businesses, and schools and government at all levels to share quality information and entertainment when and where they want it at a reasonable cost.

Applications conducted over a computer network vary in their flow of information from steady, as in a bulk file transfer, to "bursty," as in human–computer interaction via keyboard. Similarly, some applications can be carried out at relatively low communication rates, while others by their nature require high speeds. A number of scientific networking applications are characterized on the graph of speed versus burstiness in Figure 5. Traffic seen in the early days of networks appears near the bottom of the chart. More advanced applications are furthest from the origin and require more sophisticated protocols and network capabilities. This chart illustrates that a gigabit network is needed not only to carry the aggregation of low speed traffic, but also to accommodate high speed uses.

Figure 5

NREN Applications
by Bandwidth and Traffic Characteristics



The vision of the NREN is of an interconnection of the nation's educational infrastructure to its knowledge and information centers. In this system, elementary schools, high schools, two and four year colleges, and universities will be linked with research centers and laboratories so that all may share access to: libraries, databases, and diverse scientific instruments such as supercomputers, telescopes, and particle accelerators. The NREN enables communication and fosters collaboration among and within these communities. By reducing the traditional impediments of geographical isolation, the NREN improves the quality and raises the level of education nationwide. The NREN contributes to the success of the Basic Research and Human Resources component of the High Performance Computing and Communications Program. By making unique scientific and informational resources accessible beyond their physical locations, it permits widespread participation in the HPCC Program by scientists, university researchers, and students, and it enables the development of large scale distributed computing resources.

Interagency Interim NREN. The NSF will coordinate the Interim NREN activities by upgrading its backbone network, by assisting regional networks to upgrade facilities, capacity, and bandwidth, and by interconnecting the backbone networks of other agencies. A significant effort in its implementation will be the development and deployment of safeguards to enhance security and control over access to the connected computer systems, and that of the network components themselves in order to minimize vulnerability to careless or malicious attack. Coordination among participating agencies and the non Federal networking community will be expanded through the creation of a National Networking Council.

Gigabit Research and Development. DARPA will coordinate the research and development effort that will culminate in initial deployment of gigabits per second capability. Coordination of research efforts on very high speed switches, protocols, and computer interfaces will be necessary.

Basic Research and Human Resources (BRHR)

This component addresses long term national needs for more skilled personnel, enhancement of education and training, and materials and curriculum development in the high performance computing science and engineering areas. The NREN and ASTA components include support for research in the large scale project environment. The BRHR component is designed to encourage investigator initiated, long term research on experimental projects that will maintain the flow of innovative ideas and talented people into high performance computing areas.

Drawing the best and brightest of our Nation's youth into scientific and technological careers is a formidable challenge that will have profound effects on overall U.S. scientific and technological competitiveness. This component of the program will establish industry, university, and government partnerships to improve the training and utilization of personnel and to expand the base of research and development personnel in high performance computing science, technology, and applications.

The BRHR component of the HPCC Program has five goals:

> Improve the flow of human resources into high performance computing.
>
> Improve the university infrastructure to stay at the leading edge.
>
> Expand collaboration and resource sharing among the Federal, academic, and industrial sectors.
>
> Facilitate multidisciplinary research on high performance computing and communications.
>
> Create a critical mass of users by building a base with common systems, tools, and interfaces.

The BRHR component is organized across the four sub–components shown below.

Basic Research. These activities support increased participation of individual investigators in the conduct of innovative multidisciplinary research in computer science, computer engineering, and computational science and engineering related to high performance computing. The strategy is to increase the number of multi–disciplinary awards across all disciplines where computational methods are critical to achieving advances or scientific breakthroughs. Program activities include: research on scientific algorithm development for highly parallel computers;

generic computational algorithm development; scaling techniques; prediction techniques for concurrent systems; resource management strategies for highly parallel distributed systems; fault tolerant strategies for parallel and distributed systems; and research on heterogeneous software environments.

Research Participation and Training. These activities address the human resources pipeline in the computer and computational sciences, at post–doctoral (training and re–training), graduate, and undergraduate levels. Program activities include: workshops and seminars; post–doctoral fellowships in computational science and engineering; career training in medical informatics through grants to young investigators; institutional training and post–doctoral programs; knowledge transfer exchange programs at national laboratories, centers, universities, and industry; and software dissemination through national databases and libraries.

Infrastructure. These activities will improve university and government facilities for computer science, computer engineering, and computational science and engineering research related to high performance computing. Program activities include: improvement of equipment in computer science, computer engineering, and computational science and engineering academic departments, centers and institutions; development of scientific databases; and distribution of integrated system building kits and toolsets.

Education, Training and Curriculum. These activities will expand and initiate activities to improve undergraduate and pre–college education and training opportunities in high performance computing and computational science and engineering for both students and educators. The introduction of associated curriculum and training materials at all levels is an integral part of this effort. Program activities include: bringing people to national centers for training, technology transfer, and educational experiences; using professional scientists and engineers to provide curriculum development materials and instruction for high school students in the context of high school supercomputer programs, supercomputer user workshops, summer institutes, and career development informatics for health sciences.

3. PROGRAM DEVELOPMENT AND AGENCY BUDGETS

Program Planning

Leadership for the HPCC Program is provided by the Office of Science and Technology Policy (OSTP), through the Federal Coordinating Council for Science, Engineering, and Technology (FCCSET) Committee on Physical, Mathematical, and Engineering Sciences (PMES). The membership of the PMES includes senior executives of many Federal agencies.

The planning process for the HPCC Program was coordinated by a PMES working group through information exchange, the common development of interagency initiatives, and the review of individual agency HPCC proposals and budgets.

Evaluation Criteria

Each participating agency HPCC contribution was reviewed against formal evaluation criteria during the planning and budget process. A review of participating agencies was performed using these evaluation criteria to develop the FY 1992 agency requests for the Program. The evaluation criteria are shown in Figure 6.

Agency Budgets

Over the last three years of the planning process of this initiative, the participating agencies have mutually adjusted their activities within the base to achieve greater efficiency in addressing the goals of the HPCC Program. For FY 1992, $638 million is being proposed, a $149 million or 30 percent increase over the FY 1991 enacted level. The budget is shown in Figure 7.

The funds proposed from Federal sources are not intended to carry out the entire HPCC Program. Portions of this program will be cost shared by organizations from the participating sectors. The funding estimates are based on analyses of the practical experience of prior computing and computer networking programs. These estimates were then reviewed as outlined above. Cost sharing will occur with various U.S. industrial and university partners to a large extent in the HPCS and the NREN components. Cost sharing will occur in the ASTA component with agency programs and other computational applications programs, for example, in specific grand challenge areas, via multidisciplinary collaborations. This component also includes deployment of high performance systems to HPCC research centers. The close coupling of support in the Program will result in significant leverage, accelerate technology transfer, stimulate U.S. industry and markets, and enable the solution of computationally intensive applications. In addition, although HPCC is not intended to include classified programs, the technology produced will have an important impact in these national security areas. Figure 7 illustrates the relative levels of investment in the four program components.

Figure 6

Evaluation Criteria for the Federal High Performance Computing and Communications Program

Relevance/Contribution. The research must significantly contribute to the overall goals and strategy of the Federal High Performance Computing and Communications (HPCC) Program, including computing, software, networking, and basic research, to enable solution of the grand challenges.

Technical/Scientific Merit. The proposed agency program must be technically/scientifically sound and of high quality, and must be the product of a documented technical/scientific planning and review process.

Readiness. A clear agency planning process must be evident, and the organization must have demonstrated capability to carry out the program.

Timeliness. The proposed work must be technically/scientifically timely for one or more of the HPCC program components.

Linkages. The responsible organization must have established policies, programs, and activities promoting effective technical and scientific connections among government, industry, and academic sectors.

Costs. The identified resources must be adequate, represent an appropriate share of the total available HPCC resources (e.g., a balance among program components), promote prospects for joint funding, and address long–term resource implications.

Enhancements to Existing Program Research. Existing agency HPCC programs will receive adequate support before new initiatives are funded.

Agency Approval. The proposed program or activity must have policy–level approval by the submitting agency.

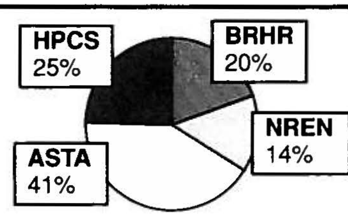
Figure 7

High Performance Computing and Communications Budgets by Agency and Program Component

(Dollars in Millions)

Agency	Base FY 1991	HPCC FY 1992	FY 1992 HPCC Component			
			HPCS	ASTA	NREN	BRHR
DARPA	**183.0**	**232.2**	103.3	38.5	32.9	57.5
DOE	**65.0**	**93.0**	15.0	58.0	12.0	8.0
NASA	**54.0**	**72.4**	14.2	49.8	7.4	1.0
NSF	**169.0**	**213.0**	24.0	103.0	32.7	53.3
DOC/NIST	**2.1**	**2.9**	0.3	0.6	2.0	0.0
DOC/NOAA	**1.4**	**2.5**	0.0	1.8	0.7	0.0
EPA	**1.4**	**5.2**	0.0	4.5	0.0	0.7
NIH/NLM	**13.5**	**17.1**	0.0	8.9	4.2	4.0
Total	**489.4**	**638.3**	**156.8**	**265.1**	**91.9**	**124.5**

FY 1992 Component Funding Comparison



HPCC — High Performance Computing and Communications

Components of HPCC:

HPCS — High Performance Computing Systems
ASTA — Advanced Software Technology and Algorithms
NREN — National Research and Education Network
BRHR — Basic Research and Human Resources

Agency Program Descriptions

The agency responsibilities under the Federal HPCC Program are outlined in Figure 8. Several agencies have been assigned a coordinating responsibility in specific technical areas:

- DOE and NASA will coordinate activities in HPC system evaluation, testbed development, and applications software capabilities.
- NASA will coordinate the accumulation of and access to the HPC software base. This will be facilitated by a wide area file system technology that is currently being deployed for early experimental use by DARPA which will be extended to include the NREN as it matures and is deployed by NSF.
- DARPA will coordinate activities in the development of scalable parallel HPC systems, including their basic units of replication, system modules, and the necessary associated systems software. DARPA will also coordinate activities in gigabit network technology research.
- NSF will coordinate activities for the broad deployment of the National Research and Education Network working with all agencies with mission specific requirements.
- NSF will coordinate activities in basic research and human resource development, while each agency retains its role as required to accomplish their own missions.
- NIST will coordinate activities in HPC system instrumentation, evaluation, and in standards issues.

Each agency participates in all of the identified activities to ensure that the resulting capabilities are a good match to user needs. DOC/NOAA, HHS/NIH/NLM, and EPA bring distinctive applications areas of broad interest and network user bases.

Figure 8

HPCC Program: Agency Responsibilities: HPCS, ASTA		
Activity / Agency	**High Performance Computing Systems**	**Advanced Software Technology & Algorithms**
DARPA	· Technology development and coordination for Teraops systems	· Technology development for parallel algorithms and systems software
DOE	· Technology development · Systems evaluation	· Energy applications centers · Energy computation research · Software tools
NASA	· Aeronautics and space application testbeds · Systems development for space flight	· Software coordination · Research in: · Aerospace computations · Information management
NSF	· Basic architecture research	· Basic research in: · Software tools, databases · Grand challenges · Computer access
DOC/ NIST		
DOC/ NOAA		· Ocean and atmospheric computation research · Software tools
EPA		· Research in environmental computations, databases, and application testbeds
NIH/ NLM		· Medical application testbeds · Software tools · Medical computation research

HPCC Program: Agency Responsibilities: NREN, BRHR		
National Research and Education Network	**Basic Research and Human Resources**	**Activity / Agency**
· Technology development and coordination for gigabits networks	· University programs	**DARPA**
· Energy deployment mission facilities · Gigabits applications research	· Basic research and education programs	**DOE**
· Aerospace deployment mission facilities	· Research initiation and university block grants	**NASA**
· Facilities coordination and deployment · Gigabits research	· Programs in: · Basic research · Education · Training / curricula	**NSF**
· High speed network research and standards		**DOC/ NIST**
· Ocean and atmospheric mission facilities		**DOC/ NOAA**
· States environmental mission assimilation	· Technology transfer to States · University programs	**EPA**
· Medical mission facilities	· University programs	**NIH/ NLM**

Defense Advanced Research Projects Agency (DARPA)

DARPA, as the DOD lead agency for advanced technology research, will focus on developing the high performance computing and networking technologies needed for the Defense and overall HPCC Programs. DARPA programs have produced both the computing and networking foundation for the HPCC Program, including the first generation of scalable parallel computing systems and large scale computer networks, and the associated system software and supporting technologies. DARPA has worked with industry to pioneer the application of these new technologies within Defense and on a cooperative basis with other Federal agencies.

The DARPA HPCC Program builds upon the DARPA Strategic Computing Program. As the HPCC Program builds up, the Strategic Computing Program integrates its results with Defense specific needs such as embedded systems, accelerators of specific problem domains, and grand challenges problems related to defense. DARPA will continue this mode of executing the HPCC Program, cooperating with various defense organizations, and working closely with AFOSR, ARO, ONR, Defense Service Laboratories, NSA, other Defense organizations, and other Federal agencies as appropriate.

High Performance Computing Systems will be produced in the four main subareas: Research for Future Generations of Computing Systems, System Design Tools, Advanced Prototype Systems, and Evaluation of Early Systems. Systems capable of sustaining 100 gigaops for large problems will available for deployment by late 1993 and the teraops systems will be available by 1996.

In *Advanced Software Technology and Algorithms*, DARPA projects will produce scalable libraries for Defense problem domains and programming and analysis tools for scalable parallel and distributed heterogeneous systems in a workstation/server configuration that will be open to the integration of embedded systems and accelerators.

For the *National Research and Education Network* component, high performance networking technologies will be produced to satisfy the gigabit technology needs of the NREN and to provide a dual use technology base for Defense. This networking technology includes development of new protocols and switch and transmission technologies, and it will be capable of supporting a wide range of advanced network services.

The *Basic Research and Human Resources* component will focus on fundamental scientific issues in these three technology development areas in cooperation with other basic research programs and provide for smaller individual investigator projects to complement the larger projects. In addition, the relevant and related basic research will also be integrated into the larger projects as it matures.

Department of Energy (DOE)

The Department of Energy will participate in all components of the High Performance Computing and Communications Program. In the area of *High Performance Computing Systems,* the DOE will be an early customer of small versions of systems with advanced architectures and will evaluate these systems on energy related applications. The DOE will consider cooperative development of advanced systems between its national laboratories and vendors, especially integration of very high speed computer and networking hardware with software systems. The DOE will support research and development on algorithms and systems software for parallel computing systems.

The *Advanced Software Technology and Algorithms* effort will include research and development of: parallel algorithms for grand challenge applications, software and tools for early prototypes of 100 gigaflops and teraflops systems, prototype computational science programming environments that meet standards and are transportable, and support for high performance computing research centers to facilitate the transition from research on parallel machines into the applications and the programming environments. The DOE will fund several grand challenge collaborations, initiate a software component and tools program with strong industrial participation, and initiate an applications driven computational research program. The DOE will evaluate proposals and make research awards related to grand challenges in global climate change, molecular biology, human genome research, materials and chemical sciences, combustion research, waste remediation, fusion energy, and other areas within its mission.

The DOE will participate in the cooperative interagency *National Research and Education Network.* The Energy Sciences Net (ESNet) will be incorporated into the NREN and will provide quality network access to the energy research facilities. ESNet will maintain compatibility and will be upgraded in concert with NREN. Gigabit network support technology will be developed for DOE applications distributed across multiple energy research centers at the national laboratories and universities.

The DOE's *Basic Research and Human Resources* activities will include: stimulating research in computational science, expanding training programs at the national laboratories for high school teachers and college students in computing techniques, initiation of a high school supercomputer access program, and provision of fellowships in computational science with internship at national laboratories.

National Aeronautics and Space Administration (NASA)

NASA's HPCC Program participates in all four components of the Federal Program, through a vertically integrated program focused on NASA's grand challenges in: Computational Aerosciences, Earth and Space Sciences, and Remote Exploration and Experimentation.

The goal of NASA's program is to accelerate the development and application of high performance computing technologies to meet NASA's science and engineering requirements. In cooperation with the other Federal agencies, NASA's program will deploy teraflops computer capabilities essential for computational design of integrated aerospace vehicle systems and for predicting long term global change, and will enable the development of massively parallel techniques for spaceborne applications.

NASA's program is focused on bringing together interdisciplinary teams of computer scientists and computational physicists to develop these technologies within three vertically integrated projects that are unique to NASA's missions. These technologies include applications algorithms and programs, systems software, peripherals, networking, and the actual high performance computing hardware. NASA will develop a suite of software tools to enhance productivity. These include: load balancing tools, run time optimizers, monitors, parallelization tools, as well as data management and visualization tools.

NASA's role includes coordinating the *Advanced Software Technology and Algorithm* component for the Federal program; acquiring experimental hardware for testbeds in computational aerosciences, earth and space systems sciences, and remote exploration and experimentation; and supporting the development of the National Research and Education Network. To encourage vigorous research into the underlying theory and concepts of high performance computing, NASA will foster interactions among academia, industry, and national laboratories and will strengthen the basic research in high performance computing at the NASA centers and research institutes and in universities.

NASA's considerable expertise in experimental parallel computer testbeds and small, scalable testbed systems will be used to demonstrate high performance computing technologies as a step toward full–scale computational capabilities. A key to successful exploitation of massively parallel computing power will be the blending of application–driven and architecture–driven computer systems to most effectively meet NASA's needs.

National Science Foundation (NSF)

The NSF HPCC Program impacts the activities of all science and engineering disciplines by providing computing and networking infrastructure support and by developing enabling technologies for advanced computing and communications platforms and paradigms.

In the area of *High Performance Computing Systems,* research will be initiated on new architectures and systems optimized for specific research applications. New tools for systems level automated design and component packaging will be supported to permit the design of application–specific devices and systems.

In *Advanced Software Technology and Algorithms,* research will be initiated on scientific database technology and implementation of prototype networked databases and associated software. Advanced applications tools for research computing environments will be supported to enable nationwide access to the full complement of parallel machines for research on grand challenge problems. Areas of research focus will include numeric and symbolic computing, algorithm development, optimization of applications software for new parallel computers, scientific visualization, automated programming tools, and new methods of scientific and technical information exchanges.

NSF coordination of the *National Research and Education Network* activities will accelerate the harmonizing of multiple agency networks and protocols into a single NREN. The number of nodes will be increased to expand distributed information resources, and to increase redundancy, capacity, control, and security. Mid–level nets will be assisted to upgrade facilities and service very high bandwidth requirements. Support will be provided for research on new protocols for gigabit networks, switch and transmission technology, routing, congestion, and flow control. The exploration of pricing mechanisms for network services and network applications and structured transition to commercial service will be initiated.

In *Basic Research and Human Resources,* support will include multidisciplinary research and university infrastructure for computer science, computer engineering, and computational science and engineering. To increase the human resource pool in computing hardware and software systems areas, support will be expanded for graduate and post–doctoral positions. Curriculum improvement, teacher training, and support for centers to provide education and training in the use of experimental and parallel supercomputers are integral parts of this component.

National Institute for Standards and Technology (DOC/NIST)

The NIST research program is directed toward developing performance monitoring tools and promoting "open systems" software. NIST has proposed to augment its current HPC research program by promoting the commercialization of protocol and security mechanisms for medium speed networks.

The goal of NIST's activities in high performance computing is to develop hardware performance monitoring tools, promote "open system" software, and support a classification system for indexing and distributing scientific software so that industry and the research community can effectively exploit the power of future generations of high performance computers. In support of *The National Research and Education Network*, NIST will develop and speed commercialization of network protocols and security mechanisms that can achieve the desired gigabit speeds on future versions of the network. NIST will participate in the HPCC Program by:

> Developing hardware monitoring methods leading to load characterization and performance measurement techniques for ultra–high–speed systems which will be made available on a publicly accessible database.
>
> Conducting research on new protocols and related security primitives necessary to sustain gigabit network speeds and standards to provide interoperability, common user interfaces to systems, and enhanced security for the network.
>
> Establishing a network testbed at NIST instrumented to collect performance data and test new network protocols, management routines, and security mechanisms for gigabit networks.

National Oceanic and Atmospheric Administration (DOC/NOAA)

The National Oceanic and Atmospheric Administration operational and research programs are directed toward weather prediction, ocean sciences, the Climate and Global Change Program, and the Coastal Oceans Program, together with data management activities for all agency programs. The HPCC Program will allow extensive development of new forecast models, studies in computational fluid dynamics, and the incorporation of evolving computer architectures and networks into the systems that carry out agency missions.

The NOAA High Performance Computing Program is focused in two components:

Advanced Software Technology and Algorithms. This component provides support for: grand challenges in atmospheric and oceanic sciences; development of advanced numerical models to simulate the general circulation of the oceans and atmosphere in support of NOAA missions and the activities of collaborating research groups; development of new computational methods for solving atmospheric, oceanic, and related problems on new computer architectures; data management R&D; support for basic research in strategies, techniques, and tools required for the management and analysis of large–scale distributed scientific databases and distributed data handling, including quality control; and development of algorithms for massively parallel processors, together with their standardized software component libraries and tools for the solution of oceanic and atmospheric analysis and forecasting problems.

NOAA will acquire, install, and operate advanced computational facilities for the evaluation of near operational prototype computers having massively parallel architectures and will develop algorithms appropriate for these architectures.

The National Research and Education Network component provides networking in support for NOAA's climate and global change research community and a wide range of agency missions in oceanography, weather prediction, and environmental sciences research. NOAA will evaluate advanced network protocols and hardware technologies related to its missions.

Environmental Protection Agency (EPA)

The EPA research program is directed toward the advancement and dissemination of computational techniques and software tools which form the core of ecosystem, atmospheric chemistry and dynamics models. The models extend the computational capability of environmental assessment tools to handle multi–pollutant interactions and optimization of control strategies.

The *Advanced Software Technology and Algorithms* component will develop new approaches for coupling numerical equation solvers to the high performance features of emerging computer architectures including specific techniques for solution of partial differential equations on massively parallel architectures. These solvers will be designed to operate on a flexible, generic grid system decoupled from specific applications to provide a testbed for evaluation and optimization. The resulting solution framework can serve as the computational engine for a variety of interdisciplinary applications.

Advancement of scientific visualization techniques can provide exceptional benefits in increasing the pace at which environmental problems are solved. EPA's program includes basic research to improve user interfaces and computational algorithms related to image rendering and manipulation of geometric images. The main focus is on simultaneous representations of multiple data sets and real–time visualization in a heterogeneous distributed processing environment. Related research on storage and access techniques for massive environmental data bases across multiple architectures will be conducted.

The *Basic Research and Human Resources* component includes the establishment of technology transfer centers to propagate the use of HPCC technology to State and local environmental groups, and Federal managers for optimization of pollution control and prevention strategies. A main goal of a Technology Transfer Center is to provide non–sophisticated users with training and guidance in the application of a variety high performance computing tools to solve important environmental challenges. The program also supports cross–discipline career training through universities and other institutions to ensure a continuing base of technical professionals knowledgeable in the use of high performance computing technology for environmental problem solving.

National Institutes of Health / National Library of Medicine (HHS/NIH/NLM)

The HHS/NIH program includes molecular biology computing, creation, and transmission of digital electronic images, the linking of academic health centers via computer networks, the creation of advanced methods to retrieve information from life sciences databases, and training in biomedical computer sciences. The HPCC program will complement the Human Genome Project by providing new methods for computer based analysis of normal and disease genes.

The *Advanced Software Technology and Algorithms* component will develop advanced software technology and algorithms in two areas of importance in biomedical research and education: biotechnology computing and digital images. In the area of biotechnology computing, the program will support development of molecular sequence comparison algorithms, new database methods, and algorithms to predict biological structure and function from genetic code. The biomedical images area will support new methods for representing, linking, and rendering images of biological structure

The *National Research and Education Network* component has two sub–components: connections among academic medical centers and their growing array of computerized information sources; and development of intelligent gateways that link conceptually related databanks in the life sciences. The academic medical centers of the country are confronted with a growing array of disconnected computer–based information sources, ranging from patient records, x–rays and laboratory systems, to basic research tools such as protein and DNA sequence databanks. Development of advanced software systems capable of representing and linking these dissimilar data types, and communicating them among centers for research and health care, is the goal of this part of the program. The focus of the Gateways HPCC program is the building of systems that are capable of translating a user 's request into multiple computer–based vocabularies, selecting appropriate databases from a wide range of widely–available resources, and retrieving information in a manner that does not require users to understand the structure and syntax of the systems being queried.

The *Basic Research and Human Resources* component addresses the need to train biomedical researchers and health care providers in the use of advanced computing and network communications to aid in their work. The successful predoctoral and postdoctoral grants program for career training in medical informatics administered by the National Institutes of Health will be expanded.

4. GRAND CHALLENGE AND SUPPORTING TECHNOLOGY CASE STUDIES

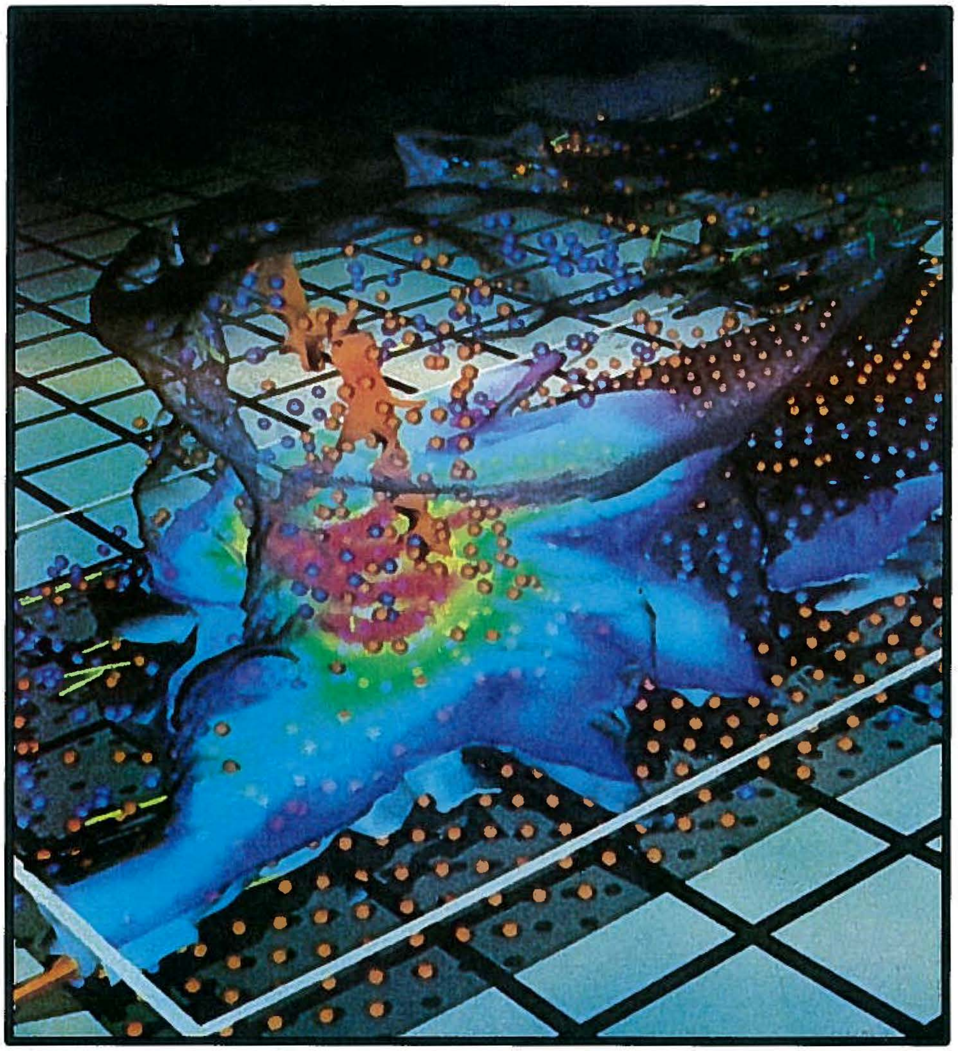
This section incorporates examples of high performance computing and computer communications technologies and several illustrative grand challenge applications in computational science and engineering, that are presented on the cover of this report. The examples were chosen to illustrate the diversity and significance of application areas that have been addressed to date. The list of grand challenges is too long to allow an example from each area, thus lack of representation of certain areas does not imply lack of importance. A more complete description of the grand challenges can be found in the Federal High Performance Computing Program report issued in September 1989.

Forecasting Severe Weather Events

Making accurate predictions about the behavior of the atmosphere is of critical importance to reducing losses due to storms, hurricanes, floods, and other weather related phenomena. Current operational small scale weather prediction models have been constrained by two broad deficiencies: inadequate size of the data sets necessary to define the detailed atmospheric structures and insufficient computer resources to support high resolution. During the next five years, major progress will be made on the first deficiency by using a combination of new ground–based and remote–observing systems. This leaves inadequate computer power as the primary stumbling block preventing high resolution models from playing operational roles in our national weather prediction efforts.

The high–resolution, operational weather prediction models of the future will represent a new generation of numerical formulations. The primary differences will be in the treatment of vertical motions and small scale processes which in the past have been considered to be "sub–grid scale." This means that physically significant events, such as convection, must be treated as parameterized processes and fine scale details in, for example, thunderstorm evolution that can affect the surroundings cannot be directly addressed by the model. Likewise, fine scale observations of local importance, such as moisture gradients around wet areas, which can provide local forcing, currently cannot be incorporated adequately.

To meet this scientific and national need, new technologies need to be developed and incorporated into existing facilities so that a less than 5 km resolution model can be operational before the end of the century. This would allow updated forecasts on a 6–hourly or shorter basis. Such models have already shown significant advances in the accuracy of predicting a wide variety of weather events, from severe thunderstorms to lake effect snowfalls in research applications. Each reduction of the model resolution by one–half requires an increase of computer power of almost an order of magnitude, as well as comparable increases in supporting memory, mass storage and networks. For example, a 5 km model could require up to a 20 teraflops computer system to meet operational schedules using this model.



Researchers have used supercomputers to model thunderstorms numerically. Wind, temperature, pressure, and other variables are calculated every few seconds at several hundred thousand locations in the area of the developing storm. Mathematical equations are then used to simulate the storm's evolution. In this graphic, particles which are released near the ground at regular intervals are colored orange when rising and blue when sinking. Yellow signifies paths of individual particles.

Cancer Genes

DNA is the blueprint of life, the molecular thread in the nucleus of each of our cells which guides the assembly of molecules and complete living organisms. When the DNA code is altered by mutation, serious diseases can result, such as cancer; this phenomenon was known to scientists studying animal tumors in the 1970's. They isolated cancer–causing genes, called "oncogenes" from animal tumors, and later found that similar genes existed in normal human DNA. This was a profound mystery. Why would we be carrying the seeds of our own destruction in our genetic blueprint?

In 1984, two separate research groups used a computerized searching algorithm to compare a newly discovered oncogene to all known genes. To their astonishment, the cancer causing gene matched a normal gene involved in growth and development. Suddenly, it became clear that cancer might be caused by a normal growth gene being switched on at the wrong time. This fundamental and unexpected insight was an early example of a field that is now known as Computational Biology, the science of using computers to store and analyze data from complex molecules in living cells.

The databases used in 1984 for these comparisons contained information about several thousand molecular units; now they contain over 30 million. Moreover, the current multi–agency genome research programs of NIH, NSF, and DOE will acquire data on tens of billions of molecular units, ranging from simple organisms to human beings. The best computer algorithms for determining the similarity of molecules require time proportional to the length of the molecules being compared; if the methods used to analyze oncogenes in 1984 were applied to the three billion base pairs of the human genome, they would require hundreds of years of computer time on today's fastest supercomputers. New computer designs and software methods will be essential to cope with the explosive growth of molecular data. Functioning as intellectual amplifiers to detect similarities and differences in molecules whose size and complexity are too vast for the unaided human mind, the computer systems to be developed by the HPCC program will be a critical tool for the life sciences in the 1990's, and the health care systems of the 21st Century.

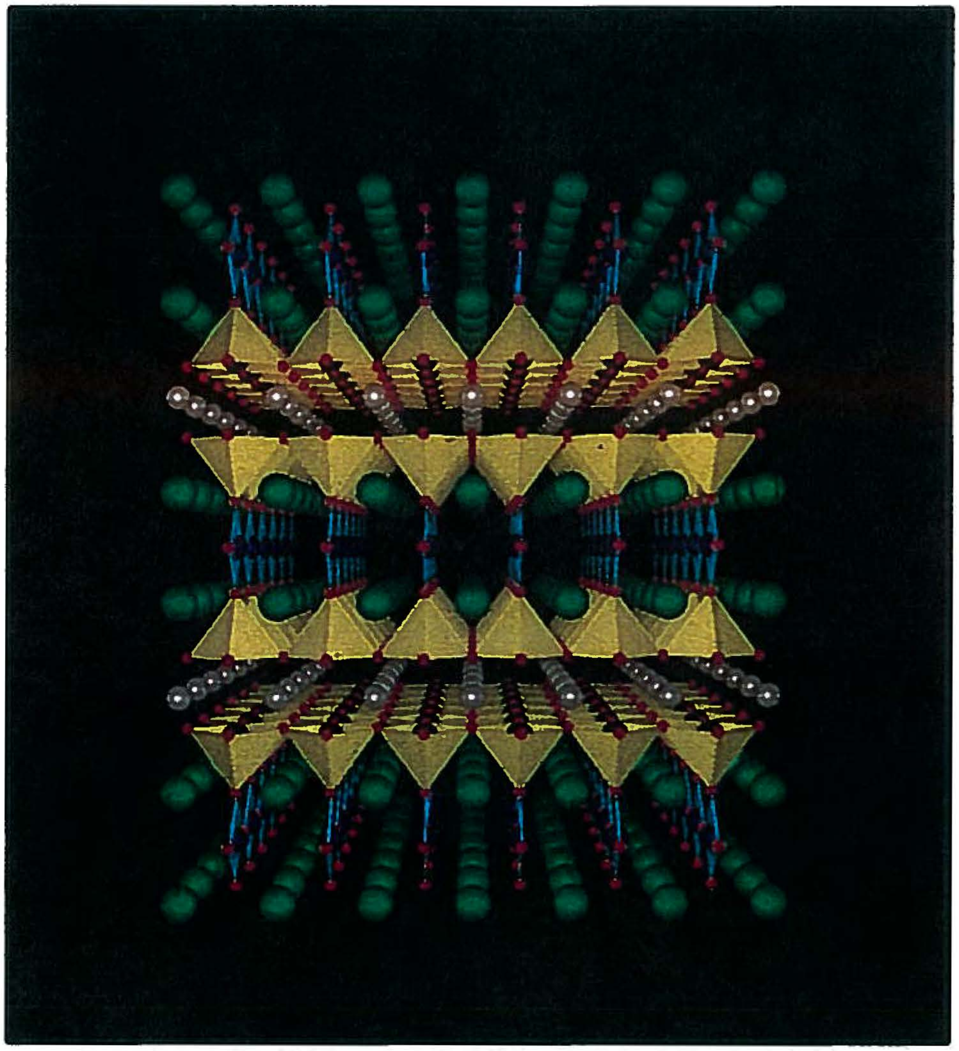
Two views are shown for the binding of a chemical carcinogen to a short stretch of DNA. The stick model on the left shows the overall structure of the DNA–carcinogen complex. The space filling model on the right shows the actual intimate binding of the carcinogen, hiding deep within the DNA. This chemical contact leads to mutations in the DNA code, and ultimately to a tumor.

Predicting New Superconductors

In 1988, the world was excited by the discovery that a particular yttrium–barium–copper–oxide compound superconducts at a temperature of 93° Kelvin, still very cold, but much warmer than any previous superconductor. This discovery sparked a worldwide effort to expand research to discover new superconducting materials. The economic benefit of a high–temperature superconductor is beyond calculation, portending the development of, for example, much more efficient power transmission and lightweight, powerful magnets to revolutionize the transportation business. Advanced computing is a central part of the arsenal of research tools which will be necessary to reach that payoff.

Despite these early successes, many questions remain before it is understood how some materials superconduct when very similar compounds do not. This understanding will be critical to predicting new superconductors, which might work at even higher temperatures, be less expensive, carry more current, or be more amenable to manufacturing processes. Increasing progress in all these areas is required before the impact of these new materials will be felt.

The solution of physical models requires intensive calculations to understand the material structure. High performance computing can shorten the discovery process by allowing the development of accurate simulations to point experimenters in the most promising directions. For example, most of the groups looking for new superconductors are trying various copper–oxide combinations. Researchers are using high performance computers to explore the possibility of various combinations of elements that may lead to new superconducting materials.



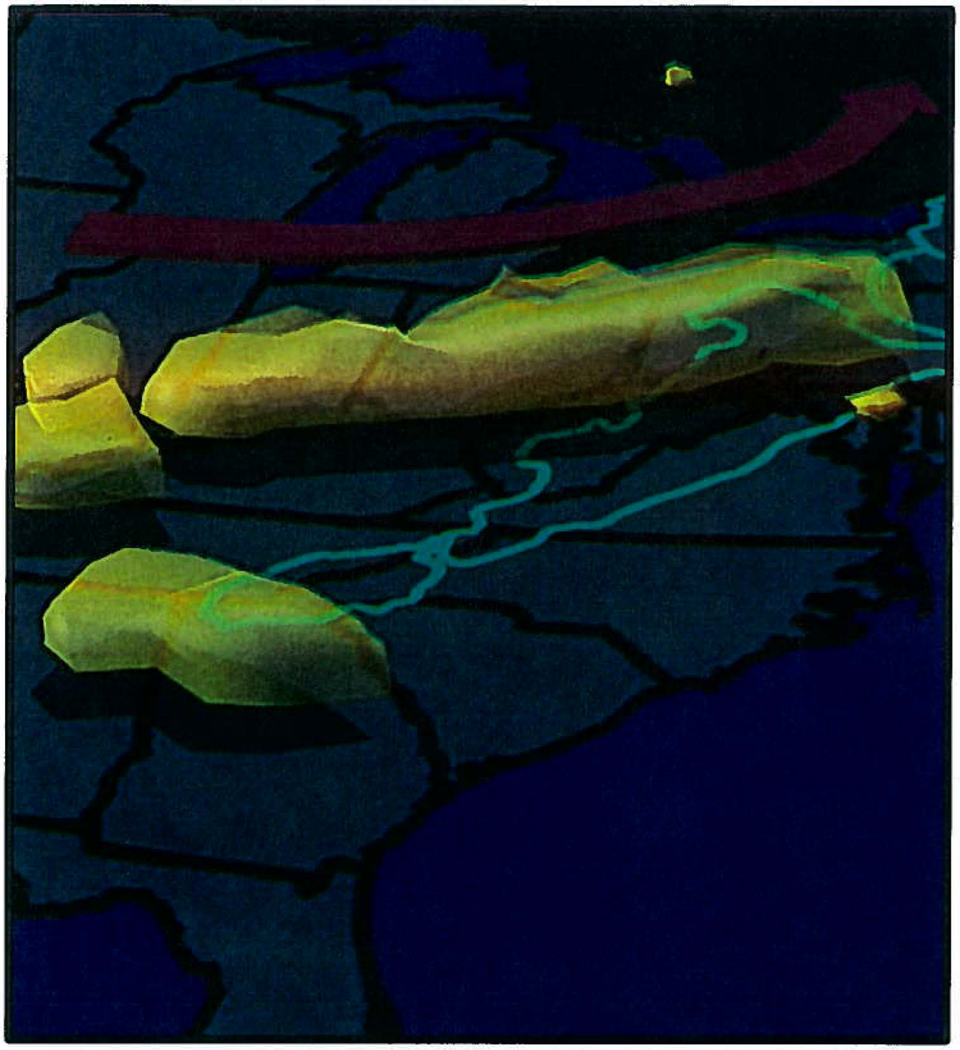
Researchers have used visualization techniques to investigate the structure of materials which are thought to be involved in the superconducting mechanism. In this graphic, barium cations (green) and yttrium cations (silver) are shown with their associated oxygen defects. The copper atoms (blue) and oxygen atoms (red) are found in two types of CuO_4 environments, one of which is depicted in yellow, the other in light blue.

Air Pollution

The ability of the atmosphere to absorb and to cleanse itself of pollutant contaminants was taken for granted until the 1950's when "killer fogs" in London, England and Donora, Pennsylvania caused the deaths of hundreds of people. Since then, technological advances in controlling source emissions have reversed the trend of steadily increasing pollution, even in the face of continued industrial growth. However, reduction strategies of individual pollutant types do not always produce the desired results. In fact, these simple solutions can even make air quality worse due to complex chemical interactions of the remaining airborne contaminants. Pollutants may travel long distances from industrial centers to sensitive areas where they are deposited in transformed products such as ozone and acid rain. These complexities of pollutant transport and transformation are costly and difficult to study experimentally, therefore, numerical models of the atmosphere have been developed to assess the effects of man–made emissions on air quality.

The new Clean Air Act mandates the use of numerical models to demonstrate the effectiveness of proposed regulatory control strategies. The potential cost to society to implement these proposed controls is estimated to reach tens of billions of dollars. Current models have been useful in evaluating alternative control strategies, but do not yet have the capability to produce optimum solutions. Present computing limitations on existing supercomputers force simplifications in the scientific descriptions of chemical and physical processes, and slow examination of alternatives. Control strategies for each pollutant are often determined independently with little evaluation of multiple pollutant interactions. Remedial solutions determined for a particular scale of space and time are not readily extendable to other scales of pollutant dispersal.

High performance computing will enable multi–scale numerical explorations with cross–pollutant interactions to be performed in a timely manner so as to be useful to legislated control and prevention requirements. Advanced computer designs and software methods will also enable cost optimization of pollutant control tradeoffs. Improved visualization techniques will enhance the interpretation and evaluation of massive amounts of environmental measurement and computer simulation data. High performance computing models will lead to a better understanding of the actions needed to minimize pollutant damage to materials and environmental damage to crops while making our air safer to breathe for future generations.



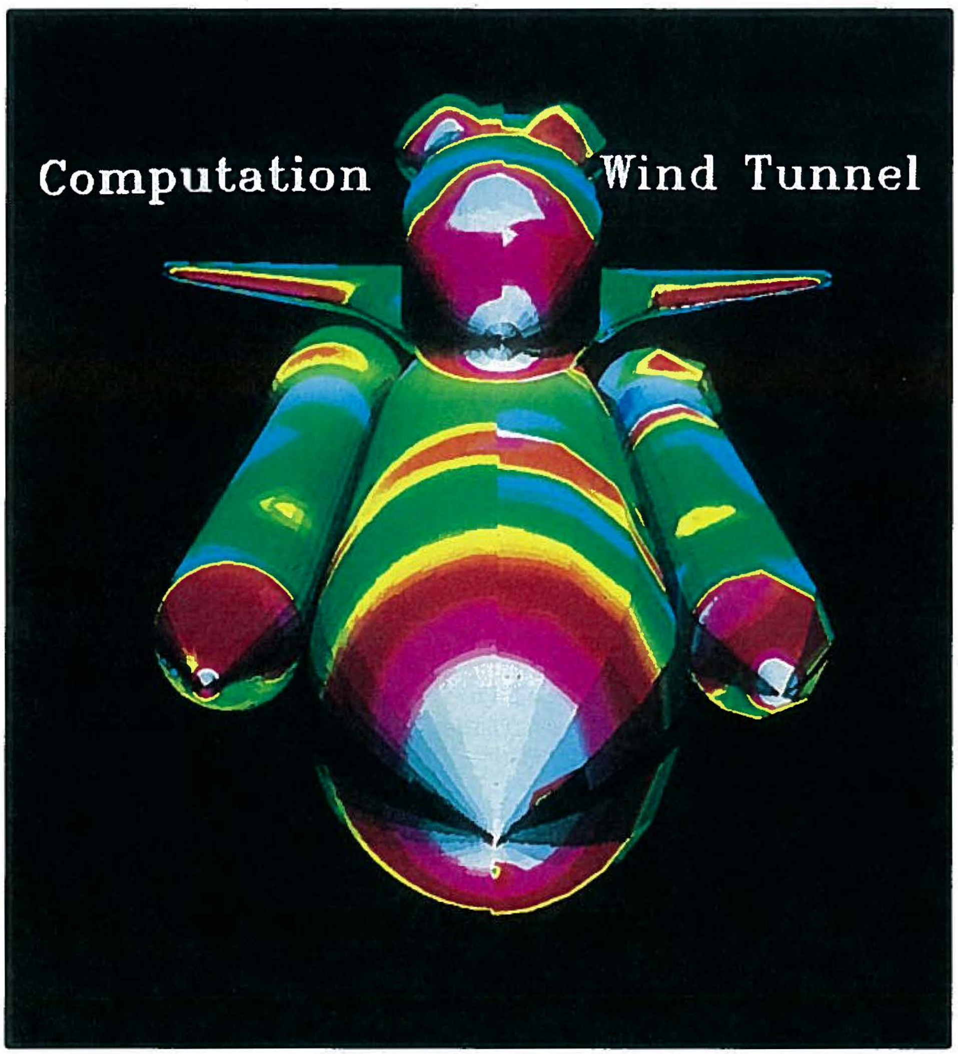
Using a supercomputer, atmospheric researchers have simulated the transport, chemical transformation, and surface deposition of sulfur compounds responsible for acid rain. Visualization specialists have depicted the movement of sulfur compounds from major sources in the Ohio river valley to sensitive lakes in the Adirondacks. The yellow cloud represents high concentrations of sulfur compounds. Several sensitive aquatic regions are outlined in green and a typical wind flow pattern is presented in red.

Aerospace Vehicle Design

Being able to predict the aerodynamic characteristics of in–flight vehicles is important for designers. Reproducing such flight regimes in wind tunnels is time consuming, costly, and in some cases impossible. Computational aerodynamics simulations are less costly and much faster than complex wind tunnel tests and are able to simulate many inaccessible flight regimes. These capabilities will be particularly important in the design of supersonic and hypersonic aircraft to serve international markets.

Computational aerosciences directly contributes to maintaining U.S. preeminence in aerospace science and engineering disciplines. The computational technology developed in such computational aerodynamics problems will directly transfer to the U.S. aerospace industry and aircraft manufacturers. Other potential beneficiaries are in diverse fields where fluid dynamics is an important design aspect such as automobile manufacture, ship design, and medicine (e.g., heart/cardiovascular flow simulation).

Massively parallel computing systems and advanced parallel software technology and algorithms will enable the development and validation of multidisciplinary, coupled models. These models will allow the numerical simulation and design optimization of complete aerospace vehicle systems throughout the flight envelope.



Shown here is a comparison between supercomputer simulation data and actual wind tunnel model data for the pressure distribution along the integrated space shuttle, solid rocket boosters and external tank flying at Mach 1.55. Note the excellent agreement between the two.

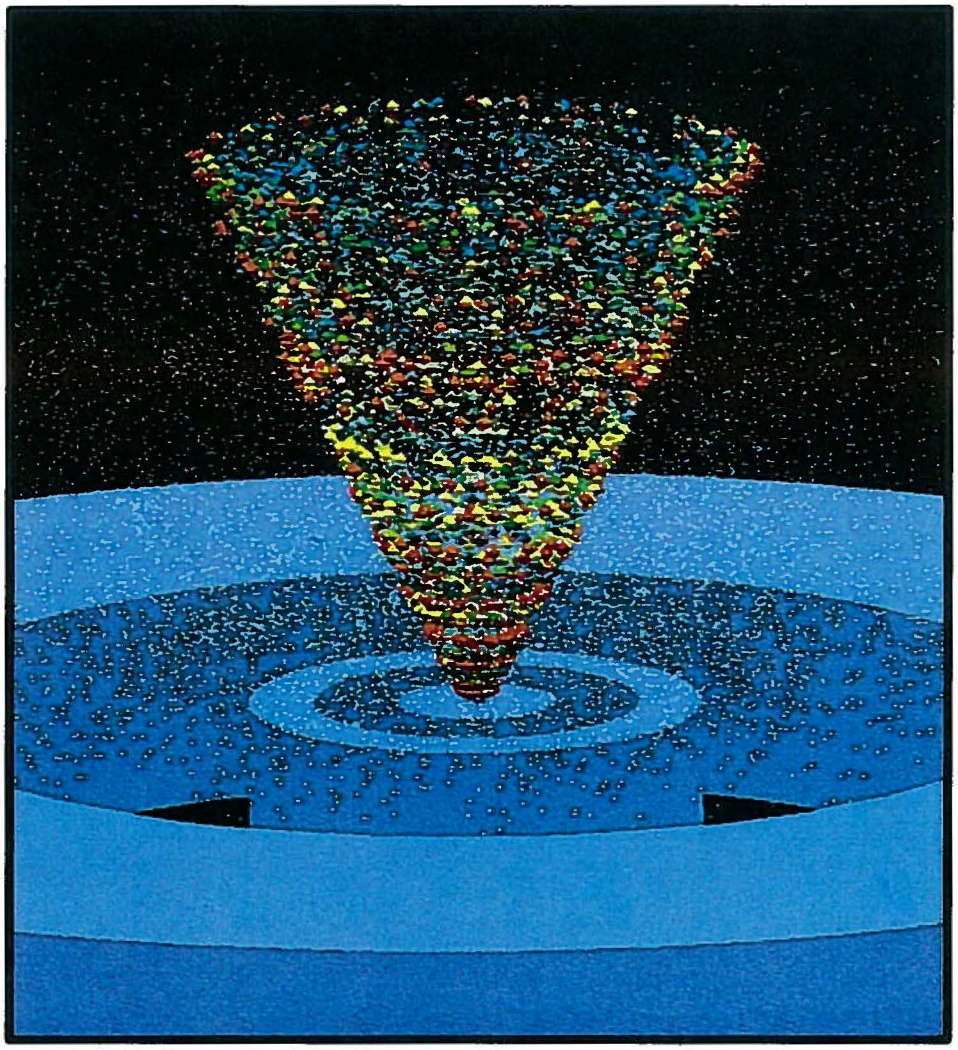
Energy Conservation and Turbulent Combustion

For the foreseeable future, 90% of the energy needs of the United States will be met by the combustion of fossil fuels. Two of the largest uses are for electrical power generation and in automobiles. Computational models offer a means of improving the design and efficiency of internal combustion engines.

Automobile engines are most efficient when run at high temperatures, but increased temperatures also lead to increased nitrogen oxide emissions. The burning of alternative fuels such as methanol is complicated by the emission of carcinogens. The effects of the emissions are influenced by local climatic conditions, making it necessary to consider the total system of fuel, engine, and atmosphere when seeking better designs. Our environment is too delicate and cannot be used as a testbed, so the atmospheric effects must be simulated.

A full three dimensional engine design code has been developed and implemented on a supercomputer. The code is designed to handle the most complex engines, such as the stratified charge and the two–stroke engines. The code represents an approximation to reality, because not all of the physical phenomena are well understood. Moreover, even if they were, the limited capabilities of existing computers would not allow this detailed information to be included.

For example, over 400 chemical processes of hydrocarbon and nitrogen chemistry are known to occur in internal combustion engines, yet only ten or less of the most significant reactions are used in simulations, in order to allow the calculation to run in a few hours on today's large supercomputers. Since the 400 hydrocarbon–nitrogen reactions are known, the real problem could be addressed with better algorithms running on a machine 10,000 times more powerful, a teraflops machine. This computational technology is needed by many private industrial engine firms as well as universities and government laboratories.



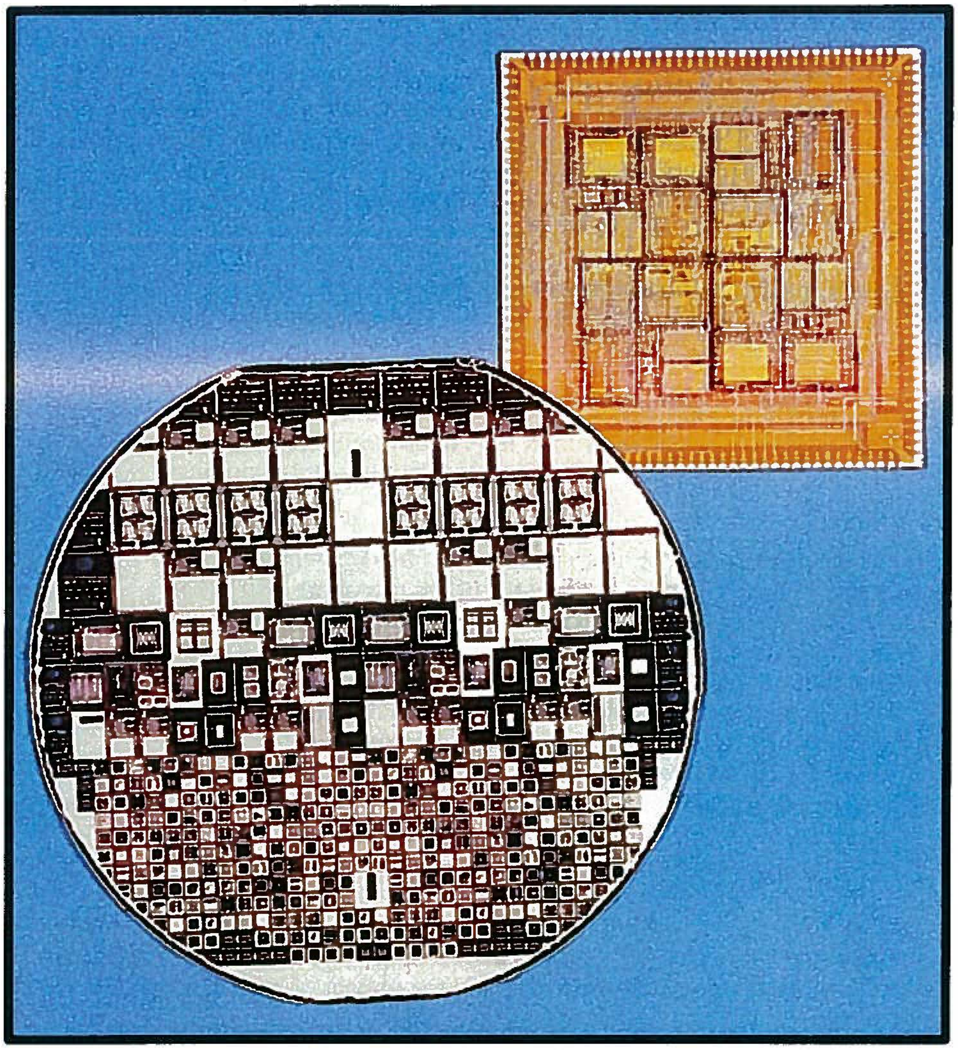
Researchers have used high performance computers to accomplish the numerical simulation of the properties of a conical fuel jet. The colors of the particles within the jet indicate the droplet size. The smallest particles are blue, intermediate size particles are shown in yellow, while the particles with the largest diameters are depicted in red. The smaller light blue particles around the jet represent the swirling air surrounding the spray.

Microsystems Design and Packaging

Since the invention of the integrated circuit in 1958, the number of transistors fabricated on a microchip has doubled every two years, providing a medium to incorporate ever–increasing complex electronic design into chips, components, and packages. By analogy, the complexity of detail incorporated in a single integrated circuit 1 cm square is equivalent to representing the map features, at a city block level, of the entire Eastern United States. A similar analogy for a 5 cm square module densely packed with a collection of chips would be the equivalent of a map of North and South America. Determining the interconnecting paths, selecting the right modules, testing the interfaces, and choosing the mix of technologies are part of the design process to build the scalable components for workstations to supercomputers.

In the computing world, scalable architectures based on the 1–2 million transistor custom structures of today will evolve in the mid 1990's to system approaches exploiting 10–million–transistor chips, standard component parts, and special interface electronics, all combined in optimally designed modules adhering to standard interfaces. System clock speeds will continue to improve, and the diversity of microchip technologies combined in a single module will allow unprecedented flexibility for designers.

Managing this complexity explosion would be overwhelming without the use of computationally based approaches that enable teams of designers to systematically reduce the time to develop such systems. Today, computational tools enable complex microcircuits to be developed on first pass, at the same time that packaging technologies such as multichip modules can be demonstrated. High performance computing applied to the technology design process will enable the exploration of design alternatives and rapid exploitation of new technologies.



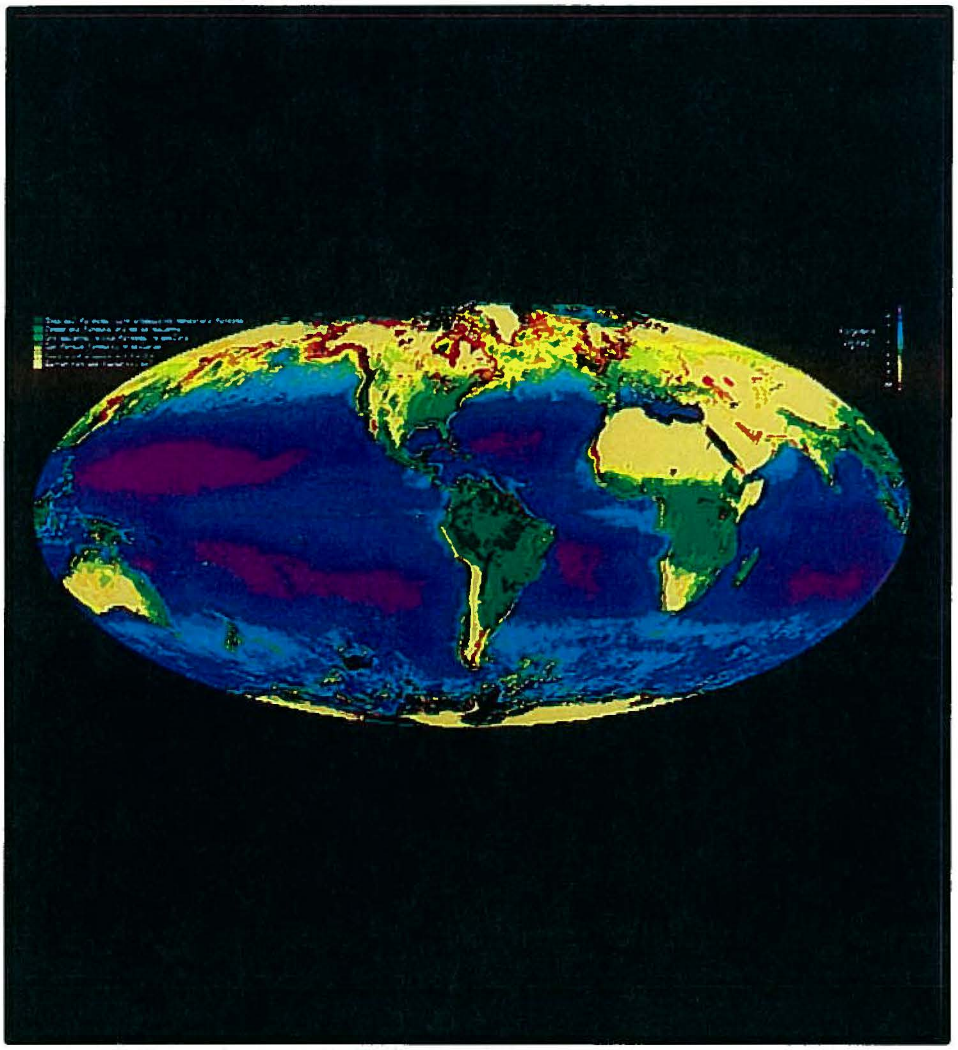
The figure shows both current and future components of the innovation design cycle. In the lower left of the photograph is a multiproject wafer using an industry/academia/government supported prototyping capability. The multiproject wafer shown is 4 inches in diameter, shares resources with 82 projects, and represents approximately 200 million transistors distributed over the wafer. Proven experimental prototypes are used as components of larger modules, interconnected with advanced technology. In the upper right hand corner is an example of an advanced interconnect module of 36 diverse microchips, packed together 1.5 inches on a side. Advanced computationally intensive design tools are essential to realize these high performance components.

Earth's Biosphere

The Earth's biosphere is a complex physical system. There are a multitude of phenomena which can change the state of the biosphere on a local, regional, and global scale. In order to predict the directions and consequences of changes in the state and condition of the biosphere, detailed scientific models are required, which in turn are constructed from massive amounts of experimental and computational data.

Experimental information is derived from satellite and ground based sensors. By the late 1990's, the sensors will have the resolution required to provide data in support of much more accurate and informed decision making on issues such as pollution and global warming. Because the sensors will generate terabytes of data each day, which will be combined with local data sets on the ecosystem, major improvements in the capability for collecting, analyzing, distributing, and archiving data are necessary.

The effort of constructing valid scientific models which describe the dynamics and underlying processes of the biosphere will involve interdisciplinary teams of experts from the geophysical, life, physical, computer, and computational sciences. They will work together to construct computational models which will validate our empirical understanding of the biosphere, and help predict how worldwide activities affect the global ecosystem. Computer and computational scientists will develop the advanced software technology and algorithms for handling massive amounts of data and working with high resolution, coupled models of the Earth's atmospheric–biospheric–oceanic interactions. Efforts in software development and experimentation will predict how local current conditions may impact future global conditions allowing the linking of ecosystem models to global climate models. The result of this collaborative effort will be a much deeper understanding of our environment and the impact of human activities upon it.



A global view of the Earth's ocean chlorophyll and land vegetation is depicted here. The image was derived from accumulated satellite data. Generating this image required over 2 terabytes (10^{12} bytes) of raw data. Data such as these are the baseline "snapshot" of our current biosphere and are vital to understanding short and long term component interactions of the Earth's biosphere for such purposes as crop productivity improvement, fishing and weather prediction.

High Speed Networks

The development of more powerful computers feeds both the demand for, as well as the growth of, more powerful data communications capability. As computing technology progresses, greater demands are placed on network performance as researchers conceive of new tasks and modalities of use that require even higher performance.

Current developments in large scale scientific computing are leading to truly distributed computing, allowing a given job to be executed on several different machines communicating partial results among themselves, sharing in different facets of calculations, and jointly assembling a final result for output. Such intermachine communication is inherently capable of taking place at speeds that stress local area network technology and are a hundred or more times faster than are possible on today's long haul networks. As an example, some of the nation's leading radiology departments have committed to digital transfer of radiology images on broadband local networks operating at speeds of 100 to 1000 million bits per second.

In order for the NREN to support these and other demanding applications, not yet contemplated, a substantial directed research and development effort is needed in the areas of protocols (the formal structure of inter–computer communications), high speed computer interfaces for computers, and network equipment, such as switches. Multi–gigabit networks represent a change in kind, not just in degree, from today's networks. For example, consider that in a coast–to–coast communication at three gigabits per second there are at any instant "in flight" nearly nine megabytes of data, which is more than the memory of most personal computers and workstations.

Some gigabit research has already begun, and several experimental facilities have been established in a productive collaboration of academic, industrial, and governmental organizations, but HPCC support will be needed to carry the research program to the stage that commercial providers can use the technology to install and operate a multi–gigabit NREN.

Education Using the NREN

Computers increasingly fill important niches in all phases of the learning process, providing flexible instruments for interactive instruction and student based learning experiences. Their use in education both provides the skills needed to function in our increasingly technology–intensive world, and aids teaching and learning of all science and engineering topics. The development of the National Research and Education Network will accelerate and transfer the technology of computer communications to the needs of educators and students. The result will be to empower them to share resources and ideas on a national scale.

In a recent project, classes in many locations chose a day to measure the length of the sun's shadow from a vertical measuring stick on the school grounds at 12 noon. Each class consulted maps and geography books to find its school's latitude, and sent the results to a shared database located in the U.S. and Canada. All schools then received the database of measurements from around the world, and each class used the complete database to calculate the curvature of the earth, and from that, the earth's diameter. A normally dry recitation of facts became an engaging problem solving exercise because the students themselves derived the answer from their shared measurements. Along the way they learned geography, geometry, statistics, and how to collect and share data over computer networks.

This project, implemented by the use of the network, is a learning laboratory without walls, similar to the way research scientists take advantage of high speed digital networks to conduct shared research that is "distance independent." The HPCC Program will face the challenge of "scaling–up" today's Internet, making it "user–friendly" and improving its services so that it is readily accessible to all U.S. educational institutions.

Glossary

ASTA
: Advanced Software Technology and Algorithms

Bit
: Acronym for binary digit

BRHR
: Basic Research and Human Resources

Byte
: A group of adjacent binary digits operated upon as a unit (usually connotes a group of eight bits)

CAD
: Acronym for "computer aided design".

Computer Engineering
: The creative application of engineering principles and methods to the design and development of hardware and software systems.

Computer Science
: The systematic study of computing systems and computation. The body of knowledge resulting from this discipline contains theories for understanding computing systems and methods; design methodology, algorithms, and tools; methods for the testing of concepts; methods of analysis and verification; and knowledge representation and implementation.

Computational Science and Engineering
: The systematic application of computing systems and computational solution techniques to mathematical models formulated to describe and simulate phenomena of scientific and engineering interest.

flops
: Acronym for "floating point operations per second". The term "floating point" refers to that format of numbers which is most commonly used for scientific calculation. Flops is used as a measure of a computing system's speed of performing basic arithmetic operations such as adding, subtracting, multiplying, or dividing two numbers.

Giga–
: 10^9 or billions of ... (e.g.: gigabits)

Grand Challenge
: A Grand Challenge is a fundamental problem in science and engineering, with broad economic and scientific impact, whose solution could be advanced by applying high performance computing techniques and resources.

HPCC

High Performance Computing and Communications

HPCS

High Performance Computing Systems

High Performance Computing

High performance computing encompasses advanced computing, communications, and information technologies, including scientific workstations, supercomputer systems, high speed networks, special purpose and experimental systems, the new generation of large scale parallel systems, and applications and systems software with all components well integrated and linked over a high speed national network.

Mega–

10^6 or millions of ... (e.g.: megaflops)

Network

Computer communications technologies that link multiple computers to share information and resources across geographically dispersed locations.

NREN

National Research and Education Network

ops

Acronym for "operations per second". Ops is used as a rating of the speed of computer systems and components. In this report ops is generally taken to mean the usual integer or floating point operations depending on what functional units are included in a particular system configuration.

Parallel Processing

Simultaneous processing by more than one processing unit on a single application.

Peta–

10^{15} or thousands of trillions of ... (e.g.: petabytes)

Supercomputer

At any given time, that class of general–purpose computers that are both faster than their commercial competitors and have sufficient central memory to store the problem sets for which they are designed. Computer memory, throughput, computational rates, and other related computer capabilities contribute to performance. Consequently, a quantitative measure of computer power in large–scale scientific processing does not exist and a precise definition of supercomputers is difficult to formulate.

Tera–

10^{12} or trillions of ... (e.g.: teraops)